С. УИЛКС

Математическая статистика

С. УИЛКС

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

ПЕРЕВОД С АНГЛИЙСКОГО
А. М. КАГАНА, Л. А. ХАЛФИНА, О. В. ШАЛАЕВСКОГО

ПОД РЕДАКЦИЕЙ
Ю. В. ЛИННИКА

ИЗДАТЕЛЬСТВО «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1967

517.8
У 36
УДК 519.24

С. Уилкс

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

М., 1967 г., 632 стр. с илл.

Редактор *Н. М. Митрофанова.*

Техн. редактор *С. Я. Шкляр.*

Корректоры *М. Л. Липелис* и *В. П. Сорокина.*

Сдано в набор 24/II 1967 г. Подписано к печати 27/VII 1967 г. Бумага $60 \times 90\ {}^1/_{16}$. Физ. печ. л. 39,5. Условн. печ. л. 39,5. Уч.-изд. л. 42,74. Тираж 19000 экз. Цена книги 3 р. 13 к. Заказ № 871.

Издательство «Наука».
Главная редакция
физико-математической литературы.
Москва, В-71, Ленинский проспект, 15.

Ордена Трудового Красного Знамени Ленинградская типография № 1 «Печатный Двор» имени А. М. Горького Главполиграфпрома Комитета по печати при Совете Министров СССР, г. Ленинград, Гатчинская ул., 26.

$\frac{2\text{-}2\text{-}3}{108\text{-}66}$

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Цель этой книги — познакомить с математической статистикой читателя, обладающего хорошей математической подготовкой. Наличие предварительных знаний у читателя в области теории вероятностей или математической статистики, вообще говоря, не предполагается, хотя знакомство с хорошими университетскими курсами в этих областях математики было бы полезным. Книга подготовлена в основном на материале, который я читал, все время его обновляя и пересматривая, аспирантам Принстонского университета в годы второй мировой войны. Первоначальный вариант, содержавший часть материала книги, был в 1943 г. издан в литографированном виде издательством Принстонского университета под названием «Математическая статистика».

За последнюю четверть века математическая статистика и ее приложения развивались таким стремительным темпом, что поток новых материалов вряд ли может быть целиком охвачен кем-либо в одиночку. Хотя наибольшее число результатов научных исследований в этой области публиковалось в специальных журналах (их около полдюжины), довольно значительная часть интересных статей публиковалась и продолжает публиковаться в других различных научных журналах.

Я не пытался дать исчерпывающего и всестороннего изложения всех основных результатов, которые публиковались в такого рода изданиях. Напротив, я отобрал фундаментальные результаты математической статистики, конечно, в связи с моими склонностями и предубеждениями, стараясь дать при этом общее и систематическое изложение, на базе современной теории вероятностей, классических; а также наиболее интересных новых результатов математической статистики. Поэтому, неизбежно, что изложение некоторых разделов математической статистики покажется энтузиастам и специалистам в этих областях недостаточно глубоким, а некоторые разделы математической статистики — незаслуженно обойденными вообще. Однако читатель, который заинтересуется вопросами, слабо освещенными в тексте книги или даже лишь упомянутыми, сможет найти ссылки для дальнейшего изучения этих вопросов. Я действительно надеюсь, что математически квалифицированный читатель, не обладавший предварительными знаниями в области математической статистики, прочитав эту книгу и заинтересовавшись дальнейшим систематическим изучением некоторых вопросов, сможет это сделать, руководствуясь, во всяком случае сначала, литературой, ссылки на которую помещены как в тексте книги, так и отдельно в библиографическом указателе в конце книги.

В книге имеется более четырехсот задач, размещенных по главам. Они дают возможность читателю лучше понять сущность проблем математической статистики. Многие из этих задач служат для того, чтобы в краткой форме осветить дополнительные разделы математической статистики, которые не могли быть подробно изложены в основном тексте книги из-за нехватки места.

Некоторые читатели будут удивлены, почему в книге не очень тесно переплетаются между собой изложения собственно математических результатов и методики статистики, которая основана на этих результатах. Обсуждение этой взаимосвязи содержится в книге умышленно в минимальном размере. Рассмотрение методики статистики и ее приложений, конечно, не менее важно, чем изложение лежащих в их основе математических результатов. Однако опыт показывает, что эти оба аспекта статистики, а именно математическая теория и методика статистики, наиболее эффективно сочетаются в оригинальных статьях, монографиях и руководствах, посвященных специальным областям. Я убежден, что в книгах по математической статистике, таких как эта, в основном содержащих общие результаты, весьма неразумно уделять одинаковое внимание обоим указанным аспектам математической статистики. Гораздо более разумный путь — тщательное изложение основ и математических теорий, на которых базируется широкий круг проблем статистики, с должным обсуждением и примерами, которые разъясняют основные концепции, как это я пытался сделать в данной книге. Я надеюсь что этот подход приведет к более полному пониманию и методики статистики, хотя я не мог ей уделить большое внимание.

Современная математическая статистика сильно связана с теорией вероятностей. Поэтому была сделана попытка дать цельное изложение оснований теории вероятностей без действительного их построения. Последнее было бы задачей, выходящей за рамки этой книги, тем более, что существуют блестящие руководства по многим вопросам теории вероятностей, затрагиваемых в книге, по которым читатель сможет углубить изучение этих разделов теории вероятностей.

В книге такого типа, как эта, в которой излагается широкий круг вопросов, введение общих обозначений и терминологии — не легкая задача. Некоторые обозначения и термины покажутся незнакомыми для специалистов по математической статистике, но это искупается получаемой при этом самосогласованностью, а также необходимостью введения новых терминов в связи с теми разделами математической статистики, которые раньше не излагались систематически в литературе.

Первые наброски большинства глав этой книги, включающих исследование порядковых статистик и непараметрических задач, были написаны, когда я был фульбрайтовским стипендиатом в Кембридже весной и летом 1951 г. В это время я запланировал написать монографию по указанным вопросам. За полученную поддержку моих исследований и попыток написания монографии я весьма благодарен. В дальнейшем, однако, было решено включить эти материалы в настоящую, более обширную книгу по математической статистике. Этот гораздо более честолюбивый проект не мог бы быть предпринят без той поддержки, которую предоставило Управление морских исследований. Я пользуюсь возможностью выразить мою глубокую благодарность за эту поддержку.

Я извлек большую пользу из ценных советов и критики моих коллег и аспирантов в Принстоне при подготовке этой книги. Широкий круг замечаний и советов, которые были сделаны Д. В. Тьюки, и обсуждения с Ф. Д. Анскомбом — особенно полезны. Критика и предложения В. С. Вардарьяна, касающиеся основных концепций теории вероятностей, были также очень полезны. Автор благодарит Д. М. Брауна, Д. А. Фридмана, И. Гутмана, А. Т. Джеймса и Ф. М. Сэнда за прочтение большей части рукописи и за предложенные улучшения. Я благодарен Д. Р. Бриллинджеру, который прорешал более 350 задач, помещенных в книге. Я благодарен также Д. А. Хартигану, который проверил доказательства и устранил ошибки в вычислениях, которые иначе остались бы в окончательном тексте книги. Советы всех этих и других моих друзей и коллег были очень полезны, хотя часть из них не была использована. Я один ответствен за ошибки и неаккуратности в изложении, которые остались в книге, и буду очень признателен читателям, которые привлекут мое внимание к замеченным ими недостаткам. Наконец, я выражаю свою благодарность Ребеке Веркман и Эмилии Соренсон за труд по напечатанию и оформлению рукописи.

Ноябрь 1961 г. *Самуэль С. Уилкс*

ГЛАВА 1

ПРЕДВАРИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

1.1. Выборочные пространства и события

Математическая статистика основывается на теории вероятностей, точное описание и трактовка которой в свою очередь зависят от теории меры и интегрирования. Теоретико-множественное описание вероятности, которое будет использовано в этой книге, было сформулировано Колмогоровым (1933а). Для наших целей здесь будет достаточно сделать небольшое вступление, охватывающее лишь основные определения, понятия и важные детали аппарата вероятностной теории в форме, полезной для математической статистики. Читателю, желающему получить более обширные и подробные сведения из теории вероятностей, следует обратиться к книгам Дуба (1953), Гнеденко и Колмогорова (1949), Феллера (1957), Колмогорова (1933а), Леви (1925, 1937) и Лоэва (1955). Подобную информацию, относящуюся к теории меры, можно найти у Халмоша (1950) и Манроу (1953).

Приступая к изложению необходимых в этой книге основ теории вероятностей, мы должны сначала рассмотреть понятия выборочного пространства и события, описание событий, а также составление и разложение событий. Аппарат для обращения с этими понятиями доставляет теория множеств.

Обозначим через R множество элементов e, которые будут называться *выборочными точками* или *случайными точками*, или, более кратко, *точками*. Число выборочных точек может быть конечным или бесконечным. R называется *выборочным пространством* или *пространством исходов*. Мы будем употреблять первое из этих двух выражений. Выборочная точка может рассматриваться как возможный исход *эксперимента, пробы* или *операции*, произведенной при данной *савокупности условий*. Однако никакой попытки формально определить подобные понятия мы предпринимать не будем. Выборочное пространство R есть просто множество *всех возможных исходов*, которые могли бы реализоваться, когда при данной совокупности условий осуществится некоторая операция.

Примеры. Несколько пояснительных примеров могут помочь закрепить эти представления.

При однократном бросании монеты, где e означает любую ее сторону, оказывающуюся сверху, выборочное пространство R состоит из двух выборочных точек: герба и решетки. При двукратном бросании монеты, где e означает произвольную комбинацию сторон, выпадающих сверху, выборочное пространство R состоит из четырех выборочных точек, в сокращенной записи ГГ, ГР, РГ, РР. Таким образом, если монета действительно подброшена дважды, одна из этих четырех выборочных точек *реализуется*.

При выборе наудачу 13 карт из колоды, содержащей 52 игральные карты, выборочное пространство R состоит из $\binom{52}{13}$ выборочных точек, причем каждая выборочная точка e есть один из возможных наборов выбранных карт. Таким образом, если 13 карт действительно выбраны, одна из этих выборочных точек из R *реализуется*.

Если электрической лампочке дают непрерывно гореть, пока не истечет срок ее службы, и e означает возможную продолжительность этого срока, то e — положительное число и (идеализированное) выборочное пространство R состоит из всех положительных чисел. Если k электрическим лампочкам $B_1, \ldots, B_k$ дают непрерывно гореть, пока все они не выйдут из строя, то e можно взять в виде множества k положительных чисел $(x_1, \ldots, x_k)$, означающих возможные сроки службы лампочек $B_1, \ldots, B_k$, и (идеализированное) выборочное пространство R будет состоять из точек в k-мерном эвклидовом пространстве, все координаты которых положительны.

Следует отметить, что во всех приведенных примерах осуществляется некоторая операция и результат этой операции изображается при помощи выборочной точки e, которая в свою очередь принадлежит выборочному пространству R.

Выборочная точка e иногда называется *элементарным событием*, а множество E выборочных точек, являющееся некоторым подмножеством точек в R, — *событием*. Конечно, событие может состоять и из одной-единственной точки. Когда мы говорим, что *происходит* событие E, мы подразумеваем, что выборочная точка e (представляющая исход эксперимента, пробы или операции) содержится в E. Вообще события и классы событий представляют больший интерес, чем отдельная выборочная точка как таковая. Выразим эту мысль несколько более точно. Обычно мы в большей степени интересуемся вероятностями, связанными с событиями, чем вероятностями, связанными с отдельными выборочными точками. Теми событиями, которые имеют для нас значение, являются *измеримые события*, т. е. множества, обладающие соответствующими вероятностями. Понятие измеримости будет рассмотрено в § 1.4.

Говоря на языке теории множеств, мы будем интересоваться некоторыми классами подмножеств R и вероятностями (функциями множеств), которые должны быть определены на множествах, принадлежащих таким классам. Тем не менее множество точек в выборочном пространстве R также называется событием. Будет действительно *удобно использовать термины «множество» и «событие» как равнозначные.*

Прежде чем перейти к дальнейшему обсуждению этих классов, мы познакомимся с основными принципами алгебры множеств.

1.2. Определения и правила для составления и разложения событий

В этом параграфе каждое из множеств E, E_1, E_2, ..., о которых будет идти речь, представляет собой событие, состоящее из выборочных точек в выборочном пространстве R. Иногда об этих множествах говорят как о e-множествах. Конечная последовательность n множеств будет записываться в виде E_1, E_2, ..., E_n, а счетная последовательность — в виде E_1, E_2, ...

Если e есть элемент множества E, мы выражаем это посредством записи

$$e \in E. \tag{1.2.1}$$

Мы также говорим, что e *принадлежит* множеству E.

Если E_1 и E_2 — два множества, причем *каждая точка множества* E_1 содержится в E_2, то E_1, называется подмножеством E_2, и мы пишем

$$E_1 \subset E_2 \quad \text{или} \quad E_2 \supset E_1. \tag{1.2.2}$$

Заметим, что если E_1 состоит в точности из одной точки, например точки e, то $E_1 \subset E_2$ эквивалентно утверждению $e \in E_2$.

Два множества E_1 и E_2 *равны*, если $E_1 \subset E_2$ и $E_1 \supset E_2$. В этом случае мы пишем

$$E_1 = E_2. \tag{1.2.3}$$

Пустое (или *нулевое*) *множество* есть множество, не содержащее никаких точек. Если E не содержит ни одной точки, мы полагаем

$$E = \varnothing. \tag{1.2.4}$$

Нулевое множество $\varnothing$ является подмножеством каждого множества $E \subset R$.

Множество всех точек, содержащихся *как в* E_1, *так и в* E_2, называется *пересечением* или *произведением* E_1 и E_2 и будет обозначаться

$$E_1 \cap E_2. \tag{1.2.5}$$

Множества E_1 и E_2 *не пересекаются*, т. е. не содержат общих точек, если

$$E_1 \cap E_2 = \varnothing. \tag{1.2.6}$$

Пересечение последовательности множеств E_1, E_2, ... обозначается символом

$$\bigcap_{\alpha=1}^{n} E_\alpha, \tag{1.2.7}$$

если имеется всего n множеств, и символом

$$\bigcap_{\alpha=1}^{\infty} E_\alpha, \tag{1.2.8}$$

если последовательность счетная. В более общем смысле для произвольной совокупности множеств $\{E_\alpha : \alpha \in T\}$ пересечение множеств обозначается через $\bigcap_{\alpha \in T} E_\alpha$ или просто через $\bigcap_\alpha E_\alpha$, если *область изменения* T значений α не вызывает сомнений. Областью T может быть, конечно, не только множество целых чисел, но и множество элементов в более общем пространстве. Почти во всех примерах, с которыми мы встретимся в этой книге, T будет последовательностью (конечным или счетным множеством целых чисел). Если $E_\alpha \bigcap E_\beta = \varnothing$, $\alpha \neq \beta$, то говорят, что $E_1, \ldots, E_n$ *попарно не пересекаются.*

Множество всех точек, содержащихся по крайней мере в одном из множеств E_1 и E_2, называется *объединением* или *суммой* E_1 и E_2 и будет обозначаться *)

$$E_1 \bigcup E_2. \tag{1.2.9}$$

Отметим, что если E_1 — любое e-множество и $E_2 = \varnothing$, то $E_1 \bigcup E_2 = E_1$.

Объединение конечной последовательности множеств $E_1, \ldots, E_n$ обозначается символом

$$\bigcup_{\alpha=1}^{n} E_\alpha \tag{1.2.10}$$

и — если последовательность счетная — символом

$$\bigcup_{\alpha=1}^{\infty} E_\alpha. \tag{1.2.11}$$

Для произвольной совокупности множеств $\{E_\alpha : \alpha \in T\}$ мы можем писать $\bigcup_\alpha E_\alpha$.

Множество, состоящее из всех точек E_1, не содержащихся в E_2, называется *разностью* между E_1 и E_2 и записывается как

$$E_1 - E_2. \tag{1.2.12}$$

Если $E_1 \subset E_2$, то, очевидно, $E_1 - E_2$ — пустое множество. Если $E_1 \bigcap E_2 = \varnothing$, то $E_1 - E_2 = E_1$. Если $E_2 \subset E_1$, $E_1 - E_2$ называется *собственной разностью* E_1 и E_2. В этом случае иногда удобно говорить, что $E_1 - E_2$ получается *вычитанием* E_2 из E_1.

Собственная разность $E_1 - E_2$ также называется *дополнением* E_2 *относительно* E_1; мы пишем

$$E_1 - E_2 = \bar{E}_2(E_1). \tag{1.2.13}$$

В частном, но важном случае, когда E_1 совпадает со всем выборочным пространством R, $R - E_2$ называется *дополнением* к E_2. Очевидно,

$$R - E_2 = \bar{E}_2. \tag{1.2.14}$$

*) Объединение (сумма) $E_1 \bigcup E_2$ иногда записывается как $E_1 + E_2$, а пересечение (произведение) $E_1 \bigcap E_2$ — как $E_1 \cdot E_2$.

Вообще говоря, мы имеем

$$E_1 - E_2 = E_1 \bigcap \overline{E}_2. \qquad (1.2.15)$$

Чтобы наглядно проиллюстрировать понятия объединения, пересечения и разности множеств, мы воспользуемся диаграммой Венна.

Пусть множество внутренних точек прямоугольника, изображенного на рис. 1.1, представляет собой выборочное пространство R и пусть E_1 и E_2 — множества внутренних точек соответственно большого и малого кружков. Тогда *объединением* $E_1 \bigcup E_2$ является множество точек, заключенных внутри по крайней мере одного из кружков; *разностью* $E_1 - E_2$ ($= E_1 \bigcap \overline{E}_2$) является множество внутренних точек большого кружка, не содержащихся внутри малого; подобную интерпретацию допускает и множество $E_2 - E_1$ ($= \overline{E}_1 \bigcap E_2$); множество $\overline{E}_1 \bigcap \overline{E}_2 = R - (E_1 \bigcup E_2)$ есть множество внутренних точек прямоугольника, не содержащихся ни внутри большого, ни внутри малого кружка.

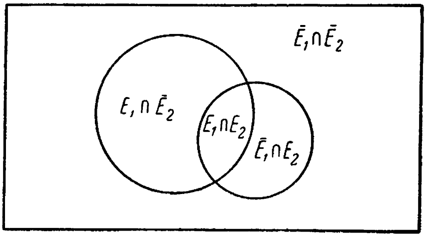

Рис. 1.1. Диаграмма Венна, иллюстрирующая четыре непересекающихся множества $E_1 \bigcap E_2$, $E_1 \bigcap \overline{E}_2$, $\overline{E}_1 \bigcap E_2$, $\overline{E}_1 \bigcap \overline{E}_2$, образованных из E_1 и E_2.

Эти понятия теории множеств допускают следующий перевод на язык событий *):

(I) $E_1 \subset E_2$ означает, что наступление события E_1 влечет наступление события E_2, т. е. выборочная точка e не может встретиться в E_1, не будучи элементом E_2.

(II) $E = \varnothing$, пустое множество, означает, что событие E произойти не может.

(III) $E = R$, все выборочное пространство, означает, что событие E обязательно должно произойти.

(IV) Пересечение $E_1 \bigcap E_2$ есть событие, состоящее в совместном наступлении событий E_1 и E_2.

(V) Объединение $E_1 \bigcup E_2$ есть событие, состоящее в наступлении по крайней мере одного из событий E_1 и E_2.

(VI) Разность $E_1 - E_2$ есть событие, обозначающее одновременно наступление E_1 и ненаступление E_2.

---

[1]) Следует заметить, что все определения и операции, начиная с (1.2.1) и кончая (1.2.15), приложимы не только к выборочным точкам и множествам выборочных точек (событиям), но и к элементам и множествам элементов любой природы. Элементами могут быть точечные множества, люди, автомобили и т. п. Например, если $\mathcal{F}$ есть совокупность множеств, содержащих множество E, мы пишем $E \in \mathcal{F}$; если каждое множество из совокупности $\mathcal{F}$ содержится в совокупности $\mathcal{S}$, мы пишем $\mathcal{F} \subset \mathcal{S}$, и т. д.

Читатель может легко проверить, что операции взятия объединения и произведения множеств *коммутативны*, т. е.

$$E_1 \bigcup E_2 = E_2 \bigcup E_1, \qquad E_1 \bigcap E_2 = E_2 \bigcap E_1, \tag{1.2.16}$$

ассоциативны, т. е.

$$\begin{aligned} (E_1 \bigcup E_2) \bigcup E_3 = E_1 \bigcup (E_2 \bigcup E_3), \\ (E_1 \bigcap E_2) \bigcap E_3 = E_1 \bigcap (E_2 \bigcap E_3), \end{aligned} \tag{1.2.17}$$

и *дистрибутивны*, т. е.

$$\begin{aligned} E_1 \bigcap (E_2 \bigcup E_3) = (E_1 \bigcap E_2) \bigcup (E_1 \bigcap E_3), \\ E_1 \bigcup (E_2 \bigcap E_3) = (E_1 \bigcup E_2) \bigcap (E_1 \bigcup E_3). \end{aligned} \tag{1.2.18}$$

Также ясно, что E_1 и собственная разность $E_2 - (E_1 \bigcap E_2)$ не пересекаются и

$$E_1 \bigcup E_2 = E_1 \bigcup [E_2 - (E_1 \bigcap E_2)]. \tag{1.2.19}$$

Подобным образом $E_1 \bigcap E_2$ и собственные разности $E_1 - (E_1 \bigcap E_2)$ и $E_2 - (E_1 \bigcap E_2)$ попарно не пересекаются, так что

$$E_1 \bigcup E_2 = (E_1 \bigcap E_2) \bigcup [E_1 - (E_1 \bigcap E_2)] \bigcup [E_2 - (E_1 \bigcap E_2)].$$

Необходимо отметить, что пересечение любых двух подмножеств E_1 и E_2 пространства R может быть выражено в терминах объединений и разностей множеств с помощью формулы

$$E_1 \bigcap E_2 = R - (\bar{E}_1 \bigcup \bar{E}_2). \tag{1.2.20}$$

Аналогичную формулу мы имеем для объединения:

$$E_1 \bigcup E_2 = R - (\bar{E}_1 \bigcap \bar{E}_2). \tag{1.2.21}$$

Справедливы также следующие соотношения:

$$(E_1 \bigcap E_2) \subset E_\alpha, \quad \alpha = 1,\ 2, \tag{1.2.22}$$

и

$$E_\alpha \subset (E_1 \bigcup E_2), \quad \alpha = 1,\ 2. \tag{1.2.23}$$

Для дальнейшего изложения будет полезно сформулировать обобщения некоторых из предшествующих результатов на конечные или бесконечные семейства множеств. Доказательства просты и предоставляются читателю.

1.2.1. *Для произвольного семейства множеств* $\{E_\alpha : \alpha \in T\}$

$$\bigcap_\alpha E_\alpha = R - \bigcup_\alpha \bar{E}_\alpha, \tag{1.2.20a}$$

$$\bigcup_\alpha E_\alpha = R - \bigcap_\alpha \bar{E}_\alpha, \tag{1.2.21a}$$

$$\bigcap_\alpha E_\alpha \subset E_\beta, \tag{1.2.22a}$$

$$\bigcup_\alpha E_\alpha \supset E_\beta. \tag{1.2.23a}$$

Обобщение формулы (1.2.19) можно сформулировать следующим образом.

1.2.2. *Если E_1, E_2, ... — последовательность (конечная или бесконечная) множеств, то E_1 и собственные разности*

$$E_2 - (E_2 \cap E_1),\ E_3 - [E_3 \cap (E_2 \cup E_1)],\ \ldots,$$
$$E_n - [E_n \cap (E_{n-1} \cup \ldots \cup E_1)],\ \ldots$$

попарно не пересекаются и объединение всех этих множеств есть $\bigcup\limits_{\alpha} E_\alpha$.

Теорема **1.2.2** дает простое правило для разложения объединения некоторой последовательности множеств на попарно не пересекающиеся множества. Необходимо отметить, что E_1 и разности $E_2 - E_1$, $E_3 - (E_1 \cup E_2)$, ..., $E_n - (E_1 \cup \ldots \cup E_{n-1})$, ... также попарно не пересекаются и их объединение есть $\bigcup\limits_{\alpha} E_\alpha$.

Если E_1, E_2, ... — счетная последовательность множеств в R, то множество E^*, состоящее из всех точек, каждая из которых принадлежит бесконечному числу множеств этой последовательности, называется *верхним пределом* последовательности, и мы пишем

$$E^* = \limsup_{\alpha} E_\alpha. \tag{1.2.24}$$

Аналогичным образом множество E_*, состоящее из всех точек, каждая из которых принадлежит всем множествам последовательности, за исключением конечного их числа, называется *нижним пределом* последовательности, и при этом используется запись

$$E_* = \liminf_{\alpha} E_\alpha. \tag{1.2.25}$$

Если $E^* = E_*$, мы будем говорить, что последовательность E_1, E_2, ... имеет предел, который мы обозначим как

$$\lim_{\alpha \to \infty} E_\alpha \quad \text{или} \quad \lim_{\alpha} E_\alpha. \tag{1.2.26}$$

Предел *расширяющейся* или *возрастающей* последовательности $E_1 \subset E_2 \subset \ldots$, очевидно, существует и равен объединению всех множеств $\{E_\alpha\}$, т. е.

$$\lim_{\alpha} E_\alpha = \bigcup_{\alpha} E_\alpha. \tag{1.2.27}$$

Существует предел и *стягивающейся* или *убывающей* последовательности, т. е. последовательности, для которой $E_1 \supset E_2 \supset \ldots$; он равен пересечению всех множеств

$$\lim_{\alpha} E_\alpha = \bigcap_{\alpha} E_\alpha. \tag{1.2.28}$$

В случае произвольной последовательности множеств $E_1, E_2, \ldots$ в R мы имеем

$$\limsup_{\alpha} E_\alpha = \bigcap_{n=1}^{\infty} \bigcup_{\alpha=n}^{\infty} E_\alpha, \quad \liminf_{\alpha} E_\alpha = \bigcup_{n=1}^{\infty} \bigcap_{\alpha=n}^{\infty} E_\alpha.$$

Расширяющиеся и стягивающиеся последовательности — несомненно самые важные последовательности в теории вероятностей и статистике. Если $E_1, E_2, \ldots$ — либо расширяющаяся, либо стягивающаяся последовательность множеств, удобно называть ее *монотонной последовательностью*. Таким образом, предел монотонной последовательности всегда существует. Если обозначить предел через E, то $E = \bigcup_\alpha E_\alpha$ или $E = \bigcap_\alpha E_\alpha$ в зависимости от того, является последовательность расширяющейся или стягивающейся. Несколько ниже мы увидим, что монотонные последовательности множеств играют фундаментальную роль в теории вероятностей.

1.3. Поля множеств

Как было указано ранее, нас в большей мере будут интересовать определенные классы подмножеств (событий) выборочного пространства R, чем отдельные выборочные точки в R. В частности, мы будем интересоваться полями множеств, т. е. классами множеств, удовлетворяющими некоторым условиям, которые должны быть сформулированы ниже.

Непустой класс $\mathcal{F}$ множеств в R называется *булевым полем* множеств, если он удовлетворяет следующим условиям:

А1. Если $E \in \mathcal{F}$, то $\bar{E} \in \mathcal{F}$.

А2. Если $E_1 \in \mathcal{F}$ и $E_2 \in \mathcal{F}$, то $E_1 \cup E_2 \in \mathcal{F}$.

Класс $\mathcal{F}$ множеств называется *борелевским полем* *) множеств, если выполняется следующее дополнительное условие:

А3. Если $E_1, E_2, \ldots$ — счетная последовательность множеств, принадлежащих $\mathcal{F}$, то $\bigcup_\alpha E_\alpha \in \mathcal{F}$.

В действительности при определении борелевского поля множеств требование **А2** является излишним, так как оно вытекает из **А3**. Полагая в **А2** $E_2 = \bar{E}_1$, мы имеем $E_1 \cup \bar{E}_1 = R \in \mathcal{F}$. Нужно также отметить, что если в **А1** в качестве E взять R, то $\bar{E}$ будет пустым множеством, и следовательно, пустое множество всегда содержится в поле множеств независимо от того, булево оно или борелевское.

В результате последовательного применения **А2** мы заключаем, что объединение любого конечного числа множеств, принадлежащих бу-

*) Булево поле множеств также называется *булевой алгеброй* или *конечно-аддитивным классом* множеств; в свою очередь борелевское поле множеств называется *σ-алгеброй* или *вполне аддитивным классом* множеств.

леву полю $\mathscr{F}$, также принадлежит $\mathscr{F}$. Можно проверить на основе **А1**, **А2** и (1.2.20), что пересечение конечной или, в случае борелевского поля, счетной последовательности множеств из $\mathscr{F}$ принадлежит $\mathscr{F}$.

Если R содержит конечное число N (различных) выборочных точек, то класс всех возможных событий конечен и является, очевидно, булевым полем. В самом деле, в этом случае булево поле $\mathscr{F}$ состоит из следующих 2^N множеств (событий): пустого множества, всех одноточечных множеств, всех двуточечных множеств, ..., и, наконец, самого R (N-точечного множества). Если мы обозначим через $\mathscr{F}_0$ класс, состоящий из всех N одноточечных множеств, то ясно, что любое конечное множество, принадлежащее булеву полю $\mathscr{F}$, можно получить, применяя **А1** и **А2** конечное число раз. Таким образом, булево поле $\mathscr{F}$ *может быть порождено начальным классом* $\mathscr{F}_0$. Следует подчеркнуть, что вообще существует много различных путей для выбора начального класса $\mathscr{F}_0$, способного породить булево поле множеств $\mathscr{F}$; класс одноточечных множеств, указанный выше, по-видимому, простейший.

Конечно, булево поле, порожденное конечным классом множеств, будет состоять только из конечного числа событий. Но если R содержит бесконечное или даже счетное число выборочных точек, то булево поле дает нам класс событий, недостаточный для рассмотрения многих представляющих интерес проблем.

Условие **А3** вводится для того, чтобы обеспечить возможность создания аппарата для исследования более широких классов событий, когда R содержит бесконечное число выборочных точек.

Предположим теперь, что R — выборочное пространство и $\mathscr{F}_0$ — произвольный начальный класс событий в R. Пусть $\mathscr{B}$ — некоторое борелевское поле множеств, содержащее все множества из $\mathscr{F}_0$. Очевидно, по крайней мере одно такое борелевское поле существует, так как в качестве $\mathscr{B}$ можно взять класс всех подмножеств R, который, разумеется, содержит $\mathscr{F}_0$.

Обозначим через $\mathscr{B}(\mathscr{F}_0)$ класс, состоящий из всех множеств, которые принадлежат каждому борелевскому полю, содержащему $\mathscr{F}_0$. $\mathscr{B}(\mathscr{F}_0)$ есть пересечение всех борелевских полей, содержащих класс $\mathscr{F}_0$. Множества из $\mathscr{B}(\mathscr{F}_0)$ удовлетворяют условиям **А1**, **А2** и **А3**, и, следовательно, $\mathscr{B}(\mathscr{F}_0)$ — борелевское поле.

Кроме того, $\mathscr{B}(\mathscr{F}_0)$ по определению содержится в любом борелевском поле, содержащем $\mathscr{F}_0$. Поэтому оно является единственным наименьшим борелевским полем, содержащим $\mathscr{F}_0$, и называется *борелевским полем, порожденным классом* $\mathscr{F}_0$.

Мы можем сформулировать следующее утверждение:

1.3.1. *Пусть* $\mathscr{F}_0$ — *произвольный класс множеств в* R. *Тогда существует единственное борелевское поле* $\mathscr{B}(\mathscr{F}_0)$, *обладающее тем свойством, что если* $\mathscr{B}$ — *любое борелевское поле, содержащее* $\mathscr{F}_0$, *то* $\mathscr{B}(\mathscr{F}_0) \subset \mathscr{B}$.

В частном, но важном случае, когда выборочное пространство R представляет собой вещественную ось R_1 (состоит из выборочных точек e, являющихся вещественными числами), а класс $\mathcal{F}_0$, который можно именовать *начальным классом*, есть класс всех полуоткрытых интервалов *) $(a, b]$, множества борелевского поля $\mathcal{B}(\mathcal{F}_0)$ называются *борелевскими множествами на вещественной прямой* и класс этих множеств $\mathcal{B}(\mathcal{F}_0)$ будет обозначаться через $\mathcal{B}_1$. Мы получим то же самое борелевское поле, если в качестве $\mathcal{F}_0$ возьмем класс интервалов вида (a, b), $[a, b)$, $(a, b]$ или вообще класс всех открытых множеств в R_1, или класс всех замкнутых множеств в R_1. Борелевские множества в R_1 образуют класс, *включающий в себя* все множества, которые могут быть получены в результате применения к конечному или счетному числу интервалов конечного или счетного числа операций объединения, пересечения, дополнения и вычитания.

В другом, более общем, случае, если выборочное пространство R является k-мерным евклидовым пространством R_k и если начальный класс $\mathcal{F}_0$ выступает как класс всех k-мерных полуоткрытых интервалов $(a; b]_k$, множества борелевского поля $\mathcal{B}(\mathcal{F}_0)$, порожденного $\mathcal{F}_0$ называются *борелевскими множествами* в R_k и само поле обозначается буквой $\mathcal{B}_k$. Имеются, конечно, другие подобные же способы выбрать начальный класс $\mathcal{F}_0$, порождающий борелевские множества в R_k. Например, $\mathcal{F}_0$ может быть классом всех интервалов вида $(a; b)_k$, $[a; b)_k$ или $[a; b]_k$. Следует также заметить, что класс борелевских множеств в R_k содержит каждое множество, которое может быть образовано с помощью конечного или счетного числа операций объединения, пересечения, разности или дополнения, примененных к интервалам любого упомянутого выше вида.

1.4. Вероятностная мера

В предыдущих параграфах мы занимались описанием событий и обучались обращению с ними. Теперь мы рассмотрим задачу, связанную с приписыванием этим событиям *вероятностей*, и установим правила оперирования с вероятностями. Общее понятие *вероятности события* есть абстракция, ведущая свое начало от идеи *относительной частоты*, с которой событие встречается в последовательности повторений эксперимента при «данной совокупности условий». Допустим, что E — некоторое событие из основного выборочного пространства R, причем каждый раз, когда «основной» эксперимент завершается, исход его соответствует некоторой выборочной точке e

*) Мы будем пользоваться обычным соглашением, по которому интервал $a < x \leq b$ обозначается символом $(a, b]$, а интервалы $a < x < b$, $a \leq x < b$ и $a \leq x \leq b$ — соответственно символами (a, b), $[a, b)$ и $[a, b]$. Если x — k-мерный вещественный вектор, т. е. если $(x_1, \ldots, x_k)$ — точка в k-мерном евклидовом пространстве, мы будем обозначать через $(a_1, \ldots, a_k; b_1, \ldots, b_k]$ или более кратко через $(a; b]_k$ k-мерный интервал $a_i < x_i \leq b_i$, $i = 1, \ldots, k$. Соответствующий смысл будут иметь и обозначения $(a; b)_k$, $[a; b)_k$ и $[a, b]_k$.

в R. При m повторениях эксперимента обозначим через m_E число случаев, когда получающаяся выборочная точка e принадлежит E, т. е. частоту, с которой наступает событие E. Относительная частота E в m опытах будет равна $\frac{m_E}{m}$. Ясно, что это отношение лежит в интервале $[0, 1]$; оно с необходимостью имеет значение 1, если $E=R$, и значение 0, если E — пустое множество. Если E_1 и E_2 — непересекающиеся события, частота, с которой наступает событие $E_1 \cup E_2$, равна $m_{E_1}+m_{E_2}$, а его относительная частота равна $\frac{m_{E_1}+m_{E_2}}{m}=\frac{m_{E_1}}{m}+\frac{m_{E_2}}{m}$. Таким образом, относительная частота объединения двух непересекающихся событий равна сумме относительных частот этих событий. Подобная же связь имеется между относительными частотами любого конечного числа попарно не пересекающихся событий и относительной частотой их объединения.

Если мы теперь представим себе относительную частоту события E в (очевидно, неосуществимом) неограниченно длинном ряду повторений нашего «основного» эксперимента, мы будем недалеко от разумной интерпретации *вероятности события* E. Однако возникнут серьезные трудности, обсуждать которые мы не можем, если мы попытаемся основать теорию вероятностей на строгой формализации идеи относительной частоты и на предположении о том, что в неограниченно длинном ряду опытов имеют место свойства сходимости. Этот подход был рассмотрен Мизесом (1931). Более простая и, по-видимому, более глубокая формулировка, первоначально предложенная Колмогоровым (1933а), сводится, по существу, к предположению, что а) в качестве вероятностей событий в некотором начальном классе относительно простых событий могут быть выбраны числа, принадлежащие интервалу $[0, 1]$, и б) эти начальные вероятности вместе с правилами определения вероятностей более сложных событий дают возможность определить вероятности любого события из класса достаточно широкого, чтобы охватить события, представляющие интерес в исходной вероятностной задаче. В процессе действительного выбора вероятностей событий обычно руководствуются гипотетическими данными, основанными на прогнозировании относительных частот событий, которые должны иметь место в длинном ряду повторений «основного» эксперимента. В предложенной Колмогоровым формулировке важная роль отводится математической теории, которая начинается *после* выбора вероятностей. Можно, конечно, поинтересоваться, правильно ли выбраны вероятности в какой-либо конкретной ситуации. Формальная постановка этого вопроса приводит нас к одной из задач теории проверки статистических гипотез, которая будет рассматриваться в более поздних главах.

Формализуем высказанные идеи, приведя определение *вероятностной меры* или *распределения вероятностей* на некотором классе событий.

Назовем функцию множества P, заданную на всех множествах булева поля $\mathcal{F}$, *вероятностной мерой на этом булевом поле* $\mathcal{F}$, если удовлетворяются следующие три условия:

B1. *Число $P(E)$, поставленное в соответствие любому событию $E \in \mathcal{F}$ и называемое вероятностью события E, является вещественным и неотрицательным. Иногда вместо $P(E)$ пишут $P(e \in E)$.*

B2. *Если $E_1, E_2, \ldots$ — счетная последовательность попарно не пересекающихся множеств из $\mathcal{F}$, объединение которых также находится в $\mathcal{F}$, то*

$$P\left(\bigcup_{\alpha=1}^{\infty} E_\alpha\right) = \sum_{\alpha=1}^{\infty} P(E_\alpha).$$

B3. $P(R) = 1$.

Если P — функция множества, определенная для всех множеств в борелевском поле $\mathcal{F}$ и удовлетворяющая условиям **B1**, **B2** и **B3**, то P называется *вероятностной мерой на борелевском поле* $\mathcal{F}$. Разумеется, в этом случае $\bigcup_{\alpha=1}^{\infty} E_\alpha \in \mathcal{F}$ по определению борелевского поля.

Тройка $(R, \mathcal{F}, P)$ называется *вероятностным пространством*. Следует отметить, что **B2** имеет место для всех конечных наборов попарно не пересекающихся множеств $E_1, \ldots, E_n$, так как в качестве $E_{n+1}, E_{n+2}, \ldots$ могут быть взяты пустые множества.

Функцию множества P, удовлетворяющую **B1** (где опущено требование неотрицательности) и **B2**, часто называют *вполне аддитивной* функцией множества, если $P(E)$ конечно для каждого $E \in \mathcal{F}$.

Обычно мы интересуемся наименьшим борелевским полем $\mathcal{B}(\mathcal{F})$, порожденным некоторым булевым полем $\mathcal{F}$. Конечно, $\mathcal{F} \subset \mathcal{B}(\mathcal{F})$. Здесь мы можем сформулировать несколько простых, но вместе с тем и полезных теорем, касающихся вероятностей событий из $\mathcal{B}(\mathcal{F})$. Естественно, что эти теоремы автоматически будут выполняться для событий из $\mathcal{F}$, так как вероятностная мера на $\mathcal{B}(\mathcal{F})$ задает вероятности для всех множеств в $\mathcal{F}$.

1.4.1. *Если E — любое событие из $\mathcal{B}(\mathcal{F})$, то $P(E) + P(\bar{E}) = 1$.*

Из теоремы **1.4.1** и условия **B3** немедленно вытекает, что $P(\emptyset) = 0$. Непустое множество E, для которого $P(E) = 0$, иногда называется *множеством* (или *событием*) *нулевой вероятности*.

1.4.2. *Если E_1 и E_2 — события из $\mathcal{B}(\mathcal{F})$ такие, что $E_1 \supset E_2$, то $0 \leqslant P(E_1 - E_2) = P(E_1) - P(E_2)$ и $P(E_1) \geqslant P(E_2)$.*

1.4.3. *Если E_1 и E_2 — события из $\mathcal{B}(\mathcal{F})$, то $P(E_1 \cup E_2) = P(E_1) + P(E_2 - E_1 \cap E_2)$ и, кроме того, $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.*

1.4.4. *Вообще, если $E_1, \ldots, E_n$ — n множеств, принадлежащих $\mathcal{B}(\mathcal{F})$, то*

$$P\left(\bigcup_{\alpha=1}^{n} E_\alpha\right) = P(E_1) + P(E_2 - E_2 \cap E_1) + \ldots$$
$$\ldots + P(E_n - E_n \cap [E_1 \cup \ldots \cup E_{n-1}]), \quad (1.4.1)$$

и, кроме того,

$$P\left(\bigcup_{\alpha=1}^{n} E_\alpha\right)=\sum_{\alpha=1}^{n} P(E_\alpha)-\sum_{\beta>\alpha=1}^{n} P(E_\alpha \cap E_\beta)+\ldots$$
$$\ldots+(-1)^{n-1} P(E_1 \cap \ldots \cap E_n). \qquad (1.4.1\text{a})$$

Доказательство теорем **1.4.1**, **1.4.2**, **1.4.3** и **1.4.4** предоставляется читателю в качестве упражнения. Если мы рассматриваем бесконечную последовательность множеств $E_1, E_2, \ldots$ из $\mathscr{B}(\mathscr{F})$, то (1.4.1) и (1.4.1a) принимают соответственно вид

$$P\left(\bigcup_{\alpha=1}^{\infty} E_\alpha\right)=P(E_1)+P(E_2-E_2 \cap E_1)+P(E_3-E_3 \cap [E_1 \cup E_2])+\ldots \qquad (1.4.1)'$$

и (предполагается сходимость сумм в правой части)

$$P\left(\bigcup_{\alpha=1}^{\infty} E_\alpha\right)=\sum_{\alpha=1}^{\infty} P(E_\alpha)-\sum_{\beta>\alpha=1}^{\infty} P(E_\alpha \cap E_\beta)+$$
$$+\sum_{\gamma>\beta>\alpha=1}^{\infty} P(E_\alpha \cap E_\beta \cap E_\gamma)-\ldots \qquad (1.4.1\text{a})'$$

Отметим, что (1.4.1) и (1.4.1a) суть частные случаи соответственно (1.4.1)' и (1.4.1a)', в чем легко убедиться, считая $E_{n+1}, E_{n+2}, \ldots$, пустыми множествами.

Следующая теорема имеет особое значение для главы 2.

1.4.5: *Если $E_1, E_2, \ldots$ — монотонная последовательность множеств, принадлежащих борелевскому полю $\mathscr{B}(\mathscr{F})$, то*

$$\lim_{\alpha\to\infty} P(E_\alpha)=P(\lim_{\alpha\to\infty} E_\alpha).$$

Чтобы доказать теорему **1.4.5**, мы рассмотрим сначала стягивающуюся последовательность $E_1 \supset E_2 \supset \ldots$ Тогда множества $E_1-E_2, \ldots, E_{n-1}-E_n, E_n, n=2, 3, \ldots$, попарно не пересекаются, и поэтому

$$E_1=(E_1-E_2) \cup (E_2-E_3) \cup \ldots \cup (E_{n-1}-E_n) \cup E_n=$$
$$=\left[\bigcup_{\alpha=1}^{n-1}(E_\alpha-E_{\alpha+1})\right] \cup E_n. \qquad (1.4.2)$$

Обращаясь к **B2**, мы имеем

$$P(E_1)=P\left(\bigcup_{\alpha=1}^{n-1}(E_\alpha-E_{\alpha+1})\right)+P(E_n), \qquad (1.4.3)$$

и после перехода здесь к пределу при $n\to\infty$ и новой ссылки на **B2** —

$$P(E_1)=\lim_{n\to\infty} P\left(\bigcup_{\alpha=1}^{n-1}(E_\alpha-E_{\alpha+1})\right)+\lim_{n\to\infty} P(E_n)=$$
$$=P\left(\lim_{n\to\infty}\bigcup_{\alpha=1}^{n-1}(E_\alpha-E_{\alpha+1})\right)+\lim_{n\to\infty} P(E_n). \qquad (1.4.4)$$

Но $\lim\limits_{n\to\infty}\bigcup\limits_{\alpha=1}^{n-1}(E_\alpha - E_{\alpha+1}) \equiv \lim\limits_{n\to\infty}(E_1 - E_n)$. Далее, $(E_1 - E_n)$, $n = 1, 2, \ldots$, — расширяющаяся последовательность, предел которой при $n \to \infty$ равен $E_1 - E$, где $E = \bigcap\limits_{\alpha=1}^{\infty} E_\alpha$. Так как E_1 и E принадлежат $\mathscr{B}(\mathscr{F})$, то $E_1 - E$ также принадлежит $\mathscr{B}(\mathscr{F})$ и мы получаем

$$P(E_1) = P(E_1 - E) + \lim_{\alpha\to\infty} P(E_\alpha). \tag{1.4.5}$$

На основании теоремы **1.4.2** $P(E_1 - E) = P(E_1) - P(E)$, что после подстановки в (1.4.5) дает $P(E) = \lim\limits_{\alpha\to\infty} P(E_\alpha)$, т. е.

$$P(\lim_{n\to\infty} E_n) = \lim_{n\to\infty} P(E_n), \tag{1.4.6}$$

тем самым завершая доказательство нашей теоремы в случае стягивающейся последовательности событий.

Если $E_1 \subset E_2 \subset \ldots$ — расширяющаяся последовательность событий, то последовательность, составленная из дополнений к этим событиям, $\bar{E}_1, \bar{E}_2, \ldots$ является стягивающейся и по только что доказанному

$$P(\lim_{n\to\infty} \bar{E}_n) = \lim_{n\to\infty} P(\bar{E}_n) \tag{1.4.7}$$

или

$$P(\lim_{n\to\infty}(R - E_n)) = \lim_{n\to\infty} P(R - E_n), \tag{1.4.8}$$

что можно переписать в виде

$$P(R - \lim_{n\to\infty} E_n) = \lim_{n\to\infty} P(R - E_n). \tag{1.4.9}$$

Воспользовавшись теоремой **1.4.2**, мы получаем соотношение (1.4.6), где, разумеется,

$$\lim_{n\to\infty} E_n = \bigcup_{\alpha=1}^{\infty} E_\alpha.$$

Следующая *теорема покрытия* будет использоваться в более поздних главах.

1.4.6. *Если $E, E_1, E_2, \ldots$ — события из $\mathscr{B}(\mathscr{F})$, причем $E \subset \bigcup\limits_{\alpha=1}^{\infty} E_\alpha$, то*

$$P(E) \leqslant \sum_{\alpha=1}^{\infty} P(E_\alpha).$$

Всякий раз, когда конечная или бесконечная последовательность множеств E, E_1, E_2 обладает тем свойством, что $E \subset \bigcup\limits_{\alpha=1}^{\infty} E_\alpha$, мы будем говорить, что $\bigcup\limits_{\alpha=1}^{\infty} E_\alpha$ есть *покрытие* множества E.

Чтобы установить справедливость утверждения **1.4.6**, достаточно заметить, что, поскольку $E \subset \bigcup_{\alpha=1}^{\infty} E_\alpha$,

$$E = E \bigcap \left(\bigcup_{\alpha=1}^{\infty} E_\alpha\right). \tag{1.4.10}$$

Применяя затем теорему **1.2.2**, мы находим

$$E = [E \bigcap E_1] \bigcup [E \bigcap (E_2 - E_2 \bigcap E_1)] \bigcup \ldots \tag{1.4.11}$$

Множества, стоящие в квадратных скобках, попарно не пересекаются и принадлежат $\mathscr{B}(\mathscr{F})$. Следовательно,

$$P(E) = P(E \bigcap E_1) + P(E \bigcap (E_2 - E_2 \bigcap E_1)) + \ldots, \tag{1.4.12}$$

но

$$E \bigcap E_1 \subset E_1, \quad E \bigcap (E_2 - E_2 \bigcap E_1) \subset E_2, \ \ldots$$

Поэтому на основании теоремы **1.4.2**

$$P(E) \leqslant P(E_1) + P(E_2) + \ldots, \tag{1.4.13}$$

и это заканчивает доказательство теоремы **1.4.6**.

Полагая $E_{n+1} = E_{n+2} = \ldots = \varnothing$, можно увидеть, что из теоремы **1.4.6** вытекает неравенство

$$P(E) \leqslant \sum_{\alpha=1}^{n} P(E_\alpha),$$

которое, конечно, также будет иметь место, если множества $E_1, \ldots$ $\ldots, E_n$ принадлежат $\mathscr{F}$.

Замечание. Если в случае выборочного пространства R, содержащего лишь конечное число выборочных точек, в качестве начального класса событий $\mathscr{F}_0$ мы возьмем все одноточечные события и зададим вероятности элементов из $\mathscr{F}_0$, то тем самым мы, очевидно, определим булеву вероятностную меру. Это означает, что мы можем вычислить вероятность любого события, находящегося в булевом поле $\mathscr{F}$, порожденном $\mathscr{F}_0$, исходя из вероятностей, приписанных множествам из $\mathscr{F}_0$.

1.5. Расширение вероятностной меры

(а) Единственность расширения вероятностной меры. Как мы видели в § 1.3, для любого начального булева поля $\mathscr{F}$ существует наименьшее борелевское поле $\mathscr{B}(\mathscr{F})$, содержащее $\mathscr{F}$. Предположим, мы имеем вероятностную меру, определенную на булевом поле $\mathscr{F}$. Возникает вопрос: нельзя ли подвергнуть вероятности, заданные на множествах из $\mathscr{F}$, некоторой операции, чтобы в результате получить вероятностную меру на $\mathscr{B}(\mathscr{F})$, сохраняющую без изменений вероятности всех множеств из $\mathcal{F}$? Ответ на этот вопрос положителен. Полученную вероятностную меру на $\mathcal{B}(\mathcal{F})$ можно рассматривать как расширение (или продолжение) вероятностной меры на $\mathcal{F}$.

Прежде чем мы укажем метод, с помощью которого осуществляется расширение меры, определенной на $\mathcal{F}$, установим следующую теорему единственности расширения.

1.5.1. *Пусть $\mathcal{F}$ — булево поле множеств и P_1 и P_2 — вероятностные меры, определенные на борелевском поле $\mathcal{B}(\mathcal{F})$. Тогда если $P_1(E)=P_2(E)$ при всех $E\in\mathcal{F}$, то $P_1(E)=P_2(E)$ при всех $E\in\mathcal{B}(\mathcal{F})$.*

Обозначим через $\mathcal{K}$ класс всех тех множеств E из $\mathcal{B}(\mathcal{F})$, для которых $P_1(E)=P_2(E)$. Легко проверить, что $\mathcal{K}$ есть борелевское поле, содержащее (наименьшее) борелевское поле $\mathcal{B}(\mathcal{F})$. Но так как $\mathcal{K}\subset\mathcal{B}(\mathcal{F})$, то $\mathcal{K}=\mathcal{B}(\mathcal{F})$ и, следовательно, $P_1(E)=P_2(E)$ при всех $E\in\mathcal{B}(\mathcal{F})$.

Теорема **1.5.1**, по существу, утверждает, что не может быть двух вероятностных мер, определенных на борелевском поле $\mathcal{B}(\mathcal{F})$, порожденном булевым полем $\mathcal{F}$, которые совпадают на множествах из $\mathcal{F}$, но не являются тождественно равными на множествах из $\mathcal{B}(\mathcal{F})-\mathcal{F}$. Таким образом, если, взяв какую-либо вероятностную меру на булевом поле $\mathcal{F}$, мы найдем способ *расширить* ее до вероятностной меры на борелевском поле $\mathcal{B}(\mathcal{F})$, то эта последняя мера, как следует из теоремы **1.5.1**, будет единственной.

(b) Расширение вероятностной меры. Расширение меры, о котором мы только что говорили, может быть осуществлено с помощью теории внешней меры, развитой Каратеодори (1927) [см. также Халмош (1950)]. Чтобы определить внешнюю меру, пригодную для наших целей, мы воспользуемся вероятностной мерой P, заданной на множествах булевского поля $\mathcal{F}$. Пусть E — произвольное множество из $\mathcal{B}(\mathcal{F})$ и $E_1, E_2, \ldots$ — последовательность множеств из $\mathcal{F}$, причем $E\subset\bigcup_{\alpha=1}^{\infty}E_\alpha$, т. е. $\bigcup_{\alpha=1}^{\infty}E_\alpha$ есть покрытие E. *Внешняя мера* множества E, обозначаемая через $P^*(E)$, определяется как точная нижняя граница значений сумм $\sum_{\alpha=1}^{\infty}P(E_\alpha)$ для всех возможных покрытий E, образованных множествами из $\mathcal{F}$,

$$P^*(E)=\inf\sum_{\alpha=1}^{\infty}P(E_\alpha). \tag{1.5.1}$$

Каково бы ни было булево поле $\mathcal{F}$, мы всегда можем взять $E_1=R$, $E_2=E_3=\ldots=\varnothing$, так что $\bigcup_{\alpha=1}^{\infty}E_\alpha=R$ будет являться покрытием любого множества E в R. Поэтому $P^*(E)$ существует для любого множества в R. Тем не менее в действительности мы будем интересоваться только значениями $P^*(E)$ на множествах из $\mathcal{B}(\mathcal{F})$.

Легко установить следующие свойства внешней меры P^*:

$$P^*(\emptyset)=0, \tag{1.5.2}$$

$$P^*(R)=1, \tag{1.5.3}$$

$$P^*(E)=P(E), \quad \text{если } E\in\mathcal{F}, \tag{1.5.4}$$

$$P^*(E)\leqslant P^*(F), \quad \text{если } E\subset F, \tag{1.5.5}$$

$$P^*(E)\leqslant\sum_{\alpha}P^*(E_\alpha), \quad \text{если } E\subset\bigcup_{\alpha}E_\alpha. \tag{1.5.6}$$

Теперь мы сформулируем основную теорему о расширении вероятностной меры P, определенной на булевом поле $\mathcal{F}$, до вероятностной меры на борелевском поле $\mathcal{B}(\mathcal{F})$.

1.5.2. *Пусть P — вероятностная мера на булевом поле $\mathcal{F}$. Тогда функция множества $P^*(E)$, определенная соотношением* (1.5.1), *есть вероятностная мера на $\mathcal{B}(\mathcal{F})$, для которой $P^*(E)=P(E)$, как только $E\in\mathcal{F}$.*

Доказательство этой теоремы читатель может найти у Халмоша (1950).

1.6. Статистическая независимость

(а) Декартовы произведения выборочных пространств. *Декартовым произведением* R выборочных пространств $R^{(1)}$ и $R^{(2)}$ называется множество всех упорядоченных пар $(e^{(1)}, e^{(2)})$, где $e^{(1)}\in R^{(1)}$ и $e^{(2)}\in R^{(2)}$; оно обозначается

$$R=R^{(1)}\times R^{(2)}. \tag{1.6.1}$$

Подобным образом, если $E^{(1)}$ и $E^{(2)}$ — события соответственно в $R^{(1)}$ и $R^{(2)}$, то декартовым произведением $E^{(1)}$ и $E^{(2)}$ называется событие E в $R=R^{(1)}\times R^{(2)}$, выборочные точки которого $e=(e^{(1)}, e^{(2)})$ удовлетворяют условию $e^{(1)}\in E^{(1)}$ и $e^{(2)}\in E^{(2)}$. Обычно пишут

$$E=E^{(1)}\times E^{(2)}. \tag{1.6.2}$$

Множество $E\in R$ представляет собой пустое множество тогда и только тогда, когда пустым является $E^{(1)}$ или $E^{(2)}$. Также очевидно, что если $E^{(1)}$ и $F^{(1)}$ — события в $R^{(1)}$ и $E^{(2)}$ и $F^{(2)}$ — события в $R^{(2)}$, то $(E^{(1)}\times E^{(2)})\subset(F^{(1)}\times F^{(2)})$ тогда и только тогда, когда $E^{(1)}\subset F^{(1)}$ и $E^{(2)}\subset F^{(2)}$. Кроме того, равенство $(E^{(1)}\times E^{(2)})=(F^{(1)}\times F^{(2)})$ равносильно равенствам $E^{(1)}=F^{(1)}$ и $E^{(2)}=F^{(2)}$.

Иногда говорят, что событие E, определенное формулой (1.6.2), выражает *совместное наступление* событий $E^{(1)}$ и $E^{(2)}$. Также удобно говорить, что $E^{(1)}$ есть *проекция* E в $R^{(1)}$ и аналогично $E^{(2)}$ есть проекция E в $R^{(2)}$. $R^{(1)}$ и $R^{(2)}$ называются *компонентами* R или *маргинальными выборочными пространствами* R. Если в R мы возьмем все те точки $(e^{(1)}, e^{(2)})$, у которых $e^{(1)}\in E^{(1)}$, то получим *цилиндрическое множество* в R, фактически — декартово произведение $E^{(1)}\times R^{(2)}$. Точно так же $R^{(1)}\times E^{(2)}$ есть цилиндрическое множество, где $e^{(2)}\in E^{(2)}$. Таким образом, $E^{(1)}\times E^{(2)}$, $E^{(1)}\times R^{(2)}$ и $R^{(1)}\times E^{(2)}$ — события в выборочном пространстве $R=R^{(1)}\times R^{(2)}$, причем декартово произведение $E^{(1)}\times E^{(2)}$ можно рассматривать как пересечение цилиндрических множеств $E^{(1)}\times R^{(2)}$ и $R^{(1)}\times E^{(2)}$, т. е.

$$E^{(1)}\times E^{(2)}=(E^{(1)}\times R^{(2)})\cap(R^{(1)}\times E^{(2)}). \tag{1.6.3}$$

Мы с особой силой подчеркиваем то обстоятельство, что события E типа декартова произведения образуют специальный класс событий в R, который играет важную роль в теории вероятностей.

Понятие декартова произведения непосредственно обобщается на случай любого конечного или счетного числа событий $E^{(1)}$, $E^{(2)}$, ... соответственно в выборочных пространствах $R^{(1)}$, $R^{(2)}$, ...

Примеры. В качестве иллюстрации понятия декартова произведения выборочных пространств и событий рассмотрим частный случай, когда $R^{(1)}$, $R^{(2)}$, $E^{(1)}$ и $E^{(2)}$ — множества точек на вещественной оси R_1. Пусть $R^{(1)}$ — замкнутый интервал $[a_1, b_1]$ на $e^{(1)}$-оси, $R^{(2)}$ — замкнутый интервал $[a_2, b_2]$ на $e^{(2)}$-оси, $E^{(1)}$ — замкнутое множество в $R^{(1)}$ и $E^{(2)}$ — замкнутое множество в $R^{(2)}$. Эти множества показаны на рис. 1.2. Декартово произведение $R=R^{(1)}\times R^{(2)}$ состоит из всех точек большого прямоугольника $ABCD$, включая его граничные точки. Цилиндрическое множество $E^{(1)}\times R^{(2)}$ есть множество точек, содержащихся в заштрихованных на рис. 1.2 вертикальных полосках.

Множество точек, содержащихся в покрытых штриховкой горизонтальных полосках, есть цилиндрическое множество $R^{(1)}\times E^{(2)}$. Наконец, декартово произведение $E=E^{(1)}\times E^{(2)}$ состоит из точек четырех зачерненных прямоугольников. Таким образом, E оказывается пересечением двух указанных цилиндрических множеств в R, т. е. $E=(E^{(1)}\times R^{(2)})\cap(R^{(1)}\times E^{(2)})$.

Следует, однако, иметь в виду, что операция образования декартова произведения событий может быть применена к событиям более общей природы, чем те, которые изображаются при помощи вещественных чисел, являясь множествами точек на прямой, плоскости или в многомерном эвклидовом пространстве. Предположим, например, что из каждой из двух колод, содержащих по 52 игральные карты, извлекается наудачу 13 карт. Пусть $e^{(1)}$ означает набор карт, взятых из первой колоды, и $e^{(2)}$ — набор карт, взятых из второй колоды. Тогда число выборочных точек $e^{(1)}$ и $e^{(2)}$, образующих соответственно пространства $R^{(1)}$ и $R^{(2)}$, в каждом случае равно $\binom{52}{13}$. Кроме того, $\binom{52}{13}^2$ пар наборов $(e^{(1)}, e^{(2)})$ могут получиться после извлечения карт из обеих колод соответственно $\binom{52}{13}^2$ выборочным точкам декартова произведения пространств $R=R^{(1)}\times R^{(2)}$. Таким образом, если $E^{(1)}$ есть множество в $R^{(1)}$, состоящее из всех тех наборов карт, которые содержат точно 2 туза, и $E^{(2)}$ есть множество в $R^{(2)}$, состоящее из всех тех наборов, которые содержат точно 3 короля, то $E=E^{(1)}\times E^{(2)}$ будет множеством в R, состоящим из всех таких пар наборов $(e^{(1)}, e^{(2)})$, что первый набор $e^{(1)}$ содержит точно 2 туза и второй набор $e^{(2)}$ содержит точно 3 короля. Подобных пар наборов, иначе выборочных точек в E, всего

$$\binom{4}{2}\binom{48}{11}\binom{4}{3}\binom{48}{10}.$$

Пусть $R^{(1)}$ и $R^{(2)}$ — два выборочных пространства и $\mathscr{B}(\mathscr{F}^{(1)})$, $\mathscr{B}(\mathscr{F}^{(2)})$ — борелевские поля, порожденные булевыми полями, множеств $\mathscr{F}^{(1)}$ и $\mathscr{F}^{(2)}$ соответственно в $R^{(1)}$ и $R^{(2)}$. В пространстве $R = R^{(1)} \times R^{(2)}$ рассмотрим класс $\mathscr{F}_0$ всех множеств вида $E = E^{(1)} \times E^{(2)}$, где $E^{(1)} \in \mathscr{B}(\mathscr{F}^{(1)})$ и $E^{(2)} \in \mathscr{B}(\mathscr{F}^{(2)})$. Класс $\mathscr{F}_0$ мы используем как начальный класс для образования наименьшего борелевского поля множеств в R, по-прежнему обозначая это поле

$$\mathscr{B}(\mathscr{F}_0). \tag{1.6.4}$$

Можно показать, что на самом деле борелевское поле (1.6.4) есть не что иное, как борелевское поле, порожденное начальным классом $\mathscr{F}_0$ множеств вида $E^{(1)} \times E^{(2)}$, где $E^{(1)} \in \mathscr{F}^{(1)}$ и $E^{(2)} \in \mathscr{F}^{(2)}$.

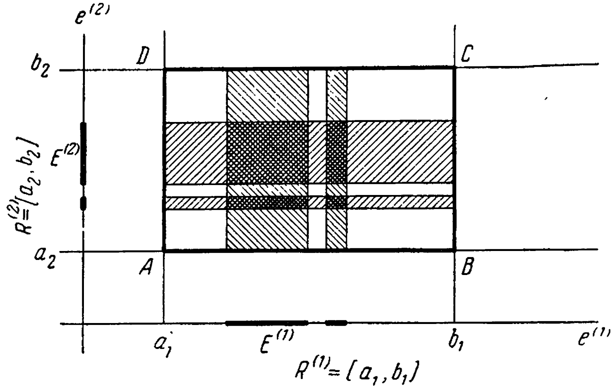

Рис. 1.2. Рисунок, иллюстрирующий прямое произведение событий и выборочных пространств.

(b) Произведения вероятностных мер. Предположим теперь, что $P^{(1)}$ и $P^{(2)}$ — вероятностные меры соответственно на $\mathscr{B}(\mathscr{F}^{(1)})$ и $\mathscr{B}(\mathscr{F}^{(2)})$. Для любого множества $E = E^{(1)} \times E^{(2)}$ из $\mathscr{F}_0$, где $E^{(1)} \in \mathscr{B}(\mathscr{F}^{(1)})$ и $E^{(2)} \in \mathscr{B}(\mathscr{F}^{(2)})$, определим число, вероятность множества E, формулой

$$P(E) = P^{(1)}(E^{(1)})\, P^{(2)}(E^{(2)}). \tag{1.6.5}$$

Можно показать, что P представляет собой булеву вероятностную меру на $\mathscr{F}_0$, допускающую единственное расширение на $\mathscr{B}(\mathscr{F}_0)$.

Всякий раз, когда некоторая вероятностная мера P, заданная на борелевском поле $\mathscr{B}(\mathscr{F}_0)$, удовлетворяет (1.6.5) для любого множества $E = E^{(1)} \times E^{(2)}$, где $E^{(1)} \in \mathscr{B}(\mathscr{F}^{(1)})$ и $E^{(2)} \in \mathscr{B}(\mathscr{F}^{(2)})$, мы будем говорить, что *составляющие* вероятностные пространства $(R^{(1)}, \mathscr{B}(\mathscr{F}^{(1)}), P^{(1)})$ и $(R^{(2)}, \mathscr{B}(\mathscr{F}^{(2)}), P^{(2)})$ являются (*статистически*) *независимыми*. Вероятностное пространство

$$(R, \mathscr{B}(\mathscr{F}_0), P), \tag{1.6.6}$$

где P есть вероятностная мера, определенная формулой (1.6.5), называется *декартовым произведением* двух составляющих вероятностных пространств

$$(R^{(1)}, \mathscr{B}(\mathscr{F}^{(1)}), P^{(1)}) \quad \text{и} \quad (R^{(2)}, \mathscr{B}(\mathscr{F}^{(2)}), P^{(2)}).$$

Понятие статистической независимости пространств естественным образом обобщается на случай более чем двух вероятностных пространств.

1.7. Случайные величины

Пусть $(R, \mathscr{B}, P)$ — вероятностное пространство. Рассмотрим вещественную однозначную функцию $x(e)$, определенную во всех выборочных точках $e \in R$, обладающую тем свойством, что для каждого вещественного числа b множество E_b всех $e \in R$, в которых $x(e) \leqslant b$, принадлежит $\mathscr{B}$. Функция $x(e)$ называется *случайной величиной* относительно $\mathscr{B}$. Часто также говорят, что $x(e)$ $\mathscr{B}$-измерима. Заметим, что если $\{E_b\}$ — класс событий в R, отвечающих всем вещественным числам b, то $\mathscr{B}(\{E_b\})$ есть борелевское поле, порожденное этим классом. Множество значений, которые может принимать $x(e)$, когда e пробегает все пространство R, удобно называть выборочным пространством случайной величины $x(e)$.

Таким образом, если $\mathscr{B}_1$ есть класс борелевских множеств на вещественной прямой R_1, то случайная величина $x(e)$ отображает выборочные точки e из R в выборочные точки x на R_1, причем, каково бы ни было борелевское множество $E' \in \mathscr{B}_1$, событие E, состоящее из всех выборочных точек в R, для которых $x(e) \in E'$, содержится в $\mathscr{B}$. Указанное событие выгодно обозначать символом $x^{-1}(E')$. Исходя из вероятностного пространства $(R, \mathscr{B}, P)$, связанного с основным выборочным пространством R, мы можем определить вероятностное пространство $(R_1, \mathscr{B}_1, P')$, связанное с вещественной прямой R_1, полагая $P'(E') = P(x^{-1}(E'))$. В этом случае можно говорить, что $(R_1, \mathscr{B}_1, P')$ *индуцируется* случайной величиной $x(e)$ из пространства $(R, \mathscr{B}, P)$. Проверка того факта, что $(R_1, \mathscr{B}_1, P')$ есть вероятностное пространство, предоставляется читателю в качестве упражнения.

Если $x_1(e), \ldots, x_k(e)$ — k случайных величин в данном вероятностном пространстве $(R, \mathscr{B}, P)$, то событие $E_{b_1, \ldots, b_k}$ в R вида

$$\{e: x_i(e) \leqslant b_i, \quad i = 1, \ldots, k\},$$

где $b_1, \ldots, b_k$ — произвольные вещественные числа, также принадлежит $\mathscr{B}$ и $P(E_{b_1, \ldots, b_k}) = P(x_1(e) \leqslant b_1, \ldots, x_k(e) \leqslant b_k)$. Кроме того, если E' есть любое множество из $\mathscr{B}_k$ (класса борелевских множеств k-мерного евклидова пространства R_k), то в $\mathscr{B}$ содержится и множество E всех точек e, для которых $(x_1(e), \ldots, x_k(e)) \in E'$, так что, полагая по определению $P'(E') = P(E)$, мы получаем вероятностное пространство $(R_k, \mathscr{B}_k, P')$, индуцированное упорядоченным набором случайных величин $(x_1(e), \ldots, x_k(e))$ из пространства $(R, \mathscr{B}, P)$. Удобно называть $(x_1(e), \ldots, x_k(e))$ *k-мерной* или *векторной случайной величиной* относительно $\mathscr{B}$ или просто $\mathscr{B}$-измеримой k-мерной случайной величиной.

Если для всякой выборочной точки $e \in R$ мы имеем $|x(e)| < M$, где $M < \infty$, то $x(e)$ называется *ограниченной случайной величиной*, k-мерная случайная величина ограничена тогда и только тогда, когда ограничена каждая из ее компонент.

Примеры. Понятие случайной величины, возможно, приобретет дальнейшую ясность в результате рассмотрения нескольких простых примеров.

Рассматривая все возможные наборы из 13 карт, вынутых наудачу из колоды в 52 карты, мы получаем множество выборочных точек e основного выборочного пространства R, состоящее из $\binom{52}{13}$ таких наборов. Если $x(e)$ — случайная величина, которая представляет число тузов, содержащих в e, то $x(e)$ оказывается определенной в каждой выборочной точке пространства R и принимает там одно из значений 0, 1, 2, 3 или 4. Таким образом, $x(e)$ отображает каждую выборочную точку $e \in R$ в одну из точек 0, 1, 2, 3, 4 на вещественной прямой R_1. Так как число выборочных точек в R конечно, класс всех возможных событий в R (включая само R и пустое множество) образует булево поле $\mathscr{F}$. Множество точек $e \in R$, в которых $x(e) \leqslant b$, очевидно, является одним из событий класса $\mathscr{F}$ при любом вещественном b. Если каждой выборочной точке из R приписана некоторая вероятность (в случае «идеально» перетасованных карт всем им приписывается значение $\frac{1}{\binom{52}{13}}$, мы имеем булево вероятностное пространство $(R, \mathscr{F}, P)$, которое позволяет определить вероятность события $E_b = \{e : x(e) \leqslant b\}$, каково бы ни было вещественное число b. Отсюда, если E' — произвольное (борелевское) множество в R_1, его вероятность определяется вероятностным пространством $(R_1, \mathscr{B}_1, P')$, индуцированным величиной $x(e)$ из $(R, \mathscr{F}, P)$. Разумеется, теми немногими множествами из $\mathscr{B}_1$, которые представляют здесь интерес, являются множество $\{0, 1, 2, 3, 4\}$, состоящее из пяти точек, и все его подмножества, т. е. выборочное пространство величины $x(e)$ вместе со всеми своими подмножествами.

Если $(x_1(e), x_2(e))$ — двумерная случайная величина, где $x_1(e)$ обозначает число тузов и $x_2(e)$ — число пик, содержащихся в выборочной точке e, то вероятность $P(x_1(e) \leqslant b_1, x_2(e) \leqslant b_2)$ может быть вычислена, исходя из булева вероятностного пространства $(R, \mathscr{F}, P)$. При соотнесении указанной вероятности интервалу $[-\infty, -\infty; b_1, b_2]$ для всех пар вещественных чисел (b_1, b_2) с помощью векторной случайной величины $(x_1(e), x_2(e))$ индуцируется вероятностное пространство $(R_1, \mathscr{B}_2, P')$, которое определяет вероятность любого события E в $\mathscr{B}_2$. Вновь обратим внимание на тот факт, что из событий $E \subset R_2$ практическое значение имеют лишь события в множестве $\{(x_1, x_2) : x_1 = 0, 1, 2, 3, 4;\ x_2 = 0, 1, \ldots, 13,\ \text{но}\ 0 \leqslant x_1 + x_2 \leqslant 14\}$, состоящем из 62 точек, причем исключаются точки (4, 0) и (0, 13), т. е. выборочное пространство двумерной величины $(x_1(e), x_2(e))$ вместе со своими подмножествами.

Возьмем другой пример. Допустим, что две электрические лампочки B_1 и B_2 непрерывно горели в течение всего срока службы, который оказался равным t_1 и t_2, соответственно. Основное выборочное пространство R есть положительный квадрант t_1t_2-плоскости, а выборочные точки есть точки этого квадранта. Событие, заключающееся в том, что ни одна из лампочек не будет гореть дольше другой более чем на b единиц времени, состоит из всех точек в R, для которых $|t_1-t_2|\leqslant b$. Таким образом, мы получаем случайную величину $x(e)$, совпадающую с $|t_1-t_2|$ в каждой точке $e=(t_1, t_2)\in R$. Вероятность неравенства $x(e)\leqslant b$ может быть вычислена как только будет определено вероятностное пространство $(R, \mathscr{B}, P)$, где $\mathscr{B}$ — борелевское поле, содержащее события $|t_1-t_2|\leqslant b$, каково бы ни было вещественное число b, и, конечно, P — вероятностная мера на $\mathscr{B}$. Случайная величина $|t_1-t_2|$ отображает точки пространства R в точки вещественной прямой R_1, так что $P'(E'_b)=P(E_b)$, где E'_b есть интервал $(-\infty, b]\subset R_1$, а E_b — множество в R, для которого $|t_1-t_2|\leqslant b$, и b — произвольное вещественное число. Начальный класс событий $\{E_b: b \text{ — вещественное}\}$ вместе с вероятностями $\{P(E_b)\}$ может быть использован для образования наименьшего борелевского поля $\mathscr{B}(\{E_b\})$ и вероятностного пространства $(P, \mathscr{B}(\{E_b\}), P)$, из которого величиной $x(e)$ индуцируется вероятностное пространство $(R_1, \mathscr{B}_1, P')$ и которое доставляет вероятность того, что $|t_1-t_2|$ принадлежит любому множеству $E'\in\mathscr{B}_1$.

Если положить $x_1(e)=t_1+t_2$ и $x_2=t_1-t_2$, то получаются две случайные величины, определенные в каждой точке R. Пусть вероятностное пространство $(R, \mathscr{B}(\{E_{b_1, b_2}\}), P)$ порождено начальным классом событий $\{E_{b_1, b_2}: b_1, b_2 \text{ — вещественные}\}$ и их вероятностями $\{P(E_{b_1, b_2})\}$, где $E_{b_1, b_2}=\{e: x_1(e)\leqslant b_1, x_2(e)\leqslant b_2\}$. Тогда величина $(x_1(e), x_2(e))$ индуцирует из $(R_1, \mathscr{B}(\{E_{b_1, b_2}\})P)$ вероятностное пространство $(R_2, \mathscr{B}_2, P')$, доставляющее вероятность того, что (t_1+t_2, t_1-t_2) принадлежит любому множеству E' из $\mathscr{B}_2$.

1.8. Интегрирование случайных величин

В этом параграфе, пользуясь вероятностной терминологией и без доказательства, мы воспроизведем некоторые основы теории интеграла Лебега — Стилтьеса. Необходимые доказательства могут быть найдены в книгах Халмоша (1950), Лоэва (1955), Макшейна и Ботса (1959) и Сакса (1937).

Предположим, что $(R, \mathscr{B}, P)$ есть вероятностное пространство и $x(e)$ — случайная величина относительно $\mathscr{B}$. Если $x(e)$ принимает только конечное число различных значений $x_1, \ldots, x_k$, так что $x(e)=x_i$ для всех $e\in E_i$, $i=1, \ldots, k$, то $x(e)$ называется простой случайной величиной. В этом случае мы полагаем

$$\int_R x(e)\,dP(e)=\sum_{i=1}^{k} x_i P(E_i). \tag{1.8.1}$$

Вообще, если E — любое множество из $\mathscr{B}$, мы пишем

$$\int_E x(e)\,dP(e)=\sum_{i=1}^{k} x_i P(E\cap E_i). \tag{1.8.2}$$

Теперь рассмотрим случай, когда $x(e)$ является ограниченной и может принимать бесконечное число значений. Можно показать, что существует такая последовательность *простых* случайных величин $x_1(e), x_2(e), \ldots$, что

$$\lim_{\alpha\to\infty} x_\alpha(e)=x(e) \tag{1.8.3}$$

равномерно для всех $e \in R$, и, кроме того, для каждого $E \in \mathscr{B}$ предел

$$\lim_{\alpha \to \infty} \int_E x_\alpha(e)\, dP(e) \tag{1.8.4}$$

существует и не зависит от выбора частной последовательности, удовлетворяющей (1.8.3). Этот предел, который мы будем обозначать

$$\int_E x(e)\, dP(e), \tag{1.8.5}$$

называется интегралом Лебега — Стилтьеса от величины $x(e)$ относительно $(R, \mathscr{B}, P)$, взятым по множеству E. В случае простой величины $x(e)$ интеграл Лебега — Стилтьеса от $x(e)$ по множеству E дается формулой (1.8.2).

В более широких условиях, когда величина $x(e)$ не обязательно является ограниченной, но является неотрицательной, она будет *интегрируемой*, если существует такая последовательность ограниченных случайных величин $x_1(e) \leqslant x_2(e) \leqslant \ldots$, где неравенства выполняются для всех $e \in R$, что

$$\lim_{\alpha \to \infty} x_\alpha(e) = x(e), \qquad e \in R, \tag{1.8.6}$$

и

$$\lim_{\alpha \to \infty} \int_E x_\alpha(e)\, dP(e) < \infty. \tag{1.8.7}$$

Этот последний предел и называется интегралом от $x(e)$ относительно $(R, \mathscr{B}, P)$, взятым по множеству E, и обозначается как

$$\int_E x(e)\, dP(e).$$

При этом можно показать, что для любого $E \in \mathscr{B}$

$$\lim_{\alpha \to \infty} \int_E x_\alpha(e)\, dP(e) \tag{1.8.8}$$

существует и не зависит от выбранной частной последовательности $x_1(e) \leqslant x_2(e) \leqslant \ldots$, которая удовлетворяет (1.8.6) и (1.8.7).

Наконец, когда величина $x(e)$ — произвольная $\mathscr{B}$-измеримая функция, то говорят, что она интегрируема, если ее можно представить в виде разности $x'(e) - x''(e)$ двух неотрицательных интегрируемых случайных величин $x'(e)$ и $x''(e)$. В этом случае по определению

$$\int_E x(e)\, dP(e) = \int_E x'(e)\, dP(e) - \int_E x''(e)\, dP(e). \tag{1.8.9}$$

Можно проверить, что значение левой части (1.8.9) не зависит от частного выбора $x'(e)$ и $x''(e)$.

Интеграл Лебега — Стилтьеса обладает следующими важными свойствами, которые мы выразим в терминах случайных величин. Пусть E есть произвольное множество из $\mathscr{B}$.

1.8.1. *Если* $x(e)=k$ *для всех* $e\in R$, *где* k — *некоторая постоянная, то*

$$\int_E x(e)\,dP(e)=kP(E).$$

1.8.2. *Если* $a\leqslant x(e)\leqslant b$ *для всех* $e\in R$ *и* a *и* b — *постоянные, то*

$$aP(E)\leqslant\int_E x(e)\,dP(e)\leqslant bP(E).$$

1.8.3. *Если* $x_1(e)$ *и* $x_2(e)$ *интегрируемы и* a *и* b — *произвольные постоянные, то случайная величина* $ax_1(e)+bx_2(e)$ *интегрируема и*

$$\int_E (ax_1(e)+bx_2(e))\,dP(e)=a\int_E x_1(e)\,dP(e)+b\int_E x_2(e)\,dP(e).$$

1.8.4. *Если* $x_1(e)\leqslant x(e)\leqslant x_2(e)$ *для всех* $e\in R$, *где* $x_1(e)$, $x_2(e)$ *и* $x(e)$ *интегрируемы, то*

$$\int_E x_1(e)\,dP(e)\leqslant\int_E x(e)\,dP(e)\leqslant\int_E x_2(e)\,dP(e).$$

1.8.5. $x(e)$ *интегрируема тогда и только тогда, когда интегрируема ее абсолютная величина* $|x(e)|$. *Кроме того,*

$$\left|\int_E x(e)\,dP(e)\right|\leqslant\int_E |x(e)|\,dP(e).$$

1.8.6. *Если* $x_1(e)$, $x_2(e)$, ... — *такая последовательность случайных величин, что для всех* $e\in R$

$$\lim_{\alpha\to\infty} x_\alpha(e)=x(e)\quad \text{и}\quad |x_\alpha(e)|\leqslant y(e),$$

где $y(e)$ *интегрируема, то* $x(e)$ *также интегрируема и*

$$\lim_{\alpha\to\infty}\int_E |x_\alpha(e)-x(e)|\,dP(e)=0.$$

В частности,

$$\lim_{\alpha\to\infty}\int_E x_\alpha(e)\,dP(e)=\int_E x(e)\,dP(e)$$

равномерно по всем E *из* $\mathscr{B}$.

1.8.7. *Пусть* $(R_k, \mathscr{B}_k, P')$ — *вероятностное пространство, индуцированное векторной случайной величиной* $(x_1(e), \ldots, x_k(e))$ *из вероятностного пространства* $(R, \mathscr{B}, P)$. *Пусть функция* $g(x_1, \ldots, x_k)$ *измерима относительно* $\mathscr{B}_k$ (*и, следовательно,* $g(x_1(e), \ldots, x_k(e))$ *измерима относительно* $\mathscr{B}$) *и* $E\subset R$ — *такое множество, для которого* $(x_1(e), \ldots, x_k(e))\in E'_k$. *Тогда*

$$\int_E g(x_1(e), \ldots, x_k(e))\,dP(e)=\int_{E'_k} g(x_1, \ldots, x_k)\,dP'(x_1, \ldots, x_k),$$

причем если один из двух интегралов конечен, то конечен и другой и оба интеграла равны.

В заключение заметим, что если в формулировке свойств **1.8.1** — **1.8.6** заменить выражение «для всех $e \in R$» выражением «для всех $e \in R$, за исключением, быть может, точек e, образующих множество F, для которого $P(F)=0$», то все утверждения останутся верными.

1.9. Условная вероятность

Пусть $(R, \mathcal{B}, P)$ — некоторое вероятностное пространство, E_1 и E_2 — события из $\mathcal{B}$ и $P(E_1)>0$. Запишем

$$P(E_2 \mid E_1)=\frac{P(E_1 \cap E_2)}{P(E_1)}. \tag{1.9.1}$$

Это отношение называется *условной вероятностью* события E_2 при условии, что событие E_1 наступило.

Можно проверить, что для любого фиксированного события $E_1 \in \mathcal{B}$, для которого $P(E_1)>0$, $(R, \mathcal{B}, P(\cdot \mid E_1))$ является вероятностным пространством. Здесь $P(\cdot \mid E_1)$ обозначает меру, принимающую на $E_2 \in \mathcal{B}$ значение $P(E_2 \mid E_1)$.

Заметим, что мы вправе переписать (1.9.1) следующим образом:

$$P(E_1 \cap E_2)=P(E_1)\,P(E_2 \mid E_1). \tag{1.9.2}$$

Если $P(E_2 \mid E_1)=P(E_2)$, то (1.9.2) приводится к случаю *независимых событий* E_1 и E_2, и тогда

$$P(E_1 \cap E_2)=P(E_1)\,P(E_2). \tag{1.9.2a}$$

Вообще же можно показать, что

1.9.1. *Если $E_1, E_2, \ldots$ — конечная или счетная последовательность событий из $\mathcal{B}_1$, для которых $P(E_1)>0$, $P(E_1 \cap E_2)>0$, $P(E_1 \cap E_2 \cap E_3)>0, \ldots$, то*

$$P(E_1 \cap E_2 \cap \ldots)=$$
$$=P(E_1)\,P(E_2 \mid E_1)\,P(E_3 \mid E_1 \cap E_2)\ldots P(E_n \mid E_1 \cap E_2 \cap \ldots \cap E_{n-1})\ldots \tag{1.9.3}$$

В случае *взаимной независимости* событий $E_1, E_2, \ldots$ (1.9.3) превращается в соотношение

$$P(E_1 \cap E_2 \cap \ldots)=P(E_1)\,P(E_2)\ldots \tag{1.9.3a}$$

Приведем еще один полезный результат.

1.9.2. *Если $E_1, E_2, \ldots$ — конечная или счетная последовательность попарно не пересекающихся множеств из $\mathcal{B}$, имеющих ненулевые вероятности, причем $\bigcup\limits_{\alpha} E_\alpha = R$ и E — любое множество из $\mathcal{B}$, то*

$$P(E)=P(E_1)\,P(E \mid E_1)+P(E_2)\,P(E \mid E_2)+\ldots \tag{1.9.4}$$

Проверка этого утверждения предоставляется читателю.

1.10. Условные случайные величины

Допустим, что $(x_1(e), x_2(e))$ является двумерной случайной величиной относительно $\mathscr{B}$ для данного вероятностного пространства $(R, \mathscr{B}, P)$. Если в качестве множеств E_1 и E_2 в (1.9.1) выбрать соответственно множества точек e, для которых $x_1(e) \in E_1'$ и $x_2(e) \in E_2'$, где E_1' и E_2' содержатся в $\mathscr{B}_1$, то (1.9.1) становится формулой условной вероятности, касающейся случайных величин. Она дает условную вероятность события $x_2(e) \in E_2'$ при условии, что $x_1(e) \in E_1'$. В частности, пусть E_1' содержит единственную точку, например x_1. Если $P(x_1(e) = x_1) > 0$, то применение формулы (1.9.1) не вызывает никаких затруднений. Однако если $P(x_1(e) = x_1) = 0$, встает вопрос: существует ли вообще какая-либо возможность придать смысл условной вероятности $P(x_2(e) \in E_2' \mid x_1(e) = x_1)$?

В большинстве случаев, представляющих интерес в математической статистике, этой условной вероятности, как мы покажем в § 2.8, можно дать истолкование с помощью весьма элементарных соображений. Но при условиях более общих ответ на наш вопрос доставляется теоремой Радона — Никодима, которая может быть сформулирована следующим образом.

1.10.1. *Если $(R, \mathscr{B}, P)$ — вероятностное пространство и Q — такая (конечная) вполне аддитивная функция множества на $\mathscr{B}$, что $Q(E) = 0$ для каждого множества $E \in \mathscr{B}$, для которого $P(E) = 0$, то существует случайная величина $g(e)$ такая, что*

$$Q(E) = \int\limits_E g(e)\, dP \tag{1.10.1}$$

для любого $E \in \mathscr{B}$. Кроме того, если $g(e)$ и $h(e)$ — две случайные величины с этим свойством, то $P(g(e) \neq h(e)) = 0$.

Мы опускаем доказательство этой теоремы, отсылая читателя к Халмошу (1950), у которого он найдет несколько более общую формулировку теоремы и ее доказательство. Следует лишь заметить, что значения Q не обязаны быть неотрицательными.

Если две вполне аддитивные функции множества P и Q, определенные на $\mathscr{B}$, таковы, что $Q(E) = 0$ для каждого множества E, для которого $P(E) = 0$, говорят, что Q *абсолютно непрерывна относительно P.*

Предположим, что $(R, \mathscr{B}, P)$ — вероятностное пространство и $x(e)$ — случайная величина. Пусть для данного $E \in \mathscr{B}$ и фиксированного числа x_1 мы хотим найти условную вероятность $P(E \mid x(e) = x_1)$. Как мы видели, определение (1.9.1) теряет смысл, если $P(x(e) = x_1) = 0$.

Тем не менее, следуя интуиции, было бы желательно определить эту условную вероятность как нечто аналогичное «пределу» вероятности $P(E \mid x(e) \in N_\alpha(x_1))$ при $\alpha \to \infty$, где $N_1(x_1)$, $N_2(x_1)$, ... представляет последовательность окрестностей точки x_1, стягивающуюся к x_1. Вообще говоря, всюду этот предел существовать не будет. Но ведь интуитивно правдоподобно, что если он существует в некоторой достаточно хорошей области, содержащей F, где $F \in \mathcal{B}$, то должна быть возможность получить $P(E \mid x(e) \in F)$ посредством подходящего усреднения $P(E \mid x(e) = x_1)$ по всем x_1, лежащим в F.

Вообще говоря, трудно установить предел такого рода, о котором шла речь выше, если не считать особо благоприятных случаев. Однако Колмогоров (1933а) предложил следующий подход.

Для любых множеств $E \in \mathcal{B}$ и $F \in \mathcal{B}_1$ положим

$$Q(F) = P(E \cap x^{-1}(F)). \tag{1.10.2}$$

Если $P(x^{-1}(F)) = 0$, то и $Q(F) = 0$. Ясно, что Q — вполне аддитивная функция множества на $\mathcal{B}_1$. Согласно теореме **1.10.1** существует такая вещественная $\mathcal{B}_1$-измеримая функция $f(x)$, что

$$Q(F) = \int_F f(x)\, dP'(x), \tag{1.10.3}$$

где $P'(F) = P(x^{-1}(F))$ для всех множеств $F \in \mathcal{B}_1$. Следовательно, если мы обозначим $g(e) = f(x(e))$, то будем иметь

$$P(E \cap x^{-1}(F)) = \int_{x^{-1}(F)} g(e)\, dP(e) \tag{1.10.4}$$

и $g(e)$, иначе, $f(x(e))$ есть «условная вероятность события E относительно $x(e)$». Следует заметить, что $f(x(e))$ единственна в том же смысле, что и функция $g(e)$ в теореме **1.10.1**. Исходя из этого определения «условной вероятности E относительно $x(e)$», можно показать, что всюду в R_1, за исключением разве лишь точек, образующих множество нулевой вероятности,

$$f(x_1) = \lim_{\alpha \to \infty} P(E \mid x(e) \in N_\alpha(x_1)), \tag{1.10.5}$$

где $\{N_\alpha(x_1),\ \alpha = 1, 2, \ldots\}$ есть последовательность измеримых окрестностей точки x_1, стягивающаяся к x_1.

Рассмотренный выше подход естественным образом распространяется на случай векторной величины $(x_1(e), \ldots, x_k(e))$.

ЗАДАЧИ

1.1. Из списка, содержащего N различных имен, «наудачу» выбираются два имени и располагаются в алфавитном порядке. Описать выборочное пространство, порождаемое этой операцией, и подсчитать, сколько выборочных точек оно содержит. Рассмотреть обобщение на случай, когда из списка N имен берутся «наудачу» n имен.

1.2. Регистрируются дни рождения (месяц и день месяца) двух лиц A и B, выбираемых «наудачу» из книги «Who is who». Опуская високосные года, описать выборочное пространство, порождаемое этой операцией, и подсчитать число выборочных точек в нем. Указать число выборочных точек следующих событий:

(a) «A и B имеют один и тот же день рождения».

(b) «Дни рождения A и B разделяют не более чем r дней».

(c) «Дни рождения A и B относятся к различным месяцам».

1.3. Магазин открывается в 9 часов утра и закрывается в 5 часов вечера. Взятый «наудачу» покупатель входит в этот магазин в момент времени x и покидает его в момент времени y (как x, так и y выражаются в часах на оси времени, где 9 часов утра служит началом отсчета). Описать выборочное пространство переменной (x, y). В терминах x и y описать следующие события:

(a) «Покупатель находится в магазине менее одного часа».

(b) «В момент времени z покупатель находится в магазине».

(c) «Покупатель вошел в магазин до момента времени u и вышел после момента времени v».

1.4. В ящике, содержащем N электрических ламп, имеется $r\,(r<N)$ ламп с дефектной нитью накала, и некто испытывает их поочередно, фиксируя при этом, зажигается лампа или нет, до тех пор, пока не обнаружит какую-нибудь неисправную лампу (т. е. лампу с дефектной нитью).

Описать выборочное пространство, порождаемое этой операцией. Сколько точек в выборочном пространстве? Дать обобщение на случай, когда лампы испытываются поочередно до тех пор, пока не будет обнаружено точно s неисправных.

1.5. Пусть R — множество всех студентов университета A. Обозначим через E_1 множество тех из них, которые подписались на журнал M_1, через E_2 — множество подписавшихся на журнал M_2 и через E_3 — множество подписавшихся на журнал M_3.

(a) Описать словами следующие множества:

$$E_1 \cup E_2 \cup E_3; \quad E_1 \cap E_2 \cap E_3; \quad R-(E_1 \cup E_2);$$

$$\bar{E}_1 \cup \bar{E}_2 \cup \bar{E}_3; \quad E_1 \cap E_2 \cap \bar{E}_3; \quad \bar{E}_1 \cap \bar{E}_2 \cap \bar{E}_3.$$

(b) Выразить посредством операций над R, E_1, E_2, E_3 следующие множества:

(I) Множество студентов, которые подписались по крайней мере на два из трех журналов.

(II) Множество студентов, которые подписались на более чем один из трех журналов.

1.6. Рассмотрим множество всех возможных наборов, содержащих 13 карт из обычной колоды игральных карт. Пусть E_1, E_2, E_3, E_4 — множества различных наборов из 13 карт, в которых находятся соответственно туз пик, туз червей, туз бубен и туз треф.

Описать словами следующие множества и подсчитать, сколько наборов заключается в каждом из них:

$$E_1 \cap E_2; \quad E_1-(E_2 \cap E_3); \quad (E_1 \cap E_2 \cap E_3) \cup E_4;$$

$$E_1 \cup (E_2 \cap E_3); \quad (E_1 \cup E_2)-\bar{E}_3; \quad \bar{E}_1 \cup \bar{E}_2; \quad \bar{E}_1 \cap (\bar{E}_2 \cup \bar{E}_3);$$

$$(E_1 \cup E_2 \cup E_3)-E_4; \quad R-[(E_1 \cup E_2) \cap (E_3 \cup \bar{E}_4)].$$

1.7. В игру крэпс играют с помощью двух обычных 6-гранных игральных костей следующим образом. Если игроку выпадает 7 или 11 очков, он выигрывает и игра оканчивается. Если ему выпадает 2, 3 или 12 очков, он проигрывает и игра также заканчивается. Если же ему выпадает 4, 5, 6, 8, 9 или 10 очков, он должен продолжить бросание костей, пока ему не выпадает

7 очков или число очков, выпавших в самом начале. Если сначала появляется 7 очков, он проигрывает. Если появляются очки, выпавшие вначале, он выигрывает. Описать соответствующее выборочное пространство и, считая кости вполне правильными, показать, что вероятность выигрыша игрока равна 244/495.

1.8. Показать, что если выборочное пространство R состоит из N выборочных точек, то общее число событий в булевом поле, порожденном этими N точками, равно 2^N.

1.9. Рассмотрим две бесконечные последовательности множеств E_1, E_2, ... и F_1, F_2, ... на xy-плоскости, где E_n — множество точек, для которых $x^2+y^2<\frac{1+n}{n}$, и F_n — множество точек, для которых $x^2+y^2\leqslant\frac{n}{1+n}$. Пусть $\bar{E}_n$ означает дополнение E_n до полной xy-плоскости; аналогичным образом определяется и $\bar{F}_n$. Описать в терминах x и y следующие множества:

$$\lim_\alpha E_\alpha;\quad \lim_\alpha \bar{E}_\alpha;\quad \lim_\alpha F_\alpha;\quad \lim_\alpha \bar{F}_\alpha;\quad \lim_\alpha E_\alpha \cup \bar{F}_\alpha.$$

1.10. Если $E_1, \ldots, E_n$ — произвольные множества в выборочном пространстве R и $\bar{E}_1, \ldots, \bar{E}_n$ — их дополнения, то множества

$$\bigcup_{\alpha=1}^{n} E_\alpha \quad \text{и} \quad \bigcap_{\alpha=1}^{n} \bar{E}_\alpha$$

не пересекаются, их объединение есть R, так что

$$P\left(\bigcup_{\alpha=1}^{n} E_\alpha\right)=1-P\left(\bigcap_{\alpha=1}^{n} \bar{E}_\alpha\right).$$

1.11. События $E_1, \ldots, E_n$ таковы, что вероятность совместного наступления любых r фиксированных среди них равна p_r, $r=1, \ldots, n$. Показать, что

a) Вероятность наступления по крайней мере одного из событий $E_1, \ldots, E_n$ равна

$$\binom{n}{1}p_1-\binom{n}{2}p_2+\ldots+(-1)^{n-1}\binom{n}{n}p_n.$$

(b) Вероятность наступления по крайней мере m из событий $E_1, \ldots, E_n$ равна

$$\binom{m-1}{m-1}\binom{n}{m}p_m-\binom{m}{m-1}\binom{n}{m+1}p_{m+1}+\ldots+(-1)^{n-m}\binom{n-1}{m-1}\binom{n}{n}p_n.$$

(c) Вероятность наступления точно m из событий $E_1, \ldots, E_n$ равна

$$\binom{m}{m}\binom{n}{m}p_m-\binom{m+1}{m}\binom{n}{m+1}p_{m+1}+\ldots+(-1)^{n-m}\binom{n}{m}\binom{n}{n}p_n.$$

1.12. На предприятии, производящем корнфлекс, в каждую коробку кладется случайная карточка с одним из чисел 1, 2, ..., r, причем все числа должны попадать в коробку с одинаковой вероятностью. Пусть куплено $n\ (>r)$ коробок корнфлекса. Тогда вероятность того, что из чисел, имеющихся на карточках в этих коробках, можно будет составить по крайней мере один полный набор, равна

$$1-\binom{r}{1}\left(1-\frac{1}{r}\right)^n+\binom{r}{2}\left(1-\frac{2}{r}\right)^n+\ldots+(-1)^{r-1}\binom{r}{r-1}\left(1-\frac{r-1}{r}\right)^n.$$

1.13. Из урны, содержащей N фишек, занумерованных числами $1, 2, \ldots, N$, одна за другой (без возвращения) извлекаются две фишки. Обозначим через e произвольную точку в выборочном пространстве R, порожденном этой операцией, и через $x(e)$ — абсолютное значение разности чисел на двух фишках, представляющих e. Пусть все точки e в выборочном пространстве R имеют одинаковые вероятности. Показать, что $x(e)$ есть случайная величина. Описать ее выборочное пространство и вычислить вероятность $P(x(e)=x')$.

1.14. Допустим, что в задаче **1.4** всем возможным последовательностям, в которых могут быть обследованы N имеющихся электрических ламп, соответствует одна и та же вероятность. Какова вероятность того, что, желая обнаружить $s\ (s\leqslant r)$ дефектных ламп, мы встретим s-ю дефектную лампу при обследовании с порядковым номером x? Определить диапазон значений x, для которых эта вероятность является положительной.

1.15. Монета бросается последовательно n раз. Пусть e — выборочная точка в пространстве R, порожденном этой операцией, которое содержит 2^n точек, и $x(e)$ — число гербов в e. При условии, что все выборочные точки в R равновероятны, показать, что $x(e)$ — случайная величина. Описать выборочное пространство величины $x(e)$ и найти выражение для вероятности $P(x(e)=x')$.

1.16. Предположим, что элементы e выборочного пространства R являются внутренними точками квадрата на uv-плоскости с вершинами $(0, 0)$, $(0, 1)$, $(1, 0)$ и $(1, 1)$. Введем функцию $x(e)=u+v$, где u и v — координаты точки e. Определяя для любого треугольника или четырехугольника $E\subset R$ вероятность $P(E)$ как площадь E, показать, что $x(e)$ есть случайная величина, для которой вероятность $P(x(e)\leqslant x')$ может быть вычислена при каждом вещественном x'. Описать выборочное пространство величины $x(e)$ и найти выражение для $P(x(e)\leqslant x')$ в виде функции от x'.

Пусть F — множество точек в R, где $\frac{u}{2}<v<u$. Вычислить значение интеграла $\int\limits_E x(e)\,dP$. Вычислить также значение интеграла $\int\limits_R x(e)\,dP$.

1.17 (*Продолжение*). Введем $y(e)=\frac{u}{v}$. Показать, что $y(e)$ является случайной величиной, для которой вероятность $P(y(e)\leqslant y')$ может быть вычислена при любом вещественном y', и найти выражение для вероятности $P(y(e)\leqslant y')$.

1.18 (*Продолжение*). Показать, что $(x(e), y(e))$ есть двумерная случайная величина, для которой вероятность $P(x(e)\leqslant x', y(e)\leqslant y')$ может быть вычислена при любых вещественных x' и y', и найти выражение для этой вероятности.

1.19 (*Продолжение*). Пусть E_n, $n=1, 2, \ldots,$ — треугольники в R, задаваемые неравенствами $u>0$, $v>0$, $u+v\leqslant\frac{1+n}{2n}$. Чему равен предел $\lim\limits_{n\to\infty}E_n$? Показать, что $\lim\limits_{n\to\infty}P(E_n)=P(\lim\limits_{n\to\infty}E_n)=\frac{1}{8}$. Убедиться, что бесконечная последовательность $E_1, E_2, \ldots$ удовлетворяет формуле (1.9.3).

1.20. Доказать теоремы **1.9.1** и **1.9.2.**

1.21. Пусть $(R, \mathscr{B}, P)$ — вероятностное пространство и $x(e)$ — неотрицательная случайная величина относительно $\mathscr{B}$. Обозначим через $I_{\delta,\alpha}$ интервал $(\alpha\delta, (\alpha+1)\delta)$, $\alpha=0, 1, 2, \ldots$ и $\delta>0$, и через $I_{\delta,\alpha}^{-1}$ то множество из $\mathscr{B}$, для точек которого $x(e)\in I_{\delta,\alpha}$.

Кроме того, для $E \in \mathscr{B}$ положим

$$A(\delta) = \sum_{\alpha=1}^{\infty} \alpha\delta P(E \cap I_{\delta,\alpha}^{-1}).$$

Предполагая $A(\delta)$ для некоторого значения δ конечным, показать, что $\lim\limits_{\delta \to 0} A(\delta)$ существует и равен интегралу Лебега — Стилтьеса от $x(e)$ по множеству E.

1.22. Пусть $E_1, \ldots, E_n$ — произвольный набор событий и $p_m(\alpha_1, \ldots, \alpha_r)$ — вероятность наступления по крайней мере m из событий $E_{\alpha_1}, \ldots, E_{\alpha_r}$. Показать, что

$$(k+1-m)\,\Sigma_{k+1}\, p_m(\alpha_1, \ldots, \alpha_{k+1}) \leqslant (n-k)\,\Sigma_k p_m(\alpha_1, \ldots, \alpha_k),$$

где $k = 1, \ldots, n-1$, $1 \leqslant m \leqslant k$ и Σ_i, $i = k, k+1$ означает, что суммирование ведется по всем $\binom{n}{i}$ возможным группам i целых чисел из множества $\{1, 2, \ldots, n\}$ (Чжун (1941)).

ГЛАВА 2

ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

2.1. Предварительные замечания

Как мы видели в § 1.7, одномерная случайная величина $x(e)$ индуцирует из основного вероятностного пространства $(R, \mathscr{B}, P)$ вероятностное пространство $(R_1, \mathscr{B}_1, P')$, связанное с вещественной прямой R_1. Аналогичным образом k-мерная случайная величина $(x_1(e), \ldots, x_k(e))$ индуцирует из $(R, \mathscr{B}, P)$ вероятностное пространство $(R_k, \mathscr{B}_k, P')$, связанное с евклидовым пространством R_k.

Имея дело со случайной величиной $x(e)$, мы обычно будем пропускать упоминание о случайной точке $e \in R$ и просто обозначать эту величину буквой x *). Пусть x — одномерная случайная величина, вероятностное пространство которой есть $(R_1, \mathscr{B}_1, P')$, и E' — произвольное множество из $\mathscr{B}_1$. Чтобы избежать путаницы, мы иногда будем обозначать событие E' символом $x \in E'$ и вместо $P(E')$ писать $P(x \in E')$. В частности, если E' представляет собой интервал $(a, b]$, мы условимся, что вероятность $P(x \in E')$ может быть записана в виде $P(a < x \leqslant b)$; если E' состоит из единственной точки x', мы можем писать $P(x = x')$, обозначая вероятность того, что в данном опыте *реализуется* значение x' случайной величины x. Если x есть k-мерная векторная случайная величина и если в целях большей ясности желательно указать координаты x, мы будем обозначать событие $E' \in \mathscr{B}_k$ символом $(x_1, \ldots, x_k) \in E'$ и вместо $P(x \in E')$ или $P(E')$ использовать запись $P((x_1, \ldots, x_k) \in E')$. Если E' есть k-мерный интервал $(a_1, \ldots, a_k; b_1, \ldots, b_k]$, иначе $E' = (a; b]_k$, то, разумеется, вероятность $P((x_1, \ldots, x_k) \in E')$ может быть также записана в виде $P(a_i < x_i \leqslant b_i,\ i = 1, \ldots, k)$.

Рассматривая k-мерную случайную величину $(x_1, \ldots, x_k)$, мы в некоторых местах предпочтем говорить о ней как о «k-мерной случайной величине с компонентами $x_1, \ldots, x_k$» или как о «случайных величинах $x_1, \ldots, x_k$».

---

*) Конечно, хотелось бы обозначить случайные величины жирными буквами или отметить их с помощью какого-либо другого характерного признака. Однако это вряд ли целесообразно делать в книге, которая содержит большое число приложений теории случайных величин, включающей множество символов (в том числе и классических) для обозначения этих величин. Всякий раз, когда рассматриваемое количество в действительности является случайной величиной, это будет ясно из самого текста.

Часто мы будем заниматься *ограниченными* множествами в R_k. Множество $E' \subset R_k$ является *ограниченным*, если оно содержится в некотором конечном k-мерном интервале $(a;\ b]_k$ (где все $a_1, \ldots, a_k$; $b_1, \ldots, b_k$ конечны). Если случайная величина $(x_1, \ldots, x_k)$ обладает тем свойством, что $P((x_1, \ldots, x_k) \in (a;\ b]_k) = 1$, где $(a;\ b]_k$ — конечный интервал, то говорят, что $(x_1, \ldots, x_k)$ является *ограниченной с вероятностью единица*.

2.2. Функции распределения одномерных случайных величин

Допустим, что x — одномерная случайная величина, вероятностное пространство*) которой есть $(R_1, \mathscr{B}_1, P)$. Мы покажем, что распределение вероятностей в выборочном пространстве R_1 величины x может быть также описано с помощью *функции распределения* $F(x)$, определенной в каждой точке из R_1 и удовлетворяющей некоторым условиям.

Для любого интервала $(-\infty, x']$ на прямой R_1, который, конечно, принадлежит $\mathscr{B}_1$, положим

$$F(x') = P(-\infty < x \leqslant x'). \tag{2.2.1}$$

Ясно, что $F(x)$ — однозначная, вещественная и неотрицательная функция от x на R_1.

Если $x'' > x'$, то из теоремы **1.4.2**

$$F(x'') - F(x') = P(-\infty < x \leqslant x'') - P(-\infty < x \leqslant x') = \\ = P(x' < x \leqslant x'') \geqslant 0. \tag{2.2.2}$$

Следовательно, $F(x)$ — неубывающая функция от x.

Если мы обозначим интервал $(-\infty, \alpha]$ через E_α, то будем иметь следующую стягивающуюся последовательность множеств:

$$E_{-1} \supset E_{-2} \supset \ldots$$

Очевидно, $\lim\limits_{\alpha \to \infty} E_{-\alpha} = \emptyset$. Следовательно, из теоремы **1.4.5**

$$F(-\infty) = \lim_{\alpha \to \infty} F(-\alpha) = \lim_{\alpha \to \infty} P(E_{-\alpha}) = P(\lim_{\alpha \to \infty} E_{-\alpha}) = P(\emptyset) = 0,$$

т. е.

$$F(-\infty) = 0. \tag{2.2.3}$$

Подобным же образом мы имеем расширяющуюся последовательность множеств

$$E_1 \subset E_2 \subset \ldots,$$

*) Начиная с этого места, если не оговорено противное, E (а не E') будет обозначать множество из $\mathscr{B}_1$ (и в более общем случае — из $\mathscr{B}_k$). Кроме того, мы опускаем штрих у P.

для которой $\lim\limits_{\alpha\to\infty} E_\alpha = R_1$. Поэтому из теоремы **1.4.5**

$$F(+\infty) = \lim_{\alpha\to\infty} F(\alpha) = \lim_{\alpha\to\infty} P(E_\alpha) = P(\lim_{\alpha\to\infty} E_\alpha) = P(R_1) = 1,$$

т. е.

$$F(+\infty) = 1. \tag{2.2.4}$$

Теперь рассмотрим убывающую последовательность вещественных чисел $x_1, x_2, \ldots$, предел которой $\lim\limits_{\alpha\to\infty} x_\alpha = x'$. Тогда $E_{x_1}, E_{x_2}, \ldots$ будет стягивающейся последовательностью множеств, имеющей предел, равный $E_{x'}$. Вновь применяя теорему **1.4.5**, мы получаем

$$\lim_{\alpha\to\infty} F(x_\alpha) = \lim_{\alpha\to\infty} P(E_{x_\alpha}) = P(\lim_{\alpha\to\infty} E_{x_\alpha}) = P(E_{x'}) = F(x')$$

или

$$F(x'+0) = F(x'). \tag{2.2.5}$$

Другими словами, функция $F(x)$ является *непрерывной справа* в каждой точке x.

Читатель должен отметить, что (2.2.5) есть следствие исключительно определения $F(x')$ как вероятности, заключенной в полузамкнутом интервале $(-\infty, x']$. Если бы мы предпочли определить $F(x')$ как $P(-\infty < x < x')$, тогда мы бы имели $F(x'-0) = F(x')$, т. е. $F(x)$ была бы непрерывной слева в каждой точке x. Таким образом, определение $F(x')$ как вероятности, заключенной в полузамкнутом интервале $(-\infty, x']$, а не в открытом интервале $(-\infty, x')$ и, как следствие (2.2.5), нужно считать результатом соглашения. В действительности это — весьма распространенное соглашение.

Итак, если x — случайная величина, имеющая вероятностное пространство $(R_1, \mathscr{B}_1, P)$, то существует функция $F(x)$, определенная формулой (2.2.1) (штрихи мы опускаем) в каждой точке $x \in R_1$ и обладающая свойствами (2.2.2) — (2.2.5).

Обратно, если имеется функция $F(x)$, определенная формулой (2.2.1) и удовлетворяющая условиям (2.2.2) — (2.2.5), она задает вероятностное пространство $(R_1, \mathscr{B}_1, P)$. В самом деле, мы можем рассмотреть в качестве начального класса $\mathscr{F}_0$ множеств в R_1 класс полузамкнутых интервалов $(-\infty, x]$, где x — произвольное вещественное число, и назначить интервалу $(-\infty, x]$ вероятность, равную $F(x)$. Этот начальный класс множеств $\mathscr{F}_0$ порождает борелевское поле $\mathscr{B}(\mathscr{F}_0)$, являющееся по определению классом $\mathscr{B}_1$. Используя свойства (2.2.2)—(2.2.5) функции $F(x)$, можно показать, что, исходя из $F(x)$, может быть построена единственная вероятностная мера на борелевском поле, порожденном $\mathscr{F}_0$.

Подытоживая предыдущие рассуждения, мы получаем следующий результат.

2.2.1. *Вероятностное пространство* $(R_1, \mathscr{B}_1, P)$ *случайной величины* x *с помощью формулы* (2.2.1) *определяет в точках*

$x \in R_1$ *единственную однозначную, вещественную и неотрицательную функцию* $F(x)$, *имеющую следующие свойства:*

$$\begin{aligned} &\text{(a)}\ F(x'') - F(x') \geqslant 0, \quad \text{если } x'' > x', \\ &\text{(b)}\ F(-\infty) = 0, \\ &\text{(c)}\ F(+\infty) = 1, \\ &\text{(d)}\ F(x+0) = F(x). \end{aligned} \tag{2.2.6}$$

Обратно, функция $F(x)$, *обладающая этими свойствами, единственным образом определяет вероятностное пространство* $(R_1, \mathscr{B}_1, P)$, *для которого* $P(E_x) = F(x)$.

$F(x)$ называется *функцией распределения* (ф. р.) или *кумулятивной функцией распределения* (к. ф. р.) одномерной случайной величины x. Обычно мы будем использовать второе наименование. Если мы считаем, что общая вероятность, распределяемая по оси x, равна единице, то $F(x)$ есть просто доля вероятности, лежащая на $(-\infty, x]$.

Таким образом, мы располагаем двумя различными путями для описания вероятностей, связанных с одномерной случайной величиной x. Мы можем описывать эти вероятности, во-первых, посредством вероятностного пространства $(R_1, \mathscr{B}_1, P)$ и, во-вторых, посредством к. ф. р. $F(x)$, определенной в каждой точке $x \in R_1$ и удовлетворяющей условиям (2.2.6). Описание с помощью к. ф. р. более удобно при анализе случайных величин в большинстве вопросов и, начиная с этого момента, будет использоваться нами почти всюду.

Следует заметить, что если мы имеем основное выборочное пространство R и случайную величину $x(e)$, определенную в точках $e \in R$, то у этой случайной величины существует к. ф. р. $F(x)$, которая дается выражением

$$P(x(e) \leqslant x) = F(x).$$

Обратно, предположим, что у нас есть некоторая к. ф. р. $F(x)$ безотносительно к какому-либо основному выборочному пространству R. Мы всегда можем определить случайную величину x, к. ф. р. которой есть $F(x)$, рассматривая вещественную ось в качестве основного выборочного пространства R. Выборочными точками будут в этом случае вещественные числа, и, каково бы ни было вещественное x', мы назначим нашей случайной величине $x(e)$ значение x', т. е. положим $x(x') = x'$. Тогда мы будем иметь $P(x(e) \leqslant x') = F(x)$.

Вероятность $P(x \in E)$, где E — произвольное множество из $\mathscr{B}_1$, существует и может быть вычислена на основе $F(x)$. Например, вероятность $P(x' < x \leqslant x'')$ получается из $F(x)$ посредством формулы (2.2.2). Другое полезное вероятностное утверждение относительно случайной величины x, которое может быть проверено читателем с помощью методов, аналогичных использованным при выводе (2.2.5), заключается в следующем. Имеют место формулы:

$$P(x = x') = F(x') - F(x' - 0), \tag{2.2.7}$$

$$P(x' < x < x'') = F(x'' - 0) - F(x'), \tag{2.2.8}$$

$$P(x' \leqslant x < x'') = F(x'' - 0) - F(x' - 0), \tag{2.2.9}$$

$$P(x' \leqslant x \leqslant x'') = F(x'') - F(x' - 0). \tag{2.2.10}$$

Заметим, что, если x — ограниченная случайная величина, то существуют такие конечные числа a и b, $a < b$, что $F(a) = 0$ и $F(b) = 1$; $b - a$ называется *размахом* величины x.

2.3. Общие типы одномерных случайных величин

Большинство одномерных случайных величин, которые встречаются в математической статистике, принадлежит к одному из двух типов: дискретному или непрерывному. Функция распределения $F(x)$ и, разумеется, вероятностная мера $P(E)$ для этих двух типов случайных величин, как мы вскоре увидим, могут быть определены в терминах, если не более простых, то во всяком случае исключающих друг друга функций.

(а) Дискретный тип. К. ф. р. $F(x)$, принадлежащая дискретному типу, является ступенчатой функцией, т. е. $F(x)$ постоянна в окрестности каждой точки $x \in R_1$, за исключением конечного или счетного числа точек $x^{(1)}, x^{(2)}, \ldots$ (в последнем случае эти точки не имеют предельной на конечном расстоянии), где происходит *скачок* или *изменение* $F(x)$. Скачки $p(x^{(1)})$, $p(x^{(2)})$, $\ldots$ даются формулой (2.2.7):

$$p(x^{(\alpha)}) = p(x = x^{(\alpha)}) = F(x^{(\alpha)}) - F(x^{(\alpha)} - 0). \tag{2.3.1}$$

Во всех других точках $x' \in R_1$, $p(x') = P(x = x') = 0$. $F(x)$ может быть выражена с помощью этих скачков следующим образом:

$$F(x) = \sum_{x^{(\alpha)} \leqslant x} p(x^{(\alpha)}), \tag{2.3.2}$$

где суммирование совершается по всем значениям α, для которых $x^{(\alpha)} \leqslant x$. Устремляя x к бесконечности в (2.3.2), мы видим, что должно иметь место равенство

$$F(\infty) = \sum_{\alpha} p(x^{(\alpha)}) = 1. \tag{2.3.3}$$

Поскольку полная вероятность 1 распределяется только по точкам $x^{(1)}$, $x^{(2)}$, $\ldots$, принято говорить об этих точках как о *точках сосредоточения вероятности или точках сосредоточения массы.* Множество точек скачков и образует выборочное пространство случайной величины x, имеющей к. ф. р., указанную в (2.3.2).

При рассмотрении какой-либо дискретной случайной величины x всегда будет ясно из контекста, каковы ее точки сосредоточения массы. Следовательно, не возникает никакой двусмысленности, если мы опустим α и будем просто писать $p(x)$. Функцию $p(x)$ мы назовем *функцией вероятностей* (ф. в.) величины x. Теперь мы можем утверждать следующее.

2.3.1. *К. ф. р. $F(x)$ дискретной случайной величины x единственным образом определяется посредством ф. в. $p(x)$ и обратно.*

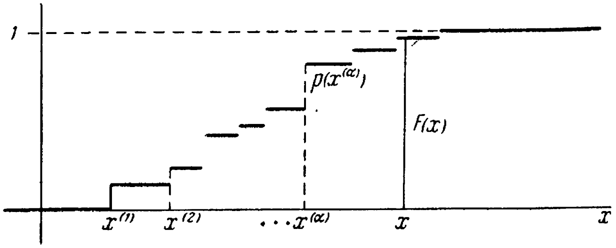

Рис. 2.1. График к. ф. р. $F(x)$ одномерной дискретной случайной величины x.

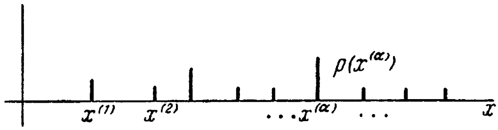

Рис. 2.2. График ф. в. $p(x)$, соответствующей к. ф. р. $F(x)$, изображенной на рис. 2.1.

В задачах, включающих дискретные случайные величины, обычно большее удобство доставляет работа с ф. в. $p(x)$, чем работа с к. ф. р. $F(x)$.

К. ф. р. $F(x)$ одномерной дискретной случайной величины x может быть графически представлена в виде ступенчатой функции, со скачками размеров $p(x^{(1)})$, $p(x^{(2)})$, ... соответственно в точках сосредоточения массы $x^{(1)}$, $x^{(2)}$, ..., как это показано на рис. 2.1.

Ф. в. $p(x)$ случайной величины, к. ф. р. которой изображена на рис. 2.1, может быть представлена графически посредством отрезков вертикальной прямой длин $p(x^{(1)})$, $p(x^{(2)})$, расположенных соответственно в точках сосредоточения массы $x^{(1)}$, $x^{(2)}$, ... Во всех других точках $p(x)$ равна нулю (см. рис. 2.2).

Если $F(x)$ имеет лишь один скачок, например равный $p(x^{(1)})$, происходящий в точке сосредоточения массы $x^{(1)}$, то величина x называется *вырожденной* одномерной случайной величиной; ее к. ф. р. мы обозначим через

$$\varepsilon(x - x^{(1)}) = \begin{cases} 1, & x \geqslant x^{(1)}, \\ 0, & x < x^{(1)}. \end{cases} \qquad (2.3.4)$$

Следует заметить, что к. ф. р. $F(x)$, определенная формулой (2.3.2), может быть выражена через ε-функцию (2.3.4). Мы имеем

$$F(x)=\sum_{\alpha} p(x^{(\alpha)})\cdot\varepsilon(x-x^{(\alpha)}). \tag{2.3.5}$$

Примеры. Число примеров одномерных дискретных случайных величин в элементарной теории вероятностей очень велико. Приведем два из них. Если x — случайная величина, представляющая число очков, выпадающих при бросании одной «правильной» шестигранной игральной кости, то точками сосредоточения массы будут точки $x^{(1)}=1,\ldots, x^{(6)}=6$ и вероятности отдельных точек будут даваться ф. в. $p(x)$, равной $\frac{1}{6}$ при $x=1, \ldots, 6$ и равной 0 при всех других значениях x. Если x — число тузов, встречающихся среди 13 карт, извлеченных «наудачу» из «хорошо перетасованной» колоды, содержащей 52 обычные игральные карты, то $x^{(1)}=0$, $x^{(2)}=1, \ldots, x^{(5)}=4$ и вероятности назначаются с помощью ф. в.

$$p(x)=\frac{\binom{4}{x}\binom{48}{13-x}}{\binom{52}{13}}, \qquad x=0,\ 1,\ 2,\ 3,\ 4,$$

и $p(x)=0$ для всех других значений x.

Дальнейшие примеры важных дискретных случайных величин будут рассмотрены в главе 6.

(b) Непрерывный тип. Для случайной величины, принадлежащей непрерывному типу, всегда существует такая измеримая по Лебегу функция $f(x)\geqslant 0$, что

$$F(x')=\int_{-\infty}^{x'} f(y)\,dy \tag{2.3.6}$$

для всех $x'\in R_1$. В этом случае производная $\frac{dF}{dx}$ существует, причем

$$\frac{dF}{dx}=f(x), \tag{2.3.7}$$

за исключением, быть может, множества значений x вероятности 0.

В действительности удовлетворяющая условиям (2.3.6) и (2.3.7) функция $f(x)$ существует тогда и только тогда, когда $F(x)$ *абсолютно непрерывна*; случайная величина, имеющая абсолютно непрерывную к. ф. р., иногда называется *абсолютно непрерывной случайной величиной.*

Может оказаться, что мы рассматриваем случайную величину x, имеющую просто *непрерывную* к. ф. р. $F(x)$, и здесь уже $F(x)$ не обязана удовлетворять (2.3.6) и (2.3.7). Этот более общий тип случайных величин не следует смешивать с тем, для которого $F(x)$ является *абсолютно непрерывной.* Однако случаи, когда $F(x)$ просто непрерывна и в то же время не абсолютно непрерывна, редки, и поэтому принято опускать наречие «абсолютно» и говорить, что соответствующая величина x есть *непрерывная* случайная величина. Так или иначе, всегда будет ясно из контекста, какой тип непрерывности имеет место в данной ситуации.

Ничего существенного не будет потеряно, если мы положим $f(x)=0$ в точках x, где (2.3.7) нарушается. Тогда $f(x)$ будет единственным образом определяться к. ф. р. $F(x)$ в каждой точке пространства R_1. Рассмотрим выражение

$$\frac{F(x'')-F(x')}{x''-x'}. \tag{2.3.8}$$

Это отношение в согласии с (2.2.6а) положительно и представляет среднее значение вероятности, приходящейся на единицу длины отрезка $(x', x'']$. Если предел этого отношения при $x'' \to x'$ существует, мы получаем $f(x') \geqslant 0$ и $f(x')$ можно понимать как плотность вероятности в точке $x=x'$. Поэтому мы будем называть $f(x)$ *функцией плотности вероятности* (ф. п. в.) случайной величины x.

Представляется полезным подвести итог сказанному.

2.3.2. *К. ф. р. $F(x)$ непрерывной случайной величины x единственным образом определяется своей ф. п. в. $f(x)$ и обратно.*

В теореме **2.3.2** должно подразумеваться, что $f(x)$ необходимо полагается равной 0 во всех точках, где $\frac{dF}{dx}$ не существует.

Если имеет место (2.3.7), то для любого борелевского множества E на прямой R_1

$$P(x \in E)=\int_E f(x)\,dx. \tag{2.3.9}$$

Из (2.3.7) немедленно вытекает, что если $E=R_1$, то $\int\limits_{R_1} f(x)\,dx=1$.

Иногда удобно использовать обычное обозначение дифференциала и писать

$$P(x'<x<x'+dx)=f(x')\,dx, \tag{2.3.10}$$

помня, конечно, о той осторожности, которая должна иметь место при обращении с дифференциалами. Количество $f(x)\,dx$ называется *вероятностным элементом* величины x (в. э.). Если $F(x)$ просто непрерывна, удобно также обозначить $P(x'<x<x'+dx)$ через $dF(x')$ и рассматривать это как в. э. величины x в точке $x=x'$.

Для любого числа p, лежащего в интервале $(0, 1)$, p-я квантиль ($100\,p$-я процентиль) $\underline{x}_p$ непрерывной случайной величины x, имеющей к. ф. р. $F(x)$, определяется как наименьшее число $\underline{x}_p$, для которого

$$F(\underline{x}_p)=p. \tag{2.3.11}$$

В частности, $\underline{x}_{0,5}$ есть *медиана*, а $\underline{x}_{0,25}$ и $\underline{x}_{0,75}$ — соответственно *нижняя квартиль* и *верхняя квартиль* величины x. Для дискретной случайной величины x, если существует по крайней мере одно значение x, для которого $F(x)=p$, p-я квантиль есть наименьшее из таких значений x. Таким образом, квантили случайной величины здесь определяются только как точки сосредоточения массы.

При исследовании задач, включающих непрерывные случайные величины x, большей частью выгоднее пользоваться ф. п. в. $f(x)$, нежели к. ф. р. $F(x)$.

Графически к. ф. р. $F(x)$ одномерной непрерывной случайной величины и соответствующая ей ф. п. в. $f(x)$ представлены соответ-

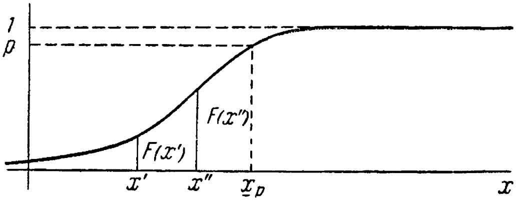

Рис. 2.3. График к. ф. р. $F(x)$ одномерной непрерывной случайной величины x.

ственно на рис. 2.3 и 2.4. На рис. 2.3 значение $P(x' < x \leqslant x'')$ изображается разностью между двумя ординатами $F(x'')$ и $F(x')$, в то время как на рис. 2.4 та же самая вероятность представлена заштрихованной площадью.

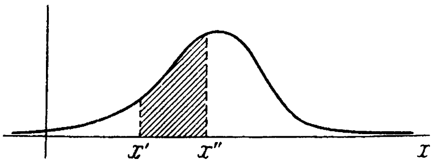

Рис. 2.4. График ф. п. в. $f(x)$ к. ф. р. $F(x)$, изображенной на рис. 2.3.

Пример. С несколькими важными специальными распределениями вероятностей непрерывного типа мы будем иметь дело в главе 7. Однако кажется уместным привести здесь следующий простой пример. Пусть вероятность того, что «случайная» точка попадает внутрь произвольной окружности радиуса δ, в свою очередь лежащей внутри некоторой данной окружности C радиуса r, равна $\frac{\delta^2}{r^2}$ и мы интересуемся расстоянием от «случайной» точки до центра окружности C. Если мы введем в рассмотрение случайную величину x, представляющую это расстояние между «случайной» точкой и центром данной окружности, и для $x' \in R_1$ определим $F(x')$ как вероятность $P(x \leqslant x')$, то к. ф. р. величины x (при условии, что точка обязательно попадет внутрь C) будет иметь вид

$$F(x) \doteq \begin{cases} 1, & x > r, \\ \frac{x^2}{r^2}, & 0 < x \leqslant r, \\ 0, & x \leqslant 0, \end{cases}$$

а ф. п. в. — вид

$$f(x) = \begin{cases} \frac{2x}{r^2}, & 0 < x \leqslant r, \\ 0, & x \leqslant 0, \quad x > r. \end{cases}$$

Можно показать, что наиболее общей формой $F(x)$ является выпуклая комбинация некоторой дискретной к. ф. р. $F_1(x)$ и непрерывной (не обязательно абсолютно непрерывной) к. ф. р. $F_2(x)$, т. е. $F(x) = aF_1(x) + bF_2(x)$, где a, $b \geqslant 0$ и $a + b = 1$.

2.4. Функции распределения двумерных случайных величин

(а) Общие свойства. Допустим, что (x_1, x_2) есть двумерная случайная величина, имеющая вероятностное пространство $(R_2, \mathscr{B}_2, P)$. Пусть $E_{(x_1, x_2)}$ означает интервал $(-\infty, -\infty; x_1, x_2] \subset R_2$ и

$$F(x_1', x_2') = P(E_{(x_1', x_2')}) = P(-\infty < x_1 \leqslant x_1', -\infty < x_2 \leqslant x_2'). \quad (2.4.1)$$

Ясно, что $F(x_1, x_2)$ — однозначная, вещественная и неотрицательная функция точки (x_1, x_2) в R_2.

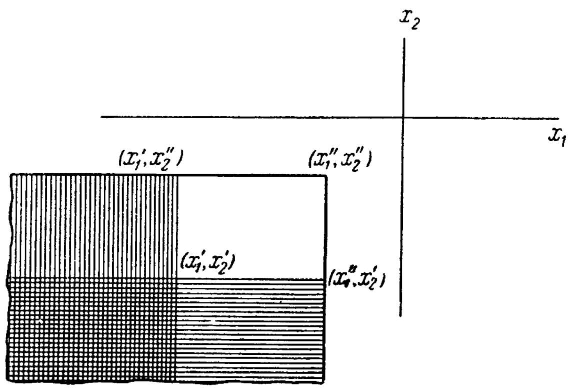

Рис. 2.5. Диаграмма, связанная с формулой для $\Delta_{I_2}^2 F(x_1, x_2)$.

Любой интервал I_2 вида $(x_1', x_2'; x_1'', x_2'']$ принадлежит $\mathscr{B}_2$, так как

$$I_2 = (E_{(x_1'', x_2'')} - E_{(x_1', x_2'')}) - (E_{(x_1'', x_2')} - E_{(x_1', x_2')}). \quad (2.4.2)$$

Кроме того, вероятность события $(x_1, x_2) \in I_2$ должна даваться формулой

$$P((x_1, x_2) \in I_2) = F(x_1'', x_2'') - F(x_1', x_2'') - \\ - F(x_1'', x_2') + F(x_1', x_2'). \quad (2.4.3)$$

Будет удобно обозначать выражение в правой части (2.4.3) через $\Delta_{I_2}^2 F(x_1, x_2)$ и называть его *второй разностью* $F(x_1, x_2)$ *над* I_2. Мы имеем

$$P((x_1, x_2) \in I_2) = \Delta_{I_2}^2 F(x_1, x_2) \geqslant 0. \quad (2.4.4)$$

На рис. 2.5 представлены различные величины, связанные с $\Delta_{I_2}^2 F(x_1, x_2)$. $F(x_1', x_2'')$ и $F(x_1'', x_2')$ суть массы вероятности, содержащиеся в бесконечных областях (включая их границы), заштрихованных соответственно вертикальными и горизонтальными линиями. $F(x_1', x_2')$ есть масса вероятности, содержащаяся в дважды заштрихованной области (включая верхнюю и правую границы), и $\Delta_{I_2}^2 F(x_1, x_2)$ есть масса вероятности, содержащаяся в незаштрихованной прямоугольной области (включая верхнюю и правую границы).

Теперь рассмотрим множества

$$E_{(-\alpha,\ x_2')}, \quad \alpha = 1,\ 2,\ \ldots,$$

очевидно, образующие стягивающуюся последовательность множеств с пределом $\lim\limits_{\alpha\to\infty} E_{(\alpha,\ x_2')} = \emptyset$. Тогда из теоремы **1.4.5** вытекает, что

$$\lim_{\alpha\to\infty} F(-\alpha,\ x_2') = \lim_{\alpha\to\infty} P(E_{(-\alpha,\ x_2')}) = P(\lim_{\alpha\to\infty} E_{(-\alpha,\ x_2')}) = P(\emptyset) = 0.$$

Иначе, опуская штрих,

$$F(-\infty,\ x_2) = 0. \tag{2.4.5}$$

Аналогично

$$F(x_1,\ -\infty) = 0. \tag{2.4.5a}$$

Возьмем множества

$$E_{(\alpha,\ \alpha)}, \quad \alpha = 1,\ 2,\ \ldots$$

Они образуют расширяющуюся последовательность множеств, причем $\lim\limits_{\alpha\to\infty} E_{(\alpha,\ \alpha)} = R_2$. Следовательно, из теоремы **1.4.5** мы получаем

$$\lim_{\alpha\to\infty} F(\alpha,\ \alpha) = \lim_{\alpha\to\infty} P(E_{(\alpha,\alpha)}) = P(\lim_{\alpha\to\infty} E_{(\alpha,\alpha)}) = P(R_2) = 1,$$

т. е.

$$F(+\infty,\ +\infty) = 1. \tag{2.4.6}$$

Рассуждая так же, как при выводе (2.2.5), можно показать, что $F(x_1,\ x_2)$ непрерывна справа по каждой переменной, другими словами, в каждой точке $(x_1,\ x_2) \in R_2$

$$F(x_1+0,\ x_2) = F(x_1,\ x_2+0) = F(x_1,\ x_2). \tag{2.4.7}$$

Из (2.4.7), как легко проверить, следует, что $F(x_1+0,\ x_2+0) = F(x_1,\ x_2)$.

Таким образом, если $(x_1,\ x_2)$ — случайная величина с вероятностным пространством $(R_2,\ \mathscr{B}_2,\ P)$, то существует единственная функция $F(x_1,\ x_2)$, определяемая формулой (2.4.1) (штрихи опущены) в каждой точке $(x_1, x_2) \in R_2$ и обладающая свойствами (2.4.4) — (2.4.7).

Обратно, подобно тому как это было в одномерном случае, функция $F(x_1,\ x_2)$, определенная формулой (2.4.1) и обладающая свойствами (2.4.4) — (2.4.7), единственным образом определяет вероятностное пространство $(R_2,\ \mathscr{B}_2,\ P)$. Собирая вместе высказанные результаты, мы получаем двумерный аналог теоремы **2.2.1.**

2.4.1. *Вероятностное пространство* $(R_2,\ \mathscr{B}_2,\ P)$ *двумерной случайной величины* $(x_1,\ x_2)$ *посредством* (2.4.1) *единственным образом определяет в точках пространства* R_2 *однозначную, веществен-*

ную и неотрицательную функцию $F(x_1, x_2)$, *обладающую следующими свойствами:*

$$\begin{aligned}
&(a) \quad \Delta^2_{I_2} F(x_1, x_2) \geqslant 0,\\
&(b) \quad F(-\infty, x_2) = F(x_1, -\infty) = 0, \qquad (2.4.8)\\
&(c) \quad F(+\infty, +\infty) = 1,\\
&(d) \quad F(x_1+0, x_2) = F(x_1, x_2+0) = F(x_1, x_2).
\end{aligned}$$

Обратно, функция $F(x_1, x_2)$, *обладающая этими свойствами, определяет единственное вероятностное пространство, для которого* $P(E_{(x_1, x_2)}) = F(x_1, x_2)$.

$F(x_1, x_2)$ называется *функцией распределения* или *кумулятивной функцией распределения* двумерной случайной величины (x_1, x_2). Иногда уместно говорить, что $F(x_1, x_2)$ есть к. ф. р. *двух случайных величин* x_1 и x_2. $F(x_1, x_2)$ называется также двумерной к. ф. р.

Как и для одномерной случайной величины, мы имеем две альтернативные схемы для описания распределения вероятностей, соответствующего двумерной случайной величине; в основу одной кладется вероятностное пространство, другая оперирует к. ф. р. Мы будем пользоваться почти исключительно последней схемой.

Различные двумерные аналоги формул (2.2.7) — (2.2.10) могут быть сформулированы и проверены читателем. В частности, имеет место следующее соотношение:

$$\begin{aligned}
P(x_1 = x_1', x_2 = x_2') = F(x_1', x_2') - F(x_1' - 0, x_2') -\\
- F(x_1', x_2' - 0) + F(x_1' - 0, x_2' - 0). \qquad (2.4.9)
\end{aligned}$$

(b) Маргинальные распределения. Рассмотрим множества

$$E_{(x', \alpha)}, \qquad \alpha = 1, 2, \ldots$$

Они образуют расширяющуюся последовательность, так что

$$\lim_{\alpha \to \infty} E_{(x_1', \alpha)} = E_{(x_1', \infty)} = E_{x_1'},$$

где $E_{x_1'}$ есть событие $x_1 \leqslant x_1'$, которое содержится в R_2. Следовательно, применяя теорему **1.4.5**, мы имеем

$$\lim_{\alpha \to \infty} F(x_1', \alpha) = \lim_{\alpha \to \infty} P(E_{(x_1', \alpha)}) = P(\lim_{\alpha \to \infty} E_{(x_1', \alpha)}) = P(E_{x_1'}),$$

т. е.

$$F(x_1', +\infty) = P(E_{x_1'}) = P(x_1 \leqslant x_1'). \qquad (2.4.10)$$

Положим

$$F(x_1, \infty) = F_1(x_1). \qquad (2.4.11)$$

Читатель может убедиться в том, что $F_1(x_1)$ удовлетворяет всем условиям (2.2.6) от (а) до (d), налагаемым на к. ф. р. одномерной случайной величины. В действительности $F_1(x_1)$ является к. ф. р.

компоненты x_1 случайной величины (x_1, x_2) и называется *маргинальной* к. ф. р. величины x_1 или просто к. ф. р. x_1.

Подобным же образом

$$F_2(x_2) = F(\infty, x_2) \tag{2.4.12}$$

есть маргинальная к. ф. р. величины x_2. Если мы представим себе, что полная вероятность, распределяемая по (x_1, x_2)-плоскости в соответствии с к. ф. р. $F(x_1, x_2)$, равна единице, и если эта вероятность ортогонально проектируется на x_1-ось $R_1^{(1)}$, то $F_1(x_1)$ будет количеством вероятности, приходящимся на интервал $(-\infty, x_1]$ x_1-оси. Аналогичная интерпретация допускается, конечно, и для $F_2(x_2)$.

Из теоремы **2.2.1** следует, что маргинальные к. ф. р. $F_1(x_1)$ и $F_2(x_2)$ определяют соответственно вероятностные пространства $(R_1^{(1)}, \mathscr{B}_1^{(1)}, P^{(1)})$ и $(R_1^{(2)}, \mathscr{B}_1^{(2)}, P^{(2)})$.

(c) Статистически независимые случайные величины. Рассмотрим двумерную случайную величину (x_1, x_2) с вероятностным пространством $(R_2, \mathscr{B}_2, P)$. Пусть компоненты этой величины x_1 и x_2 — одномерные случайные величины, имеющие соответственно вероятностные пространства $(R_1^{(1)}, \mathscr{B}_1^{(1)}, P^{(1)})$ и $(R_1^{(2)}, \mathscr{B}_1^{(2)}, P^{(2)})$. Говорят, что x_1 и x_2 *статистически независимы* (см. § 1.6), если для всякого множества E в $R_2 = R_1^{(1)} \times R_1^{(2)}$ вида $E^{(1)} \times E^{(2)}$, где $E^{(1)}$ и $E^{(2)}$ — борелевские множества соответственно в $R_1^{(1)}$ и $R_1^{(2)}$, мы имеем $P(E) = P^{(1)}(E^{(1)}) \cdot P^{(2)}(E^{(2)})$.

Отбрасывание наречия «статистически» и просто указание на *независимые* случайные величины не вызовет никакой неопределенности.

Независимость величин x_1 и x_2 может быть с большей пользой выражена посредством к. ф. р.

2.4.2. *Если (x_1, x_2) — случайная величина, имеющая к. ф. р. $F(x_1, x_2)$, то необходимое и достаточное условие для того, чтобы x_1 и x_2 были независимыми, заключается в выполнении равенства*

$$F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2), \tag{2.4.13}$$

где $F_1(x_1)$ и $F_2(x_2)$ — маргинальные к. ф. р. величин x_1 и x_2.

Чтобы установить теорему **2.4.2**, мы обозначим, как обычно, выборочные пространства величин x_1 и x_2 соответственно через $R_1^{(1)}$ и $R_1^{(2)}$. Тогда

$$R_2 = R_1^{(1)} \times R_1^{(2)}. \tag{2.4.14}$$

Пусть $E_{x_1'}$ есть событие $x_1 \leqslant x_1'$ в $R_1^{(1)}$, а $E_{x_2'}$ — событие $x_2 \leqslant x_2'$ в $R_1^{(2)}$. Очевидно,

$$E_{(x_1', x_2')} = E_{x_1'} \times E_{x_2'}. \tag{2.4.15}$$

Если x_1 и x_2 независимы, то

$$P(E_{(x_1', x_2')}) = P^{(1)}(E_{x_1'}) \cdot P^{(2)}(E_{x_2'}), \tag{2.4.16}$$

т. е.

$$F(x_1', x_2') = F_1(x_1') \cdot F_2(x_2'). \tag{2.4.17}$$

Обратно, допустим, что $F(x_1, x_2)$, $F_1(x_1)$ и $F_2(x_2)$ — к. ф. р., причем (2.4.17) имеет место для каждой точки $(x_1', x_2') \in R_2$, т. е. другими словами, (2.4.16) выполняется для всякого множества $E_{(x_1', x_2')} = E_{x_1'} \times E_{x_2'}$. В таком случае из § 1.6 вытекает, что в R_2 определяется единственное вероятностное пространство $(R_2, \mathscr{B}_2, P)$, обладающее тем свойством, что для любого множества E из $\mathscr{B}_2$ вида $E^{(1)} \times E^{(2)}$ мы имеем $P(E) = P^{(1)}(E^{(1)})\,P^{(2)}(E^{(2)})$, где $(R_1^{(1)}, \mathscr{B}_1^{(1)}, P^{(1)})$ и $(R_1^{(2)}, \mathscr{B}_1^{(2)}, P^{(2)})$ — вероятностные пространства, определенные соответственно $F_1(x_1)$ и $F_2(x_2)$. Последнее равенство эквивалентно независимости x_1 и x_2.

2.5. Общие типы двумерных случайных величин

Большая часть двумерных случайных величин, встречающихся в теории вероятностей и математической статистике, относится к трем типам: дискретному, непрерывному и смешанному. Вместе с тем величины, принадлежащие смешанному типу, встречаются гораздо менее часто, чем величины первых двух типов. Имеет смысл рассмотреть названные типы достаточно подробно.

(а) Дискретный тип. Для *дискретной* случайной величины (x_1, x_2) к. ф. р. является ступенчатой функцией, точнее, для такой величины правая часть формулы (2.4.9) обращается в нуль всюду, за исключением конечного или счетного, но не имеющего предельных точек в R_2, множества *точек сосредоточения массы* $(x_1^{(\alpha)}, x_2^{(\alpha)})$, $\alpha = 1, 2, \ldots$ В этих последних точках

$$P(x_1 = x_1^{(\alpha)},\ x_2 = x_2^{(\alpha)}) = p(x_1^{(\alpha)}, x_2^{(\alpha)}) \tag{2.5.1}$$

и, кроме того,

$$\sum_\alpha p(x_1^{(\alpha)}, x_2^{(\alpha)}) = 1. \tag{2.5.2}$$

Обратно, если имеется такая последовательность точек $(x_1^{(\alpha)}, x_2^{(\alpha)})$, что $p(x_1^{(\alpha)}, x_2^{(\alpha)}) > 0$, $\alpha = 1, 2, \ldots$ и $p(x_1, x_2) = 0$ для всех других точек в R_2, то $F(x_1, x_2)$ есть ступенчатая функция и

$$F(x_1, x_2) = \sum p(x_1^{(\alpha)}, x_2^{(\alpha)}), \tag{2.5.3}$$

где суммирование ведется по всем значениям α, для которых $x_i^{(\alpha)} \leqslant x_i$, $i = 1, 2$.

Случайная величина (x_1, x_2) является *вырожденной*, если существует одна-единственная точка сосредоточения массы $(x_1^{(1)}, x_2^{(1)})$. В этом случае $p(x_1^{(1)}, x_2^{(1)}) = 1$. Мы можем написать

$$F(x_1, x_2) = \varepsilon(x_1 - x_1^{(1)},\ x_2 - x_2^{(2)}) = \begin{cases} 1, & x_i \geqslant x_i^{(1)}, \quad i = 1, 2, \\ 0 & \text{в противном случае.} \end{cases} \tag{2.5.4}$$

Конечно, каждая из компонент (x_1, x_2) в отдельности, например x_1, может быть вырожденной случайной величиной с к. ф. р., определенной в (2.3.4). Мы будем предполагать, если не оговорено противное, что ни одна из компонент (x_1, x_2) вырожденной не является.

Не возникнет никакой путаницы, если мы опустим α и назовем $p(x_1, x_2)$ *функцией вероятностей* (ф. в.) величины (x_1, x_2).

Таким образом, имеет место следующий результат.

2.5.1. *К. ф. р. $F(x_1, x_2)$ дискретной случайной величины (x_1, x_2) единственным образом определяется ф. в. $p(x_1, x_2)$, и обратно.*

При рассмотрении дискретной случайной величины (x_1, x_2) обычно более удобно иметь дело с $p(x_1, x_2)$, нежели с $F(x_1, x_2)$.

Вообще если E — произвольное множество в R_2, то

$$P((x_1, x_2) \in E) = \sum_{(x_1^{(\alpha)},\ x_2^{(\alpha)}) \in E} p(x_1^{(\alpha)}, x_2^{(\alpha)}).$$

В частности, маргинальная к. ф. р. $F_1(x_1)$ в любой точке $x_1 = x_1'$ имеет значение

$$F_1(x_1') = P(x_1 \leqslant x_1') = \sum_{x_1^{(\alpha)} \leqslant x_1'} p(x_1^{(\alpha)}, x_2^{(\alpha)}). \tag{2.5.5}$$

Подобное же выражение определяет и значение $F_2(x_2')$.

Маргинальные к. ф. р. $F_1(x_1)$ и $F_2(x_2)$ являются дискретными одномерными к. ф. р., для которых $p_1(x_1)$ и $p_2(x_2)$ суть соответственно маргинальные ф. в.

Читатель без труда может проверить следующее утверждение.

2.5.2. *Если (x_1, x_2) — дискретная случайная величина с ф. в. $p(x_1, x_2)$, то необходимое и достаточное условие для независимости x_1 и x_2 заключается в том, что $p(x_1, x_2) = p_1(x_1) \cdot p_2(x_2)$.*

Пример. Примеров двумерных дискретных случайных величин в элементарной теории вероятностей очень много. Рассмотрим следующий пример. Пусть x_1 означает число тузов, а x_2 — число королей, появляющихся среди 13 вынутых наудачу игральных карт. Тогда точками сосредоточения массы $(x_1^{(\alpha)}, x_2^{(\alpha)})$, $\alpha = 1, 2, \ldots, 25$, случайной величины (x_1, x_2) будут точки $(0, 0)$, $(0, 1)$, $(1, 0), \ldots, (4, 4)$. При условии, что колода карт перетасована «идеально», т. е. когда всем $\binom{52}{13}$ возможным наборам карт разумно назначить равные вероятности, ф. в. есть

$$p(x_1, x_2) = \frac{\binom{4}{x_1}\binom{4}{x_2}\binom{44}{13 - x_1 - x_2}}{\binom{52}{13}}$$

для $(x_1, x_2)=(0, 0), (0, 1), \ldots, (4, 4)$ и, конечно, $p(x_1, x_2)=0$ для всех других точек в R_2. Маргинальная ф. в. x_1 дается формулой

$$p_1(x_1)=\sum_{x_2=0}^{4} p(x_1, x_2)=\frac{\binom{4}{x_1}\binom{48}{13-x_1}}{\binom{52}{13}}.$$

Эта маргинальная ф. в. является ф. в. случайной величины x_1, числа появившихся тузов. Аналогичное выражение имеет место и для $p_2(x_2)$. Заметим, что x_1 и x_2 не являются статистически независимыми, так как $p(x_1, x_2)\neq p_1(x_1)\cdot p_2(x_2)$.

Другие примеры двумерных дискретных случайных величин будут приведены в главе 6.

(b) Непрерывный тип. Для *непрерывной* двумерной случайной величины (x_1, x_2) существует такая измеримая по Лебегу функция $f(x_1, x_2)\geqslant 0$, что

$$F(x_1, x_2)=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(y_1, y_2)\,dy_1\,dy_2 \tag{2.5.6}$$

при всех $(x_1, x_2)\in R_2$. Одновременно $\frac{\partial^2 F(x_1, x_2)}{\partial x_1\,\partial x_2}$ существует и

$$\frac{\partial^2 F(x_1, x_2)}{\partial x_1\,\partial x_2}=f(x_1, x_2) \tag{2.5.7}$$

всюду в R_2, за исключением, быть может, точек, образующих множество вероятности 0.

Условия, которым должна удовлетворять $F(x_1, x_2)$, чтобы выполнялись формулы (2.5.6) и (2.5.7), есть двумерный аналог условий, указанных для одномерной случайной величины; читателю самому предоставляется разобраться в подробностях.

Полагая $f(x_1, x_2)=0$ в тех точках, где производная $\frac{\partial^2 F(x_1, x_2)}{\partial x_1\,\partial x_2}$ не существует, мы однозначно определяем $f(x_1, x_2)$ в R_2.

Возьмем теперь отношение

$$\frac{\Delta^2_{I_2} F(x_1, x_2)}{(x_1''-x_1')(x_2''-x_2')},\quad x_1''>x_1',\quad x_2''>x_2', \tag{2.5.8}$$

где выражение в числителе доставляется (2.4.4). Ясно, что это отношение дает среднее значение вероятности, приходящейся на единицу площади I_2. Если существует предел отношения (2.5.8) при $x_1''\to x_1'$ и $x_2''\to x_2'$, то он равен $f(x_1', x_2')\geqslant 0$, т. е. имеет смысл неотрицательная плотность в точке (x_1', x_2'). Если в данной точке (x_1, x_2) функция $f(x_1, x_2)$ определяется посредством (2.5.7), мы будем говорить, что *существует плотность вероятности* в этой точке. Во всех других точках R_2 $f(x_1, x_2)=0$, как мы условились по произволу. Будем называть $f(x_1, x_2)$ *функцией плотности вероятности* (ф. п. в.) величины (x_1, x_2) и $f(x_1, x_2)\,dx_1\,dx_2$ — *вероятностным элементом* (в. э.) (x_1, x_2).

Как и в одномерном случае,

2.5.3. *К. ф. р.* $F(x_1, x_2)$ *непрерывной случайной величины* (x_1, x_2) *единственным образом определяется ф. п. в.* $f(x_1, x_2)$ и обратно.

Если E — произвольное борелевское множество в R_2, вероятность $P(E)$ задается следующим лебеговским интегралом:

$$P((x_1, x_2) \in E) = \iint_E f(y_1, y_2)\,dy_1\,dy_2. \tag{2.5.9}$$

В частности, для маргинальной к. ф. р. мы имеем

$$F_1(x_1') = P(x_1 \leqslant x_1') = \int_{-\infty}^{x_1'}\int_{-\infty}^{\infty} f(y_1, y_2)\,dy_1\,dy_2. \tag{2.5.10}$$

Будем рассматривать интеграл в (2.5.10) как повторный интеграл. Тогда функция

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, y_2)\,dy_2 \tag{2.5.11}$$

называется *маргинальной* ф. п. в. величины x_1. Маргинальная ф. п. в. случайной величины x_2 определяется аналогично.

Нижеследующее утверждение доставляет полезный критерий статистической независимости двух случайных величин x_1 и x_2, который будет нетрудно проверить читателю.

2.5.4. *Если* (x_1, x_2) — *непрерывная случайная величина и* $f(x_1, x_2)$ — *ее ф. п. в., то равенство*

$$f(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$$

представляет собой необходимое и достаточное условие статистической независимости x_1 *и* x_2.

Пример. Предположим, что в интервале (0,1) независимо друг от друга «выбираются» два числа, причем все числа имеют «равную вероятность» быть выбранными. В качестве простого примера двумерной непрерывной случайной величины может служить двумерная случайная величина (x_1, x_2), где x_1 и x_2 означают соответственно меньшее и большее из выбранных чисел. Определим ф. п. в. следующим образом: $f(x_1, x_2) = 2$ внутри треугольника на x_1x_2-плоскости с вершинами (0, 0), (1, 1) и (0, 1) и $f(x_1, x_2) = 0$ во всех других точках плоскости. Читатель убедится, что маргинальная ф. п. в. величины x_1 есть

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\,dy_2 = \int_{x_1}^{1} 2 \cdot dy_2 = 2(1 - x_1),$$

и точно так же $f_2(x_2) = 2x_2$.

Ряд специальных случаев более важных двумерных непрерывных случайных величин будет обсуждаться в главе 7.

(c) **Смешанный тип.** Для случайной величины (x_1, x_2) *смешанного* типа одна из компонент дискретна, а другая непрерывна. Более точно, если $F(x_1, x_2)$ — к. ф. р. величины (x_1, x_2), где x_1 дискретна и x_2 непрерывна, и если $x_1^{(\alpha)}$, $\alpha = 1, 2, \ldots$, — точки сосредоточения массы случайной величины x_1, а $p_1(x_1)$ — ее ф. в., то существуют одномерные условные ф. п. в. $f(x_2 \mid x_1^{(\alpha)})$, $\alpha = 1, 2, \ldots$, определяемые формулой

$$f(x_2 \mid x_1^{(\alpha)}) = \frac{1}{p_1(x_1^{(\alpha)})} \cdot \frac{d}{dx_2}[F(x_1^{(\alpha)}, x_2) - F(x_1^{(\alpha)} - 0, x_2)], \tag{2.5.12}$$

такие, что

$$F(x_1, x_2) = \sum_{x_1^{(\alpha)} \leqslant x_1} p_1(x_1^{(\alpha)}) \cdot \int_{-\infty}^{x_2} f(y \mid x_1^{(\alpha)})\,dy. \tag{2.5.13}$$

Маргинальные к. ф. р. $F_1(x_1)$ и $F_2(x_2)$ даются следующими формулами:

$$F_1(x_1) = \sum_{x_1^{(\alpha)} \leqslant x_1} p_1(x_1^{(\alpha)}) \tag{2.5.14}$$

и

$$F_2(x_2) = \sum_{\alpha} p_1(x_1^{(\alpha)}) \cdot \int_{-\infty}^{x_2} f(y \mid x_1^{(\alpha)})\,dy. \tag{2.5.15}$$

Таким образом, можно сформулировать теорему

2.5.5. *К. ф. р. $F(x_1, x_2)$ случайной величины (x_1, x_2) смешанного типа, где x_1 дискретна, а x_2 непрерывна, единственным образом определяется парами $[p(x_1^{(\alpha)}), f(x_2 \mid x_1^{(\alpha)})]$, $\alpha = 1, 2, \ldots$, и обратно.*

В целях большей наглядности со случайной величиной (x_1, x_2) смешанного типа, где x_1 дискретна с ф. в. $p(x_1)$, а x_2 непрерывна с условной ф. п. в. $f(x_2 \mid x_1)$, можно связать следующую интерпретацию. Сначала вероятность 1 подразделяется на отдельные количества, именно $p_1(x_1^{(\alpha)})$, $\alpha = 1, 2, \ldots$, которые помещаются в соответствующие точки сосредоточения массы $x_1^{(\alpha)}$, $\alpha = 1, 2, \ldots$ на x_1-оси. Затем эти количества вероятности непрерывно «размазываются» вдоль вертикальных линий $x_1 = x_1^{(\alpha)}$, $\alpha = 1, 2, \ldots$, так, чтобы в любой точке $(x_1^{(\alpha)}, x_2)$ на α-й линии плотность была равна $p_1(x_1^{(\alpha)}) \cdot f(x_2 \mid x_1^{(\alpha)})$.

Следует отметить, что если $f(x_2 \mid x_1^{(\alpha)})$ для всех α совпадают, то

$$f(x_2 \mid x_1^{(\alpha)}) = f_2(x_2), \tag{2.5.16}$$

где $f_2(x_2)$ — ф. п. в. величины x_2 и $F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2)$, так что x_1 и x_2 независимы. Обратно, если x_1 и x_2 независимы, то $F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2)$ и (2.5.12) приводит к (2.5.16). Следовательно,

2.5.6. *Если (x_1, x_2) — случайная величина смешанного типа, где x_1 дискретна, а x_2 непрерывна, то равенство $f(x_2 \mid x_1^{(\alpha)}) =$*

$=f_2(x_2)$, *где* $f_2(x_2)$ — *маргинальная ф. п. в. величины* x_2, *выполняемое для всех* α, *является необходимым и достаточным условием независимости* x_1 *и* x_2.

Пример. Как утверждалось ранее, примеры случайных величин смешанного типа более редки, чем примеры величин дискретного или непрерывного типа. Однако следующий простой, хотя и несколько искусственный пример в этом направлении, возможно, послужит дальнейшему пониманию природы случайных величин смешанного типа. Предположим, что бросается игральная кость, и пусть x_1 будет случайной величиной, означающей число выпадающих очков. Далее, если оказалось, что $x_1 = x_1^{(\alpha)}$ $(x_1^{(1)} = 1, \ldots, x_1^{(6)} = 6)$, выберем независимо друг от друга $x_1^{(\alpha)}$ чисел из «равномерного» распределения на интервале (0,1). Пусть x_2 означает наибольшее из этих $x_1^{(\alpha)}$ чисел. В таком случае (x_1, x_2) будет случайной величиной смешанного типа, составленной из дискретной величины x_1, имеющей ф. в. $p_1(x_1^{(\alpha)}) = \frac{1}{6}$, $x_1^{(1)} = 1, \ldots, x_1^{(6)} = 6$, и непрерывной величины x_2 такой, что для $0 < x_2 < 1$

$$f(x_2 \mid x_1^{(\alpha)}) = \frac{d}{dx_2}(x_2)^{x_1^{(\alpha)}} = x_1^{(\alpha)}(x_2)^{x_1^{(\alpha)}-1},$$

если $x_1^{(1)} = 1, \ldots, x_1^{(6)} = 6$, и в то же время для всех других значений x_2 вне интервала (0,1) $f(x_2 \mid x_1^{(\alpha)}) = 0$. Распределением, возможно, представляющим интерес в этом примере, является маргинальная к. ф. р. $F_2(x_2)$, которая имеет вид

$$F_2(x_2) = \frac{1}{6}(x_2 + x_2^2 + \ldots + x_2^6).$$

Отсюда мы находим ф. п. в. величины x_2

$$f_2(x_2) = \frac{1}{6}(1 + 2x + \ldots + 6x_2^5).$$

Другие примеры случайных величин смешанного типа встретятся нам в задачах **8.34** и **8.35** в конце главы 8.

2.6. Функции распределения *k*-мерных случайных величин

(а) Общие свойства. Теперь мы покажем, как результаты §§ 2.4 и 2.5, относящиеся к двумерным случайным величинам, могут быть распространены на случай k-мерных величин. Достаточно набросать это обобщение лишь в общих чертах. За более детальным анализом общих многомерных распределений мы отсылаем читателя к фон Нейману (1950).

Допустим, что $(x_1, \ldots, x_k)$ — k-мерная случайная величина и $(R_k, \mathscr{B}_k, P)$ — ее вероятностное пространство. Пусть $E_{(x_1', \ldots, x_k')}$ означает множество $(-\infty, \ldots, -\infty; x_1', \ldots, x_k'] \subset R_k$. Тогда

$$F(x_1', \ldots, x_k') = P(E_{(x_1', \ldots, x_k')}) = P(-\infty < x_i \leqslant x_i', \ i = 1, \ldots, k); \tag{2.6.1}$$

ясно, что $F(x_1, \ldots, x_k)$ — однозначная, вещественная и неотрицательная функция точки $(x_1, \ldots, x_k)$ из R_k.

Любой интервал $I_k \subset R_k$ вида $(x_1', \ldots, x_k'; \; x_1'', \ldots, x_k'']$ принадлежит $\mathscr{B}_k$, так как

$$I_k = E_{(x_1'', \ldots, x_k'')} - [E_{(x_1', x_2'', \ldots, x_k'')} \cup \ldots \cup E_{(x_1'', x_2'', \ldots, x_{k-1}'', x_k')}]. \tag{2.6.2}$$

Кроме того, вероятность включения $(x_1, \ldots, x_k) \in I_k$ может быть найдена с помощью $F(x_1, \ldots, x_k)$ в результате обобщения формулы (2.4.3) на случай k переменных. Мы имеем

$$\begin{aligned} P((x_1, \ldots, x_k) \in I_k) = F(x_1'', \ldots, x_k'') - [F(x_1', x_2'', \ldots, x_k'') + \ldots + \\ + F(x_1'', \ldots, x_{k-1}'', x_k')] + [F(x_1', x_2', x_3'', \ldots, x_k'') + \ldots + \\ + F(x_1'', \ldots, x_{k-2}'', x_{k-1}', x_k')] + \ldots + (-1)^k F(x_1', \ldots, x_k'). \end{aligned} \tag{2.6.3}$$

Правую часть последнего равенства удобно обозначить через $\Delta_{I_k}^k F(x_1, \ldots, x_k)$ и назвать *k-й разностью* функции $F(x_1, \ldots, x_k)$ над I_k. Очевидно,

$$P((x_1, \ldots, x_k) \in I_k) = \Delta_{I_k}^k F(x_1, \ldots, x_k) \geqslant 0. \tag{2.6.4}$$

Посредством аргументов, подобных использованным при выводе (2.4.5), мы получаем

$$F(-\infty, x_2, \ldots, x_k) = \ldots = F(x_1, x_2, \ldots, x_{k-1}, -\infty) = 0. \tag{2.6.5}$$

Также можно обобщить рассуждения, приводящие к (2.4.6), и показать, что

$$F(+\infty, \ldots, +\infty) = 1. \tag{2.6.6}$$

Затем можно установить k-мерный аналог формулы (2.4.7), т. е. формулу

$$F(x_1, \ldots, x_{i-1}, x_i + 0, x_{i+1}, \ldots, x_k) = F(x_1, \ldots, x_k), \; i = 1, \ldots, k, \tag{2.6.7}$$

и проверить, что значение $F(x_1, \ldots, x_k)$ не изменяется, если вместо любого набора x, например, вместо $x_{i_1}, \ldots, x_{i_r}$ взять соответственно $x_{i_1} + 0, \ldots, x_{i_r} + 0$.

Таким образом, если $(x_1, \ldots, x_k)$ — k-мерная случайная величина с вероятностным пространством $(R_k, \mathscr{B}_k, P)$, то существует функция $F(x_1, \ldots, x_k)$, определенная в каждой точке $(x_1, \ldots, x_k) \in R_k$ формулой (2.6.1) и обладающая свойствами (2.6.4) — (2.6.7).

Обратно, если $F(x_1, \ldots, x_k)$ задается посредством (2.6.1) и имеет свойства, выраженные в (2.6.4) — (2.6.7), мы получаем вероятностное пространство $(R_k, \mathscr{B}_k, P)$. Резюмируя сказанное, мы находим k-мерное обобщение теоремы **2.4.1**.

2.6.1. *Вероятностное пространство $(R_k, \mathscr{B}_k, P)$ k-мерной случайной величины $(x_1, \ldots, x_k)$ единственным образом определяет* *однозначную, вещественную и неотрицательную функцию* $F(x_1, \ldots, x_k)$, *задаваемую формулой* (2.6.1) *в каждой точке пространства* R_k *и обладающую следующими свойствами*:

$$
\begin{aligned}
&\text{(a)}\quad \Delta^k_{I_k} F(x_1, \ldots, x_k) \geqslant 0,\\
&\text{(b)}\quad F(-\infty, x_2, \ldots, x_k) = \ldots = F(x_1, \ldots, x_{k-1}, -\infty) = 0,\\
&\text{(c)}\quad F(+\infty, \ldots, +\infty) = 1, \qquad\qquad (2.6.8)\\
&\text{(d)}\quad F(x_1, \ldots, x_{i-1}, x_i + 0, x_{i+1}, \ldots, x_k) =\\
&\qquad\qquad = F(x_1, \ldots, x_k),\ i = 1, \ldots, k.
\end{aligned}
$$

Обратно, $F(x_1, \ldots, x_k)$, *задаваемая равенством* $F(x_1, \ldots, x_k) = P(E_{(x_1, \ldots, x_k)})$ *и обладающая этими свойствами, определяет единственное вероятностное пространство* $(R_k, \mathcal{B}_k, P)$.

Функция $F(x_1, \ldots, x_k)$ называется *функцией распределения* (ф. р.) или *кумулятивной функцией распределения* (к. ф. р.) k-мерной случайной величины $(x_1, \ldots, x_k)$. Иногда $F(x_1, \ldots, x_k)$ называют k-мерной к. ф. р.

Заметим, что вероятность системы равенств $x_1 = x'_1, \ldots, x_k = x'_k$ получается в результате перехода к пределу в правой части (2.6.3) при $x''_i \to x'_i$, $i = 1, \ldots, k$, т. е.

$$P(x_1 = x'_1, \ldots, x_k = x'_k) = \lim_{\substack{x''_k \to x'_i \\ 1 \leqslant i \leqslant k}} \Delta^k_{I_k} F(x_1, \ldots, x_k). \qquad (2.6.9)$$

(b) Маргинальные распределения. Маргинальная к. ф. р. $F_1(x_1)$ случайной величины x_1 определяется следующим образом:

$$F_1(x_1) = F(x_1, +\infty, \ldots, +\infty), \qquad (2.6.10)$$

причем другие маргинальные к. ф. р. $F_i(x_i)$, $i = 2, \ldots, k$, определяются аналогично. Вообще же маргинальная к. ф. р. системы величин $(x_1, \ldots, x_{k_1})$, $k_1 < k$ есть функция

$$F_{1, \ldots, k_1}(x_1, \ldots, x_{k_1}) = F(x_1, \ldots, x_{k_1}, +\infty, \ldots, +\infty). \qquad (2.6.11)$$

Подобное определение имеет место, конечно, и для любого другого подмножества компонент величины $(x_1, \ldots, x_k)$.

(c) Независимость двух и более векторных случайных величин. Пусть $(x_1, \ldots, x_{k_1})$ и $(x_{k_1+1}, \ldots, x_k)$, где $k = k_1 + k_2$, — векторные случайные величины, которые имеют соответственно вероятностные пространства $(R_{k_1}, \mathcal{B}_{k_1}, P^{(1)})$ и $(R_{k_2}, \mathcal{B}_{k_2}, P^{(2)})$. Говорят, что эти две векторные случайные величины *независимы* (см. § 1.6), если для любого множества $E_k \subset R_k = R^{(1)}_{k_1} \times R^{(2)}_{k_2}$ вида $E^{(1)}_{k_1} \times E^{(2)}_{k_2}$, где $E^{(1)}_{k_1}$ и $E^{(2)}_{k_2}$ суть борелевские множества соответственно в $R^{(1)}_{k_1}$ и $R^{(2)}_{k_2}$, мы имеем $P(E_k) = P^{(1)}(E_{k_1}) \cdot P^{(2)}(E_{k_2})$.

Подобно тому как это было сделано в двумерном случае, независимость с большей пользой можно выразить в терминах к. ф. р., что показывает следующая теорема, которая может быть проверена читателем.

2.6.2. *Если* $(x_1, \ldots, x_k)$ — *случайная величина, имеющая к. ф. р.* $F(x_1, \ldots, x_k)$, *то равенство*

$$F(x_1, \ldots, x_k) = F_{1, \ldots, k_1}(x_1, \ldots, x_{k_1}) \cdot F_{k_1+1, \ldots, k}(x_{k_1+1}, \ldots, x_k), \tag{2.6.12}$$

где две функции справа суть маргинальные к. ф. р. соответственно величин $(x_1, \ldots, x_{k_1})$ *и* $(x_{k_1+1}, \ldots, x_k)$, *является необходимым и достаточным условием для независимости* $(x_1, \ldots, x_{k_1})$ *и* $(x_{k_1+1}, \ldots, x_k)$.

Понятие независимости может быть очевидным образом обобщено на случай трех или более попарно не пересекающихся подмножеств компонент величины $(x_1, \ldots, x_k)$. В частности, следует заметить, что $x_1, \ldots, x_k$ являются взаимно независимыми тогда и только тогда, когда

$$F(x_1, \ldots, x_k) = F_1(x_1) \ldots F_k(x_k). \tag{2.6.13}$$

2.7. Общие типы k-мерных случайных величин

Как и в двумерном случае, существует три общих типа k-мерных случайных величин: дискретный, непрерывный и смешанный.

Для *дискретного* типа $F(x_1, \ldots, x_k)$ является ступенчатой функцией такой, что правая часть (2.6.9) обращается в нуль во всех точках R_k, за исключением конечного или счетного, но не имеющего конечных предельных точек в R_k, множества точек $(x_1^{(\alpha)}, \ldots, x_k^{(\alpha)})$, $\alpha = 1, 2, \ldots$, в которых

$$P(x_1 = x_1^{(\alpha)}, \ldots, x_k = x_k^{(\alpha)}) = p(x_1^{(\alpha)}, \ldots, x_k^{(\alpha)}) > 0, \tag{2.7.1}$$

причем

$$\sum_{\alpha} p(x_1^{(\alpha)}, \ldots, x_k^{(\alpha)}) = 1. \tag{2.7.2}$$

Функция $p(x_1, \ldots, x_k)$, которая имеет значение 0 всюду в R_k, за исключением точек $(x_1^{(\alpha)}, \ldots, x_k^{(\alpha)})$, $\alpha = 1, 2, \ldots$, где ее значение дается (2.6.9) (что и выражается в (2.7.1)), есть ф. в. величины $(x_1, \ldots, x_k)$. В случае, если существует только одно значение α, например $\alpha = 1$, то $p(x_1^{(1)}, \ldots, x_k^{(1)}) = 1$ и $(x_1, \ldots, x_k)$ является *вырожденной* случайной величиной. Обобщение формулы (2.5.3) и теоремы **2.5.1** осуществляется непосредственно.

Читатель легко обнаружит, как нужно определить маргинальную ф. в. $p_i(x_i)$, или, более обще, $p_{1, \ldots, k_1}(x_1, \ldots, x_{k_1})$, и как можно переформулировать теорему **2.6.2** в терминах ф. в. Определение вырождения может быть, конечно, дано и для маргинальных распределений. В частности, важно отметить, что $x_1, \ldots, x_k$ *взаимно независимы* тогда и только тогда, когда

$$p(x_1, \ldots, x_k) = p_1(x_1) \ldots p_k(x_k). \tag{2.7.3}$$

В случае *непрерывной* k-мерной случайной величины $(x_1, \ldots, x_k)$ существует такая неотрицательная измеримая по Лебегу функция $f(x_1, \ldots, x_k)$, что

$$F(x_1, \ldots, x_k) = \int\limits_{-\infty}^{x_1} \ldots \int\limits_{-\infty}^{x_k} f(y_1, \ldots, y_k)\, dy_1 \ldots dy_k \tag{2.7.4}$$

для всех $(x_1, \ldots, x_k) \in R_k$, где

$$\frac{\partial^k F(x_1, \ldots, x_k)}{\partial x_1 \ldots \partial x_k}$$

существует в R_k и

$$\frac{\partial^k F(x_1, \ldots, x_k)}{\partial x_1 \ldots \partial x_k} = f(x_1, \ldots, x_k) \tag{2.7.5}$$

всюду в R_k, за исключением, быть может, точек множества вероятности 0.

В соответствии с соглашением, которое мы приняли в одно- и двумерном случаях, мы определяем $f(x_1, \ldots, x_k)$ так, чтобы иметь значение 0 во всех точках R_k, в которых не существует k-я производная левой части (2.7.4).

k-мерное обобщение теоремы **2.5.3** предоставляется читателю.

Маргинальная ф. п. в. $f_i(x_i)$ любой компоненты x_i величины $(x_1, \ldots, x_k)$ или ф. п. в. любого подмножества компонент, например $f_{1, \ldots, k_1}(x_1, \ldots, x_{k_1})$, определяется из соответствующих к. ф. р. очевидным образом.

Легко видеть, что для непрерывных случайных величин теорема **2.6.2** может быть переформулирована в терминах ф. п. в. Кроме того, компоненты $x_1, \ldots, x_k$ случайной величины $(x_1, \ldots, x_k)$ взаимно независимы тогда и только тогда, когда

$$f(x_1, \ldots, x_k) = f_1(x_1) \ldots f_k(x_k). \tag{2.7.6}$$

Если $(x_1, \ldots x_{k_1})$ — k_1-мерная дискретная случайная величина и $(x_{k_1+1}, \ldots, x_k)$, $k = k_1 + k_2$, — k_2-мерная непрерывная случайная величина, то $(x_1, \ldots, x_k)$ является k-мерной случайной величиной *смешанного* типа. Такая случайная величина определяется условной ф. п. в. $f(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1})$, которая дается формулой

$$f(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1}) = \\ = \frac{1}{p_{1, \ldots, k_1}(x_1^{(\alpha)}, \ldots, x_{k_1}^{(\alpha)})} \cdot \frac{\partial^{k_2}}{\partial x_{k_1+1} \ldots \partial x_k} [\Delta_{I_0(k_1)}^{k_1} F(x_1, \ldots, x_k)], \tag{2.7.7}$$

где $p_{1,\ldots,k_1}(x_1, \ldots, x_{k_1})$ есть ф. в. величины $(x_1, \ldots, x_{k_1})$ и

$$\Delta^{k_1}_{I_{0(k_1)}} F(x_1, \ldots, x_k) = \lim_{\substack{x'_1 \to x_1 \\ \cdots \\ x'_{k_1} \to x_{k_1}}} [\Delta^{k_1}_{I_{k_1}} F(x'_1, \ldots, x'_k)]; \qquad (2.7.8)$$

$\Delta^{k_1}_{I_{k_1}} F(x_1, \ldots, x_k)$ представляет собой k_1-ю разность функции $F(x_1, \ldots, x_k)$ относительно $(x_1, \ldots, x_{k_1})$ над k_1-мерным интервалом $(x'_1, \ldots, x'_{k_1}; x_1, \ldots, x_{k_1})$, где $x_{k_1+1}, \ldots, x_k$ считаются фиксированными.

k-мерные аналоги теорем **2.5.5** и **2.5.6** получаются путем непосредственного обобщения. Проверка этого предоставляется читателю в качестве упражнения.

2.8. Функции случайных величин

(а) Функции одной случайной величины. Напомним, что в соответствии с § 2.2 всякая одномерная случайная величина x имеет вероятностное пространство $(R_1, \mathscr{B}_1, P)$ или, что равносильно, к. ф. р. $F(x)$. Часто приходится иметь дело с вероятностной теорией измеримой функции*) $g(x)$ случайной величины x, где $g(x)$ вещественна, однозначна и определена в каждой точке $x \in R_1$, за исключением, быть может, точек, образующих множество вероятности 0, причем множество точек в R_1, для которых $g(x) \leqslant y$, каково бы ни было вещественное число y, также принадлежит классу борелевских множеств $\mathscr{B}_1$ в R_1. Например, если x — случайная величина, то такие элементарные функции, как полиномы от x, $\sin x$, e^x и т. д., очевидно, являются измеримыми.

Из того, что $(R_1, \mathscr{B}_1, P)$ — вероятностное пространство случайной величины x, а также из определения функции $g(x)$ следует, что для любого вещественного числа y множество значений x, для которых $g(x) \leqslant y$, содержится в $\mathscr{B}_1$ и, значит, вероятность, соответствующая этому множеству, доставляется пространством $(R_1, \mathscr{B}_1, P)$.

Обозначим вероятность неравенства $g(x) \leqslant y$ при фиксированном y через $H(y)$. Тогда

$$H(y) = P(g(x) \leqslant y) = P(x \in E_y), \qquad (2.8.1)$$

где E_y — множество точек $x \in R_1$, для которых $g(x) \leqslant y$.

Легко можно проверить, что $H(y)$ обладает всеми свойствами к. ф. р., и, следовательно, мы имеем

2.8.1. *Если x — некоторая случайная величина и $g(x) = y$ — новая случайная величина, то $H(y)$, определенная соотношением (2.8.1), является к. ф. р. величины y.*

Теперь допустим, что $g(x)$ строго монотонна и имеет непрерывную, не обращающуюся в нуль производную при всех x из некоторого открытого интервала A. Возьмем $y = g(x)$, и пусть B будет

*) Такие функции также называются бэровскими функциями.

образ (интервал) интервала A в пространстве y. Выберем $x' \in A$ и найдем $y' = g(x')$. Тогда для некоторого $\Delta y > 0$ существует единственное решение уравнения $g(x) = y' + \Delta y$, которое мы обозначим так: $x = g^{-1}(y' + \Delta y)$.

Если к. ф. р. величины x равна $F(x)$, то мы получаем

$$H(y' + \Delta y) - H(y') = \pm [F(g^{-1}(y' + \Delta y)) - F(g^{-1}(y'))], \quad (2.8.2)$$

причем знак плюс имеет место тогда, когда $g(x)$ монотонно возрастает, и знак минус, когда эта функция монотонно убывает.

Деля обе части (2.8.2) на Δy, мы можем написать

$$\frac{H(y' + \Delta y) - Hy'}{\Delta y} =$$

$$= \left[\frac{F(g^{-1}(y' + \Delta y)) - F(g^{-1}(y'))}{g^{-1}(y' + \Delta y) - g^{-1}(y')}\right] \left[\frac{\pm (g^{-1}(y' + \Delta y) - g^{-1}(y'))}{\Delta y}\right]. \quad (2.8.3)$$

Если $g(x)$ имеет в точке $x = x'$ не образующуюся в нуль производную, то такую же производную имеет в точке $y = y'$ и функция $g^{-1}(y)$. Переходя в (2.8.3) к пределу при $\Delta y \to 0$, предполагая затем, что $f(x)$ есть производная от $F(x)$, если $x \in A$, и опуская штрихи, мы находим

$$\frac{dH(y)}{dy} = f(x) \left| \frac{dx}{dy} \right|,$$

где справа x должно быть заменено на $g^{-1}(y)$. Здесь мы применяем символ абсолютной величины, чтобы упростить правило выбора знака. Кроме того,

$$\int_A f(x)\, dx = \int_B f(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| dy. \quad (2.8.4)$$

Итак, мы имеем теорему

2.8.2. *Предположим, что x — непрерывная случайная величина с ф. п. в. $f(x)$ и $y = g(x)$ строго монотонна и имеет непрерывную, не обращающуюся в нуль производную в некотором открытом интервале A. Пусть B — образ A в пространстве y. Если $y = g(x)$, то y есть непрерывная случайная величина, ф. п. в. $h(y)$ которой существует в B и дается формулой*

$$h(y) = f(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|, \quad (2.8.5)$$

где $g^{-1}(y)$ есть решение относительно x уравнения $g(x) = y$. Кроме того, имеет место (2.8.4).

Иногда мы будем пользоваться выражением: «преобразование $x = g^{-1}(y)$ переводит вероятностный элемент $f(x)\,dx$ в вероятностный элемент $h(y)\,dy$» или просто писать: «$f(x)\,dx \to h(y)\,dy$», где $h(y)$ определяется (2.8.5). Формула (2.8.5) представляет, конечно, хорошо известное правило замены переменной интегрирования в подынтегральном выражении (здесь функция плотности вероятности) обычного интеграла.

Пример. Чтобы пояснить правило, выражаемое (2.8.5), допустим, что x — непрерывная случайная величина, имеющая ф. п. в. $f(x)=\frac{1}{a}$ в интервале $(0, a)$ и $f(x)=0$ в других точках x.

Пусть желательно найти ф. п. в. случайной величины x^n. В этом случае $g(x)=x^n$ и $g^{-1}(y)=y^{1/n}$. Следовательно, по формуле (2.8.5)

$$h(y)=\frac{1}{n}y^{\frac{1}{n}-1}$$

для $y \in (0, a^n)$ и $h(y)=0$ в противном случае, т. е. для преобразования $x=y^{1/n}$ мы имеем

$$\frac{1}{a}dx \to \frac{1}{a}\left|\frac{dy^{1/n}}{dy}\right|dy=\frac{1}{na}y^{\frac{1}{n}-1}dy.$$

(b) Функции двух случайных величин. Пусть дана двумерная случайная величина (x_1, x_2). Функция $g(x_1, x_2)$ от x_1 и x_2 называется случайной величиной, если $g(x_1, x_2)$ вещественна, однозначна и определена в каждой точке R_2 x_1x_2-плоскости, за исключением, быть может, точек множества вероятности 0, и если для каждого вещественного y множество точек в R_2, для которых $g(x_1, x_2)\leqslant y$, принадлежит борелевскому классу множеств $\mathscr{B}_2$ в R_2. Если $g(x_1, x_2)$ — случайная величина, мы обозначим ее через y. Пусть E_y — множество точек в x_1x_2-плоскости, где $g(x_1, x_2)\leqslant y$.

Так как случайная величина характеризуется вероятностным пространством $(R_2, \mathscr{B}_2, P)$ или даже просто к. ф. р. $F(x_1, x_2)$, то это пространство (и, конечно, также $F(x_1, x_2)$) доставляет и вероятность включения $(x_1, x_2)\in E_y$ для любого вещественного числа y. Обозначим эту вероятность через $H(y)$. Мы имеем

$$H(y)=P(g(x_1, x_2)\leqslant y)=P((x_1, x_2)\in E_y). \qquad (2.8.6)$$

Можно проверить, что $H(y)$ есть к. ф. р., точнее, к. ф. р. случайной величины $y=g(x_1, x_2)$.

Рассмотрим векторную функцию $(g_1(x_1, x_2), g_2(x_1, x_2))$ случайной величины (x_1, x_2), где $g_1(x_1, x_2)$ и $g_2(x_1, x_2)$ вещественны, однозначны и определены в каждой точке R_2, за исключением, быть может, точек множества нулевой вероятности, причем множество точек на x_1x_2-плоскости, для которых $g_i(x_1, x_2)\leqslant y_i$, $i=1, 2$, принадлежит $\mathscr{B}_2$, какова бы ни была пара вещественных чисел (y_1, y_2). Обозначим последнее множество точек при данных y_1 и y_2 через $E_{(y_1, y_2)}$. Тогда вероятность включения $(x_1, x_2)\in E_{(y_1, y_2)}$ определяется вероятностным пространством $(R_2, \mathscr{B}_2, P)$ случайной величины (y_1, y_2). Пусть эта вероятность равна $H(y_1, y_2)$. Очевидно,

$$H(y_1, y_2)=P(g_i(x_1, x_2)\leqslant y_i, i=1, 2)=P((x_1, x_2)\in E_{(y_1, y_2)}). \qquad (2.8.7)$$

Легко убедиться, что $H(y_1, y_2)$ обладает всеми свойствами двумерной к. ф. р. и, следовательно,

2.8.3. *Если* (x_1, x_2) — *двумерная случайная величина и* (y_1, y_2) *обозначает случайную величину* $(g_1(x_1, x_2), g_2(x_1, x_2))$, *то* $H(y_1, y_2)$, *определенная посредством* (2.8.7), *есть к. ф. р. величины* (y_1, y_2).

Читатель должен заметить, что к. ф. р. какой-либо одной из компонент величины (y_1, y_2), например y_1, определяется следующим образом:

$$H_1(y_1) = P(g_1(x_1, x_2) \leqslant y_1) = P((x_1, x_2) \in E_{y_1}), \qquad (2.8.8)$$

где E_{y_1} есть множество точек в x_1x_2-плоскости, для которых $g_1(x_1, x_2) \leqslant y_1$.

Важным обстоятельством, касающимся $H_1(y_1)$, является то, что хотя эта функция в действительности представляет собой маргинальную к. ф. р. величины y_1, определяемую в результате перехода в к. ф. р. $H(y_1, y_2)$ к пределу при $y_2 \to \infty$, она может быть получена без предварительного определения $H(y_1, y_2)$. В самом деле (2.8.8) показывает, что введение случайной величины $y_2 = g_2(x_1, x_2)$ не имеет смысла, если мы интересуемся только величиной y_1.

Компоненты двумерной случайной величины (x_1, x_2) *линейно зависимы*, если существуют не равные нулю одновременно вещественные постоянные c_1 и c_2 такие, что случайная величина $c_1x_1 + c_2x_2$ является вырожденной (обычно очень небольшой интерес представляет случай, в котором либо c_1, либо c_2 равно нулю, ибо в этой ситуации одна из компонент x_1 или x_2 сама является вырожденной). Если $c_1x_1 + c_2x_2$ — вырожденная величина, но ни c_1, ни c_2 не равно нулю, мы будем говорить, что x_1 и x_2 *собственно линейно зависимы*; это означает, что $P(c_1x_1 + c_2x_2 = c_3) = 1$, где c_3 — некоторая константа и прямая $c_1x_1 + c_2x_2 = c_3$ не является параллельной какой-либо координатной оси. Если мы исключим случай, когда одна или обе компоненты величины (x_1, x_2) вырождены, станет ясно, что единственный тип линейной зависимости, который может иметь место, есть тип собственной линейной зависимости. Если x_1 и x_2 не являются линейно зависимыми, то говорят, что они *линейно независимы*.

В частном случае, когда $x_1 - x_2$ есть такая вырожденная случайная величина, что

$$P(x_1 - x_2 = 0) = 1,$$

x_1 и x_2 называются *эквивалентными случайными величинами*. Если $F(x_1, x_2)$ есть к. ф. р. двух эквивалентных случайных величин x_1 и x_2, то $F_1(x_1)$ и $F_2(x_2)$ тождественно совпадают. Обратное утверждение было бы, конечно, неверным.

Две векторные случайные величины эквивалентны, если их соответствующие компоненты являются эквивалентными случайными величинами.

(с) Непрерывные функции двух непрерывных случайных величин. Важный специальный случай возникает, когда (x_1, x_2) есть непрерывная случайная величина и преобразование $y_i = g_i(x_1, x_2)$,

$i=1, 2$, взаимно однозначно отображает x_1x_2-плоскость на y_1y_2-плоскость. В этом случае при некоторых условиях регулярности мы можем дать точное выражение для ф. п. в. величины (y_1, y_2) через ф. п. в. величины (x_1, x_2) и первые частные производные от $g_1(x_1, x_2)$ и $g_2(x_1, x_2)$, именно, можно сформулировать следующее утверждение:

2.8.4. *Предположим, что (x_1, x_2) есть непрерывная случайная величина с ф. п. в. $f(x_1, x_2)$. Пусть $g_i(x_1, x_2)$, $i=1, 2$ — однозначные функции, имеющие непрерывные первые частные производные в некоторой открытой области A x_1x_2-плоскости. Пусть $y_i = g_i(x_1, x_2)$, $i=1, 2$, допускают единственное решение $x_i = g_i^{-1}(y_1, y_2)$, $i=1, 2$, для всех точек в A, B — образ A, лежащий в y_1y_2-плоскости, и J — якобиан, определяемый соотношением*

$$J=\left|\frac{\partial x_i}{\partial y_j}\right|, \qquad i, j=1, 2, \tag{2.8.9}$$

имеющий во всех точках A значения, отличные от нуля. Ф. п. в. $h(y_1, y_2)$ величины (y_1, y_2) в любой точке $(y_1, y_2)\in B$ дается формулой

$$h(y_1, y_2)=f(x_1, x_2)\,|J|, \tag{2.8.10}$$

где в правой части (x_1, x_2) должна быть заменена точкой

$$(g_1^{-1}(y_1, y_2),\ g_2^{-1}(y_1, y_2)).$$

Кроме того,

$$\int\limits_A f(x_1, x_2)\,dx\,dy=\int\limits_B f(x_1, x_2)\,|J|\,dy_1\,dy_2. \tag{2.8.11}$$

Теорема 2.8.4 представляет собой просто сформулированный в терминах функций плотности вероятности вариант хорошо известной теоремы (которая может быть найдена в курсах математического анализа) о замене переменных в двойном интеграле. Читатель, интересующийся подробностями доказательства, может обратиться, например, к Уиддеру (1947).

Пример. Допустим, что (x_1, x_2) — случайная величина, имеющая ф. п. в.

$$f(x_1, x_2)=\begin{cases} e^{-x_1-x_2}, & x_1\geqslant 0,\ x_2\geqslant 0,\\ 0 & \text{в противном случае,}\end{cases}$$

и нам нужно найти ф. п. в. случайной величины $\left(x_1+x_2, \dfrac{x_1}{x_2}\right)$. Здесь мы имеем преобразование

$$y_1=x_1+x_2, \qquad y_2=\frac{x_2}{x_1}$$

и обратное преобразование

$$x_1=\frac{y_1}{1+y_2}, \qquad x_2=\frac{y_1y_2}{1+y_2},$$

которые устанавливают взаимно однозначное соответствие между точками положительного квадранта x_1x_2-плоскости и точками положительного квадранта y_1y_2-плоскости. Абсолютная величина якобиана рассматриваемого отображения для всех точек в положительном квадранте равна

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \frac{y_1}{(1+y_2)^2}.$$

Таким образом, для ф. п. в. величины (y_1, y_2) мы получаем

$$h(y_1, y_2) = \begin{cases} e^{-y_1}\dfrac{y_1}{(1+y_2)^2}, & y_1 > 0, \quad y_2 > 0, \\ 0 \text{ в противном случае.} \end{cases}$$

Попутно следует заметить, что в данном случае y_1 и y_2 (см. теорему **2.5.4**) являются независимыми случайными величинами.

(d) Функции k-мерных случайных величин. Предшествующие результаты естественным образом распространяются на большее число функций нескольких случайных величин. Если $(x_1, \ldots, x_k)$ — k-мерная случайная величина, то функция $g(x_1, \ldots, x_k)$, которую мы обозначим через y, будет также случайной величиной, если она вещественна, однозначна, определена во всех точках R_k, за исключением, быть может, точек множества нулевой вероятности, и если множество точек в R_k (x-пространстве), для которых $g(x_1, \ldots, x_k) \leqslant y$, принадлежит $\mathcal{B}_k$ — борелевским множествам в R_k — при всех вещественных значениях y. Если для данного y мы обозначим последнее множество точек через E_y, то к. ф. р. величины y будет равна

$$H(y) = P((x_1, \ldots, x_k) \in E_y). \tag{2.8.12}$$

Определение линейной зависимости также обобщается на случай k-мерной случайной величины. Величины $x_1, \ldots, x_k$ являются *линейно зависимыми*, если $c_1x_1 + \ldots + c_kx_k$ при некоторых константах $c_1, \ldots, c_k$, не обращающихся в нуль одновременно, представляет собой вырожденную случайную величину. Если ни одна из констант c_i не есть нуль, мы имеем *собственную линейную зависимость*. Если $x_1, \ldots, x_k$ не являются линейно зависимыми, говорят, что они *линейно независимы*.

Векторная функция $(g_1(x_1, \ldots, x_k), \ldots, g_{k_1}(x_1, \ldots, x_k))$ случайной величины $(x_1, \ldots, x_k)$, которую мы будем обозначать $(y_1, \ldots, y_{k_1})$, есть также k_1-мерная случайная величина, если каждая из компонент $g_1(x_1, \ldots, x_k), \ldots, g_{k_1}(x_1, \ldots, x_k)$ вещественна, однозначна и определена во всех точках $(x_1, \ldots, x_k) \in R_k$, за исключением, быть может, точек множества вероятности 0, и если множество точек $E_{(y_1, \ldots, y_{k_1})} \subset R_k$, для которых $g_l(x_1, \ldots, x_k) \leqslant y_l$, $l = 1, \ldots, k_1$ при любом наборе вещественных чисел $y_1, \ldots, y_{k_1}$ принадлежит $\mathcal{B}_k$. Таким образом, вероятность события $E_{(y_1, \ldots, y_{k_1})}$ определяется вероятностным пространством $(R_k, \mathcal{B}_k, P)$ или же к. ф. р. $F(x_1, \ldots, x_k)$ случайной величины $(x_1, \ldots, x_k)$. Функция $H(y_1, \ldots, y_{k_1})$, определяемая формулой

$$H(y_1, \ldots, y_{k_1}) = P((x_1, \ldots, x_k) \in E_{(y_1, \ldots, y_{k_1})}), \tag{2.8.13}$$

есть к. ф. р. величины $(y_1, \ldots, y_{k_1})$.

Теперь предположим, что $k_1 = k$ и $(x_1, \ldots, x_k)$ — непрерывная случайная величина с ф. п. в. $f(x_1, \ldots, x_k)$. Пусть в некоторой открытой области A пространства x уравнения $y_i = g_i(x_1, \ldots, x_k)$, $i = 1, \ldots, k$, однозначно разрешимы относительно $x_1, \ldots, x_k : x_i = g_i^{-1}(y_1, \ldots, y_k)$, $i = 1, \ldots, k$, где $g_i(x_1, \ldots, x_k)$ обладают непрерывными первыми производными, причем якобиан $J = \left|\frac{\partial x_i}{\partial y_i}\right| \neq 0$ в A. Обозначим через B образ A в пространстве $(y_1, \ldots, y_k)$. Тогда $(y_1, \ldots, y_k)$ будет непрерывной случайной величиной, имеющей ф. п. в., которая в точке $(y_1, \ldots, y_k) \in B$ равна

$$h(y_1, \ldots, y_k) = f(x_1, \ldots, x_k) \cdot |J|, \tag{2.8.14}$$

и, кроме того,

$$\int_A f(x_1, \ldots, x_k)\, dx_1 \ldots dx_k = \int_B f(x_1, \ldots x_k)\, |J|\, dy_1 \ldots dy_k. \tag{2.8.15}$$

Эти формулы представляют, конечно, k-мерное обобщение формул (2.8.10) и (2.8.11). Подразумевается, что $x_1, \ldots, x_k$ в правой части (2.8.14) и (2.8.15) должны быть выражены через $y_1, \ldots, y_k$, именно, $x_i = g_i^{-1}(y_1, \ldots, y_{k_1})$, $i = 1, \ldots, k$.

2.9. Условные функции распределения

(а) Общие замечания. В § 1.9 условная вероятность была определена для событий из общего выборочного пространства R. Так как основной интерес представляют события, определяемые случайными величинами, полезно видоизменить идеи параграфа 1.10 для случая, когда события описываются случайными величинами. Допустим, что x есть случайная величина с вероятностным пространством $(R_1, \mathscr{B}_1, P)$ или, что равносильно, с к. ф. р. $F(x)$. Пусть G — произвольное событие из $\mathscr{B}_1$, для которого $P(G) > 0$ и $E_{x'}$ означает событие $x \leqslant x'$. Положим

$$F(x' \mid G) = \frac{P(E_{x'} \cap G)}{P(G)}. \tag{2.9.1}$$

Читатель может проверить, что $F(x \mid G)$ удовлетворяет всем условиям (2.2.6) для к. ф. р.; эта функция называется *условной к. ф. р. величины x при условии, что $x \in G$*. Обратно, с помощью теоремы **2.2.1** функция $F(x \mid G)$ единственным образом определяет *условное* вероятностное пространство $(R_1, \mathscr{B}_1, P)$.

Аналогично пусть $(x_1, \ldots, x_k)$ — k-мерная случайная величина с к. ф. р. $F(x_1, \ldots, x_k)$. Мы можем определить условную к. ф. р. величины $(x_1, \ldots, x_k)$ при условии, что $(x_1, \ldots, x_k) \in G$, где G — борелевское множество в R_k, формулой

$$F(x_1, \ldots, x_k \mid G) = \frac{P(E_{(x_1, \ldots, x_k)} \cap G)}{P(G)}, \tag{2.9.2}$$

причем $P(G) > 0$ и $E_{(x_1, \ldots, x_k)}$ есть k-мерный интервал $(-\infty, x]_k$.

В действительности мы будем интересоваться (2.9.2), имея в виду главным образом ситуации, в которых событие G состоит из *сечений* $(x_1, \ldots, x_k)$-пространства, получаемых в результате закрепления одной или более компонент величины $(x_1, \ldots, x_k)$.

Одновременное обращение в нуль как числителя, так и знаменателя в правой части (2.9.2) приводит нас к трудностям, как, например, в случае, когда $(x_1, \ldots, x_k)$ есть k-мерная непрерывная случайная величина с ф. п. в. $f(x_1, \ldots, x_k)$. Рассмотрению этих трудностей посвящены последующие параграфы.

(b) Условные функции распределения двумерных случайных величин. Рассмотрим две случайные величины x_1 и x_2. Пусть I_1 — цилиндрическое множество в R_2, для которого $x_1'' < x_1 < x_1'$, и $P(I_1) > 0$. Тогда из (2.9.2) мы имеем

$$F(x_1, x_2 \mid I_1) = \frac{P(E_{(x_1, x_2)} \cap I_1)}{P(I_1)}. \tag{2.9.3}$$

Если $F(x_1, x_2)$ есть к. ф. р. величины (x_1, x_2) и $F_1(x_1)$ — маргинальная к. ф. р. величины x_1, то $F(x_1', x_2 \mid I_1)$ можно записать в виде

$$F(x_1, x_2 \mid I_1) = \frac{F(x_1', x_2) - F(x_1'', x_2)}{F_1(x_1') - F_1(x_1'')}. \tag{2.9.4}$$

Заметим, что $F(x_1, x_2 \mid I_1)$ как функция от x_2 есть к. ф. р. Если существует предел правой части при $x_1'' \to x_1'$, обозначим его через $F(x_2 \mid x_1')$, т. е. положим

$$\lim_{x_1'' \to x_1'} F(x_1, x_2 \mid I_1) = F(x_2 \mid x_1'). \tag{2.9.5}$$

Если $F(x_2 \mid x_1')$ существует для каждого x_2, можно проверить, что $F(x_2 \mid x_1')$, рассматриваемая как функция от x_2, есть к. ф. р., т. е. она обладает основными свойствами, перечисленными в (2.2.6). Она называется условной к. ф. р. величины x_2 при условии, что $x_1 = x_1'$. Для краткости, если только исключена возможность внести какую-либо двусмысленность, можно, опуская штрих, говорить, что $F(x_2 \mid x_1)$ есть к. ф. р. условной случайной величины $x_2 \mid x_1$. $F(x_2 \mid x_1)$ есть к. ф. р., для которой x_1, по существу, играет роль параметра; x_1 иногда называется *закрепленной* переменной, чтобы подчеркнуть, что при определении $F(x_2 \mid x_1)$ она не является случайной величиной. Символ $x_2 \mid x_1$ можно рассматривать как одномерную случайную величину, чья к. ф. р. $F(x_2 \mid x_1)$ при фиксированном значении x_1, например x_1', определяется во всех точках x_1x_2-плоскости, где $x_1 = x_1'$, таким образом, что, грубо говоря, $F(x_2' \mid x_1')$ представляет количество вероятности, расположенное вдоль части прямой $x_1 = x_1'$, для которой $x_2 \leqslant x_2'$, и выражается в виде доли вероятности от общего ее количества, лежащего вдоль всей линии.

Предшествующее обсуждение доставляет довольно элементарный подход к условным случайным величинам и их к. ф. р., являющийся достаточным для почти всех распределений, встречающихся в математической статистике. Более общий подход может быть развит в результате применения теоремы Радона — Никодима **1.10.1.**

В случае двух измерений особого внимания заслуживают два важных типа условных случайных величин.

Тип А. Для этого типа компонента x_1 случайной величины (x_1, x_2) дискретна. Таким образом, в (2.9.2) при $k=2$ мы считаем x_1 дискретной случайной величиной с точками сосредоточения массы $x_1^{(\alpha)}$, $\alpha=1, 2, \ldots$ Пусть G — множество в R_2 — выборочном пространстве величины (x_1, x_2), для которого $x_1=x_1^{(\beta)}$. Тогда

$$P(G)=P(x_1=x_1^{(\beta)})=p(x_1^{(\beta)})>0.$$

Используя (2.9.2) при $k=2$, мы сразу находим

$$F(x_1^{(\beta)}, x_2 \mid x_1=x_1^{(\beta)})=\frac{P(E_{(x_1^{(\beta)}, x_2)} \cap (x_1=x_1^{(\beta)}))}{P(x_1=x_1^{(\beta)})}. \qquad (2.9.6)$$

Эту величину мы будем обозначать через $F(x_2 \mid x_1^{(\beta)})$. Функция $F(x_2 \mid x_1^{(\beta)})$ есть к. ф. р., и мы будем называть ее к. ф. р. *условной случайной величины* $x_2 \mid x_1^{(\beta)}$. Таким образом, для условной случайной величины типа А можно определить $F(x_2 \mid x_1^{(\beta)})$ непосредственно из (2.9.2) при $k=2$, беря G равным множеству точек в R_2, для которых $x_1=x_1^{(\beta)}$. Формула (2.9.5) также приводит к полученному результату для $F(x_2 \mid x_1^{(\beta)})$, если в определении I_1 выбрать $x_1'=x_1^{(\beta)}$.

Заметим, что $x_2 \mid x_1^{(\beta)}$ есть случайная величина, выборочное пространство которой располагается на прямой в x_1x_2-плоскости, задаваемой уравнением $x_1=x_1^{(\beta)}$. К. ф. р. величины $x_2 \mid x_1^{(\beta)}$ есть $F(x_2 \mid x_1^{(\beta)})$. Если это не вносит путаницы, мы можем отбросить β и рассматривать $F(x_2 \mid x_1)$ как к. ф. р. условной случайной величины $x_2 \mid x_1$.

Тип А условных случайных величин включает два общих случая. Самым общим случаем является тот, в котором величина x_2, так же как и x_1, дискретна. В этом случае, очевидно,

$$F(x_2 \mid x_1^{(\beta)})=\sum_{x_2^{(\nu)} \leq x_2} \frac{p(x_1^{(\beta)}, x_2^{(\nu)})}{p_1(x_1^{(\beta)})}, \qquad (2.9.7)$$

где $p(x_1, x_2)$ — ф. в. величины (x_1, x_2) и $p_1(x_1)$ — маргинальная ф. в. величины x_1. Удобно опустить верхние индексы и положить

$$p(x_2 \mid x_1)=\frac{p(x_1, x_2)}{p_1(x_1)}, \qquad (2.9.8)$$

так что $p(x_2 \mid x_1)$ будет ф. в. условной случайной величины $x_2 \mid x_1$. Заметим, что $x_2 \mid x_1$ будет вырожденной случайной величиной, если не найдется по крайней мере двух различных точек сосредоточения массы, имеющих данную координату x_1.

Формула (2.9.8) может быть переписана следующим образом:

$$p(x_1, x_2)=p(x_2 \mid x_1) \cdot p_1(x_1). \qquad (2.9.9)$$

Это доставляет метод нахождения ф. в. величины (x_1, x_2) в два этапа, т. е. посредством определения ф. в. величин $x_2 | x_1$ и x_1 и последующего перемножения обеих ф. в. Маргинальная ф. в. величины x_2, получение которой часто является целью в такой задаче, может быть тогда определена обычным путем, исходя из $p(x_1, x_2)$.

Примеры. Предположим, что мы хотим найти вероятность $p(x_2 | x_1)$ того, что среди 13 вынутых «наудачу» карт окажется x_2 королей, если известно, что среди этих карт имеется x_1 тузов. Вероятность получить среди 13 карт точно x_1 тузов и x_2 королей равна

$$p(x_1, x_2) = \frac{\binom{4}{x_1}\binom{4}{x_2}\binom{44}{13-x_1-x_2}}{\binom{52}{13}}.$$

Далее, вероятность того, что среди 13 карт будет содержаться x_1 тузов, есть $p_1(x_1) = \sum_{x_2=0}^{4} p(x_1, x_2)$, иначе

$$p_1(x_1) = \frac{\binom{4}{x_1}\binom{48}{13-x_1}}{\binom{52}{13}}.$$

Поэтому по формуле (2.9.8) мы имеем

$$p(x_2 | x_1) = \frac{\binom{4}{x_2}\binom{44}{13-x_1-x_2}}{\binom{48}{13-x_1}}.$$

Чтобы пояснить (2.9.9), рассмотрим другой пример. Пусть x_1 есть случайная величина, означающая число очков, выпадающих при бросании «правильной» кости. Если выпало $x_1 = x_1'$ очков, то подбрасывается x_1' монет, и пусть x_2 — случайная величина, означающая число появившихся гербов. Интересующая нас задача состоит в том, чтобы определить $p_2(x_2)$, т. е. вероятность того, что в результате целого эксперимента будет наблюдаться число гербов, равное x_2. Мы имеем

$$p(x_2 | x_1) = \binom{x_1}{x_2}\left(\frac{1}{2}\right)^{x_1} \quad \text{и} \quad p_1(x_1) = \frac{1}{6}.$$

Следовательно,

$$p(x_1, x_2) = \frac{1}{6}\binom{x_1}{x_2}\left(\frac{1}{2}\right)^{x_1}.$$

Чтобы найти $p_2(x_2)$, мы складываем значения $p(x_1, x_2)$ для $x_1 = x_2, x_2+1, \ldots, 6$. Это дает

$$p_2(x_2) = \frac{63}{384}, \frac{120}{384}, \frac{99}{384}, \frac{64}{384}, \frac{29}{384}, \frac{8}{384}, \frac{1}{384},$$

для $x_2 = 0, 1, 2, 3, 4, 5, 6$ соответственно.

Второй, менее общий класс условных случайных величин типа А содержит величины, для которых $F(x_2 \mid x_1^{(\beta)})$ абсолютно непрерывна по x_2, т. е. обладает функцией плотности $f(x_2 \mid x_1^{(\beta)})$, так что

$$F(x_2 \mid x_1^{(\beta)}) = \frac{F(x_1^{(\beta)}, x_2) - F(x_1^{(\beta)} - 0, x_2)}{p_1(x_1^{(\beta)})} = \int_{-\infty}^{x_2} f(y \mid x_1^{(\beta)})\, dy. \tag{2.9.10}$$

Это следует, конечно, немедленно из (2.9.2) при $k=2$, если G означает множество точек в R_2, для которых $x_1 = x_1^{(\beta)}$. Выбирая $x_1' = x_1^{(\beta)}$, можно получить то же самое и из (2.9.5).

Таким образом, $x_2 \mid x_1^{(\beta)}$ является *непрерывной условной случайной величиной* с ф. п. в. $f(x_2 \mid x_1^{(\beta)})$. Заметим, что представление функции $f(x_2 \mid x_1^{(\beta)})$ через $F(x_1, x_2)$ дается формулой (2.5.12) при $\alpha = \beta$.

Тип В. Для этого типа x_1 является непрерывной случайной величиной с ф. п. в. $f_1(x_1)$. Мы рассмотрим два полезных случая. В более общем случае (x_1, x_2) есть непрерывная случайная величина с ф. п. в. $f(x_1, x_2)$ и (2.9.4) дает

$$F(x_1', x_2 \mid I_1) = \frac{\int\limits_{x_1''}^{x_1'} \int\limits_{-\infty}^{x_2} f(x_1, x_2)\, dx_1\, dx_2}{\int\limits_{x_1''}^{x_1'} f_1(x_1)\, dx_1}. \tag{2.9.11}$$

Если $f_1(x_1)$ и $\int\limits_{-\infty}^{x_2} f(x_1, x_2)\, dx_2$ — непрерывные функции от x_1 в точке x_1' и $f_1(x_1') > 0$, то, переходя к пределу при $x_1'' \to x_1'$, мы получаем

$$F(x_2 \mid x_1') = \frac{\int\limits_{-\infty}^{x_2} f(x_1', x_2)\, dx_2}{f_1(x_1')}. \tag{2.9.12}$$

Читатель заметит, что это выражение для $F(x_2 \mid x_1')$ в непрерывном случае подсказывается аналогичной формулой (2.9.7) в дискретном случае.

Если мы положим

$$f(x_2 \mid x_1') = \frac{f(x_1', x_2)}{f_1(x_1')}, \tag{2.9.13}$$

то $f(x_2 \mid x_1')$ будет ф. п. в., именно ф. п. в. условной случайной величины $x_2 \mid x_1'$. Если можно не опасаться путаницы, стоит отбросить штрих у x_1 и переписать (2.9.13) иначе:

$$f(x_1, x_2) = f(x_2 \mid x_1) \cdot f_1(x_1), \tag{2.9.14}$$

что, конечно, представляет аналог формулы (2.9.9) для непрерывных случайных величин.

Пример. Пусть x_1 — случайная величина, означающая число, выбранное «случайно» в интервале (0,1) (все числа интервала (0,1) должны рассматриваться равновероятными в элементарном геометрическом смысле), и x_2 — случайная величина, означающая число, выбранное «случайно» в интервале $(x_1', 1)$, где x_1' — реализованное значение x_1. Мы хотим найти ф. п. в. величины x_2, т. е. $f_2(x_2)$. Мы имеем

$$f_1(x_1) = \begin{cases} 1, & 0 < x_1 < 1, \\ 0 & \text{в противном случае;} \end{cases} \qquad f(x_2 \mid x_1) = \begin{cases} \dfrac{1}{1-x_1}, & x_1 < x_2 < 1, \\ 0 & \text{в противном случае.} \end{cases}$$

Следовательно,

$$f(x_1, x_2) = \begin{cases} \dfrac{1}{1-x_1}, & 0 < x_1 < x_2 < 1, \\ 0 & \text{во всех других точках.} \end{cases}$$

Поэтому

$$f_2(x_2) = \int_0^{x_2} \frac{dx_1}{1-x_1} = -\log(1-x_2), \qquad 0 < x_2 < 1,$$

и $f_2(x_2) = 0$ для x_2 вне промежутка (0,1).

Во втором случае условных величин типа В, встречающемся менее часто, x_1 непрерывна, а x_2 дискретна. Читатель может проверить, что если $f_1(x_1)$ непрерывна в точке $x_1 = x_1'$, $f_1(x_1') > 0$ и, кроме того, $F(x_1, x_2)$ и $F(x_1, x_2 - 0)$ обладают производными по x_1 в точке x_1', то применение (2.9.4) и (2.9.5) дает

$$F(x_2 \mid x_1') = \sum_{x_2^{(\beta)} \leqslant x_2} p^*(x_2^{(\beta)} \mid x_1'), \tag{2.9.15}$$

где

$$p^*(x_2^{(\beta)} \mid x_1') = \frac{f^*(x_1', x_2^{(\beta)})}{f_1(x_1')} \tag{2.9.16}$$

и

$$f^*(x_1, x_2) = \frac{\partial}{\partial x_1}[F(x_1, x_2) - F(x_1, x_2 - 0)]. \tag{2.9.17}$$

Заметим, что $p^*(x_2 \mid x_1)$ аналогична по структуре функции $p(x_2 \mid x_1)$, которая определена в (2.9.8), если не считать, что $p^*(x_2 \mid x_1)$ составлена из плотностей вероятности, а не самих вероятностей (в точках сосредоточения массы). Грубо говоря, мы можем вообразить, что вероятность 1 разделена на части $p_2(x_2^{(1)})$, $p_2(x_2^{(2)})$, ... и затем каждая из этих частей вероятности непрерывно «размазана» по прямым $x_2 = x_2^{(1)}$, $x_2 = x_2^{(2)}$, ... в соответствии с функциями плотности (но не ф. п. в.) $f^*(x_1, x_2^{(1)})$, $f^*(x_1, x_2^{(2)})$, ...

(с) Обобщение на k-мерные случайные величины. Определение (2.9.5) условной к. ф. р. может быть непосредственно распространено на многомерные случайные величины. Рассмотрим случайную величину $(x_1, \ldots, x_k)$ с к. ф. р. $F(x_1, \ldots, x_k)$. Пусть $F_{1,\ldots,k_1}(x_1, \ldots, x_{k_1})$ — маргинальная к. ф. р. величины $(x_1, \ldots, x_{k_1})$, $k_1 < k$, $k = k_1 + k_2$, и

I_{k_1} — цилиндрическое множество в R_k, для которого $x_i'' < x_i \leqslant x_i'$, $i = 1, \ldots, k$. Проекция этого цилиндрического множества на R_{k_1} — выборочное пространство точек $(x_1, \ldots, x_{k_1})$ — есть k_1-мерный интервал $(x_1'', \ldots, x_{k_1}''; x_1', \ldots, x_{k_1}']$, который мы также обозначим через I_{k_1}. Пусть $\Delta_{I_{k_1}}^{k_1} F_{1,\ldots,k_1}(x_1, \ldots, x_{k_1})$ — k-я разность функции $F_{1,\ldots,k_1}(x_1, \ldots \ldots, x_{k_1})$ над интервалом I_{k_1}, определенная в соответствии с (2.6.3) и (2.6.4). Эта k_1-я разность есть не что иное, как $P(I_{k_1})$, которая предполагается положительной. Теперь положим

$$F(x_1, \ldots, x_k \mid I_{k_1}) = \frac{P(E_{(x_1, \ldots, x_k)} \cap I_{k_1})}{P(I_{k_1})}. \tag{2.9.18}$$

В частности, мы имеем

$$F(x_1', \ldots, x_{k_1}', x_{k_1+1}, \ldots, x_k \mid I_{k_1}) = \frac{\Delta_{I_{k_1}}^{k_1} F(x_1, \ldots, x_k)}{\Delta_{I_{k_1}}^{k_1} F_{1,\ldots,k_1}(x_1, \ldots, x_{k_1})}, \tag{2.9.19}$$

где числитель в правой части представляет k_1-ю разность функции $F(x_1, \ldots, x_k)$ относительно $(x_1, \ldots, x_{k_1})$ при фиксированных значениях $(x_{k_1+1}, \ldots, x_k)$.

Если существует предел правой части (2.9.19), когда $x_1'' \to x_1', \ldots \ldots, x_{k_1}'' \to x_{k_1}'$, т. е. если

$$\lim_{\substack{x_i'' \to x_i' \\ i = 1, \ldots, k_1}} F(x_1'', \ldots, x_{k_1}'', x_{k_1+1}, \ldots, x_k \mid I_{k_1}) = \\ = F(x_{k_1+1}, \ldots, x_k \mid x_1', \ldots, x_{k_1}'), \tag{2.9.20}$$

то $F(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1})$ (штрихи опущены) называется к. ф. р. *условной случайной величины* $(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1})$. Можно проверить, что если эта условная случайная величина существует, $F(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1})$ обладает всеми свойствами (2.6.8) k_2-мерной к. ф. р. Если предел (2.9.20) не существует, требуется более общий подход, обсужденный в § 1.10.

В случае дискретной величины $(x_1, \ldots, x_k)$, если G в (2.9.2) выбирается в виде части R_k, для которой $x_1 = x_1', \ldots, x_{k_1} = x_{k_1}'$, причем $P(G) > 0$, то условная случайная величина $(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots \ldots, x_{k_1})$ существует и имеет ф. в., даваемую следующей формулой, являющейся обобщением (2.9.8):

$$p(x_{k_1+1}, \ldots, x_k \mid x_1', \ldots, x_{k_1}') = \frac{p(x_1', \ldots, x_{k_1}', x_{k_1+1}, \ldots, x_k)}{p_{1,\ldots,k_1}(x_1', \ldots, x_{k_1}')}. \tag{2.9.21}$$

Опустим штрихи. $p(x_1, \ldots, x_k)$ есть ф. в. величины $(x_1, \ldots, x_k)$, а $p_{1,\ldots,k_1}(x_1, \ldots, x_{k_1})$ — маргинальная ф. в. величины $(x_1, \ldots, x_{k_1})$. В условиях, которые были высказаны ранее, можно также показать, что к. ф. р. условной случайной величины $(x_{k_1+1}, \ldots, x_k \mid \ldots, x_{k_1}')$ [имеющая ф. п. в. (2.9.21)] дается (2.9.20).

Читатель может проверить, что если величина $(x_1, \ldots, x_k)$ дискретна, мы имеем следующую формулу умножения для ф. в. в точках сосредоточения масс пространства $(x_1, \ldots, x_k)$:

$$p(x_1, \ldots, x_k) = p(x_k \mid x_1, \ldots, x_{k-1}) \times \\ \times p_{1,\ldots,k-1}(x_{k-1} \mid x_1, \ldots, x_{k-2}) \ldots p_1(x_1). \tag{2.9.22}$$

В случае, если $(x_1, \ldots, x_k)$ — непрерывная величина, мы имеем в согласии с (2.9.21) ф. п. в.

$$f(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1}) = \frac{f(x_1, \ldots, x_k)}{f_{1,\ldots,k_1}(x_1, \ldots, x_{k_1})}, \tag{2.9.23}$$

где обе функции в правой части есть ф. п. в., определенные в § 2.7. Эту ф. п. в. можно получить из (2.9.20) путем непосредственного обобщения рассуждений и предположений, использованных при выводе (2.9.12).

Аналогично (2.9.22) мы имеем, конечно, и для непрерывного случая следующую формулу умножения для ф. п. в.:

$$f(x_1, \ldots, x_k) = f(x_k \mid x_1, \ldots, x_{k-1}) \times \\ \times f_{1,\ldots,k}(x_{k-1} \mid x_1, \ldots, x_{k-2}) \ldots f_1(x_1), \tag{2.9.24}$$

считая, что ф. п. в. всех рассматриваемых условных случайных величин существуют.

(d) Условные функции распределения в случае независимости. Рассмотрим сначала случайную величину (x_1, x_2). Пусть G — борелевское цилиндрическое множество в R_2, параллельное x_2-оси, и $P(G) > 0$. Тот факт, что некоторая точка (x_1, x_2) содержится в G, мы будем записывать просто: $x_1 \in G$; это не внесет никакой путаницы. Обозначим через $E_{(x_1', x_2')}$ декартово произведение $E_{x_1'} \times E_{x_2'}$, где E_{x_1} — множество в $R_1^{(1)}$, для которого $x_1 \leqslant x_1'$ и $E_{x_2'}$ — множество в $R_1^{(2)}$, для которого $x_2 \leqslant x_2'$. Тогда

$$E_{(x_1', x_2')} \cap G = (E_{x_1} \cap G) \times G_{x_2'}$$

и из (2.9.2)

$$F(x_1', x_2' \mid G) = \frac{P((E_{x_1'} \cap G) \times E_{x_1'})}{P(G)}. \tag{2.9.25}$$

Если x_1 и x_2 независимы, вместо (2.9.25) мы будем иметь

$$F(x_1', x_2' \mid G) = \frac{P(E_{x_1'} \cap G)}{P(G)} \cdot P(E_{x_1'}) \tag{2.9.26}$$

и, таким образом, мы получаем следующий результат:

2.9.1. Если (x_1, x_2) — *случайная величина с к. ф. р.* $F(x_1, x_2)$, *причем* x_1 *и* x_2 *независимы, и если* G — *произвольное (борелевское) цилиндрическое множество в* R_2, *параллельное* x_2-*оси, для которого* $P(G) > 0$, *то*

$$F(x_1, x_2 \mid G) = F(x_1 \mid G) \cdot F_2(x_2). \tag{2.9.27}$$

Другими словами, условная вероятность того, что x_2 принадлежит к какому-либо заданному множеству при условии $(x_1, x_2) \in G$, где G — цилиндрическое множество, параллельное x_2-оси, не зависит от G. Допустим, в частности, что G есть множество, для которого x_1 принадлежит множеству I_1, встречавшемуся в (2.9.3). Тогда, если x_1 и x_2 независимы, то из теоремы **2.4.2** вытекает, что вместо (2.9.4) можно написать

$$F(x_1', x_2' \mid I_1) = F_2(x_2). \tag{2.9.28}$$

В этом случае $\lim\limits_{x_1'' \to x_1'} F(x_1', x_2' \mid I_1)$ существует всюду, за исключением, быть может, множества точек x_2 вероятности нуль, и равен $F_2(x_2)$. Таким образом, мы получаем следствие теоремы **2.9.1.**

2.9.1a. *Если x_1 и x_2 независимы, тогда всюду, за исключением, быть может, множества вероятности* 0,

$$F(x_2 \mid x_1) = F_2(x_2). \tag{2.9.29}$$

Ясно, что в случае, когда x_1 и x_2 независимы, (2.9.8) приводится к соотношению

$$p(x_2 \mid x_1) = p_2(x_2), \tag{2.9.29a}$$

а (2.9.13) — к соотношению

$$f(x_2 \mid x_1) = f_2(x_2). \tag{2.9.29b}$$

Вообще

2.9.2. *Если $(x_1, \ldots, x_k)$ — случайная величина с к. ф. р. $F(x_1, \ldots, x_k)$, причем случайные величины $(x_1, \ldots, x_{k_1})$ и $(x_{k_1+1}, \ldots, x_k)$ независимы, и G — произвольное борелевское цилиндрическое множество в R_k, определяемое некоторым множеством значений $x_1, \ldots, x_{k_1}$, так что $P(G) > 0$, то*

$$F(x_1, \ldots, x_k \mid G) = \\ = F_{1,\ldots,k_1}(x_1, \ldots, x_{k_1} \mid G)\, F_{k_1+1,\ldots,k}(x_{k_1+1}, \ldots, x_k). \tag{2.9.30}$$

Если G является интервалом, для которого $(x_1, \ldots, x_{k_1})$ есть точка множества I_{k_1}, использованного в (2.9.19), и, кроме того, $(x_1, \ldots, x_{k_1})$ и $(x_{k_1+1}, \ldots, x_k)$ независимы, мы имеем

$$F(x_1, \ldots, x_k \mid I_{k_1}) = F_{1,\ldots,k_1}(x_1, \ldots, x_{k_1} \mid I_{k_1}) \cdot F_{k_1+1,\ldots,k}(x_{k_1+1}, \ldots, x_k),$$

так что (2.9.19) дает

$$F(x_1', \ldots, x_{k_1}', x_{k_1+1}, \ldots, x_k \mid I_{k_1}) = F_{k_1+1,\ldots,k}(x_{k_1+1}, \ldots, x_k). \tag{2.9.31}$$

Переходя к пределам при $x_1'' \to x_1', \ldots, x_{k_1}'' \to x_{k_1}'$, существующим почти везде, мы получаем как следствие теоремы **2.9.2** теорему

2.9.2a. *Если $(x_1, \ldots, x_k)$ — случайная величина с к.ф.р. $F(x_1, \ldots, x_k)$, причем $(x_1, \ldots, x_{k_1})$ и $(x_{k_1+1}, \ldots, x_k)$ независимы, то всюду,* *за исключением, быть может, множества вероятности* 0,

$$F(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1}) = F_{k_1+1, \ldots, k}(x_{k_1+1}, \ldots, x_k). \quad (2.9.32)$$

Если $(x_1, \ldots, x_{k_1})$ и $(x_{k_1+1}, \ldots, x_k)$ независимы, из (2.9.32) следует, что вместо (2.9.21) и (2.9.23) мы будем иметь соответственно

$$p(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1}) = p_{k_1+1, \ldots, k}(x_{k_1+1}, \ldots, x_k) \quad (2.9.32a)$$

и

$$f(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1}) = f_{k_1+1, \ldots, k}(x_{k_1+1}, \ldots, x_k). \quad (2.9.32b)$$

2.10. Конечные случайные процессы

k-мерная случайная величина $(x_1, \ldots, x_k)$ часто называется *конечным случайным процессом*. Особенно это принято в приложениях, где компоненты $x_1, \ldots, x_k$ соответствуют измерениям, произведенным на исходах последовательности физических операций таким, конечно, образом, что k-мерная к. ф. р. определяется.

Пример. Пусть x_1 — случайная величина, означающая число, взятое «наудачу» в интервале (0,1), причем все числа предполагаются «равновероятными», x_2 есть число, взятое «наудачу» из $(x_1, 1)$, и т. д., k таких чисел дают нам k-мерную случайную величину $(x_1, \ldots, x_k)$ с ф. п. в. $f(x_1, \ldots, x_k)$, которую можно найти, прилагая (2.9.24). Мы имеем

$$f(x_1, \ldots, x_k) = \frac{1}{(1 - x_{k-1})} \cdot \frac{1}{(1 - x_{k-2})} \cdots \frac{1}{1 - x_1}$$

для $0 < x_1 < \ldots < x_k < 1$ и $f(x_1, \ldots, x_k) = 0$ в противном случае. Мы можем тогда считать $(x_1, \ldots, x_k)$ (конечным) случайным процессом для описания результатов последовательности «зарубок» интервала (0,1), произведенных в соответствии с вышеизложенным.

Продолжая знакомство с этой книгой, читатель найдет в ней много примеров конечных случайных процессов. Бесконечные случайные процессы, т. е. многомерные случайные величины с бесконечным числом компонент, также встретятся в ряде мест. Они будут определены и изучены в главе 4.

ЗАДАЧИ

2.1. Пусть $F(x)$ — к.ф.р. случайной величины x. Рассматривая подходящие последовательности полуоткрытых интервалов и опираясь на теорему **1.4.5**, установить формулы (2.2.7) — (2.2.10).

2.2. Рассматривая подходящую последовательность двумерных полуоткрытых интервалов, установить формулу (2.4.9).

2.3. Пусть $F_1(x)$ и $F_2(x)$ — к. ф. р. соответственно дискретной и непрерывной случайной величины. Показать, что если a и b — неотрицательные числа, сумма которых равна единице, то функция

$$aF_1(x) + bF_2(x)$$

обладает всеми свойствами к. ф. р.

2.4. Смешанная двумерная случайная величина (x_1, x_2) такова, что x_1 дискретна с ф. в. $p(x_1)=\left(\frac{1}{2}\right)^{x_1}$, $x_1=1, 2, \ldots$, и x_2 непрерывна, причем условная величина $x_2 \mid x_1$ имеет на интервале (0,1) ф. п. в. $x_1(1-x_2)^{x_1-1}$. Показать, что ф. п. в. (безусловной) случайной величины x_2 равна на отрезке (0,1) $2(1+x_2)^{-2}$.

2.5. Дискретная случайная величина x_1 имеет ф. в. qp^{x_1-1}, $x_1=1, 2, \ldots$, где $0<p<1$, $q=1-p$. Кроме того, x_2 — непрерывная случайная величина, причем ф. п. в. для $x_2 \mid x_1$ на (0,1) равна $x_1 \cdot x_2^{x_1-1}$ и вне (0,1) — нулю. Определить безусловную ф. п. в. величины x_2, ф. в. величины $x_1 \mid x_2$, к. ф. р. величины x_1 и к. ф. р. величины x_2.

2.6. Показать, что если

$$G(x_1, x_2)=\begin{cases}[1-(1+x_1)^{-k}-(1+x_2)^{-k}+(1+x_1+x_2)^{-k}], & x_1, x_2 \geqslant 0,\\ 0 \text{ в противном случае,}\end{cases}$$

где $k>0$, то $G(x_1, x_2)$ удовлетворяет всем условиям для к. ф. р. двумерной непрерывной случайной величины (x_1, x_2). Найти ф. п. в. этой величины.

2.7. Показать, что если (x, y) — двумерная случайная величина с ф. п. в., которая равна $f(x, y)$ при $x>0$, $y>0$ и равна 0 в противном случае, то

a) ф. п. в. величины $u=\frac{y}{x}$ есть

$$\int_0^\infty xf(x, ux)\,dx,$$

b) ф. п. в. величины $v=x+y$ есть

$$\int_0^\infty f(x, v-x)\,dx,$$

c) ф. п. в. величины $w=xy$ есть

$$\int_0^\infty \frac{1}{x} f\left(x, \frac{w}{x}\right) dx.$$

2.8. Допустим, что случайная величина (x_1, x_2) имеет ф. п. в., равную $\frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2+x_2^2)}$ в каждой точке (x_1, x_2) x_1x_2-плоскости. Показать, что если случайные величины y_1, y_2 связаны с x_1 и x_2 формулами

$$x_1=y_1 \cos y_2, \qquad x_2=y_1 \sin y_2,$$

где ф. п. в. величины y_1 равна $y_1 e^{-\frac{1}{2}y_1^2}$ для $y_1>0$ и 0 для $y_1 \leqslant 0$, а ф. п. в. величины y_2 равна $\frac{1}{2\pi}$ для $0<y_2<2\pi$ и 0 в противном случае, то y_1 и y_2 независимы.

2.9. Ф. п. в. трехмерной случайной величины (x_1, x_2, x_3) равна 6 в тетраэдре с вершинами (0,0,0), (0,0,1), (0,1,0), (1,0,0). Найти к. ф. р. и ф. п. в. следующих величин: (x_1, x_2), x_1, $x_2 \mid x_1$, x_1+x_2, $x_1+x_2+x_3$.

2.10. Пусть (x_1, x_2, x_3) — случайная величина с неотрицательными компонентами и ф. п. в., равной $e^{-(x_1+x_2+x_3)}$. Найти ф. п. в. величины (u, x_2, x_3), где $u=x_1+x_2+x_3$. Исходя из ф.п.в. величины (x_1, x_2, x_3), найти ф.п.в. величин

a) u,

b) $u \mid x_2$,

c) $u \mid x_2, x_3$,

d) $(x_2 \mid u, x_3 \mid u)$.

2.11. Показать, что если $(x_1, \ldots, x_k)$ — случайная величина с к.ф.р. $F(x_1, \ldots, x_k)$ и все компоненты $x_1, \ldots, x_k$ независимы, то в (2.6.4)

$$\Delta_{I_k}^k F(x_1, \ldots, x_k) = \Delta F_1(x_1) \ldots \Delta F_k(x_k),$$

где

$$\Delta F_i(x_i) = F_i(x_i'') - F_i(x_i')$$

и $F_i(x_i)$ — маргинальная к.ф.р. величины x_i.

2.12. Показать, что если $(x_1, \ldots, x_k)$ — k-мерная случайная величина с ф.п.в., симметричной относительно $x_1, \ldots, x_k$, то $P(x_1<x_2<\ldots<x_k)=\frac{1}{k!}$.

2.13. Известно, что k-мерная случайная величина $(x_1, \ldots, x_k)$ имеет ф.п.в. вида

$$C(A+x_1+\ldots+x_k)^{-k-B}$$

для $x_1>0, \ldots, x_k>0$, где A и B — положительные числа, и равна 0 в противном случае. Определить C. Найти ф.п.в. маргинального распределения величины $(x_1, \ldots, x_r)$, $r<k$.

2.14. Показать, что если $x_1, \ldots, x_k$ — независимые случайные величины с тождественно равными непрерывными к.ф.р. $F(x_1), \ldots, F(x_k)$, и

$$u=\min(x_1, \ldots, x_k), \qquad v=\max(x_1, \ldots, x_k),$$

то к.ф.р. величины (u, v) равна

$$[F(v)]^k - g(u, v)[F(v)-F(u)]^k,$$

где

$$g(u, v)=\begin{cases}1 & \text{для } u<v,\\ 0 & \text{для } u\geqslant v.\end{cases}$$

Найти маргинальные к.ф.р. величин u и v. Полагая, что $F(x)$ имеет производную $f(x)$, найти также ф.п.в. величин (u, v), u и v.

2.15. В урне находятся N шаров, занумерованных числами $1, \ldots, N$. Извлекается один шар, и его номер регистрируется как значение случайной величины x_1, после чего шар возвращается обратно в урну. Затем извлекается второй шар, его номер регистрируется как значение случайной величины x_2 и шар возвращается в урну. И так далее до тех пор, пока не будут получены значения k величин $x_1, \ldots, x_k$. Приписывая равные вероятности появления всем номерам на каждом шагу и предполагая взаимную независимость величин $x_1, \ldots, x_k$, выписать к.ф.р. k-мерной случайной величины $(x_1, \ldots, x_k)$. Найти к.ф.р. и ф.в. величины $y=\max(x_1, \ldots, x_k)$.

2.16. (*Продолжение*). Допустим, что последовательно без возвращения извлекается k $(<N)$ шаров, причем считается, что вероятность извлечения какого-нибудь шара, остающегося в урне на любом данном шагу, должна равняться вероятности извлечения любого другого шара, остающегося в урне на этом шагу. Сколько точек сосредоточения массы имеет величина $(x_1, \ldots, x_k)$? Определить ф.в. величины $(x_1, \ldots, x_k)$. Показать, что ф.в. маргинального распределения любого набора r $(<k)$ x-ов остается неизменной при переходе к любому другому набору r x-ов.

2.17. Показать, что если $(x_1, y_1), \ldots, (x_k, y_k)$, $k \geqslant 2$, — независимые двумерные случайные величины с тождественно равными к. ф. р. $F(x_1, y_1), \ldots$ $\ldots, F(x_k, y_k)$ (и тождественно равными ф. п. в. $f(x_1, y_1), \ldots, f(x_k, y_k)$) и

$$u = \max(x_1, \ldots, x_k), \qquad v = \max(y_1, \ldots, y_k),$$

то ф. п. в. величины (u, v) равна

$$k(k-1) F^{k-2}(u, v) \frac{\partial F}{\partial u} \cdot \frac{\partial F}{\partial v} + kF^{k-1}(u, v) f(u, v).$$

2.18. (*Продолжение*). Показать, что если $w = \max(x_1 + y_1, \ldots, x_k + y_k)$, то ф. п. в. величины w есть

$$k \left[\int_{-\infty}^{\infty} \int_{-\infty}^{w-x} f(x, y)\, dy\, dx \right]^{k-1} \cdot \int_{-\infty}^{\infty} f(x, w-x)\, dx.$$

2.19. Показать, что если каждая из случайных величин $x_1, \ldots, x_k$ независима от остальных, то эти величины взаимно независимы.

2.20. Показать, что если $F(x_1, \ldots, x_k)$ — k-мерная к.ф.р. с маргинальными к. ф. р. $F_1(x_1), \ldots, F_k(x_k)$, то

$$F(x_1, \ldots, x_k) \leqslant [F_1(x_1) \ldots F_k(x_k)]^{\frac{1}{k}}.$$

ГЛАВА 3

СРЕДНИЕ ЗНАЧЕНИЯ И МОМЕНТЫ СЛУЧАЙНЫХ ВЕЛИЧИН

3.1. Введение

В § 1.8 мы определили для данного вероятностного пространства $(R, \mathscr{B}, P)$ интеграл Лебега — Стилтьеса от случайной величины $x(e)$ относительно P, взятый по множеству $E \in \mathscr{B}$.

Пусть E' — борелевское множество в R_1 и E — множество в R, для которого $x(e) \in E'$. Тогда, если $F(x)$ есть к. ф. р. величины $x(e)$, мы можем написать интеграл Лебега — Стилтьеса (1.8.5) от $x(e)$ по множеству E несколько иначе, именно

$$\int_E x(e)\,dP(e) = \int_{E'} x\,dF(x). \tag{3.1.1}$$

Аналогично, пусть E'_k — борелевское множество в R_k и E — множество в R, для которого $(x_1(e), \ldots, x_k(e)) \in E'_k$. Если $F(x_1, \ldots, x_k)$ — к. ф. р. величины $(x_1(e), \ldots, x_k(e))$ и функция $g(x_1, \ldots, x_k)$ измерима относительно $\mathscr{B}_k$ (так что $g(x_1(e), \ldots, x_k(e))$ становится измеримой относительно $\mathscr{B}$), то

$$\int_E g(x_1(e), \ldots, x_k(e))\,dP(e) = \int_{E'_k} g(x_1, \ldots, x_k)\,dF(x_1, \ldots, x_k). \tag{3.1.2}$$

Допустим, что в (3.1.2) E'_k есть множество точек R_k, для которых $g(x_1, \ldots, x_k) \in E'_1$, где E'_1 — борелевское множество в R_1. В таком случае, если $H(y)$ — к. ф. р. величины $g(x_1, \ldots, x_k)$, мы можем также написать

$$\int_{E'_k} g(x_1, \ldots, x_k)\,dF(x_1, \ldots, x_k) = \int_{E'_1} y\,dH(y). \tag{3.1.3}$$

3.2. Среднее значение случайной величины

В том случае, если в формуле (3.1.1) мы полагаем $E' = R_1$, получающийся интеграл,

$$\mathscr{E}(x) = \int_{-\infty}^{\infty} x\,dF(x), \tag{3.2.1}$$

определяет *среднее значение* случайной величины x.

Для дискретной случайной величины x с ф. в. $p(x)$, введенной в § 2.3 (а), $\mathscr{E}(x)$ сводится к следующей сумме:

$$\mathscr{E}(x)=\sum_{\alpha} x^{(\alpha)} p(x^{(\alpha)}), \tag{3.2.2}$$

где $x^{(\alpha)}$ — соответствующие точки сосредоточения массы. Для непрерывной случайной величины x с ф. п. в. $f(x)$ мы имеем

$$\mathscr{E}(x)=\int_{-\infty}^{\infty} x\,dF(x)=\int_{-\infty}^{\infty} xf(x)\,dx. \tag{3.2.3}$$

В более общем случае, когда в качестве множества E_k' в формуле (3.1.2) берется все пространство R_k, мы получаем

$$\mathscr{E}(g(x_1, \ldots, x_k))=\int_{R_k} g(x_1, \ldots, x_k)\,dF(x_1, \ldots, x_k). \tag{3.2.4}$$

Кроме того, из (3.1.3) следует, что если $E_k'=R_k$, то $E_1'=R_1$ и

$$\mathscr{E}(g(x_1, \ldots, x_k))=\int_{-\infty}^{\infty} y\,dH(y). \tag{3.2.5}$$

Обозначив $g(x_1, \ldots, x_k)$ через y, мы находим, что интеграл в (3.2.4) есть просто $\mathscr{E}(y)$ и, следовательно,

$$\mathscr{E}(g(x_1, \ldots, x_k))=\mathscr{E}(y). \tag{3.2.6}$$

Как и выше, для дискретной случайной величины $(x_1, \ldots, x_k)$ (3.2.4) сводится к сумме, а для непрерывной представляет k-мерный интеграл по пространству R_k.

Средние значения, если они существуют, обладают следующими полезными свойствами:

3.2.1. *Если c — константа, то*

$$\mathscr{E}(cy)=c\mathscr{E}(y).$$

3.2.2. $\mathscr{E}(ay_1+by_2)=a\mathscr{E}(y_1)+b\mathscr{E}(y_2)$.

3.2.3. *Если $m\leqslant y\leqslant M$, то $m\leqslant \mathscr{E}(y)\leqslant M$.*

3.2.4. *Если $y_1\leqslant y_2$, то $\mathscr{E}(y_1)\leqslant \mathscr{E}(y_2)$.*

3.2.5. $|\mathscr{E}(y)|\leqslant \mathscr{E}(|y|)$.

Допустим теперь, что (x_1, x_2) есть случайная величина с к. ф. р. $F(x_1, x_2)$, удовлетворяющей условию независимости x_1 и x_2, и рассмотрим среднее значение произведения x_1x_2. Можно проверить, что

$$\mathscr{E}(x_1x_2)=\int_{-\infty}^{\infty} x_1\,dF_1(x_1)\int_{-\infty}^{\infty} x_2\,dF_2(x_2)=\mathscr{E}(x_1)\mathscr{E}(x_2). \tag{3.2.7}$$

Таким образом, мы приходим к свойству

3.2.6. *Если (x_1, x_2) — случайная величина, компоненты которой x_1 и x_2 независимы, то*

$$\mathcal{E}(x_1 x_2) = \mathcal{E}(x_1)\,\mathcal{E}(x_2).$$

Аналогичный результат, конечно, справедлив и для k взаимно независимых случайных величин.

3.3. Моменты одномерных случайных величин

В математической статистике имеется много задач, в которых трудно или даже невозможно полностью определить к. ф. р. случайной величины. В таких случаях часто, и не без пользы, удается в общих чертах описать распределение случайной величины посредством ее моментов и некоторых функций этих моментов.

Пусть x — случайная величина, имеющая к. ф. р. $F(x)$. *Среднее значение* величины x, доставляемой формулой (3.2.1), обычно обозначается через $\mathcal{E}(x)$ или $\mu(x)$. Таким образом,

$$\mathcal{E}(x) = \mu(x) = \int_{-\infty}^{\infty} x\,dF(x). \tag{3.3.1}$$

Дисперсия $\sigma^2(x)$ величины x определяется как среднее значение квадрата ее отклонения от $\mu(x)$, т. е.*)

$$\sigma^2(x) = \mathcal{E}(x - \mu(x))^2. \tag{3.3.2}$$

Число $\sigma(x)$, положительное значение квадратного корня из $\sigma^2(x)$, называется *стандартным отклонением* случайной величины x. Отношение $[\sigma(x)]/\mu(x)$ называется *коэффициентом изменчивости* величины x.

На языке элементарной механики, среднее случайной величины x можно истолковать как *центр тяжести* в R_1 распределения вероятностей этой величины. В свою очередь дисперсию можно истолковать как *момент инерции* того же самого распределения вероятностей относительно центра тяжести. Дисперсия указывает на степень собранности (или концентрации) массы вероятности около центра тяжести.

При отсутствии каких-либо сомнений насчет того, о какой случайной величине идет речь, мы будем писать μ вместо $\mu(x)$ и σ вместо $\sigma(x)$.

Используя (3.3.1) и свойства средних значений, выраженные в **3.2.1** и **3.2.2**, мы получаем

$$\sigma^2 = \mathcal{E}(x^2 - 2x\mu + \mu^2) = \mathcal{E}(x^2) - \mu^2. \tag{3.3.3}$$

---

*) Иногда среднее значение обозначают символом ave (x), а дисперсию — символом var (x), используя начальные буквы английских слов average — среднее и variance — дисперсия.

Если мы рассмотрим среднее значение $\mathscr{E}(x-a)^2$, где a — произвольная постоянная, то точно так же найдем

$$\mathscr{E}(x-a)^2 = \mathscr{E}[(x-\mu)+(\mu-a)]^2 = \sigma^2 + (\mu-a)^2. \qquad (3.3.4)$$

Последнее соотношение позволяет сделать следующее утверждение.

3.3.1. *Минимум среднего $\mathscr{E}(x-a)^2$ достигается при значении постоянной a, равном μ, и этот минимум есть σ^2.*

Полезная оценка вероятности, сосредоточенной на «хвостах» данного распределения, доставляется приведенным ниже неравенством Чебышева (1867).

3.3.2. *Если x — случайная величина, обладающая средним μ и дисперсией σ, то*

$$P(|x-\mu| \geqslant \lambda\sigma) \leqslant \frac{1}{\lambda^2}, \qquad (3.3.5)$$

где λ — любая постоянная величина.

Чтобы доказать это утверждение, мы сначала разобьем ось R_1 на три непересекающихся интервала

$$I = (-\infty,\ \mu-\lambda\sigma], \quad I' = (\mu-\lambda\sigma,\ \mu+\lambda\sigma), \quad I'' = [\mu+\lambda\sigma,\ \infty).$$

Тогда дисперсия величины x запишется в виде

$$\sigma^2 = \int_I (x-\mu)^2\,dF(x) + \int_{I'} (x-\mu)^2\,dF(x) + \int_{I''} (x-\mu)^2\,dF(x). \qquad (3.3.6)$$

Отбрасывая средний член в правой части (3.3.6), мы имеем

$$\sigma^2 \geqslant \int_I (x-\mu)^2\,dF(x) + \int_{I''} (x-\mu)^2\,dF(x). \qquad (3.3.7)$$

Поскольку в интервале I $x-\mu \leqslant -\lambda\sigma$, а в интервале I'' $x-\mu \geqslant \lambda\sigma$, неравенство (3.3.7) только усилится, если в обоих его интегралах заменить $(x-\mu)^2$ на $\lambda^2\sigma^2$. Таким образом,

$$\sigma^2 \geqslant \sigma^2\lambda^2 P(|x-\mu| \geqslant \lambda\sigma), \qquad (3.3.8)$$

а это эквивалентно неравенству (3.3.5).

r-й момент $\mu'_r(x)$ случайной величины x, где r — целое положительное число, определяется как среднее значение, если оно существует, величины x^r, т. е.

$$\mu'_r(x) = \mathscr{E}(x^r). \qquad (3.3.9)$$

Для удобства мы изредка будем допускать, что $r=0$; в этом случае $\mu'_0(x)=1$. Вопрос о том, существует ли и является ли конечным r-й момент $\mu'_r(x)$, есть просто вопрос об интегрируемости переменной x^r по пространству R_1.

r-м центральным моментом $\mu_r(x)$ случайной величины x называется среднее значение r-й степени ее отклонения от μ, иначе,

$$\mu_r(x) = \mathscr{E}[(x-\mu)^r]. \qquad (3.3.10)$$

Всякий раз, когда это не сможет вызвать никакой путаницы, мы будем несколько сокращать обозначение и писать μ'_r и μ_r соответственно вместо $\mu'_r(x)$ и $\mu_r(x)$.

Ясно, что центральные моменты μ_r, $r=1, 2, \ldots$, можно представить в виде полиномов от μ'_r и наоборот. Учитывая равенство $\mu'_1=\mu$, мы имеем

$$\begin{aligned} &\mu_1=0,\\ &\mu_2=\mu'_2-\mu^2,\\ &\mu_3=\mu'_3-3\mu'_2\mu+2\mu^3,\\ &\ldots\ldots\ldots\ldots \end{aligned} \tag{3.3.11}$$

Обратно,

$$\begin{aligned} &\mu'_1=\mu,\\ &\mu'_2=\mu_2+\mu^2,\\ &\mu'_3=\mu_3+3\mu_2\mu+\mu^3,\\ &\ldots\ldots\ldots\ldots \end{aligned} \tag{3.3.12}$$

Очевидно также, что $\mu_2=\sigma^2$.

r-й абсолютный момент $\nu'_r(x)$ задается соотношением

$$\nu'_r(x)=\mathscr{E}(|x|^r). \tag{3.3.13}$$

Вместо $\nu'_r(x)$ мы будем просто писать ν'_r. Величину

$$\nu_r(x)=\mathscr{E}(|x-\mu|^r) \tag{3.3.13a}$$

мы назовем *абсолютным центральным моментом.*

В общем случае, имея функцию $g(x)$, мы можем выразить ее дисперсию и моменты более высоких порядков через к. ф. р. случайной величины x. Например, используя (3.3.1) и (3.3.2), мы получаем для среднего и дисперсии величины $g(x)$ выражения

$$\mu(g(x))=\mathscr{E}[g(x)] \quad \text{и} \quad \sigma^2(g(x))=\mathscr{E}\{(g(x))^2-[\mu(g(x))]^2\}. \tag{3.3.14}$$

В задачах определенного класса, включающих дискретные случайные величины, часто удобно находить моменты μ'_r, предварительно вычисляя так называемые *факториальные моменты.* Положим

$$x^{[r]}=x(x-1)\ldots(x-r+1). \tag{3.3.15}$$

Тогда r-й *факториальный момент* есть число

$$\mu'_{[r]}(x)=\mathscr{E}(x^{[r]}). \tag{3.3.16}$$

Соблюдая известную осторожность, мы будем обозначать это число через $\mu'_{[r]}$. Факториальные моменты легко выражаются через обычные, именно,

$$\begin{aligned} &\mu'_{[1]}=\mu'_1,\\ &\mu'_{[2]}=\mu'_2-\mu'_1,\\ &\mu'_{[3]}=\mu'_3-3\mu'_2+2\mu'_1,\\ &\ldots\ldots\ldots\ldots \end{aligned} \tag{3.3.17}$$

Обратно,

$$\begin{aligned} \mu_1' &= \mu_{[1]}', \\ \mu_2' &= \mu_{[2]}' + \mu_{[1]}', \\ \mu_3' &= \mu_{[3]}' + 3\mu_{[2]}' + \mu_{[1]}', \\ &\ldots\ldots\ldots \end{aligned} \tag{3.3.18}$$

Заметим, что мы определили различного вида моменты, рассматривая их как моменты случайной величины x. Иногда их называют моментами распределения вероятностей величины x, а если $F(x)$ — к. ф. р. данной случайной величины, эти различного вида моменты выступают также как моменты к. ф. р. $F(x)$.

3.4. Моменты двумерных случайных величин

Допустим, что (x_1, x_2) — двумерная случайная величина с к. ф. р. $F(x_1, x_2)$. *Средние значения* величин x_1 и x_2 соответственно равны

$$\mu(x_1) = \mathscr{E}(x_1), \qquad \mu(x_2) = \mathscr{E}(x_2) \tag{3.4.1}$$

и обычно обозначаются просто через μ_1 и μ_2. *Дисперсии* этих величин равны

$$\sigma^2(x_1) = \mathscr{E}(x_1 - \mu_2)^2, \qquad \sigma^2(x_2) = \mathscr{E}(x_2 - \mu_2)^2 \tag{3.4.2}$$

и также кратко обозначаются через σ_1^2 и σ_2^2.

Конечно, полагая μ_1 вместо $\mu(x_1)$ и μ_2 вместо $\mu(x_2)$, мы вступаем в конфликт с обозначением (3.3.10). Однако опасность недоразумения представляется незначительной, так как в смысле (3.3.10) $\mu_1 = 0$ и $\mu_2 = \mu_2' - \mu^2$, и это обстоятельство будет ясно указываться при всяком рассмотрении, в котором могут появиться введенные символы.

Ковариация $\operatorname{cov}(x_1, x_2)$ между величинами x_1 и x_2 определяется равенством

$$\operatorname{cov}(x_1, x_2) = \mathscr{E}[(x_1 - \mu_1)(x_2 - \mu_2)]. \tag{3.4.3}$$

Правую часть этого равенства можно упростить, после чего оно записывается в виде

$$\operatorname{cov}(x_1, x_2) = \mathscr{E}(x_1, x_2) - \mathscr{E}(x_1)\mathscr{E}(x_2). \tag{3.4.4}$$

Обратим внимание, что $\operatorname{cov}(x_1, x_2) = \operatorname{cov}(x_2, x_1)$.

Если x_1 и x_2 независимы, из **3.2.6** вытекает, что $\mathscr{E}(x_1, x_2) = \mathscr{E}(x_1)\mathscr{E}(x_2)$, и мы получаем следующий результат:

3.4.1. *Если x_1 и x_2 независимы, то*

$$\operatorname{cov}(x_1, x_2) = 0. \tag{3.4.5}$$

Замечание. Следует заметить, что условие $\operatorname{cov}(x_1, x_2) = 0$ отнюдь не означает независимости величин x_1 и x_2. Например, пусть компоненты двумерной случайной величины (x_1, x_2) с вероятностью единица связаны функциональной зависимостью $x_2 = \sin x_1$ и ф. п. в. величины x_1 есть

$$f_1(x_1) = \begin{cases} \frac{1}{2\pi}, & -\pi < x_1 < \pi, \\ 0 & \text{в противном случае.} \end{cases}$$

Можно подсчитать, что $\operatorname{cov}(x_1, x_2) = 0$, однако с вероятностью единица значение x_2 можно точно определить, исходя из значения x_1.

Коэффициентом корреляции между x_1 и x_2 называется величина

$$\rho(x_1, x_2) = \frac{\operatorname{cov}(x_1, x_2)}{\sigma_1 \sigma_2}, \tag{3.4.6}$$

которая в условиях, исключающих какую-либо двусмысленность, будет обозначаться через ρ_{12} или просто буквой ρ. Если $\rho(x_1, x_2) = 0$, величины x_1 и x_2 называются *некоррелированными*.

Рассмотрим теперь среднее значение случайной величины

$$\left[t\left(\frac{x_1 - \mu_1}{\sigma_1}\right) + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)\right]^2,$$

где t — вещественная постоянная. Мы имеем

$$\mathscr{E}\left[t\left(\frac{x_1 - \mu_1}{\sigma_1}\right) + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)\right]^2 = t^2 + 2t\rho + 1 \geqslant 0. \tag{3.4.7}$$

Очевидно, $t^2 + 2t\rho + 1 \geqslant 0$, при всех вещественных t тогда и только тогда, когда $\rho^2 - 1 \geqslant 0$. Следовательно,

3.4.2. *Коэффициент корреляции* ρ *удовлетворяет условию* $-1 \leqslant \rho \leqslant 1$.

Если величины x_1 и x_2 собственно линейно зависимы [см. § 2.8 (b)], то $P(x_2 = \beta_0 + \beta_1 x_1) = 1$, где $\beta_1 \neq 0$. Отсюда $\sigma^2(x_2 - \beta_0 - \beta_1 x_1) = 0$ и тем самым $\rho = +1$ или $\rho = -1$ в зависимости от того, какой знак имеет число β_1, плюс или минус. Обратно, если $\rho = \pm 1$, случайная величина

$$\mp\left(\frac{x_1 - \mu_1}{\sigma_1}\right) + \left(\frac{x_2 - \mu_2}{\sigma_2}\right) \tag{3.4.8}$$

является вырожденной со средним значением нуль. Это означает, что

$$P\left(\mp\left(\frac{x_1 - \mu_1}{\sigma_1}\right) + \left(\frac{x_2 - \mu_2}{\sigma_2}\right) = 0\right) = 1. \tag{3.4.9}$$

Таким образом,

3.4.3. *Необходимое и достаточное условие для того, чтобы невырожденные случайные величины* x_1 *и* x_2 *были (собственно) линейно зависимы, дается равенством* $\rho^2 = 1$.

Моменты $\mu'_{r_1 r_2}$ *и центральные моменты* $\mu_{r_1 r_2}$ определяются следующим образом:

$$\mu'_{r_1 r_2} = \mathscr{E}(x_1^{r_1} x_2^{r_2}) \tag{3.4.10}$$

и

$$\mu_{r_1 r_2} = \mathscr{E}[(x_1 - \mu_1)^{r_1} (x_2 - \mu_2)^{r_2}]. \tag{3.4.11}$$

В этих обозначениях μ'_{10} и μ'_{01} являются соответственно *средними*, а μ_{20} и μ_{02} — *дисперсиями* величин x_1 и x_2. *Ковариация* между x_1 и x_2 равна μ_{11} и коэффициент корреляции $\rho_{12} = \mu_{11}/\sqrt{\mu_{20}\mu_{02}}$.

Наконец, *абсолютные моменты* $\nu'_{r_1 r_2}$ и *факториальные моменты* $\mu'_{[r_1][r_2]}$ вводятся посредством соотношений

$$\nu'_{r_1 r_2} = \mathcal{E}(|x_1|^{r_1}\,|x_2|^{r_2}), \tag{3.4.12}$$

$$\nu_{r_1 r_2} = \mathcal{E}(|x_1 - \mu_1|^{r_1}\,|x_2 - \mu_2|^{r_2}), \tag{3.4.13}$$

$$\mu'_{[r_1][r_2]} = \mathcal{E}(x_1^{[r_1]} x_2^{[r_2]}). \tag{3.4.14}$$

3.5. Моменты k-мерных случайных величин

Все предшествующие определения допускают непосредственное распространение на k-мерные случайные величины. Так, например, если $(x_1, \ldots, x_k)$ — k-мерная случайная величина, имеющая к. ф. р. $F(x_1, \ldots, x_k)$, *среднее значение* для x_i дается формулой (3.2.6) при $g(x_1, \ldots, x_k) = x_i$ или же формулой (3.2.5), в которой $H(y)$ есть маргинальная к. ф. р. величины x_i. Таким образом,

$$\mu_i = \mathcal{E}(x_i) = \int\limits_{R_k} x_i\, dF(x_1, \ldots, x_k) = \int\limits_{-\infty}^{\infty} x_i\, dF_i(x_i). \tag{3.5.1}$$

Аналогично этому *дисперсия* σ_i^2 величины x_i определяется через маргинальную к. ф. р. $F_i(x_i)$ согласно (3.3.2) и *ковариация* между x_i и x_j — через маргинальную к. ф. р. $F_{ij}(x_i, x_j)$ согласно (3.4.3).

Совокупность дисперсий k компонент случайной величины $(x_1, \ldots, x_k)$ и всех ковариаций между ними образует симметричную матрицу

$$\|\sigma_{ij}\|, \qquad i, j = 1, \ldots, k, \tag{3.5.2}$$

порядка $k \times k$, называемую *ковариационной* матрицей. Здесь σ_{ii} — дисперсия величины x_i и σ_{ij}, $i \neq j$, — ковариация между x_i и x_j. Иногда, если вообще не исключена возможность внести путаницу в рассматриваемые случайные величины, мы будем писать ковариационную матрицу в виде $\|\sigma(x_i, x_j)\|$.

Удобно говорить, что случайная величина $(x_1, \ldots, x_k)$ имеет *среднее* $(\mu_1, \ldots, \mu_k)$ и *ковариационную матрицу* $\|\sigma_{ij}\|$.

Мы будем существенно использовать не только ковариационную матрицу (3.5.2), но и обратную к ней матрицу, именно $\|\sigma_{ij}\|^{-1}$, которая будет обозначаться символом

$$\|\sigma^{ij}\|, \tag{3.5.3}$$

где формально σ^{ij} можно представить как отношение алгебраического дополнения элемента σ_{ij} матрицы (3.5.2) к $|\sigma_{ij}|$ — определителю этой матрицы. Заметим, что $\|\sigma^{ij}\|$ будет существовать, только если $|\sigma_{ij}| \neq 0$. Заметим также, что если матрица $\|\sigma_{ij}\|$ имеет обратную $\|\sigma^{ij}\|$, то

$|\sigma^{ij}| = \frac{1}{|\sigma_{ij}|}$. Аналогичные утверждения справедливы для *корреляционной матрицы* $\|\rho_{ij}\|$.

Напомним, что приведенное в **3.4.3** условие, необходимое и достаточное для того, чтобы невырожденные случайные величины x_1 и x_2 были собственно линейно зависимы, есть $\rho^2 = 1$. Равносильное необходимое и достаточное условие можно выразить равенством

$$\begin{vmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{vmatrix} = 0.$$

Полезно иметь критерий линейной зависимости и в случае k-мерной случайной величины. Как было указано в § 2.8 (d), компоненты случайной величины $(x_1, \ldots, x_k)$ собственно линейно зависимы, если существует такой набор констант $c_1, \ldots, c_k$, каждая из которых отлична от нуля, что величина $c_1x_1 + \ldots + c_kx_k$ является вырожденной. Сформулируем критерий собственной линейной зависимости компонент $(x_1, x_2, \ldots, x_k)$ как результат следующего обобщения теоремы **3.4.3**.

3.5.1. *Необходимое и достаточное условие для того, чтобы компоненты случайной величины* $(x_1, \ldots, x_k)$, *ни одна из которых не является вырожденной, были собственно линейно зависимы, заключается в выполнении соотношения* $|\sigma_{ij}| = 0$.

Сначала рассмотрим д о с т а т о ч н о с т ь условия. Так как ни одна из компонент не является вырожденной, $\sigma_{ii} > 0$, $i = 1, \ldots, k$. Тогда, если $|\sigma_{ij}| = 0$, согласно хорошо известному свойству определителей существует набор вещественных отличных от нуля констант c_i, $i = 1, \ldots, k$, таких, что $\sum_{i=1}^{n} \sigma_{ij}c_i = 0$ для $j = 1, \ldots, k$. Умножая эти уравнения на c_j и затем суммируя по j, мы получаем

$$\sum_{i,j=1}^{k} \sigma_{ij}c_ic_j = 0, \tag{3.5.4}$$

что может быть переписано в виде

$$\mathcal{E}\left[\sum_{i=1}^{k} (x_i - \mu_i)c_i\right]^2 = 0. \tag{3.5.5}$$

Но левая часть (3.5.5) представляет собой дисперсию суммы $\sum_{i=1}^{k} c_ix_i$.

Поскольку она равна нулю, величина $\sum_{i=1}^{k} c_ix_i$ является вырожденной. К тому же $c_i \neq 0$, $i = 1, \ldots, k$, т. е. по определению величины $x_1, \ldots, x_k$ собственно линейно зависимы.

Теперь покажем, что условие теоремы необходимо. Если $x_1, \ldots, x_k$ находятся в собственной линейной зависимости, то по определению при некоторых вещественных и отличных от нуля константах c_i случайная величина $\sum_{i=1}^{k} x_i c_i$ является вырожденной. Следовательно, для этой величины

$$\sigma^2\left(\sum_{i=1}^{k} c_i x_i\right) = \sum_{i,\,j=1}^{k} \sigma_{ij} c_i c_j = 0. \tag{3.5.6}$$

Рассмотрим, далее, величину $\psi(c_1, \ldots, c_k) = \sum_{i,\,j=1}^{k} \sigma_{ij} c_i c_j$ как функцию от c_i. Так как $\psi(c_1, \ldots, c_k)$ не может принимать отрицательных значений, она, очевидно, достигает своего минимума, равного нулю, в некоторой точке пространства всех $c = (c_1, \ldots, c_k)$, имеющей по предположению ненулевые координаты c_i. Но значения c_i, минимизирующие функцию $\psi(c_1, \ldots, c_k)$, удовлетворяют условиям

$$\frac{\partial \psi}{\partial c_i} = 0, \qquad i = 1, \ldots, k. \tag{3.5.7}$$

Это приводит к системе уравнений

$$\sum_{j=1}^{k} \sigma_{ij} c_j = 0, \qquad i = 1, \ldots, k, \tag{3.5.8}$$

и для того чтобы эта система имела решение, отличное от нулевого, необходимо выполнение условия $|\sigma_{ij}| = 0$, что и завершает доказательство теоремы **3.5.1.**

Если для вещественных $c_1, \ldots, c_k$ квадратичная форма $\sum_{i,\,j=1}^{k} \sigma_{ij} c_i c_j$ обращается в нуль (т. е. достигает абсолютного минимума) только при $c_1 = \ldots = c_k = 0$, она называется *положительно определенной формой* и ее матрица $\|\sigma_{ij}\|$ — *положительно определенной матрицей.*

Один из простых критериев положительной определенности ковариационной матрицы можно сформулировать следующим образом.

3.5.2. *Необходимое и достаточное условие для того, чтобы ковариационная матрица $\|\sigma_{ij}\|$ k-мерной случайной величины $(x_1, \ldots, x_k)$ с невырожденными компонентами была положительно определена, состоит в отсутствии линейной зависимости между этими компонентами.*

Доказательство этого утверждения предоставляется читателю.

Приведем теперь одно свойство квадратичных форм, которое будет использовано в дальнейшем.

3.5.3. *Если форма $\sum_{i,\,j=1}^{k} \sigma_{ij} c_i c_j$ положительно определена, то также положительно определена и квадратичная форма $\sum_{i,\,j=1}^{k} \sigma^{ij} c_i c_j$.*

Доказательство этой теоремы очевидно и потому опускается.

Моменты более высоких порядков $\mu'_{r_1 \ldots r_k}$, $\mu_{r_1 \ldots r_k}$, $\nu'_{r_1 \ldots r_k}$, $\nu_{r_1 \ldots r_k}$ и $\mu'_{[r_1] \ldots [r_k]}$, относящиеся к k случайным величинам $(x_1, \ldots, x_k)$, определяются путем естественного обобщения формул (3.4.10) — (3.4.14).

3.6. Средние, дисперсии и ковариации линейных функций случайных величин

Наиболее часто встречающимся типом функции нескольких случайных величин является линейная функция. Ниже следует ряд полезных фактов относительно средних, дисперсий и ковариаций для таких функций.

3.6.1. *Если $(x_1, \ldots, x_k)$ — k-мерная случайная величина, имеющая среднее $(\mu_1, \ldots, \mu_k)$ и ковариационную матрицу $\|\sigma_{ij}\|$, то среднее и дисперсия линейной функции*

$$L = \sum_{i=1}^{k} c_i x_i, \tag{3.6.1}$$

где $c_1, \ldots, c_k$ — некоторые константы, соответственно равны

$$\mu(L) = \sum_{i=1}^{k} c_i \mu_i \tag{3.6.2}$$

и

$$\sigma^2(L) = \sum_{i,\, j=1}^{k} \sigma_{ij} c_i c_j. \tag{3.6.3}$$

Доказательство предоставляется читателю.

В том случае, когда все x_i независимы, так что согласно **3.4.1** $\sigma_{ij} = 0$, $i \neq j$, мы находим следующее важное следствие теоремы **3.6.1**.

3.6.1а. *Если $x_1, \ldots, x_k$ — независимые случайные величины с дисперсиями $\sigma_1^2, \ldots, \sigma_k^2$ соответственно, то дисперсия линейной функции* (3.6.1) *дается формулой*

$$\sigma^2(L) = \sum_{i=1}^{k} c_i^2 \sigma_i^2. \tag{3.6.4}$$

Сформулируем более общее утверждение.

3.6.2. *Если $L_p = \sum_{i=1}^{k} c_{ip} x_i$, $p = 1, \ldots, s$, есть линейные функции случайных величин $x_1, \ldots, x_k$, то величина $(L_1, \ldots, L_s)$ имеет среднее $(\mu(L_1), \ldots, \mu(L_s))$, где*

$$\mu(L_p) = \sum_{i=1}^{k} c_{ip} \mu_i, \qquad p = 1, \ldots, s, \tag{3.6.5}$$

и ковариационную матрицу

$$\| \sigma^2(L_p, L_q) \| = \left\| \sum_{i,\, j=1}^{k} \sigma_{ij} c_{ip} c_{jq} \right\|. \tag{3.6.6}$$

Следует отметить, что если $s > k$, то величины L_p необходимо линейно зависимы.

Если x_i статистически независимы и имеют дисперсии σ_i^2, $i = 1, \ldots$ $\ldots, k$, то ковариационная матрица $\| \sigma(L_p, L_q) \|$ сводится к матрице $\left\| \sum_{i=1}^{k} \sigma_i^2 c_{ip} c_{iq} \right\|$.

3.7. Средние значения условных случайных величин

(a) Случай двух переменных. Пусть $x_2 \mid x_1$ — условная случайная величина и $F(x_2 \mid x_1)$ — ее к. ф. р., как это было определено в § 2.9 (b). Среднее значение величины $x_2 \mid x_1$, если оно вообще существует, есть число

$$\mu(x_2 \mid x_1) = \mathscr{E}(x_2 \mid x_1) = \int_{-\infty}^{\infty} x_2 \, dF(x_2 \mid x_1). \tag{3.7.1}$$

Это же число, рассматриваемое как функция от x_1, называется *функцией регрессии* величины x_2 на x_1. Образно говоря, она представляет положение центра тяжести распределения условной случайной величины $x_2 \mid x_1$ в зависимости от значения x_1. В частности, если

$$\mu(x_2 \mid x_1) = \beta_0 + \beta_1 x_1, \tag{3.7.2}$$

мы имеем *функцию линейной регрессии* величины x_2 на x_1; β_0 и β_1 называются *коэффициентами регрессии.*

Более общо, среднее значение условной случайной величины $g(x_2) \mid x_1$ выражается в виде интеграла

$$\mathscr{E}(g(x_2) \mid x_1) = \int_{-\infty}^{\infty} g(x_2) \, dF(x_2 \mid x_1). \tag{3.7.3}$$

Например, дисперсия величины $x_2 \mid x_1$ есть число

$$\sigma^2(x_2 \mid x_1) = \mathscr{E}\left[(x_2 - \mu(x_2 \mid x_1))^2 \mid x_1\right]. \tag{3.7.4}$$

Иногда для обозначения $\mu(x_2 \mid x_1)$ и $\sigma^2(x_2 \mid x_1)$ используются соответственно термины *условное среднее* и *условная дисперсия* величины x_2 при данном x_1. Кроме того, $\sigma^2(x_2 \mid x_1)$ иногда называют *остаточной дисперсией* величины x_2 на x_1.

Нижеследующее утверждение является следствием теоремы **3.3.1** и несет полезную информацию о взаимоотношении между $\sigma^2(x_2 \mid x_1)$ и $\mu(x_2 \mid x_1)$.

3.7.1. *Если $x_2 | x_1$ — условная случайная величина, введенная нами в § 2.9 (b), то наименьшее значение условного среднего $\mathcal{E}[(x_2 - u(x_1))^2 | x_1]$ в классе вещественных однозначных функций $u(x_1)$ достигается при $u(x_1) = \mu(x_2 | x_1)$, и это наименьшее значение равно $\sigma^2(x_2 | x_1)$.*

Если $\sigma^2(x_2 | x_1) = 0$ для всех точек x_1, в которых величина $x_2 | x_1$ определена, то, очевидно, $P[x_2 = \mu(x_2 | x_1) | x_1] = 1$. Это, конечно, означает, что если значение случайной величины x_1 известно, тогда для всех исходов, которым это значение отвечает, за исключением множества их вероятности нуль, величина x_2 равна $\mu(x_2 | x_1)$.

Полагая в формуле (3.7.3) $g(x_2) = x_2^r$, мы получаем r-й момент условной случайной величины $x_2 | x_1$. r-й центральный и факториальный моменты величины $x_2 | x_1$ определяются аналогичным образом.

При фактическом вычислении среднего значения функции $g(x_1, x_2)$ двумерной величины (x_1, x_2) в отдельных простых случаях удобно переходить к повторному интегрированию. Условные случайные величины играют здесь важную роль, как это видно из теоремы, приводимой ниже без доказательства.

3.7.2. *Пусть (x_1, x_2) — случайная величина с к. ф. р. $F(x_1, x_2)$ и, кроме того, пусть $x_2 | x_1$ и $x_1 | x_2$ — введенные в § 2.9 (b) случайные величины с к. ф. р. соответственно $F(x_2 | x_1)$ и $F(x_1 | x_2)$. Тогда, если $g(x_1, x_2)$ является случайной величиной, то*

$$\int_{R_2} g(x_1, x_2)\, dF(x_1, x_2) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(x_1, x_2)\, dF(x_2 | x_1) \right] dF_1(x_1) =$$
$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(x_1, x_2)\, dF(x_1 | x_2) \right] dF_2(x_2) \qquad (3.7.5)$$

или в более компактной форме

$$\mathcal{E}(g(x_1, x_2)) = \mathcal{E}_{(x_1)}[\mathcal{E}_{(x_2)}(g(x_1, x_2) | x_1)] =$$
$$= \mathcal{E}_{(x_2)}[\mathcal{E}_{(x_1)}(g(x_1, x_2) | x_2)]. \qquad (3.7.5\text{a})$$

Читатель легко может убедиться в справедливости этой теоремы для более простых и вместе с тем часто встречающихся типов условных случайных величин из числа рассмотренных в § 2.9.

Распространение утверждений **3.7.1** и **3.7.2** на более общий класс условных случайных величин можно осуществить на основе теоремы Радона — Никодима **1.10.1.**

(b) Случай нескольких переменных. Если $x_k | x_1, \ldots, x_{k-1}$ — условная случайная величина с к. ф. р. $F(x_k | x_1, \ldots, x_{k-1})$ в смысле параграфа 2.9 (c), то среднее значение условной случайной величины $g(x_k) | x_1, \ldots, x_{k-1}$ дается равенством

$$\mathcal{E}(g(x_k) | x_1, \ldots, x_{k-1}) = \int_{-\infty}^{\infty} g(x_k)\, dF(x_k | x_1, \ldots, x_{k-1}). \qquad (3.7.6)$$

В частности, среднее и дисперсия величины $x_k \mid x_1, \ldots, x_{k-1}$ будут равны

$$\mu(x_k \mid x_1, \ldots, x_{k-1}) = \mathscr{E}(x_k \mid x_1, \ldots, x_{k-1}) \tag{3.7.7}$$

и

$$\sigma^2(x_k \mid x_1, \ldots, x_{k-1}) = \\ = \mathscr{E}\left[(x_k - \mu(x_k \mid x_1, \ldots, x_{k-1}))^2 \mid x_1, \ldots, x_{k-1}\right]. \tag{3.7.8}$$

Читатель заметит, что теорема **3.7.1** может быть немедленно перефразирована для случайной величины $x_k \mid x_1, \ldots, x_{k-1}$.

Вообще, если $x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1}$, $k = k_1 + k_2$, есть k_2-мерная условная случайная величина с к. ф. р. $F(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1})$, соотношение

$$\mathscr{E}\left[g(x_{k_1+1}, \ldots, x_k) \mid x_1, \ldots, x_{k_1}\right] = \\ = \int\limits_{R_{k_2}} g(x_{k_1+1}, \ldots, x_k)\, dF(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1}) \tag{3.7.9}$$

определяет среднее значение случайной величины

$$g(x_{k_1+1}, \ldots, x_k) \mid x_1, \ldots, x_{k_1}.$$

Ясно, каким образом вводятся теперь такие количества, как ковариация между двумя условными случайными величинами, например, $x_k \mid x_1, \ldots, x_{k-2}$ и $x_{k-1} \mid x_1, \ldots, x_{k-2}$, смешанные моменты и т. д.

Число $\mu(x_k \mid x_1, \ldots, x_{k-1})$, рассматриваемое как функция от $x_1, \ldots, x_{k-1}$, называется *функцией регрессии* величины x_k на $x_1, \ldots, x_{k-1}$. В том частном случае, когда

$$\mu(x_k \mid x_1, \ldots, x_{k-1}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_{k-1} x_{k-1}, \tag{3.7.10}$$

величина x_k обладает *функцией линейной регрессии* на $x_1, \ldots, x_{k-1}$.

Наконец, читателю следует обратить внимание, что рамки теоремы **3.7.2** можно расширить так, что в качестве обобщения формулы (3.7.5) мы будем иметь

$$\int\limits_{R_k} g(x_1, \ldots, x_k)\, dF(x_1, \ldots, x_k) = \\ = \int\limits_{R_{k_1}} \left[\int\limits_{R_{k_2}} g(x_1, \ldots, x_k)\, dF(x_{k_1+1}, \ldots, x_k \mid x_1, \ldots, x_{k_1})\right] \times \\ \times dF_{1, \ldots, k_1}(x_1, \ldots, x_{k_1}) \tag{3.7.11}$$

или, переходя к повторному интегралу,

$$\int\limits_{R_k} g(x_1, \ldots, x_k)\, dF(x_1, \ldots, x_k) = \\ = \int\limits_{-\infty}^{\infty} \ldots \int\limits_{-\infty}^{\infty} g(x_1, \ldots, x_k)\, dF(x_k \mid x_1, \ldots, x_{k-1}) \times \\ \times dF_{1, \ldots, k-1}(x_{k-1} \mid x_1, \ldots, x_{k-2}) \ldots dF_1(x_1). \tag{3.7.12}$$

Здесь, конечно, существует $k!$ мыслимых перемен порядка итерации

Отметим, что путем использования теоремы Радона — Никодима формулам (3.7.6) — (3.7.12) можно придать смысл и при более общих условиях, чем те, которые были наложены на $F(x_1, \ldots, x_k)$ в § 2.9 (с).

(с) Корреляционное отношение. Дисперсия условной случайной величины $x_2 \mid x_1$ доставляет некоторую информацию о том, насколько точно может быть оценена компонента x_2 выборочной точки в R_2, когда нам известно значение компоненты x_1 этой точки. Если $\sigma^2(x_2 \mid x_1) = 0$, компонента x_2 определяется с вероятностью единица. Если $\sigma^2(x_2 \mid x_1) \neq 0$, то известное представление о точности оценивания можно получить, беря отношение $[\sigma^2(x_2 \mid x_1)]/\sigma^2(x_2)$, где, разумеется, $\sigma^2(x_2)$ считается отличной от нуля. Однако, за редким исключением, это отношение зависит от x_1. Более полезный критерий получается иногда в результате рассмотрения среднего значения этого отношения относительно x_1. Обозначая это среднее значение через $\eta^2_{2\cdot 1}$, мы имеем

$$\eta^2_{2\cdot 1} = \frac{\mathcal{E}\,[\sigma^2(x_2 \mid x_1)]}{\sigma^2(x_2)}, \tag{3.7.13}$$

где

$$\mathcal{E}\,[\sigma^2(x_2 \mid x_1)] = \int_{-\infty}^{\infty} \sigma^2(x_2 \mid x_1)\, dF_1(x_1). \tag{3.7.14}$$

Число $\eta^2_{2\cdot 1}$ называется *корреляционным отношением* величины x_2 на x_1. Очевидно, $0 \leqslant \eta^2_{2\cdot 1} \leqslant 1$. Кроме того, $\eta^2_{2\cdot 1} = 0$, только если $\sigma^2(x_2 \mid x_1) = 0$, т. е. если $P[(x_2 = \mu(x_2 \mid x_1)) \mid x_1] = 1$. С другой стороны, $\eta^2_{2\cdot 1} = 1$ тогда и только тогда, когда $\mu(x_2 \mid x_1)$ не зависит от x_1, например, для независимых x_1 и x_2.

Множественное корреляционное отношение есть число

$$\eta^2_{k\cdot 12\ldots(k-1)} = \frac{\mathcal{E}\,[\sigma^2(x_k \mid x_1, \ldots, x_{k-1})]}{\sigma^2(x_k)}, \tag{3.7.15}$$

где

$$\mathcal{E}\,[\sigma^2(x_k \mid x_1, \ldots, x_{k-1})] = \\ = \int_{R_{k-1}} \sigma^2(x_k \mid x_1, \ldots, x_{k-1})\, dF_{1\ldots(k-1)}(x_1, \ldots, x_{k-1}). \tag{3.7.16}$$

Как и в случае корреляционного отношения $\eta^2_{2\cdot 1}$, $0 \leqslant \eta^2_{k\cdot 12\ldots(k-1)} \leqslant 1$, причем значение 0 достигается, только если $P[(x_k = \mu(x_k \mid x_1, \ldots$ $\ldots, x_{k-1})) \mid x_1, \ldots, x_{k-1}] = 1$, и значение 1, если, например, x_k и $(x_1, \ldots, x_{k-1})$ независимы.

3.8. Линейная средняя квадратическая регрессия

(а) Случай двух переменных. Предположим, что в качестве функций $u(x_1)$, упоминаемых в теореме **3.7.1**, мы рассматриваем только линейные функции вида $\beta_0 + \beta_1 x_1$. Определим величины β_0 и β_1, при которых функция $\mathcal{E}\,(x_2 - \beta_0 - \beta_1 x_1)^2$ принимает наименьшее значение.

Как обычно, мы считаем, что ни x_1, ни x_2 не являются вырожденными, т. е. обе дисперсии σ_1^2 и σ_2^2 положительны.

Обозначая эту функцию через $\psi(\beta_0, \beta_1)$, мы можем написать

$$\psi(\beta_0, \beta_1) = \mathcal{E}\,[(x_2 - \mu_2) - (\beta_0 - \mu_2 + \beta_1\mu_1) - \beta_1(x_1 - \mu_1)]^2, \quad (3.8.1)$$

где $\mu_1 = \mathcal{E}(x_1)$ и $\mu_2 = \mathcal{E}(x_2)$. Величины β_0 и β_1, минимизирующие $\psi(\beta_0, \beta_1)$, будут удовлетворять уравнениям

$$\frac{\partial\psi}{\partial\beta_0} = 0, \qquad \frac{\partial\psi}{\partial\beta_1} = 0. \quad (3.8.2)$$

Используя обозначения § 3.4 и упрощая эти уравнения, мы получаем систему

$$\beta_0 - \mu_2 + \beta_1\mu_1 = 0, \qquad \rho\sigma_1\sigma_2 - \beta_1\sigma_1^2 = 0, \quad (3.8.3)$$

которая имеет единственное решение

$$\beta_0^* = \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1, \qquad \beta_1^* = \rho\frac{\sigma_2}{\sigma_1}. \quad (3.8.4)$$

Таким образом,

3.8.1. *Линейная функция $u(x_1)$, минимизирующая среднее $\mathcal{E}(x_2 - u(x_1))^2$, дается формулой*

$$u(x_1) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1). \quad (3.8.5)$$

Прямая, имеющая уравнение

$$x_2 = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1) \quad (3.8.6)$$

или, в более симметричном виде,

$$\frac{(x_2 - \mu_2)}{\sigma_2} = \rho\,\frac{(x_1 - \mu_1)}{\sigma_1}, \quad (3.8.7)$$

называется *прямой средней квадратической регрессии* величины x_2 на x_1.

Легко проверить следующее утверждение.

3.8.2. *Минимум среднего $\mathcal{E}(x_2 - \beta_0 - \beta_1 x_1)^2$, который достигается при значениях β_0 и β_1, определяемых формулой* (3.8.4), *равен*

$$\sigma_{2\cdot1}^2 = \sigma_2^2(1 - \rho^2). \quad (3.8.8)$$

Число $\sigma_{2\cdot1}^2$ называется *средней квадратической остаточной дисперсией* величины x_2 на x_1. Оно обращается в нуль тогда и только тогда, когда $\rho = \pm 1$, т. е. когда в согласии с теоремой **3.4.3** x_1 и x_2 (по предположению невырожденные) собственно линейно зависимы.

Меняя x_1 и x_2 ролями, мы получаем прямую средней квадратической регрессии величины x_1 на x_2

$$\frac{(x_1 - \mu_1)}{\sigma_1} = \rho\,\frac{(x_2 - \mu_2)}{\sigma_2}. \quad (3.8.9)$$

Так как ни x_1, ни x_2 не являются вырожденными, ясно, что две линии регрессии (3.8.7) и (3.8.9) будут совпадать тогда и только тогда, когда $\rho = \pm 1$.

Естественно возникает вопрос, существуют ли специального вида двумерные распределения вероятностей, обладающие тем свойством, что для них функции средней квадратической регрессии тождественны с функциями действительной регрессии. Ответ на этот вопрос утвердительный. Одним из наиболее важных распределений в математической статистике, которое обладает этим свойством, является двумерное нормальное или гауссовское распределение. Это распределение будет рассмотрено в § 7.3. К числу его свойств относится также независимость условной дисперсии $\sigma^2(x_2 \mid x_1)$ от x_1; фактически она равна $\sigma^2_{2\cdot 1}$.

Критерий, показывающий, насколько эффективно может быть использована прямая средней квадратической регрессии для определения значения x_2 при некотором исходе, когда x_1 задано, состоит в свободном от влияния x_1 сравнении дисперсии величины x_2 относительно функции ее средней квадратической регрессии на x_1 с обычной дисперсией этой величины, т. е. в сравнении $\sigma^2_{2\cdot 1}$ и σ^2_2. Отношение этих двух количеств, которое мы записываем следующим образом:

$$\eta^2_{2\cdot 1(L)} = \frac{\sigma^2_{2\cdot 1}}{\sigma^2_2} = 1 - \rho^2, \tag{3.8.10}$$

называется *линейным корреляционным отношением*. Если x_1 линейно связана с x_2, т. е. если x_1 и x_2 суть случайные величины, находящиеся в собственной линейной зависимости, то $\rho^2 = 1$ и $\eta^2_{2\cdot 1(L)} = 0$. Обратно, если $\rho^2 = 1$ (и $\eta^2_{2\cdot 1(L)} = 0$), то x_1 и x_2 линейно зависимы. В другом крайнем случае, когда x_1 и x_2 независимы, $\rho = 0$ и $\eta^2_{2\cdot 1(L)} = 1$, так что x_1 не дает никакой информации для определения x_2.

(b) Случай нескольких переменных. Рассмотрим теперь линейные функции вида

$$\beta_0 + \beta_1 x_1 + \ldots + \beta_{k-1} x_{k-1} \tag{3.8.11}$$

и определим $\beta_0, \ldots, \beta_{k-1}$ так, чтобы они минимизировали функцию

$$\psi(\beta_0, \ldots, \beta_{k-1}) = \mathscr{E}\,[x_k - \beta_0 - \beta_1 x_1 - \ldots - \beta_{k-1} x_{k-1}]^2. \tag{3.8.12}$$

Будет удобно переписать (3.8.12) подобно тому, как это было сделано в двумерном случае:

$$\psi = \mathscr{E}\,[(x_k - \mu_k) - (\beta_0 - \mu_k + \beta_1\mu_1 + \ldots + \beta_{k-1}\mu_{k-1}) - \\ - \beta_1(x_1 - \mu_1) - \ldots - \beta_{k-1}(x_{k-1} - \mu_{k-1})]^2. \tag{3.8.13}$$

Значения $\beta_0, \ldots, \beta_{k-1}$, которые минимизируют (3.8.13), получаются в результате решения уравнений

$$\frac{\partial\psi}{\partial\beta_\alpha} = 0, \qquad \alpha = 0, 1, \ldots, k-1. \tag{3.8.14}$$

Первое уравнение $\frac{\partial\psi}{\partial\beta_0} = 0$ приводит к соотношению

$$\beta_0 - \mu_k + \beta_1 \mu_1 + \ldots + \beta_{k-1} \mu_{k-1} = 0. \tag{3.8.15}$$

Учитывая этот результат, мы получаем из остальных $(k-1)$ уравнений (3.8.14) следующие равенства:

$$\begin{aligned} &\beta_1\sigma_{11}+\ldots+\beta_{k-1}\sigma_{1\,k-1}=\sigma_{1k},\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &\beta_1\sigma_{1\,k-1}+\ldots+\beta_{k-1}\sigma_{k-1,\,k-1}=\sigma_{k-1,\,k}, \end{aligned} \tag{3.8.16}$$

где $\|\sigma_{ij}\|$, $i, j=1, \ldots, k$, есть ковариационная матрица компонент случайной величины $(x_1, \ldots, x_k)$, введенная выше, в § 3.5. Отсюда следует, что

3.8.3. *Если* $|\sigma_{pq}|\neq 0$, $p, q=1, \ldots, k-1$, *и значения* $\beta_0, \beta_1, \ldots, \beta_{k-1}$, *минимизирующие* ψ, *соответственно равны* $\beta_0^*, \beta_1^*, \ldots, \beta_{k-1}^*$, *то*

$$\begin{aligned} &\beta_0^*=\mu_k-\beta_1^*\mu_1-\ldots-\beta_{k-1}^*\mu_{k-1},\\ &\beta_p^*=\sum_{q=1}^{k-1}\sigma_{(k)}^{pq}\sigma_{qk},\quad p=1, \ldots, k-1, \end{aligned} \tag{3.8.17}$$

где $\|\sigma_{(k)}^{pq}\|$ — *матрица, обратная к ковариационной матрице* $\|\sigma_{pq}\|$, $p, q=1, \ldots, k-1$.

Уместно вспомнить, что согласно теореме **3.5.1** условие $|\sigma_{pq}|\neq 0$ означает отсутствие линейной зависимости между $x_1, \ldots, x_{k-1}$.

Наконец, мы можем утверждать, что *гиперплоскость средней квадратической регрессии* величины x_k на $x_1, \ldots, x_{k-1}$ имеет уравнение

$$x_k=\mu_k+\sum_{p=1}^{k-1}\beta_p^*(x_p-\mu_p). \tag{3.8.18}$$

Уравнение (3.8.18) можно также представить в более удобной для запоминания детерминантной форме

$$\begin{vmatrix} \sigma_{11} & \ldots\ldots & \sigma_{1k}\\ \ldots & \ldots & \ldots\\ \sigma_{k-1,\,1} & \ldots\ldots & \sigma_{k-1,\,k}\\ (x_1-\mu_1) & \ldots\ldots & (x_k-\mu_k) \end{vmatrix}=0. \tag{3.8.19}$$

Читатель докажет это, разлагая определитель по элементам нижней строки.

Рассмотрим, далее, задачу нахождения минимума функции $\psi(\beta_0, \ldots$ $\ldots, \beta_{k-1})$ относительно $\beta_0, \ldots, \beta_{k-1}$. Подставляя в (3.8.13) значения $\beta_0^*, \ldots$ $\ldots, \beta_{k-1}^*$, взятые из (3.8.17), и обозначая минимум ψ через $\sigma_{k\cdot 12\ldots(k-1)}^2$, мы имеем

$$\begin{aligned} \sigma_{k\cdot 12\ldots(k-1)}^2&=\mathcal{E}\,[(x_k-\mu_k)-\beta_1^*(x_1-\mu_1)-\ldots-\beta_{k-1}^*(x_{k-1}-\mu_{k-1})]^2=\\ &=\sigma_{kk}-2\sum_{p=1}^{k-1}\sigma_{pk}\beta_p^*+\sum_{p,\,q=1}^{k-1}\sigma_{pq}\beta_p^*\beta_q^*, \end{aligned} \tag{3.8.20}$$

причем

$$\sum_{p=1}^{k-1} \sigma_{pk}\beta_p^* = \sum_{p,\,q=1}^{k-1} \sigma_{(k)}^{pq}\sigma_{pk}\sigma_{qk}, \tag{3.8.21}$$

$$\sum_{p,\,q=1}^{k-1} \sigma_{pq}\beta_p^*\beta_q^* = \sum_{p,\,q=1}^{k-1}\sum_{r,\,s=1}^{k-1} \sigma_{pq}\sigma_{(k)}^{ps}\sigma_{(k)}^{qr}\sigma_{rk}\sigma_{sk} = \sum_{r,\,s=1}^{k-1} \sigma_{(k)}^{rs}\sigma_{rk}\sigma_{sk}, \tag{3.8.22}$$

ибо $\sum\limits_{p=1}^{k-1} \sigma_{pq}\sigma_{(k)}^{ps} = \delta_{qs}$, где δ_{qs} — символ Кронекера, равный 1, если $q = s$, и 0, если $q \neq s$. Крайний член справа в (3.8.22), очевидно, совпадает с правой частью (3.8.21). Следовательно,

$$\sigma_{k\cdot 12\,\ldots\,(k-1)}^2 = \sigma_{kk} - \sum_{p,\,q=1}^{k-1} \sigma_{(k)}^{pq}\sigma_{pk}\sigma_{qk}. \tag{3.8.23}$$

Но

$$\sigma_{kk} - \sum_{p,\,q=1}^{k-1} \sigma_{(k)}^{pq}\sigma_{pk}\sigma_{qk} = \frac{|\sigma_{ij}|}{|\sigma_{pq}|}, \tag{3.8.24}$$

где $i, j = 1, \ldots, k$ и $p, q = 1, \ldots, k-1$, что можно усмотреть, разлагая определитель $|\sigma_{ij}|$ по элементам последней строки, а затем и алгебраические дополнения этих элементов (за исключением алгебраического дополнения элемента σ_{kk}) — по элементам их последнего столбца [см., например, Бохер (1929)]. Таким образом, в результате мы получаем

3.8.4. *Минимальное значение функции* $\psi(\beta_0, \beta_1, \ldots, \beta_{k-1})$, *обозначаемое через* $\sigma_{k\cdot 12\,\ldots\,(k-1)}^2$, *определяется равенством*

$$\sigma_{k\cdot 12\,\ldots\,(k-1)}^2 = \frac{|\sigma_{ij}|}{|\sigma_{pq}|}. \tag{3.8.25}$$

Число $\sigma_{k\cdot 12\,\ldots\,(k-1)}^2$ называется *средней квадратической остаточной дисперсией* величины x_k на $x_1, \ldots, x_{k-1}$; это — дисперсия величины x_k относительно плоскости средней квадратической регрессии, уравнение которой дается (3.8.18).

Коэффициент корреляции между случайной величиной x_k и функцией регрессии $\mu_k + \sum\limits_{p=1}^{k-1} \beta_p^*(x_p - \mu_p)$ называется *множественным коэффициентом корреляции* между x_k и $(x_1, \ldots, x_{k-1})$. Он обозначается символом $\rho_{k\cdot 12\,\ldots\,(k-1)}$ и выражается через элементы матрицы $\|\sigma_{ij}\|$ посредством следующей формулы:

$$\rho_{k\cdot 12\,\ldots\,(k-1)}^2 = 1 - \frac{|\sigma_{ij}|}{\sigma_{kk}\,|\sigma_{pq}|}. \tag{3.8.26}$$

Чтобы установить эту формулу, определим дисперсии величин x_k и $\mu_k + \sum_{p=1}^{k-1} \beta_p^* (x_p - \mu_p)$ и ковариацию между ними. Мы находим

$$\sigma^2 (x_k) = \sigma_{kk},$$
$$\sigma^2 [\mu_k + \sum_{p=1}^{k-1} \beta_p^* (x_p - \mu_p)] = \sum_{p, q=1}^{k-1} \sigma_{pq} \beta_p^* \beta_q^*,$$
$$\operatorname{cov} [x_k, \mu_k + \sum_{p=1}^{k-1} \beta_p^* (x_p - \mu_p)] = \sum_{p=1}^{k-1} \sigma_{pk} \beta_p^*.$$

Обращаясь к определению (3.4.6) коэффициента корреляции и используя формулы (3.8.21) и (3.8.22), мы можем написать

$$\rho_{k \cdot 12 \ldots (k-1)}^2 = \sqrt{\sum_{p, q=1}^{k-1} \sigma_{(k)}^{pq} \sigma_{pk} \sigma_{qk} / \sigma_{kk}}. \tag{3.8.27}$$

Заменяя здесь правую часть по формуле (3.8.24), приходим к (3.8.26).

Из (3.8.25) и (3.8.26) вытекает, что

$$\sigma_{k \cdot 12 \ldots (k-1)}^2 = \sigma_{kk} (1 - \rho_{k \cdot 12 \ldots (k-1)}^2). \tag{3.8.28}$$

Линейное корреляционное отношение есть величина

$$\eta_{k \cdot 12 \ldots (k-1), (L)}^2 = \frac{\sigma_{k \cdot 12 \ldots (k-1)}^2}{\sigma_{kk}} = 1 - \rho_{k \cdot 12 \ldots (k-1)}^2. \tag{3.8.29}$$

Так же как в случае линейного корреляционного отношения для двух случайных величин, $\sigma_{k \cdot 12 \ldots (k-1)}^2$ будет равна нулю тогда и только тогда, когда $\rho_{k \cdot 12 \ldots (k-1)}^2 = 1$, т. е. когда равна единице вероятность, сосредоточенная на плоскости регрессии величины x_1 на $(x_1, \ldots, x_{k-1})$.

ЗАДАЧИ

3.1. Показать, что если x — случайная величина, у которой существует первый абсолютный центральный момент, то среднее μ этой величины конечно и при $\lambda > 0$

$$P(|x - \mu| < \lambda \nu_1) | \geqslant 1 - \frac{1}{\lambda}.$$

3.2. Показать, что если $g(x)$ — измеримая функция случайной величины x, то $P(|g(x)| \geqslant \lambda) \leqslant \mathscr{E} |g(x)| / \lambda$.

3.3. Показать, что если случайная величина обладает первыми r моментами $\mu'_1, \ldots, \mu'_r$ (и первыми r центральными моментами $\mu_1, \ldots, \mu_r$), то

$$\mu'_r = \sum_{i=0}^{r} \binom{r}{i} \mu_{r-i} \mu_i$$

и

$$\mu_r = \sum_{i=0}^{r} (-1)^{r+i} \binom{r}{i} \mu'_{r-i} \mu_i.$$

3.4. Показать, что у любой дискретной случайной величины с точками сосредоточения массы 0, 1, 2, ..., r все факториальные моменты порядка выше, чем $[r]$, равны нулю.

3.5. Показать, что в случае существования моментов μ_{2r}, μ_{2r+1}, μ_{2r+2}

$$(\mu_{2r+1})^2 \leqslant \mu_{2r}\mu_{2r+2}.$$

3.6. Доказать теоремы **3.5.2** и **3.5.3**.

3.7. Показать, что если $(x_1, \ldots, x_k)$ — k-мерная случайная величина, обладающая средними $(\mu_1, \ldots, \mu_k)$ и (положительно определенной) ковариационной матрицей $\|\sigma_{ij}\|$, то

$$P\left(\sum_{i,j=1}^{k} \sigma^{ij}(x_i - \mu_i)(x_j - \mu_j) < \lambda^2\right) \geqslant 1 - \frac{k}{\lambda^2},$$

где σ^{ij} — элементы обратной к $\|\sigma_{ij}\|$ матрицы.

3.8. Доказать теорему **3.6.1**.

3.9. Показать, что если $x_1, \ldots, x_k$ — независимые случайные величины со средними, равными нулю, и дисперсиями, равными единице, то

$$P\left(\sum_{i=1}^{k} x_i^2 \geqslant \lambda k\right) \leqslant \frac{1}{\lambda}.$$

3.10. Показать, что если x — случайная величина со средним μ, дисперсией σ^2 и к. ф. р. $F(x)$, то

$$F(x-\mu)\begin{cases} \leqslant \dfrac{1}{1+\left(\dfrac{x-\mu}{\sigma}\right)^2}, & \text{если} \quad x<\mu, \\[2ex] \geqslant \dfrac{1}{1+\left(\dfrac{\sigma}{x-\mu}\right)^2}, & \text{если} \quad x>\mu, \end{cases}$$

— результат, принадлежащий Крамеру (1946).

3.11. Показать, что если моменты порядка $2r$ случайных величин x и y конечны, то также конечен момент порядка $2r$ величины $x+y$.

3.12. Показать, что если x — случайная величина со средним μ, дисперсией σ^2 и ф. п. в., достигающей абсолютного максимума в точке c, то

$$P\left(|x-c| \geqslant \lambda\sqrt{\sigma^2+(\mu-c)^2}\right) \leqslant \left(\frac{2}{3\lambda}\right)^2$$

— результат, принадлежащий Гауссу.

3.13. (*Продолжение*). Показать, что

$$P(|x-\mu| \geqslant \lambda\sigma) \leqslant \frac{4(1+\delta^2)}{9[\lambda - |\delta|]^2},$$

где $\delta = \dfrac{\mu - c}{\sigma}$ и $\lambda > |\delta|$. [Крамер (1946)].

3.14. Опираясь на неравенство Чебышева, показать, что если $x_1, \ldots, x_n$ — независимые случайные величины со средними, равными μ, и дисперсиями, равными σ^2, то для любого $\delta > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{x_1+\ldots+x_n}{n} - \mu\right| < \delta\right) = 1.$$

3.15. Показать, что если $(x_1, \ldots, x_k)$ — k-мерная случайная величина, причем коэффициент корреляции между любыми двумя ее компонентами равен одному и тому же числу ρ, то $-\dfrac{1}{k-1} \leqslant \rho \leqslant +1$.

3.16. Пусть $(x_1, \ldots, x_m, y_1, \ldots, y_n)$ — $(m+n)$-мерная случайная величина, причем дисперсии всех ее компонент равны единице, $\operatorname{cov}(x_i, x_j) = \rho_1$, $\operatorname{cov}(y_p, y_q) = \rho_2$ и $\operatorname{cov}(x_i, y_q) = \rho_3$. Показать, что если $u = x_1 + \ldots + x_m$ и $v = y_1 + \ldots + y_n$, то коэффициент корреляции между u и v равен

$$\rho(u, v) = \frac{\sqrt{mn}\,\rho_3}{\sqrt{1+(m-1)\rho_1}\sqrt{1+(n-1)\rho_2}}.$$

3.17. Показать, что если $x_1, \ldots, x_p, y_1, \ldots, y_q, z_1, \ldots, z_r$ — случайные величины и все дисперсии их равны единице, а все ковариации — нулю, то коэффициент корреляции между u и v, где $u = x_1 + \ldots + x_p + y_1 + \ldots + y_q$, $v = x_1 + \ldots + x_p + z_1 + \ldots + z_r$, дается равенством

$$\rho(u, v) = \frac{p}{\sqrt{(p+q)(p+r)}}.$$

3.18. Пусть $x_1, \ldots, x_n$ — случайные величины, причем x_1 и $x_2 \mid x_1, \ldots, x_n \mid x_{n-1}$ независимы. Показать, что если

$$\mathscr{E}(x_1) = \mu, \quad \mathscr{E}(x_\xi \mid x_{\xi-1}) = x_{\xi-1}, \quad \xi = 2, \ldots, n,$$

и

$$\mathscr{E}(x_1 - \mu)^2 = \mathscr{E}[(x_\xi - x_{\xi-1})^2 \mid x_{\xi-1}] = \sigma^2, \quad \xi = 2, \ldots, n,$$

то безусловное среднее и безусловная дисперсия величины x_n соответственно равны μ и $n\sigma^2$.

3.19. Показать, что если случайная величина обладает первыми $2k$ моментами $\mu_1, \ldots, \mu_{2k}$, то матрица $\|\mu_{i+j}\|$, $i, j = 1, \ldots, k$, положительно определена.

3.20. Показать, что если x — неотрицательная случайная величина, то $\mathscr{E}\left(\frac{1}{x}\right) \geqslant \frac{1}{\mathscr{E}(x)}$.

3.21. Пусть $(x_1, \ldots, x_m)$ — m-мерная случайная величина, представляющая зачеты ответов студента (взятого «наудачу» на курсе G) на m вопросов экзамена по предмету A, и $(y_1, \ldots, y_n)$ — n-мерная случайная величина, представляющая зачеты его ответов на экзамене по предмету B. Пусть, далее, $T_1 = x_1 + \ldots + x_m$ и $T_2 = y_1 + \ldots + y_n$ — общие оценки на экзаменах по A и B. Обозначим через $\sigma^2_{1,m}$ и $\sigma^2_{1,m}\rho_{1,m}$ соответственно средние арифметические дисперсий компонент x_i и ковариаций между этими компонентами, а через $\sigma^2_{2,n}$ и $\sigma^2_{2,n}\rho_{2,n}$ — такие же характеристики величины $(y_1, \ldots, y_n)$. Кроме того, обозначим через $\sigma_{1,m}\sigma_{2,n}\rho_{3,m,n}$ среднее арифметическое всех ковариаций между x_i и y_j. Показать, что если при $m, n \to \infty$ $\sigma^2_{1,m} \to \sigma^2_1$, $\sigma^2_{2,n} \to \sigma^2_2$, $\rho_{1,m} \to \rho_1$, $\rho_{2,n} \to \rho_2$ и $\rho_{3,m,n} \to \rho_3$, где $\sigma^2_1, \sigma^2_2, \rho_1, \rho_2, \rho_3$ — положительные числа, то

$$\lim_{m, n \to \infty} \rho(T_1, T_2) = \frac{\rho_3}{\sqrt{\rho_1\rho_2}}.$$

(Следовательно, если T и T^* — оценки, полученные студентом на двух *весьма продолжительных* экзаменах по тому же самому предмету (A или B), то $\rho(T, T^*) \cong 1$.)

3.22. Рассмотрим трехмерную случайную величину (x_1, x_2, x_3), имеющую (конечную) ковариационную матрицу $\|\sigma_{ij}\|$ и корреляционную матрицу $\|\rho_{ij}\|$, $i, j = 1, 2, 3$. Пусть $\beta^*_{02} + \beta^*_{12}x_1$ — прямая средней квадратической регрессии величины x_2 на x_1 и $\beta^*_{03} + \beta^*_{13}x_1$ — прямая средней квадратической регрессии величины x_3 на x_1. Положим $y_2 = x_2 - \beta^*_{02} - \beta^*_{12}x_1$ и $y_3 = x_3 - \beta^*_{03} - \beta^*_{13}x_1$. Коэффициент корреляции между y_2 и y_3, обозначаемый символом $\rho_{23\cdot 1}$, называется ***частным коэффициентом корреляции между x_2 и***

x_3 *при фиксированном значении* x_1. Показать, что

$$\rho_{23\cdot 1}=\frac{\rho_{23}-\rho_{12}\rho_{13}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{12}^2)}}.$$

3.23. Обобщая предыдущую задачу, рассмотрим $(k+1)$-мерную случайную величину $(x_1, \ldots, x_k, x_{k+1})$ с конечной ковариационной матрицей $\|\sigma_{ij}\|$ и корреляционной матрицей $\|\rho_{ij}\|$, $i, j=1, \ldots, k, k+1$. Пусть $\beta_{0k}^*+\beta_{1k}^*x_1+\ldots+\beta_{k-1,k}^*x_{k-1}$ — плоскость средней квадратической регрессии величины x_k на $x_1, \ldots, x_{k-1}$ и $\beta_{0,k+1}^*+\beta_{1,k+1}^*x_1+\ldots+\beta_{k-1,k+1}^*x_{k-1}$ — плоскость средней квадратической регрессии величины x_{k+1} на $x_1, \ldots, x_{k-1}$. Положим $y_1=x_k-\beta_{0k}^*-\beta_{1k}^*x_1-\ldots-\beta_{k-1,k}^*x_{k-1}$ и $y_2=x_{k-1}-\beta_{0,k+1}^*-\beta_{1,k+1}^*x_1-\ldots-\beta_{k-1,k+1}^*x_{k-1}$. *Частный коэффициент корреляции* $\rho_{k(k+1)\cdot 12\ldots(k-1)}$ *между* x_k *и* x_{k+1} *при фиксированных значениях* $x_1, \ldots, x_{k-1}$ есть обычный коэффициент корреляции между y_1 и y_2. Показать, что

$$\rho_{k,(k+1)\cdot 12\cdots(k-1)}=\frac{\Delta_{k,k+1}}{\sqrt{\Delta_{kk}\Delta_{k+1,k+1}}},$$

где Δ_{pq}, $p, q=k, k+1$, — минор, получаемый вычеркиванием p-й строки и q-го столбца в определителе

$$\begin{vmatrix} 1 & \rho_{12} & \cdots & \rho_{1,k+1} \\ \rho_{21} & 1 & \cdots & \rho_{2,k+1} \\ \cdot & \cdot & \cdot\cdot\cdot & \cdot \\ \rho_{k+1,1} & \rho_{k+1,2} & \cdots & 1 \end{vmatrix}.$$

3.24. Показать, что если $(x_1, \ldots, x_k)$ — k-мерная случайная величина, причем коэффициент корреляции между любыми двумя ее компонентами равен одному и тому же числу ρ, то множественный коэффициент корреляции между данной компонентой и всеми остальными компонентами есть

$$\rho\frac{\sqrt{k-1}}{\sqrt{1+(k-2)\rho}}.$$

3.25. Пусть в предыдущей задаче коэффициент корреляции между x_1 и каждой из остальных компонент $x_2, \ldots, x_k$ есть ρ_1, тогда как коэффициент корреляции между любыми двумя компонентами величины $(x_2, \ldots, x_k)$ есть ρ. Доказать неравенство $\rho_1\rho \geqslant 0$ и тот факт, что множественный коэффициент корреляции между x_1 и $x_2, \ldots, x_k$ равен

$$\frac{\sqrt{(k-1)\rho_1^2}}{\sqrt{1+(k-2)\rho}}.$$

3.26. Будем считать в задаче **3.25** коэффициент корреляции между x_1 и x_r равным ρ_r, $r=2, \ldots, k$, и величины $x_2, \ldots, x_k$ взаимно независимыми. Показать, что множественный коэффициент корреляции между x_1 и $x_2, \ldots, x_k$ есть $\sqrt{\rho_2^2+\cdots+\rho_k^2}$.

3.27. Привести к наиболее простому виду уравнения гиперплоскостей средней квадратической регрессии величины x_1 на $x_2, \ldots, x_k$ в задачах **3.25** и **3.26**.

3.28. Убедиться, что уравнения (3.8.18) и (3.8.19) эквивалентны.

ГЛАВА 4

ПОСЛЕДОВАТЕЛЬНОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН

4.1. Определение случайного процесса

Важный класс задач математической статистики образуют задачи, где требуется определить предельные распределения вероятностей некоторых функций n случайных величин, когда $n \to \infty$. Более точно, пусть $(x_1, \ldots, x_n)$ — n-мерная случайная величина и $g_n(x_1, \ldots, x_n)$ — функция от $(x_1, \ldots, x_n)$, которая сама является случайной величиной. Задача заключается в том, чтобы найти предельную к. ф. р. величины $g_n(x_1, \ldots, x_n)$, когда $n \to \infty$, или по крайней мере выяснить некоторые свойства этой к. ф. р., если последняя существует. Разумеется, это означает, что мы должны рассматривать случайные величины, имеющие бесконечно много компонент. Такие случайные величины называются *случайным процессом.* Случайный процесс со строго счетным числом компонент часто называют *последовательностью* случайных величин.

Строго говоря, *случайный процесс* есть семейство случайных величин $\{x_\alpha;\ \alpha \in A\}$, где параметрическим множеством A может являться некоторый интервал вещественной оси или некоторая последовательность точек вещественной оси, как, например, последовательность целых чисел, или даже вообще произвольное множество точек, для которого каждый конечный набор компонент $(x_{\alpha_1}, \ldots, x_{\alpha_m})$ есть множество случайных величин с заданной к. ф. р. $F_{\alpha_1, \ldots, \alpha_m}(x_{\alpha_1}, \ldots, x_{\alpha_m})$. Естественно, что к. ф. р. конечных множеств случайных величин в целом должны быть согласованы в том смысле, что к. ф. р. любого конечного множества случайных величин, скажем, $(x_{\alpha_1}, \ldots, x_{\alpha_m})$ должна совпадать с маргинальной к. ф. р. этого множества, получаемой из к. ф. р. каждого конечного множества случайных величин, которое содержит $(x_{\alpha_1}, \ldots, x_{\alpha_m})$ в качестве подмножества.

4.2. Вероятностная мера для случайного процесса

Так как обычно мы будем иметь дело со случайными процессами $(x_1, x_2, \ldots)$ со строго счетным числом компонент, можно, не опасаясь какой-либо двусмысленности, обозначить через R_∞ произведение пространств $R_1^{(1)} \times R_1^{(2)} \times \ldots$, где $R_1^{(1)}, R_1^{(2)}, \ldots$ суть (одномерные) выборочные пространства соответственно величин $x_1, x_2, \ldots$ Тогда любая счетная последовательность вещественных чисел $(b_1, b_2, \ldots)$ будет точкой пространства R_∞, и обратно, всякая точка в R_∞ есть некоторая последовательность вещественных чисел. Пусть b означает такую точку из R_∞. Таким образом, каждая реализация случайного процесса $(x_1, x_2, \ldots)$, задаваемая значениями счетного числа компонент, представляет собой выборочную точку в R_∞. Мы можем называть R_∞ выборочным пространством нашего случайного процесса.

Рассмотрим какое-либо конечное множество $(b_{\alpha_1}, \ldots, b_{\alpha_m})$ координат точки b. Это множество изображается точкой b' в m-мерном евклидовом пространстве $R_m^{(\alpha_1, \ldots, \alpha_m)} = R_1^{(\alpha_1)} \times \ldots \times R_1^{(\alpha_m)}$. Точка b' есть проекция точки $b \in R_\infty$ в $R_m^{(\alpha_1, \ldots, \alpha_m)}$. Этот факт мы можем записать следующим образом:

$$b' = h_{\alpha_1, \ldots, \alpha_m}(b). \tag{4.2.1}$$

Множество E всех точек в R_∞, проектирующихся в данное множество E' пространства $R_m^{(\alpha_1, \ldots, \alpha_m)}$, называется *цилиндрическим множеством* и может быть записано в виде

$$E = h^{-1}_{\alpha_1, \ldots, \alpha_m}(E'). \tag{4.2.2}$$

Если E' — борелевское множество в $R_m^{(\alpha_1, \ldots, \alpha_m)}$, то цилиндрическое множество E в R_∞, соответствующее E' (прообраз множества E' в R_∞), называется *борелевским цилиндрическим множеством.* Читатель должен заметить, что E' есть просто проекция в $R_m^{(\alpha_1, \ldots, \alpha_m)}$ множества E пространства R_∞. Если мы распространим определение борелевского цилиндрического множества на случай, когда $m = \infty$, то, очевидно, R_∞ само будет борелевским цилиндрическим множеством. Кроме того, если для конечного m множество E — борелевское цилиндрическое, то таково же и множество $\bar{E}$.

Допустим теперь, что E' и E'' — борелевские множества в $R_m^{(\alpha_1, \ldots, \alpha_m)}$ и $R_n^{(\beta_1, \ldots, \beta_n)}$, а E_1 и E_2 — соответствующие им цилиндрические множества в R_∞. Рассмотрим множество $E_1 \cup E_2 \subset R_\infty$. Это — цилиндрическое множество в R_∞, соответствующее (по определению) множеству E_*, точек в $R_s^{(\gamma_1, \ldots, \gamma_s)}$, где $\gamma_1, \ldots, \gamma_s$ — все отличные друг от друга целые числа ряда $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_n$. Тогда $R_m^{(\alpha_1, \ldots, \alpha_m)}$ и $R_n^{(\beta_1, \ldots, \beta_n)}$ суть маргинальные выборочные пространства для $R_s^{(\gamma_1, \ldots, \gamma_s)}$ и $E_* = E'_* \cup E''_*$, где E'_* — цилиндрическое множество в $R_s^{(\gamma_1, \ldots, \gamma_s)}$, соответствующее $E' \subset R_m^{(\alpha_1, \ldots, \alpha_m)}$, и E''_* — цилиндрическое множество в $R_s^{(\gamma_1, \ldots, \gamma_s)}$, соответствующее $E'' \subset R_n^{(\beta_1, \ldots, \beta_n)}$. Множество $E_1 \cap E_2$ есть цилиндрическое множество в R_∞, соответствующее $E'_* \cap E''_* \subset R_s^{(\gamma_1, \ldots, \gamma_s)}$. Следует отметить, что E_1 и E_2 не пересекаются тогда и только тогда, когда $E'_* \cap E''_* = \varnothing$.

Если E' и E'' являются борелевскими множествами соответственно в $R_m^{(\alpha_1, \ldots, \alpha_m)}$ и $R_n^{(\beta_1, \ldots, \beta_n)}$ и их прообразами в $R_s^{(\gamma_1, \ldots, \gamma_s)}$ служат множества E'_* и E''_*, то $E'_* \bigcup E''_*$ также будет борелевским множеством в $R_s^{(\gamma_1, \ldots, \gamma_s)}$. Следовательно, $E_1 \bigcup E_2$ как цилиндрическое множество в R_∞, соответствующее $E'_* \bigcup E''_*$, есть борелевское цилиндрическое множество. Таким образом,

4.2.1. *Класс борелевских цилиндрических множеств в R_∞ образует булево поле множеств $\mathcal{F}$.*

Однако мы в первую очередь заинтересованы в борелевском расширении этого поля $\mathcal{F}$, которое мы назовем $\mathcal{B}_\infty$. Такое борелевское поле $\mathcal{B}_\infty$ определяется как наименьший борелевский класс множеств в R_∞, который содержит булево поле $\mathcal{F}$. $\mathcal{B}_\infty$ называется *классом борелевских множеств* пространства R_∞ и включает все борелевские цилиндрические множества в R_∞.

Далее, пусть E' — борелевское множество в пространстве $R^{(\alpha_1, \ldots, \alpha_m)}$ и E — прообраз этого множества в R_∞. Для того чтобы назначить вероятность борелевскому цилиндрическому множеству E, мы воспользуемся обычным правилом, именно положим

$$P(E) = P_{\alpha_1, \ldots, \alpha_m}(E'), \tag{4.2.3}$$

где $P_{\alpha_1, \ldots, \alpha_m}(E')$ — вероятность, сопоставляемая множеству E' в результате подсчета по к. ф. р. величины $(x_{\alpha_1}, \ldots, x_{\alpha_m})$, которая образуется выделенными компонентами случайного процесса $(x_1, x_2, \ldots)$. Возникает вопрос, как определяются вероятности множеств из $\mathcal{B}_\infty$, которые не являются борелевскими цилиндрическими множествами. Ответ на этот вопрос дается следующей теоремой расширения, принадлежащей Колмогорову (1933а).

4.2.2. *Пусть $(x_{\alpha_1}, \ldots, x_{\alpha_m})$ — какой-либо конечный набор компонент случайного процесса $(x_1, x_2, \ldots)$ и $R_m^{(\alpha_2, \ldots, \alpha_m)}$ — выборочное пространство величины $(x_{\alpha_1}, \ldots, x_{\alpha_m})$.* Пусть *$E$ — борелевское цилиндрическое множество в R_∞, соответствующее некоторому борелевскому множеству $E' \subset R_m^{(\alpha_1, \ldots, \alpha_m)}$. Назначим множеству E вероятность $P(E)$ согласно правилу* (4.2.3). *Тогда на $\mathcal{B}_\infty$, борелевских множествах пространства R_∞, существует единственная вероятностная мера, сужение которой на классе цилиндрических множеств в R_∞ есть функция множества, определенная* (4.2.3).

Доказательство этой теоремы мы опускаем. Читатель, заинтересованный в доказательстве, может найти его в книгах Дуба (1953) и Колмогорова (1933а).

В практических задачах, предлагая случайный процесс со счетным числом компонент в качестве математической модели, мы часто (как и в случае конечномерной величины) действительно имеем дело с физическим процессом определенного рода, который «порождает» последовательность компонент $x_1, x_2, \ldots$ таким образом, что условия теоремы **4.2.2** можно считать выполненными.

Примеры. Пусть многократно бросается «правильная» игральная кость. Тогда мы можем определить очень простой случайный процесс $(x_1, x_2, \ldots)$, где x_α есть случайная величина, обозначающая число очков, выпавших при α-м бросании, и к тому же при любом выборе r целых чисел $\alpha_1, \ldots, \alpha_r$ r-мерная случайная величина $(x_{\alpha_1}, \ldots, x_{\alpha_r})$ имеет ф. в. р. $p(x_{\alpha_1}, \ldots, x_{\alpha_r})$, равную $\frac{1}{6^r}$ в каждой из 6^r своих точек сосредоточения массы.

Если мы допустим, что x_1 — случайная величина, имеющая равномерное распределение на интервале $(0, 1)$, x_2 — случайная величина, имеющая равномерное распределение на интервале $(x_1, 1)$, и вообще x_α — случайная величина, имеющая равномерное распределение на интервале $(x_{\alpha-1}, 1)$, $\alpha = 2, 3, \ldots$, мы получим пример важного вида случайного процесса $(x_1, x_2, \ldots)$, называемого *цепью Маркова порядка 1.* Заметим, что при любом α ф. п. в. условной случайной величины $x_\alpha \mid x_{\alpha-1}$, именно $\frac{1}{1 - x_{\alpha-1}}$, в точности совпадает с ф. п. в. условной случайной величины $x_\alpha \mid x_{\alpha-1}, x_{\beta_1}, \ldots, x_{\beta_r}$, где $\beta_1, \ldots, \beta_r$ — произвольный набор $r < \alpha - 1$ целых чисел из ряда $1, 2, \ldots, \alpha - 2$. *Цепь Маркова порядка k* есть случайный процесс, обладающий тем свойством, что при любом α условная случайная величина $x_\alpha \mid x_{\alpha-1}, \ldots, x_{\alpha-k}$ имеет то же самое распределение, что и условная случайная величина $x_\alpha \mid x_{\alpha-1}, \ldots, x_{\alpha-k}, x_{\beta_1}, \ldots, x_{\beta_r}$, где $\beta_1, \ldots, \beta_r$ — произвольный набор $r < \alpha - k$ целых чисел из ряда $1, 2, \ldots, \alpha - k - 1$. За сведениями, касающимися общей теории цепей Маркова или процессов Маркова, читатель отсылается к книгам Дуба (1953), Феллера (1957) и Лоэва (1955).

Когда мы имеем случайный процесс $(x_1, x_2, \ldots)$ и вынуждены рассматривать вероятности, сопоставленные борелевским множествам в (евклидовых) пространствах, отвечающих конечным наборам x, мы всегда можем перейти к прообразам этих множеств в R_∞ и, следовательно, всегда можем говорить о множествах в R_∞ и о сопоставленных им вероятностях. Например, вместо того чтобы говорить о (борелевском) множестве на $x_m x_n$-плоскости, для которого $|x_m - x_n| < c$, мы можем говорить о (борелевском цилиндрическом) множестве в R_∞, для которого $|x_m - x_n| < c$.

Вероятности, назначаемые двум этим множествам, равны по определению. В некоторых ситуациях возможность рассмотрения прообразов событий в R_∞, минуя события в особых конечномерных пространствах, отвечающих конечным наборам компонент процесса $(x_1, x_2, \ldots)$, представляет известное удобство. R_∞ есть фиксированное выборочное пространство, в котором имеют место события относительно $(x_1, x_2, \ldots)$, в то время как выборочное пространство величины $(x_1, \ldots, x_n)$ меняется вместе с n. Однако, помимо удобства, существуют и другие причины для рассмотрения R_∞. В самом деле, как будет показано, только в пространстве R_∞ и можно сформулировать некоторые задачи и теоремы.

4.3. Сходимость по вероятности

(а) Некоторые критерии сходимости по вероятности. Допустим, что $(x, x_1, x_2, \ldots)$ — случайный процесс, причем для любого $\varepsilon > 0$

$$\lim_{n \to \infty} P(|x_n - x| \geqslant \varepsilon) = 0. \tag{4.3.1}$$

Тогда говорят, что $(x_1, x_2, \ldots)$ *сходится стохастически или сходится по вероятности* к случайной величине x. Факт сходимости этого типа мы иногда кратко записываем следующим образом:

$$p \lim_{n \to \infty} x_n = x. \tag{4.3.1a}$$

Если x — вырожденная случайная величина, так что $P(x = x_0) = 1$, то случайный процесс $(x_1, x_2, \ldots)$ сходится по вероятности к постоянной x_0. Одним из простейших и наиболее важных примеров сходимости случайного процесса к постоянной величине служит *слабый закон больших чисел*, выражаемый теоремой

4.3.1. *Пусть $(x_1, x_2, \ldots)$ — последовательность независимых случайных величин, имеющих нулевые средние и дисперсии $\sigma_1^2, \sigma_2^2, \ldots$ Положим $\bar{x}_n = \frac{1}{n}(x_1 + \ldots + x_n)$ и $c_n^2 = \sigma_1^2 + \ldots + \sigma_n^2$. Тогда, если $\lim\limits_{n \to \infty} \frac{c_n^2}{n^2} = 0$, то случайный процесс $(\bar{x}_1, \bar{x}_2, \ldots)$ сходится по вероятности к нулю.*

Чтобы доказать эту теорему, заметим, что $\mathscr{E}(\bar{x}_n) = 0$ и $\mathscr{E}(\bar{x}_n^2) = \frac{c_n^2}{n^2}$. Следовательно, по неравенству Чебышева (3.3.5) при любом $\varepsilon > 0$

$$P(|\bar{x}_n| < \varepsilon) \geqslant 1 - \frac{c_n^2}{n^2 \varepsilon^2}.$$

Таким образом, если $\frac{c_n^2}{n^2} \to 0$, когда $n \to \infty$, мы имеем $\lim\limits_{n \to \infty} P(|\bar{x}_n| < \varepsilon) = 1$, что и завершает рассуждения.

Теорема **4.3.1** фактически остается справедливой и в том случае, когда равны нулю все ковариации между компонентами процесса $(x_1, x_2, \ldots)$.

Если $F_1(x), F_2(x), \ldots$ суть к. ф. р. компонент процесса $(x_1, x_2, \ldots)$ и $\lim\limits_{n \to \infty} F_n(x) = F(x)$ в каждой точке непрерывности $F(x)$, мы говорим, что $(x_1, x_2, \ldots)$ *сходится к $F(x)$ по распределению.*

Ниже следует несколько теорем относительно сходимости по вероятности и сходимости по распределению. Они используются в более поздних главах.

Один из простейших критериев сходимости по вероятности можно сформулировать в виде следующего утверждения:

4.3.2. *Пусть* $(x,\ x_1,\ x_2,\ \ldots)$ — *случайный процесс и*

$$\lim_{n\to\infty} \mathscr{E}(x_n - x)^2 = 0. \tag{4.3.2}$$

Тогда процесс $(x_1,\ x_2,\ \ldots)$ *сходится по вероятности к случайной величине* x.

Из неравенства Чебышева, установленного в § 3.3, следует, что

$$P(|x_n - x| \geqslant \varepsilon) \leqslant \frac{\mathscr{E}(x_n - x)^2}{\varepsilon^2}, \tag{4.3.3}$$

но, поскольку (4.3.2) предполагается выполненным, мы имеем $\lim\limits_{n\to\infty} P(|x_n - x| \geqslant \varepsilon) \leqslant 0$, что в свою очередь влечет (4.3.1).

Говорят, что $(x_1,\ x_2,\ \ldots)$ *сходится в среднем* к величине x, если $(x,\ x_1,\ x_2, \ldots)$ есть случайный процесс, который удовлетворяет (4.3.2), причем $\mathscr{E}(x_n^2) < \infty$, $n = 1,\ 2,\ \ldots$ и $\mathscr{E}(x^2) < \infty$. Для краткого обозначения этого вида сходимости иногда применяется запись

$$\underset{n\to\infty}{\mathrm{l.i.m.}}\ x_n = x. \tag{4.3.2a}$$

4.3.3. *Если* $(x,\ x',\ x_1,\ x_2, \ldots)$ — *случайный процесс и* $(x_1,\ x_2,\ \ldots)$ *сходится по вероятности как к случайной величине* x, *так и к величине* x', *то эти последние являются эквивалентными.*

Чтобы это установить, мы прежде всего заметим, что

$$P\left(|x - x'| > \frac{1}{N}\right) = P\left(|(x_n - x') - (x_n - x)| > \frac{1}{N}\right). \tag{4.3.4}$$

Пусть теперь $E,\ E',\ E''$ — множества точек в выборочном пространстве R_∞ процесса $(x,\ x',\ x_1,\ x_2,\ \ldots)$, для которых соответственно $|(x_n - x') - (x_n - x)| > \frac{1}{N}$, $|x - x_n| > \frac{1}{2N}$ и $|x_n - x'| > \frac{1}{2N}$. Тогда, очевидно, $E \subset (E' \cup E'')$. Следовательно, $P(E) \leqslant P(E') + P(E'')$, т. е.

$$P\left(|x - x'| > \frac{1}{N}\right) \leqslant P\left(|x_n - x| > \frac{1}{2N}\right) + P\left(|x_n - x'| > \frac{1}{2N}\right). \tag{4.3.5}$$

По предположению при $n \to \infty$ оба члена в правой части этого неравенства имеют предел, равный нулю. Таким образом,

$$P\left(|x - x'| > \frac{1}{N}\right) = 0.$$

Обозначив через G_N множество в R_∞, для которого $|x - x'| > \frac{1}{N}$, мы находим, что $G_1 \subset G_2 \subset \ldots$, причем $\lim\limits_{N\to\infty} G_N = G$, где G есть множество всех точек в R_∞, для которых $x \neq x'$. Поэтому по теореме **1.4.6** для случая множеств из $\mathscr{B}_\infty$

$$P(x \neq x') = P(G_1 \cup G_2 \cup \ldots) \leqslant P(G_1) + P(G_2) + \ldots \tag{4.3.6}$$

Но $P(G_1)=P(G_2)=\ldots=0$. Отсюда $P(x\neq x')=0$ и, следовательно, случайные величины x и x' эквивалентны [см. § 2.8 (b)].

Другой полезный результат состоит в том, что сходимость по вероятности влечет сходимость по распределению, или в более точной формулировке:

4.3.4. *Пусть $(x, x_1, x_2, \ldots)$ — случайный процесс с компонентами, имеющими к. ф. р. соответственно $F(x)$, $F_1(x)$, $F_2(x)$, ... Если $(x_1, x_2, \ldots)$ сходится по вероятности к случайной величине x, то последовательность к. ф. р. $F_1(x)$, $F_2(x)$, ... сходится к $F(x)$ в каждой точке непрерывности этой функции.*

Как обычно, пусть R_∞ означает выборочное пространство случайного процесса $(x, x_1, x_2, \ldots)$. Предположим, что $F(x)$ непрерывна в точке $x=x_0$. Выберем константу $x'<x_0$ и рассмотрим следующие три события в R_∞: $x\leqslant x'$, $x_n\leqslant x_0$ и $|x_n-x|>(x_n-x')$. Первое событие содержится в объединении второго и третьего. Отсюда

$$F(x')=P(x\leqslant x')\leqslant P(x_n\leqslant x_0)+P(|x_n-x|>(x_0-x')). \tag{4.3.7}$$

Теперь перейдем к пределу при $n\to\infty$. Поскольку $(x_1, x_2, \ldots)$ сходится по вероятности к случайной величине x, $\lim\limits_{n\to\infty} P(|x_n-x|>(x_0-x'))=0$, так что

$$F(x')\leqslant \liminf_{n\to\infty} F_n(x_0). \tag{4.3.8}$$

Подобным же образом, выбирая константу $x''>x_0$, мы находим, что

$$F(x'')\geqslant \limsup_{n\to\infty} F_n(x_0) \tag{4.3.9}$$

и, следовательно,

$$F(x')\leqslant \liminf_{n\to\infty} F_n(x_0)\leqslant \limsup_{n\to\infty} F_n(x_0)\leqslant F(x''). \tag{4.3.10}$$

Но так как $F(x)$ непрерывна в точке $x=x_0$, мы имеем $\lim\limits_{x'\to x_0} F(x')=\lim\limits_{x''\to x_0} F(x'')=F(x_0)$. Таким образом,

$$\lim_{n\to\infty} F_n(x_0)=F(x_0), \tag{4.3.11}$$

что и завершает доказательство теоремы **4.3.4.**

(b) Сходимость функций от компонент случайных процессов. Рассмотрим случайный процесс $(x_1, x_2, \ldots)$, сходящийся по вероятности к случайной величине x. В этой ситуации нам иногда требуется знать, какие условия будут обеспечивать сходимость по вероятности нового процесса $(g(x_1), g(x_2), \ldots)$ к величине $g(x)$. Мы приведем несколько теорем, относящихся к этому и подобным этому вопросам, которые будут использованы в дальнейших главах.

4.3.5. *Пусть $(x, x_1, x_2, \ldots)$ — случайный процесс, для которого $(x_1, x_2, \ldots)$ сходится по вероятности к случайной величине x, и пусть $g(x)$ — непрерывная в R_1 функция от x. Тогда*

$$p\lim_{n\to\infty} g(x_n)=g(x). \tag{4.3.12}$$

Если x — вырожденная случайная величина, именно постоянная c, то (4.3.12) просто означает, что $(g(x_1), g(x_2), \ldots)$ сходится по вероятности к постоянной $g(c)$.

Чтобы доказать теорему **4.3.5**, мы заметим, что $g(x)$ равномерно непрерывна в любом замкнутом интервале, например в интервале $[-M, M]$. Так как x является случайной величиной, мы можем для произвольного $\varepsilon > 0$ выбрать M, удовлетворяющее условию

$$P(|x| > M) < \frac{\varepsilon}{2}. \tag{4.3.13}$$

Для выбранных ε и M найдется такое $\delta(\varepsilon, M)$, что, как только

$$\text{(I)} \quad |x| \leqslant M,$$

$$\text{(II)} \quad |x_n - x| < \delta(\varepsilon, M),$$

так сейчас же

$$\text{(III)} \; |g(x_n) - g(x)| < \varepsilon.$$

Если мы обозначим через E_1, E_2 и E_3 множества в R_∞, для которых имеют место соответственно I, II и III, то сможем написать $\overline{E}_3 \subset (\overline{E}_1 \cup \overline{E}_2)$. Отсюда $P(\overline{E}_3) \leqslant P(\overline{E}_1) + P(\overline{E}_2)$, т. е.

$$P(|g(x_n) - g(x)| \geqslant \varepsilon) \leqslant P(|x_n - x| \geqslant \delta(\varepsilon, M)) + P(|x| > M). \tag{4.3.14}$$

Но по условию теоремы существует столь большое $n(\varepsilon, \delta, M)$, что при $n > n(\varepsilon, \delta, M)$

$$P(|x_n - x| \geqslant \delta(\varepsilon, M)) < \frac{\varepsilon}{2}. \tag{4.3.15}$$

Поскольку M было выбрано так, чтобы удовлетворить (4.3.13), мы, в конце концов, получаем для $n > n(\varepsilon, \delta, M)$

$$P(|g(x_n) - g(x)| \geqslant \varepsilon) < \varepsilon, \tag{4.3.16}$$

что эквивалентно (4.3.12) и доказывает, таким образом, теорему.

Следует отметить, что теорема **4.3.5** может быть сформулирована в более общей форме, если потребовать, чтобы функция $g(x)$ была непрерывна в замкнутом интервале I, для которого $P(x \in I) = 1$. Так как это повлекло бы за собой только незначительные изменения вышеприведенных рассуждений, соответствующее рассмотрение предоставляется читателю.

Иногда нам приходится заниматься проблемами сходимости, касающимися случайных процессов, компоненты которых суть векторы. Например, если $(x, y;\ x_1, y_1;\ x_2, y_2;\ \ldots)$ — такой случайный процесс, нас могут интересовать условия, при которых $(g(x_1, y_1), g(x_2, y_2), \ldots)$ сходится по вероятности к $g(x, y)$. В этом случае R_∞ есть выборочное пространство процесса $(x, y;\ x_1, y_1;\ x_2, y_2;\ \ldots)$.

Если $(z;\ x_1,\ y_1;\ x_2,\ y_2;\ \ldots)$ — случайный процесс, причем для любого $\varepsilon > 0$
$$\lim_{n\to\infty} P(|x_n - y_n| > \varepsilon) = 0,$$
в то время как одна из последовательностей $(x_1,\ x_2,\ \ldots)$ или $(y_1,\ y_2,\ \ldots)$ сходится по вероятности к случайной величине z, легко можно убедиться, что и другая последовательность также сходится по вероятности к случайной величине z. При таком положении вещей удобно говорить, что $(x_1,\ x_2,\ \ldots)$ и $(y_1,\ y_2,\ \ldots)$ *одновременно сходятся по вероятности к случайной величине* z. Подобным же образом, если имеет место вышеуказанное предельное соотношение и одна из двух последовательностей сходится по распределению к $F(x)$, мы будем говорить, что обе последовательности *одновременно сходятся по распределению к* $F(x)$.

Мы можем следующим образом сформулировать обобщение теоремы **4.3.5** на случай нескольких последовательностей случайных величин:

4.3.6. *Предположим, что* $(x^{(1)},\ \ldots,\ x^{(k)};\ x_1^{(1)},\ \ldots,\ x_1^{(k)};\ x_2^{(1)},\ \ldots\ \ldots,\ x_2^{(k)},\ \ldots)$ — *k-мерный векторный случайный процесс, который сходится по вероятности к k-мерной случайной величине* $(x^{(1)},\ \ldots\ \ldots,\ x^{(k)})$. *Пусть* $g(x^{(1)},\ \ldots,\ x^{(k)})$ — *непрерывная в R_k функция от* $(x^{(1)},\ \ldots,\ x^{(k)})$. *Тогда*
$$p\lim_{n\to\infty} g(x_n^{(1)},\ \ldots,\ x_n^{(k)}) = g(x^{(1)},\ \ldots,\ x^{(k)}). \tag{4.3.17}$$

Доказательство этой теоремы аналогично доказательству теоремы **4.3.5** и предоставляется читателю.

Снова следует отметить, что мы можем несколько ослабить условия теоремы **4.3.6**, требуя, чтобы функция $g(x^{(1)},\ \ldots,\ x^{(k)})$ была непрерывна в замкнутом k-мерном интервале I_k, для которого $P((x_1^{(1)},\ \ldots,\ x^{(k)}) \in I_k) = 1$.

Ниже приводится следствие теоремы **4.3.6**, которое будет особенно полезно в дальнейшем.

4.3.6a. *Если векторный случайный процесс* $(x,\ c;\ x_1,\ y_1;\ x_2,\ y_2;\ \ldots)$ *таков, что* $(x_1,\ y_1;\ x_2,\ y_2;\ \ldots)$ *сходится по вероятности к* $(x,\ c)$, *причем x имеет к.ф.р. $F(x)$ и c — некоторая положительная константа, то в каждой точке непрерывности функции* $F(x)$
$$\lim_{n\to\infty} P(x_n + y_n \leqslant z) = F(z - c), \tag{4.3.18}$$
$$\lim_{n\to\infty} P(x_n y_n \leqslant z) = F\left(\frac{z}{c}\right), \tag{4.3.19}$$
$$\lim_{n\to\infty} P\left(\frac{x_n}{y_n} \leqslant z\right) = F(cz), \tag{4.3.20}$$
$$\lim_{n\to\infty} P(ax_n + by_n \leqslant z) = F\left(\frac{z - bc}{a}\right), \tag{4.3.21}$$
где a и b — константы и $a > 0$.

Позже представит интерес следующая теорема:

4.3.7. *Допустим, что* $(x_1^{(1)}, \ldots, x_1^{(k)}, y_1^{(1)}, \ldots, y_1^{(k)}; x_2^{(1)}, \ldots, x_2^{(k)}, y_2^{(1)}, \ldots, y_2^{(k)}; \ldots)$ — *векторный случайный процесс, причем*

$$|y_n^{(i)} - c^{(i)}| \leqslant |x_n^{(i)} - c^{(i)}|, \quad i = 1, \ldots, k; \; n = 1, 2, \ldots, \tag{4.3.22}$$

и $(x_1^{(1)}, \ldots, x_1^{(k)}; x_2^{(1)}, \ldots, x_2^{(k)}; \ldots)$ *сходится по вероятности к векторной постоянной* $(c^{(1)}, \ldots, c^{(k)})$. *Пусть, кроме того,* $g(x^{(1)}, \ldots, x^{(k)})$ — *однозначная функция, определенная во всех точках пространства* R_k *и непрерывная в некотором открытом* k-*мерном параллелепипеде, содержащем точку* $(c^{(1)}, \ldots, c^{(k)})$. *Тогда*

$$p \lim_{n\to\infty} g(y_n^{(1)}, \ldots, y_n^{(k)}) = g(c^{(1)}, \ldots, c^{(k)}). \tag{4.3.23}$$

Доказательство этой теоремы не содержит никаких особых трудностей и потому опускается.

Следует указать на то обстоятельство, что условия теоремы **4.3.7** можно ослабить, введя предположение, согласно которому функция $g(x^{(1)}, \ldots, x^{(k)})$ должна быть определена и непрерывна только в некотором замкнутом k-мерном интервале I_k, содержащем точку $(c^{(1)}, \ldots, c^{(k)})$, для которого $P((x^{(1)}, \ldots, x^{(k)}) \in I_k) = 1$.

Пусть при каждом θ из некоторого интервала (θ', θ'') последовательность $(g(x, \theta), g(x_1, \theta), g(x_2, \theta), \ldots)$ есть случайный процесс. Говорят, что $(g(x_1, \theta), g(x_2, \theta), \ldots)$ *сходится по вероятности к* $g(x, \theta)$ *равномерно относительно* $\theta \in (\theta', \theta'')$, если для любого $\varepsilon > 0$ существует такое n_ε, что при $n > n_0$ для каждого θ в интервале (θ', θ'')

$$P(|g(x_n, \theta) - g(x, \theta)| > \varepsilon) < \varepsilon. \tag{4.3.24}$$

Это понятие очевидным образом можно обобщить на случай, когда $g(x, \theta)$ является вырожденной, а также на случай, когда x_n — векторная величина. Наконец, докажем еще одну теорему, которая будет использоваться в главах, посвященных теории статистического оценивания.

4.3.8. *Пусть случайный процесс* $(x_1, x_2, \ldots)$ *таков, что процесс* $(f_1(x_1, \theta), f_2(x_1, x_2, \theta), \ldots)$ *также случайный и, кроме того, сходится по вероятности равномерно относительно* $\theta \in (\theta', \theta'')$ *к конечному числу* $g(\theta)$, *причем функция* $g(\theta)$ *непрерывна в точке* $\theta = \theta_0$, $\theta_0 \in (\theta', \theta'')$. *Пусть, далее,* $(\theta_1^*(x_1), \theta_2^*(x_1, x_2), \ldots)$ — *случайный процесс, который сходится по вероятности к* θ_0. *Тогда процесс* $(f_1(x_1, \theta_1^*(x_1)), f_2(x_1, x_2, \theta_2^*(x_1, x_2)), \ldots)$ *сходится по вероятности к* $g(\theta_0)$.

Так как процесс $(\theta_1^*(x_1), \theta_2^*(x_1, x_2), \ldots)$ сходится по вероятности к θ_0, то для каждого интервала $(\theta_0 - \varepsilon, \theta_0 + \varepsilon) \subset (\theta', \theta'')$ найдется такое n_ε, что при $n > n_\varepsilon$

$$P(\theta_0 - \varepsilon < \theta_n^*(x_1, \ldots, x_n) < \theta_0 + \varepsilon) > 1 - \varepsilon. \tag{4.3.25}$$

Пусть теперь $g_u(\varepsilon)$ означает точную верхнюю грань, $g_l(\varepsilon)$ — точную нижнюю грань значений $g(\theta)$ в интервале $(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$. Так как в свою очередь $(f_1(x, \theta), f_2(x_1, x_2, \theta), \ldots)$ сходится по вероятности к $g(\theta)$ равномерно относительно $\theta \in (\theta', \theta'')$, то для произвольного ε' найдется такое $n_{\varepsilon'}$, что при $n > n_{\varepsilon'}$ и при любом θ в интервале $(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$

$$P(g_l(\varepsilon) - \varepsilon' < f_n(x_1, \ldots, x_n, \theta) < g_u(\varepsilon) + \varepsilon') > 1 - \varepsilon'. \tag{4.3.26}$$

Далее, для любого $n > \max(n_\varepsilon, n_{\varepsilon'})$ обозначим через E_n то множество в выборочном пространстве R_∞ процесса $(x_1, x_2, \ldots)$, для которого имеют место оба неравенства, фигурирующие в (4.3.25) и (4.3.26) под знаком вероятности. Очевидно, $P(E_n) \geqslant 1 - \varepsilon - \varepsilon'$. Но всякая точка из E_n будет удовлетворять неравенству *)

$$g_l(\varepsilon) - \varepsilon < f_n(x_1, \ldots, x_n, \theta_n^*) < g_u(\varepsilon) + \varepsilon'^*), \tag{4.3.27}$$

где $\theta_n^*(x_1, \ldots, x_n)$ кратко записана как θ_n^*. Следовательно, при $n > \max(n_\varepsilon, n_{\varepsilon'})$ вероятность этого неравенства по меньшей мере равна $1 - \varepsilon - \varepsilon'$. Поскольку ε и ε' произвольны и $\lim\limits_{\varepsilon \to 0}(g_l(\varepsilon) - g_u(\varepsilon)) = 0$ вследствие непрерывности $g(\theta)$ в точке θ_0, мы приходим к заключению, что $(f_n(x_1, \ldots, x_n, \theta_n^*), n = 1, 2, \ldots)$ сходится по вероятности к $g(\theta_0)$. Таким образом, теорема **4.3.8** доказана.

Принимая во внимание теорему **4.3.4**, можно получить варианты теорем **4.3.5**, **4.3.6**, **4.3.7** и **4.3.8**, в которых вместо сходимости по вероятности утверждается только сходимость по распределению.

4.4. Сходимость почти наверное

(а) Определение. Пусть $(x_1, x_2, \ldots)$ — случайный процесс и $(R_\infty, \mathscr{B}_\infty, P)$ — соответствующее ему вероятностное пространство. Обозначим через E множество всех сходящихся последовательностей $(b_1, b_2, \ldots)$. Чтобы лучше представить себе это множество, рассмотрим в R_∞ множества $E_{n,p,N}$ точек, удовлетворяющих условию

$$|b_{n+j} - b_n| < \frac{1}{N}, \qquad j = 1, 2, \ldots, p, \tag{4.4.1}$$

где N — целое положительное число.

При любом выборе чисел n, p и N множество $E_{n,p,N}$ является борелевским цилиндрическим в R_∞ и, конечно, принадлежит $\mathscr{B}_\infty$. Множества $E_{(N)} = \bigcup\limits_{n=1}^{\infty} \bigcap\limits_{p=1}^{\infty} E_{n,p,N}$, $N = 1, 2, \ldots$, также принадлежат $\mathscr{B}_\infty$ и образуют убывающую последовательность. Множество E, введенное выше, есть предел этой убывающей последовательности, т. е.

$$E = \lim_{N \to \infty} E_{(N)}. \tag{4.4.2}$$

---

*) Это утверждение не вполне точно. Читателю следует с особым вниманием отнестись ко всем последующим ссылкам на него. (*Прим. перев.*)

Следовательно, множество E принадлежит $\mathscr{B}_\infty$ и, таким образом, обладает некоторой вероятностью $P(E)$. Эта вероятность может быть названа *вероятностью сходимости* случайного процесса $(x_1, x_2, \ldots)$. Само E называется *множеством сходимости* в пространстве R_∞.

Предположим теперь, что $P(E)=1$. Пусть x — случайная величина, принимающая значение $\lim\limits_{n\to\infty} b_n$, если $(b_1, b_2, \ldots)$ принадлежит $\mathscr{B}_\infty$, и значение 0, если $(b_1, b_2, \ldots)$ классу $\mathscr{B}_\infty$ не принадлежит. В этом случае мы говорим, что последовательность случайных величин сходится *почти наверное* к случайной величине x.

Замечание. Понять смысл этого определения читателю поможет интерпретация сказанного на языке событий из исходного основного вероятностного пространства $(R, \mathscr{B}, P)$. Обозначая, как обычно, выборочную точку R через e, можно записать данный случайный процесс в виде $(x_1(e), x_2(e), \ldots)$ (причем все компоненты измеримы относительно $\mathscr{B}$). Таким образом, мы имеем счетную последовательность случайных величин, которая отображает каждую выборочную точку пространства R в некоторую точку пространства R_∞. Если существует такая случайная величина $x(e)$ (измеримая относительно $\mathscr{B}$), что $\lim\limits_{n\to\infty} x_n(e)=x(e)$ для любой точки $e\in E'$, где E' — прообраз в R множества E в R_∞, мы говорим, что $(x_1(e), x_2(e), \ldots)$ сходится к $x(e)$ с вероятностью $P(E')$. E' есть *множество сходимости* в пространстве R. Если $P(E')=P(E)=1$, мы говорим, что $(x_1(e), x_2(e), \ldots)$ сходится к $x(e)$ *почти наверное* или *сходится с вероятностью единица.*

(b) Соотношение между сходимостью почти наверное и сходимостью по вероятности. Необходимо отметить, что сходимость почти наверное относится к более сильному типу сходимости, чем сходимость по вероятности. В самом деле,

4.4.1. *Если случайный процесс $(x, x_1, x_2, \ldots)$ таков, что процесс $(x_1, x_2, \ldots)$ сходится почти наверное к случайной величине x, то последний сходится к этой величине и по вероятности.*

Пусть E — множество сходимости процесса $(x_1, x_2, \ldots)$ к величине x. Тогда E при любом $\varepsilon>0$ содержится в множестве, равном пределу расширяющейся последовательности следующих множеств в R_∞: $[|x_{n+j}-x|<\varepsilon,\ j=0, 1, 2, \ldots]$, $n=1, 2, \ldots$ Обозначим через E_n и F_n соответственно множества $[|x_{n+j}-x|<\varepsilon,\ j=0, 1, 2, \ldots]$ и $[|x_n-x|<\varepsilon]$. Поскольку $E_n\subset F_n$,

$$P(|x_n-x|<\varepsilon)\geqslant P(|x_{n+j}-x|<\varepsilon,\quad j=0, 1, 2, \ldots). \tag{4.4.3}$$

Устремляя теперь n к бесконечности, мы получаем

$$\lim_{n\to\infty} P(|x_n-x|<\varepsilon)\geqslant \lim_{n\to\infty} P(|x_{n+j}-x|<\varepsilon,\ j=0, 1, 2, \ldots)= \\ =P(\lim_{n\to\infty} E_n)\geqslant P(E)=1, \tag{4.4.4}$$

т. е.

$$\lim_{n\to\infty} P(|x_n-x|<\varepsilon)=1. \tag{4.4.5}$$

Последнее соотношение равносильно (4.3.1), что и доказывает теорему.

4.5. Неравенство Колмогорова

Следующее интересное распространение идеи, заложенной в неравенстве Чебышева, на целую совокупность неравенств предложено Колмогоровым (1928):

4.5.1. *Пусть* $x_1, \ldots, x_n$ — *множество независимых случайных величин, имеющих нулевые средние и дисперсии* $\sigma_1^2, \ldots, \sigma_n^2$. *Положим* $c_n^2 = \sigma_1^2 + \ldots + \sigma_n^2$. *Тогда вероятность одновременного осуществления неравенств*

$$|x_1 + \ldots + x_\alpha| < \lambda c_n, \qquad \alpha = 1, \ldots, n, \tag{4.5.1}$$

по меньшей мере равна $1 - \frac{1}{\lambda^2}$.

Чтобы доказать этот результат, выделим в выборочном пространстве R_n события F_α, $\alpha = 1, \ldots, n$, отвечающие соответственно неравенствам

$$|x_1 + \ldots + x_\alpha| \geqslant \lambda c_n, \qquad \alpha = 1, \ldots, n.$$

Положим $E_1 = F_1$, $E_2 = \overline{F}_1 \cap F_2$, $E_3 = \overline{F}_1 \cap \overline{F}_2 \cap F_3, \ldots, E_n = \overline{F}_1 \cap \ldots \cap \overline{F}_{n-1} \cap F_n$, и пусть G_n означает дополнение множества $E_1 \cup E_2 \cup \ldots \cup E_n$. G_n есть множество в R_n, для точек которого имеют место все неравенства (4.5.1). Множества $E_1, \ldots, E_n$, G_n не пересекаются и их объединение образует полное выборочное пространство R_n. Таким образом,

$$P(G_n) = 1 - [P(E_1) + \ldots + P(E_n)]. \tag{4.5.2}$$

Пусть z_α — случайная величина, принимающая значение 1 в каждой точке множества E_α и значение 0 в каждой точке множества $\overline{E}_\alpha$, $\alpha = 1, \ldots, n$, а z_{n+1} — случайная величина, аналогичным образом связанная с G_n. Очевидно, можно написать

$$\mathscr{E}(x_1 + \ldots + x_n)^2 = \mathscr{E}[z_1(x_1 + \ldots + x_n)^2] + \ldots \mathscr{E}[z_n(x_1 + \ldots + x_n)^2] + \\ + \mathscr{E}[z_{n+1}(x_1 + \ldots + x_n)^2]. \tag{4.5.3}$$

Любая функция g_α от переменных $x_1, \ldots, x_\alpha$ является независимой от $(x_{\alpha+1}, \ldots, x_n)$. Следовательно,

$$\mathscr{E}(g_\alpha x_\beta) = \mathscr{E}(g_\alpha)\,\mathscr{E}(x_\beta) = 0, \qquad \beta = \alpha + 1, \ldots, n,$$

так что мы имеем

$$\mathscr{E}[z_\alpha(x_1 + \ldots + x_n)^2] = \mathscr{E}[z_\alpha(x_1 + \ldots + x_\alpha)^2] + \\ + \mathscr{E}[z_\alpha(x_{\alpha+1} + \ldots + x_n)^2] \geqslant \mathscr{E}[z_\alpha(x_1 + \ldots + x_\alpha)^2]. \tag{4.5.4}$$

Но $(x_1 + \ldots + x_\alpha)^2 \geqslant \lambda^2 c_n^2$ в каждой точке множества E_α. В силу этого для $\alpha = 1, \ldots, n$

$$\mathscr{E}[z_\alpha(x_1 + \ldots + x_n)^2] \geqslant \lambda^2 c_n^2 P(E_\alpha). \tag{4.5.5}$$

Теперь вернемся к формуле (4.5.3). Отбрасывая в правой части последнее слагаемое, применяя к остальным неравенство (4.5.5) и, наконец, используя тот факт, что $\mathscr{E}(x_1+\ldots+x_n)^2=c_n^2$, мы получаем

$$c_n^2 \geqslant \lambda^2 c_n^2 [P(E_1)+\ldots+P(E_n)]. \tag{4.5.6}$$

Из (4.5.6) и (4.5.2) вытекает неравенство

$$P(G_n) \leqslant 1-\frac{1}{\lambda^2}, \tag{4.5.7}$$

что и требовалось доказать.

4.6. Усиленный закон больших чисел

В § 4.3 мы уже упоминали о слабом законе больших чисел, согласно которому при некоторых условиях среднее арифметическое первых n компонент последовательности независимых случайных величин с нулевыми средними сходится по вероятности к 0. Возникает вопрос, можем ли мы сделать какое-либо вероятностное утверждение о том, что *все* средние арифметические, начиная с некоторого достаточно удаленного места в своей последовательности, произвольно близки к нулю? Ответ на этот вопрос дает *усиленный закон больших чисел*, который можно сформулировать следующим образом:

4.6.1. *Пусть* $(x_1, x_2, \ldots)$ — *последовательность независимых случайных величин, имеющих нулевые средние и дисперсии* $(\sigma_1^2, \sigma_2^2, \ldots)$, *причем* $\sum_{i=1}^{\infty}\frac{\sigma_i^2}{i^2}<\infty$. *Положим* $\bar{x}_n=\frac{1}{n}(x_1+\ldots+x_n)$. *Тогда* $(\bar{x}_1, \bar{x}_2, \ldots)$ *сходится почти наверное к* 0, *т. е. для любых* $\delta>0$ *и* $\varepsilon>0$ *всегда найдется такое* $N_{\delta,\varepsilon}$, *что*

$$P(|\bar{x}_n|<\varepsilon,\ n=N,\ N+1, \ldots, N+k) \geqslant 1-\delta \tag{4.6.1}$$

при всех $N>N_{\delta,\varepsilon}$ *и для каждого* k.

Доказывая теорему **4.6.1**, мы будем следовать цепи рассуждений, подобной той, которая была использована Феллером (1957). Обозначим через E_n, $n=N, N+1, \ldots, N+k$, соответственно события $|\bar{x}_n|<\varepsilon$, $n=N, N+1, \ldots, N+k$. Вероятность, указанная в (4.6.1) есть

$$P(E_N \cap E_{N+1} \cap \ldots \cap E_{N+k}), \tag{4.6.2}$$

что в свою очередь равно

$$1-P(\bar{E}_N \cup \bar{E}_{N+1} \cup \ldots \cup \bar{E}_{N+k}). \tag{4.6.3}$$

Таким образом, мы должны показать, что при $N>N_{\delta,\varepsilon}$ и любом k

$$P(\bar{E}_N \cup \bar{E}_{N+1} \cup \ldots \cup \bar{E}_{N+k}) \leqslant \delta. \tag{4.6.4}$$

Разобьем совокупность всех целых положительных чисел на множества $I_1, I_2, \ldots$, где $I_\alpha=\{2^{\alpha-1}+1,\ 2^{\alpha-1}+2, \ldots, 2^\alpha\}$. Пусть F_α —

событие, состоящее в том, что по крайней мере одно из неравенств $\{|\bar{x}_n| < \varepsilon,\ n \in I_\alpha\}$ не имеет места. Тогда для некоторых α и β

$$F_\alpha \cup F_{\alpha+1} \cup \ldots \cup F_{\alpha+\beta} \supset \bar{E}_N \cup \bar{E}_{N+1} \cup \ldots \cup \bar{E}_{N+k}. \tag{4.6.5}$$

Но

$$P(F_\alpha \cup F_{\alpha+1} \cup \ldots \cup F_{\alpha+\beta}) \leqslant P(F_\alpha) + \ldots + P(F_{\alpha+\beta}). \tag{4.6.6}$$

Тем самым достаточно установить сходимость ряда $\sum\limits_{\alpha=1}^{\infty} P(F_\alpha)$. Согласно определению F_α по крайней мере для одного $n \in I_\alpha$ мы имеем $|\bar{x}_n| \geqslant \varepsilon$, иначе,

$$|x_1 + \ldots + x_n| \geqslant n\varepsilon.$$

Из последнего неравенства вытекает неравенство

$$|x_1 + \ldots + x_n| > \left(\frac{\varepsilon \cdot 2^{\alpha-1}}{c_{2^\alpha}}\right) c_{2^\alpha},$$

где $c_{2^\alpha}^2 = \sigma_1^2 + \ldots + \sigma_{2^\alpha}^2$, как это требуется в теореме **4.5.1.** Используя эту теорему (неравенство Колмогорова), мы находим

$$P(F_\alpha) \leqslant \frac{4c_{2^\alpha}^2}{\varepsilon^2 \cdot 2^{2\alpha}},$$

так что

$$\sum_{\alpha=1}^{\infty} P(F_\alpha) \leqslant \frac{4}{\varepsilon^2} \sum_{\alpha=1}^{\infty} 2^{-2\alpha} \sum_{i=1}^{2^\alpha} \sigma_i^2. \tag{4.6.7}$$

Меняя здесь порядок суммирования по i и по α и замечая, что сумма $\sum 2^{-2\alpha}$, взятая по всем положительным целым α, для которых $2^\alpha \geqslant i$, не превосходит $2i^{-2}$, мы получаем

$$\sum_{\alpha=1}^{\infty} P(F_\alpha) \leqslant \frac{8}{\varepsilon^2} \sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2}.$$

Следовательно, ряд $\sum\limits_{\alpha=1}^{\infty} P(F_\alpha)$ сходится, если сходится ряд $\sum\limits_{i=1}^{\infty} \frac{\delta_i^2}{i^2}$. Но в таком случае путем выбора достаточно большого α правая часть неравенства (4.6.6) может быть сделана меньше δ для всех значений β, а тогда из (4.6.5) и (4.6.6) следует, что имеет место (4.6.4) и, значит, (4.6.1). Эти аргументы завершают доказательство теоремы **4.6.1.**

Необходимо отметить, что если все компоненты процесса $(x_1, x_2, \ldots)$ имеют одинаковые (конечные) дисперсии, условие $\sum\limits_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} < \infty$ удовлетворяется автоматически. В действительности можно показать, что $(\bar{x}_1, \bar{x}_2, \ldots)$ сходится почти наверное к значению 0, если величины x_1, x_2, ... независимы и одинаково распределены с нулевыми средними [см. Хинчин (1929)].

ЗАДАЧИ

4.1. Показать, что если $(x, y; x_1, y_1; x_2, y_2; \ldots)$ — случайный процесс и процессы $(x_1, x_2, \ldots)$ и $(y_1, y_2, \ldots)$ сходятся по вероятности соответственно к случайным величинам x и y, причем x и y эквивалентны, то для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P(|x_n - y_n| < \varepsilon) = 1.$$

4.2. (*Продолжение*). Показать, что если $\lim\limits_{n\to\infty} \mathscr{E}(x_n - y_n)^2 = 0$ и процесс $(x_1, x_2, \ldots)$ сходится по вероятности к случайной величине x, то процесс $(y_1, y_2, \ldots)$ также сходится по вероятности к этой величине.

4.3. Пусть $(x_1, x_2, \ldots)$ — независимые случайные величины, имеющие одинаковые к. ф. р.

$$F(x) = \begin{cases} 0, & x \leqslant 0, \\ x, & 0 < x \leqslant 1, \\ 1, & x > 1. \end{cases}$$

Образуем новый случайный процесс $(u_1, u_2, \ldots)$, где

$$u_n = \max(x_1, \ldots, x_n),$$

и положим $v_n = n(1 - u_n)$. Показать, что последовательность $(v_1, v_2, \ldots)$ сходится по распределению к случайной величине x, к. ф. р. которой

$$F_1(v) = 1 - e^{-v}.$$

4.4. (*Продолжение*). Показать, что если $g(x)$ — произвольная однозначная дифференцируемая возрастающая функция от x в интервале $(0, \infty)$ и $w_n = g(n(1 - u_n))$, то $(w_1, w_2, \ldots)$ сходится по распределению к случайной величине w, имеющей ф. п. в., равную

$$e^{-g^{-1}(w)} \frac{d}{dw} g^{-1}(w)$$

при $w \in (g(0), g(+\infty))$ и нулю в противном случае; $g^{-1}(w)$ есть обратная функция для $g(x)$.

4.5. Известно, что дискретная случайная величина x_n, возникающая в теории крайних наблюденных значений, имеет к. ф. р.

$$F_n(x) = 1 - \frac{n(n-1)\ldots(n-r+1)}{(n+nx)(n+nx-1)\ldots(n+nx-r+1)},$$

где $r \leqslant n+1$, причем точки сосредоточения массы величины x_n суть $\frac{1}{n}, \frac{2}{n}, \ldots$ Показать, что последовательность случайных величин $(x_1, x_2, \ldots)$ сходится по распределению к случайной величине x, имеющей *непрерывную* к. ф. р.

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - \dfrac{1}{(1+x)^r}, & x > 0. \end{cases}$$

4.6. Пользуясь теоремой **4.3.6а**, показать, что если d — произвольная отличная от нуля константа, то

$$\lim_{n\to\infty} P(x_n + ny_n \leqslant d) = 0.$$

4.7. Пусть $(x_1, x_2, \ldots)$ — случайный процесс, причем $\mathscr{E}(x_n)=0$, $\sigma^2(x_n)=\sigma^2$, $n=1, 2, \ldots$, и $\operatorname{cov}(x_i, x_j)=\rho\sigma^2$, $i\neq j$, $i, j=1, 2, \ldots$, где σ^2 конечно и $\rho\in(0, 1)$. Положим

$$u_k=\frac{1}{k}(x_1+x_3+\ldots+x_{2k-1})$$

и

$$v_k=\frac{1}{k}(x_2+x_4+\ldots+x_{2k}).$$

Показать, что векторный случайный процесс $(u_1, v_1; u_2, v_2; \ldots)$ сходится по распределению к (u, v), где u и v — эквивалентные случайные величины, имеющие средние 0 и дисперсии $\sigma^2\rho$.

4.8. Пусть $x_1, \ldots, x_n$ — n независимых случайных величин, имеющих нулевые средние и единичные дисперсии. Рассмотрим неравенства

$$|\bar{x}_\alpha|<\lambda\frac{1}{\sqrt{\alpha}}, \qquad \alpha=1, \ldots, n,$$

где

$$\bar{x}_\alpha=\frac{1}{\alpha}(x_1+\ldots+x_\alpha),$$

и обозначим через G_n событие, состоящее в их совместном выполнении. Тогда $\bar{G}_n$ эквивалентно невыполнению по крайней мере одного из этих неравенств. Показать с помощью метода, подобного использованному при рассмотрении неравенства Колмогорова, что для любого $\lambda>0$

$$P(\bar{G}_n)\leqslant\frac{n}{n+(\lambda^2-1)\,\mathscr{E}(\alpha\,|\,\bar{G}_n)},$$

где α — случайная величина, обозначающая номер первого из указанных выше неравенств, которое не имеет места, и $\mathscr{E}(\alpha\,|\,\bar{G}_n)$ — условное среднее значение величины α при условии, что событие $\bar{G}_n$ уже произошло.

4.9. (*Продолжение*). Рассмотрим неравенства

$$x_1^2+\ldots+x_\alpha^2<\alpha\lambda^2, \qquad \alpha=1, \ldots, n.$$

Обозначив через $\bar{G}_n$ событие, состоящее в невыполнении по крайней мере одного из этих неравенств, показать, что $P(\bar{G}_n)$ оценивается сверху точно так же, как и в предыдущей задаче.

4.10. Показать, что если $(x_1, \ldots, x_n)$ — независимые положительные случайные величины, средние значения которых равны одному и тому же числу μ, то для любого $\lambda>0$ вероятность совместного осуществления неравенств

$$(x_1x_2\ldots x_\alpha)<\lambda\mu^\alpha, \qquad \alpha=1, \ldots, n,$$

не меньше чем $1-\dfrac{1}{\lambda}$.

4.11. Распространить усиленный закон больших чисел на тот случай, когда компоненты последовательности независимых случайных величин $(x_1, x_2, \ldots)$ имеют (конечные) средние $\mu_1, \mu_2, \ldots$, а также (конечные) дисперсии $\sigma_1^2, \sigma_2^2, \ldots$

4.12. Показать, что если $x_1, \ldots, x_n$ — n независимых случайных величин, медианы которых равны нулю, то

$$\mathscr{E}(|x_1+\ldots+x_n|)\geqslant\varphi(n)\,d,$$

где

$$d=\frac{1}{n}[\mathscr{E}|x_1|+\ldots+\mathscr{E}|x_n|]$$

и

$$\varphi(2k+1)=\varphi(2k+2)=\frac{(2k+1)!}{2^{2k}(k!)^2}$$

— результат, принадлежащий Тьюки (1946).

4.13. Пусть $(x_1, \ldots, x_k)$ — k-мерная случайная величина с вектором средних $(0, \ldots, 0)$ и ковариационной матрицей $\|\sigma_{ij}\|$, где $\sigma_{ii}=\sigma^2$ и $\sigma_{ij}=\sigma^2\rho$. Показать, что

$$P(\max_{\alpha}|x_\alpha|\geqslant t\sigma)\leqslant\frac{\sigma}{kt^2}\left[(k-1)\sqrt{1-\rho}+\sqrt{1+(k-1)\rho}\right]$$

— результат, принадлежащий Олкину и Прэтту (1958).

4.14. Показать, что если два случайных процесса $(x_1, x_2, \ldots)$ и $(y_1, y_2, \ldots)$ сходятся по вероятности соответственно к случайной величине x и положительной константе c, то имеют место четыре утверждения **4.3.6a:** (4.3.18), (4.3.19), (4.3.20) и (4.3.21).

4.15. Доказать, что если $(x_1, x_2, \ldots)$ — последовательность случайных величин с ф. п. в. $(f_1(x), f_2(x), \ldots)$, причем $\lim\limits_{\alpha\to\infty} f_\alpha(x)=f(x)$ для всех $x\in R_1$, за исключением, быть может, множества их вероятности 0, где $f(x)$ — также некоторая ф. п. в., то

$$\lim_{\alpha\to\infty}\int_E f_\alpha(x)\,dx=\int_E f(x)\,dx$$

равномерно по множествам $E\subset R$. Этот результат принадлежит Шеффе (1947). Следует отметить, что условия теоремы Шеффе сильнее, чем условия теоремы **4.3.4**. Рассмотрение различных условий сходимости распределений и их сравнение были осуществлены Роббинсом (1948).

4.16. Пусть случайный процесс $(x, x_1, x_2, \ldots)$ таков, что $(x_1, x_2, \ldots)$ сходится по вероятности к случайной величине x и $g_1(x), g_2(x), \ldots$ — последовательность непрерывных функций, сходящаяся равномерно к функции $g(x)$ в некотором ограниченном интервале. Показать, что $(g_1(x), g_2(x), \ldots)$ есть случайный процесс, который сходится по вероятности к величине $g(x)$.

4.17. Пусть $x_1, \ldots, x_n$ — независимые случайные величины. Положим $y_1=x_1$, $y_2=x_1+x_2$, $\ldots$, $y_n=x_1+\ldots+x_n$ и обозначим через $F_\xi(y_\xi)$ к. ф. р. величины y_ξ, $\xi=1, \ldots, n$, а через $F(y_1, \ldots, y_n)$ — к. ф.р. величины $(y_1, \ldots, y_n)$.

Показать, что

$$F(y_1, \ldots, y_n)\geqslant F_1(y_1)\ldots F_n(y_n)$$

— результат, принадлежащий Роббинсу (1954).

4.18. Доказать теорему **4.3.7.**

ГЛАВА 5

ХАРАКТЕРИСТИЧЕСКИЕ И ПРОИЗВОДЯЩИЕ ФУНКЦИИ

5.1. Случай одномерной величины

Одним из наиболее важных классов задач в математической статистике является класс задач на отыскание функций распределения измеримых функций случайных величин, т. е. таких функций случайных величин, которые сами суть случайные величины. Несколько методов исследования подобных задач было представлено в § 2.7. Однако часто эти методы технически трудны, чтобы их применять в конкретных случаях. В то же время отдельные ситуации, в частности включающие линейные функции независимых случайных величин, можно изящным образом исследовать, вводя характеристические функции данных функций рассматриваемых величин. Характеристические функции и связанный с ними аппарат в некоторых случаях оказываются полезными при решении таких задач, как задача воспроизведения моментов и кумулянтов распределений, а также задача проверки гипотезы независимости двух или более функций случайных величин. Настоящая глава будет посвящена методу характеристических функций и его приложениям.

Характеристическая функция $\varphi(t)$ случайной величины x, имеющей к. ф. р. $F(x)$, определяется следующим образом:

$$\varphi(t) = \mathscr{E}(e^{\mathrm{i}tx}) = \int_{-\infty}^{\infty} e^{\mathrm{i}tx} dF(x), \qquad (5.1.1)$$

где $\mathrm{i} = \sqrt{-1}$ *) и t — вещественное число. Иногда удобно говорить, что $\varphi(t)$ является характеристической функцией, соответствующей $F(x)$, или, более кратко, характеристической функцией величины x.

Так как

$$e^{\mathrm{i}tx} = \cos tx + \mathrm{i} \sin tx$$

и обе функции $\cos tx$ и $\sin tx$ интегрируемы по пространству R_1, то $\varphi(t)$ есть комплексное число, вещественная и мнимая части которого конечны при любом значении t.

---

*) На всем протяжении этой книги мы обозначаем $\sqrt{-1}$ через i, чтобы избежать путаницы в связи с использованием в различных параграфах курсивной буквы i в качестве индекса суммирования.

Производящая функция моментов $\psi(t)$ определяется соотношением

$$\psi(t)=\mathscr{E}(e^{tx})=\varphi\left(\frac{t}{\mathrm{i}}\right). \tag{5.1.2}$$

Обычно мы интересуемся значениями $\psi(t)$ в некоторой окрестности точки $t=0$. Но в отличие от $\varphi(t)$ функция $\psi(t)$ при всех значениях t существует не для каждой к. ф. р.

Производящая функция факториальных моментов $\theta(t)$, если она существует, получается из (5.1.2) путем замены t на $\ln t$, т. е.

$$\theta(t)=\psi(\ln t)=\mathscr{E}(t^x). \tag{5.1.3}$$

Если случайная величина x дискретна и ее точки сосредоточения массы суть целые положительные числа, то $\theta(t)$, если она существует, также называется *производящей функцией вероятностей* случайной величины x. Ибо если $\theta(t)$ разлагается в ряд по степеням t, то коэффициент при t^x равен $p(x)$, представляя ф. в. величины x.

Если r-й момент $\mu'_r(x)$ существует, то мы можем продифференцировать функцию (5.1.1) h раз по переменной t и получить

$$\varphi^{(h)}(t)=\mathrm{i}^h\int\limits_{-\infty}^{\infty}x^h e^{\mathrm{i}tx}dF(x). \tag{5.1.4}$$

Отсюда мы находим, что

$$\mu'_h(x)=\frac{\varphi^{(h)}(0)}{\mathrm{i}^h},\qquad 0<h\leqslant r. \tag{5.1.5}$$

Кроме того, для удобства положим $\varphi^{(0)}(t)\equiv\varphi(t)$.

Аналогичным образом, если $\psi(t)$ существует при всех значениях t в некоторой окрестности нулевой точки, мы имеем

$$\mu'_h(x)=\psi^{(h)}(0). \tag{5.1.6}$$

Если все моменты величины x до порядка $2s$ включительно существуют, мы можем записать

$$\varphi(t)=\int\limits_{-\infty}^{\infty}\left[1+\sum_{h=1}^{2s-1}\frac{(\mathrm{i}tx)^h}{h!}+\frac{(\mathrm{i}tx)^{2s}}{(2s)!}(\cos t'x+\mathrm{i}\sin t''x)\right]dF(x), \tag{5.1.7}$$

где t' и t'' — числа, принадлежащие интервалу $(0, t)$. Если теперь случайная величина x не является вырожденной, то $\mu'_{2s}\neq 0$ и, следовательно,

$$\varphi(t)=1+\sum_{h=1}^{2s-1}\frac{(\mathrm{i}t)^h}{h!}\mu'_h+\frac{(\mathrm{i}t)^{2s}\mu'_{2s}}{(2s)!}g_1(t, s), \tag{5.1.8}$$

где $g_1(t, s)$ — комплексная функция, вещественная и мнимая, компоненты которой соответственно равны средним значениям $\mathscr{E}\left(x^{2s}\cos t'x\frac{1}{\mu'_{2s}}\right)$ и $\mathscr{E}\left(x^{2s}\sin t''x\frac{1}{\mu'_{2s}}\right)$. Ясно, что $|g_1(t, s)|\leqslant 1$.

Более того, поскольку оба числа, t' и t'', лежат в интервале $(0, t)$, $\lim\limits_{t\to 0} g_1(t, s) = 1$.

Точно так же, если существуют все моменты до порядка $(2s-1)$ включительно, мы можем написать

$$\varphi(t) = 1 + \sum_{h=1}^{2s-2} \frac{(\mathrm{i}\,t)^h}{h!} \mu_h + \frac{(\mathrm{i}\,t)^{2s-1}}{(2s-1)!} \nu'_{2s-1} g''_2(t, s), \tag{5.1.9}$$

где $g_2(t, s)$ — комплексная функция, вещественная и мнимая компоненты которой соответственно равны $\mathscr{E}\left(x^{2s-1}\cos t_1 x \frac{1}{\nu'_{2s-1}}\right)$ и $\mathscr{E}\left(x^{2s-1}\sin t_2 x \frac{1}{\nu'_{2s-1}}\right)$, причем числа t_1 и t_2^{2s-1} принадлежат интервалу $(0, t)$ и ν'_{2s-1} есть определяемый формулой (3.3.13) абсолютный момент порядка $(2s-1)$, отличный от нуля, если x — невырожденная величина. Снова мы имеем $|g_2(t, s)| \leqslant 1$ и, сверх того, $\lim\limits_{t\to 0} g_2(t, s) = \frac{\mu'_{2s-1}}{\nu'_{2s-1}}$, где этот предел, разумеется, конечен.

Независимо от того, является момент самого высокого порядка четным или нет, мы можем утверждать следующее:

5.1.1. *Если r-й момент случайной величины x существует, то в некоторой окрестности точки $t = 0$ функцию $\varphi(t)$ можно представить в виде*

$$\varphi(t) = 1 + \sum_{h=1}^{r} \frac{(\mathrm{i}\,t)^h}{h!} \mu_h + o(t^r), \tag{5.1.10}$$

где

$$\lim_{t\to 0} \frac{o(t^r)}{t^r} = 0.$$

Если представление (5.1.10) функции $\varphi(t)$ законно, можно получить аналогичное представление логарифма этой функции

$$\ln \varphi(t) = \sum_{h=1}^{r} \frac{\varkappa_h}{h!} (\mathrm{i}\,t)^h + o(t^r). \tag{5.1.11}$$

Величины $\varkappa_h$, первоначально введенные и изученные Сайилом (1903), называются *семиинвариантами* или *кумулянтами* к. ф. р. $F(x)$.

Заметим, что любой семиинвариант $\varkappa_h$ есть полином относительно $\mu'_1, \mu'_2, \ldots$, и наоборот. Например,

$$\begin{aligned} \varkappa_1 &= \mu'_1 = \mu, \\ \varkappa_2 &= \mu'_2 - (\mu'_1)^2 = \sigma^2, \\ \varkappa_3 &= \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3, \\ &\ldots\ldots\ldots\ldots \end{aligned} \tag{5.1.12}$$

Обратно,

$$\begin{aligned} \mu_1' &= \varkappa_1, \\ \mu_2' &= \varkappa_2 + \varkappa_1^2, \\ \mu_3' &= \varkappa_3 + 3\varkappa_1\varkappa_2 + \varkappa_1^3, \\ &\dots\dots\dots \end{aligned} \tag{5.1.13}$$

Существует много задач в теории вероятностей и ее приложениях, особенно в теории выборочного метода, в которых совсем просто определить характеристическую функцию данной функции компонент многомерной случайной величины, в частности линейной функции. Возникает естественный вопрос: как найти к. ф. р. случайной величины по ее характеристической функции? Во многих случаях ответ на этот вопрос может быть дан следующей теоремой, принадлежащей Леви (1925):

5.1.2. *Пусть x — случайная величина, имеющая характеристическую функцию $\varphi(t)$ и к. ф. р. $F(x)$. Тогда, если $x = x' \pm \delta$, $\delta > 0$ — точки непрерывности функции $F(x)$, то*

$$F(x'+\delta) - F(x'-\delta) = \lim_{A\to\infty} \frac{1}{\pi} \int_{-A}^{A} \frac{\sin \delta t}{t} e^{-\mathrm{i}tx'} \varphi(t)\, dt. \tag{5.1.14}$$

Кроме того, если выполнено условие $\int_{-\infty}^{\infty} |\varphi(t)|\, dt < \infty$, то в точке $x = x'$ существует ф. п. в. $f(x)$, причем

$$f(x') = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\mathrm{i}tx'} \varphi(t)\, dt. \tag{5.1.15}$$

Чтобы доказать эту теорему, рассмотрим интеграл в правой части (5.1.14) и заменим в нем функцию $\varphi(t)$ ее интегральным выражением согласно (5.1.1). Пусть

$$G(x', A, \delta) = \frac{1}{\pi} \int_{-A}^{A} \frac{\sin \delta t}{t} e^{-\mathrm{i}tx'} \int_{-\infty}^{+\infty} e^{\mathrm{i}tx}\, dF(x)\, dt. \tag{5.1.16}$$

Так как $\left| \frac{\sin \delta t}{t} e^{\mathrm{i}t(x'-x)} \right| < \delta$, мы можем изменить в (5.1.16) порядок интегрирования и переписать эту формулу в виде

$$G(x', A, \delta) = \int_{-\infty}^{\infty} m(x, x', A, \delta)\, dF(x), \tag{5.1.17}$$

где

$$m(x, x', A, \delta) = \frac{1}{\pi} \int_{-A}^{A} \frac{\sin \delta t}{t} e^{it(x-x')}\, dt =$$

$$= \frac{2}{\pi} \int_{0}^{A} \frac{\sin \delta t}{t} \cos [t(x-x')]\, dt =$$

$$= \frac{1}{\pi} \int_{0}^{A} \frac{\sin (x-x'+\delta) t}{t}\, dt - \frac{1}{\pi} \int_{0}^{A} \frac{\sin (x-x'-\delta) t}{t}\, dt. \quad (5.1.18)$$

Если теперь перейти к пределу при $A \to \infty$, то

$$\lim_{A\to\infty} G(x', A, \delta) = \int_{-\infty}^{\infty} \lim_{A\to\infty} m(x, x', A, \delta)\, dF(x). \quad (5.1.19)$$

Но, используя известное соотношение $\lim\limits_{\nu\to\infty} \int\limits_0^{\nu} \frac{\sin u}{u}\, du = \frac{\pi}{2}$, можно показать, что

$$\lim_{A\to\infty} m(x, x', A, \delta) = \begin{cases} 0, & x < x' - \delta,\ x > x' + \delta, \\ \frac{1}{2}, & x = x' - \delta,\ x = x' + \delta, \\ 1, & x' - \delta < x < x' + \delta. \end{cases} \quad (5.1.20)$$

После подстановки найденных значений $\lim\limits_{A\to\infty} m(x, x', A, \delta)$ в (5.1.19) и учета непрерывности функции $F(x)$ в точках $x = x' \pm \delta$, мы получаем

$$\lim_{A\to\infty} G(x', A, \delta) = \int_{x'-\delta}^{x'+\delta} dF(x) = F(x'+\delta) - F(x'-\delta),$$

что доказывает формулу (5.1.14).

Далее, деля обе части (5.1.14) на 2δ, мы имеем

$$\frac{F(x'+\delta) - F(x'-\delta)}{2\delta} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin \delta t}{\delta t} e^{-itx'} \varphi(t)\, dt. \quad (5.1.21)$$

Очевидно, $\left| \frac{\sin \delta t}{\delta t} e^{-itx'} \varphi(t) \right| < |\varphi(t)|$. В таком случае, если $\int\limits_{-\infty}^{\infty} |\varphi(t)|\, dt < \infty$, и, кроме того, функция $F(x)$ обладает в точке $x = x'$ производной $f(x')$,

$$\lim_{\delta\to 0} \left[\frac{F(x'+\delta) - F(x'-\delta)}{2\delta} \right] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{\delta\to 0} \left(\frac{\sin \delta t}{\delta t} \right) e^{-itx'} \varphi(t)\, dt.$$

Но $\lim\limits_{\delta \to 0} \left(\frac{\sin \delta t}{\delta t}\right) = 1$ и, следовательно,

$$f(x') = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-\mathrm{i}\, tx'} \varphi(t)\, dt,$$

что совпадает с формулой (5.1.15).

Если две случайные величины x_1 и x_2 имеют одинаковые к. ф. р., то ясно, что их характеристические функции также одинаковы. Обратно, пусть случайные величины имеют одну и ту же характеристическую функцию $\varphi(t)$. Тогда, если $F_1(x)$ и $F_2(x)$ суть к. ф. р. этих величин и $[x' - \delta,\ x' + \delta]$ — любой интервал, на концах которого обе к. ф. р. непрерывны, из формулы (5.1.14) следует, что

$$F_1(x' + \delta) - F_1(x' - \delta) = F_2(x' + \delta) - F_2(x' - \delta).$$

Вместе с нашим соглашением считать все к. ф. р. непрерывными справа это приводит к тождеству $F_1(x) \equiv F_2(x)$.

Таким образом, имеет место теорема.

5.1.3. *К. ф. р. двух случайных величин x_1 и x_2 совпадают тогда и только тогда, когда совпадают характеристические функции этих величин.*

Это взаимно однозначное соответствие между к. ф. р. и характеристическими функциями весьма полезно для теории вероятностей, так как составляет основу, с помощью которой к. ф. р. можно опознать по соответствующим им характеристическим функциям.

Пример. Пусть $x_1, \ldots, x_k$ — статистически независимые случайные величины, имеющие ф. п. в.

$$f(x_1, \ldots, x_k) = \begin{cases} e^{-\sum\limits_{i=1}^{k} x_i}, & x_i \geqslant 0,\ i = 1, \ldots, k, \\ 0 & \text{в противном случае.} \end{cases}$$

Положим $L = \sum\limits_{i=1}^{k} x_i$ и найдем моменты и ф. п. в. величины L.

Характеристическая функция величины L равна

$$\varphi(t) = \int\limits_0^{\infty} \ldots \int\limits_0^{\infty} e^{-\sum\limits_i x_i + \mathrm{i} t \sum\limits_i x_i} dx_1 \ldots dx_k = (1 - \mathrm{i}t)^{-k}. \qquad (5.1.22)$$

Применяя (5.1.5), мы получаем момент r-го порядка

$$\mu'_r(L) = \frac{\varphi^{(r)}(0)}{\mathrm{i}^r} = \frac{\Gamma(k + r)}{\Gamma(k)}.$$

Ф. п. в. величины L согласно (5.1.15) дается выражением

$$f(L) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-\mathrm{i}tL} (1 - \mathrm{i}\, t)^{-k}\, dt. \qquad (5.1.23)$$

Этот интеграл можно вычислить путем контурного интегрирования в комплексной плоскости. Сделаем преобразование

$$z = -L(1 - \mathrm{i}\,t).$$

Тогда

$$f(L) = L^{k-1} e^{-L} H,$$

где H — интеграл Ханкеля,

$$H = \frac{\mathrm{i}}{2\pi} \int_{-L+\mathrm{i}\infty}^{-L-\mathrm{i}\infty} e^{-z} (-z)^{-k}\, dz, \tag{5.1.24}$$

значение которого равно $\frac{1}{\Gamma(k)}$. [См., например, Уиттекер и Ватсон (1946).] Следовательно, ф. п. в. величины L имеет вид

$$\begin{aligned} f(L) &= \frac{1}{\Gamma(k)} L^{k-1} e^{-L}, \quad && L \geqslant 0, \\ f(L) &= 0, && L < 0. \end{aligned} \tag{5.1.25}$$

5.2. Случай k-мерной величины

Допустим, что $(x_1, \ldots, x_k)$ — k-мерная случайная величина, имеющая к. ф. р. $F(x_1, \ldots, x_k)$. *Характеристическая функция* $\varphi(t_1, \ldots, t_k)$ величины $(x_1, \ldots, x_k)$ определяется формулой

$$\varphi(t_1, \ldots, t_k) = \mathscr{E}\left[\exp\left(\mathrm{i}\sum_{i=1}^{k} t_i x_i\right)\right] =$$

$$= \int_{R_k} \exp\left(\mathrm{i}\sum_{i=1}^{k} t_i x_i\right) dF(x_1, \ldots, x_k). \tag{5.2.1}$$

Формула

$$\psi(t_1, \ldots, t_k) = \varphi\left(\frac{t_1}{\mathrm{i}}, \ldots, \frac{t_k}{\mathrm{i}}\right) \tag{5.2.2}$$

определяет *производящую функцию моментов* величины $(x_1, \ldots, x_k)$.

Подобно предыдущему, заменяя в (5.2.2) t_i на $\ln t_i$, $i = 1, \ldots, k$, мы приходим к производящей функции факториальных моментов $\theta(t_1, \ldots, t_k)$.

Если смешанный момент $\mu'_{r_1, \ldots, r_k}$ существует, он может быть получен в результате дифференцирования функции $\varphi(t_1, \ldots, t_k)$, именно,

$$\mu'_{r_1, \ldots, r_k} = \frac{\varphi^{(r_1 + \ldots + r_k)}(0, \ldots, 0)}{\mathrm{i}^{(r_1 + \ldots + r_k)}}. \tag{5.2.3}$$

Если $\mu'_{r_1, \ldots, r_k}$ существует, то также существуют все смешанные моменты $\mu'_{h_1, \ldots, h_k}$, где $0 < h_i \leqslant r$, $i = 1, \ldots, k$.

Следует заметить, что характеристическая функция любого подмножества компонент случайной величины $(x_1, \ldots, x_k)$ получается из характеристической функции этой величины приравниванием нулю значений t_i, соответствующих компонентам, не входящим в данное подмножество. Например, характеристическая функция случайной величины $(x_1, \ldots, x_{k_1})$, $k_1 < k$, есть $\varphi(t_1, \ldots, t_{k_1}, 0, \ldots, 0)$.

Распространение теоремы Леви на случай k величин несложно. Мы сформулируем его следующим образом.

5.2.1. *Пусть $(x_1, \ldots, x_k)$ — k-мерная случайная величина с характеристической функцией $\varphi(t_1, \ldots, t_k)$ и к. ф. р. $F(x_1, \ldots, x_k)$. Обозначим через I_k интервал в пространстве R_k, задаваемый системой неравенств $x_i' - \delta < x_i \leqslant x_i' + \delta$, $i = 1, \ldots, k$, $\delta > 0$, и предположим, что функция $F(x_1, \ldots, x_k)$ непрерывна на границе этого интервала. Тогда*

$$P((x_1, \ldots, x_k) \in I_k) = \\ = \lim_{A \to \infty} \frac{1}{\pi^k} \int_{-A}^{A} \ldots \int_{-A}^{A} \prod_{i=1}^{k} \left[\frac{\sin \delta_i t_i}{t_i} e^{-\mathrm{i}\, t_i x_i}\right] \varphi(t_1, \ldots, t_k)\, dt_1 \ldots dt_k. \quad (5.2.4)$$

Кроме того, если выполнено условие $\int_{R_k} |\varphi(t_1, \ldots, t_k)|\, dt_1 \ldots dt_k < \infty$, то в точке $(x_1', \ldots, x_k')$ существует ф. п. в. $f(x_1, \ldots, x_k)$, причем

$$f(x_1', \ldots, x_k') = \\ = \left(\frac{1}{2\pi}\right)^k \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\left(-\mathrm{i} \sum_{i=1}^{k} t_i x_i\right) \varphi(t_1, \ldots, t_k)\, dt_1 \ldots dt_k. \quad (5.2.5)$$

Доказательство теоремы **5.2.1** аналогично доказательству теоремы **5.1.2** и потому опускается.

Теорема **5.1.3** о взаимно однозначном соответствии между к. ф. р. и характеристическими функциями может быть распространена на случай k-мерных величин без каких-либо новых трудностей.

5.3. Характеристические функции независимых случайных величин

Характеристические функции иногда используются при выяснении вопроса о независимости двух случайных величин, позволяя избежать предварительного определения соответствующей двумерной к. ф. р. Приведем основной результат, на который при этом опираются.

5.3.1. *Пусть (x_1, x_2) — двумерная случайная величина и $\varphi(t_1, t_2)$ — ее характеристическая функция. Величины x_1 и x_2 независимы тогда и только тогда, когда выполнено условие*

$$\varphi(t_1, t_2) = \varphi(t_1, 0)\, \varphi(0, t_2). \quad (5.3.1)$$

Заметим, что $\varphi(t_1, 0)$ есть характеристическая функция величины x_1, а $\varphi(0, t_2)$ — характеристическая функция величины x_2. Удобно обозначить $\varphi(t_1, 0)$ и $\varphi(0, t_2)$ соответственно через $\varphi_1(t_1)$ и $\varphi_2(t_2)$.

Чтобы убедиться в *необходимости* условия (5.3.1), мы предположим, что величины x_1 и x_2 независимы, т. е. $F(x_1, x_2) = F_1(x_1) F_2(x_2)$. Тогда

$$\varphi(t_1, t_2) = \int\limits_{R_2} e^{it_1x_1 + it_2x_2}\, dF(x_1, x_2) = \int\limits_{R_2} e^{it_1x_1 + it_2x_2}\, d[F_1(x_1) F_2(x_2)] =$$

$$= \int\limits_{-\infty}^{\infty} e^{it_1x_1}\, dF_1(x_1) \int\limits_{-\infty}^{\infty} e^{it_2x_2}\, dF_2(x_2), \tag{5.3.2}$$

иначе

$$\varphi(t_1, t_2) = \varphi_1(t_1)\, \varphi_2(t_2).$$

Теперь покажем *достаточность*. Предполагая, что условие (5.3.1) выполнено, мы согласно формуле (5.2.4) при $k = 2$ имеем

$$P(x_i' - \delta_i < x_i \leqslant x_i' + \delta_i;\ i = 1, 2) =$$

$$= \lim_{A \to \infty} \frac{1}{\pi^2} \int\limits_{-A}^{A} \int\limits_{-A}^{A} \prod_{i=1}^{2} \left[\frac{\sin \delta_i t_i}{t_i} e^{-it_ix_i'}\right] \varphi_1(t_1)\, \varphi_2(t_2)\, dt_1\, dt_2 =$$

$$= \prod_{i=1}^{2} \left[\lim_{A \to \infty} \frac{1}{\pi} \int\limits_{-A}^{A} \frac{\sin \delta_i t_i}{t_i} e^{-it_ix_i'} \varphi_i(t_i)\, dt_i\right]. \tag{5.3.3}$$

Или, в другой записи,

$$P((x_1, x_2) \in I_2) = P(x_1 \in I_1^{(1)})\, P(x_2 \in I_1^{(2)}), \tag{5.3.4}$$

где $I_1^{(i)}$ есть интервал $x_i' - \delta_i < x_i \leqslant x_i' + \delta_i$, $i = 1, 2$, и I_2 — декартово произведение $I_1^{(1)} \times I_1^{(2)}$. Формула (5.3.4) приводит к соотношению

$$F(x_1, x_2) = F_1(x_1) F_2(x_2),$$

из которого по теореме **2.4.2** вытекает независимость x_1 и x_2.

Распространение теоремы **5.3.1** на любое (конечное) число случайных величин очевидно и предоставляется читателю.

Другое полезное свойство характеристических функций независимых случайных величин связано с рассмотрением линейных функций этих величин. Главный результат в этом направлении, который легко может быть проверен читателем, дается следующей теоремой.

5.3.2. *Пусть*

$$L = \sum_{i=1}^{k} c_i x_i \tag{5.3.5}$$

— линейная функция k независимых случайных величин $x_1, \ldots, x_k$. Пусть $\varphi(t)$ — характеристическая функция величины L и $\varphi_i(t_i)$ — характеристическая функция величины x_i, $i = 1, \ldots, k$. Тогда

$$\varphi(t) = \prod_{i=1}^{k} \varphi_i(c_i t_i). \tag{5.3.6}$$

Предположим, что x_1 и x_2 — независимые случайные величины, которые имеют соответственно к. ф. р. $F(x_1;\ \theta_1)$ и $F(x_2;\ \theta_2)$, где θ_1 и θ_2 — значения параметра θ. Обозначим через L случайную величину x_1+x_2. Тогда, если к. ф. р. величины L равна $F(L, \theta_1+\theta_2)$, то к. ф. р. $F(x, \theta)$ называется *воспроизводящей по* θ. Подобным же образом можно говорить, что ф. в., ф. п. в. или распределение, зависящее от θ, являются воспроизводящими по θ. Воспроизводимость представляет важное свойство, которое часто будет использоваться в более поздних главах. Понятие воспроизводимости можно без труда распространить на тот случай, когда либо величина x, либо параметр θ, либо и то и другое многомерны. В терминах характеристических функций можно сформулировать следующий полезный критерий воспроизводимости к. ф. р.

5.3.3. *Пусть x_1 и x_2 — независимые случайные величины, имеющие соответственно к. ф. р. $F(x_1;\ \theta_1)$ и $F(x_2;\ \theta_2)$ и характеристические функции $\varphi(t;\ \theta_1)$ и $\varphi(t;\ \theta_2)$, где θ_1 и θ_2 — параметры. Для того чтобы функция $F(x;\ \theta)$ была воспроизводящей, необходимо и достаточно выполнение условия*

$$\varphi(t;\ \theta_1)\,\varphi(t;\ \theta_2)=\varphi(t;\ \theta_1+\theta_2). \tag{5.3.7}$$

Доказательство этой теоремы несложно и предоставляется читателю.

5.4. Характеристические функции последовательности случайных величин

Как мы указали в § 4.1, одной из важных задач в теории вероятностей, а также в ее приложениях к теории выборочного метода и другим близким разделам статистики является задача сходимости по вероятности последовательности случайных величин. Так как к. ф. р. единственным образом определяются характеристическими функциями, задача сходимости по вероятности последовательности случайных величин часто может быть с большей легкостью решена путем рассмотрения сходимости соответствующей последовательности характеристических функций, нежели путем непосредственного рассмотрения к. ф. р. случайных величин. Применяемый здесь фундаментальный принцип выражается нижеследующей теоремой, принадлежащей Леви (1937) и Крамеру (1937).

5.4.1. *Пусть $(x_1, x_2, \ldots)$ — некоторая последовательность случайных величин и $\varphi_1(t)$, $\varphi_2(t)$, ... — соответствующая последовательность характеристических функций. Последовательность $(x_1, x_2, \ldots)$ сходится по распределению к случайной величине x тогда и только тогда, когда при любом t последовательность $\varphi_1(t), \varphi_2(t), \ldots$ сходится к пределу $\varphi(t)$ и предельная функция непрерывна в точке $t=0$. При этом условии предельная функция $\varphi(t)$ совпадает с характеристической функцией величины x.*

Сначала покажем, что условие теоремы *необходимо*. Пусть $F_1(x)$, $F_2(x)$, ... — к. ф. р. соответственно величин x_1, x_2, ... и $F(x)$ — к. ф. р. величины x. Предположим, что последовательность $(x_1, x_2, \ldots)$ сходится по распределению к величине x. Мы должны установить, что при любом t $\lim\limits_{n\to\infty} \varphi_n(t) = \varphi(t)$. Очевидно, можно написать

$$\varphi_n(t) = \int\limits_{-\infty}^{\infty} \cos tx\, dF_n(x) + \mathrm{i} \int\limits_{-\infty}^{\infty} \sin tx\, dF_n(x). \tag{5.4.1}$$

Так как $\cos tx$ для каждого t есть ограниченная функция от $x \in (-\infty, +\infty)$ и, кроме того, последовательность $F_1(x), F_2(x), \ldots$ сходится к к. ф. р. $F(x)$ в каждой ее точке непрерывности, то по теореме Хелли — Брея [см., например, Лоэв (1955)]

$$\lim_{n\to\infty} \int\limits_{-\infty}^{\infty} \cos tx\, dF_n(x) = \int\limits_{-\infty}^{\infty} \cos tx\, dF(x). \tag{5.4.2}$$

Подобным же образом мы находим

$$\lim_{n\to\infty} \int\limits_{-\infty}^{\infty} \sin tx\, dF_n(x) = \int\limits_{-\infty}^{\infty} \sin tx\, dF(x). \tag{5.4.3}$$

Но (5.4.2) и (5.4.3), взятые вместе, эквивалентны соотношению

$$\lim_{n\to\infty} \int\limits_{-\infty}^{\infty} e^{\mathrm{i}tx} dF_n(x) = \int\limits_{-\infty}^{\infty} e^{\mathrm{i}tx} dF(x),$$

а это и означает, что

$$\lim_{n\to\infty} \varphi_n(t) = \varphi(t).$$

Теперь рассмотрим *достаточность* условия. Пусть $\lim\limits_{n\to\infty} \varphi_n(t) = \varphi(t)$ при любом t и функция $\varphi(t)$ непрерывна в точке $t=0$. Можно показать [см., например, Крамер (1946)], что последовательность к. ф. р. $F_1(x)$, $F_2(x)$, ... всегда содержит подпоследовательность $F_{n_1}(x)$, $F_{n_2}(x)$, ..., которая сходится к неубывающей непрерывной справа функции $F(x)$ в каждой ее точке непрерывности. Наша ближайшая задача состоит в том, чтобы убедиться, что функция $F(x)$, обладающая указанными свойствами, удовлетворяет и остальным условиям, налагаемым на к. ф. р., именно, $F(-\infty) = 0$ и $F(+\infty) = 1$. То, что $0 \leqslant F(x) \leqslant 1$, сомнений не вызывает.

С помощью аргументов, аналогичных использованным при доказательстве формулы (5.1.14), можно показать, что для $c > 0$

$$c\left[\frac{1}{c}\int\limits_0^c F_{n_i}(y)\,dy - \frac{1}{c}\int\limits_{-c}^{0} F_{n_i}(y)\,dy\right] = \frac{1}{\pi}\int\limits_{-\infty}^{\infty} \frac{1-\cos ct}{t^2}\varphi_{n_i}(t)\,dt. \tag{5.4.4}$$

Следует отметить, что если в первом интеграле слева y рассматривается как случайная величина непрерывного типа, имеющая ф. п. в., равную $\frac{1}{c}$ в интервале $(0, c)$, то $F_{n_i}(y)$ будет случайной величиной, зависящей от параметра n_i. Так как последовательность $F_{n_1}(x)$, $F_{n_2}(x)$, ... сходится к $F(x)$ в каждой точке непрерывности последней и все значения этих функций заключены в интервале $[0, 1]$, мы имеем

$$\lim_{i\to\infty} \frac{1}{c}\int_0^c F_{n_i}(y)\,dy = \frac{1}{c}\int_0^c F(y)\,dy.$$

Точно так же

$$\lim_{i\to\infty} \frac{1}{c}\int_{-c}^0 F_{n_i}(y)\,dy = \frac{1}{c}\int_{-c}^0 F(y)\,dy.$$

Аналогичным образом можно оправдать и переход к пределу при $i\to\infty$ под знаком интеграла в правой части (5.4.4). В результате из (5.4.4) вытекает

$$c\left[\frac{1}{c}\int_0^c F(y)\,dy - \frac{1}{c}\int_c^0 F(y)\,dy\right] = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos ct}{t^2}\varphi(t)\,dt. \quad (5.4.5)$$

Полагая $t = \frac{u}{c}$ и деля обе части (5.4.5) на c, мы получаем

$$\frac{1}{c}\int_0^c F(y)\,dy - \frac{1}{c}\int_{-c}^0 F(y)\,dy = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos u}{u^2}\varphi\left(\frac{u}{c}\right)du. \quad (5.4.6)$$

Устремим теперь c к бесконечности. Поскольку $F(y)$ — неубывающая функция, предел левой части (5.4.6) равен $F(+\infty) - F(-\infty)$. Далее, так как $\varphi(t)$ непрерывна в точке $t = 0$, мы имеем $\lim\limits_{c\to\infty}\varphi\left(\frac{u}{c}\right) = \varphi(0)$ при всяком u. Но последовательность $\varphi_1(t)$, $\varphi_2(t)$, ... всюду сходится к $\varphi(t)$. В частности, $\lim\limits_{n\to\infty}\varphi_n(0) = \varphi(0)$. Замечая, что при любом n $\varphi_n(0) = 1$, мы находим $\varphi(0) = 1$. В итоге переход к пределу при $c\to\infty$ в формуле (5.4.6) дает нам соотношение

$$F(+\infty) - F(-\infty) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{1-\cos u}{u^2}\,du = 1. \quad (5.4.7)$$

Так как функция $F(x)$ неотрицательна, неубывающая и не может превосходить единицу, мы должны заключить, что $F(+\infty) = 1$ и $F(-\infty) = 0$. Следовательно, $F(x)$ (предел последовательности $F_{n_1}(x)$, $F_{n_2}(x)$, ...) есть к. ф. р., а $\varphi(t)$ (предел последовательности $\varphi_{n_1}(t)$, $\varphi_{n_2}(t)$, ...) есть характеристическая функция этой к. ф. р. С другой стороны, пусть какая-либо другая подпоследовательность последовательности $F_1(x)$, $F_2(x)$, ... сходится к некоторой неубывающей функции, которую мы обозначим через $F^*(x)$, в каждой ее точке непрерывности. Тогда, повторяя предыдущие рассуждения, можно показать, что $F^*(x)$ есть к. ф. р., а $\varphi(t)$ есть ее характеристическая функция. Вследствие того, что $F(x)$ и $F^*(x)$ оказываются к. ф. р. с одинаковыми характеристическими функциями, по теореме **5.1.3** они обязаны совпадать, $F(x) \equiv F^*(x)$. Это означает, что каждая сходящаяся подпоследовательность последовательности $F_1(x)$, $F_2(x)$, ... сходится к к. ф. р. $F(x)$, что, разумеется, эквивалентно сходимости к $F(x)$ самой последовательности $F_1(x)$, $F_2(x)$, ... Последние аргументы завершают доказательство теоремы **5.4.1.**

Приведем следствие теоремы **5.4.1**, представляющее интерес при решении задач, в которых требуется установить сходимость по вероятности к константе.

5.4.1а. *Условие, необходимое и достаточное для сходимости последовательности случайных величин* x_1, x_2, ... , *имеющих соответственно характеристические функции* $\varphi_1(t)$, $\varphi_2(t)$, ... , *к константе* c, *заключается в выполнении предельного соотношения* $\lim\limits_{n\to\infty} \varphi_n(t) = e^{ict}$.

Нетрудно получить k-мерные аналоги теорем **5.4.1** и **5.4.1а.** Их формулировка и доказательство предоставляются читателю.

5.5. Определение функций распределения моментами

(а) Определение к. ф. р. последовательностью моментов. В § 5.1 было показано, что моменты функций распределения, если они существуют, можно найти путем дифференцирования характеристических функций; формула (5.1.10) выражает связь между характеристическими функциями и теми моментами, которые действительно существуют. Однако иногда моменты к. ф. р. случайной величины могут быть более эффективно найдены методами, отличными от дифференцирования характеристической функции. Имеется ряд ситуаций, где последовательности моментов случайных величин и функций случайных величин легко определяются тем или иным способом. С этими ситуациями мы встретимся в дальнейших главах. Но основной вопрос заключается в следующем: при каких условиях последовательность моментов μ_1', μ_2', ... случайной величины x единственным образом определяет к. ф. р. этой величины? Ниже формулируется полезный достаточный критерий, принадлежащий Крамеру (1946).

5.5.1. *Пусть* $F(x)$ — *к. ф. р. с моментами* μ_r', $r = 0, 1, 2, \ldots$, *каждый из которых конечен. Если при некотором* $c > 0$ *ряд* $\sum\limits_{r=0}^{\infty} \frac{\mu_r'}{r!} c^r$ *сходится абсолютно, то* $F(x)$ *является единственной к. ф. р., обладающей указанными моментами.*

Исходя из определения (5.1.1) характеристической функции, соответствующей $F(x)$, мы подобно формуле (5.1.7) можем написать

$$\varphi(t+u)=\int_{-\infty}^{\infty}\left[1+\sum_{r=1}^{n-1}\frac{(\mathrm{i}u)^r}{r!}x^r+\right.$$
$$\left.+\frac{(\mathrm{i}u)^n}{n!}x^n(\cos u'x+\mathrm{i}\sin u''x)\right]e^{\mathrm{i}tx}\,dF(x), \qquad (5.5.1)$$

где u' и u'' — (вещественные) числа, лежащие в интервале $(0, u)$. Если n-й момент существует, можно теперь воспользоваться (5.1.4) и получить

$$\varphi(t+u)=\sum_{r=0}^{n-1}\frac{u^r}{r!}\varphi^{(r)}(t)+\frac{\nu'_n u^n}{n!}q, \qquad (5.5.2)$$

где q — комплексная функция, причем $|q|\leqslant 1$, и ν'_n — n-й абсолютный момент величины x. Пусть n четно. Тогда $\nu'_n=\mu'_n$, и из предположения нашей теоремы вытекает, что $\frac{\nu'_n u^n}{n!}q\to 0$, если $|u|<c$ и $n\to\infty$. Рассмотрим, далее, остаточный член при нечетном n. Заметим, что для любого вещественного λ

$$\mathcal{E}\left[\lambda|x|^{\frac{n-1}{2}}+|x|^{\frac{n+1}{2}}\right]^2=\lambda^2\nu'_{n-1}+2\lambda\nu'_n+\nu'_{n+1}\geqslant 0, \qquad (5.5.3)$$

т. е. мы должны иметь

$$\nu'^2_n\leqslant\nu'_{n-1}\nu'_{n+1}. \qquad (5.5.4)$$

Но в таком случае можно написать

$$\frac{\nu'_n}{n!}u^n\leqslant\left[\left(\frac{\nu'_{n-1}}{(n-1)!}u^{n-1}\right)\left(\frac{\nu'_{n+1}}{(n+1)!}u^{n+1}\right)\left(\frac{n+1}{n}\right)\right]^{\frac{1}{2}}. \qquad (5.5.5)$$

Так как $\nu'_{n-1}=\mu'_{n-1}$ и $\nu'_{n+1}=\mu'_{n+1}$, правая часть этого неравенства стремится к нулю, когда $n\to\infty$. Тем самым стремится к нулю и остаточный член в формуле (5.5.2), когда $n\to\infty$ (n четно). Следовательно,

$$\varphi(t+u)=1+\sum_{r=1}^{\infty}\frac{u^r}{r!}\varphi^{(r)}(t), \qquad (5.5.6)$$

причем ряд сходится по крайней мере для $|u|<c$. Полагая $t=0$ и используя (5.1.5), мы находим соотношение

$$\varphi(u)=1+\sum_{r=1}^{\infty}\frac{\mu'_r}{r!}(\mathrm{i}u)^r, \qquad (5.5.7)$$

которое означает, что для значений $|u|<c$ функция $\varphi(u)$ единственным образом определяется моментами μ'_r. С помощью процесса ана-

литического продолжения можно показать, что формула (5.5.7) имеет место для *всех* значений u. В самом деле, например, в точке $u = \pm\frac{1}{2}c$ существуют и могут быть вычислены из (5.5.7) все производные функции $\varphi(u)$. Тогда при $|u| < c$ функции $\varphi\left(\frac{1}{2}c + u\right)$ и $\varphi\left(-\frac{1}{2}c + u\right)$ можно единственным образом представить в виде ряда (5.5.6), где t соответственно равно $\frac{1}{2}c$ и $-\frac{1}{2}c$. Последнее равносильно утверждению, что (5.5.7) справедливо при $|u| < \frac{3}{2}c$. Последовательно продолжая эту процедуру, мы, очевидно, придем к заключению, что (5.5.7) имеет место для всех u. Таким образом, в условиях нашей теоремы характеристическая функция единственным образом определяется всей последовательностью моментов, а характеристическая функция в свою очередь единственным образом определяет к. ф. р. $F(x)$. Теорема **5.5.1** доказана.

Приводимое ниже следствие теоремы **5.5.1** представляет интерес при решении вопроса об определении к. ф. р., заданных на конечных интервалах, последовательностью моментов соответствующих распределений.

5.5.1. (а) *Если x — ограниченная случайная величина, то ее к. ф. р. $F(x)$ единственным образом определяется своими моментами μ'_r, $r = 0, 1, 2, \ldots$*

Если x — ограниченная величина, то найдутся такие конечные числа a и b, $a < b$, что $F(a) = 0$ и $F(b) = 1$. Обозначив через M наибольшее из чисел $|a|$ и $|b|$, мы имеем

$$|\mu'_r| \leqslant \int_a^b |x|^r dF(x) = \nu'_r \leqslant M^r \tag{5.5.8}$$

и

$$\left|\sum_{r=1}^{\infty} \frac{\mu'_r c^r}{r!}\right| \leqslant \sum_{r=0}^{\infty} \frac{|\mu'_r|\, c^r}{r!} \leqslant \sum_{r=0}^{\infty} \frac{\nu'_r c^r}{r!} \leqslant \sum_{r=0}^{\infty} \frac{|Mc|^r}{r!} = e^{|Mc|}. \tag{5.5.9}$$

Поскольку величина справа в (5.5.9) конечна при всех значениях c, достаточное условие теоремы **5.5.1** оказывается выполненным, что доказывает нашу теорему.

Обобщая теоремы **5.5.1** и **5.5.1а**, можно указать условия, при которых к. ф. р. k-мерной случайной величины $(x_1, \ldots, x_k)$ единственным образом определяется моментами $\mu'_{r_1, \ldots, r_k}$. Формулировка и доказательство этих обобщений предоставляются читателю.

(b) Определение предела последовательности к. ф. р. моментами. Допустим, что $(x_1, x_2, \ldots)$ есть последовательность случайных величин, у которой заданы последовательности моментов каждой компоненты. Существует ли простой критерий, основанный на этих моментах, для того, чтобы решить вопрос о сходимости по распределению последовательности $(x_1, x_2, \ldots)$ к некоторой случайной величине x? Простая теорема Кендалла и Рао (1950), касающаяся сходимости по распределению к некоторым к. ф. р. подпоследовательностей последовательности $(x_1, x_2, \ldots)$ и использующая только моменты второго порядка, может быть сформулирована следующим образом:

5.5.2. *Пусть $\mu_2'(x_n)$ — второй момент n-й компоненты последовательности случайных величин $(x_1, x_2, \ldots)$. Если для $n = 1, 2, \ldots$*

$$\mu_2'(x_n) < K < \infty, \tag{5.5.10}$$

то всегда существует подпоследовательность данной последовательности, которая сходится по распределению.

Докажем эту теорему. Пусть $F_n(x)$ означает к. ф. р. величины x_n. Так как для любого $x_1 > 0$

$$K > \mu_2'(x_n) = \int_{-\infty}^{\infty} x^2\,dF_n(x) \geqslant x_1^2 \int_{-\infty}^{-x_1} dF_n(x) + x_1^2 \int_{x_1}^{\infty} dF_n(x), \tag{5.5.11}$$

мы можем записать

$$\frac{K}{x_1^2} > F_n(-x_1) + 1 - F_n(x_1), \qquad n = 1, 2, \ldots \tag{5.5.12}$$

Следовательно, для данного $\varepsilon > 0$ можно выбрать такое $x_1 > 0$, что $1 - [F_n(x) - F_n(-x)] < \varepsilon$ для всех n, если только $x > x_1$. Далее, последовательность к. ф. р. $F_1(x), F_2(x), \ldots$ всегда содержит подпоследовательность, которая сходится к некоторой неубывающей функции $G(x)$ во всех ее точках непрерывности [см., например, Крамер, (1946)]. Ясно, что при $x > x_1$ мы имеем $1 - [G(x) - G(-x)] < \varepsilon$, откуда $G(-\infty) = 0$ и $G(+\infty) = 1$. Таким образом, $G(x)$ сама есть к. ф. р., т. е. действительно некоторая подпоследовательность последовательности $(x_1, x_2, \ldots)$ сходится по распределению к случайной величине.

Предположим теперь, что для каждой компоненты x_n последовательности случайных величин $(x_1, x_2, \ldots)$ мы имеем полную последовательность моментов $\mu_r'(x_n) = \mu_{r,n}'$, $r = 0, 1, 2, \ldots$, причем все эти моменты конечны и $\lim\limits_{n\to\infty} \mu_{r,n}' = \mu_r'$, $r = 0, 1, 2, \ldots$, где последовательность пределов μ_r', $r = 0, 1, 2, \ldots$, единственным образом определяет к. ф. р. $F(x)$. Спрашивается, какие условия будут гарантировать сходимость по распределению последовательности $(x_1, x_2, \ldots)$ к случайной величине x, имеющей к. ф. р., тождественную с $F(x)$?

Ответ на этот вопрос дается теоремой, принадлежащей Кендаллу и Рао (1950).

5.5.3. *Пусть каждая компонента x_n последовательности случайных величин $(x_1, x_2, \ldots)$ имеет конечный момент r-го порядка $\mu_{r,n}'$, $r = 0, 1, 2, \ldots$, и $\lim\limits_{n\to\infty} \mu_{r,n}' = \mu_r'$, где все эти числа также* *конечны. Тогда, если* $(x_1, x_2, \ldots)$ *сходится по распределению к* $F(x)$, $\mu'_0, \mu'_1, \ldots$ *представляет собой последовательность моментов* $F(x)$. *Обратно, если эта последовательность моментов единственным образом определяет некоторую к. ф. р.* $F(x)$, *последняя является предельной для* $(x_1, x_2, \ldots)$ *в смысле сходимости по распределению.*

Чтобы установить теорему **5.5.3**, мы сначала допустим, что последовательность $(x_1, x_2, \ldots)$ сходится по распределению к некоторой к. ф. р. $F(x)$, и покажем, что $\mu'_r = \lim\limits_{n\to\infty} \mu'_{r,n}$, $r = 1, 2, \ldots$, являются моментами $F(x)$. Другими словами, мы должны доказать, что если $F_n(x)$ — к. ф. р. величины x_n, $n = 1, 2, \ldots$, то

$$\lim_{n\to\infty} \left| \int_{-\infty}^{\infty} x^r\, dF_n(x) - \int_{-\infty}^{\infty} x^r\, dF(x) \right| = 0 \qquad \text{для } r = 1, 2, \ldots \tag{5.5.13}$$

Заметим, что для любого $K > 0$

$$\left| \int_{-\infty}^{\infty} x^r\, dF_n(x) - \int_{-\infty}^{\infty} x^r\, dF(x) \right| \leqslant A_1 + A_2 + A_3,$$

где

$$\begin{aligned} A_1 &= \left| \int_{-K}^{K} x^r\, dF_n(x) - \int_{-K}^{K} x^r\, dF(x) \right|, \\ A_2 &= \left| \int_{E_K} x^r\, dF_n(x) \right|, \\ A_3 &= \left| \int_{E_K} x^r\, dF(x) \right|, \end{aligned} \tag{5.5.14}$$

и E_K есть множество значений x, удовлетворяющих неравенству $|x| > K$. На основании неравенства Шварца

$$A_2^2 \leqslant \int_{E_K} x^{2r}\, dF_n(x) \int_{E_K} dF_n(x), \tag{5.5.15}$$

причем оба интеграла в правой части неотрицательны. Так как моменты $\mu'_{2r,n}$ при любом r имеют конечный предел, когда $n \to \infty$, первый интеграл является ограниченным относительно n и K. Но $F_n(x) \to F(x)$, когда $n \to \infty$, и, следовательно, второй интеграл справа, а вместе с ним и A_2, можно сделать сколь угодно малым одновременно для всех n, выбирая достаточно большое K.

Учитывая конечность μ'_{2r} и выбирая достаточно большое K, можно сделать сколь угодно малой и величину A_3.

Поскольку $F_n(x) \to F(x)$, когда $n \to \infty$, и как $\mu'_{r,n}$, так и μ'_r конечны, очевидно, что, каково бы ни было K, величина A_1 может быть сделана произвольно малой, если только взять достаточно большое n.

Из всего этого мы заключаем, что формула (5.5.13) имеет место и $\mu'_1, \mu'_2, \ldots$ есть последовательность моментов $F(x)$ — к. ф. р., к которой сходится по распределению последовательность $(x_1, x_2, \ldots)$.

Теперь рассмотрим обратное утверждение. Предположим, что μ_1', μ_2', ... единственным образом определяют некоторую к. ф. р. $F(x)$. Мы должны показать, что последовательность $F_1(x)$, $F_2(x)$, ... сходится к $F(x)$ во всех ее точках непрерывности. Из теоремы **5.5.2** мы знаем, что всякая сходящаяся подпоследовательность данной последовательности $F_1(x)$, $F_2(x)$, ... сходится к некоторой к. ф. р. и по только что доказанному предельная к. ф. р. должна обладать той же самой последовательностью моментов, именно μ_1', μ_2', ... Но предполагается, что эта последовательность моментов определяет к. ф. р. единственным образом. Поэтому всевозможные предельные к. ф. р. тождественны с $F(x)$, т. е. с к. ф. р., имеющей моменты μ_1', μ_2', ...

Наконец, следует отметить, что условие конечности моментов $\mu_{r,n}'$ при всех n и r можно, не нарушая справедливости теоремы **5.5.3**, заменить условием конечности $\mu_{r,n}'$ при каждом r и всех n, бо́льших некоторого целого числа n^*, возможно зависящего от r.

ЗАДАЧИ

5.1. Показать, что если случайная величина x имеет ф. п. в. (или ф. в.), которая симметрична относительно начала координат, то ее характеристическая функция $\varphi(t)$ принимает только вещественные значения.

5.2. Случайная величина x имеет характеристическую функцию

$$\varphi(t)=\frac{1}{1+t^2}.$$

Показать, что ф. п. в. величины x при любом значении $x \in R_1$ равна $\frac{1}{2}e^{-|x|}$.

5.3. Характеристическая функция случайной величины x равна

$$\frac{e^{it}(1-e^{nit})}{n(1-e^{it})}.$$

Показать, что x является дискретной случайной величиной с ф. п. в. $p(x)=\frac{1}{n}$ для $x=1, \ldots, n$.

5.4. Показать, что если $e^{-\frac{1}{2}t^2}$ — характеристическая функция случайной величины x, то ф. п. в. этой величины равна $(2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}$.

5.5. Доказать теорему **5.2.1**.

5.6. Показать, что если $x_1, \ldots, x_k$ — независимые случайные величины, к. ф. р. которых равна одной и той же функции $F(x)$, то характеристическая функция суммы $x_1+\ldots+x_k$ есть $[\varphi(t)]^k$, где $\varphi(t)$ — характеристическая функция любой из величин x_i, $i=1, \ldots, k$.

5.7. Говорят, что случайная величина x имеет *распределение Коши* (1853), если ее ф. п. в. равна

$$f(x)=\frac{k}{\pi[k^2+(x-\mu)^2]}, \qquad -\infty<x<\infty,$$

где k и μ — вещественные константы и $k>0$. Показать, что характеристическая функция величины x дается соотношением

$$\varphi(t)=e^{i\mu t-k|t|},$$

и, следовательно, если $x_1, \ldots, x_n$ — независимые случайные величины, каждая из которых имеет указанное распределение Коши, случайная величина $\frac{1}{n}(x_1+\ldots+x_n)$ также имеет то же самое распределение Коши.

5.8. Используя характеристические функции, показать, что если $x_1, \ldots$ $\ldots, x_n$ — независимые случайные величины в R_1, имеющие одинаковые ф. п. в., именно $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, то ту же самую ф. п. в. имеет и случайная величина $\frac{1}{\sqrt{n}}(x_1+\ldots+x_n)$.

5.9. Пусть x — случайная величина, означающая число бросаний игральной кости, которые должны быть сделаны, чтобы первый раз получить одно очко. Определить производящую функцию вероятностей $\theta(t)$ этой величины. Исходя из $\theta(t)$, найти среднее и дисперсию случайной величины x.

5.10. Пусть x — случайная величина, означающая число бросаний «правильной» монеты, необходимых для k выпадений гéрба. Показать, что если $F(x, k)$ — к. ф. р. величины x, $F(x, k)$ является воспроизводящей по k.

5.11. Показать, что если x — случайная величина, означающая число пик в наборе из 13 обыкновенных игральных карт, то производящая функция вероятностей $\theta(t)$ величины x дается коэффициентом при v^{13} в разложении выражения $\frac{(1+tv)^{13}(1+v)^{39}}{C_{52}^{13}}$ по степеням v. Исходя из $\theta(t)$, найти среднее и дисперсию величины x.

5.12. Случайная величина x имеет моменты

$$\mu_r' = \frac{k}{k+r}, \qquad r = 1, 2, \ldots,$$

где $k > 0$. Показать, что ф. п. в. величины x дается выражением

$$f(x) = \begin{cases} x^{k-1} & \text{для } 0 \leqslant x \leqslant 1, \\ 0 & \text{в противном случае} \end{cases}$$

и что это распределение определяется единственным образом.

5.13. Показать, что если x — случайная величина, имеющая моменты

$$\mu_r' = \frac{(k+r)!}{k!},$$

причем k — положительное целое число, то ф. п. в. этой величины равна

$$f(x) = \begin{cases} 0, & x \leqslant 0, \\ \frac{x^k}{k!}e^{-x}, & x > 0, \end{cases}$$

и это распределение определяется единственным образом.

5.14. Пусть x — случайная величина, означающая общее число очков, которые выпадают при одновременном бросании n «правильных» игральных костей. Показать, что производящая функция моментов $\psi(t)$ величины x равна $\left(e^t \cdot \frac{1}{6}\right)^n (1-e^t)^n (1-e^{6t})^n$ и что ее распределение является воспроизводящим по n.

5.15. В последовательности $(x_1, x_2, \ldots)$ случайная величина x_n имеет ф. п. в., равную $n(1-x)^{n-1}$, если x принадлежит интервалу $(0, 1)$ и равную нулю в противном случае. Пусть $y_n = nx_n$. Используя характеристические функции, показать, что последовательность $(y_1, y_2, \ldots)$ сходится по распределению к случайной величине y с ф. п. в.

$$f(y) = \begin{cases} e^{-y}, & 0 < y < \infty, \\ 0 & \text{в противном случае.} \end{cases}$$

5.16. В последовательности случайных величин $(x_1, x_2, \ldots)$ величина x_n имеет характеристическую функцию

$$\varphi_n(t) = \frac{\sin nt}{nt}.$$

Показать, что ф. п. в. величины x_n равна $\frac{1}{2n}$ в интервале $(-n, n)$ и 0 в противном случае и, следовательно, хотя последовательность характеристических функций и сходится к некоторому пределу $\varphi(t)$, последовательность к. ф. р. не сходится к к. ф. р. Какое условие теоремы **5.4.1** здесь нарушено? [Крамер (1946).]

5.17. Используя характеристические функции, показать, что если $x_1, \ldots, x_k$ — независимые дискретные случайные величины, причем ф. в. величины x_i равна

$$p_i(x) = \frac{\mu_i^x e^{-\mu_i}}{x!}, \qquad x = 0, 1, 2, \ldots, \mu_i > 0,$$

то ф. в. случайной величины $z = x_1 + \ldots + x_n$ дается формулой

$$p(z) = \frac{\left(\sum_{i=1}^{k} \mu_i\right)^z e^{-\sum_{i=1}^{k} \mu_i}}{z!}, \qquad z = 0, 1, 2, \ldots$$

5.18. Последовательность неотрицательных случайных величин $(x_1, x_2, \ldots)$ такова, что r-й момент величины x_n дается формулой

$$\mu'_{r,n} = \frac{r!\, n^r (n-r-1)!}{(n-1)!}, \qquad n > r = 1, 2, \ldots$$

Показать, что данная последовательность случайных величин сходится по распределению к случайной величине x, имеющей ф. п. в.

$$f(x) = \begin{cases} 0, & x \leqslant 0, \\ xe^{-x}, & x > 0. \end{cases}$$

Показать, что это предельное распределение определяется единственным образом.

5.19. Рассмотрим случайный ветвящийся процесс, где некоторый объект может порождать другие объекты, каждый из которых в свою очередь может порождать новые объекты и т. д., подобно тому как это имеет место для самцов (или самок) в последовательно сменяющих друг друга поколениях или для нейтронов в процессе ядерного деления.

Пусть x_1 — случайная величина, обозначающая число объектов, которые производит некоторый начальный объект (нулевого поколения). Пусть x_2 — случайная величина, обозначающая число объектов, которые производит каждый из объектов первого поколения. Вообще пусть x_{n+1} — случайная величина, обозначающая число объектов, которые производит каждый из объектов n-го поколения. Предположим, что ф. в. величин $x_1, x_2, \ldots$ тождественны между собой и равны $p(x)$, $x=0, 1, 2, \ldots$ Введем случайную величину y_n — общее число объектов, произведенных в n-м поколении, и обозначим через $\theta_n(t)$ производящую функцию вероятностей этой величины. $\left[\text{Заметим, что } \theta_0(t)=t \text{ и } \theta_1(t)=\sum_{x=0}^{\infty} t^x p(x).\right]$ Считая все ряды, представляющие производящие функции, сходящимися, показать, что

$$\theta_{n+1}(t)=\theta_1[\theta_n(t)].$$

Показать, что если μ — среднее, а σ^2 — дисперсия величины y_1, то

$$\mathscr{E}(y_n)=\mu^n,$$
$$\sigma^2(y_n)=n\sigma^2,\ \mu=1,$$
$$\sigma^2(y_n)=\sigma^2\mu^{n-1}(\mu^n-1)/(\mu-1),\ \mu\neq 1.$$

[Харрис (1948).]

5.20. (*Продолжение*). Пусть случайные величины $x_1, \ldots, x_n$ имеют различные ф. в., например, соответственно $p_1(x), \ldots, p_n(x)$. Обозначим через $\theta_1^*(t), \ldots, \theta_n^*(t)$ соответствующие производящие функции вероятностей. Показать, что производящая функция вероятностей $\theta_n(t)$ величины y_n дается формулой

$$\theta_n(t)=\theta_1^*[\theta_2^*\{\ldots\theta_n^*(t)\ldots\}],$$

причем

$$\mathscr{E}(y_n)=\mu_1\mu_2\ldots\mu_n,$$

где $\mu_1, \ldots, \mu_n$ суть соответственно средние величин $x_1, \ldots, x_n$.

5.21. Установить формулу (5.4.4).

5.22. Пусть x — случайная величина с ф. в. $p(x)$ и к. ф. р. $F(x)$. Пусть выборочным пространством этой величины является множество неотрицательных целых чисел. Обозначим через $\theta^*(t)=\sum_{x=0}^{\infty} t^x F(x)$ *производящую функцию* для $F(x)$. Показать, что если $\theta(t)$ — производящая функция вероятностей величины x, то при $|t|<1$

$$\theta^*(t)=\frac{\theta(t)}{(1-t)}.$$

5.23. Показать, что если $x_1, \ldots, x_k$ — независимые случайные величины, выборочные пространства которых суть множества неотрицательных целых чисел и производящие функции вероятностей которых равны соответственно $\theta_1(t), \ldots, \theta_k(t)$, то производящая функция вероятностей суммы $x_1+\ldots+x_k$ есть $\theta_1(t)\ldots\theta_k(t)$.

Г Л А В А 6

НЕКОТОРЫЕ СПЕЦИАЛЬНЫЕ ДИСКРЕТНЫЕ РАСПРЕДЕЛЕНИЯ

Цель настоящей главы — представить несколько наиболее важных в математической статистике распределений вероятностей и при этом не только сообщить основную информацию о самих распределениях, но и проиллюстрировать ряд понятий, принципов и методов из предшествующих глав с большей полнотой, чем это было сделано до сих пор посредством примеров. Рассмотренные в данной главе распределения и некоторые их основные свойства будут также использованы во многих местах на протяжении оставшейся части книги. Кроме того, их обсуждение покажет, что, вопреки общим принципам и методам, введенным в более ранних главах, исследование специальных распределений часто требует особых методов и аппарата.

6.1. Гипергеометрическое распределение

(а) Случай одномерной величины. Функция вероятности $p(x)$, приведенная в конце § 2.3 (а), является простым случаем *гипергеометрического распределения*. Вообще пусть мы имеем совокупность элементов Π, причем каждый элемент принадлежит либо классу C, либо классу $\bar{C}$. Обозначим через $N \cdot p$ число элементов, принадлежащих C, и через $N \cdot q$ — число элементов, принадлежащих $\bar{C}$; $p+q=1$. Предположим, что из Π извлекается $n (\leqslant N)$ элементов, и пусть x есть случайная величина, означающая число тех из них, которые принадлежат C. Мы хотим найти ф. в. величины x. Всего имеется C_N^n различных множеств из n элементов совокупности Π и среди них $C_{N \cdot p}^x \cdot C_{N \cdot q}^{n-x}$ будут насчитывать точно x элементов, принадлежащих C.

Следует отметить, что в данных условиях основное выборочное пространство R, рассмотренное в главе 1, состоит из выборочных точек, причем каждая выборочная точка является множеством, содержащим n элементов из Π. Событие, описываемое неравенством $x \leqslant x'$, включает все выборочные точки в R, для которых насчитывается не более чем x' элементов, принадлежащих C.

Назначая всем выборочным точкам в R равные вероятности, мы получаем следующее выражение для ф. в. величины x:

$$p(x)=\frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}, \qquad (6.1.1)$$

причем точки сосредоточения массы вероятности для x (точки выборочного пространства R_1 этой величины) являются целыми числами, удовлетворяющими как неравенству $0 \leqslant x \leqslant Np$, так и неравенству $0 \leqslant n - x \leqslant Nq$.

Функция (6.1.1) называется *гипергеометрической функцией вероятности*. Будет удобно называть распределение, имеющее такую ф. в., *гипергеометрическим распределением* $H(N, n; p)$.

Чтобы убедиться в том, что сумма значений $p(x)$ по всей указанной области изменения x равна единице, рассмотрим тождество

$$(u+v)^{A+B} \equiv (u+v)^A (u+v)^B,$$

где A и B — целые положительные числа. Раскрывая обе части этого выражения, мы имеем

$$\sum_{r=0}^{A+B} \binom{A+B}{r} u^r v^{A+B-r} \equiv$$

$$\equiv \left[\sum_{s=0}^{A} \binom{A}{s} u^s v^{A-s} \right] \left[\sum_{r=s}^{B+s} \binom{B}{r-s} u^{r-s} v^{B-r+s} \right].$$

Так как это — тождество по u и v, коэффициент при $u^r v^{A+B-r}$ в левой части должен совпадать с коэффициентом при $u^r v^{A+B-r}$ в правой части. Поэтому

$$\binom{A+B}{r} = \sum_s \binom{A}{s} \binom{B}{r-s}, \tag{6.1.2}$$

где $\sum\limits_s$ означает суммирование по всем целым s, удовлетворяющим неравенствам $0 \leqslant s \leqslant A$ и $0 \leqslant r - s \leqslant B$.

Теперь ясно, что

$$\sum_x \binom{Np}{x} \binom{Nq}{n-x} = \binom{N}{n} \tag{6.1.2a}$$

и, следовательно, сумма значений $p(x)$, распространенная на все выборочное пространство величины x, равна единице.

Это гипергеометрическое распределение доставляет нам пример распределения, для которого характеристическая функция как инструмент для нахождения моментов фактически ничего не стоит. Но мы можем определить факториальные моменты $\mu'_{[r]}$ следующим довольно естественным способом.

Сначала заметим, что

$$\sum_x x^{[r]} \binom{Np}{x} \binom{Nq}{n-x} = (Np)^{[r]} \cdot \sum_x \binom{Np-r}{x-r} \binom{Nq}{n-x}, \tag{6.1.3}$$

причем область изменения x совпадает с указанной для формулы (6.1.2а).

На основании (6.1.2) сумма в правой части равна $\binom{N-r}{n-r}$. Если мы разделим (6.1.3) на $\binom{N}{n}$, выражение в левой части определит нам $\mu'_{[r]}$, т. е.

$$\mu'_{[r]}=\frac{(N\cdot p)^{[r]}\cdot\binom{N-r}{n-r}}{\binom{N}{n}}. \tag{6.1.4}$$

В частности, мы имеем

$$\mu'_{[1]}=np, \quad \mu'_{[2]}=\frac{np\,(Np-1)\,(Np-2)}{N\,(N-1)}$$

и, кроме того, выражения для среднего и дисперсии

$$\mu(x)=np, \quad \sigma^2(x)=\frac{N-r}{N-1}\,npq. \tag{6.1.5}$$

Гипергеометрическое распределение является одним из наиболее важных распределений в элементарной теории вероятностей. Оно выступает как основное в связи с вычислением вероятностей в задачах, возникающих в большинстве карточных игр, при извлечении выборок из партий изделий массового производства, среди которых есть дефектные, и из других конечных совокупностей и т. д.

Ф. в. и к. ф. р. гипергеометрического распределения (6.1.1) были табулированы в широких пределах Либерманом и Оуэном (1961).

(b) Случай k-мерной величины. Распределение вероятностей (6.1.1) может быть обобщено на случай k величин. Пример в конце § 2.5(а) доставляет частный случай двумерного гипергеометрического распределения. Вообще говоря, пусть Π есть совокупность N элементов, причем $N\cdot p_i$ элементов принадлежит классу C_i, $i=1, 2, \ldots, k+1$; $p_1+p_2+\ldots+p_{k+1}=1$. Это означает, что классы $C_1, \ldots, C_{k+1}$ взаимно исключающие, т. е. каждый элемент из Π принадлежит одному и только одному из этих классов. Далее, предположим, что из Π извлекается n элементов. Пусть $x_1, x_2, \ldots, x_{k+1}$ — случайные величины, обозначающие числа извлеченных элементов, принадлежащих соответственно $C_1, C_2, \ldots, C_{k+1}$. Эти случайные величины линейно зависимы, так как их сумма равна n. Мы можем считать первые k величин $x_1, \ldots, x_k$ линейно независимыми и $(k+1)$-ю случайную величину x_{k+1} писать в виде $n-x_1-x_2-\ldots-x_k$.

Имеется всего $\binom{N}{n}$ различных n-элементных множеств, которые могли бы быть образованы из Π, иначе, C_N^n выборочных точек в основном выборочном пространстве R^N. Из этого числа выборочных точек

$\binom{Np_1}{x} \ldots \binom{Np_{k+1}}{x_{k+1}}$ насчитывают точно x_1 элементов, принадлежащих $C_1, \ldots, x_{k+1}$ элементов, принадлежащих C_{k+1}. Назначая всем этим $\binom{N}{n}$ выборочным точкам равные вероятности, мы получаем следующую ф. в. величины $(x_1, \ldots, x_k)$:

$$p(x_1, \ldots, x_k) = \frac{\binom{Np_1}{x_1} \ldots \binom{Np_{k+1}}{x_{k+1}}}{\binom{N}{n}}, \tag{6.1.6}$$

где $x_{k+1} = n - x_1 - \ldots - x_k$. Это — *$k$-мерная гипергеометрическая функция вероятности.* Точками сосредоточения массы (выборочными точками) этого распределения являются те точки пространства R_k с неотрицательными целыми координатами, которые находятся в *симплексе* $x_i \geqslant 0$, $i = 1, \ldots, k$, $\sum_{i=1}^{k} x_i \leqslant n$ и удовлетворяют системе неравенств $0 \leqslant x_i \leqslant Np_i$, $i = 1, \ldots, k+1$. Будет удобно называть распределение вероятностей, имеющее ф. в. (6.1.6), *k-мерным гипергеометрическим распределением* $H(N, n; p_1, \ldots, p_k)$.

Средние, дисперсии, ковариации и моменты высшего порядка распределения (6.1.6) могут быть определены через факториальные моменты. Обозначая факториальный момент $\mathcal{E}(x_1^{[r_1]} \ldots x_k^{[r_k]})$ через $\mu'_{[r_1], \ldots, [r_k]}$, мы находим в результате прямого обобщения метода, использованного при выводе (6.1.4),

$$\mu'_{[r_1], \ldots, [r_k]} = \frac{(Np_1)^{[r_1]} \ldots (Np_k)^{[r_k]} \cdot \binom{N - r_1 - \ldots - r_k}{n - r_1 - \ldots - r_k}}{\binom{N}{n}}, \tag{6.1.7}$$

где $0 \leqslant r_i \leqslant Np_i$, $i = 1, \ldots, k$ и $r_1 + \ldots + r_k \leqslant n$. Исходя из формулы (6.1.7), можно проверить, что

$$\mu(x_i) = np_i, \quad \sigma^2(x_i) = np_i(1 - p_i)\frac{N-n}{N-1},$$
$$\operatorname{cov}(x_i, x_j) = -np_ip_j\frac{N-n}{N-1}. \tag{6.1.8}$$

6.2. Биномиальное распределение

Допустим, что некоторая «операция» или «испытание» при своем осуществлении будет иметь один из двух возможных исходов: C или $\bar{C}$. Например, при бросании игральной кости мы можем обозначить через C выпадение одного очка и через $\bar{C}$ — выпадение любого другого количества очков. Предположим, что p есть вероятность исхода C и q — вероятность исхода $\bar{C}$. Тогда $p + q = 1$.

Если «испытание» будет осуществлено n раз, мы получим последовательность, состоящую из n чередующихся исходов C и $\bar{C}$. Всего имеется 2^n различных последовательностей; они образуют основное выборочное пространство R, причем каждая выборочная точка представляется одной из возможных последовательностей. Пусть x — случайная величина, значение которой для любой данной последовательности равно числу исходов C в этой последовательности. Если мы предположим, что все n «испытаний» статистически независимы, то вероятность, приписываемая любой из последовательностей, для которой $x = x'$, будет равна

$$p^{x'} \cdot q^{n-x'}, \tag{6.2.1}$$

как это можно увидеть, заменяя в последовательности C на p и $\bar{C}$ на q и перемножая полученные числа. Всего имеется $\binom{n}{x}$ последовательностей, для которых $x = x'$, и, значит,

$$P(x = x') = \binom{n}{x'} \cdot p^{x'} \cdot q^{n-x'}, \quad x' = 0, 1, \ldots, n. \tag{6.2.2}$$

Итак, x есть дискретная случайная величина, ф. в. $p(x)$ которой дается правой частью формулы (6.2.2). Следует отметить, что $p(x)$ представляет собой общий член в разложении бинома $(q+p)^n$ по степеням p и q. Отсюда сумма значений $p(x)$, взятых из (6.2.2), по всему выборочному пространству случайной величины x равна $(p+q)^n$, что, конечно, есть единица, так как $p+q=1$. В соответствии с этим распределение, имеющее ф. в. вида (6.2.2), называется *биномиальным распределением* и будет обозначаться $\mathrm{Bi}(n, p)$. (Мы используем символ $\mathrm{Bi}(a, b)$ для обозначения биномиального распределения и символ $\mathrm{Be}(a, b)$ для обозначения бета-распределения, которое должно быть введено в главе 7.) Последовательности независимых испытаний, из которых каждое приводит к одному из двух возможных исходов с постоянными вероятностями, не зависящими от номера испытания, часто называются *испытаниями Бернулли* по имени Я. Бернулли (1713), изучавшего их впервые.

Характеристическая функция распределения (6.2.2) равна

$$\varphi(t) = \mathscr{E}(e^{itx}) = \sum_{x=0}^{n} \binom{n}{x} e^{itx} \cdot p^x q^{n-x} = (q + pe^{it})^n, \tag{6.2.3}$$

откуда, прилагая (5.1.5), можно найти моменты μ'_r. В частности, мы имеем

$$\mu'_1 = \frac{\varphi'(0)}{i} = np, \qquad \mu'_2 = \frac{\varphi''(0)}{i^2} = np(q + np), \tag{6.2.4}$$

что дает возможность определить среднее и дисперсию распределения (6.2.2)

$$\mu(x) = np, \quad \sigma^2(x) = npq. \tag{6.2.5}$$

Будем теперь рассматривать n в (6.2.3) как параметр и вместо $\varphi(t)$ писать $\varphi(t;\ n)$. Функция $\varphi(t;\ n)$ удовлетворяет (5.3.7), и потому, если x_1 и x_2 — две независимые случайные величины, имеющие соответственно биномиальные распределения $\mathrm{Bi}(n_1,\ p)$ и $\mathrm{Bi}(n_2,\ p)$, сумма $x = x_1 + x_2$ имеет биномиальное распределение $\mathrm{Bi}(n_1 + n_2,\ p)$. Интуитивно этот результат очевиден. Следовательно,

6.2.1. *Биномиальное распределение* $\mathrm{Bi}(n,\ p)$ *является воспроизводящим по* n.

Необходимо указать на то, что биномиальное распределение (6.2.2) является предельным для гипергеометрического распределения (6.1.1), когда $N \to \infty$. В самом деле, (6.1.1) можно переписать в следующем виде:

$$p(x) = \binom{n}{x} \frac{p\left(p-\frac{1}{N}\right)\ldots\left(p-\frac{x-1}{N}\right)\cdot q\left(q-\frac{1}{N}\right)\ldots\left(q-\frac{n-x-1}{N}\right)}{\left(1-\frac{1}{N}\right)\ldots\left(1-\frac{n-1}{N}\right)} \underset{N\to\infty}{\longrightarrow} \binom{n}{x} p^x q^{n-x}. \tag{6.2.6}$$

Отсюда, полагая x и n фиксированными, мы заключаем, что

6.2.2. *Пределом ф. в. гипергеометрического распределения* $H(N,\ n;\ p)$ *при* $N \to \infty$ *является ф. в. биномиального распределения* $\mathrm{Bi}(n,\ p)$.

Поскольку число точек сосредоточения массы для биномиального и гипергеометрического распределений конечно, из теоремы **6.2.2** вытекает, что если $P_H(x \in E)$ и $P_B(x \in E)$ суть вероятности некоторого события E, вычисленные, исходя соответственно из $H(N,\ n;\ p)$ и $\mathrm{Bi}(n,\ p)$, то $\lim\limits_{N\to\infty} P_H(x \in E) = P_B(x \in E)$. Для последовательностей случайных величин это означает, что если x_N — случайная величина, имеющая гипергеометрическое распределение $H(N,\ n;\ p)$, то последовательность случайных величин $(x_1,\ x_2,\ \ldots)$ сходится по распределению к случайной величине x, имеющей биномиальное распределение $\mathrm{Bi}(n,\ p)$.

Биномиальное распределение является одним из весьма важных распределений в статистике. Как ф. в. (6.2.2), так и к. ф. р. были широко табулированы, причем наиболее обширное табулирование было осуществлено Артиллерийско-технической службой армии США (1952), Вычислительной лабораторией Гарвардского университета (1955), Национальным бюро стандартов (1949) и Ромигом (1953).

6.3. Мультиномиальное распределение

Теперь предположим, что каждая «операция» или «испытание» будет приводить к одному и только одному из взаимно исключающих событий $C_1,\ \ldots,\ C_{k+1}$ и вероятности, соответствующие этим событиям, равны $p_1, \ldots, p_{k+1}$, где $p_1+\ldots+p_{k+1}=1$. Пусть осуществляется n испытаний. Введем $(k+1)$-мерную случайную величину $(x_1, \ldots, x_{k+1})$, компоненты которой $x_1, \ldots, x_{k+1}$ означают числа испытаний, приводящих соответственно к $C_1, \ldots, C_{k+1}$. Компоненты $x_1, \ldots, x_{k+1}$ линейно зависимы, так как $x_1+\ldots+x_{k+1}=n$. Основное выборочное пространство R состоит из $(k+1)^n$ выборочных точек, каждая из которых есть последовательность n выборов из множества $\{C_1, \ldots \ldots, C_{k+1}\}$, причем допускаются повторения. Вероятность, связанная с любой выборочной точкой, представляемой x_1 событиями $C_1, \ldots, x_{k+1}$ событиями C_{k+1}, есть

$$p_1^{x_1} \ldots p_{k+1}^{x_{k+1}}, \tag{6.3.1}$$

а число таких выборочных точек в R, очевидно, равно

$$\frac{n!}{x_1! \ldots x_{k+1}!}. \tag{6.3.2}$$

Отсюда следует, что ф. в. случайной величины $(x_1, \ldots, x_k)$ дается формулой

$$p(x_1, \ldots, x_k)=\frac{n!}{x_1! \ldots x_{k+1}!} p_1^{x_1} \ldots p_{k+1}^{x_{k+1}}, \tag{6.3.3}$$

где необходимо положить $x_{k+1}=n-x_1-\ldots-x_k$ и $p_{k+1}=1-p_1-\ldots-p_k$. Можно заметить, что выражение для $p(x_1, \ldots, x_k)$ представляет собой общий член в разложении мультинома $(p_1+\ldots+p_{k+1})^n$ по степеням $p_1, p_2, \ldots, p_{k+1}$. В соответствии с этим распределение, имеющее ф. в. (6.3.3), называется *k-мерным мультиномиальным распределением*. Будет удобно обозначить его через $M(n; p_1, \ldots, p_k)$. Точки сосредоточения массы для этого распределения, т. е. точки выборочного пространства величины $(x_1, \ldots, x_k)$, совпадают с узлами решетки в R_k, содержащимися в симплексе $x_i \geqslant 0$, $i=1, \ldots, k$, $\sum_{i=1}^{k} x_i \leqslant n$, иначе, с точками этого симплекса, координаты которых суть целые положительные числа или нуль.

Характеристическая функция мультиномиального распределения имеет вид

$$\varphi(t_1, \ldots, t_k)=\sum e^{it_1x_1+\ldots+it_kx_k} \cdot p(x_1, \ldots, x_k)=$$
$$=\sum \frac{n!}{x_1! \ldots x_{k+1}!} (p_1e^{it_1})^{x_1} \ldots (p_ke^{it_k})^{x_k} (p_{k+1})^{x_{k+1}},$$

где суммирование совершается по всем точкам выборочного пространства величины $(x_1, \ldots, x_k)$. Но это просто равно сумме всех членов в разложении мультинома $(p_1e^{it_1}+\ldots+p_ke^{it_k}+p_{k+1})^n$ по степеням $(p_1e^{it_1}), \ldots, (p_ke^{it_k}), p_{k+1}$, т. е.

$$\varphi(t_1, \ldots, t_k)=(p_1e^{it_1}+\ldots+p_ke^{it_k}+p_{k+1})^n. \tag{6.3.4}$$

Прилагая (5.2.3), мы можем найти смешанные моменты величин $x_1, \ldots$ $\ldots, x_k$. В частности,

$$\mu(x_i) = np_i, \qquad \sigma^2(x_i) = np_i(1 - p_i),$$
$$\operatorname{cov}(x_i, x_j) = -np_ip_j. \tag{6.3.5}$$

Если n в (6.3.4) рассматривается как параметр, то можно доказать, что

6.3.1. Мультиномиальное распределение $M(n; p_1, \ldots, p_k)$ является воспроизводящим по n.

6.4. Распределение Пуассона

Предположим, что при фиксированном x в биномиальном распределении (6.2.2) мы устремляем n к бесконечности, а p к нулю таким образом, чтобы для соответствующих значений n и p было $np = \mu$, где μ — некоторая константа. Тогда пределом для (6.2.2) оказывается следующая функция вероятности:

$$p(x) = \frac{\mu^x e^{-\mu}}{x!}. \tag{6.4.1}$$

Это — ф. в. так называемого *распределения Пуассона*, с именем которого связано его первое (1837) исследование. Будет удобно обозначить распределение Пуассона, имеющее ф. в. (6.4.1), через $\mathrm{Po}(\mu)$. [Мы вставляем букву о, т. е. пишем $\mathrm{Po}(\mu)$, чтобы избежать путаницы с обозначением вероятности $P(\mu)$.] Так как при получении (6.4.1) допускается, что n неограниченно возрастает, мы можем при желании положить x равным любому целому числу. Таким образом, 0, 1, 2, ... суть точки сосредоточения массы величины x. Чтобы установить (6.4.1), заметим, что правая часть формулы (6.2.2) может быть записана в виде

$$\frac{\frac{n}{n}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{x-1}{n}\right)}{x!}(np)^x\left(1 - \frac{np}{n}\right)^{n-x}. \tag{6.4.2}$$

Устремляя n к бесконечности, а p к нулю так, чтобы было $np = \mu$, мы действительно находим, что пределом распределения (6.4.2) является распределение (6.4.1). Можно легко проверить, что сумма вероятностей, даваемых (6.4.1) для $x = 0, 1, 2, \ldots$, равна единице.

Характеристическая функция распределения Пуассона есть

$$\varphi(t) = \sum_{x=0}^{\infty} \frac{e^{itx}\mu^x e^{-\mu}}{x!} e^{-\mu(1 - e^{it})}, \tag{6.4.3}$$

откуда мы получаем

$$\mu(x) = \mu, \qquad \sigma^2(x) = \mu. \tag{6.4.4}$$

Считая в выражении (6.4.3) для характеристической функции μ параметром, можно обнаружить, что функция (6.4.3) удовлетворяет (5.3.7) и, следовательно,

6.4.1. *Распределение Пуассона* $\mathrm{Po}(\mu)$ *является воспроизводящим по* μ.

Замечание. Так как гипергеометрическое распределение при больших N является приближением к биномиальному, а биномиальное распределение при больших n и малых p является приближением к распределению Пуассона, можно ожидать, что при некоторых условиях гипергеометрическое распределение должно быть близким к относительно простому распределению Пуассона. Фактически, грубо говоря, это имеет место, если (I) n велико, (II) p мало (и $np=\mu$) и (III) N значительно больше n. Эти условия достаточно хорошо выполняются в таких важных приложениях, как извлечение выборок из партий изделий массового производства. Здесь N — объем партии, n — объем выборки из партии и p — доля дефектных изделий в партии.

Подобно биномиальному распределению, распределение Пуассона является очень важным в приложениях теории вероятностей и статистики. Имеется много задач, связанных не только с распределением чисел дефектных изделий в выборках из партий промышленной продукции, но и, например, с распределением колоний бактерий, приходящихся на единицу объема выращиваемой культуры или жидкости, числа телефонных вызовов, поступивших в единицу времени, числа бомбовых осколков, поражающих цель, на единицу площади и т. д., где появляется распределение Пуассона. Во всех этих задачах мы, по существу, имеем большое число n независимых или почти независимых «испытаний», малую вероятность p «успеха» в любом данном испытании, и нам требуется определить вероятность получения x «успехов» в n испытаниях.

Ф. в. и к. ф. р. распределения Пуассона были в широких пределах табулированы Молина (1942).

6.5. Дискретные распределения времени ожидания

(а) Гипергеометрический случай. Предположим, что совокупность Π состоит из $N\cdot p$ элементов C и $N\cdot q$ элементов $\overline{C}$, где $p>0$, $q>0$ и $p+q=1$. Пусть из Π один за другим извлекаются элементы до тех пор, пока не появится точно k элементов C. Мы хотим определить вероятность того, что нужно будет извлечь точно x элементов, чтобы достигнуть этой цели. Случайные точки нашего основного выборочного пространства R будут представлять собой все возможные последовательности элементов C и $\overline{C}$, которые могли бы быть образованы в результате последовательного извлечения из Π всех N элементов. В R имеется всего

$$\binom{N}{Np} \tag{6.5.1}$$

выборочных точек.

В каждой выборочной точке значение нашей случайной величины x равно числу элементов, которые нужно было извлечь из Π, чтобы получить точно k элементов C. Для любого значения x', удовлетворяющего неравенствам $0 \leqslant k-1 \leqslant x'-1$ и $0 \leqslant Np-k \leqslant N-x'$, число выборочных точек, где $x=x'$, очевидно, должно быть равно

$$\binom{x'-1}{k-1}\binom{N-x'}{Np-k}, \tag{6.5.2}$$

как это следует из комбинаторного анализа. Если всем выборочным точкам в R приписываются равные вероятности, именно $\frac{1}{\binom{N}{Np}}$, мы имеем

$$P(x=x')=\frac{\binom{x'-1}{k-1}\binom{N-x'}{Np-k}}{\binom{N}{Np}}, \tag{6.5.3}$$

что и дает нам ф. в. $p(x)$ случайной величины x. Распределение, имеющее в качестве своей ф. в. функцию (6.5.3), будет называться гипергеометрическим распределением времени ожидания. Заданное значение x' случайной величины x есть, по существу, число испытаний, которые мы должны *пронаблюдать*, чтобы получить точно k элементов C.

Для того чтобы убедиться в том, что

$$p(x)=\frac{\binom{x-1}{k-1}\binom{N-x}{Np-k}}{\binom{N}{Np}} \tag{6.5.4}$$

является ф. в., где выборочное пространство величины x есть множество целых чисел $k, k+1, \ldots, Nq+k$, $1 \leqslant k \leqslant Np$, мы должны показать, что

$$\sum_{x=k}^{Nq+k}\binom{x-1}{k-1}\binom{N-x}{Np-k}=\binom{N}{Np}. \tag{6.5.5}$$

Заметим, что для $|t|<1$ мы имеем

$$(1-t)^{-k}=\sum_{x=k}^{\infty}\binom{x-1}{k-1}t^{x-k},$$

$$(1-t)^{-(Np-k+1)}=\sum_{y=-\infty}^{Nq+k}\binom{N-y}{Np-k}t^{Nq-y+k} \tag{6.5.6}$$

и

$$(1-t)^{-(Np+1)}=\sum_{z=0}^{\infty}\binom{Np+z}{Np}t^{z}. \tag{6.5.7}$$

Из этой формулы можно увидеть, что коэффициент при t^{Nq} в разложении бинома $(1-t)^{-(Np+1)}$ по степеням t равен $\binom{N}{Np}$. Перемножая оба равенства в (6.5.6), мы получаем

$$(1-t)^{-(Np+1)}=\sum_x\sum_y\binom{x-1}{k-1}\binom{N-y}{Np-k}t^{Nq+(x-y)}. \tag{6.5.8}$$

Коэффициент при t^{Nq} в разложении бинома $(1-t)^{-(Np+1)}$, определяемом формулой (6.5.8), равен сумме коэффициентов $\binom{x-1}{k-1}\times$ $\times\binom{N-y}{Np-k}$, взятой по *всем возможным* парам целых чисел (x, y), для которых $x=y$, т. е. равен

$$\sum_{x=k}^{Nq+k}\binom{x-1}{k-1}\binom{N-x}{Np-k}. \tag{6.5.9}$$

Поскольку одно разложение (6.5.7) дает в качестве коэффициента при t^{Nq} число $\binom{N}{Np}$, тогда как другое дает (6.5.9), мы заключаем, что (6.5.5) имеет место и, следовательно, $p(x)$, представляемая (6.5.4), есть ф. в.

Распределение, имеющее ф. в. (6.5.4), доставляет нам еще один пример дискретного распределения, характеристическая функция которого не приносит никакой пользы при нахождении моментов. Но моменты могут быть найдены методом, подобным использованному в § 6.1. В этом случае возможно проще вычислить $\mathscr{E}[(x-k)^{[r]}]$, *r-й факториальный момент величины x относительно k*, а затем, исходя из этих факториальных моментов, найти обычные моменты. Мы имеем

$$\mathscr{E}[(x-k)^{[r]}]=\sum_{x=k}^{Nq+k}(x-k)^{[r]}\binom{x-1}{k-1}\binom{N-x}{Np-k}\Big/\binom{N}{Np}=$$

$$=(k+r-1)^{[r]}\left\{\sum_{x=k+r}^{Nq+k}\binom{x-1}{k+r}\binom{N-x}{Np-k}\right\}\Big/\binom{N}{Np}. \tag{6.5.10}$$

Если мы теперь воспользуемся формулой (6.5.5), то увидим, что сумма, заключенная в фигурные скобки, равна $\binom{N}{Np+r}$. Следовательно,

$$\mathscr{E}[(x-k)^{[r]}]=(k+r-1)^{[r]}\binom{N}{Np+r}\Big/\binom{N}{Np}. \tag{6.5.11}$$

Полагая в этом выражении $r=1, 2$, можно найти, что среднее значение случайной величины x равно

$$\mu(x)=\frac{k(N+1)}{Np+1} \tag{6.5.12}$$

и дисперсия

$$\sigma^2(x)=\frac{Nq(Nq-1)k(k+1)}{(Np+2)(Np+1)}+\frac{k(2k+1)(N+1)}{Np+1}-\frac{k^2(N+1)^2}{(Np+1)^2}-k(k+1). \tag{6.5.13}$$

(b) Биномиальный случай. Предположим, что последовательно осуществляются некоторые «испытания» и исход каждого из них есть либо C, либо $\overline{C}$, причем все испытания независимы. Пусть $P(C)=p$ и $P(\overline{C})=q$, $p+q=1$. Наше основное выборочное пространство представляет собой множество всех возможных последовательностей (выборочных точек) исходов C и $\overline{C}$, которые могли бы наступить в неограниченно длинном ряду испытаний. Введем случайную величину x, определяемую для каждой выборочной точки как число испытаний, необходимых для того, чтобы получить точно k исходов C.

Пусть E — событие в R, состоящее в наступлении $(k-1)$ исходов C в $(x-1)$ первых испытаниях, и F — событие, состоящее в наступлении исхода C в испытании с номером x. Если событие G обозначает тот факт, что потребуется точно x испытаний, чтобы получить точно k исходов C, то $G=E\cap F$ и $P(G)=P(E\cap F)=P(E)\times P(F\mid E)$.

Но $P(F\mid E)=p$ и

$$P(E)=\binom{x-1}{k-1}p^{k-1}q^{x-k}. \tag{6.5.14}$$

Таким образом, если $p(x)$ есть ф. в. величины x, мы имеем $p(x)=P(G)$, т. е.

$$p(x)=\binom{x-1}{k-1}\cdot p^k q^{x-k}, \qquad x=k,\ k+1,\ \ldots \tag{6.5.15}$$

Заметим, что выражение для $p(x)$ совпадает с $(x-k+1)$-м членом разложения функции $p^k(1-q)^{-k}$ в ряд по степеням q. Сумма значений $p(x)$ по $x=k$, $k+1$, ... равна $p^k(1-q)^{-k}=p^kp^{-k}=1$.

Распределение, имеющее ф. в., определенную формулой (6.5.15), будет называться *биномиальным распределением времени ожидания*. Ф. в. $p(x)$ есть просто вероятность того, что мы должны ждать окончания испытания с номером x, чтобы получить k исходов C. Это распределение также называется *отрицательным биномиальным распределением или распределением Паскаля*.

Характеристическая функция распределения (6.5.15) равна

$$\varphi(t) = \sum_{x=k}^{\infty} \binom{x-1}{k-1} (pe^{it})^k (qe^{it})^{x-k} = (pe^{it})^k (1 - qe^{it})^{-k},$$

что приводит к формуле

$$\varphi(t) = p^k (e^{-it} - q)^{-k}. \tag{6.5.16}$$

С помощью обычной процедуры дифференцирования можно убедиться в том, что

$$\mu(x) = \frac{k}{p}, \qquad \sigma^2(x) = \frac{kq}{p^2}. \tag{6.5.17}$$

Обратим внимание, что среднее и дисперсия являются здесь предельными выражениями для (6.5.12) и (6.5.13), когда $N \to \infty$. На самом деле ф. в. биномиального распределения времени ожидания (6.5.15), как можно предполагать, есть предел ф. в. гипергеометрического распределения времени ожидания (6.5.3) при $N \to \infty$. Для читателя будет поучительно проверить эти утверждения.

Мы рассмотрели лишь простейший случай дискретного распределения времени ожидания. Возникают задачи, в которых мы хотим знать не только функцию вероятностей числа испытаний, требующихся, чтобы получить заданное число исходов C, но и функцию вероятностей числа испытаний, необходимых для получения заданных чисел исходов C и $\bar{C}$. Далее, в случае более чем двух классов мы можем хотеть определить функцию вероятностей числа испытаний, необходимых для получения заданных чисел исходов из каждого класса. Задачи этого типа являются значительно более сложными, чем те, которые мы уже обсудили. Некоторые из этих задач были рассмотрены Гиршиком, Мостеллером и Сэвиджем (1946), Холдэйном (1945), Лапласом (1814) и Маккарти (1947).

6.6. Распределения, встречающиеся в теории серий

Теория серий, как мы увидим в последующих главах, играет важную роль в некоторых задачах непараметрического статистического вывода. Но основные распределения, встречающиеся в теории серий, можно с известным удобством ввести в настоящий момент. Результаты, которые будут представлены ниже, принадлежат главным образом Муду (1940).

(а) Серии, образованные из элементов двух видов. Рассмотрим выборочное пространство R, выборочные точки которого суть множества всех $\binom{n}{n_1}$ перестановок n_1 элементов C и n_2 элементов $\bar{C}$, где $n_1 + n_2 = n$. Любая данная выборочная точка есть последовательность элементов C и $\bar{C}$, которая состоит из чередующихся *серий* этих элементов. *Длиной* серии называется число элементов в ней. Обозначим через r_{1j} число серий длины j, образованных из элементов C, и через r_{2j} число серий длины j, образованных из элементов $\bar{C}$. Например, для последовательности

$$C\,C\,C\,\bar{C}\,\bar{C}\,C\,C\,\bar{C}\,C\,C\,C\,\bar{C}\,\bar{C}\,C\,\bar{C}$$

мы имеем: $n_1=8$, $n_2=6$, $r_{11}=1$, $r_{12}=2$, $r_{13}=1$, $r_{21}=2$, $r_{22}=2$, причем все другие значения r_{ij} равны нулю.

Из определения чисел n_1, n_2, r_{1j} и r_{2j} следует, что $\sum\limits_j jr_{1j}=n_1$ и $\sum\limits_j jr_{2j}=n_2$. Пусть $r_1=\sum\limits_j r_{1j}$ и $r_2=\sum\limits_j r_{2j}$ означают общие числа серий, образованных соответственно из элементов C и $\bar{C}$. Каковы бы ни были вероятности, назначенные выборочным точкам в R, мы можем ввести случайные величины r_{1j} и r_{2j}, которые будут иметь определенные значения, установленные выше для любой данной выборочной точки в R. Наша задача заключается в том, чтобы найти ф. в. этих случайных величин, когда всем выборочным точкам в R назначаются равные вероятности. Для данного множества значений r_{1j} число способов расположения r_1 серий элементов C равно

$$\frac{r_1!}{r_{11}!\dots r_{1n_1}!}. \tag{6.6.1}$$

Аналогично этому число способов расположения r_2 серий элементов $\bar{C}$ равно

$$\frac{r_2!}{r_{21}!\ \dots\ r_{2n_2}!}. \tag{6.6.2}$$

Заметим, что r_1 и r_2 не могут отличаться друг от друга более чем на единицу, ибо в противном случае по крайней мере две серии элементов одного вида должны были бы оказаться примыкающими, а это не соответствует определению серии. Если $r_1=r_2$, данное расположение серий элементов C можно двумя способами сочетать с данным расположением серий элементов $\bar{C}$, так что полная последовательность элементов C и $\bar{C}$ будет начинаться либо с серии элементов C, либо с серии элементов $\bar{C}$. Если мы определим функцию $\gamma(r_1, r_2)$, значение которой в точке (r_1, r_2) есть число способов расположения r_1 неразличимых объектов одного вида и r_2 неразличимых объектов другого вида, где никакие два объекта одного и того же вида не находятся рядом, то мы будем иметь

$$\gamma(r_1, r_2)=\begin{cases} 0, & |r_1-r_2|>1, \\ 1, & |r_1-r_2|=1, \\ 2, & r_1=r_2. \end{cases} \tag{6.6.3}$$

Таким образом, общее число способов получить r_{1j} серий элементов C длин $j=1, \ldots, n_1$ и r_{2j} серий элементов $\overline{C}$ длин $j=1, \ldots, n_2$ равно

$$\frac{r_1!\, r_2!\, \gamma(r_1, r_2)}{r_{11}! \ldots r_{1n_1}!\, r_{21}! \ldots r_{2n_2}!}. \tag{6.6.4}$$

Но существует всего $\binom{n}{n_1}$ возможных расположений n_1 элементов C и n_2 элементов $\overline{C}$, т. е. $\binom{n}{n_1}$ выборочных точек в R. Если всем этим расположениям назначаются равные вероятности, то (n_1+n_2)-мерная случайная величина $(r_{ij};\ j=1, \ldots, n_i,\ i=1, 2)$ будет иметь следующую ф. в.:

$$p(\{r_{ij}\}) = \frac{r_1!\, r_2!}{r_{11}! \ldots r_{1n_1}!\, r_{21}! \ldots r_{2n_2}!} \cdot \frac{n_1!\, n_2!}{n!} \cdot \gamma(r_1, r_2). \tag{6.6.5}$$

Если мы интересуемся лишь ф. в. величин r_{1j} — чисел серий элементов C длин $j=1, \ldots, n_1$, — мы должны определить из (6.6.5) маргинальное распределение величины $(r_{11}, \ldots, r_{1n_1})$, т. е. просуммировать $p(\{r_{ij}\})$ по $r_{21}, \ldots, r_{2n_2}$. Это означает, что мы должны взять сумму значений (6.6.2) по всем $r_{21}, \ldots, r_{2n_2}$, для которых $\sum\limits_j jr_{2j}=n_2$ и $\sum\limits_j r_{2j}=r_2$. Для того чтобы сделать это, мы воспользуемся следующим тождеством относительно s, имеющим место для значений s, близких к нулю:

$$(s+s^2+\ldots)^{r_2} \equiv s^{r_2}(1-s)^{-r_2} \equiv s^{r_2} \sum_{t=0}^{\infty} \frac{(r_2+t-1)!}{(r_2-1)!\, t!} s^t. \tag{6.6.6}$$

Теперь коэффициент при s^{n_2} в левой части (6.6.6) есть сумма значений (6.6.2) по $r_{21}, \ldots, r_{2n_2}$ при ограничениях $\sum\limits_j jr_{2j}=n_2$ и $\sum\limits_j r_{2j}=r_2$. Но коэффициент при s^{n_2} в левой части (6.6.6) должен быть равен коэффициенту при s^{n_2} в правой части этого тождества, который имеет значение

$$\frac{(n_2-1)!}{(r_2-1)!\,(n_2-r_2)!}. \tag{6.6.7}$$

Следовательно, ф. в. величин $(r_{1j},\ j=1, \ldots, n_1)$ и r_2 есть

$$p(\{r_{1j}\}, r_2) = \frac{r_1!}{r_{11}! \ldots r_{1n_1}!} \cdot \frac{(n_2-1)!}{(r_2-1)!\,(n_2-r_2)!} \cdot \frac{n_1!\, n_2!}{n!}\, \gamma(r_1, r_2). \tag{6.6.8}$$

Наконец, чтобы получить ф. в. величин $\{r_{1j}\}$, мы должны просуммировать (6.6.8) по r_2. Применяя (6.6.3), мы находим после некоторых упрощений

$$\sum_{r_2=1}^{n_2} \frac{(n_2-1)!}{(r_2-1)!\,(n_2-r_2)!}\,\gamma(r_1, r_2) = \frac{n_2+1}{r_1!\,(n_2-r_1+1)!} = \binom{n_2+1}{r_1}. \tag{6.6.9}$$

Таким образом, ф. в. величин $(r_{1j},\ j=1, \ldots, n_1)$ дается формулой

$$p(\{r_{1j}\}) = \frac{r_1!}{r_{11}! \ldots r_{1n_1}!} \cdot \frac{\binom{n_2+1}{r_1}}{\binom{n}{n_1}}. \tag{6.6.10}$$

Подобный же результат имеет место, конечно, и для ф. в. величин $\{r_{2j}\}$.

Другим важным распределением является распределение величин r_1 и r_2. Ф. в. этого распределения получается суммированием значений (6.6.8) по r_{1j} при условиях $\sum\limits_j jr_{1j} = n_1$ и $\sum\limits_j r_{1j} = r_1$. Порядок действий здесь подобен тому, который был использован при выводе (6.6.8) с помощью суммирования (6.6.5). Этим путем мы приходим к формуле

$$p(r_1, r_2) = \frac{\binom{n_1-1}{r_1-1}\binom{n_2-1}{r_2-1}}{\binom{n}{n_1}}\,\gamma(r_1, r_2). \tag{6.6.11}$$

Просуммировав (6.6.11) по r_2, мы заключаем, что ф. в. величины r_1 — общего числа серий элементов C — есть

$$p(r_1) = \frac{\binom{n_1-1}{r_1-1}\binom{n_2+1}{r_1}}{\binom{n}{n_1}}. \tag{6.6.12}$$

Аналогичное выражение справедливо для ф. в. величины r_2.

Чтобы проверить, что сумма значений $p(r_1)$ по всему выборочному пространству величины r_1 равна единице, мы рассмотрим соотношение

$$\sum_{i=0}^{B} \binom{A}{k+i}\binom{B}{i} = \binom{A+B}{k+B}, \tag{6.6.13}$$

полученное в результате приравнивания коэффициентов при s^k в тождестве

$$(1+s)^A\left(1+\frac{1}{s}\right)^B \equiv \frac{(1+s)^{A+B}}{s^B}. \tag{6.6.14}$$

Из (6.6.13) следует, что

$$\sum_{r_1=1}^{n_1} \binom{n_1-1}{r_1-1}\binom{n_2+1}{r_1} = \binom{n}{n_1},$$

а это равносильно утверждению

$$\sum_{r_1=1}^{n_1} p(r_1) = 1.$$

Самый легкий путь найти моменты величины r_1 состоит в использовании факториальных моментов. Для g-го факториального момента $\mu'_{[g]}$ мы имеем

$$\mu'_{[g]} = \mathcal{E}(r_1^{[g]}) = \frac{(n_2+1)^{[g]}}{\binom{n}{n_1}} \sum_{r_1=g}^{n_1} \binom{n_1-1}{r_1-1}\binom{n_2+1-g}{r_1-g}. \quad (6.6.15)$$

Но из (6.6.13) вытекает, что

$$\sum_{r_1=g}^{n_1} \binom{n_1-1}{r_1-1}\binom{n_2+1-g}{r_1-g} = \binom{n-g}{n_1-g}.$$

Поэтому

$$\mu'_{[g]} = (n_2+1)^{[g]} \frac{\binom{n-g}{n_1-g}}{\binom{n}{n_1}}, \quad (6.6.16)$$

откуда мы находим

$$\mu(r_1) = \frac{n_1(n_2+1)}{n}, \qquad \sigma^2(r_1) = \frac{(n_2+1)^{[2]}(n_1)^{[2]}}{n(n)^{[2]}}. \quad (6.6.17)$$

Такие же формулы существуют и для факториальных моментов, среднего и дисперсии величины r_2.

Аналогичным образом, учитывая (6.6.11), можно показать, что смешанный факториальный момент величин (r_1-1) и (r_2-1) дается выражением

$$\mathcal{E}\left[(r_1-1)^{[g_1]}(r_2-1)^{[g_2]}\right] = \frac{(n_1-1)^{[g_1]}(n_2-1)^{[g_2]}}{\binom{n}{n_1}} \binom{n-g_1-g_2}{n_1-g_2}. \quad (6.6.18)$$

В некоторого рода задачах представляет особый интерес случайная величина $u = r_1 + r_2$ — *общее число серий.*

Среднее и дисперсия величины u могут быть немедленно получены из (6.6.17) с помощью формул (3.4.2) и (3.4.3). Мы имеем

$$\mu(u) = \mu(r_1) + \mu(r_2) = \frac{2n_1n_2}{n} + 1,$$

$$\sigma^2(u) = \sigma^2(r_1) + \sigma^2(r_2) + 2\operatorname{cov}(r_1, r_2) = \frac{2n_1n_2(2n_1n_2-n)}{n^2(n-1)}. \quad (6.6.19)$$

На самом деле ф. в. величины u можно определить из (6.6.11), суммируя $p(r_1, r_2)$ вдоль прямой линии $u = r_1 + r_2$ на r_1r_2-плоскости. Можно показать, что если u четно, то существует только одна точка, в которой $p(r_1, r_2)$ имеет значение, отличное от нуля. Если u нечетно, то существуют две точки, обладающие этим свойством. Ф. в. величины u может быть сразу записана в виде

$$p(u) = \begin{cases} \dfrac{2\dbinom{n_1-1}{\frac{1}{2}u-1}\dbinom{n_2-1}{\frac{1}{2}u-1}}{\dbinom{n}{n_1}}, & \text{если } u \text{ четно,} \\[3ex] \dfrac{\dbinom{n_1-1}{\frac{1}{2}u-\frac{1}{2}}\dbinom{n_2-1}{\frac{1}{2}u-\frac{3}{2}}+\dbinom{n_1-1}{\frac{1}{2}u-\frac{3}{2}}\dbinom{n_2-1}{\frac{1}{2}u-\frac{1}{2}}}{\dbinom{n}{n_1}}, & \text{если } u \text{ нечетно.} \end{cases} \tag{6.6.20}$$

Этот результат был первоначально получен Стивенсом (1939). Разнообразные таблицы, касающиеся $p(u)$, были составлены Сведом и Эйзенхартом (1943).

Во всех предыдущих рассуждениях n_1 и n_2 были постоянными числами. Если мы условимся считать их случайными величинами, то распределения, ф. в. которых выражаются формулами (6.6.5), (6.6.10), (6.6.11) и (6.6.12), будут условными распределениями при фиксированных значениях n_1 и n_2. Если бы ф. в. величин n_1 и n_2 равнялась $p^*(n_1, n_2)$, то произведение каждой из указанных четырех ф. в. на $p^*(n_1, n_2)$ дало бы ф. в. соответствующих величин r и n_1 и n_2. Чтобы получить ф. в. величин r в любом случае, мы должны были бы просуммировать относящееся к этому случаю произведение по всей области изменения n_1 и n_2. В частности, если бы мы рассмотрели данное множество независимых испытаний, каждое из которых приводит либо к C, либо к $\bar{C}$, то n_1 и n_2 были бы линейно зависимыми случайными величинами, удовлетворяющими условию $n_1 + n_2 = n$ и n_1 имело бы биномиальное распределение $\mathrm{Bi}(n, p)$, где p есть вероятность появления C в отдельном испытании. В этом примере

$$p^*(n_1, n_2) = \binom{n}{n_1} p^{n_1} q^{n_2}, \qquad n_2 = n - n_1. \tag{6.6.21}$$

Чтобы найти здесь g-й факториальный момент величины r_1, мы умножили бы выражение в правой части (6.6.16) на (6.6.21) и просуммировали бы по n_1 от 0 до n, разумеется, понимая, что члены этой суммы, отвечающие $n = 1, \ldots, g-1$, равны нулю.

(b) Серии, образованные из элементов k типов. Предыдущие результаты допускают непосредственное распространение на случай серий, образованных из элементов нескольких типов. Предположим, что имеется некоторое множество n элементов, состоящее из n_1 элементов $C_1, \ldots, n_k$ элементов C_k, где $n_1+\ldots+n_k=n$. Пусть r_{ij} — случайная величина, означающая число серий элементов C_i длины j, и $r_i=\sum_j r_{ij}$ — общее число серий элементов C_i. Муд (1940) показал, что ф. в. множества величин r_{ij} равна

$$p(\{r_{ij}\})=\frac{n_1!\ldots n_k!}{n!}\left[\prod_{i=1}^{k}\frac{r_i!}{r_{i1}!\ldots r_{in_i}!}\right]\gamma(r_1, \ldots, r_k), \qquad (6.6.22)$$

где $\gamma(r_1, \ldots, r_k)$ есть число способов, которыми можно так расположить r_1 объектов первого вида, ..., r_k объектов k-го вида, что никакие два соседних объекта не будут иметь один и тот же вид. Функция $\gamma(r_1, \ldots, r_k)$ совпадает с коэффициентом при $x_1^{r_1}\ldots x_k^{r_k}$ в разложении произведения

$$(x_1+\ldots+x_k)^k(x_2+x_3+\ldots+x_k)^{r_1-1}\times$$
$$\times(x_1+x_3+\ldots+x_k)^{r_2-1}\ldots(x_1+x_2+\ldots+x_{k-1})^{r_k-1} \qquad (6.6.23)$$

по степеням $x_1, \ldots, x_k$. Доказательство формулы (6.6.22) заключается в прямом обобщении доказательства формулы (6.6.5) и предоставляется читателю. Подобным же образом обобщение рассуждений, использованных при переходе от (6.6.5) к (6.6.11), дает ф. в. величины $(r_1, \ldots, r_k)$

$$p(\{r_i\})=\frac{n_1!\ldots n_k!}{n!}\left[\prod_{i=1}^{k}\binom{n_i-1}{r_i-1}\right]\gamma(r_1, \ldots, r_k). \qquad (6.6.24)$$

Моменты величин могут быть найдены из (6.6.22) и (6.6.24) аналогично тому, как это было сделано в случае $k=2$.

ЗАДАЧИ

6.1. Показать, что если x — случайная величина, имеющая гипергеометрическое распределение $H(N, n; p)$, то

$$\mathscr{E}[(n-x)^{[r]}]=(Nq)^{[r]}\binom{N-r}{n-r}\Big/\binom{N}{n}.$$

6.2. Урна содержит $m+n$ фишек, занумерованных числами $1, 2, \ldots, m+n$. Случайным образом из урны берется множество n фишек. Показать, что вероятность того, что x из взятых фишек будут иметь номера, превосходящие номера всех оставшихся в урне фишек, равна

$$\binom{m+n-x-1}{m-1}\Big/\binom{m+n}{m}.$$

Показать также, что

$$\mathscr{E}[(m+n-x+r-1)^{[r]}]=(m+r-1)^{[r]}\binom{m+n+r}{m+r}\Big/\binom{m+n}{n}.$$

Определить отсюда среднее и дисперсию случайной величины x.

6.3. Показать, что если $(x_1, \ldots, x_k)$ — k-мерная случайная величина, имеющая k-мерное гипергеометрическое распределение $H(N, n; p_1, \ldots, p_k)$, то маргинальное распределение величины $(x_1, \ldots, x_{k_1})$, $k_1 < k$, есть гипергеометрическое распределение

$$H(N, n; p_1, \ldots, p_{k_1}).$$

6.4. (*Продолжение*). Показать, что условная случайная величина

$$x_k \mid x_1, \ldots, x_{k-1}$$

имеет одномерное гипергеометрическое распределение

$$H\left(N(p_k+p_{k+1}),\ n-x_1-\ldots-x_{k-1};\ \frac{p_k}{p_k+p_{k+1}}\right).$$

6.5. (*Продолжение*). Показать, что величина $x_1+\ldots+x_k$ имеет одномерное гипергеометрическое распределение $H(N, n; p_1+\ldots+p_k)$.

6.6. Показать, что результатом формального суммирования значений $p(x_1, \ldots, x_k)$, определяемых формулой (6.1.6), по всему выборочному пространству является единица.

6.7. Доказать теорему **6.2.1.**

6.8. Доказать теорему **6.3.1.**

6.9. Доказать теорему **6.4.1.**

6.10. Показать, что если $n \to \infty$ и каждое из $p_1, \ldots, p_k \to 0$, причем $np_1 = \mu_1, \ldots, np_k = \mu_k$, то предел ф. в., определенной в (6.3.3), есть произведение ф.в. независимых распределений Пуассона Po $(\mu_1), \ldots,$ Po (μ_k).

6.11. Показать, что биномиальное распределение времени ожидания, ф.в. которого дается формулой (6.5.15), является воспроизводящим по k.

6.12. Используя характеристические функции, показать, что если $(x_1, \ldots, x_k)$ имеет k-мерное мультиномиальное распределение

$$M(n; p_1, \ldots, p_k),$$

то маргинальное распределение величины $(x_1, \ldots, x_{k_1})$, $k_1 < k$, есть k_1-мерное мультиномиальное распределение

$$M(n; p_1, \ldots, p_{k_1}).$$

6.13. (*Продолжение*). Показать, что распределением условной случайной величины $x_k \mid x_1, \ldots, x_{k-1}$ является биномиальное распределение

$$\mathrm{Bi}\left(n-x_1-x_2-\ldots-x_{k-1};\ \frac{p_k}{p_k+p_{k+1}}\right).$$

6.14. (*Продолжение*). Показать, что величина $x_1+\ldots+x_{k_1}$, $k_1<k$, имеет биномиальное распределение Bi $(n; p_1+\ldots+p_{k_1})$.

6.15. Показать, что предел гипергеометрического распределения времени ожидания, определенного в (6.5.3), при $N \to \infty$ есть биномиальное распределение времени ожидания (6.5.15).

6.16. Проверить формулу (6.6.12).

6.17. Проверить формулу (6.6.18).

6.18. Показать, что если x — случайная величина, имеющая распределение Пуассона Po (μ), и условная случайная величина $y \mid x$ имеет биномиальное распределение Bi (x, p), то безусловное распределение величины y есть распределение Пуассона Po (μp).

6.19. Показать, что если x имеет биномиальное распределение времени ожидания (6.5.15) и $y = x - k$, то предельное распределение величины y, когда $k \to \infty$, $q \to 0$ и $kq \to \mu$, есть распределение Пуассона Po (μ).

6.20. Показать, что если x имеет распределение Пуассона $\mathrm{Po}(\mu)$, то к. ф. р. величины x дается формулой

$$\frac{1}{x!}\int_{\mu}^{\infty} e^{-z}\cdot z^{x}\,dz.$$

6.21. *Зависящее от случайных параметров распределение Неймана типа А.* Предположим, что x есть случайная величина, имеющая распределение Пуассона $\mathrm{Po}(\mu_1)$, и y — случайная величина, для которой $y\,|\,x$ имеет распределение Пуассона $\mathrm{Po}(\mu_2 x)$. Показать, что характеристическая функция безусловной случайной величины y дается формулой

$$\varphi(t)=\exp\{-\mu_1[1-e^{-\mu_2(1-e^{it})}]\}.$$

Показать также, что

$$\mathscr{E}(y)=\mu_1\mu_2,$$
$$\sigma^2(y)=\mu_1\mu_2(1+\mu_2).$$

6.22. Показать, что если x — случайная величина, имеющая биномиальное распределение $\mathrm{Bi}(n, p)$, то к. ф. р. величины x дается формулой

$$(n-x)\binom{n}{x}\int_{0}^{q} y^{n-x-1}(1-y)^{x}\,dy.$$

6.23. *Задача проверки крови.* Лица, составляющие большую совокупность, проходят анализ крови группами по k человек одновременно следующим образом. У всех k лиц берутся образцы крови и смешиваются. Если анализ смешанной крови отрицателен, одна проверка оказывается достаточной для данных k лиц. Если анализ положителен, кровь каждого лица проверяется отдельно. Показать, что если q — вероятность отрицательного анализа крови любого выбранного наудачу лица, то значение k, минимизирующее ожидаемое значение общего числа анализов, есть положительное целое число, для которого интервал

$$\left(\sqrt[k]{\frac{1}{k(k+1)(1-q)}},\ \sqrt[k]{\frac{q}{k(k+1)(1-q)}}\right)$$

содержит q. Показать,что если q близко к единице, то известным приближением к k служит решение уравнения

$$k^2g^k\ln g+1=0$$

[Дорфман (1943)].

6.24. Предположим, что урна содержит n фишек, занумерованных числами 1, 2, ..., n. Некоторое лицо вытягивает одну фишку, возвращает ее обратно в урну, вытягивает снова одну фишку, опять возвращает ее обратно и т. д., пока не появится фишка, которая уже была вынута раньше, и тогда процесс извлечения фишек заканчивается. Пусть x — случайная величина, означающая число извлечений, необходимых для достижения этой цели. Показать, что ф. в. величины x имеет вид

$$p(x)=(x-1)!\binom{n}{x-1}\frac{(x-1)}{n^k},\qquad x=2,\ \ldots,\ n+1.$$

Убедиться, что $\sum_{x=2}^{n+1} p(x) = 1$. Показать, что

$$\mathscr{E}(x) = 2 + \left(1 - \frac{1}{n}\right) + \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \ldots + \left(1 - \frac{1}{n}\right) \times$$
$$\times \left(1 - \frac{2}{n}\right) \ldots \left(1 - \frac{n-1}{n}\right).$$

6.25. В рамках биномиального случая, рассмотренного в § 6.5, предположим, что x есть число испытаний, необходимых для получения k последовательных успехов. Показать, что производящая функция моментов величины x равна

$$\psi(t) = (pe^t)^k (1 - pe^t)(1 - e^t + p^k q e^{(k+1)t})^{-1}.$$

Показать также, что

$$\mathscr{E}(x) = \frac{1 - p^k}{p^k q}.$$

6.26. *Задача распределения случайных объемов групп — мультиномиальный случай.* В мультиномиальном распределении, имеющем ф. в. (6.3.3), положим

$$p_1 = \ldots = p_{k+1} = \frac{1}{k+1}.$$

Пусть $r_0, r_1, \ldots, r_n$ означают соответственно количества компонент величины $(x_1, \ldots, x_{k+1})$, которые равны $0, 1, \ldots, n$. Показать, что ф. в. величины $(r_0, r_1, \ldots, r_n)$ дается формулой

$$\frac{n!\,(k+1)\,(k+1)^{-n}}{(0!)^{r_0}(1!)^{r_1}\ldots(n!)^{r_n}\, r_0! \ldots r_n!},$$

причем $r_0, r_1, \ldots, r_n$ подчинены условиям

$$r_0 + r_1 + \ldots + r_n = k + 1,$$
$$r_1 + 2r_2 + \ldots + nr_n = n.$$

Кроме того, показать, что

$$\mathscr{E}\left[r_0^{[s_0]} r_1^{[s_1]} \ldots r_n^{[s_n]}\right] = \frac{n!\,(k+1)!\,(k+1)^{-n}(k+1-B)^{n-A}}{(0!)^{s_0}(1!)^{s_1}\ldots(n!)^{s_n}(n-A)!\,(k+1-B)!},$$

где $A = s_1 + 2s_2 + \ldots + ns_n$ и $B = s_0 + s_1 + \ldots + s_n$. Найти отсюда значения $\mathscr{E}(r_i)$, $\sigma^2(r_i)$ и $\operatorname{cov}(r_i, r_u)$ [Тьюки (1949b)].

6.27. *Задача распределения случайных объемов групп — гипергеометрический случай.* Предположим, что колода, содержащая $M(k+1)$ карты, имеет M карт каждой из $k+1$ отличных друг от друга «мастей» и после тщательного тасования «игроку» сдается n карт. Пусть r_i, $i = 0, 1, \ldots, n$, означает число мастей, каждая из которых представлена i картами, доставшимися игроку. В частности, r_0 есть число отсутствующих мастей. Показать, что ф. в. $(n+1)$-мерной случайной величины $(r_0, r_1, \ldots; r_n)$ дается формулой

$$\frac{(M!)^{k+1}(k+1)!\binom{M(k+1)}{n}^{-1}}{r_0!\,r_1! \ldots r_n!\,[0!\,(M-0)!]^{r_0}\,[1!\,(M-1)!]^{r_1}\ldots[n!\,(M-n)!]^{r_n}}.$$

Кроме того, показать, что

$$\mathscr{E}\,[r_0^{[s_0]} r_1^{[s_1]} \ldots r_n^{[s_n]}] =$$

$$= \frac{(M!)^B (k+1)! \binom{M(k+1-B)}{n-A}}{(k+1-B)! \binom{M(k+1)}{n} [0!\,(M-0)!]^{s_0} \ldots [n!\,(M-n)!]^{s_n}},$$

где $A = s_1 + 2s_2 + \ldots + ns_n$ и $B = s_0 + s_1 + \ldots + s_n$. Найти отсюда значения $\mathscr{E}\,(r_t)$, $\sigma^2\,(r_t)$ и $\operatorname{cov}(r_t, r_u)$ [Тьюки (1949b)].

6.28. Задача совпадения для двух колод карт. Пусть A — колода, содержащая N карт, каждая из которых имеет одну и только одну «масть» $S_1, \ldots, S_k$, причем количества карт, имеющих эти масти, равны соответственно $m_1, \ldots, m_k$. Аналогично пусть B — вторая колода, содержащая N карт, из которых $n_1, \ldots, n_k$ имеют соответственно масти $S_1, \ldots, S_k$. Колода A перетасовывается и карты сдаются, располагаясь открыто в одну линию. Точно так же и колода B перетасовывается и карты сдаются, располагаясь открыто в линию непосредственно ниже линии, образованной картами колоды A. Пусть x — случайная величина, означающая число пар карт в двух колодах, которые оказываются сравнимыми по масти. Назначая всем возможным перестановкам карт равные вероятности, показать, что ф. в. величины x дается коэффициентом при

$$e^{tx} a_1^{m_1} \ldots a_k^{m_k} b_1^{n_1} \ldots b_k^{n_k}$$

в разложении функции

$$\Phi = \frac{1}{M} \left(\sum_{i,j=1}^{k} a_i b_j e^{\delta_{ij} t} \right)^N,$$

где

$$\delta_{ij} = \begin{cases} 0, & i \neq j, \\ 1, & i = j, \end{cases}$$

и

$$M = \frac{(N!)^2}{m_1! \ldots m_k! n_1! \ldots n_k!}.$$

Вывести отсюда, что r-й момент μ_r' величины x есть коэффициент при $a_1^{m_1} \ldots a_k^{m_k} b_1^{n_1} \ldots b_k^{n_k}$ в разложении функции

$$\left[\frac{\partial^r \Phi}{\partial t^r}\right]_{t=0}.$$

В частности, установить, что

$$\mathscr{E}\,(x) = \sum_{i=1}^{k} \frac{m_i n_i}{N},$$

$$\sigma^2\,(x) = \frac{1}{N^2 (N-1)} \left[\left(\sum_{i=1}^{k} m_i n_i \right)^2 - N \sum_{i=1}^{k} (m_i^2 n_i + m_i n_i^2) + N^2 \sum_{i=1}^{k} m_i n_i \right]$$

[Бэттин (1942) и Капланский и Риордан (1945)].

ГЛАВА 7

НЕКОТОРЫЕ СПЕЦИАЛЬНЫЕ НЕПРЕРЫВНЫЕ РАСПРЕДЕЛЕНИЯ

В настоящей главе мы рассмотрим ряд важных непрерывных распределений вероятностей, встречающихся в математической статистике, вместе с некоторыми их свойствами. Многие из приведенных результатов будут использованы в последующих главах.

7.1. Равномерное распределение

Простейшее непрерывное распределение, которое мы считаем полезным здесь определить, задается следующей ф. п. в.:

$$f(x)=\begin{cases} \frac{1}{\omega}, & \mu-\frac{\omega}{2}\leqslant x\leqslant\mu+\frac{\omega}{2}, \\ 0, & x<\mu-\frac{\omega}{2},\quad x>\mu+\frac{\omega}{2}. \end{cases} \tag{7.1.1}$$

Будет удобно называть распределение вероятностей с данной ф. п. в. *равномерным распределением* $R(\mu,\ \omega)$. Среднее и дисперсия этого распределения равны соответственно

$$\mu(x)=\mu,\qquad \sigma^2(x)=\frac{\omega^2}{12}. \tag{7.1.2}$$

Параметр ω называется *размахом* распределения.

Следует отметить, что если x — случайная величина, имеющая равномерное распределение $R(\mu,\ \omega)$, то

$$y=\frac{x-\mu+\frac{1}{2}\omega}{\omega}$$

является случайной величиной, имеющей равномерное распределение $R\left(\frac{1}{2},\ 1\right)$, ф. п. в. которого отлична от нуля на интервале $[0,\ 1]$.

Важный случай величины, имеющей равномерное распределение $R\left(\frac{1}{2},\ 1\right)$, находит выражение в следующем утверждении.

7.1.1. *Если x — случайная величина, имеющая непрерывную к. ф. р. $F(x)$, то случайная величина $y=F(x)$ имеет равномерное распределение* $R\left(\frac{1}{2},\ 1\right)$.

Доказательство этой теоремы немедленно вытекает из того факта, что к. ф. р. величины y равна

$$H(y)=P(F(x)\leqslant y)=\begin{cases} 1, & y>1,\\ y, & 0<y\leqslant 1,\\ 0, & y\leqslant 0,\end{cases}$$

т. е. является к. ф. р. равномерного распределения $R\left(\frac{1}{2},\ 1\right)$.

7.2. Нормальное распределение

Наиболее важным распределением непрерывной случайной величины является *нормальное* или *гауссовское распределение*. Ф. п. в. этого распределения может быть записана в виде

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \tag{7.2.1}$$

при $\infty<x<\infty$, где μ и σ^2 — *параметры*. Ниже будет показано, что эти два параметра фактически равны соответственно *среднему* и *дисперсии* ф. п. в. (7.2.1). Будет удобно называть распределение, имеющее ф. п. в. (7.2.1), нормальным распределением $N(\mu,\ \sigma^2)$ или просто

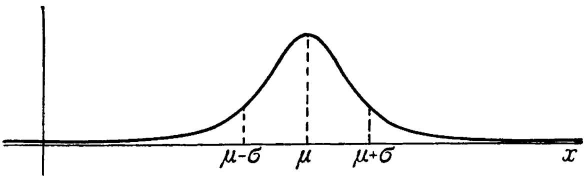

Рис. 7.1. График ф. п. в. нормального распределения (7.2.1).

распределением $N(\mu,\ \sigma^2)$, так как это распределение встречается весьма часто. График функции (7.2.1), изображенный на рис. 7.1, симметричен относительно прямой $x=\mu$ и имеет максимальную ординату $\frac{1}{\sqrt{2\pi}\sigma}$ в точке $x=\mu$. При $x=\mu\pm\sigma$ график содержит две точки перегиба.

Очень удобной для целей табулирования является форма нормального распределения, соответствующая случайной величине $y=\frac{x-\mu}{\sigma}$. Ф. п. в. величины y определяет нормированное нормальное распределение $N(0,\ 1)$. Заметим, что

$$P(x\leqslant x')=P(y\leqslant y')=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{y'}e^{-\frac{y^2}{2}}dy,$$

где $y'=\frac{x'-\mu}{\sigma}$.

К. ф. р. $\Phi(x)$ нормированного нормального распределения, задаваемая формулой

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}y^2}\,dy, \tag{7.2.1a}$$

протабулирована в широких пределах. Превосходные таблицы подготовлены Национальным бюро стандартов (1942). Обширный указатель различного вида табуляций дали Гринвуд и Хартли (1961).

Сначала мы покажем, что интеграл от нормальной функции распределения (7.2.1) по всей x-оси равен единице.

Полагая $y=\frac{x-\mu}{\sigma}$, мы можем написать

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\,dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2}\,dy. \tag{7.2.2}$$

Пусть I означает интеграл в правой части (7.2.2) $\left(\text{без константы } \frac{1}{\sqrt{2\pi}}\right)$. Наша задача состоит в том, чтобы показать, что $I=\sqrt{2\pi}$. Очевидно,

$$I^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(y_1^2+y_2^2)}\,dy_1\,dy_2. \tag{7.2.3}$$

Вводя полярные координаты $y_1=r\cos\theta$, $y_2=r\sin\theta$, мы получаем

$$I^2=\int_{0}^{2\pi}\int_{0}^{\infty} re^{-\frac{1}{2}r^2}\,dr\,d\theta=2\pi, \tag{7.2.3a}$$

и, следовательно, $I=\sqrt{2\pi}$.

Рассмотрим теперь характеристическую функцию распределения (7.2.1). Мы имеем

$$\varphi(t)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{\mathrm{i}tx-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\,dx, \tag{7.2.4}$$

что можно переписать в виде

$$\varphi(t)=e^{\mathrm{i}t\mu-\frac{1}{2}\sigma^2t^2}\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\mu-\mathrm{i}\sigma^2t)^2/\sigma^2}\,dx. \tag{7.2.4a}$$

Интеграл в (7.2.4a) преобразовывается в интеграл от функции $e^{-\frac{z^2}{2\sigma^2}}$, где z — комплексная переменная, взятый в комплексной плоскости вдоль прямой, параллельной вещественной оси, именно прямой $y=-\mathrm{i}\sigma^2t$, и можно показать, что такой интеграл равен интегралу от той же самой функции, взятому вдоль вещественной оси, т. е.

равен $\int\limits_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}}\,dx$. Но последний интеграл имеет значение $\sigma\sqrt{2\pi}$. Таким образом,

7.2.1. *Характеристическая функция распределения* $N(\mu, \sigma^2)$ *есть*

$$\varphi(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}. \tag{7.2.5}$$

С помощью обычной процедуры дифференцирования мы находим, что среднее и дисперсия распределения $N(\mu, \sigma^2)$ равны соответственно

$$\mu(x) = \mu, \qquad \sigma^2(x) = \sigma^2. \tag{7.2.6}$$

Считая μ и σ^2 в выражении для характеристической функции (7.2.5) параметрами, мы очевидным образом обнаруживаем, что характеристическая функция (7.2.5) удовлетворяет (5.3.7) и, следовательно,

7.2.2. *Распределение* $N(\mu, \sigma^2)$ *является воспроизводящим по* (μ, σ^2).

На самом деле распределение $N(\mu, \sigma^2)$ обладает более сильным свойством, чем просто воспроизводительность. Это свойство можно сформулировать следующим образом.

7.2.3. *Предположим, что* x_1 *и* x_2 *— независимые случайные величины, имеющие соответственно распределения* $N(\mu_1, \sigma_1^2)$ *и* $N(\mu_2, \sigma_2^2)$. *Пусть* $L = c_1x_1 + c_2x_2$, *где* c_1 *и* c_2 *— вещественные константы, не равные нулю одновременно. Тогда величина* L *имеет распределение* $N(c_1\mu_1 + c_2\mu_2,\ c_1^2\sigma_1^2 + c_2^2\sigma_2^2)$.

Самый быстрый путь проверить это утверждение — найти характеристическую функцию величины L. Согласно (5.3.6),

$$\varphi_L(t) = \varphi_1(c_1t)\cdot\varphi_2(c_2t),$$

где $\varphi_1(t)$, $\varphi_2(t)$ и $\varphi_L(t)$ суть соответственно характеристические функции величин x_1, x_2 и L. Поскольку

$$\varphi_j(c_jt) = e^{itc_j\mu_j - \frac{1}{2}c_j^2\sigma_j^2t^2}, \qquad j = 1,\ 2,$$

мы имеем

$$\varphi_L(t) = \exp\left[i(c_1\mu_1 + c_2\mu_2)t - \frac{1}{2}(c_1^2\sigma_1^2 + c_2^2\sigma_2^2)t^2\right], \tag{7.2.7}$$

что, как видно из (7.2.5), является характеристической функцией нормального распределения $N(c_1\mu_1 + c_2\mu_2,\ c_1^2\sigma_1^2 + c_2^2\sigma_2^2)$. Доказательство теоремы **7.2.3** завершает ссылка на теорему **5.1.3**. Доказанная теорема может быть немедленно перенесена на случай k независимых величин, имеющих нормальное распределение. Это обобщение предоставляется читателю в качестве упражнения.

7.3. Двумерное нормальное распределение

Нормальное распределение занимает такое важное место в теории вероятностей и статистике, что, прежде чем перейти к k-мерному случаю, будет полезно достаточно подробно рассмотреть случай двух измерений. Наибольшее удобство представляет следующая форма ф. п. в. *двумерного нормального распределения*:

$$f(x_1, x_2) = \frac{\sqrt{|\sigma^{ij}|}}{2\pi} e^{-\frac{1}{2} Q(x_1, x_2)}, \tag{7.3.1}$$

где (x_1, x_2) — произвольная точка в R_2 и

$$Q(x_1, x_2) = \sigma^{11}(x_1 - \mu_1)^2 + 2\sigma^{12}(x_1 - \mu_1)(x_2 - \mu_2) + \sigma^{22}(x_2 - \mu_2)^2. \tag{7.3.2}$$

Ниже мы покажем, что μ_1 и μ_2 суть *средние* соответственно величин x_1 и x_2 и матрица $\|\sigma^{ij}\|$, $i, j = 1, 2$, где $\sigma^{12} = \sigma^{21}$, является обратной [см. (3.5.3)] к ковариационной матрице $\|\sigma_{ij}\|$ этих величин. Излишне упоминать, что мы предполагаем матрицу $\|\sigma_{ij}\|$ положительно определенной. Это влечет положительную определенность матрицы $\|\sigma^{ij}\|$, т. е. положительную определенность квадратичной формы $Q(x_1, x_2)$.

Мы считаем удобным обозначить двумерное нормальное распределение, имеющее ф. п. в. (7.3.1), через $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j = 1, 2$.

Сначала проверим, что

$$\int\limits_{R_2} f(x_1, x_2)\, dx_1\, dx_2 = 1, \tag{7.3.3}$$

т. е. убедимся в справедливости соотношения

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} e^{-\frac{1}{2} Q(x_1, x_2)}\, dx_1\, dx_2 = \frac{2\pi}{\sqrt{|\sigma^{ij}|}}. \tag{7.3.4}$$

Полагая $y_1 = x_1 - \mu_1$ и $y_2 = x_2 - \mu_2$, мы можем записать левую часть (7.3.4) в виде

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\sqrt{\sigma^{11}}\,(y_1 + \sigma^{12}/\sigma^{11}\cdot y_2)\right]^2 - \right.$$
$$\left. -\frac{1}{2}\left[y_2\sqrt{|\sigma^{ij}|}/\sigma^{11}\right]^2\right\} dy_1\, dy_2. \tag{7.3.5}$$

Делая преобразование

$$z_1 = \sqrt{\sigma^{11}}\left(y_1 + \frac{\sigma^{12}}{\sigma^{11}} y_2\right), \qquad z_2 = \sqrt{\frac{|\sigma^{ij}|}{\sigma^{11}}}\, y_2, \tag{7.3.6}$$

якобиан которого равен

$$\frac{\partial(y_1, y_2)}{\partial(z_1, z_2)} = \left|\frac{\partial(z_1, z_2)}{\partial(y_1, y_2)}\right|^{-1} = \frac{1}{\sqrt{|\sigma^{ij}|}},$$

мы приводим левую часть (7.3.4) к интегралу

$$\frac{1}{\sqrt{|\sigma^{ij}|}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z_1^2 + z_2^2)}\, dz_1\, dz_2.$$

Но из (7.2.3) и (7.2.3а) мы знаем, что написанный двойной интеграл имеет значение 2π. Таким образом, соотношение (7.3.4) действительно справедливо.

Чтобы показать, что μ_1 и μ_2 суть средние величин x_1 и x_2, продифференцируем обе части (7.3.3) по μ_1 и μ_2. Мы получим

$$\mathscr{E}\,[\sigma^{11}(x_1 - \mu_1) + \sigma^{12}(x_2 - \mu_2)] = 0,$$
$$\mathscr{E}\,[\sigma^{21}(x_1 - \mu_1) + \sigma^{22}(x_2 - \mu_2)] = 0,$$

что можно переписать следующим образом:

$$\begin{aligned} \sigma^{11}\mathscr{E}\,(x_1 - \mu_1) + \sigma^{12}\mathscr{E}\,(x_2 - \mu_2) &= 0, \\ \sigma^{21}\mathscr{E}\,(x_1 - \mu_1) + \sigma^{22}\mathscr{E}\,(x_2 - \mu_2) &= 0. \end{aligned} \qquad (7.3.7)$$

Это — система однородных линейных уравнений относительно $\mathscr{E}\,(x_1 - \mu_1)$ и $\mathscr{E}\,(x_2 - \mu_2)$, и поскольку $|\sigma^{ij}| \neq 0$, она допускает единственное решение

$$\mathscr{E}\,(x_1 - \mu_1) = 0, \qquad \mathscr{E}\,(x_2 - \mu_2) = 0. \qquad (7.3.8)$$

Из этих формул вытекает, что

$$\mu_1 = \mathscr{E}\,(x_1), \qquad \mu_2 = \mathscr{E}\,(x_2), \qquad (7.3.9)$$

т. е. μ_1 и μ_2 *суть средние соответственно величин* x_1 и x_2.

Теперь рассмотрим дисперсии и ковариации. Если мы продифференцируем обе части (7.3.4) по σ^{11} и затем умножим их на $\sqrt{|\sigma^{ij}|}/\pi$, мы получим

$$\frac{\sqrt{|\sigma^{ij}|}}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_1)^2 e^{-\frac{1}{2} Q(x_1, x_2)} dx_1 dx_2 = \frac{\sigma^{22}}{|\sigma^{ij}|}.$$

Но левая часть есть выражение для $\mathscr{E}\,(x_1 - \mu_1)^2$, дисперсии величины x_1. Обозначая дисперсию через σ_{11}, мы имеем

$$\sigma_{11} = \frac{\sigma^{22}}{|\sigma^{ij}|}.$$

Обозначая ковариацию между x_1 и x_2 через σ_{12} и дисперсию величины x_2 через σ_{22}, мы подобным же образом находим

$$\sigma_{12} = -\frac{\sigma^{12}}{|\sigma^{ij}|}, \qquad \sigma_{22} = \frac{\sigma^{11}}{|\sigma^{ij}|}.$$

Следовательно, ковариационная матрица $\|\sigma_{ij}\|$ выражается через параметры σ^{11}, σ^{12}, σ^{22} в (7.3.2) посредством формулы

$$\|\sigma_{ij}\| = \|\sigma^{ij}\|^{-1}. \qquad (7.3.10)$$

Таким образом, мы имеем следующий результат.

7.3.1. *Постоянные* μ_1 и μ_2 *в* (7.3.2) *суть средние величин* x_1 и x_2, *а матрица* $\|\sigma^{ij}\|$ *является обратной к ковариационной матрице* $\|\sigma_{ij}\|$ *этих величин.*

Если мы обозначим дисперсии величин x_1 и x_2 через σ_1^2 и σ_2^2, а ковариацию через σ_1, σ_2, ρ, где ρ — коэффициент корреляции между x_1 и x_2, то двумерную нормальную ф. п. в. (7.3.1) можно будет переписать в другой часто используемой форме:

$$f(x_1,\ x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\,e^{-\frac{1}{2}Q(x_1,\ x_2)}, \tag{7.3.11}$$

где

$$Q(x_1,\ x_2)=\frac{1}{(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-2\rho\,\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]. \tag{7.3.12}$$

Следует отметить, что нормальная ф. п. в. постоянна на любом эллипсе вида $Q(x_1,\ x_2)=\text{const}$. Кроме того, наибольшее значение функции $f(x_1,\ x_2)$ достигается в *центре тяжести* распределения, именно в точке $(\mu_1,\ \mu_2)$. Найдем, далее, характеристическую функцию $\varphi(t_1,\ t_2)$ величины $(x_1,\ x_2)$. Очевидно,

$$\varphi(t_1,\ t_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{it_1x_1+it_2x_2}f(x_1,\ x_2)\,dx_1\,dx_2. \tag{7.3.13}$$

Снова полагая $y_1=x_1-\mu_1$ и $y_2=x_2-\mu_2$, мы получаем

$$\varphi(t_1,\ t_2)=\frac{e^{i\mu_1t_1+i\mu_2t_2}}{2\pi\sqrt{|\sigma_{ij}|}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{1}{2}Q'(y_1,\ y_2)}\,dy_1\,dy_2, \tag{7.3.14}$$

где

$$Q'(y_1,\ y_2)=\sigma^{11}y_1^2+\sigma^{22}y_2^2+2\sigma^{12}y_1y_2-2it_1y_1-2it_2y_2.$$

Но $Q'(y_1,\ y_2)$ можно записать в виде

$$Q'(y_1,\ y_2)=z_1^2+z_2^2+(\sigma_{11}t_1^2+2\sigma_{12}t_1t_2+\sigma_{22}t_2^2), \tag{7.3.15}$$

где

$$z_1=\sqrt{\sigma^{11}}\left(y_1+\frac{\sigma^{12}}{\sigma^{11}}y_2-\frac{it_1}{\sigma^{11}}\right),$$

$$z_2=\sqrt{\frac{|\sigma^{ij}|}{\sigma^{11}}}\left(y_2+\frac{i\sigma^{12}t_1-i\sigma^{11}t_2}{|\sigma^{ij}|}\right).$$

Якобиан преобразования, которое связывает $(y_1,\ y_2)$ и $(z_1,\ z_2)$, равен $\left|\frac{\partial(y_1,\ y_2)}{\partial(z_1,\ z_2)}\right|=\frac{1}{\sqrt{|\sigma^{ij}|}}$.

Используя (7.2.3) и (7.2.3а), мы из (7.3.14) находим, что

7.3.2. *Характеристическая функция двумерного нормального распределения* (7.3.1) *или* (7.3.11) *дается формулой*

$$\varphi(t_1,\ t_2)=\exp\left[i(\mu_1t_1+\mu_2t_2)-\frac{1}{2}(\sigma_{11}t_1^2+\sigma_{22}t_2^2+2\sigma_{12}t_1t_2)\right]. \tag{7.3.16}$$

Мы должны обратить внимание, что обе переменные z_1 и z_2 являются комплексными и каждый из интегралов $\int e^{-\frac{z_1^2}{2}}dz_1$ и $\int e^{-\frac{z_2^2}{2}}dz_2$, полученных в результате перехода от (y_1, y_2) к (z_1, z_2), фактически берется вдоль прямой в комплексной плоскости, параллельной вещественной оси. Но все такие интегралы совпадают с интегралом, взятым вдоль вещественной оси.

Следует, в частности, отметить, что матрица квадратичной формы переменных t_1 и t_2 в выражении для характеристической функции (7.3.16) есть ковариационная матрица $\|\sigma_{ij}\|$, в то время как матрица квадратичной формы переменных x_1 и x_2 в выражении для ф. п. в. (7.3.1) есть матрица $\|\sigma^{ij}\|$, обратная ковариационной.

Для читателя будет поучительно проверить, применяя обычную процедуру дифференцирования к (7.3.16), что μ_1 и μ_2 — средние величин x_1 и x_2, а $\|\sigma_{ij}\|$ — их ковариационная матрица.

Если в формуле (7.3.16) мы положим $t_2=0$, то получим характеристическую функцию величины x_1, именно

$$\varphi(t_1, 0)=e^{i\mu_1 t_1-\frac{1}{2}\sigma_{11}t_1^2}, \tag{7.3.17}$$

откуда в силу теоремы **7.2.1** вытекает

7.3.3. *Маргинальное распределение величины x_1 для распределения $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j=1, 2$, является распределением $N(\mu_1, \sigma_{11})$.*

Аналогичное утверждение, конечно, справедливо и для маргинального распределения величины x_2.

В более общем случае характеристическая функция величины $L=c_1x_1+c_2x_2$, где постоянные c_1 и c_2 обе не равны нулю, дается формулой

$$\varphi(c_1t, c_2t)=\exp\left[i(c_1\mu_1+c_2\mu_2)t-\frac{1}{2}(\sigma_{11}c_1^2+2\sigma_{12}c_1c_2+\sigma_{22}c_2^2)t^2\right], \tag{7.3.18}$$

и мы имеем следующий результат:

7.3.4. *Если x_1 и x_2 — случайные величины, имеющие распределение $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j=1, 2$, то величина $L=c_1x_1+c_2x_2$ имеет распределение $N\left(\sum_{i=1}^{2} c_i\mu_i, \sum_{i,j=1}^{2}\sigma_{ij}c_ic_j\right)$.*

Заметим, что теорема **7.2.3** есть частный случай теоремы **7.3.4**, возникающий, если $\sigma_{12}=0$ при обозначениях $\sigma_{11}=\sigma_1^2$ и $\sigma_{22}=\sigma_2^2$.

Рассмотрим теперь условную случайную величину $x_2 \mid x_1$, когда (x_1, x_2) имеет двумерное нормальное распределение. Маргинальная ф. п. в. величины x_1 равна

$$f_1(x_1)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{1}{2}\frac{(x_1-\mu_1)^2}{\sigma_1^2}}, \tag{7.3.19}$$

где $\sigma_1^2=\sigma_{11}=\sigma^2(x_1)$.

Используя выражение (7.3.11) двумерной нормальной ф. п. в., мы из (2.9.13) получаем

$$f(x_2\,|\,x_1)=\frac{1}{\sqrt{2\pi}\,\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2}Q(x_1,\,x_2)+\frac{1}{2}\frac{(x_1-\mu_1)^2}{\sigma_1^2}},$$

где $Q(x_1, x_2)$ определяется равенством (7.3.12). После простых алгебраических преобразований мы находим

$$f(x_2\,|\,x_1)=\frac{1}{\sqrt{2\pi}\,\sigma_2\sqrt{1-\rho^2}}\times$$
$$\times\exp\left\{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left[(x_2-\mu_2)-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)\right]^2\right\}, \qquad (7.3.20)$$

и, таким образом,

7.3.5. *Если* (x_1, x_2) — *случайная величина, имеющая двумерное нормальное распределение с ф. п. в.* (7.3.11), *то условная случайная величина* $x_2\,|\,x_1$ *имеет распределение*

$$N\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1),\ \sigma_2^2(1-\rho^2)\right).$$

Из соображений симметрии ясно, что аналогичное утверждение имеет место и для условного распределения величины $x_1\,|\,x_2$.

Следует обратить внимание, что среднее значение величины $x_2\,|\,x_1$, т. е. функция регрессии величины x_2 на x_1, в случае ф. п. в. (7.3.20) является линейной функцией от x_1, именно

$$\mu(x_2\,|\,x_1)=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1).$$

Отсюда уравнение *линии регрессии величины* x_2 *на* x_1,

$$x_2=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1), \qquad (7.3.21)$$

как легко видеть, в точности совпадает с уравнением (3.8.6) прямой средней квадратической регрессии величины x_2 на x_1. Все это подтверждает замечание, сделанное в § 3.8 (a), которое мы теперь можем сформулировать более точно.

7.3.6. *Двумерное нормальное распределение обладает тем свойством, что обе его функции регрессии линейны и тождественно совпадают с прямыми регрессии, полученными по методу наименьших квадратов. Кроме того, дисперсии условных случайных величин* $x_2\,|\,x_1$ и $x_1\,|\,x_2$ *совпадают соответственно со средними квадратическими остаточными дисперсиями* $\sigma_{2\,.\,1}^2$ и $\sigma_{1\,.\,2}^2$, *определенными формулой* (3.8.8).

7.4. Многомерное нормальное распределение

(а) Структура ф. п. в. k-мерного нормального распределения. Многомерное, или k-мерное, нормальное распределение и его свойства получаются в результате прямого обобщения случая $k=2$, рассмотренного в § 7.3.

Ф. п. в. k-мерного нормального распределения задается соотношением

$$f(x_1, \ldots, x_k) = \frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{\frac{1}{2}k}} e^{-\frac{1}{2} Q(x_1, \ldots, x_k)}, \tag{7.4.1}$$

где $(x_1, \ldots, x_k)$ — любая точка в R_k и

$$Q(x_1, \ldots, x_k) = \sum_{i, j=1} \sigma^{ij}(x_i - \mu_i)(x_j - \mu_j), \tag{7.4.2}$$

причем $\|\sigma^{ij}\|$ — *матрица, обратная ковариационной матрице* $\|\sigma_{ij}\|$ и μ_i — *средние величин* x_i.

Мы считаем, что матрица $\|\sigma^{ij}\|$ положительно определена, т. е. $|\sigma^{ij}| \neq 0$, а следовательно, и $|\sigma_{ij}| \neq 0$. Распределение, имеющее ф. п. в. (7.4.1), будет называться *k-мерным нормальным распределением* и будет обозначаться $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j = 1, \ldots, k$. Выборочное пространство соответствующей случайной величины $(x_1, \ldots, x_k)$ совпадает со всем k-мерным пространством R_k.

Пусть $y_i = x_i - \mu_i$, $i = 1, \ldots, k$. Тогда, чтобы показать, что интеграл от функции (7.4.1) по пространству R_k равен единице, нужно убедиться в справедливости равенства

$$\int_{R_k} \exp\left(-\frac{1}{2} \sum_{i, j=1}^{k} \sigma^{ij} y_i y_j\right) dy_1 \ldots dy_k = \frac{(2\pi)^{\frac{1}{2}k}}{\sqrt{|\sigma^{ij}|}}. \tag{7.4.3}$$

Очевидно,

$$\sum_{i, j=1}^{k} \sigma^{ij} y_i y_j = \sigma^{11}\left(y_1 + \frac{\sum_{j=2}^{k} \sigma^{1j} y_j}{\sigma^{11}}\right)^2 + \sum_{i, j=2}^{k}\left(\sigma^{ij} - \frac{\sigma^{1j}\sigma^{1i}}{\sigma^{11}}\right) y_i y_j.$$

Полагая

$$z_1 = \sqrt{\sigma^{11}}\left(y_1 + \frac{1}{\sigma^{11}} \sum_{j=2}^{k} \sigma^{1j} y_j\right), \quad \sigma_{(1)}^{ij} = \sigma^{ij} - \frac{\sigma^{1i}\sigma^{1j}}{\sigma^{11}},$$

$$i, j = 2, \ldots, k,$$

мы имеем

$$\sum_{i, j=1}^{k} \sigma^{ij} y_i y_j = z_1^2 + \sum_{i, j=2}^{k} \sigma_{(1)}^{ij} y_i y_j.$$

Продолжая этот процесс и вводя обозначения

$$\sigma_{(p)}^{ij} = \sigma_{(p-1)}^{ij} - \frac{\sigma_{(p-1)}^{pi}\sigma_{(p-1)}^{pj}}{\sigma_{(p-1)}^{pp}}, \qquad i, j = p+1, \ldots, k; \ p = 1, \ldots, k-1,$$

причем

$$\sigma_{(0)}^{ij} = \sigma^{ij}, \qquad i, j = 1, \ldots, k,$$

мы приходим к равенству

$$\sum_{i,\,j=1}^{k} \sigma^{ij} y_i y_j = \sum_{i=1}^{k} z_i^2,$$

где

$$z_i = \sqrt{\sigma^{ii}_{(i-1)}}\left(y_i + \frac{1}{\sigma^{ii}_{(i-1)}} \sum_{j=i+1}^{k} \sigma^{ij}_{(i-1)} y_j\right), \qquad i = 1, \ldots, k. \tag{7.4.4}$$

Можно проверить, что поскольку $\|\sigma^{ij}\|$ — положительно определенная матрица, все величины σ^{11}, $\sigma^{22}_{(1)}$, ..., $\sigma^{kk}_{(k-1)}$ положительны.

Процедура, которую мы только что описали, фактически доставляет линейное преобразование, приводящее положительно определенную квадратичную форму, фигурирующую в (7.4.3), к сумме квадратов, и известна как метод Лагранжа. Заметим, что существует целое семейство линейных преобразований, обладающих этим свойством.

Якобиан преобразования (7.4.4) есть

$$\left|\frac{\partial (y_1, \ldots, y_k)}{\partial (z_1, \ldots, z_k)}\right| = \left|\frac{\partial (z_1, \ldots, z_k)}{\partial (y_1, \ldots, y_k)}\right|^{-1} = \frac{1}{\sqrt{\sigma^{11}\sigma^{22}_{(1)} \ldots \sigma^{kk}_{(k-1)}}},$$

вследствие чего левая часть (7.4.3) принимает вид

$$\frac{1}{\sqrt{\sigma^{11}\sigma^{22}_{(1)} \ldots \sigma^{kk}_{(k-1)}}} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\sum_{i=1}^{k} z_i^2\right) dz_1 \ldots dz_k. \tag{7.4.5}$$

Применяя формулу (7.2.3а), мы видим, что k-кратный интеграл (7.4.5) равен $(2\pi)^{\frac{k}{2}}$. Наша задача сводится, таким образом, к доказательству соотношения

$$\sigma^{11}\sigma^{22}_{(1)} \ldots \sigma^{kk}_{(k-1)} = |\sigma^{ij}|. \tag{7.4.6}$$

Но определитель $|\sigma^{ij}|$ можно записать следующим образом:

$$|\sigma^{ij}| = \begin{vmatrix} \sigma^{11} & \sigma^{12} & \ldots & \sigma^{1k} \\ \sigma^{21} & \sigma^{22} & \ldots & \sigma^{2k} \\ \cdot & \cdot & \cdot & \cdot \\ \sigma^{k1} & \sigma^{k2} & \ldots & \sigma^{kk} \end{vmatrix} = \sigma^{11} \begin{vmatrix} 1 & \sigma^{12} & \ldots & \sigma^{1k} \\ \dfrac{\sigma^{21}}{\sigma_1} & \sigma^{22} & \ldots & \sigma^{2k} \\ \cdot & \cdot & \cdot & \cdot \\ \dfrac{\sigma^{k1}}{\sigma^{11}} & \sigma^{k2} & \ldots & \sigma^{kk} \end{vmatrix}. \tag{7.4.7}$$

Умножим 1-й столбец определителя, стоящего в последней части равенства (7.4.7), на σ^{12} и вычтем из 2-го столбца, затем умножим 1-й столбец на σ^{13} и вычтем из 3-го и т. д. В результате мы получим

$$|\sigma^{ij}| = \sigma^{11} \begin{vmatrix} \sigma^{22}_{(1)} & \sigma^{23}_{(1)} & \ldots & \sigma^{2k}_{(1)} \\ \sigma^{32}_{(1)} & \sigma^{33}_{(1)} & \ldots & \sigma^{3k}_{(1)} \\ \cdot & \cdot & \cdot & \cdot \\ \sigma^{k2}_{(1)} & \sigma^{k3}_{(1)} & \ldots & \sigma^{kk}_{(1)} \end{vmatrix}. \tag{7.4.8}$$

Продолжая этот процесс далее, мы находим

$$|\sigma^{ij}| = \sigma^{11}\sigma^{22}_{(1)} \ldots \sigma^{kk}_{(k-1)},$$

что и требовалось доказать.

Теперь убедимся в том, что μ_i — действительно средние случайных величин, имеющих ф. п. в. (7.4.1). Для этого, как в двумерном случае, возьмем частные производные по μ_i, $i = 1, \ldots, k$, от обеих частей тождества

$$\frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{\frac{1}{2}k}} \int\limits_{R_k} e^{-\frac{1}{2}Q(x_1, \ldots, x_k)} dx_1 \ldots dx_k = 1.$$

Это даст k уравнений

$$\mathscr{E}\left[\sum_{j=1}^{k} \sigma^{ij}(x_j - \mu_j)\right] = 0, \qquad i = 1, \ldots, k,$$

которые можно переписать в виде

$$\sum_{j=1}^{k} \sigma^{ij}\, \mathscr{E}(x_j - \mu_j) = 0. \qquad (7.4.9)$$

Так как мы предположили, что матрица $\|\sigma_{ij}\|$, а следовательно и $\|\sigma^{ij}\|$, положительно определена, то $|\sigma^{ij}| = 0$, и поэтому (7.4.9) имеет только нулевое решение

$$\mathscr{E}(x_i - \mu_i) = 0 \quad \text{или} \quad \mathscr{E}(x_i) = \mu_i, \qquad i = 1, \ldots, k, \qquad (7.4.10)$$

т. е. числа $\mu_1, \ldots, \mu_k$ в формуле (7.4.1) *суть средние соответственно величин* $x_1, \ldots, x_k$.

Для того чтобы проверить, что $\|\sigma^{ij}\|$ — матрица, обратная к ковариационной, мы воспользуемся соотношением

$$\int\limits_{R_k} e^{-\frac{1}{2}Q(x_1, \ldots, x_k)} dx_1 \ldots dx_k = \frac{(2\pi)^{\frac{1}{2}k}}{\sqrt{|\sigma^{ij}|}}. \qquad (7.4.11)$$

Дифференцируя обе части (7.4.11) по σ^{ij} и затем умножая на $-(1+\delta_{ij}) \times \times \frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{\frac{k}{2}}}$, где δ_{ij} — символ Кронекера, мы получаем формулу

$$\mathscr{E}[(x_i - \mu_i)(x_j - \mu_j)] = \sigma_{ij}, \qquad (7.4.12)$$

справедливую как при $i = j$, так и при $i \neq j$. Таким образом, имеет место теорема

7.4.1. *Константы* $\mu_1, \ldots, \mu_k$ *в выражении* (7.4.1) *являются средними величин* $x_1, \ldots, x_k$, *а матрица* $\|\sigma^{ij}\|$ *квадратичной формы — обратной ковариационной матрице* $\|\sigma_{ij}\|$ *этих величин.*

Если мы обозначим дисперсии величин x_i через σ_i^2 и ковариации между x_i и x_j через $\sigma_i\sigma_j\rho_{ij}$, то (7.4.1) можно будет записать в форме, впрочем, довольно громоздкой, представляющей собой k-мерный аналог (7.3.12). Оставляем это читателю в качестве упражнения.

(b) Характеристическая функция k-мерного нормального распределения. Рассмотрим теперь характеристическую функцию k-мерного нормального распределения. Мы имеем

$$\varphi(t_1, \ldots, t_k) = \exp\left\{ i\sum_{i=1}^{k} \mu_i t_i \right\} \cdot \frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{\frac{1}{2}k}} \times$$
$$\times \int\limits_{R_k} \exp\left\{ -\frac{1}{2}\left[Q(x_1, \ldots, x_k) - 2i\sum_{i=1}^{k}(x_i - \mu_i)t_i \right] \right\} dx_1 \ldots dx_k, \quad (7.4.13)$$

где $Q(x_1, \ldots, x_k)$ дается формулой (7.4.2).

Чтобы вычислить интеграл (7.4.13), осуществим k-мерный вариант «дополнения до полного квадрата», которое мы использовали при переходе от (7.2.4) к (7.2.4a). С этой целью, положив $y_i = x_i - \mu_i$, $i = 1, \ldots, k$, заметим, что

$$\sum_{i,\,j=1}^{k} \sigma^{ij}\left[y_i - i\sum_{h=1}^{k}\sigma_{ih}t_h\right]\left[y_j - i\sum_{g=1}^{k}\sigma_{gj}t_g\right] \equiv$$
$$\equiv \sum_{i,\,j=1}^{k}\sigma^{ij}y_iy_j - i\sum_{i,\,j,\,h=1}^{k}\sigma^{ij}\sigma_{ih}t_hy_j - i\sum_{i,\,j,\,g=1}^{k}\sigma^{ij}\sigma_{gj}t_gy_i -$$
$$- \sum_{i,\,j,\,h,\,g=1}^{k}\sigma^{ij}\sigma_{ih}\sigma_{gj}t_gt_h. \quad (7.4.14)$$

Так как $\sigma^{ij} = \sigma^{ji}$ и, разумеется, $\sigma_{ij} = \sigma_{ji}$, суммы $\sum_{i=1}^{k}\sigma^{ij}\sigma_{ih}$ и $\sum_{j=1}^{k}\sigma^{ij}\sigma_{gj}$ равны соответственно δ_{jh} и δ_{ig}, и поэтому правая часть равенства (7.4.14) сводится к выражению

$$Q(x_1, \ldots, x_k) - 2i\sum_{i=1}^{k}t_iy_i - \sum_{i,\,j=1}^{k}\sigma_{ij}t_it_j.$$

Следовательно,

$$Q(x_1, \ldots, x_k) - 2i\sum_{i=1}^{k}t_iy_i = Q'(x_1, \ldots, x_k) + \sum_{i,\,j=1}^{k}\sigma_{ij}t_it_j, \quad (7.4.15)$$

где через $Q'(x_1, \ldots, x_k)$ обозначена квадратичная форма, стоящая в левой части (7.4.14). Пользуясь этим, (7.4.13) можно записать в виде

$$\varphi(t_1, \ldots, t_k) = \exp\left(i\sum_{i=1}^{k}\mu_it_i - \frac{1}{2}\sum_{i,\,j=1}^{k}\sigma_{ij}t_it_j \right) H, \quad (7.4.16)$$

где

$$H = \frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{\frac{1}{2}k}} \int\limits_{R_k} e^{-\frac{1}{2}Q'(x_1, \ldots, x_k)} dx_1 \ldots dx_k.$$

$Q'(x_1, \ldots, x_k)$ есть квадратичная форма комплексных переменных $(x_i - \mu_i) + iB_i$, $i = 1, \ldots, k$, где B_i — вещественные числа, с матрицей $\|\sigma^{ij}\|$, тогда как $Q(x_1, \ldots, x_k)$ — квадратичная форма, получаемая из $Q'(x_1, \ldots, x_k)$ при $B_i = 0$. Далее, если мы обозначим $(x_i - \mu_i) + iB_i$ через y_i' и в (7.4.4) заменим y_i на y_i', то $Q'(x_1, \ldots, x_k)$, очевидно, приведется этим преобразованием к сумме квадратов $\sum\limits_{i=1}^{k} z_i^2$. Величины z_i в этом случае будут, конечно, комплексными переменными, и интегрирование по каждому z_i будет вестись параллельно вещественной оси этого переменного. Но, как мы указывали для одномерного и двумерного нормальных распределений, такие интегралы равны интегралам от тех же самых функций вдоль соответствующих вещественных осей. Следовательно, H сохраняет свою величину при замене $Q'(x_1, \ldots, x_k)$ на $Q(x_1, \ldots, x_k)$, т. е. $H = 1$. Таким образом,

7.4.2. *Характеристическая функция распределения $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j = 1, \ldots, k$, равна*

$$\varphi(t_1, \ldots, t_k) = \exp\left(i \sum_{i=1}^{k} \mu_i t_i - \frac{1}{2} \sum_{i,\, j=1}^{k} \sigma_{ij} t_i t_j\right). \tag{7.4.17}$$

Положив в (7.4.17) $t_{k_1+1} = \ldots = t_k = 0$, мы будем иметь

$$\varphi(t_1, \ldots, t_{k_1}, 0, \ldots, 0) = \exp\left(i \sum_{i=1}^{k_1} \mu_i t_i - \frac{1}{2} \sum_{i,\, j=1}^{k_1} \sigma_{ij} t_i t_j\right). \tag{7.4.18}$$

Следовательно,

7.4.3. *Если $(x_1, \ldots, x_k)$ — векторная случайная величина, имеющая k-мерное нормальное распределение $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j = 1, \ldots, k$, то маргинальное распределение величины $(x_1, \ldots, x_{k_1})$, $k_1 < k$, является k_1-мерным нормальным распределением, $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j = 1, \ldots, k_1$.*

(с) Распределение линейных функций нормальных случайных величин. Если в (7.4.17) мы положим $t_i = c_i t$, $i = 1, \ldots, k$, где константы c_i не равны нулю одновременно, то получим выражение для характеристической функции $\varphi(t)$ линейной функции $L = \sum\limits_{i=1}^{k} c_i x_i$:

$$\varphi(t) = \exp\left[i\left(\sum_{i=1}^{k} c_i \mu_i\right) t - \frac{1}{2}\left(\sum_{i,\, j=1}^{k} \sigma_{ij} c_i c_j\right) t^2\right]. \tag{7.4.19}$$

Применяя теперь теорему **7.2.1**, мы приходим к следующему результату:

7.4.4. *Если* $(x_1, \ldots, x_k)$ *имеет k-мерное распределение* $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j = 1, \ldots, k$, *то форма* $L = c_1x_1 + \ldots + c_kx_k$ *имеет распределение*

$$N\Big(\sum_{i=1}^{k} c_i\mu_i,\ \sum_{i,\,j=1}^{k} \sigma_{ij}c_ic_j\Big).$$

Аналогичным образом можно установить более общую теорему

7.4.5. *Если* $L_p = c_{p1}x_1 + \ldots + c_{pk}x_k$, $p = 1, \ldots, s$, $s \leqslant k$, — *линейно независимые случайные величины и* $(x_1, \ldots, x_k)$ *имеет распределение* $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j = 1, \ldots, k$, *то величина* $(L_1, \ldots, L_s)$ *имеет s-мерное распределение*

$$N\Big(\Big\{\sum_{i=1}^{k} c_{pi}\mu_i\Big\},\ \Big\|\sum_{i,\,j=1}^{k} c_{ij}c_{pi}c_{qj}\Big\|\Big), \qquad p, q = 1, \ldots, s.$$

(d) Условные распределения, связанные с k-мерным нормальным распределением. В заключение рассмотрим ф. п. в. условной случайной величины $x_k \mid x_1, \ldots, x_{k-1}$ для k-мерного нормального распределения. По определению

$$f(x_k \mid x_1, \ldots, x_{k-1}) = \frac{f(x_1, \ldots, x_k)}{f_{12\ldots(k-1)}(x_1, \ldots, x_{k-1})}, \tag{7.4.20}$$

где $f(x_1, \ldots, x_k)$ — ф. п. в. (7.4.1) и $f_{12\ldots(k-1)}(x_1, \ldots, x_{k-1})$ — маргинальная ф. п. в. величины $(x_1, \ldots, x_{k-1})$, т. е.

$$f_{12\ldots(k-1)}(x_1, \ldots, x_{k-1}) = \\ = \frac{\sqrt{|\sigma_{(k)}^{pq}|}}{(2\pi)^{\frac{1}{2}(k-1)}} \exp\Big[-\frac{1}{2}\sum_{p,\,q=1}^{k-1} \sigma_{(k)}^{pq}(x_p - \mu_p)(x_q - \mu_q)\Big], \tag{7.4.21}$$

причем $\|\sigma_{(k)}^{pq}\|$ означает матрицу, обратную $\|\sigma_{pq}\|$, $p, q = 1, \ldots, k-1$. Заметим, что в случае матрицы $\|\sigma_{ij}\|$ и обратной ей матрицы $\|\sigma^{ij}\|$ мы имеем $i, j = 1, \ldots, k$.

Пусть $y_i = x_i - \mu_i$, $i = 1, \ldots, k$. Легко можно проверить тождество

$$\sum_{i,\,j=1}^{k} \sigma^{ij}y_iy_j \equiv \sigma^{kk}\Big[y_k - \sum_{p=1}^{k-1} \beta_q^* y_p\Big]^2 + \sum_{p,\,q=1}^{k-1} \sigma_{(k)}^{pq}y_py_q, \tag{7.4.22}$$

где

$$\beta_p^* = \sum_{q=1}^{k-1} \sigma_{(k)}^{pq}\sigma_{qp}, \qquad p = 1, \ldots, k-1. \tag{7.4.23}$$

Кроме того,

$$\frac{|\sigma_{ij}|}{|\sigma_{(k)}^{pq}|} = \frac{|\sigma_{pq}|}{|\sigma_{ij}|} = \sigma^{kk}, \tag{7.4.24}$$

где $i, j = 1, \ldots, k$ и $p, q = 1, \ldots, k-1$.

Подставив выражения для $f(x_1, \ldots, x_k)$ и $f_{12\ldots(k-1)}(x_1, \ldots, x_{k-1})$ в (7.4.20) и воспользовавшись (7.4.22) и (7.4.24), мы найдем

$$f(x_k \mid x_1, \ldots, x_{k-1}) = \\ = \frac{\sqrt{\sigma^{kk}}}{\sqrt{2\pi}} \exp\left\{-\frac{\sigma^{kk}}{2}\left[(x_k - \mu_k) - \sum_{p=1}^{k-1} \beta_p^* (x_p - \mu_p)\right]^2\right\}. \quad (7.4.25)$$

Последняя формула приводит нас к теореме

7.4.6. *Если* $(x_1, \ldots, x_k)$ — *векторная случайная величина, имеющая k-мерное распределение* $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j = 1, \ldots, k$, *то распределение условной случайной величины* $x_k \mid x_1, \ldots, x_{k-1}$ *есть*

$$N\left(\mu_k + \sum_{p=1}^{k-1} \beta_p^* (x_p - \mu_p), \frac{1}{\sigma^{kk}}\right),$$

где величины β_p^* *даются формулой* (7.4.23).

Следует отметить, что $\mu(x_k \mid x_1, \ldots, x_{k-1})$ для распределения (7.4.25) является линейной функцией от $x_1, \ldots, x_{k-1}$ и тождественно совпадает с правой частью выражения (3.8.18), определяющего плоскость средней квадратической регрессии величины x_k на $x_1, \ldots, x_{k-1}$. Кроме того, $\sigma^2(x_k \mid x_1, \ldots, x_{k-1})$ в этом случае имеет значение, равное $\frac{1}{\sigma^{kk}}$, что в свою очередь совпадает с величиной средней квадратической остаточной дисперсии x_k на $x_1, \ldots, x_{k-1}$, определяемой формулой (3.8.25).

Таким образом, мы установили следующий важный факт.

7.4.7. *В случае k-мерного нормального распределения среднее и дисперсия условной случайной величины $x_k \mid x_1, \ldots, x_{k-1}$ совпадают соответственно с функцией средней квадратической регрессии и средней квадратической остаточной дисперсией величины x_k на $x_1, \ldots, x_{k-1}$.*

7.5. Гамма-распределение

Обычно *гамма-функция* $\Gamma(g)$ определяется для комплексных чисел g, вещественная часть которых положительна, с помощью интеграла

$$\Gamma(g) = \int_0^\infty x^{g-1} e^{-x} dx. \quad (7.5.1)$$

На самом деле мы будем интересоваться только вещественными и положительными значениями g.

В результате интегрирования по частям мы находим

$$\Gamma(g) = (g-1)\Gamma(g-1),$$

откуда следует, что если g — положительное целое число, то

$$\Gamma(g) = (g-1)!.$$

Если $g > 0$, но не является целым числом, то

$$\Gamma(g) = (g-1)(g-2)\ldots\delta\Gamma(\delta),$$

где $0 < \delta < 1$. В частности, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. В этом можно убедиться, заметив, что $I = 2\pi\Gamma\left(\frac{1}{2}\right)$, где I берется из формулы (7.2.3).

В задачах математической статистики наиболее важными значениями g, как мы увидим в последующих главах, являются значения, кратные $\frac{1}{2}$.

В задачах, где рассматриваются суммы квадратов и квадратичные формы нормально распределенных случайных величин, часто появляется распределение, которое имеет ф. п. в.

$$f(x) = \begin{cases} \dfrac{x^{\mu-1}e^{-x}}{\Gamma(x)}, & x > 0, \\ 0, & x \leqslant 0 \end{cases} \tag{7.5.2}$$

при $\mu > 0$. Будет удобно называть это распределение *гамма-распределением* $G(\mu)$. Его также называют распределением Пирсона типа III.

r-ый момент гамма-распределения равен

$$\mu'_r = \frac{1}{\Gamma(\mu)} \int_0^\infty x^{\mu+r-1}e^{-x}\,dx = \frac{\Gamma(\mu+r)}{\Gamma(\mu)}, \tag{7.5.3}$$

откуда

$$\mathscr{E}(x) = \mu, \qquad \sigma^2(x) = \mu, \tag{7.5.4}$$

т. е. *как среднее, так и дисперсия гамма-распределения равны* μ. Как мы помним, этим свойством обладает также распределение Пуассона (6.4.1), которое, конечно, является дискретным распределением.

Полагая в (7.5.2) $\mu = p$, мы получаем функцию

$$I(u, p) = \frac{1}{\Gamma(p+1)} \int_0^{u\sqrt{p+1}} x^p e^{-x}\,dx,$$

названную Карлом Пирсоном (1905) *неполной гамма-функцией*. Он протабулировал (1922) ее для комбинаций значений u и p при u от 0 до 12 и p от 0 до 50.

Рассмотрим характеристическую функцию гамма-распределения $G(\mu)$:

$$\varphi(t) = \frac{1}{\Gamma(\mu)} \int_0^\infty x^{\mu-1}e^{-x(1-it)}\,dx. \tag{7.5.5}$$

Полагая $x(1-it)=y$, мы находим

$$\varphi(t)=(1-it)^{\mu}. \tag{7.5.6}$$

Если считать μ параметром гамма-распределения, то, очевидно, характеристическая функция удовлетворяет (5.3.7) и, следовательно,

7.5.1. *Гамма-распределение* $G(\mu)$ *является воспроизводящим по* μ.

Как мы уже указали, одно из важнейших приложений гамма-распределения связано с рассмотрением задач нахождения распределений некоторых квадратичных форм нормально распределенных случайных величин. В общем случае k-мерной нормально распределенной случайной величины, имеющей ф. п. в. (7.4.1), представляет интерес квадратичная форма, стоящая в экспоненте самой ф. п. в., именно форма $Q(x_1, \ldots, x_k)$, определяемая (7.4.2).

Рассмотрим характеристическую функцию формы $\frac{1}{2}Q(x_1, \ldots, x_k)$

$$\varphi(t)=\frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{\frac{1}{2}k}}\int\limits_{R_k} e^{-\frac{1}{2}(1-it)Q(x_1,\ldots,x_k)}\,dx_1\ldots dx_k. \tag{7.5.7}$$

Сделав преобразование $y_i=\sqrt{1-it}\,(x_i-\mu_i)$, $i=1, \ldots, k$, мы приведем (7.5.7) к виду

$$\varphi(t)=\frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{\frac{1}{2}k}}(1-it)^{-\frac{1}{2}k}\int\limits_{R_k}\exp\left(-\frac{1}{2}\sum_{i,j=1}^{k}\sigma^{ij}y_iy_j\right)dy_1\ldots dy_k. \tag{7.5.8}$$

Используя теперь (7.4.3), мы получаем

$$\varphi(t)=(1-it)^{-\frac{1}{2}k}, \tag{7.5.9}$$

что является характеристической функцией гамма-распределения $G\left(\frac{k}{2}\right)$. Таким образом,

7.5.2. *Если* $(x_1, \ldots, x_k)$ — *векторная случайная величина, имеющая* k-*мерное распределение* $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j=1, \ldots, k$, *то форма* $\frac{1}{2}Q(x_1, \ldots, x_k)$ *имеет гамма-распределение* $G\left(\frac{k}{2}\right)$.

Задачи другого типа, в которых встречается гамма-распределение, связаны с непрерывным распределением времени ожидания. Как мы знаем из § 6.5, (6.5.14) представляет собой вероятность того, что необходимо осуществить x испытаний, прежде чем получить k событий C. Теперь предположим, что события C могут наступать в любой момент времени, и найдем, как долго мы должны ожидать, чтобы получить k событий C. Пусть $(0, t)$ — интервал времени, в течение которого событие C наступило точно $(k-1)$ раз, и $(t, t+\Delta t)$ — интервал, в течение которого C наступило k-й раз. Разобьем интервал $(0, t)$ на $\frac{t}{\Delta t}$ интервалов длины Δ. Условимся считать, что вероятность наступления C в течение заданного интервала времени Δt равна

$\lambda\Delta t + o(\Delta t)$, где λ — некоторая постоянная, именно равная *среднему числу наступлений C в единицу времени.* Кроме того, предположим, что наступление любого числа событий C в интервале I не зависит от наступления какого-либо числа событий C в интервале I', не перекрывающемся с I. Мы хотим определить вероятность того, что необходимо будет ждать в течение времени от t до $t+\Delta t$, чтобы наступило последнее из k событий C. Эта вероятность равна произведению вероятности наступления $(k-1)$ событий C в $\frac{t}{\Delta t}$ интервалах, на которые мы разбили $(0, t)$, и вероятности наступления k-го события C в течение интервала $(t, t+\Delta t)$. С точностью до членов порядка Δt мы получаем эту вероятность, полагая в (6.5.15) $x=\frac{t}{\Delta t}$ и $p=\lambda\Delta t$. Это дает

$$\binom{\frac{t}{\Delta t}-1}{k-1}(\lambda\Delta t)^k(1-\lambda\Delta t)^{(t/\Delta t)-k}=$$
$$=\frac{(t-\Delta t)(t-2\Delta t)\dots(t-(k-1)\Delta t)}{(k-1)!}\lambda^{k-1}(1-\lambda\Delta t)^{(t/\Delta t)-k}\lambda\Delta t. \qquad (7.5.10)$$

Разделив (7.5.10) на Δt и устремив Δt к нулю, мы находим ф. п. в.

$$f(t)=\begin{cases}\dfrac{\lambda(\lambda t)^{k-1}e^{-\lambda t}}{(k-1)!}, & t>0,\\ 0, & t\leqslant 0.\end{cases} \qquad (7.5.11)$$

Функция $f(t)$ при тех предположениях, которые были сделаны, является ф. п. в. времени ожидания, необходимого для того, чтобы наступило точно k событий C.

Из (7.5.11) ясно, что λt есть *случайная величина, имеющая гамма-распределение $G(k)$.*

Наконец, будет полезно сформулировать следующий результат, оставляя его проверку читателю.

7.5.3. *Если x — случайная величина, имеющая равномерное распределение $R\left(\frac{1}{2}, 1\right)$, то случайная величина $y=-\ln x$ имеет гамма-распределение $G(1)$.*

7.6. Бета-распределение

Другим непрерывным распределением, которое часто встречается в математической статистике, является распределение, имеющее следующую ф. п. в.:

$$f(x)=\begin{cases}\dfrac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)}x^{\nu_1-1}(1-x)^{\nu_2-1}, & 0<x<1,\\ 0 & \text{в противном случае,}\end{cases} \qquad (7.6.1)$$

где ν_1 и ν_2 — положительные вещественные числа. Мы будем называть это распределение *бета-распределением* Be (ν_1, ν_2).

Убедимся, что интеграл от $f(x)$, взятый по интервалу (0, 1), равен единице. Применяя (7.5.1), мы имеем

$$\Gamma(\nu_1)\Gamma(\nu_2)=\int_0^\infty\int_0^\infty x_1^{\nu_1-1}x_2^{\nu_2-1}e^{-x_1-x_2}dx_1\,dx_2, \tag{7.6.2}$$

что после преобразования

$$x_1=r^2\cos^2\theta,\qquad x_2=r^2\sin^2\theta$$

с якобианом

$$\frac{\partial(x_1, x_2)}{\partial(r, \theta)}=4r^3\sin\theta\cos\theta$$

дает

$$\Gamma(\nu_1)\Gamma(\nu_2)=4\int_0^{\frac{\pi}{2}}\int_0^\infty(\cos\theta)^{2\nu_1-1}(\sin\theta)^{2\nu_2-1}r^{2\nu_1+2\nu_2-1}e^{-r^2}\,d\theta\,dr. \tag{7.6.3}$$

Производя в интеграле

$$2\int_0^\infty r^{2\nu_1+2\nu_2-1}e^{-r^2}\,dr$$

замену $r=\sqrt{y}$, мы приводим его к выражению, определяющему $\Gamma(\nu_1+\nu_2)$. Поэтому (7.6.3) приобретает вид

$$\frac{\Gamma(\nu_1)\Gamma(\nu_2)}{\Gamma(\nu_1+\nu_2)}=2\int_0^{\frac{\pi}{2}}(\cos\theta)^{2\nu_1-1}(\sin\theta)^{2\nu_2-1}\,d\theta. \tag{7.6.4}$$

Наконец, делая в (7.6.4) замену переменной интегрирования по формуле $\cos\vartheta=\sqrt{x}$, мы получаем

$$\frac{\Gamma(\nu_1)\Gamma(\nu_2)}{\Gamma(\nu_1+\nu_2)}=\int_0^1 x^{\nu_1-1}(1-x)^{\nu_2-1}dx, \tag{7.6.5}$$

т. е. действительно интеграл от $f(x)$ по интервалу (0, 1) равен единице.

К. ф. р. $F(x)$ бета-распределения $\mathrm{Be}(\nu_1, \nu_2)$, обозначаемая через $I_x(\nu_1, \nu_2)$ и называемая *неполной бета-функцией*, была табулирована под руководством Карла Пирсона (1934) для значений x от 0,01 до 1,00 и ν_1, ν_2 от 0,05 до 50.

Функция от ν_1 и ν_2, равная определенному интегралу в правой части (7.6.5), называется *бета-функцией* от ν_1 и ν_2; она имеет классическое обозначение $\mathrm{B}(\nu_1, \nu_2)$.

Бета-распределение доставляет еще один пример непрерывного распределения, для которого характеристическая функция не приносит пользы при нахождении моментов. Простейший путь найти эти моменты — обратиться к непосредственному вычислению. Так, если μ'_r — r-й момент распределения (7.6.1), мы имеем

$$\mu'_r = \frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} \int_0^1 x^{\nu_1+r-1}(1-x)^{\nu_2-1}\,dx. \tag{7.6.6}$$

Используя (7.6.5), мы получаем следующее выражение для r-го момента:

$$\mu'_r = \frac{\Gamma(\nu_1+\nu_2)\,\Gamma(\nu_1+r)}{\Gamma(\nu_1+\nu_2+r)\,\Gamma(\nu_1)}, \tag{7.6.7}$$

откуда

$$\mu(x) = \frac{\nu_1}{\nu_1+\nu_2}, \quad \sigma^2(x) = \frac{\nu_1\nu_2}{(\nu_1+\nu_2)^2(\nu_1+\nu_2+1)}. \tag{7.6.8}$$

Бета-распределение возникает в различных ситуациях, встречающихся в математической статистике, в частности, в теории порядковых статистик и некоторых статистических критериях. Мы обсудим эти ситуации в более поздних главах. Один из наиболее важных случаев возникновения бета-распределения выражается теоремой

7.6.1. *Если x_1 и x_2 — независимые случайные величины, имеющие соответственно гамма-распределения $G(\nu_1)$ и $G(\nu_2)$, то случайная величина*

$$u = \frac{x_1}{x_1+x_2} \tag{7.6.9}$$

имеет бета-распределение $\mathrm{Be}(\nu_1, \nu_2)$.

Это утверждение проверяется непосредственно. В. э. двумерной случайной величины (x_1, x_2) равен

$$\frac{1}{\Gamma(\nu_1)\Gamma(\nu_2)} x_1^{\nu_1-1} x_2^{\nu_2-1} e^{-x_1-x_2}\,dx_1\,dx_2. \tag{7.6.10}$$

Если в (7.6.10) сделать замену переменных

$$u = \frac{x_1}{x_1+x_2}, \quad v = x_2, \tag{7.6.11}$$

мы получим в. э. величины (u, v)

$$\frac{1}{\Gamma(\nu_1)\Gamma(\nu_2)} u^{\nu_1-1} v^{\nu_1+\nu_2-1}(1-u)^{-(\nu_1+1)} e^{-\frac{v}{1-u}}\,du\,dv. \tag{7.6.12}$$

Распределение u является маргинальным для распределения, имеющего в. э. (7.6.12). Интегрируя (7.6.12) по v от 0 до ∞, мы обнаруживаем, что u имеет бета-распределение $\mathrm{Be}(\nu_1, \nu_2)$.

Замечания. Прежде чем закончить параграф, посвященный бета-распределению, будет полезно указать на две важные формулы, касающиеся гамма-функции, полученные соответственно Лежандром и Стирлингом. Эти формулы можно вывести из бета-распределения путем элементарных рассуждений.

Формула удвоения Лежандра для гамма-функции. Предположим, что в (7.6.5) $\nu_1 = \nu_2 = \nu$. Тогда

$$\frac{\Gamma^2(\nu)}{\Gamma(2\nu)} = \int_0^1 x^{\nu-1}(1-x)^{\nu-1}\,dx = 2\int_0^{\frac{1}{2}} (x-x^2)^{\nu-1}\,dx. \tag{7.6.13}$$

Переходя в последнем интеграле к новой переменной $y = 4(x - x^2)$ при $0 < x < \frac{1}{2}$, т. е. полагая $x = \frac{1}{2}(1 - \sqrt{1-y})$, мы находим

$$\frac{\Gamma^2(\nu)}{\Gamma(2\nu)} = 2^{1-2\nu}\int_0^1 y^{\nu-1}(1-y)^{-\frac{1}{2}}\,dy = 2^{1-2\nu}\,\frac{\Gamma(\nu)\,\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\nu+\frac{1}{2}\right)} \tag{7.6.14}$$

и, поскольку $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$,

$$\Gamma(2\nu) = \frac{2^{2\nu-1}}{\sqrt{\pi}}\,\Gamma(\nu)\,\Gamma\left(\nu+\frac{1}{2}\right). \tag{7.6.15}$$

Эта формула и называется *формулой удвоения Лежандра для гамма-функции;* она понадобится нам в последующих главах.

Формула Стирлинга для больших факториалов. Используя некоторые предельные свойства бета-распределения, мы можем установить с помощью элементарного анализа чрезвычайно важную аппроксимацию для $\Gamma(g)$ для больших вещественных значений g.

Возьмем в формуле (7.6.1) $\nu_1 = g$ и $\nu_2 = n+1$ и сделаем замену переменной $x = \frac{y}{n}$. Тогда

$$\frac{\Gamma(g+n+1)}{n^g\,\Gamma(g)\,\Gamma(n+1)}\int_0^n y^{g-1}\left(1-\frac{y}{n}\right)^n dy = 1 \tag{7.6.16}$$

для всех значений n. Поэтому предел левой части (7.6.16) при $n\to\infty$ равен единице. Следует заметить, что подынтегральное выражение в (7.6.16) сходится к $y^{g-1}e^{-y}$ при $n\to\infty$ равномерно в любом конечном интервале $(0, K)$. Выбрав n достаточно большим, мы можем сделать разность

$$\int_0^{\min(n,K)} y^{g-1}\left(1-\frac{y}{n}\right)^n dy - \int_0^K y^{g-1}e^{-y}\,dy$$

произвольно малой. Но K можно заранее фиксировать так, чтобы интеграл

$$\int_0^{\min(n,K)} y^{g-1}\left(1-\frac{y}{n}\right)^n dy$$

также был произвольно мал. Следовательно,

$$\lim_{n\to\infty}\int_0^n y^{g-1}\left(1-\frac{y}{n}\right)^n dy = \int_0^\infty y^{g-1}e^{-y}\,dy = \Gamma(g). \tag{7.6.17}$$

Таким образом,

$$\lim_{n\to\infty}\frac{\Gamma(n+1)\,n^g}{\Gamma(g+n+1)} = 1,$$

что можно переписать в виде

$$\Gamma(g)=\lim_{n\to\infty}\frac{n!n^g}{g(g+1)\ldots(g+n)}. \tag{7.6.18}$$

Пусть теперь

$$S_n(g)=g\ln n+\sum_{\alpha=1}^{n}\ln\alpha-\sum_{\alpha=0}^{n}\ln(g+\alpha). \tag{7.6.19}$$

Дифференцируя (7.6.19) по g, мы находим

$$S'_n(g)=\ln n-\sum_{\alpha=0}^{n}\frac{1}{g+\alpha}. \tag{7.6.20}$$

Очевидно,

$$\lim_{n\to\infty}S_n(g)=\ln\Gamma(g)\quad\text{и}\quad\lim_{n\to\infty}S'_n(g)=\frac{\Gamma'(g)}{\Gamma(g)}.$$

Далее, как можно показать посредством элементарных вычислений,

$$-B_n(g)<-\sum_{\alpha=0}^{n}\frac{1}{g+\alpha}+\frac{1}{2g}+\frac{1}{2(g+n)}<-C_n(g), \tag{7.6.21}$$

где

$$B_n(g)=\frac{1}{2}\int_{\frac{1}{2}}^{n+\frac{1}{2}}\left(\frac{1}{g+x-1}+\frac{1}{g+x}\right)dx,\qquad C_n(g)=\int_{\frac{1}{2}}^{n+\frac{1}{2}}\frac{dx}{g+x-\frac{1}{2}}.$$

Обозначая $\ln n-\frac{1}{2}\left(\frac{1}{g}+\frac{1}{g+n}\right)$ через $A_n(g)$ и прибавляя $A_n(g)$ ко всем частям неравенства (7.6.21), мы имеем

$$A_n(g)-B_n(g)<S'_n(g)<A_n(g)-C_n(g). \tag{7.6.22}$$

Устремляя здесь n к бесконечности и учитывая вид $A_n(g)$, $B_n(g)$ и $C_n(g)$, приходим к неравенству

$$\ln g-\frac{1}{2g}+\frac{1}{2}\ln\left(1-\frac{1}{4g^2}\right)<\frac{\Gamma'(g)}{\Gamma(g)}<\ln g-\frac{1}{2g}, \tag{7.6.23}$$

из которого вытекает соотношение

$$\frac{\Gamma'(g)}{\Gamma(g)}=\ln g-\frac{1}{2g}+O\left(\frac{1}{g^2}\right). \tag{7.6.24}$$

Интегрирование (7.6.24) по g дает

$$\ln\Gamma(g)=g\ln g-n-\frac{1}{2}\ln g+C+O\left(\frac{1}{g}\right), \tag{7.6.25}$$

где C — не зависящая от g постоянная. Мы можем определить значение C, воспользовавшись формулой (7.6.15). Для этого заменим в этой формуле ν на g и возьмем логарифм. Тогда

$$\ln\Gamma(2g)=(2g-1)\ln 2+\ln\Gamma(g)+\ln\Gamma\left(g+\frac{1}{2}\right)-\frac{1}{2}\ln\pi. \tag{7.6.26}$$

Подставив сюда вместо $\ln\Gamma(2g)$, $\ln\Gamma(g)$ и $\ln\Gamma\left(g+\frac{1}{2}\right)$ их выражения из (7.6.25) и перейдя к пределу при $g\to\infty$, мы получаем $C=\frac{1}{2}\ln 2\pi$.

Таким образом, окончательно

$$\Gamma(g)=\sqrt{2\pi}\, g^{g-\frac{1}{2}}e^{-g}\left(1+O\left(\frac{1}{g}\right)\right). \tag{7.6.27}$$

Неравенство (7.6.21) слишком грубо, чтобы дать нам коэффициенты при различных степенях $\frac{1}{g}$ в $O\left(\frac{1}{g}\right)$ из формулы (7.6.27). В действительности $O\left(\frac{1}{g}\right)=\frac{1}{12g}+\frac{1}{288g^2}+O\left(\frac{1}{g^3}\right)$, но чтобы установить это, требуются более сильные аналитические методы, чем использованные нами. По этому поводу можно, например, обратиться к Уиттекеру и Ватсону (1946) или Крамеру (1946). Для больших значений g асимптотическая формула

$$\Gamma(g)\cong\sqrt{2\pi}\, g^{g-\frac{1}{2}}e^{-g} \tag{7.6.28}$$

достаточна для большинства приложений. Если g — целое положительное число и, следовательно, $\Gamma(g+1)=g!$, мы имеем формулу Стирлинга для больших факториалов

$$g!\cong\sqrt{2\pi g}\, g^g e^{-g}. \tag{7.6.29}$$

7.7. Распределение Дирихле

Многомерным аналогом (7.6.1) является распределение, имеющее ф. п. в., равную

$$f(x_1, \ldots, x_k)=\frac{\Gamma(\nu_1+\ldots+\nu_{k+1})}{\Gamma(\nu_1)\ldots\Gamma(\nu_{k+1})}x_1^{\nu_1-1}\ldots x_k^{\nu_k-1}\times$$
$$\times(1-x_1-\ldots-x_k)^{\nu_{k+1}-1} \tag{7.7.1}$$

в любой точке симплекса $S_k=\{(x_1, \ldots, x_k): x_i\geqslant 0,\ i=1, \ldots, k,\ \sum_{i=1}^{k}x_i\leqslant 1\}$ и равную нулю в других точках R_k. Здесь все ν_i — вещественные положительные числа. Мы будем называть распределение с ф. п. в. (7.7.1) *k-мерным распределением Дирихле* $D(\nu_1, \ldots, \nu_k;\ \nu_{k+1})$. Заметим, что если $k=1$, (7.7.1) сводится к ф. п. в. бета-распределения, т. е. $D(\nu_1;\ \nu_2)$ совпадает с $\mathrm{Be}(\nu_1,\ \nu_2)$. Распределение Дирихле является основным в вероятностной теории порядковых статистик, как мы увидим это в § 8.7.

С целью убедиться в том, что интеграл от $f(x_1, \ldots, x_k)$ по симплексу S_k равен единице, сделаем в (7.7.1) замену переменных

$$x_1=\theta_1,\quad x_2=\theta_2(1-x_1),\quad \ldots,\quad x_k=\theta_k(1-x_1-\ldots-x_{k-1}).$$

В результате мы получим

$$f(x_1, \ldots, x_k)\,dx_1 \ldots dx_k =$$
$$= \frac{\Gamma(\nu_1 + \ldots + \nu_{k+1})}{\Gamma(\nu_1) \ldots \Gamma(\nu_{k+1})} \theta_1^{\nu_1 - 1} (1-\theta_1)^{\nu_2 + \ldots + \nu_{k+1} - 1} \theta_2^{\nu_2 - 1} \times$$
$$\times (1-\theta_2)^{\nu_3 + \ldots + \nu_{k+1} - 1} \ldots \theta_k^{\nu_k - 1} (1-\theta_k)^{\nu_{k+1} - 1}\, d\theta_1 \ldots d\theta_k, \qquad (7.7.2)$$

где $(\theta_1, \ldots, \theta_k)$ принадлежит *k-мерному единичному кубу* $\{(\theta_1, \ldots, \theta_k): 0 < \theta_i < 1,\ i = 1, \ldots, k\}$. Применяя (7.6.5), мы имеем

$$\int_{S_k} f(x_1, \ldots, x_k)\,dx_1 \ldots dx_k = \frac{\Gamma(\nu_1 + \ldots + \nu_{k+1})}{\Gamma(\nu_1) \ldots \Gamma(\nu_{k+1})} \times$$
$$\times \frac{\Gamma(\nu_1)\Gamma(\nu_2 + \ldots + \nu_{k+1})}{\Gamma(\nu_1 + \ldots + \nu_{k+1})} \ldots \frac{\Gamma(\nu_k)\Gamma(\nu_{k+1})}{\Gamma(\nu_k + \nu_{k+1})}. \qquad (7.7.3)$$

Ясно, что правая часть после сокращения дает единицу.

Интеграл

$$\int_{S_k} x_1^{\nu_1 - 1} \ldots x_k^{\nu_k - 1} (1 - x_1 - \ldots - x_k)^{\nu_{k+1} - 1}\, dx_1 \ldots dx_k, \qquad (7.7.4)$$

равный по доказанному

$$\frac{\Gamma(\nu_1) \ldots \Gamma(\nu_{k+1})}{\Gamma(\nu_1 + \ldots + \nu_{k+1})}, \qquad (7.7.5)$$

был впервые исследован Дирихле (1839) и известен как *интеграл Дирихле*. Этим и объясняется название, данное нами распределению, имеющему ф. п. в. (7.7.1). Аналогичным образом ранее был введен термин «бета-распределение».

Нетрудно проверить, что смешанный момент $\mu'_{r_1, \ldots, r_k}$ k-мерного распределения Дирихле выражается формулой

$$\mu'_{r_1, \ldots, r_k} = \frac{\Gamma(\nu_1 + r_1) \ldots \Gamma(\nu_k + r_k)\,\Gamma(\nu_1 + \ldots + \nu_{k+1})}{\Gamma(\nu_1 + \ldots + \nu_{k+1} + r_1 + \ldots + r_k)\,\Gamma(\nu_1) \ldots \Gamma(\nu_k)}, \qquad (7.7.6)$$

откуда мы находим средние, дисперсии и ковариации величин x_i:

$$\mu(x_i) = \frac{\nu_i}{\nu_1 + \ldots + \nu_{k+1}}, \qquad i = 1, \ldots, k,$$
$$\sigma^2(x_i) = \frac{\nu_i(\nu_1 + \ldots + \nu_{k+1} - \nu_i)}{(\nu_1 + \ldots + \nu_{k+1})^2 (\nu_1 + \ldots + \nu_{k+1} + 1)}, \qquad i = 1, \ldots, k, \qquad (7.7.7)$$
$$\sigma(x_i, x_j) = -\frac{\nu_i \nu_j}{(\nu_1 + \ldots + \nu_{k+1})^2 (\nu_1 + \ldots + \nu_{k+1} + 1)}, \qquad i \neq j = 1, \ldots, k.$$

Можно сформулировать следующий k-мерный аналог теоремы **7.6.1**:

7.7.1. *Пусть $x_1, \ldots, x_{k+1}$ — независимые случайные величины, имеющие соответственно гамма-распределения $G(\nu_1), \ldots, G(\nu_{k+1})$, и*

$$y_i = \frac{x_i}{x_1 + \ldots + x_{k+1}}, \qquad i = 1, \ldots, k. \qquad (7.7.8)$$

Тогда $(y_1, \ldots, y_k)$ *имеет* k*-мерное распределение Дирихле* $D(\nu_1, \ldots$ $\ldots, \nu_k; \nu_{k+1})$.

Теорема **7.7.1** доказывается непосредственно. Это доказательство мы оставляем читателю в качестве упражнения.

Положим в формуле (7.7.6) $r_{k_1+1} = \ldots = r_k = 0$, $k_1 < k$. Мы получим смешанный момент $\mu'_{r_1, \ldots, r_{k_1}}$ маргинального распределения случайной величины $(x_1, \ldots, x_{k_1})$ для k-мерного распределения Дирихле $D(\nu_1, \ldots, \nu_k; \nu_{k+1})$, причем

$$\mu'_{r_1, \ldots, r_{k_1}} = \frac{\Gamma(\nu_1 + r_1) \ldots \Gamma(\nu_{k_1} + r_{k_1}) \Gamma(\nu_1 + \ldots + \nu_{k+1})}{\Gamma(\nu_1 + \ldots + \nu_{k+1} + r_1 + \ldots + r_{k_1}) \Gamma(\nu_1) \ldots \Gamma(\nu_{k_1})}. \quad (7.7.9)$$

Но именно так записываются смешанные моменты k_1-мерного распределения Дирихле $D(\nu_1, \ldots, \nu_{k_1}; \nu_{k_1+1} + \ldots + \nu_{k+1})$. С другой стороны, согласно многомерному аналогу теоремы **5.5.1a** распределение Дирихле однозначно определяется своими моментами. Таким образом,

7.7.2. *Если* $(x_1, \ldots, x_k)$ *— векторная случайная величина, имеющая* k*-мерное распределение Дирихле* $D(\nu_1, \ldots, \nu_k; \nu_{k+1})$, *то маргинальное распределение случайной величины* $(x_1, \ldots, x_{k_1})$, $k_1 < k$, *является* k_1*-мерным распределением Дирихле* $D(\nu_1, \ldots, \nu_{k_1}; \nu_{k_1+1} + \ldots + \nu_{k+1})$.

Теперь мы можем выписать ф. п. в. условной случайной величины $x_k \mid x_1, \ldots, x_{k-1}$ для k-мерного распределения Дирихле, взяв отношение ф. п. в. распределения $D(\nu_1, \ldots, \nu_k; \nu_{k+1})$ и ф. п. в. распределения $D(\nu_1, \ldots, \nu_{k-1}; \nu_k + \nu_{k+1})$. После некоторого упрощения мы получаем следующее выражение для в. э. условной величины $x_k \mid x_1, \ldots, x_{k-1}$:

$$dF(x_k \mid x_1, \ldots, x_{k-1}) = \frac{\Gamma(\nu_k + \nu_{k+1})}{\Gamma(\nu_k)\Gamma(\nu_{k+1})} \left(\frac{x_k}{1 - x_1 - \ldots - x_{k-1}}\right)^{\nu_k - 1} \times$$
$$\times \left(1 - \frac{x_k}{1 - x_1 - \ldots - x_{k-1}}\right)^{\nu_{k+1} - 1} d\left(\frac{x_k}{1 - x_1 - \ldots - x_{k-1}}\right). \quad (7.7.10)$$

Из (7.7.10), очевидно, вытекает

7.7.3. *Если* $(x_1, \ldots, x_k)$ *— векторная случайная величина, имеющая* k*-мерное распределение Дирихле* $D(\nu_1, \ldots, \nu_k; \nu_{k+1})$, *то условная случайная величина*

$$\frac{x_k}{1 - x_1 - \ldots - x_{k-1}} \,\Big|\, x_1, \ldots, x_{k-1}$$

имеет бета-распределение $\mathrm{Be}(\nu_k, \nu_{k+1})$.

Следует заметить, что в соответствии с (7.6.8) среднее и дисперсия величины $x_k \mid x_1, \ldots, x_{k-1}$ даются формулами

$$\mu(x_k \mid x_1, \ldots, x_{k-1}) = \frac{\nu_k}{\nu_k + \nu_{k+1}} (1 - x_1 - \ldots - x_{k-1}),$$
$$\sigma^2(x_k \mid x_1, \ldots, x_{k-1}) = \frac{\nu_k \nu_{k+1}}{(\nu_k + \nu_{k+1})^2 (\nu_k + \nu_{k+1} + 1)} (1 - x_1 - \ldots - x_{k-1})^2. \quad (7.7.11)$$

Другое полезное свойство k-мерного распределения Дирихле можно сформулировать в виде теоремы

7.7.4. *Если* $(x_1, \ldots, x_k)$ — *векторная случайная величина, имеющая k-мерное распределение Дирихле* $D(\nu_1, \ldots, \nu_k; \nu_{k+1})$, *то сумма* $x_1+\ldots+x_k$ *имеет бета-распределение* $\mathrm{Be}(\nu_1+\ldots+\nu_k; \nu_{k+1})$.

Эта теорема следует из того, что r-й момент величины $1-(x_1+\ldots+x_k)$ равен

$$\frac{\Gamma(\nu_{k+1}+r)\,\Gamma(\nu_1+\ldots+\nu_{k+1})}{\Gamma(\nu_1+\ldots+\nu_{k+1}+r)\,\Gamma(\nu_{k+1})}, \qquad r=1, 2, \ldots$$

и, значит, по теореме **5.5.1a** величина $1-(x_1+\ldots+x_k)$ имеет бета-распределение $\mathrm{Be}(\nu_{k+1}; \nu_1+\ldots+\nu_k)$. Поэтому сумма $x_1+\ldots+x_k$ имеет бета-распределение $\mathrm{Be}(\nu_1+\ldots+\nu_k; \nu_{k+1})$.

Теорема **7.7.4** может быть обобщена следующим образом.

7.7.5. *Если* $(x_1, \ldots, x_k)$ — *векторная случайная величина, имеющая k-мерное распределение Дирихле* $D(\nu_1, \ldots, \nu_k; \nu_{k+1})$, *то случайная величина* $(z_1, \ldots, z_s)$, *где* $z_1=x_1+\ldots+x_{k_1}$, $z_2=x_{k_1+1}+\ldots+x_{k_1+k_2}, \ldots, z_s=x_{k_1+\ldots+k_{s-1}+1}+\ldots+x_{k_1+\ldots+k_s}$ *и* $k_1+\ldots+k_s\leqslant k$, *имеет s-мерное распределение Дирихле* $D(\nu_{(1)}, \ldots, \nu_{(s)}; \nu_{(s+1)})$, *где* $\nu_{(1)}=\nu_1+\ldots+\nu_{k_1}, \ldots, \nu_{(s)}=\nu_{k_1+\ldots+k_{s-1}+1}+\ldots+\nu_{k_1+\ldots+k_s}$, $\nu_{(s+1)}=\nu_{k_1+\ldots+k_s+1}+\ldots+\nu_{k+1}$.

Действительно, в результате прямых вычислений, использующих (7.7.1), мы имеем

$$\mathcal{E}\left[x_{k_1+1}^{r_{k_1+1}}\ldots x_k^{r_k}(1-z_1-x_{k_1+1}-\ldots-x_k)^{r_{(1)}}\right]=$$
$$=G\frac{\Gamma(\nu_{(1)})\,\Gamma(\nu_{k_1+1}+r_{k_1+1})\ldots\Gamma(\nu_k+r_k)\,\Gamma(\nu_{k+1}+r_{(1)})}{\Gamma(\nu_{(1)}+\nu_{k_1+1}+\ldots+\nu_k+\nu_{k+1}+r_{k_1+1}+\ldots+r_k+r_{(1)})}, \tag{7.7.12}$$

где

$$G=\frac{\Gamma(\nu_{(1)}+\nu_{k_1+1}+\ldots+\nu_{k+1})}{\Gamma(\nu_{(1)})\,\Gamma(\nu_{k_1+1})\ldots\Gamma(\nu_{k+1})}. \tag{7.7.13}$$

Но правая часть формулы (7.7.12) равна среднему

$$\mathcal{E}\left[x_{k_1+1}^{r_{k_1+1}}\ldots x_k^{r_k}(1-z_1-x_{k_1+1}-\ldots-x_k)^{r_{(1)}}\right],$$

вычисленному, исходя из ф. п. в.

$$f(z_1, x_{k_1+1}, \ldots, x_k)=Gz_1^{\nu_{(1)}-1}x_{k_1+1}^{\nu_{k_1+1}-1}\ldots x_k^{\nu_k-1}\times$$
$$\times(1-z_1-x_{k_1+1}-\ldots-x_k)^{\nu_{k+1}-1}, \tag{7.7.14}$$

причем последняя определяется моментами единственным образом, так как $(z_1, x_{k_1+1}, \ldots, x_k)$ — ограниченная случайная величина.

Мы можем, далее, повторить рассуждения, вводя z_2 и т. д. В конце концов мы получим

$$f(z_1, \ldots, z_s)=\frac{\Gamma(\nu_{(1)}+\ldots+\nu_{(s+1)})}{\Gamma(\nu_{(1)})\ldots\Gamma(\nu_{(s+1)})}\times$$
$$\times z_1^{\nu_{(1)}-1}\ldots z_s^{\nu_{(s)}-1}(1-z_1-\ldots-z_s)^{\nu_{(s+1)}-1}, \tag{7.7.15}$$

что и доказывает теорему **7.7.5**.

В приложениях к порядковым статистикам, в частности в § 8.7, оказывается полезным другое распределение, тесно связанное с k-мерным распределением Дирихле. Пусть $(x_1, \ldots, x_k)$ — векторная случайная величина, имеющая k-мерное распределение Дирихле $D(\nu_1, \ldots, \nu_k; \nu_{k+1})$. Введем новые величины

$$\begin{aligned} y_1 &= x_1, \\ y_2 &= x_1 + x_2, \\ &\ldots\ldots\ldots \\ y_k &= x_1 + \ldots + x_k. \end{aligned} \tag{7.7.16}$$

Так как якобиан этого преобразования равен единице, ф. п. в. величины $(y_1, \ldots, y_k)$ будет иметь вид

$$f(y_1, \ldots, y_k) = \frac{\Gamma(\nu_1 + \ldots + \nu_{k+1})}{\Gamma(\nu_1)\ldots\Gamma(\nu_{k+1})} y_1^{\nu_1 - 1} \times \\ \times (y_2 - y_1)^{\nu_2 - 1} \ldots (y_k - y_{k-1})^{\nu_k - 1} (1 - y_k)^{\nu_{k+1} - 1}, \tag{7.7.17}$$

где область изменения y описывается неравенствами $0 < y_1 < \ldots < y_k < 1$. Удобно называть распределение, имеющее ф. п. в. (7.7.17), *порядковым k-мерным распределением Дирихле* $D^*(\nu_1, \ldots, \nu_k; \nu_{k+1})$. Заметим, что при $k = 1$ (7.7.17) сводится к ф. п. в. бета-распределения $\mathrm{Be}(\nu_1, \nu_2)$.

Иногда представляет интерес маргинальное распределение некоторого подмножества s величин y_i. Приведем результат, который найдет применение в задачах, касающихся порядковых статистик.

7.7.6. *Если $(y_1, \ldots, y_k)$ — векторная случайная величина, имеющая порядковое k-мерное распределение Дирихле $D^*(\nu_1, \ldots, \nu_k; \nu_{k+1})$, то маргинальное распределение величины $(y_{k_1}, y_{k_1+k_2}, \ldots, y_{k_1+\ldots+k_s})$ является порядковым s-мерным распределением Дирихле $D^*(\nu_1, \ldots, \nu_{(s)}; \nu_{(s+1)})$, где $\nu_{(1)}, \ldots, \nu_{(s+1)}$ — числа, определенные в теореме* **7.7.5.**

В самом деле, случайные величины $y_{k_1}, \ldots, y_{k_1+\ldots+k_s}$, которые определены в (7.7.16), имеют то же самое распределение, что и случайные величины $z_1, z_1 + z_2, \ldots, z_1 + \ldots + z_s$, определенные в теореме **7.7.5.** Так как $(z_1, \ldots, z_s)$ имеет s-мерное распределение Дирихле $D(\nu_{(1)}, \ldots, \nu_{(s)}; \nu_{(s+1)})$, то по определению $(z_1, z_1 + z_2, \ldots, z_1 + \ldots + z_s)$ имеет порядковое s-мерное распределение Дирихле $D^*(\nu_{(1)}, \ldots, \nu_{(s)}; \nu_{(s+1)})$.

Следует отметить, что теорема **7.7.6** также может быть установлена прямым интегрированием ф. п. в. (7.7.17) по всем y_i, за исключением $y_{k_1}, y_{k_1+k_2}, \ldots, y_{k_1+\ldots+k_s}$. Интегрирование ведется сначала последовательно по $y_1, \ldots, y_{k_1-1}$ в области $0 < y_1 < \ldots < y_{k_1}$, затем последовательно по $y_{k_1+1}, \ldots, y_{k_1+k_2-1}$ в области $y_{k_1} < y_{k_1+1} < \ldots < y_{k_1+k_2}$ и т. д.

7.8. Распределения, встречающиеся в дисперсионном анализе

В дисперсионном анализе, а также в других основанных на нормально распределенных случайных величинах статистических процедурах, которые будут рассмотрены в дальнейших главах, фундаментальную роль играют три распределения, тесно связанные с гамма- и бета-распределениями, именно, χ^2-распределение, распределение Стьюдента и распределение Снедекора. Удобно привести здесь эти распределения в их основных формах в качестве новых примеров важных непрерывных распределений.

(a) χ^2-распределение. Сделав замену переменной

$$\chi^2 = 2x$$

в в. э. гамма-распределения $G(\mu)$, мы получим

$$dF_{2\mu}(\chi^2) = \frac{\left(\frac{\chi^2}{2}\right)^{\mu-1} e^{-\frac{1}{2}\chi^2}}{2\Gamma(\mu)} d\chi^2. \tag{7.8.1}$$

Это выражение определяет в. э. *χ^2-распределения с 2μ степенями свободы.* Если случайная величина имеет распределение с в. э. (7.8.1), мы будем говорить, что она имеет *χ^2-распределение* $C(2\mu)$. В большинстве статистических приложений этого распределения 2μ является целым положительным числом. Значения χ^2_α, для которых $\int\limits_{\chi^2_\alpha}^{\infty} dF_{2\mu}(\chi^2) = \alpha$, для различных значений α от 0,001 до 0,99 при $2\mu = 1, 2, \ldots, 30$ были табулированы Фишером (1925a). Более обширные таблицы изданы Пирсоном и Хартли (1954).

Следует отметить, что характеристической функцией χ^2-распределения $C(2\mu)$ будет функция

$$\varphi(t) = (1 - 2it)^{-\mu}, \tag{7.8.2}$$

которая получается путем замены в (7.5.6) t на $2t$.

Из (7.8.2), очевидно, вытекает теорема

7.8.1. *χ^2-распределение $C(2\mu)$ является воспроизводящим по μ.*

Поскольку $\frac{\chi^2}{2}$ имеет гамма-распределение $G(\mu)$, применяя (7.5.3), мы находим r-й момент величины χ^2

$$\mu'_r = \frac{2^r\Gamma(\mu + r)}{\Gamma(\mu)}. \tag{7.8.3}$$

Отсюда для χ^2-распределения $C(2\mu)$

$$\mathscr{E}(\chi^2) = 2\mu, \quad \sigma^2(\chi^2) = 4\mu.$$

Обозначение χ^2, введенное К. Пирсоном (1905), довольно неудобно, но мы будем продолжать им пользоваться, так как оно глубоко укоренилось в статистической литературе.

Возвращаясь к задаче распределения формы $Q(x_1, \ldots, x_k)$ и учитывая теорему **7.5.2**, можно установить следующий факт.

7.8.2. *Если $(x_1, \ldots, x_k)$ — векторная случайная величина, имеющая k-мерное распределение $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j = 1, \ldots, k$, то форма $\sum_{i,j=1}^{k} \sigma^{ij}(x_i - \mu_i)(x_j - \mu_j)$ имеет χ^2-распределение $C(k)$.*

Приведем важное следствие этой теоремы.

7.8.2а. *Если x — случайная величина, имеющая распределение $N(0, 1)$, то x^2 имеет χ^2-распределение $C(1)$.*

Причина, по которой с параметром 2μ связывается термин степени свободы, становится понятной, если заметить, что указанное в теореме **7.8.1** число степеней свободы совпадает с числом случайных величин в рассматриваемом (невырожденном) нормальном распределении. Разумность выбора этого термина станет еще яснее, когда мы познакомимся в дальнейших главах с некоторыми приложениями χ^2-распределения.

(b) Распределение Стьюдента. Математическую сущность этого распределения и его отношение к распределениям других случайных величин можно выразить следующим образом.

7.8.3. *Пусть u — случайная величина, имеющая распределение $N(0, 1)$, и v — случайная величина, имеющая χ^2-распределение $C(k)$. Если u и v независимы, случайная величина*

$$t = \frac{u}{\sqrt{\frac{v}{k}}}$$

имеет ф. п. в.

$$f_k(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}} \left(1 + \frac{t^2}{k}\right)^{-\frac{1}{2}(k+1)}. \tag{7.8.4}$$

Формула (7.8.4) и определяет ф. п. в. *распределения Стьюдента с k степенями свободы*. Для краткости мы будем называть это распределение *распределением Стьюдента* $S(k)$. Различные его приложения рассматриваются в §§ 8.4 и 10.4.

Чтобы получить распределение (7.8.4), применим к в. э.

$$\frac{1}{2\sqrt{2\pi}\,\Gamma\left(\frac{k}{2}\right)} \left(\frac{v}{2}\right)^{\frac{1}{2}(k-2)} e^{-\frac{1}{2}(u^2+v)}\, du\, dv \tag{7.8.5}$$

случайной величины (u, v) преобразование

$$s = v, \quad t = \frac{u}{\sqrt{\frac{v}{k}}} \tag{7.8.6}$$

и найдем затем маргинальное распределение t.

Якобиан преобразования (7.8.6) равен $\sqrt{\frac{s}{k}}$, и, следовательно, в. э. величины (s, t) будет иметь вид

$$\frac{1}{2\sqrt{\pi k}\,\Gamma\left(\frac{k}{2}\right)}\left(\frac{s}{2}\right)^{\frac{1}{2}(k-1)} e^{-\frac{1}{2}\left[1+\frac{t^2}{k}\right]s}\,ds\,dt. \qquad (7.8.7)$$

ф. п. в. величины t, и притом тождественную (7.8.4), мы получим, вычислив маргинальный в. э. t, т. е. проинтегрировав (7.8.7) по s от 0 до ∞.

Величины t_α, для которых $P(|t|>t_\alpha)=\alpha$, были табулированы Фишером (1925a) для различных значений α от 0,1 до 0,99 и $k = 1, 2, \ldots, 30$. Соответствующую таблицу можно найти также у Пирсона и Хартли (1954).

Отметим, что все нечетные моменты распределения, имеющего ф. п. в. (7.8.4), если они существуют, равны нулю. Что же касается существующих четных моментов, то

$$\mu'_{2r}=\mu_{2r}=k^r\mathscr{E}\left(\frac{u^{2r}}{v^r}\right).$$

Далее, так как u и v независимы,

$$\mathscr{E}\left(\frac{u^{2r}}{v^r}\right)=\mathscr{E}(u^{2r})\,\mathscr{E}(v^{-r}).$$

Но u^2 и v имеют соответственно χ^2-распределения $C(1)$ и $C(k)$. Отсюда, используя (7.8.3), мы находим

$$\mu'_{2r}=\mu_{2r}=k^r\frac{\Gamma\left(\frac{1}{2}+r\right)\Gamma\left(\frac{k}{2}-r\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{k}{2}\right)}. \qquad (7.8.8)$$

Одновременно обнаруживается, что μ'_{2r} существует в том и только том случае, если $-1<2r<k$.

Среднее и *дисперсия* распределения Стьюдента даются равенствами

$$\mathscr{E}(t)=0, \qquad \sigma^2(t)=\frac{k}{k-2}.$$

Нижеследующее утверждение устанавливает связь между распределением Стьюдента и бета-распределением.

7.8.4. *Если t — случайная величина, имеющая распределение Стьюдента $S(k)$, то $x=\frac{1}{1+\frac{t^2}{k}}$ имеет бета-распределение* $\mathrm{Be}\left(\frac{1}{2}k, \frac{1}{2}\right)$.

(c) Распределение Снедекора. Математическую сущность и происхождение этого распределения выясняет теорема

7.8.5. *Пусть u и v — независимые случайные величины, имеющие соответственно χ^2-распределения $C(k_1)$ и $C(k_2)$. Тогда случайная величина*

$$\mathcal{F}=\frac{u}{k_1}:\frac{v}{k_2}$$

имеет в. э.

$$dF_{k_1,k_2}(\mathcal{F})=\frac{\Gamma\left(\frac{1}{2}k_1+\frac{1}{2}k_2\right)}{\Gamma\left(\frac{1}{2}k_1\right)\Gamma\left(\frac{1}{2}k_2\right)}\left(\frac{k_1}{k_2}\right)^{\frac{1}{2}k_1}\mathcal{F}^{\frac{1}{2}k_1-1}\left[1+\frac{k_1\mathcal{F}}{k_2}\right]^{-\frac{1}{2}(k_1+k_2)}d\mathcal{F}. \tag{7.8.9}$$

Это — в. э. *распределения Снедекора со степенями свободы* k_1 и k_2. Для краткости мы будем называть это распределение *распределением Снедекора* $S(k_1, k_2)$. Его приложения рассматриваются в §§ 8.4, 10.4 и 10.6.

Распределение Снедекора можно получить, применяя к в. э.

$$\frac{\left(\frac{1}{2}u\right)^{\frac{1}{2}k_1-1}\left(\frac{1}{2}v\right)^{\frac{1}{2}k_2-1}}{4\Gamma\left(\frac{1}{2}k_1\right)\Gamma\left(\frac{1}{2}k_2\right)}e^{-\frac{1}{2}(u+v)}\,du\,dv \tag{7.8.10}$$

случайной величины (u, v) преобразование

$$\mathcal{F}=\frac{u}{k_1}:\frac{v}{k_2},\qquad \mathcal{G}=v \tag{7.8.11}$$

и вычисляя маргинальное распределение $\mathcal{F}$. В. э. величины $(\mathcal{F}, \mathcal{G})$ будет равен

$$\frac{\left(\frac{k_1}{k_2}\right)^{\frac{1}{2}k_1}}{2\Gamma\left(\frac{1}{2}k_1\right)\Gamma\left(\frac{1}{2}k_2\right)}\mathcal{F}^{\frac{1}{2}k_1-1}\left(\frac{\mathcal{G}}{2}\right)^{\frac{1}{2}(k_1+k_2)-1}e^{-\frac{1}{2}\left[1+\frac{k_1\mathcal{F}}{k_2}\right]\mathcal{G}}\,d\mathcal{F}\,d\mathcal{G}. \tag{7.8.12}$$

Интегрируя (7.8.12) по $\mathcal{G}$ от 0 до ∞, мы придем к в. э. величины $\mathcal{F}$, совпадающему с (7.8.9).

При условии $-k_1<2r<k_2$ распределение Снедекора обладает r-м моментом

$$\mu_r'=\left(\frac{k_2}{k_1}\right)^r \mathcal{E}(u^r)\,\mathcal{E}(v^{-r})=\left(\frac{k_2}{k_1}\right)^r\frac{\Gamma\left(\frac{1}{2}k_1+r\right)\Gamma\left(\frac{1}{2}k_2-r\right)}{\Gamma\left(\frac{1}{2}k_1\right)\Gamma\left(\frac{1}{2}k_2\right)}. \tag{7.8.13}$$

Для среднего и дисперсии величины $\mathcal{F}$ мы имеем

$$\mu(\mathcal{F})=\frac{k_2}{k_2-2},\qquad \sigma^2(\mathcal{F})=\frac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}. \tag{7.8.14}$$

Между распределением Снедекора и бета-распределением существует связь, которая может быть выражена следующей теоремой.

7.8.6. *Если $\mathscr{F}$ — случайная величина, имеющая распределение Снедекора $S(k_1, k_2)$, то случайная величина $\left[1+\frac{k_1\mathscr{F}}{k_2}\right]^{-1}$ имеет бета-распределение* $\mathrm{Be}\left(\frac{1}{2}k_2, \frac{1}{2}k_1\right)$, *а случайная величина* $\frac{k_1\mathscr{F}}{k_2+k_1\mathscr{F}}$ — *бета-распределение* $\mathrm{Be}\left(\frac{1}{2}k_1, \frac{1}{2}k_2\right)$.

Распределение Снедекора с практической точки зрения оказывается несколько более удобным по сравнению с распределением, первоначально введенным Фишером (1924) для целей дисперсионного анализа. Случайная величина z, предложенная Фишером, связана с величиной $\mathscr{F}$ равенством

$$z=\frac{1}{2}\ln\mathscr{F}. \tag{7.8.15}$$

Фишер (1925а) табулировал значения z_α, для которых $P(z>z_\alpha)=\alpha$, для $\alpha=0{,}01$, $0{,}05$, $k_1=1, 2, 3, 4, 5, 6, 8, 12, 24$ и $k_2=1, 2, \ldots$ $\ldots, 30, 60, \infty$. Снедекор (1937) табулировал значения $\mathscr{F}_\alpha$, соответствующие z_α (т. е. $z_\alpha=\frac{1}{2}\ln\mathscr{F}_\alpha$) для $\alpha=0{,}01$, $0{,}05$, $k_1=1, 2, \ldots, 12, 14, 16, 20, 24, 30, 40, 50, 75, 100, 200, 500, \infty$ и $k_2=1, 2, \ldots, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 55, 60, 65, 70, 80, 100, 125, 150, 200, 400, 1000, \infty$. По вопросу дальнейшего табулирования см. Гринвуд и Хартли (1961).

Если $\mathscr{F}$ имеет распределение Снедекора $S(k_1, k_2)$, то, как это видно из теоремы **7.8.6**, $\mathscr{F}_\alpha$ можно найти по таблицам неполной бета-функции Карла Пирсона (1934), поскольку $P(\mathscr{F}>\mathscr{F}_\alpha)=I_{x_\alpha}\left(\frac{1}{2}k_1, \frac{1}{2}k_2\right)$ и $\mathscr{F}_\alpha=\frac{k_2x_\alpha}{k_1(1-x_\alpha)}$.

ЗАДАЧИ

7.1. Доказать теорему **7.2.2.**

7.2. Показать, что если x есть случайная величина, имеющая нормальное распределение $N(\mu, \sigma^2)$, то

$$\mathscr{E}(|x-\mu|)=\sqrt{\frac{2}{\pi}}\sigma.$$

7.3. Показать, что если x есть случайная величина, имеющая ф. п. в. $f(x)$, то $\mathscr{E}(|x-c|)$ достигает минимума $c=\underline{x}_{0,5}$, где $\underline{x}_{0,5}$ — медиана.

7.4. Пусть x — случайная величина, имеющая нормальное распределение $N(\mu, \sigma^2)$ и $x_2 | x_1$ — условная случайная величина, имеющая нормальное распределение $N(x_1, \sigma^2)$. Показать, что (x_1, x_2) имеет двумерное нормальное распределение $N(\mu, \mu; \|\sigma_{ij}\|)$, где

$$\|\sigma_{ij}\|=\begin{Vmatrix}\sigma^2 & \sigma^2\\ \sigma^2 & 2\sigma^2\end{Vmatrix}.$$

7.5. Пусть $f(x_1, x_2)$ — ф. п. в. двумерного нормального распределения $N(\{0\}, \|\sigma_{ij}\|)$ с круговой симметрией. Показать, что интеграл от $f(x_1, x_2)$

по квадрату $(-k, -k)$, $(-k, +k)$, $(k, -k)$, (k, k) меньше интеграла по кругу $x_1^2+x_2^2<4k^2/\pi$, и, следовательно,

$$\frac{1}{\sqrt{2\pi}}\int_0^k e^{-\frac{1}{2}t^2}\,dt<\frac{1}{2}\sqrt{1-e^{2k^2/\pi}}$$

— результат, принадлежащий Уильямсу (1946).

7.6. Пусть $f(x_1, x_2)$ — ф. п. в. двумерного нормального распределения $N(\{0\}, \|\sigma_{ij}\|)$, где

$$\|\sigma_{ij}\|=\begin{Vmatrix}1 & \rho\\ \rho & 1\end{Vmatrix}.$$

Пусть $F_1(x)$ и $F_2(x)$ — к. ф. р. маргинальных распределений соответственно величин x_1 и x_2. Показать, что коэффициент корреляции между случайными величинами $F_1(x_1)$ и $F_2(x_2)$ равен $\frac{6}{\pi}\sin^{-1}\left(\frac{1}{2}\rho\right)$.

7.7. Показать, что если (x_1, x_2) — двумерная случайная величина, имеющая распределение $N(\{0\}, \|\sigma_{ij}\|)$, где

$$\|\sigma_{ij}\|=\begin{Vmatrix}1 & \rho\\ \rho & 1\end{Vmatrix},$$

то отношение

$$z=\frac{x_1}{x_2}$$

имеет ф. п. в.

$$\frac{\sqrt{1-\rho^2}}{\pi[1-2\rho z+z^2]}\qquad\text{[Филлер (1932)].}$$

7.8. Показать, что если x_1 и x_2 — независимые случайные величины, каждая из которых имеет распределение $R\left(\frac{1}{2}, 1\right)$, то величины $\sqrt{-2\ln x_1}\cos 2\pi x_2$ и $\sqrt{-2\ln x_1}\sin 2\pi x_2$ независимы и каждая имеет распределение $N(0, 1)$ [Бокс и Маллер (1958)].

7.9. Показать, что если x_1 и x_2 — независимые случайные величины, имеющие соответственно гамма-распределения $G(k_1)$ и $G(k_2)$, то величины x_1+x_2 и $\frac{x_1}{x_1+x_2}$ независимы и имеют соответственно гамма-распределение $G(k_1+k_2)$ и бета-распределение $\mathrm{Be}(k_1+k_2)$.

7.10. Показать, что если x_1 и x_2 — независимые случайные величины с гамма-распределениями $G(\nu)$ и $G\left(\nu+\frac{1}{2}\right)$, то случайная величина $y=2\sqrt{x_1, x_2}$ имеет гамма-распределение $G(2\nu)$.

7.11. Двумерная случайная величина имеет ф. п. в.

$$f(x)=\begin{cases}\dfrac{1}{\Gamma(k_1)\Gamma(k_2)}x_1^{k_1-1}(x_2-x_1)^{k_2-1}e^{-x_2}, & 0<x_1<x_2<\infty,\\ 0 & \text{в противном случае.}\end{cases}$$

Показать, что маргинальные распределения величин x_1 и x_2 суть соответственно гамма-распределения $G(k_1)$ и $G(k_1+k_2)$.

7.12. Пусть непрерывная случайная величина x такова, что для некоторого положительного целого k величина kx имеет гамма-распределение $G(k)$. Пусть, кроме того, $y \mid x$ — дискретная условная случайная величина, имеющая распределение Пуассона $\mathrm{Po}(x)$. Показать, что безусловное распределение величины y есть биномиальное распределение времени ожидания.

7.13. Показать, что если x_1 и x_2 — независимые случайные величины, имеющие бета-распределения $\mathrm{Be}(\nu_1, \nu_2)$ и $\mathrm{Be}\left(\nu_1+\frac{1}{2}, \nu_2\right)$, то $\sqrt{x_1x_2}$ имеет бета-распределение $\mathrm{Be}(2\nu_1, 2\nu_2)$.

7.14. Показать, что если $x_1, \ldots, x_k$ — независимые случайные величины, каждая из которых имеет равномерное распределение $\mathrm{Re}\left(\frac{1}{2}, 1\right)$, то $-\ln(x_1\cdots x_k)$ имеет гамма-распределение $G(k)$.

7.15. Воспользовавшись характеристическими функциями, показать, что если u — случайная величина, имеющая χ^2-распределение $C(k)$, то предельное распределение величины $\frac{u-k}{\sqrt{2k}}$ при $k\to\infty$ есть $N(0, 1)$.

7.16. Пусть x — случайная величина с к. ф. р. $F(x)$ и характеристической функцией $\varphi(t)$. Если для каждого положительного целого числа n $[\varphi(t)]^{\frac{1}{n}}$ является характеристической функцией некоторой случайной величины, то x называется *безгранично делимой* случайной величиной, т. е. для каждого n $F(x)$ представляет собой распределение суммы n независимых случайных величин, из которых каждая имеет $[\varphi(t)]^{\frac{1}{n}}$ своей характеристической функцией. Показать, что случайная величина, имеющая распределение Пуассона, гамма- или нормальное распределение, является безгранично делимой. [По поводу теории безгранично делимых случайных величин см. Гнеденко и Колмогоров (1949).]

7.17. Показать, что при $k\to\infty$ ф. п. в. t-распределения Стьюдента (7.8.4) сходится к ф. п. в. распределения $N(0, 1)$ для любого t.

7.18. Показать, что если $x_1, \ldots, x_k$ — независимые случайные величины, каждая из которых имеет распределение $N(\mu, \sigma^2)$, и, кроме того, $c_1, \ldots, c_k$ — вещественные константы, причем $\sum_{i=1}^{k} c_i = 1$ и $\sum_{i=1}^{k} c_i^2 = 1$, то форма $c_1x_1 + \ldots + c_kx_k$ также имеет распределение $N(\mu, \sigma^2)$. Показать, что этим условиям не может удовлетворять никакой набор строго положительных констант c_i.

7.19. Пусть $x_1, \ldots, x_k$ — независимые случайные величины, каждая из которых имеет распределение $N(0, 1)$. Пусть, далее, $y_1, \ldots, y_s$, $s \leqslant k$, — новые случайные величины, определенные формулой $y_p = c_{p1}x_1 + \ldots + c_{pk}x_k$, $p = 1, \ldots, s$, где $\sum_{i=1}^{k} c_{pi}c_{qi} = \delta_{pq}$ (символ Кронекера). Показать, что $y_1, \ldots, y_s$ независимы и все имеют распределение $N(0, 1)$.

7.20. С помощью соответствующего дифференцирования характеристической функции (7.4.17) k-мерной случайной величины $(x_1, \ldots, x_k)$, имеющей распределение $N(\{\mu_i\}, \|\sigma_{ij}\|)$, показать, что $\mathcal{E}(x_i) = \mu_i$ и $\mathcal{E}(x_i-\mu_i) \times (x_j-\mu_j) = \sigma_{ij}$.

7.21. Используя характеристические функции, показать, что если $(x_1, \ldots, x_k)$ и $(x_1', \ldots, x_k')$ — независимые k-мерные случайные величины, имеющие распределения $N(\{\mu_i\}, \|\sigma_{ij}\|)$ и $N(\{\mu_i'\}, \|\sigma_{ij}'\|)$, то k-мерная случайная величина $(x_1+x_1', \ldots, x_k+x_k')$ имеет нормальное распределение $N(\{\mu_i+\mu_i'\}, \|\sigma_{ij}+\sigma_{ij}'\|)$. Другими словами, k-мерное нормальное распределение является воспроизводящим по вектору средних и своей ковариационной матрице.

7.22. Пусть μ — «истинная» длина основного стержня. Обозначим через x_1 случайную величину, равную длине копии стержня, которую будем считать копией первого порядка. Обозначим через x_2 случайную величину, равную длине копии x_1, которую будем считать копией второго порядка, и т. д. Величина x_k будет, таким образом, длиной копии основного стержня

k-го порядка. Показать, что если $(x_1 - \mu)$, $(x_2 - x_1)$, ..., $(x_k - x_{k-1})$ — независимые случайные величины, каждая из которых имеет нормальное распределение $N(0, 1)$, то величина $(x_1, \ldots, x_k)$ имеет k-мерное нормальное распределение $N(\{\mu_i\}, \|\sigma_{ij}\|)$, где $\mu_i = \mu$ и

$$\sigma_{ij} = \begin{Vmatrix} 1 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 2 & 2 & \ldots & 2 & 2 \\ 1 & 2 & 3 & \ldots & 3 & 3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 2 & 3 & \ldots & k-1 & k-1 \\ 1 & 2 & 3 & \ldots & k-1 & k \end{Vmatrix}.$$

7.23. Показать, что если $(x_1, \ldots, x_k)$ — векторная случайная величина, имеющая k-мерное сферическое нормальное распределение $N(\{\mu_i\}, \|\delta_{ij}\sigma^2\|)$, где δ_{ij} — символ Кронекера, и, кроме того,

$$r = [(x_1 - \mu_1)^2 + \ldots + (x_k - \mu_k)^2]^{\frac{1}{2}},$$

то

$$\mathscr{E}(r) = \sqrt{2}\,\sigma\,\frac{\Gamma\left(\frac{1}{2}k + \frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}k\right)}.$$

7.24. Пусть в некотором классе объектов действует «закон смертности» так, что вероятность взятому наудачу объекту «умереть» в течение интервала времени $(t, t + dt)$ (в соответствующим образом выбранных единицах времени) равна

$$\frac{t^{k-1}e^{-t}}{\Gamma(k)}\,dt.$$

Как только объект «умирает», он заменяется другим того же самого типа. Как только этот новый объект «умирает», он снова заменяется объектом того же типа и т. д. Показать, что вероятность «смерти» в течение интервала $(t, t + dt)$ r-го объекта в такой последовательности есть

$$\frac{t^{rk-1}e^{-t}}{\Gamma(rk)}\,dt.$$

7.25. *(Продолжение). Процесс восстановления с фиксированным общим «законом смертности».* Пусть $f(t)\,dt$ есть вероятность того, что некоторый объект в начальном (нулевом) поколении должен быть заменен «наследником» в течение интервала $(t, t + dt)$, и $g_n(\tau)\,d\tau$ есть вероятность того, что некоторый объект в n-м поколении должен быть заменен «наследником» в течение интервала $(\tau, \tau + d\tau)$. Предполагая, что объекты в любом поколении имеют тот же «закон смертности», что и в начальном поколении, показать, что

$$g_{n+1}(t) = \int_0^t g_n(\tau) f(t - \tau)\,d\tau,$$

и, следовательно, если существует «установившийся» закон замещения объектов $g(t)_t$, он должен удовлетворять интегральному уравнению $g(t) = \int_0^t g(\tau) f(t - \tau)\,d\tau$ [Лотка (1939) и Смит (1958)].

7.26. Пусть $x_1, \ldots, x_{k+1}$ — независимые случайные величины, каждая из которых имеет распределение $N(0, 1)$. Введем новые величины

$$y_i = \frac{x_i^2}{x_1^2 + \ldots + x_{k+1}^2}, \qquad i = 1, \ldots, k.$$

Показать, что $(y_1, \ldots, y_k)$ имеет ф. п. в., равную $k!$ внутри симплекса $\{(y_1, \ldots, y_k): y_1 > 0, \ldots, y_k > 0, y_1 + \ldots + y_k < 1\}$ и равную 0 вне его. Другими словами, $(y_1, \ldots, y_k)$ имеет k-мерное распределение Дирихле $D(1, 1, \ldots, 1; 1)$.

7.27. Доказать теорему **7.7.1.**

7.28. Пусть $y_{k_2} = k_1 \tau$, где τ — случайная величина, имеющая распределение Снедекора $S(k_1, k_2)$. Показать, что последовательность случайных величин $(y_1, y_2, \ldots)$ сходится по распределению к случайной величине, имеющей χ^2-распределение с k_1 степенями свободы.

7.29. Пусть $(x_1, \ldots, x_k)$ — k-мерная случайная величина, ф. п. в. которой равна $g(a_1 x_1 + \ldots + a_k x_k)$ в области $x_1 > 0, \ldots, x_k > 0$ и равна 0 вне ее, где $a_1, \ldots, a_k$ суть положительные константы. Показать, что ф. п. в. случайной величины $y = a_1 x_1 + \ldots + a_k x_k$ дается выражением

$$\frac{1}{(a_1 \ldots a_k)\, \Gamma(k)}\, y^{k-1} g(y), \qquad y \geqslant 0.$$

7.30. Пусть $(x_1, \ldots, x_k)$ — k-мерная случайная величина, имеющая ф. п. в. вида $g(y)$, где $y = \sum_{i,j=1}^{k} a_{ij} x_i x_j$, причем матрица констант $\|a_{ij}\|$ симметрична и положительно определена. Показать, что случайная величина y имеет ф. п. в.

$$\frac{\pi^{\frac{k}{2}}}{\sqrt{|a_{ij}|}\, \Gamma\left(\frac{k}{2}\right)}\, y^{\frac{k}{2}-1} g(y), \qquad y \geqslant 0.$$

7.31. Показать, что если $(x_1, \ldots, x_k)$ — k-мерная случайная величина, имеющая распределение $N(\{\mu_i\}, \|\sigma_{ij}\|)$, то условная ф. п. в. $f(x_1, \ldots, x_s \mid x_{s+1}, \ldots, x_k)$ есть ф. п. в. s-мерного распределения $N(\mu_1^*, \ldots, \mu_s^*; \|\sigma_{pq}^*\|)$, где значение μ_p^* дается уравнением

$$\begin{vmatrix} (\mu_p^* - \mu_p) & \sigma_{ps+1} & \cdots & \sigma_{pk} \\ (x_{s+1} - \mu_{s+1}) & \sigma_{s+1,\, s+1} & \cdots & \sigma_{s+1,\, k} \\ \cdot & \cdot & \cdot & \cdot \\ (x_k - \mu_k) & \sigma_{k,\, s+1} & \cdots & \sigma_{kk} \end{vmatrix} = 0, \qquad p = 1, \ldots, s, \text{ и где}$$

$$\sigma_{pq}^* = \frac{\begin{vmatrix} \sigma_{pq} & \sigma_{ps+1} & \cdots & \sigma_{pk} \\ \sigma_{s+1q} & \sigma_{s+1,\, s+1} & \cdots & \sigma_{s+1,\, k} \\ \sigma_{kq} & \sigma_{ks+1} & \cdots & \sigma_{kk} \end{vmatrix}}{\begin{vmatrix} \sigma_{s+1,\, s+1} & \cdots & \sigma_{s+1,\, k} \\ \cdot & \cdot & \cdot \\ \sigma_{k,\, s+1} & \cdots & \sigma_{kk} \end{vmatrix}}, \qquad p, q = 1, \ldots, s.$$

7.32. *Процесс Пуассона.* Рассмотрим следующий подход к формуле (7.5.11). Обозначим через E событие, состоящее в том, что в течение интервала времени $(0, t + \Delta t)$ события C произошли точно k раз. Тогда E будет объединением взаимно исключающих событий $E_0, E_1, \ldots, E_k$, где E_i состоит в наступлении C точно i раз в интервале времени $(t, t + \Delta t)$ и точно $(k - i)$ раз в интервале времени $(0, t)$. Поэтому

$$P(E) = P(E_0) + P(E_1) + \ldots + P(E_k).$$

Если мы предположим, что наступление любого числа событий C в течение интервала I не зависит от наступления того или иного числа событий C в течение интервала I', где $I \cap I' = \varnothing$, то вероятности наступления точно 0, 1, ..., k событий C в течение интервала $(t, t+\Delta t)$ будут равны соответственно

$$1-\lambda\Delta t + o(\Delta t), \quad \lambda\Delta t + o(\Delta t), \ldots, (\lambda\Delta t)^k + o((\Delta t)^2),$$

и, обозначив $P(E)$ через $f_k(t+\Delta t)$, мы имеем

$$P(E_0) = f_k(t)(1-\lambda\Delta t) + o(\Delta t), \quad P(E_1) = f_{k-1}(t)(\lambda\Delta t) + o(\Delta t),$$

тогда как $P(E_2), \ldots, P(E_k)$ будут величинами высшего порядка малости по сравнению с Δt. Отсюда вытекает уравнение

$$f_k(t+\Delta t) = f_k(t)(1-\lambda\Delta t) + f_{k-1}(t)\lambda\Delta t + o(\Delta t)$$

для $k = 1, 2, \ldots$ (Заметим, что для $k=1$ $f_{-1}(t) = 0$.)

Показать, что решениями этого множества дифференциальных уравнений служат функции $f_k(t)$, даваемые (7.5.11). Обратим внимание, что $\{f_k(t): t>0\}$ доставляет нам пример случайного процесса с непрерывным параметром t. Этот частный процесс называется *процессом Пуассона*.

7.33. *Процесс рождения* (Юл (1924)). Допустим, что мы имеем совокупность объектов, которые могут производить (или «рождать») новые объекты, причем никакие объекты из совокупности не исчезают (не «умирают»). Обозначим через $f_k(t)$ вероятность того, что в совокупности в момент времени t будет находиться точно k объектов. Пусть E — событие, состоящее в том, что в момент $t+\Delta t$ совокупность насчитывает k объектов. Тогда E составляется из несовместных событий $E_0, E_1, \ldots, E_k$, где E_i есть событие, состоящее в том, что в момент времени t в совокупности имеется $k-i$ объектов и в течение интервала $(t+\Delta t)$ рождается еще i объектов. Введя предположения относительно независимости наступления рождений в двух неперекрывающихся интервалах времени, подобные тем, которые были сделаны при рассмотрении наступлений событий C в задаче **7.32**, показать, что $f_k(t)$ удовлетворяет системе дифференциальных уравнений

$$f'_k(t) = (k-1)\lambda f_{k-1}(t) - k\lambda f_k(t), \qquad k = 1, 2, \ldots$$

Показать, что если $k=m$ в момент $t=0$ и $f_m(0)=1$, $f_k(0)=0$, $k=m, m+1, \ldots$, то решение системы дифференциальных уравнений дается формулой

$$f_k(t) = \binom{k-1}{m-1} e^{-m\lambda t}(1-e^{-\lambda t})^{k-m}.$$

7.34. *Простая очередь.* Предположим, что среднее число посещений клиентами пункта обслуживания, приходящихся на единицу времени, есть λ, а среднее число уходов клиентов (после завершения обслуживания) на единицу времени есть μ. Пусть $f_k(t)$ означает вероятность того, что в момент t на пункте будет существовать очередь *), состоящая из k клиентов. Таким образом, $f_k(t+\Delta t)$ есть вероятность, что в момент $t+\Delta t$ очередь будет насчитывать k клиентов. Если E — событие, состоящее в том, что в момент $t+\Delta t$ очередь насчитывает k клиентов, и $E_{ij(k-i+j)}$ — событие, состоящее в том, что в момент t очередь насчитывает $k-i+j$ клиентов и в течение интервала времени $(t, t+\Delta t)$ на пункт прибудут точно i клиентов, а точно j клиентов покинут его (после обслуживания), то события $E_{ij(k-i+j)}$ не пере-

---

*) Под очередью мы понимаем здесь тех клиентов, которые либо ожидают обслуживания, либо уже обслуживаются. Некоторые авторы используют этот термин, обозначая только тех клиентов, которые ждут обслуживания.

секаются и их соединение равно E. При этом (i, j) суть пары неотрицательных целых чисел таких, что $k-i+j \geqslant 0$. Следовательно, мы имеем

$$f_k(t+\Delta t)=P(E)=\sum_{i,j} P(E_{ij(k-i+j)}).$$

Допустим, мы теперь сделаем предположения относительно независимости прибытий и отбытий клиентов в неперекрывающихся интервалах времени, подобные тем, которые были сделаны для наступлений событий C в задаче **7.32.** При этих предположениях единственными событиями, вероятностями которых нельзя пренебречь, будут события

$$E_{00(k)}, \quad E_{10(k-1)}, \quad E_{01(k+1)}.$$

Их вероятности равны $f_k(t)[1-(\lambda+\mu)\Delta t]+o(\Delta t)$, $f_{k-1}(t)\lambda\Delta t+o(\Delta t)$ и $f_{k+1}(t)\mu\Delta t+o(\Delta t)$. Вероятности всех других событий будут иметь порядок $(\Delta t)^2$ или выше. Показать, что $f_k(t)$ удовлетворяют следующей системе дифференциальных уравнений:

$$f'_k(t)=[\lambda f_{k-1}(t)+\mu f_{k+1}(t)-(\lambda+\mu)f_k(t)], \qquad k=1, 2, \ldots$$

За более подробной теорией очередей читатель отсылается к работам Феллера (1957), Кендалла (1951, 1953) и Морса (1958). Книга Морса содержит важную библиографию.

7.35. *(Продолжение).* Показать, что в случае установившегося режима, где $f'_k(t)=0$, решение соответствующего разностного уравнения

$$\lambda f_{k-1}+\mu f_{k+1}-(\lambda+\mu)f_k=0,$$

где $f_{-1}=0$, имеет вид $f_k=\rho^k f_0$, где $\rho=\dfrac{\lambda}{\mu}$. Если условия позволяют иметь очередь лишь длины n, то

$$f_0=\frac{1-\rho}{1-\rho^{n+1}}.$$

Показать также, что средняя длина очереди в случае установившегося режима равна

$$\mathfrak{E}(k)=\frac{\rho-(n+1)\rho^{n+1}+n\rho^{n+2}}{(1-\rho)(1-\rho^{n+1})}$$

и средняя длина очереди для случая $\rho=1$ (средние скорости прибытий и отбытий равны) равна $\dfrac{1}{2}n$, а для $\rho<1$ при $n=\infty$ равна $\dfrac{\rho}{1-\rho}$. Кроме того, показать, что среднее $\mathfrak{E}(l)$ числа клиентов l, *ждущих обслуживания*, дается формулой

$$\mathfrak{E}(l)=\sum_{i=1}^{n}(i-1)f_i=\frac{\rho^2-n\rho^{n+1}+(n-1)\rho^{n+2}}{(1-\rho)(1-\rho^n)}.$$

ГЛАВА 8

ТЕОРИЯ ВЫБОРОЧНОГО МЕТОДА

8.1. Определение случайной выборки

Пусть x — одномерная случайная величина, заданная в R_1, с к. ф. р. $F(x)$. *Случайной выборкой объема n из совокупности с к. ф. р. $F(x)$* называется n-мерная случайная величина $(x_1, \ldots, x_n)$ с к. ф. р.

$$\prod_{\xi=1}^{n} F(x_\xi), \tag{8.1.1}$$

определенная в выборочном пространстве $R_n = R_1^{(1)} \times R_1^{(2)} \times \ldots \times R_1^{(n)}$, где $R_1^{(\xi)}$ — одномерное выборочное пространство (вещественная прямая) переменной x_ξ.

Следует специально отметить, что элементы или компоненты $x_1, \ldots, x_n$ выборки взаимно независимы и имеют одинаковые к. ф. р. В математической статистике такое определение случайной выборки было впервые введено Фишером (1915), хотя он в действительности не использовал термин «выборочное пространство» по отношению к R_n. Однако до 1940 года термин «выборочное пространство» употреблялся в этом смысле, а затем уже в более широком смысле, указанном в главе 1. Слово «случайная» было введено в основном с описательной целью и в дальнейшем обычно будет опускаться. Мы также часто будем сокращать запись, обозначая выборку $(x_1, \ldots, x_n)$ через O_n. Для краткости мы будем обычно говорить, что $(x_1, \ldots, x_n)$ (или O_n) есть выборка объема n из $F(x)$. Заметим, что выборка объема n из $F(x)$ является простым примером конечного стохастического процесса. Иногда она называется *простым случайным выбором.*

Если x — случайная величина дискретного типа (дискретная случайная величина) с ф. в. $p(x)$, то выборка имеет ф. в.

$$\prod_{\xi=1}^{n} p(x_\xi), \tag{8.1.2}$$

определенную на R_n.

Если x — случайная величина непрерывного типа (непрерывная случайная величина), то выборка имеет в. э.

$$f(x_1) \ldots f(x_n)\, dx_1 \ldots dx_n, \tag{8.1.3}$$

определенный на R_n.

Для краткости мы будем часто ссылаться на $(x_1, \ldots, x_n)$ как на выборку из $p(x)$ в дискретном случае и как на выборку из $f(x)$ в непрерывном случае.

Замечание. x_1 удобно понимать как случайную величину, обозначающую значение x, получающееся при первом «выборе» из данной совокупности, x_2 — при втором «выборе» и т. д.

Например, если мы бросаем «*правильную*» игральную кость последовательно n раз, то мы можем рассматривать x_ξ как случайную величину, обозначающую число очков, выпадающих при ξ-м бросании, $\xi = 1, \ldots, n$. Наша выборка O_n в этом случае будет состоять из независимых случайных величин $x_1, \ldots, x_n$, каждая из которых имеет ф. в.

$$p(x) = \frac{1}{6}, \qquad x = 1, \ldots, 6.$$

Функция вероятности выборки O_n определена в каждой из 6^n точек сосредоточения массы R_n, координаты которых соответствуют 6^n *возможным* последовательностям граней, получающимся при n-кратном бросании игральной кости, и равна в этих точках

$$p(x_1) \ldots p(x_n) = \left(\frac{1}{6}\right)^n.$$

В теории выборочного метода обычно интересуются распределениями различного рода функций от n случайных величин, входящих в выборку. Этими функциями могут быть, например, *выборочная сумма* $z = \sum_{\xi=1}^{n} x_\xi$, *выборочное среднее значение* $\bar{x} = \frac{1}{n} \sum_{\xi=1}^{n} x_\xi$, *выборочная дисперсия* $s^2 = \frac{1}{n-1} \sum_{\xi=1}^{n} (x_\xi - \bar{x})^2$, наименьший элемент выборки $\min(x_1, \ldots, x_n)$, наибольший элемент выборки $\max(x_1, \ldots, x_n)$ и т. п. Вообще, если $g(x_1, \ldots, x_n)$ — функция, которая сама является случайной величиной, то мы будем обычно заинтересованы в нахождении ее функции распределения. Такая функция $g(x_1, \ldots, x_n)$ называется *статистикой*, и ее к. ф. р. (будем обозначать ее $H(y)$), представляет собой частный случай (2.8.12), т. е.

$$H(y) = P(g(x_1, \ldots, x_n) \leqslant y) = \int\limits_{g^{-1}(y)} dF(x_1) \ldots dF(x_n), \tag{8.1.4}$$

где $g^{-1}(y)$ есть множество в R_n, для которого $g(x_1, \ldots, x_n) \leqslant y$. Аналогично, если $g_i(x_1, \ldots, x_n)$, $i = 1, \ldots, s \leqslant n$, есть s (функционально независимых) статистик, то мы заинтересованы в определении к. ф. р. этих статистик. Функция распределения $g_i(x_1, \ldots, x_n)$ задается формулой

$$H(y_1, \ldots, y_s) = P(g_i(x_1, \ldots, x_n) \leqslant y_i,\ i = 1, \ldots, s) = \\ = \int\limits_{g^{-1}(y_1, \ldots, y_n)} dF(x_1) \ldots dF(x_n), \tag{8.1.5}$$

где $g^{-1}(y_1, \ldots, y_n)$ есть множество в R_n', для которого

$$g_i(x_1, \ldots, x_n) \leqslant y_i, \qquad i = 1, \ldots, s.$$

Случайная величина x может быть k-мерной случайной величиной $(x_1, \ldots, x_k)$ с к. ф. р. $F(x_1, \ldots, x_k)$, определенной в R_k. В этом случае выборка O_n есть kn-мерная случайная величина $(x_{1_\xi}, \ldots, x_{k_\xi};$ $\xi = 1, \ldots, n)$ с к. ф. р.

$$\prod_{\xi=1}^{n} F(x_{1_\xi}, \ldots, x_{k_\xi}), \tag{8.1.6}$$

определенной в $R_{nk} = R_k^{(1)} \times \ldots \times R_k^{(n)}$, где R_k^ξ — выборочное пространство вектора $(x_{1_\xi}, \ldots, x_{k_\xi})$, $\xi = 1, \ldots, n$. И снова задача теории выборочного метода состоит в нахождении функции распределения различных функций от nk случайных величин x_{i_ξ}, $i = 1, \ldots, k$; $\xi = 1, \ldots, n$. Например, мы можем интересоваться распределением таких статистик, как *выборочная сумма* $z_i = \sum_{\xi=1}^{n} x_{i_\xi}$, *выборочное среднее значение* $\bar{x}_i = \frac{1}{n} \sum_{\xi=1}^{n} x_{i_\xi}$ или элементы

$$s_{ij} = \frac{1}{n-1} \sum_{\xi=1}^{n} (x_{i_\xi} - \bar{x}_i)(x_{j_\xi} - \bar{x}_j)$$

выборочной ковариационной матрицы и т. д.

Иногда мы будем иметь дело с теорией выборочного метода для функций двух или более случайных выборок. Например, предположим, что $O_{n_1}: (x_{11}, \ldots, x_{1n_1})$ и $O_{n_2}: (x_{21}, \ldots, x_{2n_2})$ суть выборки объема n_1 и n_2 из совокупностей с к. ф. р. $F_1(x)$ и $F_2(x)$ соответственно. Мы рассматриваем тогда $(x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{2n_2})$ как (n_1+n_2)-мерную случайную величину с к. ф. р., равной

$$\prod_{\xi_1=1}^{n_1} F_1(x_{1\xi_1}) \prod_{\xi_2=1}^{n_2} F_2(x_{2\xi_2}) \tag{8.1.7}$$

в $R_{n_1+n_2}$. Одной из существенных проблем теории выборочного метода, с которой мы столкнемся, будет нахождение распределения одной или более функций от компонент этой (n_1+n_2)-мерной случайной величины, или по крайней мере выяснение некоторых свойств этого распределения. Подобные замечания имеют место и для случаев трех и более выборок.

При решении задач математической статистики, включающих нахождение функций распределения, мы обычно встречаемся с относительно простыми статистиками, такими как: средние, суммы квадратов, частные, ковариации и т. д. Простые и явные выражения для ф. в. и ф. п. в. таких выборочных распределений существуют, как это будет видно из последующих разделов, только для некоторых специальных распределений. Однако можно найти среднее значение, дисперсию, ковариацию (смешанный второй момент) и некоторые другие младшие моменты этих статистик и для гораздо более общих распределений совокупностей, применяя результаты §§ 3.3, 3.4 и 3.5. Мы рассмотрим некоторые из наиболее важных результатов этого типа в следующем параграфе.

8.2. Средние значения и дисперсии выборочных среднего, дисперсии и других симметрических функций

(а) Среднее значение и дисперсия выборочного среднего. Пусть $(x_1, \ldots, x_n)$ — выборка из популяции, распределение которой имеет среднее значение μ и дисперсию σ^2. Рассмотрим задачу нахождения среднего значения и дисперсии выборочного среднего значения $\bar{x}$.

По определению

$$\bar{x} = \frac{1}{n}(x_1 + \ldots + x_n). \tag{8.2.1}$$

Беря среднее значение от обеих частей (8.2.1), найдем

$$\mu(\bar{x}) = \mathcal{E}(\bar{x}) = \frac{1}{n}[\mathcal{E}(x_1) + \ldots + \mathcal{E}(x_n)]. \tag{8.2.2}$$

Но так как $x_1, \ldots, x_n$ — независимые случайные величины, имеющие одну и ту же к. ф. р., то

$$\mathcal{E}(x_1) = \ldots = \mathcal{E}(x_n) = \mu.$$

Подставляя эти значения в (8.2.2), получаем

$$\mu(\bar{x}) = \mu. \tag{8.2.3}$$

Не останавливаясь здесь специально на обсуждении основных понятий и принципов статистического оценивания, мы заметим только, что (8.2.3) утверждает, что $\bar{x}$ есть *несмещенная оценка* для μ. Обсуждению оценок будут посвящены главы 10, 11 и 12.

Теперь рассмотрим дисперсию $\bar{x}$. Имеем

$$\sigma^2(\bar{x}) = \mathcal{E}[\bar{x} - \mu(\bar{x})]^2,$$

но

$$\bar{x} - \mu(\bar{x}) = \frac{1}{n}\sum_{\xi}(x_\xi - \mu).$$

Следовательно,

$$\sigma^2(\bar{x}) = \mathcal{E}\left[\frac{1}{n}\sum_{\xi}(x_\xi - \mu)\right]^2 =$$

$$= \frac{1}{n^2}\sum_{\xi}\mathcal{E}(x_\xi - \mu)^2 + \frac{1}{n^2}\sum_{\xi \neq \eta}\mathcal{E}[(x_\xi - \mu)(x_\eta - \mu)].$$

Так как $x_1, \ldots, x_n$ независимы и одинаково распределены, то

$$\begin{aligned} &\mathscr{E}(x_\xi - \mu)^2 = \sigma^2, \qquad \xi = 1, \ldots, n, \\ &\mathscr{E}[(x_\xi - \mu)(x_\eta - \mu)] = 0, \quad \xi \neq \eta. \end{aligned} \tag{8.2.4}$$

Поэтому

$$\sigma^2(\bar{x}) = \frac{\sigma^2}{n}. \tag{8.2.5}$$

(b) Среднее значение и дисперсия выборочной дисперсии. Рассмотрим теперь *выборочную дисперсию*, которая определяется как

$$s^2 = \frac{1}{n-1}\sum_{\xi=1}^{n}(x_\xi - \bar{x})^2. \tag{8.2.6}$$

Мы можем записать (8.2.6) в виде

$$s^2 = \frac{1}{n-1}\sum_{\xi}\left[(x_\xi - \mu) - \frac{1}{n}\sum_{\eta}(x_\eta - \mu)\right]^2 =$$

$$= \frac{1}{n}\sum_{\xi}(x_\xi - \mu)^2 - \frac{1}{n(n-1)}\sum_{\xi \neq \eta}(x_\xi - \mu)(x_\eta - \mu). \tag{8.2.6a}$$

Беря среднее значение и используя (8.2.4), имеем

$$\mathscr{E}(s^2) = \sigma^2. \tag{8.2.7}$$

Заметим, что причиной, по которой в знаменателе (8.2.6) стоит $n-1$, а не n, является желание сделать $\mu(s^2)$ в точности равным σ^2, т. е. сделать тем самым s^2 *несмещенной оценкой* для σ^2.

Проводя вычисления аналогично тому, как было сделано при определении среднего значения, найдем после некоторых преобразований, что

$$\mathscr{E}[(s^2)^2] = \frac{\mu_4}{n} + \frac{(n-1)^2 + 2}{n(n-1)}\sigma^4, \tag{8.2.8}$$

где μ_4 — четвертый центральный момент распределения совокупности.

Подставляя (8.2.7) и (8.2.8) в

$$\sigma^2(s^2) = \mathscr{E}[(s^2)^2] - [\mu(s^2)]^2,$$

находим для дисперсии s^2 выражение

$$\sigma^2(s^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right). \tag{8.2.9}$$

Мы можем теперь сделать следующий вывод.

8.2.1. *Если $(x_1, \ldots, x_n)$ — выборка из распределения со средним значением μ, дисперсией σ^2 и четвертым центральным моментом μ_4, то выборочное среднее $\bar{x}$ имеет среднее значение и дисперсию*

$$\mu(\bar{x}) = \mu, \qquad \sigma^2(\bar{x}) = \frac{\sigma^2}{n},$$

а среднее значение и дисперсия выборочной дисперсии равны соответственно

$$\mu(s^2) = \sigma^2, \qquad \sigma^2(s^2) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right).$$

(c) k-статистики Фишера. Необходимо отметить, что первые два *выборочных момента*, а именно выборочное среднее значение и выборочная дисперсия, обладают тем свойством, что они являются несмещенными оценками для среднего значения и дисперсии соответственно. Возникает вопрос: обладают ли этим свойством более старшие *центральные выборочные моменты*? r-й центральный выборочный момент определяется как

$$m_r = \frac{1}{n}\sum_{\xi=1}^{n}(x_\xi - \bar{x})^r.$$

Ответ на этот вопрос отрицательный, независимо от того, какая функция от n стоит вместо множителя $\frac{1}{n}$.

Фишер (1928а) ввел функции $k_r(x_1, \ldots, x_n)$, $r = 1, 2, \ldots$, являющиеся наиболее общими однородными полиномами степени r от переменных $x_1, \ldots, x_n$ и подчиняющиеся при этом следующим условиям: 1) функция $k_r(x_1, \ldots, x_n)$ симметрична относительно $x_1, \ldots, x_n$; 2) $\mathfrak{E}(k_r) = k_r$, где k_r согласно определению (5.1.11) — r-й семиинвариант распределения, из которого взята выборка. Такие функции $k_r(x_1, \ldots, x_n)$ называются *k-статистиками.* Для того чтобы выяснить метод определения коэффициентов в k_r в общем случае, достаточно привести построение первых трех k-статистик. Читатель, интересующийся дальнейшими деталями построения k-статистик и теорий выборочного метода для них, а также связанными с этим проблемами, отсылается для консультации к работам Крейга (1928), Фишера (1928а), Корниша и Фишера (1937), Двайера (1938) и Кендалла (1943).

Первые три k-статистики, как можно заметить, имеют вид

$$\begin{aligned} k_1 &= a_1 \sum_\xi x_\xi, \\ k_2 &= a_{21}\sum_\xi x_\xi^2 + a_{22}\sum_{\xi\neq\eta} x_\xi x_\eta, \\ k_3 &= a_{31}\sum_\xi x_\xi^3 + a_{32}\sum_{\xi\neq\eta} x_\xi^2 x_\eta + a_{33}\sum_{\xi\neq\eta\neq\zeta} x_\xi x_\eta x_\zeta. \end{aligned} \tag{8.2.10}$$

Беря средние значения этих трех выражений и приравнивая их первым трем семиинвариантам распределения совокупности, а именно:

$$\begin{aligned} k_1 &= \mu_1' = \mu, \\ k_2 &= \mu_2' - (\mu_1')^2 = \sigma^2, \\ k_3 &= \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3, \end{aligned} \tag{8.2.11}$$

где μ'_1, μ'_2 и μ'_3 — первые три момента распределения совокупности, мы определим искомые коэффициенты. Подставляя таким образом найденные коэффициенты в (8.2.10), мы получаем в качестве первых трех k-статистик

$$
\begin{aligned}
k_1 &= \frac{1}{n} z_1 = \bar{x}, \\
k_2 &= \frac{1}{n(n-1)} (nz_2 - z_1^2) = s^2, \\
k_3 &= \frac{1}{n(n-1)(n-2)} (n^2 z_3 - 3nz_1 z_2 + 2z_1^3),
\end{aligned}
\tag{8.2.12}
$$

где $z_i = \sum\limits_{\xi} x_\xi^i$, $i = 1, 2, 3$.

Отметим, что первые две k-статистики — это просто $\bar{x}$ и s^2, чьи средние значения есть μ и σ^2, являющиеся в свою очередь первыми двумя семиинвариантами k_1 и k_2.

(d) Среднее и дисперсия некоторых симметрических функций выборки. Необходимо заметить, что выборочное среднее значение $\bar{x}$, выборочная дисперсия s^2, выборочные моменты и выборочные k-статистики все являются симметрическими функциями элементов выборки $(x_1, \ldots, x_n)$. Теперь рассмотрим задачу нахождения среднего значения и дисперсии более общего класса симметрических функций выборки.

Пусть $(x_1, \ldots, x_r)$ — выборка объема r из к. ф. р. $F(x)$ и $g(x_1, \ldots \ldots, x_r)$ — функция от $(x_1, \ldots, x_r)$ такая, что

$$
\mathscr{E}(g^i(x_1, \ldots, x_r)) = \begin{cases} \theta_1, & i = 1, \\ \theta_2, & i = 2, \end{cases}
\tag{8.2.13}
$$

где θ_1 и θ_2 конечны. Для выборки $(x_1, \ldots, x_n)$ объема n, где $n \geqslant r$, определим функцию

$$
g_0(x_{\eta_1}, \ldots, x_{\eta_r}) = \frac{1}{r!} \sum_p g(x_{\xi_1}, \ldots, x_{\xi_r}),
\tag{8.2.14}
$$

где $\eta_1, \ldots, \eta_r$ — набор r целых чисел из $1, \ldots, n$ таких, что $\eta_1 < \ldots \ldots < \eta_r$, и $\sum\limits_p$ означает суммирование по всем $r!$ перестановкам $(\xi_1, \ldots \ldots, \xi_r)$ набора $(\eta_1, \ldots, \eta_r)$. Заметим, что $g_0(x_{\eta_1}, \ldots, x_{\eta_r})$ — симметрическая функция $(x_{\eta_1}, \ldots, x_{\eta_r})$. Пусть

$$
Q_{[r]}(x_1, \ldots, x_n) = \binom{n}{r}^{-1} \sum_c g_0(x_{\eta_1}, \ldots, x_{\eta_r}),
\tag{8.2.15}
$$

где $\sum\limits_c$ означает суммирование по всем $\binom{n}{r}$ наборам $(\eta_1, \ldots, \eta_r)$ из $(1, \ldots, n)$. $Q_{[r]}(x_1, \ldots, x_n)$ есть симметрическая функция компонент выборки $(x_1, \ldots, x_n)$. Тогда $\mathscr{E}(g(x_{\xi_1}, \ldots, x_{\xi_r}))$, которое предполагается конечным, имеет одно и то же значение, а именно θ_1, для всех перестановок $(x_{\xi_1},\ldots, x_{\xi_r})$ элементов выборки $(x_1,\ldots, x_n)$. Более того, очевидно, что

$$\mathscr{E}(g_0(x_{\eta_1},\ldots, x_{\eta_r}))=\frac{1}{r!}\sum_P \mathscr{E}(g(x_{\xi_1},\ldots, x_{\xi_r}))=\theta_1, \quad (8.2.16)$$

а также

$$\mathscr{E}(Q_{[r]}(x_1,\ldots, x_n))=\binom{n}{r}^{-1}\sum_C \mathscr{E}(g_0(x_{\eta_1},\ldots, x_{\eta_r}))=\theta_1. \quad (8.2.17)$$

Рассмотрим теперь дисперсию $Q_{[r]}(x_1,\ldots, x_n)$. Имеем

$$\sigma^2(Q)=\mathscr{E}(Q_{[r]}^2(x_1,\ldots, x_n))-\theta_1^2. \quad (8.2.18)$$

Но

$$\mathscr{E}(Q_{[r]}^2(x_1,\ldots, x_n))=\binom{n}{r}^{-2}{\sum_C}' \mathscr{E}(g_0(x_{\eta_1},\ldots, x_{\eta_r})\,g_0(x_{\zeta_1},\ldots, x_{\zeta_r})), \quad (8.2.19)$$

где $\eta_1,\ldots, \eta_r$ и $\zeta_1,\ldots, \zeta_r$ — два набора из целых чисел $1,\ldots, n$ такие, что $\eta_1<\ldots<\eta_r$ и $\zeta_1<\ldots<\zeta_r$, а ${\sum\limits_C}'$ означает суммирование по всем парам таких наборов. Очевидно, что если $\eta_1,\ldots, \eta_r$ и $\zeta_1,\ldots, \zeta_r$ — пара наборов с j общими целыми числами, то тогда $\mathscr{E}(g_0(x_{\eta_1},\ldots, x_{\eta_r})g_0(x_{\zeta_1},\ldots, x_{\zeta_r}))$ имеет одну и ту же величину, скажем φ_j, для всех таких пар наборов. Из неравенства Шварца следует, что если θ_1 и θ_2 в (8.2.13) конечны, то все φ_j, $j=0,1,\ldots, r$, также конечны. Из комбинаторных рассмотрений очевидно, что имеется $\binom{n}{r}\binom{r}{j}\binom{n-r}{r-j}$ таких пар наборов и j может принимать значения $0,1,\ldots, r$. Поэтому

$$\mathscr{E}(Q_{[r]}^2(x_1,\ldots, x_n))=\binom{n}{r}^{-1}\sum_{j=0}^{r}\binom{r}{j}\binom{n-r}{r-j}\varphi_j. \quad (8.2.20)$$

Таким образом, получаем

$$\sigma^2(Q_{[r]})=\binom{n}{r}^{-1}\sum_{j=0}^{r}\binom{r}{j}\binom{n-r}{r-j}\varphi_j-\theta_1^2. \quad (8.2.21)$$

Если $g(x_{\xi_1},\ldots, x_{\xi_r})$ выбрать так, чтобы $\theta_1=0$, то тогда, так как $\varphi_0=\theta_1^2$, $\sigma^2(Q_{[r]})$ приводится к виду

$$\sigma^2(Q_{[r]})=\binom{n}{r}^{-1}\sum_{j=0}^{r}\binom{r}{j}\binom{n-r}{r-j}\varphi_j. \quad (8.2.22)$$

Этот результат принадлежит Гофдингу (1948а).

Суммируя, получаем следующую теорему.

8.2.2. *Предположим, что $(x_1, \ldots, x_n)$ — выборка из произвольного распределения, и пусть $g(x_1, \ldots, x_r)$ — функция, имеющая* *конечные первый и второй моменты θ_1 и θ_2. Для выборки $(x_1, \ldots, x_n)$, $n \geqslant r$, из этого же распределения пусть $Q_{[r]}(x_1, \ldots, x_n)$ определено согласно* (8.2.15). *Тогда $Q_{[r]}(x_1, \ldots, x_n)$ есть несмещенная оценка для θ_1 и дисперсия этой оценки дается* (8.2.21).

Гофдинг показал, что если θ_1 и θ_2 конечны, то асимптотическое распределение $\sqrt{n}(Q_{[r]} - \theta)/\sigma(Q_{[r]})$ при $n \to \infty$ есть $N(0,1)$. Он обобщил этот результат, так же как и **8.2.2**, на случай, когда $Q_{[r]}(x_1, \ldots, x_n)$ есть вектор.

8.3. Теория выборочного метода для выборочных сумм и средних значений

(а) Итеративный метод. Мы рассмотрим сначала итеративный метод для нахождения функции распределения выборочной суммы или среднего значения. Это даст нам прямой подход к исследованию проблемы, который в некоторых специальных случаях довольно просто приводит к решению.

Рассмотрим случай выборки объема 2. Если $G_2(z)$ — к. ф. р. случайной величины z, то

$$G_2(z) = P(x_1 + x_2 \leqslant z) = \int_E d(F(x_1)\ F(x_2)), \tag{8.3.1}$$

где E — событие в R_2, для которого $x_1 + x_2 \leqslant z$. Если теперь мы применим **3.7.2** для случая, когда x_1 и x_2 независимы, и положим $g(x_1, x_2) = 1$ для всех точек в E и $g(x_1, x_2) = 0$ для всех точек в $\bar{E}$, то сможем написать

$$\int_E d(F(x_1)\ F(x_2)) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-x_1} dF(x_2) \right] dF(x_1) = \\ = \int_{-\infty}^{\infty} F(z - x_1)\, dF(x_1). \tag{8.3.2}$$

Таким образом, для выборки объема 2

$$G_2(z) = \int_{-\infty}^{\infty} F(z - x_1)\, dF(x_1). \tag{8.3.3}$$

Обобщая эти рассуждения на выборку объема n и обозначая через $G_n(z)$ к. ф. р. z, аналогично находим, что $G_n(z)$ может быть выражено следующим повторным интегралом:

$$G_n(z) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{z-x_1-\ldots-x_{n-3}} \int_{-\infty}^{z-x_1-\ldots-x_{n-2}} F(z - x_1 - \ldots - x_{n-1}) \times \\ \times dF(x_{n-1})\, dF(x_{n-2}) \ldots dF(x_1) \tag{8.3.4}$$

функция распределения выборочного среднего значения $\bar{x}$ (обозначим ее через $H_n(\bar{x})$) определяется соотношением $H_n(\bar{x}) \equiv G_n(n\bar{x})$. Поэтому

8.3.1. *Если* $(x_1, \ldots, x_n)$ — *выборка объема* n *из к. ф. р.* $F(x)$, *то* $G_n(x)$ — *к. ф. р. выборочной суммы* z *дается выражением* (8.3.4), *а* $H_n(\bar{x})$ — *к. ф. р. выборочного среднего значения — соответственно соотношением* $H_n(\bar{x}) \equiv G_n(n\bar{x})$.

О функции распределения $G_n(z)$ часто говорят как о *свертке* функций распределения $F(x_1), \ldots, F(x_n)$. Аналогичный процесс применяется и непосредственно для получения свертки n независимых случайных величин с произвольными к. ф. р. $F_1(x_1), \ldots, F_n(x_n)$. Свертка в этом случае есть к. ф. р. суммы n независимых случайных величин.

В случае выборки из совокупности с дискретным распределением, ф. в. которого равна $p(x)$, можно утверждать, что ф. в. z, скажем $p_n(z)$, удовлетворяет следующему уравнению:

$$p_n(z) = \sum_x p(z-x)\, p_{n-1}(x), \tag{8.3.4a}$$

где $p_1(x) \equiv p(x)$.

Аналогично, если мы имеем выборку из популяции с непрерывным распределением, ф. п. в. которого равна $f(x)$, то ф. п. в. z, назовем ее $f_n(z)$, дается следующим соотношением:

$$f_n(z) = \int_{-\infty}^{\infty} f(z-x)\, f_{n-1}(x)\, dx, \tag{8.3.4b}$$

где $f_1(x) \equiv f(x)$.

Пример. Интересен частный случай (8.3.4b), в котором $f(x)$ — ф. п. в. равномерного распределения $R\left(\frac{1}{2}, 1\right)$. В этом случае

$$f_1(z) = f(z) = \begin{cases} 0, & z \leqslant 0, \\ 1, & 0 < z \leqslant 1, \\ 0, & z > 1, \end{cases}$$

$$f_2(z) = \begin{cases} 0, & z \leqslant 0, \\ z, & 0 < z \leqslant 1, \\ z - 2(z-1), & 1 < z \leqslant 2, \\ 0, & z > 2, \end{cases}$$

и, применяя метод математической индукции, получаем

$$f_n(z) = \frac{1}{(n-1)!}\left[z^{n-1} - \binom{n}{1}(z-1)^{n-1} + \binom{n}{2}(z-2)^{n-1} - \ldots + (-1)^k \binom{n}{k}(z-k)^{n-1} \right]$$

для $k < z \leqslant k+1$, $k = 0, 1, \ldots, n-1$; и $f_n(z) = 0$ для $z \leqslant 0$ и $z > n$. Функция вероятности (ф. в.) $\bar{x}$ равна, как можно увидеть, $nf_n(n\bar{x})$ для $k/n < \bar{x} \leqslant (k+1)/n$, $k = 0, 1, \ldots, n-1$. Этот результат принадлежит Лапласу (1814).

(b) Применение характеристических функций. Характеристические функции дают во многих случаях простой и мощный метод нахождения функций распределения выборочных сумм и средних значений.

Обращаясь к (5.3.2), мы, очевидно, получим как следствие теорему

8.3.2. *Если* $(x_1, \ldots, x_n)$ — *выборка из к. ф. р.* $F(x)$ *и характеристическая функция* $F(x)$ *есть* $\varphi(t)$, *то характеристическая функция выборочной суммы* z *равна*

$$[\varphi(t)]^n$$

и характеристическая функция выборочного среднего значения $\bar{x}$ *равна*

$$\left[\varphi\left(\frac{t}{n}\right)\right]^n.$$

Функция же распределения (и ф. п. в. в непрерывном случае) z или $\bar{x}$ может быть определена, во всяком случае теоретически, с помощью своей характеристической функции применением **5.1.2.**

В реальных случаях, за исключением некоторых специальных, вычисление интеграла в правых частях (5.1.14) или (5.1.15) затруднительно. Ирвин (1930) интенсивно использовал технику характеристических функций для нахождения распределений выборочных средних значений для различных к. ф. р. Однако для более важных специальных случаев, которые возникают в теории выборочного метода, мы можем найти функцию распределения z и $\bar{x}$ применением **5.3.3,** используя воспроизводящее свойство функций распределения. Предположим, что $(x_1, \ldots, x_n)$ — выборка из к. ф. р. $F(x;\ \theta)$, характеристическая функция которой равна $\varphi(t;\ \theta)$. Если z — выборочная сумма, то, как мы знаем из **8.3.2,** характеристическая функция z есть $[\varphi(t;\ \theta)]^n$. Но, если $\varphi(t;\ \theta)$ удовлетворяет критерию воспроизводимости по отношению к θ согласно (5.3.7), то

$$[\varphi(t;\ \theta)]^n = \varphi(t;\ n\theta). \qquad (8.3.5)$$

Это означает, что z имеет к. ф. р. $F(z;\ n\theta)$. Но так как $z = n\bar{x}$, где $\bar{x}$ — выборочное среднее значение, то к. ф. р. $\bar{x}$ есть $F(n\bar{x};\ n\theta)$.

Мы можем суммировать результат следующим образом.

8.3.3. Пусть $(x_1, \ldots, x_n)$ — *выборка из к. ф. р.* $F(x;\ \theta)$ *и характеристическая функция* $F(x;\ \theta)$ *есть* $\varphi(t;\ \theta)$. *Тогда, если* $[\varphi(t;\ \theta)]^n = \varphi(t;\ n\theta)$, *то функция распределения выборочной суммы* z *и выборочного среднего значения* $\bar{x}$ *есть* $F(z;\ n\theta)$ *и* $F(n\bar{x};\ n\theta)$ *соответственно.*

Необходимо отметить, что справедливо обобщение этого результата для случая выборок из совокупности с k-мерной функцией распределения. В этом случае z есть вектор $(z_1, \ldots, z_k)$, *где* $z_i = \sum_{\xi=1}^{n} x_{i\xi}$. Параметр θ также может быть вектором с некоторым числом компонент.

Приводимые ниже следствия **8.3.3** дают информацию о теории выборочного метода для выборочной суммы в случае некоторых важных специальных распределений, которые будут использоваться в дальнейшем.

8.3.3a. *Если* $(x_1, \ldots, x_n)$ — *выборка из биномиального распределения* $\mathrm{Bi}(m, p)$, *то распределение выборочной суммы* z *есть биномиальное распределение* $\mathrm{Bi}(mn, p)$.

Характеристическая функция биномиального распределения $\mathrm{Bi}(m, p)$ равна $(q+e^{it}p)^m$, откуда следует, что характеристическая функция для z есть $(q+e^{it}p)^{mn}$ и, следовательно, действительно, z имеет распределение $\mathrm{Bi}(mn, p)$. Отметим, что для $m=1$ мы делаем выборку из *биномиальной совокупности*, имеющей ф. в.

$$p(x)=\begin{cases} q, & x=0, \\ p, & x=1, \end{cases} \tag{8.3.6}$$

и, следовательно, биномиальное распределение, имеющее ф. в. согласно (6.2.2), есть, по сути дела, распределение выборочной суммы $z=\sum_{\xi=1}^{n} x_\xi$, где каждое x_ξ принимает значения 0 или 1 с вероятностями q или p соответственно.

8.3.3b. *Если* $(x_{1\xi}, \ldots, x_{k\xi};\ \xi=1, \ldots, n)$ — *выборка из мультиномиального распределения* $M(m; p_1, \ldots, p_k)$, *то распределение вектора выборочных сумм* $(z_1, \ldots, z_k)$ *тоже мультиномиальное* $M(mn; p_1, \ldots, p_k)$.

Это утверждение проверяется аналогично **8.3.3a.** Заметим, что для $m=1$ мультиномиальное распределение $M(n; p_1, \ldots, p_k)$ представляет собой распределение выборочной суммы из мультиномиального распределения $M(1; p_1, \ldots, p_k)$ для k случайных величин, имеющих ф. в.

$$p(x_1, \ldots, x_k)=p_1^{x_1}\ldots p_k^{x_k}(1-p_1-\ldots-p_k)^{1-x_1-\ldots-x_k}, \tag{8.3.7}$$

где $x_1, \ldots, x_k$ — случайные величины такие, что $x_1+\ldots+x_k\leqslant 1$, принимающие каждая значения либо 0, либо 1.

8.3.3c. *Если* $(x_1, \ldots, x_n)$ — *выборка из распределения Пуассона* $\mathrm{Po}(\mu)$, *то выборочное распределение* z *есть тоже распределение Пауссона* $\mathrm{Po}(n\mu)$.

Действительно, характеристическая функция распределения Пуассона $\mathrm{Po}(\mu)$ равна $e^{-\mu(1-e^{it})}$, откуда следует, что характеристическая функция z есть $e^{-n\mu(1-e^{it})}$. А это — как раз характеристическая функция распределения Пуассона $\mathrm{Po}(n\mu)$, откуда следует, что z имеет распределение Пуассона $\mathrm{Po}(n\mu)$.

8.3.3d. *Если* $(x_1, \ldots, x_n)$ — *выборка из нормального распределения* $N(\mu, \sigma^2)$, *то выборочное распределение* z *есть* $N(n\mu, n\sigma^2)$. *Выборочное распределение* $\bar{x}$ *есть* $N(\mu, \sigma^2/n)$.

Действительно, вспоминая из раздела 7.2, что характеристическая функция нормального распределения $N(\mu, \sigma^2)$ равна $\exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right)$, получаем, что характеристическая функция z равна $\exp\left(i(n\mu)t - \frac{1}{2}(\sigma^2 n)t^2\right)$. А это не что иное, как характеристическая функция $N(n\mu, n\sigma^2)$, и, следовательно, $N(n\mu, n\sigma^2)$ — выборочное распределение z. Характеристическая функция $\bar{x}$ получается заменой t на t/n в характеристической функции z, что дает $\exp\left(i\mu t - \frac{1}{2}(\sigma^2/n)t^2\right)$. А это есть характеристическая функция $N\left(\mu, \frac{\sigma^2}{n}\right)$, и, следовательно, $N\left(\mu, \frac{\sigma^2}{n}\right)$ — выборочное распределение $\bar{x}$.

Сформулируем соответствующее утверждение для нормального распределения k-мерных случайных величин.

8.3.3e. *Если* $(x_{1\xi}, \ldots, x_{k\xi};\ \xi = 1, \ldots, n)$ — *выборка из* k-*мерной нормальной совокупности* $N(\{\mu_i\}, \|\sigma_{ij}\|)$, *то распределение вектора выборочных сумм* $(z_1, \ldots, z_k)$ *есть* $N(\{n\mu_i\}, \|n\sigma_{ij}\|)$. *Распределение вектора выборочных средних* $(\bar{x}_1, \ldots, \bar{x}_k)$ *есть* $N(\{\mu_i\}, \|\sigma_{ij}/n\|)$.

Теорема **8.3.3e** проверяется аналогично **8.3.3d**, и мы оставляем ее как упражнение для читателя.

Приведем, наконец, следствие **8.3.3** для гамма-распределения, доказательство которого мы также оставляем читателю.

8.3.3f. *Если* $(x_1, \ldots, x_n)$ — *выборка из гамма-распределения* $G(\mu)$, *то выборочное распределение* z *тоже есть гамма-распределение* $G(n\mu)$.

Если функция распределения совокупности, из которой берется выборка $(x_1, \ldots, x_n)$, не воспроизводящая, то характеристическая функция не очень удобна, вообще говоря, для нахождения функции распределения выборочной суммы или среднего значения.

Иногда мы встречаемся с линейными функциями от средних значений выборок из нормальной совокупности. Следующее утверждение дает полезный результат относительно распределения таких линейных функций.

8.3.4. *Пусть* $(x_{11}, \ldots, x_{1n_1}), \ldots, (x_{k1}, \ldots, x_{kn_k})$ — k (*независимых*) *выборок из* $N(\mu_1, \sigma_1^2), \ldots, N(\mu_k, \sigma_k^2)$ *соответственно. Пусть* $\bar{x}_1, \ldots, \bar{x}_k$ — *средние значения этих выборок соответственно и пусть* $c_1, \ldots, c_k$ — *постоянные, не все равные нулю. Тогда* $c_1\bar{x}_1 + \ldots + c_k\bar{x}_k$ *имеет выборочное распределение* $N\left(\sum_{i=1}^{k} c_i\mu_i, \sum_{i=1}^{k}\frac{c_i^2\sigma_i^2}{n_i}\right)$.

Доказательство этого утверждения прямо следует из метода характеристических функций, и мы оставляем его как упражнение для читателя. Заметим особо, что если $c_1 = 1$, $c_2 = -1$ и $c_3 = \ldots = c_k = 0$, то легко получить как следствие, что

8.3.4a. *Выборочное распределение разности* $\bar{x}_1 - \bar{x}_2$ *между средними значениями двух (независимых) выборок* $(x_{11}, \ldots, x_{1n_1})$ *и* $(x_{21}, \ldots, x_{2n_2})$ *из* $N(\mu_1, \sigma_1^2)$ *и* $N(\mu_2, \sigma_2^2)$ *соответственно есть* $N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$.

8.4. Теория выборочного метода для некоторых квадратичных форм в выборках из нормального распределения

Основные факты теории выборочного метода для сумм и средних значений выборок из нормальных распределений были даны в **8.3.3d**, **8.3.3e**, **8.3.4** и **8.3.4a.** Настоящий параграф содержит теорию выборочного метода для сумм квадратов, выборочных дисперсий и других квадратичных форм выборок из одномерного нормального распределения. Теория выборочного метода, соответствующая выборкам из нормального распределения k-мерных случайных величин, более сложна, она принадлежит к *многомерному статистическому анализу* и будет рассмотрена в главе 18.

Одним из основных результатов теории выборочного метода для сумм квадратов элементов выборки является следующий:

8.4.1. *Если* $(x_1, \ldots, x_n)$ — *выборка из нормального распределения* $N(\mu, \sigma^2)$, *то выборочное распределение суммы квадратов* $\frac{1}{\sigma^2}\sum_{\xi=1}^{n}(x_\xi - \mu)^2$ *есть* χ^2*-распределение* $C(n)$.

Чтобы доказать это утверждение, достаточно заметить, что если ф. п. в. нормального распределения $N(\mu, \sigma^2)$ подставить в (8.1.3), то в экспоненте плотности функции распределения выборки появится квадратичная форма $\frac{1}{\sigma^2}\sum_{\xi=1}^{n}(x_\xi - \mu)^2$. Как следует из **7.8.2**, эта квадратичная форма имеет χ^2-распределение $C(n)$. Результат, сформулированный в **8.4.1**, впервые был получен Хельмертом (1876a). В определении (8.2.6) выборочной дисперсии s^2 участвует сумма квадратов $\sum_{\xi=1}^{n}(x_\xi - \bar{x})^2$, где $\bar{x}$ — выборочное среднее. Результат теории выборочного метода для этой суммы квадратов в выборках из нормального распределения может быть изложен следующим образом:

8.4.2. *Если* $(x_1, \ldots, x_n)$ — *выборка из нормального распределения* $N(\mu, \sigma^2)$, *то* $\sqrt{n}(\bar{x} - \mu)/\sigma$ *и* $(n-1)s^2/\sigma^2$ *статистически независимы и их выборочные распределения суть* $N(0, 1)$ *и* χ^2*-распределение* $C(n-1)$ *соответственно.*

Мы уже видели на основании **8.3.3d**, что $\bar{x}$ имеет распределение $N(\mu, \sigma^2/n)$, что эквивалентно утверждению о том, что $\sqrt{n}(\bar{x} - \mu)/\sigma$

имеет распределение $N(0, 1)$. Чтобы установить справедливость **8.4.2,** остается показать, что $\sqrt{n}\,(\bar{x}-\mu)/\sigma$ и $(n-1)\,s^2/\sigma^2$ независимы и последнее имеет χ^2-распределение $C(n-1)$. Для этого рассмотрим характеристическую функцию этих двух величин, а именно

$$\varphi(t_1, t_2) = \mathscr{E} \exp\left(it_1(\bar{x}-\mu)\frac{\sqrt{n}}{\sigma} + it_2\frac{(n-1)\,s^2}{\sigma^2}\right) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \times$$
$$\times \int\limits_{R_n} \exp\left\{-\frac{1}{2}\left[Q_0(x_1, \ldots, x_n) - 2it_1(\bar{x}-\mu)\frac{\sqrt{n}}{\sigma}\right]\right\} dx_1, \ \ldots, \ dx_n, \tag{8.4.1}$$

где

$$Q_0(x_1, \ldots, x_n) = \frac{1}{\sigma^2}\sum_{\xi=1}^{n}(x_\xi-\mu)^2 - \frac{2it_2}{\sigma^2}\sum_{\xi=1}^{n}(x_\xi-\bar{x})^2. \tag{8.4.2}$$

Введя

$$\begin{aligned} \tau^{\xi\xi} &= \frac{1}{\sigma^2}\left[1 - 2it_2\left(1-\frac{1}{n}\right)\right], & \xi &= 1, \ldots, n, \\ \tau^{\xi\eta} &= \tau^{\eta\xi} = \frac{2it_2}{n\sigma^2}, & \xi &\neq \eta = 1, \ldots, n, \end{aligned} \tag{8.4.3}$$

мы получаем

$$Q_0(x_1, \ldots, x_n) = \sum_{\xi,\eta}^{n}\tau^{\xi\eta}(x_\xi-\mu)(x_\eta-\mu). \tag{8.4.4}$$

Используя (7.4.15), замечаем, что

$$Q_0(x_1, \ldots, x_n) - \frac{2it_1}{\sigma}(\bar{x}-\mu)\sqrt{n} =$$
$$= Q_0^1(x_1, \ldots, x_n) + \left(\sum_{\xi,\ \eta=1}^{n}\tau_{\xi\eta}\right)\frac{t_1^2}{n\sigma^2}, \tag{8.4.5}$$

где

$$Q_0^1(x_1, \ldots, x_n) = Q_0(x_1 - ig_1, \ldots, x_n - ig_n) \tag{8.4.6}$$

и

$$\|\tau_{\xi\eta}\| = \|\tau^{\xi\eta}\|^{-1}; \qquad g_\xi = \frac{t_1}{\sqrt{n}\,\sigma}\sum_{\eta=1}^{n}\tau_{\xi\eta}, \qquad \xi = 1, \ldots, n.$$

Необходимо отметить при этом, что матрица $\|\tau^{\xi\eta}\|$ имеет вид

$$\left\|\begin{matrix} a & b & \ldots & b \\ b & a & \ldots & b \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ b & b & \ldots & a \end{matrix}\right\|, \tag{8.4.7}$$

так что ее детерминант равен $(a-b)^{n-1}[a+(n-1)b]$ и ее обратная матрица $\|\tau^{\xi\eta}\|^{-1}=\|\tau_{\xi\eta}\|$ имеет вид

$$\left\|\begin{matrix} A & B & \ldots & B \\ B & A & \ldots & B \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ B & B & \ldots & A \end{matrix}\right\|, \tag{8.4.8}$$

где

$$A=\frac{a+(n-2)b}{(a-b)[a+(n-1)b]}, \qquad B=\frac{-b}{(a-b)[a+(n-1)b]}. \tag{8.4.9}$$

Поэтому

$$\begin{aligned} \tau_{\xi\xi}&=\frac{\sigma^2\left(1-2i\dfrac{t_2}{n}\right)}{(1-2it_2)}, \qquad \xi=1,\ldots,n; \\ \tau_{\xi\eta}&=\frac{-\sigma^2\left(2i\dfrac{t_2}{n}\right)}{(1-2it_2)}, \qquad \xi\neq\eta=1,\ldots,n \end{aligned} \tag{8.4.10}$$

и

$$\sum_{\xi,\ \eta=1}^{n}\tau_{\xi\eta}=n\sigma^2. \tag{8.4.11}$$

Подставляя (8.4.11) в (8.4.5) и затем выражение, полученное в правой части (8.4.5), в интеграл в (8.4.1), получаем

$$\varphi(t_1,\ t_2)=e^{-\frac{1}{2}t_1^2}\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\int\limits_{R_n}e^{-\frac{1}{2}Q_0^1(x_1,\ \ldots,\ x_n)}\,dx_1\ldots dx_n. \tag{8.4.12}$$

Следуя тому же методу, с помощью которого было показано, что в (7.4.16) $H=1$, можно увидеть, что интеграл в (8.4.12) равен $\sqrt{(2\pi)^n}\times$ $\times|\tau^{\xi\eta}|^{-\frac{1}{2}}$. Но

$$|\tau^{\xi\eta}|=\frac{1}{\sigma^{2n}}\,(1-2it_2)^{n-1}.$$

Поэтому

$$\varphi(t_1,\ t_2)=\varphi_1(t_1)\ \varphi_2(t_2), \tag{8.4.13}$$

где

$$\varphi_1(t_1)=e^{-\frac{1}{2}t_1^2} \tag{8.4.14}$$

и

$$\varphi_2(t_2)=(1-2it_2)^{-\frac{1}{2}(n-1)}. \tag{8.4.15}$$

Так как $\varphi(t_1,\ t_2)$ разлагается на множители, как указано в (8.4.13), то из **5.3.1** следует, что $\sqrt{n}(\bar{x}-\mu)/\sigma$ и $(n-1)s^2/\sigma^2$ независимы, а их характеристические функции (8.4.14) и (8.4.15) такие же, как у распределения $N(0,1)$ (см. (7.2.1)) и χ^2-распределения $C(n-1)$ (см. (7.8.2)) соответственно. Так как эти распределения однозначно определяются своими характеристическими функциями, то тем самым доказательство **8.4.2** закончено.

Следующий важный результат следует из **8.4.2** и § **7.8 (b)**:

8.4.3. *Если* $(x_1, \ldots, x_n)$ — *выборка из* $N(\mu, \sigma^2)$, то

$$t = \sqrt{n}(\bar{x} - \mu)/s \tag{8.4.16}$$

имеет распределение Стьюдента $S(n-1)$.

Действительно, если в (7.8.3) мы положим $u = \sqrt{n}(x - \mu)/\sigma$ и $v = (n-1)s^2/\sigma^2$, то будем иметь $t = \frac{u}{\sqrt{v/(n-1)}} = \sqrt{n}(\bar{x} - \mu)/s$. Но если выборка взята из $N(\mu, \sigma^2)$, то, как мы знаем из **8.4.2**, u и v независимы и имеют соответственно распределения $N(0,1)$ и χ^2-распределение $C(n-1)$. Итак, применяя **7.8.3**, мы получаем **8.4.3.**

Замечание. Результат, сформулированный в **8.4.3**, был впервые предсказан Госсетом (1908), который писал под псевдонимом «Стьюдент». Этот результат, так же как сформулированный в **8.4.2**, был позднее доказан Фишером (1926a), который связал имя «Стьюдент» с отношением (8.4.16), так же как и с распределением, имеющим плотность (7.8.4). Тот факт, что v имеет χ^2-распределение с $(n-1)$ степенями свободы, был впервые установлен Хельмертом (1876b).

Необходимо заметить, что утверждение, обратное **8.4.2**, также справедливо. Это утверждение, по существу, гласит, что, если $\bar{x}$ и s^2 суть среднее значение и дисперсия выборки объема n из плотности функции распределения $f(x)$ и если $\bar{x}$ и s^2 статистически независимы, то $f(x)$ — плотность нормального распределения. Этот результат впервые был получен Гири (1936) и сравнительно недавно Кавата и Сакамото (1949) и Лукачем (1942). См. задачу **8.3.3.**

Еще более глубокий результат был получен Крамером (1936), который показал, что если x_1 и x_2 — независимые случайные величины, сумма которых имеет нормальное распределение, то x_1 и x_2 каждая имеет нормальное распределение. Этот результат по индукции обобщается на случай нескольких (более чем 2) случайных величин.

Как мы увидим в следующих главах книги, важной особенностью отношения t, определенного формулой (8.4.16), для статистических выводов является независимость от σ^2 как его самого, так и его распределения (в предположении, что выборка взята из распределения $N(\mu, \sigma^2)$).

Вспомним теперь, что плотность функции распределения выборки из $N(\mu, \sigma^2)$ содержит под знаком экспоненты сумму квадратов $Q = \frac{1}{\sigma^2}\sum_{\xi=1}^{n}(x_\xi - \mu)^2$, и, как мы знаем из **8.4.1**, Q имеет χ^2-распределение $C(n)$. Заметим, что Q может быть разложено следующим образом:

$$Q = Q_2 + Q_2, \tag{8.4.17}$$

где
$$Q_1 = \frac{1}{\sigma^2}\sum_{\xi=1}^{n}(x_\xi - \bar{x})^2, \quad Q_2 = \frac{n}{\sigma^2}(\bar{x} - \mu)^2. \tag{8.4.18}$$

Как мы видели, $\sqrt{n}(\bar{x} - \mu)/\sigma^2$ имеет распределение $N(0,1)$ и, как следует из **7.8.2a**, $[\sqrt{n}(\bar{x} - \mu)/\sigma]^2$, т. е. Q_2 имеет χ^2-распределение $C(1)$. Так как Q_1 и $\sqrt{n}(\bar{x} - \mu)/\sigma$ независимы, то Q_1 и Q_2 также независимы. Таким образом, случайная величина Q, имеющая χ^2-распределение $C(n)$, разложена на сумму двух компонент Q_1 и Q_2, которые статистически независимы и имеют χ^2-распределения $C(n-1)$ и $C(1)$ соответственно. Это приводит к вопросу о том, при каких условиях сумма квадратов под знаком экспоненты ф. п. в. выборки из нормального распределения может быть разложена на сумму нескольких независимых компонент, имеющих χ^2-распределения. Ответ на этот вопрос содержится в следующей теореме, принадлежащей Кочрэну (1934). Достаточно рассмотреть случай, когда выборка взята из нормального распределения $N(0,1)$.

8.4.4. *Если $(x_1, \ldots, x_n)$ — выборка из распределения $N(0,1)$ и $\sum_{\xi=1}^{n} x_\xi^2 \equiv \sum_{i=1}^{k} Q_i$, где Q_i — неотрицательная квадратичная форма от $x_1, \ldots, x_n$, матрица которой имеет ранг n_i, то для того, чтобы Q_i были независимы и имели распределения $C(n_i)$, $i = 1, \ldots, k$, необходимо и достаточно, чтобы $\sum_{i=1}^{k} n_i = n$.*

Во-первых, покажем, что это условие *необходимо*. Итак, предположим, что Q_i независимы, имеют χ^2-распределения $C(n_i)$, $i = 1, \ldots, k$, и мы должны показать, что $\sum_{i=1}^{k} n_i = n$. Но, как следует из свойства воспроизводимости χ^2- распределения, сумма $Q_1 + \ldots + Q_k$ имеет χ^2-распределение $C(n_1 + \ldots + n_k)$. Но $Q_1 + \ldots + Q_k \equiv \sum_{\xi=1}^{n} x_\xi^2$, и так как выборка взята из нормального распределения $N(0,1)$, то на основании **8.4.1** $\sum_{\xi=1}^{n} x_\xi^2$ имеет χ^2-распределение $C(n)$, которое должно поэтому совпадать с $C(n_1 + \ldots + n_k)$. Следовательно, $\sum_{i=1}^{k} n_i = n$.

Для установления *достаточности* условия, сформулированного в **8.4.4**, предположим, что $n_1 + \ldots + n_k = n$, и тогда мы должны показать, что существует невырожденное линейное преобразование
$$x_\xi = \sum_{\eta=1}^{n} b_{\xi\eta}\, y_\eta, \qquad \xi = 1, \ldots, n, \tag{8.4.19}$$

которое преобразует $Q_1, \ldots, Q_k$ к виду

$$\begin{aligned} Q_1 &= y_1^2 + \ldots + y_{n_1}^2, \\ Q_2 &= y_{n_1+1}^2 + \ldots + y_{n_1+n_2}^2, \\ &\cdots\cdots\cdots\cdots \\ Q_k &= y_{n_1+\ldots+n_{k-1}+1}^2 + \ldots + y_{n_1+\ldots+n_k}^2, \end{aligned} \tag{8.4.20}$$

где $n_1 + \ldots + n_k = n$ и $y_1, \ldots, y_n$ — n независимых случайных величин, имеющих распределение $N(0,1)$. Если это действительно можно сделать, то такое преобразование переводит э. в. $(x_1, \ldots, x_n)$ в э. в. выборки $(y_1, \ldots, y_n)$ из распределения $N(0,1)$. Таким образом, $Q_1, \ldots, Q_k$, будучи суммами квадратов соответственно $n_1, \ldots, n_k$ различных независимых одинаково распределенных нормальных случайных величин, независимы и согласно **8.4.1** имеют χ^2-распределения $C(n_1), \ldots, C(n_k)$ соответственно.

Так как Q_1 — неотрицательная квадратичная форма с матрицей ранга n_1, то существует n_1 линейно независимых комбинаций $y_1, \ldots, y_{n_1}$ переменных $x_1, \ldots, x_n$, именно

$$y_{\xi_1} = \sum_{\eta=1}^{n} b^{\xi_1\eta} x_\eta, \qquad \xi_1 = 1, \ldots, n_1, \tag{8.4.21}$$

таких, что

$$Q_1 = y_1^2 + \ldots + y_{n_1}^2. \tag{8.4.22}$$

(См., например, Бохер (1929) или Биркгоф и Маклейн (1953).)

Аналогично для $Q_2, \ldots, Q_k$ имеем соответственно наборы $n_2, \ldots, n_k$ линейно независимых линейных комбинаций переменных $x_1, \ldots, x_n$:

$$\begin{aligned} y_{\xi_2} &= \sum_{\eta=1}^{n} b^{\xi_2\eta} x_\eta, \qquad \xi_2 = n_1 + 1, \ldots, n_1 + n_2, \\ &\vdots \\ y_{\xi_k} &= \sum_{\eta=1}^{n} b^{\xi_k\eta} x_\eta, \qquad \xi_k = n_1 + \ldots + n_{k-1} + 1, \ldots, n_1 + \\ &\qquad\qquad\qquad\qquad \ldots + n_{k-1} + n_k. \end{aligned} \tag{8.4.23}$$

Мы показали, что не может существовать линейной зависимости среди y, связанных с данным Q. Мы должны теперь показать, что не может быть линейной зависимости и среди y, связанных с двумя или более разными Q. Если бы это было не так, то можно было бы выразить линейные формы (8.4.22), связанные с одной из матриц $Q_1, \ldots, Q_k$, например с Q^*, как однородную линейную комбинацию форм, связанных с одной или более из матриц $Q_1, \ldots, Q_k$, отличных от Q^*. Это означало бы, что $Q_1 + \ldots + Q_k$ можно представить как неотрицательную квадратичную форму от (самое большее) $(n-1)$ форм (8.4.22). Так как каждая форма представляет собой однородную линейную комбинацию $x_1, \ldots, x_n$, то $Q_1 + \ldots + Q_k$ была бы тогда неотрицательной квадратичной формой относительно x с матрицей ранга не более $n-1$. Но мы знаем, что матрица $Q_1+\ldots+Q_k$ есть матрица квадратичной формы $\sum\limits_1^n x_\xi^2$ ранга n. Полученное противоречие и доказывает, что не может быть линейной зависимости между формами, связанными с различными матрицами из $Q_1, \ldots, Q_k$. Следовательно, матрица

$$\| b^{\xi\eta} \|, \qquad \xi,\eta=1,\ldots,n,$$

невырожденная. Поэтому преобразование

$$y_\xi=\sum_{\eta=1}^n b^{\xi\eta}x_\eta, \qquad \xi=1, \ldots, n,$$

также невырожденное и имеет единственное обратное

$$x_\eta=\sum_{\xi=1}^n b_{\xi\eta}y_\xi, \qquad \eta=1,\ldots, n,$$

переводящее $\sum\limits_1^n x_\xi^2$, т. е. $Q_1+\ldots+Q_k$ в

$$(y_1^2+\ldots+y_{n_1}^2)+(y_{n_1+1}^2+\ldots+y_{n_1+n_2}^2)+\ldots+(y_{n_1+\ldots+n_{k-1}+1}^2+\ldots$$
$$\ldots+y_n^2).$$

Это преобразование, как легко видеть, ортогональное и, следовательно, $y_1,\ldots,y_n$ — независимые случайные величины с распределением $N(0,1)$. Поэтому $Q_1,\ldots, Q_k$ имеют распределения, совпадающие с распределениями $(y_1^2+\ldots+y_{n_1}^2), \ldots, (y_{n_1+\ldots+n_{k-1}+1}^2+\ldots+y_n^2)$ соответственно, которые в свою очередь являются независимыми случайными величинами с χ^2-распределениями $C(n_1), \ldots, C(n_k)$. Это завершает доказательство достаточности условия и тем самым доказательство **8.4.4.**

Теорема Кочрэна имеет важные приложения в дисперсионном анализе и в теории нормальной регрессии, как мы увидим в главе 10.

8.5. Выборочный метод для конечной совокупности

(a) Одномерный случай. При определении случайной выборки $(x_1, \ldots, x_n)$ в § 8.1 мы предполагали, что ее элементы являются *независимыми* случайными величинами с одной и той же к. ф. р. $F(x)$. Теория выборочного метода, основанная на этом определении, иногда называется теорией *простого выборочного метода для бесконечной совокупности* π_∞ с к. ф. р. $F(x)$. Эта терминология обусловлена тем, что элементы выборки $x_1,\ldots,x_n$ рассматриваются как последовательность независимых случайных величин с тем же распределением, что и совокупность. Поэтому $x_1,\ldots,x_n$ можно понимать как результаты следующего один за другим «извлечения» из совокупности, причем распределение совокупности остается *неизменным* при этих «извлечениях». Говоря о выборочном методе для бесконечной совокупности, мы, как правило, будем иметь в виду выборку объема n из этой совокупности.

Существует также теория *простого выборочного метода для конечной совокупности* π_N, состоящей из конечного числа N объектов $o_1, \dots, o_N$. Если мы рассматриваем π_N как основное выборочное пространство с выборочными точками $o_1, \dots, o_N$ и полагаем, что $x(o)$ — случайная величина, определенная в каждой выборочной точке, т. е. $x(o_t) = x_{ot}$, $t = 1, \dots, N$, то $x(o)$ отображает $o_1, \dots, o_N$ на точки $x_{o1}, \dots, x_{oN}$ в R_1. Заметим, что $x_{o1}, \dots, x_{oN}$ могут рассматриваться просто как значения функции x в N точках пространства π_N (x-значения N объектов в π_N). Удобно, и это не ограничивает общности, предположить, что объекты в π_N занумерованы так, что $x_{o1} \leqslant \dots \leqslant x_{oN}$. Если мы припишем точкам $o_1, \dots, o_N$ равные вероятности $\frac{1}{N}$ и обозначим случайную величину $x(o)$ через x, а ее ф. в. через $p(x)$, то точками сосредоточения массы (т. е. выборочное пространство x) будут $x_{o1}, \dots$ $\dots, x_{oN}$ и в каждой из этих точек

$$p(x) = \frac{1}{N}. \tag{8.5.1}$$

В случае, если $x(o)$ имеет одно и то же значение для двух или более элементов из π_N, то $p(x)$ для этого $x(o)$ есть произведение $\frac{1}{N}$ на соответствующее число элементов. Конкретнее, если $x(o) = x'$ для r элементов из π_N, то $p(x') = \frac{r}{N}$. Мы можем думать об x как о случайной величине, обозначающей x-значение одного объекта, «выбранного случайно» из π_N. Удобно называть $p(x)$, определенную согласно (8.5.1), функцией вероятности (ф. в.) *конечной популяции*.

Теперь предположим, что $(o_{\gamma_1}, \dots, o_{\gamma_n})$ есть некоторая перестановка n объектов из N в π_N. Существует $\frac{N!}{(N-n!)}$ таких *n-перестановок*. Рассмотрим теперь эти n-перестановки как выборочные точки в основном выборочном пространстве R. Если $(o_{\gamma_1}, \dots, o_{\gamma_n})$ — выборочная точка e в R, то пусть $x_1(e), \dots, x_n(e)$ — случайные величины такие, что $x_1(e) = x_{o\gamma_1}, \dots, x_n(e) = x_{o\gamma_n}$. Обозначив $(x_1(e), \dots, x_n(e))$ через $(x_1, \dots, x_n)$, будем говорить об этой n-мерной случайной величине как *о выборке объема n из конечной совокупности* π_N. Мы можем тогда думать об x_1 как о случайной величине, означающей x-значение первого объекта, «выбранного случайно» из π_N, о x_2 — соответственно для второго объекта, ..., о x_n для n-го объекта, *причем объекты выбираются из π_N без возвращения.*

Очевидно, что $(x_1, \ldots, x_n)$ — дискретная n-мерная случайная величина. Если мы припишем всем выборочным точкам e в R одинаковые вероятности, а именно

$$\frac{(N-n)!}{N!}, \tag{8.5.2}$$

то $p(x_1, \ldots, x_n)$, ф. в. $(x_1, \ldots, x_n)$, равна $\frac{(N-n)!}{N!}$ в каждой точке сосредоточения массы.

На самом деле удобно, не ограничивая при этом общности, рассмотреть x-значения N объектов разными и объекты обозначить тогда как $x_{o1} < \ldots < x_{oN}$. Тогда точки сосредоточения массы случайной величины x_ξ есть $x_{o1}, \ldots, x_{oN}$, $\xi = 1, \ldots, n$. Пусть $R_1^{(\xi)}$ — вещественная ось, соответствующая случайной величине x_ξ, и пусть $E_1^{(\xi)}$ обозначает последовательность точек $x_{o1}, \ldots, x_{oN}$ в $R_1^{(\xi)}$. Пусть E_n — прямое произведение $E_1^{(1)} \times \ldots \times E_1^{(n)}$ и R_n — прямое произведение $R_1^{(1)} \times \ldots \times R_1^{(n)}$. Тогда точки сосредоточения массы $(x_1, \ldots, x_n)$ образуют множество E'_n, состоящее из тех точек E_n, у которых нет двух одинаковых координат. В E'_n точно $\frac{N!}{(N-n)!}$ точек и имеется взаимно однозначное соответствие между точками в E'_n и e-точками в R. Другими словами, n-мерная случайная величина $(x_1(e), \ldots, x_n(e))$ отображает e-точки R в точки множества $E'_n \subset R_n$. Таким образом, ф. в. $(x_1, \ldots, x_n)$ определены для точек в своем выборочном пространстве R_n выражением

$$p(x_1, \ldots, x_n) = \begin{cases} \dfrac{(N-n)!}{N!}, & x_1 \neq \ldots \neq x_n, \\ 0 \text{ в остальных случаях.} \end{cases} \tag{8.5.3}$$

Точка сосредоточения массы $(x_1, \ldots, x_n)$, соответствующая объектам $o_{\gamma_1}, \ldots, o_{\gamma_n}$, есть $(x_{o\gamma_1}, \ldots, x_{o\gamma_n})$.

Заметим, что маргинальная функция вероятности $(x_1, \ldots, x_{n-1})$ в (8.5.3) есть сумма $p(x_1, \ldots, x_n)$ по x_n, что дает

$$p_{1 \ldots (n-1)}(x_1, \ldots, x_{n-1}) = (N-n+1)\frac{(N-n)!}{N!} = \frac{(N-n+1)!}{N!}$$

в каждой точке $E_1^{(1)} \times \ldots \times E_1^{(n-1)}$, у которой $(n-1)$ координат различны, и $p_{1 \ldots (n-1)}(x_1, \ldots, x_{n-1})$, равное нулю во всех остальных точках $E_1^{(1)} \times \ldots \times E_1^{(n-1)}$. В более общей форме — маргинальная функция вероятности $(x_{\xi_1}, \ldots, x_{\xi_r})$ есть

$$p_{\xi_1 \ldots \xi_r}(x_{\xi_1}, \ldots, x_{\xi_r}) = \frac{(N-r)!}{N!} \tag{8.5.4}$$

в тех точках $E_1^{(\xi_1)} \times \ldots \times E_1^{(\xi_r)}$, у которых все координаты разные, и нуль во всех остальных точках множества $E_1^{(\xi_1)} \times \ldots \times E_1^{(\xi_r)}$. Функция вероятности компоненты x_ξ вектора $(x_1, \ldots, x_n)$ дается выражением (8.5.1), так что

$$p_\xi(x_\xi) = \frac{1}{N}, \tag{8.5.5}$$

при этом точки сосредоточения массы x_ξ суть $x_{o1}, \ldots, x_{oN}$. Таким образом, любая компонента $(x_1, \ldots, x_n)$ имеет такое же распределение, что и совокупность.

Среднее значение μ_N *и дисперсию* σ^2_N *распределения совокупности* определим следующим образом:

$$\mu_N = \frac{1}{N}\sum_{t=1}^{N} x_{ot}, \qquad \sigma^2_N = \frac{1}{N-1}\sum_{t=1}^{N}(x_{ot} - \mu_N)^2. \tag{8.5.6}$$

Мы будем обычно опускать значок N и вместо μ_N и σ^2_N писать соответственно μ и σ^2. Очевидно, что для $\xi = 1, \ldots, n$

$$\mu(x_\xi) = \mu, \qquad \sigma^2(x_\xi) = \frac{N-1}{N}\sigma^2. \tag{8.5.7}$$

Полезно найти маргинальную функцию вероятности (8.5.4) при $r = 2$, т. е. для каких-либо двух компонент (x_ξ, x_η) из $(x_1, \ldots, x_N)$. Находим

$$p_{\xi\eta}(x_\xi, x_\eta) = \frac{1}{N(N-1)}, \tag{8.5.8}$$

а точки сосредоточения массы суть $(x_{ot}, x_{ot'})$, $t \neq t' = 1, \ldots, N$. Ковариация между x_ξ и x_η дается выражением

$$\operatorname{cov}(x_\xi, x_\eta) = \sum_{x_\xi, x_\eta}(x_\xi - \mu)(x_\eta - \mu)p_{\xi\eta}(x_\xi, x_\eta) =$$

$$= \sum_{t \neq t' = 1}^{N} \frac{(x_{ot} - \mu)(x_{ot'} - \mu)}{N(N-1)} =$$

$$= -\sum_{t=1}^{N}\frac{(x_{ot} - \mu)^2}{N(N-1)} + \frac{1}{N(N-1)}\left[\sum_{t=1}^{N}(x_{ot} - \mu)\right]^2. \tag{8.5.9}$$

Используя же (8.5.6), находим, что

$$\operatorname{cov}(x_\xi, x_\eta) = -\frac{\sigma^2}{N}. \tag{8.5.10}$$

Заметим, что коэффициент корреляции между x_ξ и x_η имеет величину

$$\rho(x_\xi, x_\eta) = -\frac{1}{N-1}. \tag{8.5.11}$$

Напомним, что на основании (8.2.4) $\operatorname{cov}(x_\xi, x_\eta) = 0$ в случае выборки из бесконечной совокупности.

В итоге получаем следующий результат:

8.5.1. *Если* $(x_1, \ldots, x_n)$ — *выборка объема* n *из конечной совокупности* π_N, *то*

$$\mu(x_\xi) = \mu, \quad \sigma^2(x_\xi) = \frac{N-1}{N}\sigma^2, \qquad \xi = 1, \ldots, n;$$

$$\operatorname{cov}(x_\xi, x_\eta) = -\frac{\sigma^2}{N}, \quad \rho(x_\xi, x_\eta) = -\frac{1}{N-1}, \qquad \xi \neq \eta = 1, \ldots, n,$$

где μ *и* σ^2 *определены согласно* (8.5.6).

Построение теории выборочного метода для некоторой функции $g(x_1,\ldots,x_n)$ от выборки $(x_1,\ldots,x_n)$ из конечной совокупности π_N сводится к нахождению функции распределения для $g(x_1,\ldots,x_n)$, где $(x_1,\ldots,x_n)$ — случайная величина, имеющая ф. в. (8.5.3). Аналогичное замечание справедливо, конечно, и для теории выборочного метода для двух или более функций выборки.

(b) Среднее значение выборочного среднего значения и дисперсии. Для иллюстративных целей сначала рассмотрим две специфические функции от $(x_1,\ldots,x_n)$, а именно, $\bar{x}$ и s^2. Так как $\bar{x}=\frac{1}{n}\sum_{\xi=1}^{n} x_\xi$, то получаем

$$\mu(\bar{x})=\mathscr{E}\left(\frac{1}{n}\sum_{\xi} x_\xi\right)=\frac{1}{n}\sum_{\xi}\mathscr{E}(x_\xi). \tag{8.5.12}$$

Но $\mathscr{E}(x_\xi)=\mu$ при $\xi=1,\ldots,n$. Следовательно,

$$\mu(\bar{x})=\mu, \tag{8.5.13}$$

т. е. $\bar{x}$ — несмещенная оценка для μ.

Для $\sigma^2(\bar{x})$ получаем

$$\sigma^2(\bar{x})=\mathscr{E}(\bar{x}-\mu)^2=\mathscr{E}\left[\frac{1}{n}\sum_{\xi}(x_\xi-\mu)\right]^2=$$

$$=\frac{1}{n^2}\sum_{\xi}\mathscr{E}(x_\xi-\mu)^2+\frac{1}{n^2}\sum_{\xi\neq\eta}\mathscr{E}(x_\xi-\mu)(x_\eta-\mu).$$

Но $\mathscr{E}(x_\xi-\mu)^2=\frac{N-1}{N}\sigma^2$ при $\xi=1,\ldots, n$, а для $\xi\neq\eta$

$$\mathscr{E}(x_\xi-\mu)(x_\eta-\mu)=\operatorname{cov}(x_\xi, x_\eta)=-\frac{\sigma^2}{N}.$$

Следовательно,

$$\sigma^2(\bar{x})=\left(\frac{1}{n}-\frac{1}{N}\right)\sigma^2. \tag{8.5.14}$$

Рассмотрим теперь среднее значение s^2. Имеем

$$s^2=\frac{1}{n-1}\sum_{\xi}(x_\xi-\bar{x})^2=\frac{1}{n}\sum_{\xi}(x_\xi-\mu)^2-\frac{1}{n(n-1)}\sum_{\xi\neq\eta}(x_\xi-\mu)(x_\eta-\mu).$$

Беря от этого выражения математическое ожидание и используя явно **8.5.1**, находим, что

$$\mathscr{E}(s^2)=\mu(s^2)=\sigma^2. \tag{8.5.15}$$

Следовательно, s^2 — несмещенная оценка для σ^2.

В итоге получаем следующий результат.

8.5.2. *Пусть* $(x_1, \ldots, x_n)$ — *выборка из конечной совокупности* π_N, *имеющей среднее значение* μ *и дисперсию* σ^2. *Тогда*

$$\mu(\bar{x}) = \mu, \qquad \sigma^2(\bar{x}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sigma^2, \qquad \mu(s^2) = \sigma^2.$$

Необходимо заметить, что среднее значение и дисперсия выборочной суммы определены выражениями

$$\mu(z) = n\mu, \qquad \sigma^2(z) = n^2\left(\frac{1}{n} - \frac{1}{N}\right)\sigma^2. \tag{8.5.16}$$

Наконец, необходимо заметить, что если μ_N и $\sigma^2{}_N$, определенные согласно (8.5.6), имеют конечные пределы при $N \to \infty$ (обозначим их через μ и σ^2), то можно показать, что пределы $\mu(\bar{x})$ и $\sigma^2(\bar{x})$ в **8.5.2** при $N \to \infty$ совпадают с величинами $\mu(\bar{x})$ и $\sigma^2(\bar{x})$ для выборок из бесконечной совокупности, которые определены согласно **8.2.1.**

(с) Средние значения некоторых симметрических функций от выборки из конечной совокупности π_N. Основные идеи § 8.5(b) могут быть обобщены, чтобы получить средние значения некоторых общих симметрических функций от выборки. Итак, пусть $(x_1, \ldots, x_n)$ — выборка из конечной совокупности π_N и $g(x_{\xi_1}, \ldots, x_{\xi_r})$ — такая функция, что $\xi_1 \neq \xi_2 \neq \ldots \neq \xi_r$ и $r < n$. Тогда с очевидностью из вида функции вероятности $(x_{\xi_1}, \ldots, x_{\xi_r})$, даваемого (8.5.4), следует, что

$$\mathscr{E}[g(x_{\xi_1}, \ldots, x_{\xi_r})] = \frac{(N-r)!}{N!} \sum\nolimits_t' g(x_{ot_1}, \ldots, x_{ot_r}), \tag{8.5.17}$$

где $\sum_t'$ означает суммирование по всем величинам $t_1, \ldots, t_r$ таким, что $t_1 \neq t_2 \neq \ldots \neq t_r$, и каждое t пробегает все значения $1, \ldots, N$. Образуем теперь следующую симметрическую функцию от $(x_1, \ldots, x_n)$:

$$Q(x_1, \ldots, x_n) = \frac{(n-r)!}{n!} \sum\nolimits_\xi' g(x_{\xi_1}, \ldots, x_{\xi_r}), \tag{8.5.18}$$

где $\sum_\xi'$ имеет аналогичное значение в терминах $\xi_1, \ldots, \xi_r$, что и $\sum_t'$ для $t_1, \ldots, t_r$.

Если вычислить математическое ожидание от $Q(x_1, \ldots, x_n)$, то получим

$$\mathscr{E}(Q(x_1, \ldots, x_n)) = \frac{(n-r)!}{n!} \sum\nolimits_\xi' \mathscr{E}[g(x_{\xi_1}, \ldots, x_{\xi_r})]. \tag{8.5.19}$$

Но $\mathscr{E}[g(x_{\xi_1}, \ldots, x_{\xi_r})]$ для всех $\xi_1 \neq \xi_2 \neq \ldots \neq \xi_r$ равняется правой части (8.5.17). Так как имеется $\frac{n!}{(n-r)!}$ таких наборов величин $\xi_1, \ldots$ $\ldots, \xi_r$, то получаем

$$\mathscr{E}(Q(x_1, \ldots, x_n)) = \frac{(N-r)!}{N!} \sum\nolimits_t' g(x_{ot_1}, \ldots, x_{ot_r}) = Q(x_{o1}, \ldots, x_{oN}), \tag{8.5.20}$$

так что $Q(x_1, \ldots, x_n)$ — несмещенная оценка для $Q(x_{o1}, \ldots, x_{oN})$.

Таким образом, получаем следующий результат:

8.5.3. *Если* $(x_1, \ldots, x_n)$ — *выборка из конечной совокупности* π_N *и если* $Q(x_1, \ldots, x_n)$ — *некоторая симметрическая функция* $(x_1, \ldots$ $\ldots, x_n)$ *типа* (8.5.18), *тогда* $Q(x_1, \ldots, x_n)$ — *несмещенная оценка для* $Q(x_{o1}, \ldots, x_{oN})$.

Более того, как следует из **3.2.1** и **3.2.2**, справедлива следующая теорема:

8.5.4. *Если* $Q_1(x_1, \ldots, x_n), \ldots, Q_s(x_1, \ldots, x_n)$ — *симметрические функции выборки* $(x_1, \ldots, x_n)$ *типа* (8.5.18) *из конечной совокупности* π_N, *то для любого набора постоянных* $d_1, \ldots, d_s$ $d_1 Q_1(x_1, \ldots, x_n) + \ldots + d_s Q_s(x_1, \ldots, x_n)$ — *несмещенная оценка для* $d_1 Q_1(x_{o1}, \ldots$ $\ldots, x_{oN}) + \ldots + d_s Q_s(x_{o1}, \ldots, x_{oN})$.

Если в (8.5.18) возьмем $g(x_{\xi_1}, \ldots, x_{\xi_r}) \equiv x_{\xi_1}^{c_1} \ldots x_{\xi_r}^{c_r}$, где $c_1, \ldots, c_r$ — положительные целые числа (или нули), то получим класс симметрических функций $\{Q(x_1, \ldots, x_n)\}$ такой, что выборочное среднее значение, выборочная дисперсия, k-статистики и другие полиномы от выборочных моментов могут быть выражены как линейные комбинации от функций Q этого класса. Этот факт вместе с теоремами **8.5.3** и **8.5.4** делает, как показал Тьюки (1950, 1956а), теорию выборочного метода для среднего значения, дисперсии и более старших моментов, особенно k-статистик, весьма простой.

Пример. Для иллюстрации определим величину $\mathcal{E}(\bar{x}^3)$. Имеем

$$\bar{x}^3 = \frac{1}{n^3} \sum_{\xi_1, \xi_2, \xi_3 = 1}^{n} x_{\xi_1} x_{\xi_2} x_{\xi_3} =$$

$$= \frac{1}{n^3} \left[\sum_{\xi_1 = 1}^{n} x_{\xi_1}^3 + 3 \sum_{\xi_1 \neq \xi_2} x_{\xi_1}^2 x_{\xi_2} + \sum_{\xi_1 \neq \xi_2 \neq \xi_3} x_{\xi_1} x_{\xi_2} x_{\xi_3} \right] =$$

$$= \frac{1}{n^2} Q_1(x_1, \ldots, x_n) + \frac{3(n-1)}{n^2} Q_2(x_1, \ldots, x_n) + \frac{(n-1)(n-2)}{n^2} Q_3(x_1, \ldots, x_n),$$

где

$$Q_1 = \frac{1}{n} \sum_{\xi_1 = 1}^{n} x_{\xi_1}^3,$$

$$Q_2 = \frac{1}{n(n-1)} \sum_{\xi_1 \neq \xi_2} x_{\xi_1}^2 x_{\xi_2},$$

$$Q_3 = \frac{1}{n(n-1)(n-2)} \sum_{\xi_1 \neq \xi_2 \neq \xi_3} x_{\xi_1} x_{\xi_2} x_{\xi_3},$$

т. е. функции типа (8.5.18).

Применяя **8.5.4**, мы получаем тогда

$$\mathcal{E}(\bar{x}^3) = \frac{1}{n^2} Q_1(x_{o1}, \ldots, x_{oN}) + \frac{3(n-1)}{n^2} Q_2(x_{o1}, \ldots, x_{oN}) + \\ + \frac{(n-1)(n-2)}{n^2} Q_3(x_{o1}, \ldots, x_{oN}).$$

Если положим

$$\mu'_r = \frac{1}{N} \sum_{t=1}^{N} x^r_{ot},$$

то нетрудно заметить, что

$$Q_1(x_{o1}, \dots, x_{oN}) = \mu'_3,$$

$$Q_2(x_{o1}, \dots, x_{oN}) = \frac{1}{N-1} \, (N\mu'_2\mu'_1 - \mu'_3),$$

$$Q_3 = (x_{o1}, \dots, x_{oN}) = \frac{1}{(N-1)(N-2)} \, [N^2 \mu'^3_1 + 2\mu'_3 - 3N\mu'_2\mu'_1],$$

и, следовательно, $\mathscr{E}(\bar{x}^3)$ может быть выражено полиномом от трех первых моментов совокупности.

(d) k-мерный случай. Теперь рассмотрим тот случай, когда случайная величина $x(o)$, введенная в § 8.5(a), есть вектор с k компонентами $x_1(o), \dots, x_k(o)$. Тогда каждый обьект в π_N «измеряется» k числами. Таким образом, если $x_i(o_t) = x_{oit}$, то мы можем говорить о x_{oit} как об x_i-величине o_t, и если векторы для N объектов в π_N все различны, то наша векторная случайная величина $(x_1(o), \dots, x_k(o))$ отображает объекты из π_N в N точек из R_k, причем координаты образов точек o_t есть $(x_{o1t}, \dots, x_{okt})$, $t = 1, \dots, N$. Если вероятности, соответствующие всем этим N точкам, выбрать равными, а именно $\frac{1}{N}$, то получим k-мерную случайную величину $(x_1, \dots, x_k)$ с ф. в.

$$p(x_1, \dots, x_k) = \frac{1}{N} \tag{8.5.21}$$

и точками сосредоточения массы $(x_{o1t}, \dots, x_{okt})$, $t = 1, \dots, N$. Мы можем говорить тогда о $p(x_1, \dots, x_k)$ как о ф. в. конечной совокупности π_N.

Рассмотрим теперь основное выборочное пространство R $N!/(N-n)!$ перестановок, введенное в § 8.5(a), так что $(o_{\gamma_1}, \dots, o_{\gamma_n})$ — типичная выборочная точка e. Пусть $(x_{i1}(e), \dots, x_{in}(e); i = 1, \dots, k)$ — nk случайных величин таких, что $x_{i1}(e) = x_{oi\gamma_1}, \dots, x_{in}(e) = x_{oi\gamma_n}$. Если обозначить $(x_{i1}(e), \dots, x_{in}(e); i = 1, \dots, k)$ через $(x_{i\xi}, i = 1, \dots, k; \xi = 1, \dots, n)$, то мы будем говорить об этой nk-мерной случайной величине как о *выборке объема n из k-мерной конечной совокупности* π_N, с функцией вероятности (8.5.3). Для этих nk-мерных случайных величин в R_{nk} имеется $N!/(N-n)!$ точек сосредоточения масс, в каждой из которых выбранная нами вероятность равна $(N-n)!/N!$. Точке $(o_{\gamma_1}, \dots, o_{\gamma_n})$ соответствует точка $(x_{oi\gamma_1}, \dots, x_{oi\gamma_n}, i = 1, \dots, k)$.

Сказанного выше достаточно, чтобы построить теорию выборочного метода для различных функций от нашей выборки. А именно, для таких функций, как векторы выборочных *сумм* и *средних* *значений*

$$z_i = \sum_{\xi=1}^{n} x_{i_\xi}, \qquad \bar{x}_i = \frac{1}{n}\sum_{\xi=1}^{n} x_{i_\xi} \tag{8.5.22}$$

и *выборочная ковариационная матрица* $\|s_{ij}\|$, где

$$s_{ij} = \frac{1}{n-1}\sum_{\xi=1}^{n}(x_{i_\xi} - \bar{x}_i)(x_{j_\xi} - \bar{x}_j), \qquad i, j = 1, \ldots, k. \tag{8.5.23}$$

Определим вектор *средних значений* $(\mu_1, \ldots, \mu_k)$ из π_N как

$$\mu_i = \frac{1}{N}\sum_{t=1}^{N} x_{oit}, \qquad i = 1, \ldots, k, \tag{8.5.24}$$

и *ковариационную матрицу* $\|\sigma_{ij}\|$ из π_N как

$$\sigma_{ij} = \frac{1}{N-1}\sum_{t=1}^{N}(x_{oit} - \mu_i)(x_{ojt} - \mu_j), \qquad i, j = 1, \ldots, k. \tag{8.5.25}$$

Тогда, вычисляя среднее значение, аналогично тому как это было сделано в (8.5.12) — (8.5.15), можно показать, что

$$\begin{aligned} &\mu(\bar{x}_i) = \mu_i, \quad \sigma^2(\bar{x}) = \left(\frac{1}{n} - \frac{1}{N}\right)\sigma_{ii}, && i = 1, \ldots, k; \\ &\operatorname{cov}(\bar{x}_i, \bar{x}_j) = \left(\frac{1}{n} - \frac{1}{N}\right)\sigma_{ij}, && i \neq j = 1, \ldots, k; \\ &\mu(s_{ij}) = \sigma_{ij}, && i, j = 1, \ldots, k. \end{aligned} \tag{8.5.26}$$

8.6. Матричный выборочный метод

(а) Матричные выборки второго порядка. Предположим, что u и v — независимые случайные величины, имеющие к. ф. р. $F_1(u)$ и $F_2(v)$. Пусть $x(u, v)$ — случайная величина, для которой

$$\begin{aligned} \mu &= \mathscr{E}(x(u, v)), \\ \mu_{u\cdot} &= \int_{-\infty}^{+\infty} x(u, v)\, dF_2(v), \\ \mu_{\cdot v} &= \int_{-\infty}^{+\infty} x(u, v)\, dF_1(u). \end{aligned} \tag{8.6.1}$$

Далее, пусть $\varepsilon_{u\cdot}$, $\varepsilon_{\cdot v}$ и ε_{uv} — случайные величины такие, что

$$\begin{aligned} \varepsilon_{u\cdot} &= \mu_{u\cdot} - \mu, \\ \varepsilon_{\cdot v} &= \mu_{\cdot v} - \mu, \\ \varepsilon_{uv} &= x(u, v) - \mu_{u\cdot} - \mu_{\cdot v} + \mu. \end{aligned} \tag{8.6.2}$$

Тогда, очевидно, среднее значение каждой из этих случайных величин равно нулю и ковариация между любыми двумя из них тоже равна нулю.

Таким образом, имеем следующую теорему разложения:

8.6.1. *Если случайная величина $x(u, v)$ есть функция независимых случайных величин u и v, то имеет место следующее разложение:*

$$x(u, v) = \mu + \varepsilon_{u\cdot} + \varepsilon_{\cdot v} + \varepsilon_{uv}, \tag{8.6.3}$$

где $\varepsilon_{u\cdot}$, $\varepsilon_{\cdot v}$, ε_{uv} определены согласно (8.6.2) *и имеют среднее значение, равное нулю, и нулевые ковариации. Более того,*

$$\sigma^2(x(u, v)) = \sigma^2(\varepsilon_{u\cdot}) + \sigma^2(\varepsilon_{\cdot v}) + \sigma^2(\varepsilon_{uv}). \tag{8.6.4}$$

Замечание. Необходимо отметить, что более общая форма этой теоремы (и последующих теорем этого раздела) справедлива, когда u и v рассматриваются как независимые выборочные точки в произвольных выборочных пространствах $R^{(1)}$ и $R^{(2)}$, а не как просто случайные величины. В таких пространствах мы должны, конечно, иметь вероятностные меры и $x(u, v)$ должна быть случайной величиной, измеримой по отношению к прямому произведению этих двух вероятностных мер.

Удобно сократить обозначения для дисперсий следующим образом:

$$\begin{aligned} \sigma^2(x(u, v)) &= \sigma^2, \\ \sigma^2(\varepsilon_{u\cdot}) &= \sigma^2_{\cdot 0}, \\ \sigma^2(\varepsilon_{\cdot v}) &= \sigma^2_{0\cdot}, \\ \sigma^2(\varepsilon_{uv}) &= \sigma^2_{\cdot\cdot}. \end{aligned} \tag{8.6.5}$$

Тогда (8.6.4) переходит в

$$\sigma^2 = \sigma^2_{\cdot 0} + \sigma^2_{0\cdot} = \sigma^2_{\cdot\cdot}. \tag{8.6.4a}$$

Пусть теперь $(u_1, \ldots, u_r)$ и $(v_1, \ldots, v_s)$ — независимые выборки с к.ф.р. $F_1(u)$ и $F_2(v)$ соответственно. Функция распределения этих двух выборок тогда равна

$$\prod_{\xi=1}^{r} F_1(u_\xi) \prod_{\eta=1}^{s} F_2(v_\eta). \tag{8.6.6}$$

Пусть $(x(u_\xi, v_\eta);\ \xi = 1, \ldots, r;\ \eta = 1, \ldots, s)$ — новое множество случайных величин, которое может быть представлено прямоугольной таблицей с r строчками и s столбцами. Эта таблица будет называться *матричной выборкой второго порядка.* Необходимо отметить, что эта выборка зависит только от $r + s$ случайных величин, а именно, $u_1, \ldots, u_r$ и $v_1, \ldots, v_s$.

Замечание. Матричные выборки играют фундаментальную роль в таких областях, как психометрика и планирование эксперимента. В психометрике, например, строчки могут представлять данную совокупность экзаменующихся, которые получили определенный тест, столбцы—данную совокупность вопросов. Таким образом, $x(u_\xi, v_\eta)$ будет представлять отметку лица, для которого $u = u_\xi$, за ответ на вопрос, для которого $v = v_\eta$. Если тест из s вопросов задается группе из r экзаменующихся, то совокупность отметок этих лиц на каждый из s вопросов и составляет матричную выборку $x(u_\xi, v_\eta)$, $\xi = 1, \ldots, r$; $\eta = 1, \ldots, s$.

Лорд (1955) рассмотрел теорию выборочного метода для различных линейных и квадратичных функций от $x(u_\xi, v_\eta)$, которые встречаются в психометрике.

В случае планирования эксперимента строчки могут представлять совокупность операторов, работающих на некоторого типа машинах, а столбцы — совокупность этих машин. Таким образом, если мы выберем случайно r операторов и s машин, то можем считать $x(u_\xi, v_\eta)$ случайной величиной, представляющей выход (продукцию) рабочего дня оператора с номером ξ на машине с номером η. Такой эксперимент требует, конечно, для своего выполнения s рабочих дней. Приложение теории матричного выборочного метода в планировании эксперимента будет обсуждаться в разделах 10.6 и 10.7.

Чтобы сократить наши обозначения, положим

$$\begin{aligned} x_{\xi\eta} &= x(u_\xi, v_\eta), \\ x'_{\xi\eta} &= x_{\xi\eta} - \mu, \\ \bar{x}_{..} &= \frac{1}{rs}\sum_{\xi,\,\eta} x_{\xi\eta}, \\ \bar{x}_{\xi.} &= \frac{1}{s}\sum_{\eta} x_{\xi\eta}, \\ \bar{x}_{.\eta} &= \frac{1}{r}\sum_{\xi} x_{\xi\eta}, \end{aligned} \tag{8.6.7}$$

где $\xi = 1, \ldots, r$; $\eta = 1, \ldots, s$.

Необходимо заметить, что для каждого ξ и η (8.6.3) можно переписать в виде

$$x_{\xi\eta} = \mu + \varepsilon_{\xi.} + \varepsilon_{.\eta} + \varepsilon_{\xi\eta}. \tag{8.6.8}$$

Пусть теперь

$$\begin{aligned} S_T &= \sum_{\xi,\,\eta} (x_{\xi\eta} - \bar{x}_{..})^2, \\ S_{.0} &= \sum_{\xi,\,\eta} (\bar{x}_{\xi.} - \bar{x}_{..})^2, \\ S_{0.} &= \sum_{\xi,\,\eta} (\bar{x}_{.\eta} - \bar{x}_{..})^2, \\ S_{..} &= \sum_{\xi,\,\eta} (\bar{x}_{\xi\eta} - \bar{x}_{\xi.} - \bar{x}_{.\eta} + \bar{x}_{..})^2. \end{aligned} \tag{8.6.9}$$

Легко видеть, что при этом

$$S_T = S_{.0} + S_{0.} + S_{..}\,.$$

Иногда удобно называть $S_{.0}$, $S_{0.}$ и $S_{..}$ *строчной, столбцовой и остаточной компонентами* полной суммы квадратов S_T. Мы будем рассматривать далее среднее значение и дисперсию $\bar{x}_{..}$ и среднее значение S_T, $S_{\cdot 0}$, $S_{0\cdot}$ и $S_{\cdot\cdot}$, при этом все средние значения берутся относительно к. ф. р. (8.6.6). Так как

$$\mathcal{E}(x_{\xi\eta})=\mu, \qquad \xi=1, \ldots, r; \qquad \eta=1, \ldots, s,$$

то, очевидно,

$$\mathcal{E}(\bar{x}_{\cdot\cdot})=\mu.$$

При нахождении средних значений величин S из (8.6.9) мы увидим, что они могут быть получены как линейные функции от следующих выражений четырех типов:

$$\begin{aligned} \mathcal{E}(x'_{\xi\eta}, x'_{\xi'\eta'}) &= \sigma^2_{\cdot\cdot}+\sigma^2_{\cdot 0}+\sigma^2_{0\cdot} & \xi&=\xi'; & \eta&=\eta', \\ &=\sigma^2_{\cdot 0} & \xi&=\xi'; & \eta&\neq\eta', \\ &=\sigma^2_{0\cdot} & \xi&\neq\xi'; & \eta&=\eta', \\ &=0 & \xi&\neq\xi'; & \eta&\neq\eta'. \end{aligned} \tag{8.6.10}$$

Вычисляя средние значения, мы действительно находим

$$\begin{aligned} \mathcal{E}(S_{\cdot 0}) &= s(r-1)\left(\sigma^2_{\cdot 0}+\frac{\sigma^2_{\cdot\cdot}}{s}\right), \\ \mathcal{E}(S_{0\cdot}) &= r(s-1)\left(\sigma^2_{0\cdot}+\frac{\sigma^2_{\cdot\cdot}}{r}\right), \\ \mathcal{E}(S_{\cdot\cdot}) &= (r-1)(s-1)\sigma^2_{\cdot\cdot}, \end{aligned} \tag{8.6.11}$$

в то время как $\mathcal{E}(S_T)$ может быть определено из уравнения

$$\mathcal{E}(S_T)=\mathcal{E}(S_{\cdot 0})+\mathcal{E}(S_{0\cdot})+\mathcal{E}(S_{\cdot\cdot}).$$

Аналогично дисперсия $\bar{x}_{\cdot\cdot}$ есть линейная функция четырех типов средних значений, упомянутых в (8.6.10), что дает

$$\sigma^2(\bar{x}_{\cdot\cdot})=\frac{\sigma^2_{\cdot 0}}{r}+\frac{\sigma^2_{0\cdot}}{s}+\frac{\sigma^2_{\cdot\cdot}}{rs}. \tag{8.6.12}$$

В итоге

8.6.2. *Пусть* $(u_1, \ldots, u_r)$ и $(v_1, \ldots, v_s)$ — *независимые выборки из бесконечных совокупностей. Пусть* $(x_{\xi\eta}; \xi=1, \ldots, r; \eta=1, \ldots, s)$ — *матричная выборка, среднее значение которой определено согласно* (8.6.7), *а строчная, столбцовая и остаточная компоненты полной суммы квадратов* S_T, $S_{\cdot 0}$, $S_{0\cdot}$ *и* $S_{\cdot\cdot}$ *соответственно определены формулой* (8.6.9). *Тогда* $\mathcal{E}(\bar{x}_{\cdot\cdot})=\mu$, *а* $\sigma^2(\bar{x}_{\cdot\cdot})$ *определено согласно* (8.6.12), *и средние значения* $S_{\cdot 0}$, $S_{0\cdot}$, $S_{\cdot\cdot}$ даются выражениями (8.6.11).

(b) Случай конечных совокупностей строчек и столбцов. Предыдущие результаты могут быть перенесены с помощью аналогичного метода на случай, когда $(u_1, \ldots, u_r)$ и $(v_1, \ldots, v_s)$ — независимые случайные выборки из конечных совокупностей π_{N_1} и π_{N_2}. Этот случай был рассмотрен Хуком (1956а) и Тьюки (1950).

В случае конечных совокупностей мы можем, без ограничения общности, считать, что элементы в π_{N_1} имеют различные значения, так что $u_{01}<\ldots<u_{0N_1}$. Аналогично мы можем рассматривать раз-

личные $v_{o1} < \ldots < v_{oN_2}$. Удобно использовать в дальнейшем следующие обозначения:

$$x(u_{oi}, v_{oj}) = x_{oij},$$
$$x'_{oij} = x_{oij} - \mu_f, \tag{8.6.13}$$

и тогда в соответствии с (8.6.1)

$$\mu_f = \frac{1}{N_1 N_2} \sum_{i,\, j} x_{oij},$$
$$\mu_{fi\cdot} = \frac{1}{N_2} \sum_{j} x_{oij}, \tag{8.6.14}$$
$$\mu_{f\cdot j} = \frac{1}{N_1} \sum_{i} x_{oij}.$$

В случае конечных совокупностей строчек и столбцов мы можем существенно упростить формулы для средних значений, определив σ_f^2, $\sigma_{f\cdot 0}^2$, $\sigma_{f0\cdot}^2$, $\sigma_{f\cdot\cdot}^2$ следующим образом:

$$\sigma_f^2 = \frac{1}{N_1 N_2 - 1} \sum_{i,\, j} x'^2_{oij},$$
$$\sigma_{f\cdot 0}^2 = \frac{1}{N_2 (N_1 - 1)} \sum_{i,\, j} (\mu_{fi\cdot} - \mu_f)^2, \tag{8.6.15}$$
$$\sigma_{f0\cdot}^2 = \frac{1}{N_1 (N_2 - 1)} \sum_{i,\, j} (\mu_{f\cdot j} - \mu_f)^2,$$
$$\sigma_{f\cdot\cdot}^2 = \frac{1}{(N_1 - 1)(N_2 - 1)} \sum_{i,\, j} (x_{oij} - \mu_{fi\cdot} - \mu_{f\cdot j} + \mu_f)^2.$$

Заметим, что уравнение, связывающее σ_f^2, $\sigma_{f\cdot 0}^2$, $\sigma_{f0\cdot}^2$ и $\sigma_{f\cdot\cdot}^2$, именно

$$\frac{N_1 N_2 - 1}{N_1 N_2} \sigma_f^2 = \frac{(N_1 - 1)}{N_1} \sigma_{f\cdot 0}^2 + \frac{(N_2 - 1)}{N_2} \sigma_{f0\cdot}^2 + \frac{(N_1 - 1)(N_2 - 1)}{N_1 N_2} \sigma_{f\cdot\cdot}^2, \tag{8.6.16}$$

переходит в пределе при $N_1, N_2 \to \infty$ в (8.6.4а).

Нас интересуют среднее значение и дисперсия $\bar{x}_{\cdot\cdot}$ и средние значения $S_{0\cdot}$, $S_{\cdot 0}$ и $S_{\cdot\cdot}$. Чтобы вычислить эти средние значения, оказывается удобным определить следующие величины:

$$T_{(\cdot\cdot)} = \sum_{i,\, j} (x'_{oij})^2,$$
$$T_{(\cdot 0)} = \sum_{i,\, j \neq j'} x'_{oij} x'_{oij'}, \tag{8.6.17}$$
$$T_{(0\cdot)} = \sum_{i \neq i',\, j} x'_{oij} x'_{oi'j},$$
$$T_{(00)} = \sum_{i \neq i',\, j \neq j'} x'_{oij} x'_{oi'j'},$$

при этом, как можно показать,

$$T_{(\cdot\cdot)}+T_{(\cdot 0)}+T_{(0\cdot)}+T_{(00)}=\Big[\sum_{i,j} x'_{0ij}\Big]^2=0. \tag{8.6.18}$$

Поскольку $\sigma^2_{f\cdot 0}$, $\sigma^2_{f0\cdot}$ и $\sigma^2_{f\cdot\cdot}$ — линейные функции от $T_{(\cdot\cdot)}$, $T_{(\cdot 0)}$, $T_{(0\cdot)}$ и $T_{(00)}$, то, выражая T через σ и используя (8.6.18), получаем

$$\begin{aligned}
T_{(\cdot\cdot)}&=(N_1-1)(N_2-1)\left[\sigma^2_{f\cdot\cdot}+\frac{N_1}{N_1-1}\sigma^2_{f0\cdot}+\frac{N_2}{N_2-1}\sigma^2_{f\cdot 0}\right],\\
T_{(\cdot 0)}&=(N_1-1)(N_2-1)\left[-\sigma^2_{f\cdot\cdot}-\frac{N_1}{N_1-1}\sigma^2_{f0\cdot}+N_2\sigma^2_{f\cdot 0}\right],\\
T_{(0\cdot)}&=(N_1-1)(N_2-1)\left[-\sigma^2_{f\cdot\cdot}+N_1\sigma^2_{f0\cdot}-\frac{N_2}{N_2-1}\sigma^2_{f\cdot 0}\right],\\
T_{(00)}&=(N_1-1)(N_2-1)\left[\sigma^2_{f\cdot\cdot}-N_1\sigma^2_{f0\cdot}-N_2\sigma^2_{f\cdot 0}\right].
\end{aligned} \tag{8.6.19}$$

Как и в случае бесконечных совокупностей строк и столбцов, пусть ($x_{\xi\eta}$; $\xi=1, \ldots, r$; $\eta=1, \ldots, s$) — матричная выборка величины $r+s$ и $T_{\cdot\cdot}$, $T_{\cdot 0}$, $T_{0\cdot}$ и T_{00} выражены через $x'_{\xi\eta}$, определенные в (8.6.7), аналогично тому как $T_{(\cdot\cdot)}$, $T_{(\cdot 0)}$, $T_{(0\cdot)}$ и $T_{(00)}$ выражены через x'_{0ij}. Заметим при этом, что выборочные случайные переменные $T_{\cdot\cdot}$, $T_{\cdot 0}$, $T_{0\cdot}$ и T_{00} не удовлетворяют, конечно, (8.6.18).

Можно легко показать, что

$$\begin{aligned}
\mathscr{E}(T_{\cdot\cdot})&=\frac{rs}{N_1N_2}T_{(\cdot\cdot)},\\
\mathscr{E}(T_{\cdot 0})&=\frac{rs(s-1)}{N_1N_2(N_2-1)}T_{(\cdot 0)},\\
\mathscr{E}(T_{0\cdot})&=\frac{rs(r-1)}{N_1N_2(N_1-1)}T_{(0\cdot)},\\
\mathscr{E}(T_{00})&=\frac{rs(r-1)(s-1)}{N_1N_2(N_1-1)(N_2-1)}T_{(00)}.
\end{aligned} \tag{8.6.20}$$

Эти формулы являются основными в проблеме нахождения средних значений $S_{\cdot 0}$, $S_{0\cdot}$ и $S_{\cdot\cdot}$. Действительно, $S_{\cdot 0}$, $S_{0\cdot}$ и $S_{\cdot\cdot}$ являются линейными функциями от $T_{\cdot\cdot}$, $T_{\cdot 0}$, $T_{0\cdot}$, T_{00}:

$$\begin{aligned}
S_{\cdot 0}&=\frac{r-1}{rs}(T_{\cdot\cdot}+T_{\cdot 0})-\frac{1}{rs}(T_{0\cdot}+T_{00}),\\
S_{0\cdot}&=\frac{s-1}{rs}(T_{\cdot\cdot}+T_{0\cdot})-\frac{1}{rs}(T_{\cdot 0}+T_{00}),\\
S_{\cdot\cdot}&=\frac{(r-1)(s-1)}{rs}T_{\cdot\cdot}-\frac{(r-1)}{rs}T_{\cdot 0}-\frac{(s-1)}{rs}T_{0\cdot}+\frac{1}{rs}T_{00}.
\end{aligned} \tag{8.6.21}$$

Вычисляя математические ожидания от этих выражений, используем (8.6.20) и подставим величины $T_{(\cdot\cdot)}$, $T_{(\cdot 0)}$ и $T_{(00)}$ из (8.6.19). Тогда получаем следующие выражения для искомых средних значений:

$$\begin{aligned}
\mathscr{E}(S_{\cdot 0})&=s(r-1)\left[\sigma^2_{f\cdot 0}+\left(\frac{1}{s}-\frac{1}{N_2}\right)\sigma^2_{f\cdot\cdot}\right],\\
\mathscr{E}(S_{0\cdot})&=r(s-1)\left[\sigma^2_{f0\cdot}+\left(\frac{1}{r}-\frac{1}{N_1}\right)\sigma^2_{f\cdot\cdot}\right],\\
\mathscr{E}(S_{\cdot\cdot})&=(r-1)(s-1)\sigma^2_{f\cdot\cdot}.
\end{aligned} \tag{8.6.22}$$

Значение же $\mathscr{E}(S_{\text{т}})$ может быть найдено из уравнения

$$\mathscr{E}(S_{\text{т}}) = \mathscr{E}(S_{\cdot 0}) + \mathscr{E}(S_{0\cdot}) + \mathscr{E}(S_{\cdot\cdot}).$$

Заметим, что эти уравнения переходят в (8.6.11) при $N_1, N_2 \to \infty$, если $\sigma_{f\cdot\cdot}^2$, $\sigma_{f0\cdot}^2$, $\sigma_{f\cdot0}^2$ имеют своими пределами $\sigma_{\cdot\cdot}^2$, $\sigma_{0\cdot}^2$, $\sigma_{\cdot0}^2$ соответственно.

Чтобы определить дисперсию $\bar{x}_{\cdot\cdot}$, мы должны найти среднее значение $(\bar{x}_{\cdot\cdot} - \mu_f)^2$, которое в свою очередь может быть представлено в следующем виде:

$$(\bar{x}_{\cdot\cdot} - \mu_f)^2 = \frac{1}{rs}[T_{\cdot\cdot} + T_{\cdot 0} + T_{0\cdot} + T_{00}]. \tag{8.6.23}$$

Вычисляя математические ожидания от обеих сторон (8.6.23) и явно используя (8.6.19) и (8.6.20), находим

$$\sigma^2(\bar{x}_{\cdot\cdot}) = \left(\frac{1}{r} - \frac{1}{N_1}\right)\sigma_{f\cdot0}^2 + \left(\frac{1}{s} - \frac{1}{N_2}\right)\sigma_{f0\cdot}^2 + \left(\frac{1}{r} - \frac{1}{N_1}\right)\left(\frac{1}{s} - \frac{1}{N_2}\right)\sigma_{f\cdot\cdot}^2. \tag{8.6.24}$$

Читатель может заметить, что эта формула переходит в (8.6.12) при $N_1, N_2 \to \infty$.

Формулировку итогового утверждения, аналогичного **8.6.2**, для теории матричного выборочного метода в случае конечных совокупностей строк и столбцов мы оставляем читателю.

(с) Матричная выборка третьего порядка. Идеи, изложенные на предыдущих страницах, распространяются без особых трудностей на случай матричных выборок третьего и более высоких порядков. Чтобы отметить возникающие особенности, достаточно кратко рассмотреть матричные выборки третьего порядка. Здесь мы будем иметь дело со случайной переменной $x(u, v, w)$, где u, v, w — в свою очередь независимые случайные величины с к. ф. р. $F_1(u)$, $F_2(v)$, $F_3(w)$ соответственно.

Для $x(u, v, w)$ определим следующие средние значения:

$$\begin{aligned} \mu &= \mathscr{E}(x(u, v, w)), \\ \mu_{u\cdot\cdot} &= \int_{R_2} x(u, v, w)\, dF_2(v)\, dF_3(w), \\ \mu_{uv\cdot} &= \int_{-\infty}^{\infty} x(u, v, w)\, dF_3(w) \end{aligned} \tag{8.6.25}$$

($\mu_{\cdot v\cdot}$, $\mu_{\cdot\cdot w}$, $\mu_{u\cdot w}$, $\mu_{\cdot vw}$ определяются аналогично (8.6.25)) и случайные величины:

$$\begin{aligned} \varepsilon_{u\cdot\cdot} &= \mu_{u\cdot\cdot} - \mu, \\ \varepsilon_{uv\cdot} &= \mu_{uv\cdot} - \mu_{u\cdot\cdot} - \mu_{\cdot v\cdot} + \mu, \\ \varepsilon_{uvw} &= x(u, v, w) - \mu_{uv\cdot} - \mu_{u\cdot w} - \mu_{\cdot vw} + \mu_{u\cdot\cdot} + \mu_{\cdot v\cdot} + \mu_{\cdot\cdot w} - \mu. \end{aligned} \tag{8.6.26}$$

($\varepsilon_{\cdot v\cdot}$, $\varepsilon_{\cdot\cdot w}$, $\varepsilon_{u\cdot w}$, $\varepsilon_{\cdot vw}$ определяются аналогично (8.6.26)).

Тогда получаем следующее обобщение **8.6.1.**

8.6.3. *Если u, v, w — независимые случайные величины и $x(u, v, w)$ — некоторая случайная величина, то*

$$x(u, v, w) = \mu + \varepsilon_{u\cdot\cdot} + \varepsilon_{\cdot v\cdot} + \varepsilon_{\cdot\cdot w} + \varepsilon_{uv\cdot} + \varepsilon_{u\cdot w} + \varepsilon_{\cdot vw} + \varepsilon_{uvw},$$

где все ε — случайные величины с нулевыми средними значениями и *нулевыми ковариациями, определенные согласно* (8.6.26). *Более того,*

$$\sigma^2(x(u, v, w)) = \sigma^2(\varepsilon_{u\cdot\cdot}) + \sigma^2(\varepsilon_{\cdot v\cdot}) + \sigma^2(\varepsilon_{\cdot\cdot w}) + \\ + \sigma^2(\varepsilon_{uv\cdot}) + \sigma^2(\varepsilon_{u\cdot w}) + \sigma^2(\varepsilon_{\cdot vw}) + \sigma^2(\varepsilon_{uvw}). \qquad (8.6.27)$$

Удобно ввести следующие величины:

$$\sigma^2(x(u, v, w)) = \sigma^2,$$
$$\sigma^2(\varepsilon_{u\cdot\cdot}) = \sigma^2_{\cdot 00}, \quad \sigma^2(\varepsilon_{\cdot v\cdot}) = \sigma^2_{0\cdot 0}, \quad \sigma^2(\varepsilon_{\cdot\cdot w}) = \sigma^2_{00\cdot},$$
$$\sigma^2(\varepsilon_{uv\cdot}) = \sigma^2_{\cdot\cdot 0}, \quad \sigma^2(\varepsilon_{u\cdot w}) = \sigma^2_{\cdot 0\cdot}, \quad \sigma^2(\varepsilon_{\cdot vw}) = \sigma^2_{0\cdot\cdot},$$
$$\sigma^2(\varepsilon_{uvw}) = \sigma^2_{\cdot\cdot\cdot}.$$

Тогда (8.6.27) может быть записано как

$$\sigma^2 = \sigma^2_{\cdot 00} + \sigma^2_{0\cdot 0} + \sigma^2_{00\cdot} + \sigma^2_{\cdot\cdot 0} + \sigma^2_{\cdot 0\cdot} + \sigma^2_{0\cdot\cdot} + \sigma^2_{\cdot\cdot\cdot} \qquad (8.6.28)$$

Пусть теперь $(u_1, \ldots, u_r)$, $(v_1, \ldots, v_s)$, $(w_1, \ldots, w_t)$ — независимые выборки из бесконечных совокупностей, имеющих к.ф.р. $F_1(u)$, $F_2(v)$ и $F_3(w)$ соответственно. Функция распределения выборки равна

$$\prod_{\xi=1}^{r} F_1(u_\xi) \prod_{\eta=1}^{s} F_2(v_\eta) \prod_{\zeta=1}^{t} F_3(w_\zeta). \qquad (8.6.29)$$

Наша матричная выборка третьего порядка есть множество случайных величин

$$(x(u_\xi, v_\eta, w_\zeta); \qquad \xi = 1, \ldots, r; \quad \eta = 1, \ldots, s; \quad \zeta = 1, \ldots, t),$$

которое может быть рассмотрено как матрица третьего порядка с r строками, s столбцами и t слоями.

Для этой выборки определим $x_{\xi\eta\zeta}$, $x'_{\xi\eta\zeta}$ и различные выборочные средние $\bar{x}_{\cdot\cdot\cdot}$, $\bar{x}_{\xi\cdot\cdot}$, $\bar{x}_{\cdot\eta\cdot}$, $\bar{x}_{\cdot\cdot\zeta}$, $\bar{x}_{\xi\eta\cdot}$, $\bar{x}_{\xi\cdot\zeta}$, $\bar{x}_{\cdot\eta\zeta}$ с помощью непосредственного обобщения определений (8.6.7). Каждая $x_{\xi\eta\zeta}$ есть сумма случайных величин, имеющих свойства, сформулированные в **8.6.3,** и мы можем написать

$$x_{\xi\eta\zeta} = \mu + \varepsilon_{\xi\cdot\cdot} + \varepsilon_{\cdot\eta\cdot} + \varepsilon_{\cdot\cdot\zeta} + \varepsilon_{\xi\eta\cdot} + \varepsilon_{\xi\cdot\zeta} + \varepsilon_{\cdot\eta\zeta} + \varepsilon_{\xi\eta\zeta}. \qquad (8.6.30)$$

Более того, определим следующие суммы квадратов:

$$S_T = \sum_{\xi, \eta, \zeta} (x_{\xi\eta\zeta} - \bar{x}_{...})^2,$$

$$S_{\cdot 00} = \sum_{\xi, \eta, \zeta} (\bar{x}_{\xi \cdot\cdot} - \bar{x}_{...})^2 \text{ (аналогично определяются } S_{0 \cdot 0} \text{ и } S_{00 \cdot}),$$

$$S_{\cdot\cdot 0} = \sum_{\xi, \eta, \zeta} (\bar{x}_{\xi\eta \cdot} - \bar{x}_{\xi \cdot\cdot} - \bar{x}_{\cdot \eta \cdot} + \bar{x}_{...})^2 \text{ (аналогично определяются } S_{\cdot 0 \cdot} \text{ и } S_{0 \cdot\cdot}), \tag{8.6.31}$$

$$S_{...} = \sum_{\xi, \eta, \zeta} (x_{\xi\eta\zeta} - \bar{x}_{\xi\eta \cdot} - \bar{x}_{\xi \cdot \zeta} - \bar{x}_{\cdot \eta\zeta} + \bar{x}_{\xi \cdot\cdot} + \bar{x}_{\cdot \eta \cdot} + \bar{x}_{\cdot\cdot \zeta} - \bar{x}_{...})^2.$$

Как нетрудно увидеть, они удовлетворяют соотношению

$$S_T = S_{\cdot 00} + S_{0 \cdot 0} + S_{00 \cdot} + S_{\cdot\cdot 0} + S_{\cdot 0 \cdot} + S_{0 \cdot\cdot} + S_{...}. \tag{8.6.32}$$

Обычно $S_{\cdot 00}$ называют *строчной компонентой* S_T, $S_{0\cdot 0}$ и $S_{00\cdot}$; $S_{\cdot\cdot 0}$ — *строчно-столбцовой компонентой взаимодействия* S_T (аналогично называются $S_{\cdot 0 \cdot}$ и $S_{0 \cdot\cdot}$) и $S_{...}$ — *остаточной* компонентой S_T.

Очевидно, что

$$\mathscr{E}(\bar{x}_{...}) = \mu.$$

При нахождении средних значений всех S можно увидеть, что они представляют собой линейные функции от величин типа:

$$\begin{aligned}
\mathscr{E}(x_{\xi\eta\zeta} x_{\xi'\eta'\zeta'}) &= \sigma^2; & \xi &= \xi', & \eta &= \eta', & \zeta &= \zeta', \\
&= \sigma^2_{\cdot\cdot 0} + \sigma^2_{\cdot 00} + \sigma^2_{0 \cdot 0}; & \xi &= \xi', & \eta &= \eta', & \zeta &\neq \zeta', \\
&= \sigma^2_{\cdot 0 \cdot} + \sigma^2_{\cdot 00} + \sigma^2_{00 \cdot}; & \xi &= \xi', & \eta &\neq \eta', & \zeta &= \zeta', \\
&= \sigma^2_{0 \cdot\cdot} + \sigma^2_{0 \cdot 0} + \sigma^2_{00 \cdot}; & \xi &\neq \xi', & \eta &= \eta', & \zeta &= \zeta', \\
&= \sigma^2_{\cdot 00}; & \xi &= \xi', & \eta &\neq \eta', & \zeta &\neq \zeta', \\
&= \sigma^2_{0 \cdot 0}; & \xi &\neq \xi', & \eta &= \eta', & \zeta &\neq \zeta', \\
&= \sigma^2_{00 \cdot}; & \xi &\neq \xi', & \eta &\neq \eta', & \zeta &= \zeta', \\
&= 0\ ; & \xi &\neq \xi', & \eta &\neq \eta', & \zeta &\neq \zeta',
\end{aligned} \tag{8.6.33}$$

где σ^2 определено выражением (8.6.28).

Действительно, производя вычисления, получаем

$$\begin{aligned}
\mathscr{E}(S_{00 \cdot}) &= rs(t-1)\left[\sigma^2_{00 \cdot} + \frac{1}{s}\sigma^2_{0 \cdot\cdot} + \frac{1}{r}\sigma^2_{\cdot 0 \cdot} + \frac{1}{rs}\sigma^2_{...}\right], \\
\mathscr{E}(S_{0 \cdot 0}) &= rt(s-1)\left[\sigma^2_{0 \cdot 0} + \frac{1}{t}\sigma^2_{0 \cdot\cdot} + \frac{1}{r}\sigma^2_{\cdot\cdot 0} + \frac{1}{rt}\sigma^2_{...}\right], \\
\mathscr{E}(S_{\cdot 00}) &= st(r-1)\left[\sigma^2_{\cdot 00} + \frac{1}{t}\sigma^2_{\cdot 0 \cdot} + \frac{1}{s}\sigma^2_{\cdot\cdot 0} + \frac{1}{st}\sigma^2_{...}\right], \\
\mathscr{E}(S_{0 \cdot\cdot}) &= r(s-1)(t-1)\left[\sigma^2_{0 \cdot\cdot} + \frac{1}{r}\sigma^2_{...}\right], \\
\mathscr{E}(S_{\cdot 0 \cdot}) &= s(r-1)(t-1)\left[\sigma^2_{\cdot 0 \cdot} + \frac{1}{s}\sigma^2_{...}\right], \\
\mathscr{E}(S_{\cdot\cdot 0}) &= t(r-1)(s-1)\left[\sigma^2_{\cdot\cdot 0} + \frac{1}{t}\sigma^2_{...}\right], \\
\mathscr{E}(S_{...}) &= (r-1)(s-1)(t-1)\sigma^2_{...},
\end{aligned} \tag{8.6.34}$$

и, конечно, значение $\mathscr{E}(S_T)$ есть сумма величин правых частей (8.6.34).

Аналогично дисперсия $\bar{x}...$ есть линейная функция величин типа (8.6.33) и оказывается равной

$$\sigma^2(\bar{x}...) = \frac{1}{r}\sigma^2_{\cdot 00} + \frac{1}{s}\sigma^2_{0\cdot 0} + \frac{1}{t}\sigma^2_{00\cdot} + \frac{1}{rs}\sigma^2_{\cdot\cdot 0} + \frac{1}{rt}\sigma^2_{\cdot 0\cdot} + \\ + \frac{1}{st}\sigma^2_{0\cdot\cdot} + \frac{1}{rst}\sigma^2_{\cdot\cdot\cdot} \qquad (8.6.35)$$

Мы оставляем читателю возможность сформулировать обобщение **8.6.2** на случай, когда в матричной выборке третьего порядка совокупности строчек, столбцов и слоев бесконечны.

Если читатель сравнит структуру соотношений (8.6.11) и (8.6.22), то ему нетрудно будет догадаться, почему справедливы соотношения (8.6.34) в том случае, когда в матричной выборке $\{x(u_\xi, v_\eta, w_\zeta)\}$ $(u_1, \ldots, u_r)$, $(v_1, \ldots, v_s)$ и $(w_1, \ldots, w_t)$ — независимые выборки из конечных совокупностей строк, столбцов и слоев соответственно. Аналогично, сравнивая (8.6.12) и (8.6.24), читатель легко сделает вывод формулы (8.6.35) для $\sigma^2(\bar{x}\ldots)$ в случае конечной совокупности. Мы оставляем читателю в качестве упражнения доказательство и исследования этих формул.

(d) Сбалансированные неполные матричные выборки. Вернемся теперь к случаю матричных выборок второго порядка ($x_{\xi\eta}$; $\xi = 1, \ldots, r$; $\eta = 1, \ldots, s$), которые мы обозначаем через $\{x_{\xi\eta}\}$. Выделим подвыборку $\{x_{\xi\eta}\}^*$, состоящую из s' элементов из каждой строки и r' элементов из каждого столбца $\{x_{\xi\eta}\}$. Если n — объем выборки, то n, r, s, r', s' должны быть положительными целыми числами, удовлетворяющими соотношению

$$n = rs' = r's. \qquad (8.6.36)$$

Пусть множество пар величин (ξ, η) для такого выбора обозначено через G^*. Мы будем называть такую подвыборку *сбалансированной неполной матричной выборкой*. Приведем примеры, в которых могут быть сделаны такие выборы G^*: $n = 12$, $r = 3$, $s = 6$, $r' = 2$, $s' = 4$; $n = 16$, $r = 4$, $s = 8$, $r' = 2$, $s' = 4$; $n = 40$, $r = 5$, $s = 10$, $r' = 4$, $s' = 8$. Читатель, интересующийся деталями конструкции специальных образцов выбора в таких выборках, отсылается к статьям Боуза (1939), Коннора (1952), Шрикхэнда (1952) и к книгам Кочрена и Кокса (1957), Кемпсорна (1952).

Пусть среднее значение неполной выборки, среднее значение ξ-й строки и η-го столбца обозначены через $\bar{x}^*_{\cdot\cdot}$, $\bar{x}^*_{\xi\cdot}$, $\bar{x}^*_{\cdot\eta}$ соответственно, и пусть

$$\begin{aligned} S^*_T &= \sum_{\xi,\eta}{}^* (x_{\xi\eta} - \bar{x}^*_{\cdot\cdot})^2, \\ S^*_{\cdot 0} &= \sum_{\xi,\eta}{}^* (\bar{x}_{\xi\cdot} - \bar{x}^*_{\cdot\cdot})^2, \\ S^*_{0\cdot} &= \sum_{\xi,\eta}{}^* (\bar{x}^*_{\cdot\eta} - \bar{x}^*_{\cdot\cdot})^2, \\ S^*_E &= S^*_T - S^*_{\cdot 0} - S^*_{0\cdot}, \end{aligned} \qquad (8.6.37)$$

где $\sum\limits_{\xi,\eta}^{*}$ означает суммирование по всем $(\xi, \eta) \in G^*$, а S_E^* — *остаточная* сумма квадратов.

Рассмотрим дисперсию $\bar{x}_{..}^*$ и средние значения величин S в случае бесконечных совокупностей строк и столбцов. Тогда все эти средние значения есть линейные функции от величин типа тех, что мы имеем в (8.6.10). Вычисление $\sigma^2(\bar{x}_{..}^*)$, $\mathcal{E}(S_T^*)$, $\mathcal{E}(S_{.0}^*)$ и $\mathcal{E}(S_{0.}^*)$ приводит к соотношению

$$\mathcal{E}(S_E^*) = \mathcal{E}(S_T^*) - \mathcal{E}(S_{.0}^*) - \mathcal{E}(S_{0.}^*). \tag{8.6.38}$$

Выбирая $r's = rs' = n$ и опуская детали выкладок, получаем

$$\begin{aligned}
&\mathcal{E}(\bar{x}_{..}^*) = \mu,\\
&\sigma^2(\bar{x}_{..}^*) = \frac{1}{n}\sigma_{..}^2 + \frac{1}{r}\sigma_{.0}^2 + \frac{1}{s}\sigma_{0.}^2,\\
&\mathcal{E}(S_{.0}^*) = (r-1)\sigma_{..}^2 + (r-1)s'\sigma_{.0}^2 + (r-r')\sigma_{0.}^2,\\
&\mathcal{E}(S_{0.}^*) = (s-1)\sigma_{..}^2 + (s-s')\sigma_{.0}^2 + (s-1)r'\sigma_{0.}^2,\\
&\mathcal{E}(S_E^*) = (n-r-s+1)\sigma_{..}^2 - (s-s')\sigma_{.0}^2 - (r-r')\sigma_{0.}^2.
\end{aligned} \tag{8.6.39}$$

Заметим, что если $r' = r$ и $s' = s$, то эти уравнения естественно переходят в (8.6.12) и (8.6.11).

Сбалансированные неполные матричные выборки, взятые из матричных выборок третьего и более высоких порядков, в общем случае должны удовлетворять некоторым дополнительным условиям и описываются значительно сложнее. Поэтому мы ограничимся обсуждением только одного специального случая матричных выборок третьего порядка, который, однако, играет основную роль в теории планирования экспериментов.

А именно, рассмотрим матричную выборку третьего порядка с r строками, r столбцами и r слоями, т. е. $\{x_{\xi\eta\zeta}\}$, $\xi, \eta, \zeta = 1, \ldots, r$. Наша сбалансированная неполная матричная выборка $\{x_{\xi\eta\zeta}\}^*$ связана тогда с таким выбором из $\{x_{\xi\eta\zeta}\}$, что она включает один элемент из каждой комбинации величин ξ, η, один элемент из каждой комбинации величин ξ, ζ и один элемент из каждой комбинации величин η, ζ.

Число элементов в так построенной выборке $\{x_{\xi\eta\zeta}\}^*$ равно r^2, так что $\{x_{\xi\eta\zeta}\}^*$ включает в себя только «скелетный» выбор из $\{x_{\xi\eta\zeta}\}$. Однако сбалансированность, соответствующая такой выборке $\{x_{\xi\eta\zeta}\}^*$, очень примечательна. Действительно, если все элементы из $\{x_{\xi\eta\zeta}\}^*$ спроектировать на «плоскость» $\xi\eta$ (плоскость строк-столбцов), то в каждой строке спроектированной $\{x_{\xi\eta\zeta}\}^*$ найдется один и только один элемент из каждого из r слоев $\{x_{\xi\eta\zeta}\}$. Аналогично в каждом столбце «плоскости» $\xi\eta$ найдется один и только один элемент из каждого из r слоев $\{x_{\xi\eta\zeta}\}$. Более того, аналогичные утверждения справедливы, если $\{x_{\xi\eta\zeta}\}^*$ спроектировать на «плоскость» $\xi\zeta$ или на «плоскость» $\eta\zeta$, заменив при этом соответствующим образом слова «строка», «столбец» и «слой». Каждая из этих спроектированных форм «выборки»

$\{x_{\xi\eta\zeta}\}^*$ может быть названа выбором строк, столбцов и слоев согласно методу *латинского квадрата*.

Пусть G^* — множество величин (ξ, η, ζ), использованных в некотором специальном отборе неполной выборки. Существует много способов выбора элементов в G^*. Читатель, интересующийся построением различных G^* (построение латинских квадратов), отсылается к работам Боуза (1938) и Манна (1943).

Мы будем интересоваться только средним значением $\bar{x}^*_{\cdot\cdot\cdot}$ и средними значениями $\bar{x}^*_{\xi\cdot\cdot}$, $\bar{x}^*_{\cdot\eta\cdot}$ и $\bar{x}^*_{\cdot\cdot\zeta}$ элементов ξ-й строки, η-го столбца и ζ-го слоя соответственно, в случае неполной выборки, и величинами

$$\begin{aligned} S_T^* &= \sum_{\xi,\eta,\zeta}{}^* (x_{\xi\eta\zeta} - \bar{x}^*_{\cdot\cdot\cdot})^2, \\ S^*_{\cdot 00} &= \sum_{\xi,\eta,\zeta}{}^* (\bar{x}^*_{\xi\cdot\cdot} - \bar{x}^*_{\cdot\cdot\cdot})^2, \\ S^*_{0\cdot 0} &= \sum_{\xi,\eta,\zeta}{}^* (\bar{x}^*_{\cdot\eta\cdot} - \bar{x}^*_{\cdot\cdot\cdot})^2, \\ S^*_{00\cdot} &= \sum_{\xi,\eta,\zeta}{}^* (\bar{x}^*_{\cdot\cdot\zeta} - \bar{x}^*_{\cdot\cdot\cdot})^2, \\ S_E^* &= S_T^* - S^*_{\cdot 00} - S^*_{0\cdot 0} - S^*_{00\cdot}, \end{aligned} \tag{8.6.40}$$

где $\sum\limits_{\xi,\eta,\zeta}{}^*$ означает суммирование по всем $(\xi, \eta, \zeta) \in G^*$.

Но дисперсия $\bar{x}^*_{\cdot\cdot\cdot}$ и средние значения всех S^* в (8.6.40) суть линейные функции *только* величин первых четырех типов, указанных в (8.6.33). Вычисляя средние значения, опуская при этом детали выкладок, находим

$$\begin{aligned} \mathscr{E}(S^*_{\cdot 00}) &= (r-1)\sigma_E^2 + (r-1)\sigma^2_{\cdot 00}, \\ \mathscr{E}(S^*_{0\cdot 0}) &= (r-1)\sigma_E^2 + (r-1)\sigma^2_{0\cdot 0}, \\ \mathscr{E}(S^*_{00\cdot}) &= (r-1)\sigma_E^2 + (r-1)\sigma^2_{00\cdot}, \\ \mathscr{E}(S_E^*) &= (r-1)(r-2)\sigma_E^2 + (r-1)^2[\sigma^2_{\cdot 00} + \sigma^2_{0\cdot 0} + \sigma^2_{00\cdot}], \end{aligned} \tag{8.6.41}$$

где σ_E^2 удовлетворяет условию

$$\sigma^2 = \sigma_E^2 + \sigma^2_{\cdot 00} + \sigma^2_{0\cdot 0} + \sigma^2_{00\cdot}. \tag{8.6.42}$$

Необходимо отметить на основании (8.6.27) и (8.6.28), что σ_E^2 есть сумма четырех компонент, а именно

$$\sigma_E^2 = \sigma^2_{\cdot\cdot\cdot} + \sigma^2_{\cdot\cdot 0} + \sigma^2_{\cdot 0\cdot} + \sigma^2_{0\cdot\cdot}. \tag{8.6.43}$$

8.7. Теория выборочного метода для порядковых статистик

Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности, имеющей *непрерывную* к. ф. р. $F(x)$. Расположим элементы $x_1, \ldots, x_n$ в порядке их возрастания и обозначим полученные упорядоченные величины через

$x_{(1)}, \ldots, x_{(n)}$, где $x_{(1)} < \ldots < x_{(n)}$. Эти новые случайные величины называются *порядковыми статистиками выборки*, $x_{(k)}$ называется *k-й порядковой статистикой*, часто k называют *номером* порядковой статистики и статистику $(x_{(1)}, \ldots, x_{(n)})$ — *вариационным рядом*. Интервалы $(-\infty, x_{(1)}]$, $[x_{(1)}, x_{(2)}], \ldots, (x_{(n)}, +\infty)$ называются *выборочными блоками* $B_1^{(1)}, \ldots, B_1^{(n+1)}$ соответственно, а функции $F(x_{(1)})$, $F(x_{(2)}) - F(x_{(1)}), \ldots, 1 - F(x_{(n)})$ этих блоков называются *долями* $u_1, \ldots, u_{n+1}$ соответственно. Поскольку сумма всех u равна 1, то мы будем обычно опускать u_{n+1}. Значок у B обозначает размерность. Выборочные блоки для размерностей 2 и более будут определены в § 8.7 (c). Заметим, что доля для данного выборочного блока есть просто сумма вероятностей в распределении популяций (P-мера), попадающих в этот выборочный блок, и является случайной величиной.

Теория выборочного метода для порядковых статистик и долей имеет фундаментальное значение при исследовании *непараметрических статистических выводов*, которые будут обсуждаться в главах 11 и 13. В этом разделе мы рассмотрим часть основных результатов теории выборочного метода для порядковых статистик. Читатель, интересующийся более подробными деталями, может почерпнуть их из книг Фрезера (1957), Гамбела (1958) и Кендалла (1953). Полная библиография публикаций в этой области приведена в работе Сэвиджа (1953).

(а) Выборочные распределения порядковых статистик. Если $F(x)$ непрерывна, то, как установленно в **7.1.1**, случайная величина $y = F(x)$ имеет *равномерное распределение* $R\left(\frac{1}{2}, 1\right)$. Пусть теперь $y_\xi = F(x_\xi)$, $\xi = 1, \ldots, n$. Тогда $(y_1, \ldots, y_n)$ есть выборка объема n из совокупности, имеющей равномерное распределение $R\left(\frac{1}{2}, 1\right)$.

Элемент вероятности (э. в.) этой выборки равен

$$1 \cdot dy_1 \ldots dy_n \tag{8.7.1}$$

внутри единичного «куба» $\{(y_1, \ldots, y_n) : 0 < y_\xi < 1, \ \xi = 1, \ldots, n\}$ в выборочном пространстве R_n и $0 \cdot dy_1 \ldots dy_n$ вне этого куба. Пусть $y_{(1)} < < \ldots < y_{(n)}$ — *порядковые статистики* выборки. Рассмотрим теперь задачу нахождения э. в. этих порядковых статистик. Вероятность того, что два или более элементов выборки $(y_1, \ldots, y_n)$ равны между собой, есть нуль. Значит, мы можем рассматривать только те точки в единичном кубе, координаты которых не равны между собой. Допустим, что $(y_1, \ldots, y_n)$ — такая точка P. Элемент вероятности, связанный с этой точкой, есть $1 \cdot dy_1 \ldots dy_n$. Но при перестановках координат P мы получаем в единичном кубе $n!$ таких точек. Если взять новые случайные величины $y_{(k)}$, такие, что $y_{(1)}$ — наименьшая координата этих $n!$ точек, $y_{(2)}$ — следующая наименьшая координата и т. д., то ф. п. в. случайной величины $(y_{(1)}, \ldots, y_{(n)})$, компоненты которой представляют собой порядковые статистики выборки $(y_1, \ldots, y_n)$, равна сумме ф. п. в. для этих $n!$ точек. В результате э. в. $(y_{(1)}, \ldots, y_{(n)})$ есть

$$n!\, dy_{(1)} \ldots dy_{(n)} \tag{8.7.2}$$

внутри $\{(y_{(1)}, \ldots, y_{(n)}): 0 < y_{(1)} < \ldots < y_{(n)} < 1\}$ и $0 \cdot dy_{(1)} \ldots dy_{(n)}$ вне этой области. Это есть n-мерное упорядоченное распределение Дирихле $D^*(1, \ldots, 1; 1)$, определенное в (7.7.17). Поэтому мы получаем следующий результат:

8.7.1. *Если* $(x_{(1)}, \ldots, x_{(n)})$ — *порядковые статистики для непрерывной к. ф. р.* $F(x)$, *то случайные величины* $F(x_{(1)}), \ldots, F(x_{(n)})$ *имеют упорядоченное n-мерное распределение Дирихле* $D^*(1, 1, \ldots, 1; 1)$.

Распределение для любой подпоследовательности из этих случайных величин можно получить отсюда с помощью **7.7.6.** Так что для одной из этих величин справедливо утверждение

8.7.2. *Случайная величина* $F(x_{(k)})$, $1 < k < n$, *имеет бета-распределение* $\mathrm{Be}(k, n-k+1)$.

Для s таких случайных величин можем получить результат:

8.7.3. *Случайные величины* $F(x_{(k_1)})$, $F(x_{(k_1+k_2)}), \ldots, F(x_{k_1+\ldots+k_s})$ *имеют упорядоченное s-мерное распределение Дирихле* $D^*(k_1, \ldots, k_s; n-k_1-\ldots-k_s+1)$.

Необходимо еще раз подчеркнуть, что в **8.7.1, 8.7.2** и **8.7.3** мы предполагали только, что $F(x)$ непрерывна. Если же мы теперь сделаем более сильные предположения, а именно, что существует ф. п. в. $f(x)$, то сможем найти э. в. порядковых статистик $x_{(1)}, \ldots, x_{(n)}$ и подпоследовательностей этих статистик. Так что в соответствии с **8.7.1, 8.7.2** и **8.7.3** получаем следующий результат:

8.7.1a. *Если* $(x_{(1)}, \ldots, x_{(n)})$ — *порядковые статистики выборки объема n из совокупности, имеющей к. ф. р.* $F(x)$ *и ф. п. в.* $f(x)$, *то элементы вероятности порядковых статистик суть*

$$n!\, f(x_{(1)}) \ldots f(x_{(n)})\, dx_{(1)} \ldots dx_{(n)}, \tag{8.7.3}$$

и выборочное пространство $(x_{(1)}, \ldots, x_{(n)})$ *представляет собой область в* R_n, *для которой* $-\infty < x_{(1)} < \ldots < x_{(n)} < +\infty$.

8.7.2a. *Вероятность* $x_{(k)}$, $1 \leqslant k \leqslant n$, *при условиях, сформулированных в* **8.7.1a**, *есть*

$$\frac{\Gamma(n+1)}{\Gamma(k)\,\Gamma(n-k+1)} [F(x_{(k)})]^{k-1} [1-F(x_{(k)})]^{n-k} f(x_{(k)})\, dx_{(k)} \tag{8.7.4}$$

и выборочное пространство для $x_{(k)}$ *есть* R_1.

8.7.3a. *Плотность вероятности для* $(x_{(k_1)}, x_{(k_1+k_2)}, \ldots, x_{(k_1+\ldots+k_s)})$ *при условиях, сформулированных в* **8.7.1a**, *равна*

$$\frac{\Gamma(n+1)}{\Gamma(k_1)\ldots\Gamma(k_s)\,\Gamma(n-k_1-\ldots-k_{s+1})} \cdot [F(x_{(k_1)})]^{k_1-1} [F(x_{(k_1+k_2)}) - \\ - F(x_{(k_1)})]^{k_2-1} \ldots [1-F(x_{(k_1+\ldots+k_s)})]^{n-k_1-\ldots-k_s} f(x_{(k_1)}) \ldots \times \\ \times f(x_{(k_1+\ldots+k_s)})\, dx_{(k_1)} \ldots dx_{(k_1+\ldots+k_s)}, \tag{8.7.5}$$

и выборочное пространство для $(x_{(k_1)}, x_{(k_1+k_2)}, \ldots, x_{(k_1+\ldots+k_s)})$ *есть область* R_s, *в которой*

$$-\infty < x_{(k_1)} < \ldots < x_{(k_1+\ldots+k_s)} < +\infty.$$

Формулы (8.7.3), (8.7.4) и (8.7.5) впервые были получены Крейгом (1932) и представляют интерес в различных частных случаях. Например, если в (8.7.4) мы положим $k=1$, то получим элемент вероятности для *наименьшего элемента из выборки*, а при $k=n$ находим элемент вероятности *наибольшего элемента из выборки*. Если n — нечетное и $k=n+\frac{1}{2}$, то получаем элемент вероятности для *выборочной медианы*. Если в (8.7.5) $s=2$, $k_1=1$, $k_2=n-1$, то получим э. в. *наибольшего и наименьшего* элементов в выборке:

$$n(n-1)[F(x_{(n)})-F(x_{(1)})]^{n-2} f(x_{(1)}) f(x_{(n)})\, dx_{(1)}\, dx_{(n)}. \quad (8.7.6)$$

Распределение *выборочного размаха* w может быть получено из (8.7.6). Для этого надо использовать преобразование

$$x_{(1)}=v, \quad x_{(n)}=v+w$$

и взять безусловное распределение w, т. е. проинтегрировать по v.

(b) Выборочное распределение одномерных долей. Распределения, появляющиеся в **8.7.1, 8.7.2** и **8.7.3**, имеют более широкое значение, чем это может показаться с первого взгляда. Допустим, что мы сделали преобразования

$$\begin{aligned} y_{(1)} &= u_1, \\ y_{(2)} &= u_1+u_2, \\ &\ldots\ldots\ldots\ldots \\ y_{(n)} &= u_1+\ldots+u_n. \end{aligned} \quad (8.7.7)$$

Элемент вероятности случайных величин $u_1, \ldots, u_n$ будет равен

$$n!\, du_1 \ldots du_n \quad (8.7.8)$$

в симплексе $S_n: \{(u_1, \ldots, u_n): u_\xi > 0,\ \xi=1, \ldots, n,\ \sum_{\xi=1}^{n} u_\xi < 1\}$ и $0 \cdot du_1 \ldots du_n$ вне S_n. А это есть не что иное, как n-мерное распределение Дирихле $D(1, \ldots, 1; 1)$, определенное в (7.7.1). Заметим теперь, что случайные величины $u_1, \ldots, u_n$ есть доли, введенные в рассмотрение в пункте 8.7 (a). Распределение долей, как мы видим, не зависит от к. ф. р. $F(x)$ совокупности, которая по предположению непрерывна. На основании этого можно сказать, что доли ведут себя *независимо от распределения*.

Мы можем в итоге сформулировать следующее утверждение.

8.7.4. *Пусть* $(x_{(1)}, \ldots, x_{(n)})$ — *порядковые статистики выборки с непрерывной к. ф. р.* $F(x)$. *Тогда доли* $u_1=F(x_{(1)})$, $u_2=F(x_{(2)})-F(x_{(1)})$, $\ldots$, $u_n=F(x_{(n)})-F(x_{(n-1)})$ *есть случайные величины, имеющие n-мерные распределения Дирихле* $D(1, \ldots, 1; 1)$.

Отметим, что распределение долей, как видно из **8.7.4**, полностью симметрично относительно своих переменных. Этот факт дает основание говорить иногда о выборочных блоках $B_1^{(1)}, \ldots, B_1^{(n+1)}$, соответствующих долям $u_1, \ldots, u_{n+1}$, где $u_{n+1} = 1 - u_1 - \ldots - u_n$, как о *статистически эквивалентных блоках*. Из упомянутого факта и из **7.7.2** следует, что

8.7.5. *Любая k-я ($k \leqslant n$) доля, упомянутая в* **8.7.4**, *имеет k-мерное распределение Дирихле* $D(1, \ldots, 1;\ n-k+1)$.

Применяя также **7.7.4**, получаем

8.7.6. *Сумма любых k долей из упомянутых в* **8.7.4** *имеет бета-распределение* $\mathrm{Be}(k,\ n-k+1)$.

Можно заметить, что **8.7.2** есть частный случай **8.7.6**, так как $F(x_{(k)})$ есть сумма *первых* k долей $u_1, \ldots, u_k$.

Окончательно находим, что применение **7.7.5** ведет к следующему результату:

8.7.7. *Если $v_1, \ldots, v_s$ — суммы $k_1, \ldots, k_s$ долей, упомянутых в* **8.7.4**, *причем нет долей, принадлежащих более чем к одному v, тогда распределение $(v_1, \ldots, v_s)$ есть s-мерное распределение Дирихле* $D(k_1, \ldots, k_s;\ n-k_1-\ldots-k_{s+1})$.

(с) Выборочные распределения многомерных долей. До сих пор мы рассматривали порядковые статистики выборок из одномерных распределений. Результаты, которые были сформулированы в **8.7.4**, **8.7.5**, **8.7.6** и **8.7.7**, остаются в силе, если обобщить определения выборочных блоков и долей на случаи распределений двух или более измерений. Рассмотрим случай двух измерений. Мы предположим, что $(x_{1\xi}, x_{2\xi};\ \xi = 1, \ldots, n)$ есть выборка из непрерывной двумерной к. ф. р. $F(x_1, x_2)$. Мы введем *упорядочивающую функцию* $h(x_1, x_2)$ такую, что $w = h(x_1, x_2)$ есть случайная величина, которая имеет непрерывную к. ф. р. $H(w)$. Тогда случайные величины $w_\xi = h(x_{1\xi}, x_{2\xi})$, $\xi = 1, \ldots, n$ составляют выборку из совокупности, распределение которой равно $u(x_1, x_2)$, и эта выборка может быть упорядочена. Пусть порядковые статистики выборки $(w_1, \ldots, w_n)$ есть $(w_{(1)}, \ldots, w_{(n)})$. Доли $u'_1 = H(w_{(1)})$, $u'_2 = H(w_{(2)}) - H(w_{(1)}), \ldots, u'_n = H(w_{(n)}) - H(w_{(n-1)})$ и случайные величины, связанные соответственно с двумерными блоками $B_2^{(1)}, \ldots, B_2^{(n+1)}$, на которые разделена упорядочивающими «кривыми» $w_{(\xi)} = h(x_1, x_2)$, $\xi = 1, \ldots, n$, плоскость x_1, x_2, изображены на рис. 8.1.

Из предыдущих исследований очевидно, как надо определить доли $u'_1, \ldots, u'_n$ для k-мерных выборочных блоков $B_k^{(1)}, \ldots, B_k^{(n+1)}$, полученных с помощью k-мерной упорядочивающей функции $h(x_1, \ldots, x_k)$ с непрерывной к. ф. р. $H(w)$. Если $w_\xi = h(x_{1\xi}, \ldots, x_{k\xi})$, $\xi = 1, \ldots, n$, и $(w_{(1)}, \ldots, w_{(n)})$ — порядковые статистики, соответствующие выборке $(w_1, \ldots, w_n)$, т. е. $(w_{(1)}, \ldots, w_{(n)})$ — порядковые статистики совокупности, распределение которой совпадает с распределением $h(x_1, \ldots, x_k)$, то получаем следующий результат:

8.7.8. *Доли $u'_1, \ldots, u'_n$ для k-мерных блоков, определенных с помощью упорядочивающей функции $h(x_1, \ldots, x_n)$, имеют те же свойства распределений, что и те, которые характеризуют распределения долей $u_1, \ldots, u_n$ и которые были сформулированы в* **8.7.4**, **8.7.5**, **8.7.6** *и* **8.7.7**.

Определение долей для двух и более измерений более общо, чем может показаться сначала. Вальд (1943) дал метод построения

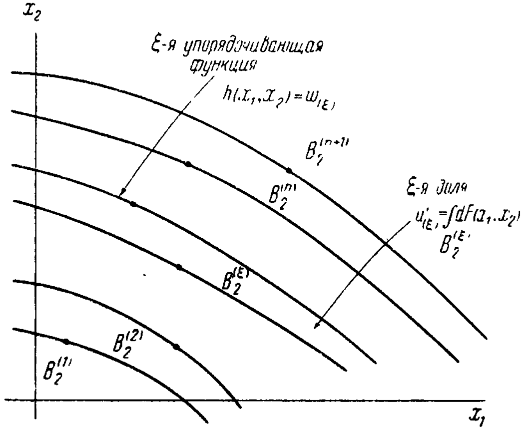

Рис. 8.1. Выборочные блоки и доли для случая двух измерений
а) ξ-я упорядочивающая функция: $h(x_1, x_2) = \omega_{(\xi)}$;
б) ξ-я доля: $u'_{(\xi)} = \int\limits_{B_2(\xi)} dF(x_1, x_2)$.

долей, основанный на прямоугольных выборочных блоках, который был существенно обобщен Тьюки (1947).

Мы рассмотрим метод Тьюки для случая двух измерений. Вместо одной упорядочивающей функции $h(x_1, x_2)$ для n выборочных точек (x_{1_ξ}, x_{2_ξ}), $\xi = 1, \ldots, n$, мы можем использовать n таких функций $h_\eta(x_1, x_2)$, $\eta = 1, \ldots, n$. Предположим, что каждая из функций $h_\eta(x_1, x_2)$ имеет непрерывную функцию распределения и функция распределения совокупности $F(x_1, x_2)$ также непрерывна. Некоторые или все эти упорядочивающие функции могут совпадать между собой. Если мы возьмем $w_\xi^\eta = h_\eta(x_{1_\xi}, x_{2_\xi})$, то для каждого значения η будем иметь множество порядковых статистик $w_{(\xi)}^\eta$, $\xi = 1, \ldots, n$. Разделим теперь плоскость x_1, x_2 следующим образом: кривая $w_{(1)}^1 = h_1(x_1, x_2)$ делит x_1x_2-плоскость на два множества $B_2^{(1)}$ и $\overline{B}_2^{(1)}$, где $B_2^{(1)}$ — множество точек, для которых $h_1(x_1, x_2) < w_{(1)}^1$. Определим первую долю, назовем ее u_1^*, как P-меру (вероятность, определенная из $F(x_1, x_2)$) выборочного блока $B_2^{(1)}$. Кривая $w_{(2)}^2 = h_2(x_1, x_2)$ делит множество $\overline{B}_2^{(1)}$ на два множества $B_2^{(2)}$ и $\overline{B}_2^{(2)}$ таких, что для $B_2^{(2)}$

$h_2(x_1, x_2) < w_{(2)}^2$. Пусть доля u_2^* есть P-мера $B_2^{(2)}$. Продолжая этот процесс, мы соответствующим образом определим множества $B_2^{(1)}$, $B_2^{(2)}, \ldots, B_2^{(n)}$ такие, что $B_2^{(2)} \subset \overline{B}_2^{(1)}, \ldots, B_2^{(n)} \subset \overline{B}_2^{(n-1)}$. Останется, конечно, остаточное множество $B_2^{(n+1)}$ такое, что $B_2^{(1)} \cup \ldots \cup B_2^{(n)} \cup B_2^{(n+1)} = R_2$. P-меры $B_2^{(1)}, \ldots, B_2^{(n)}$ есть определенные общим образом доли $u_1^*, \ldots, u_n^*$. Рис. 8.2. показывает пример для $n = 10$, когда упорядочивающие функции $h_\eta(x_1, x_2)$, $\eta = 1, \ldots, 10$, суть прямые линии, составляющие с x_1-осью углы $(\eta - 1)\,90°$.

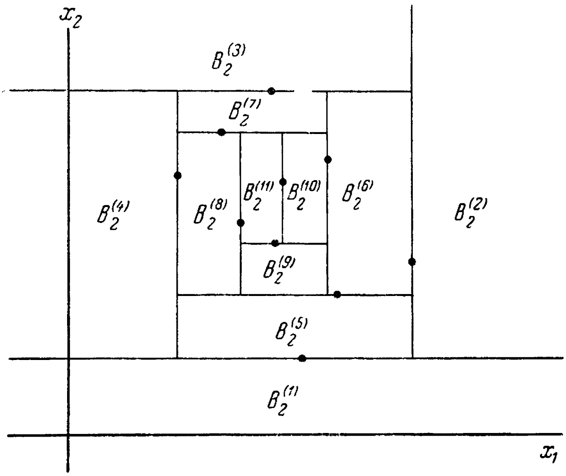

Рис. 8.2. Пример двумерных выборочных блоков, порожденных горизонтальными и вертикальными линиями, являющихся по отдельности графиками упорядочивающих функций.

Определение k-мерных аналогий долей $u_1^*, \ldots, u_n^*$ получается самым прямым путем, и мы оставляем это читателю. Справедливо следующее утверждение.

8.7.9. *Распределения долей $u_1^*, \ldots, u_n^*$ для k-мерных блоков (определенных указанным выше способом) обладают теми же свойствами, что и распределения долей $u_1, \ldots, u_n$, именно, указанными в теоремах* **8.7.4**, **8.7.5**, **8.7.6** *и* **8.7.7**.

Для доказательства **8.7.9** достаточно показать, что $u_1^*, \ldots, u_n^*$ имеют n-мерное распределение Дирихле $D(1, \ldots, 1; 1)$ для $k = 2$. Если $F(x_1, x_2)$ — функция распределения совокупности, из которой взята выборка $(x_{1\xi}, x_{2\xi};\ \xi = 1, \ldots, n)$, то э. в. этой выборки, которую мы обозначим через O_n, есть

$$\prod_{\xi=1}^{n} dF(x_{1\xi}, x_{2\xi}). \qquad (8.7.9)$$

Пусть $(x_{1*}^{(1)}, x_{2*}^{(1)})$ — элемент выборки, который доставляет наименьшее значение $h_1(x_1, x_2)$ среди всех n элементов выборки, и пусть $(x_{1\xi_1}^{(1)}, x_{2\xi_1}^{(1)};\ \xi_1 = 1, \ldots, n-1)$ суть $n-1$ элементов выборки, обозначенные через $O_{n-1}^{(1)}$, полученные выбрасыванием $(x_{1*}^{(1)}, x_{2*}^{(1)})$ из O_n. Элемент вероятности $(x_{1*}^{(1)}, x_{2*}^{(1)})$ и $(x_{1\xi_1}^{(1)}, x_{2\xi_1}^{(1)};\ \xi_1 = 1, \ldots, n-1)$ есть

$$n\,dF(x_{1*}^{(1)}, x_{2*}^{(1)}) \prod_{\xi_1=1}^{n-1} dF(x_{1\xi_1}^{(1)}, x_{2\xi_1}^{(1)}). \tag{8.7.10}$$

Пусть теперь u_1'' — доля, связанная с множеством $B_1^{(1)}$ в плоскости (x_1, x_2), для которого $h_1(x_1, x_2) < h_1(x_{1*}^{(1)}, x_{2*}^{(1)})$. Из **8.8.2** следует, что распределение u_1'' дается распределением $\mathrm{Be}(1, n)$, т. е. э. в. u_1'' есть

$$n(1-u_1'')^{n-1}du_1'', \tag{8.7.11}$$

где

$$u_1'' = \int\limits_{B_2^{(1)}} dF(x_1, x_2), \qquad du_1'' = dF(x_{1*}^{(1)}, x_{2*}^{(1)}). \tag{8.7.12}$$

Рассмотрим теперь $(2n-2)$-мерную условную случайную величину $(x_{1\xi_1}^{(1)}, x_{2\xi_2}^{(2)} \mid x_{1*}^{(1)}, x_{2*}^{(2)};\ \xi = 1, \ldots, n)$. Элемент ее вероятности дается отношением (8.7.10) к (8.7.11), что сводится к

$$\prod_{\xi_1=1}^{n-1} dF^{(1)}(x_{1\xi_1}^{(1)}, x_{2\xi_1}^{(1)}), \tag{8.7.13}$$

где

$$F^{(1)}(x_{1\xi_1}^{(1)}, x_{2\xi_1}^{(1)}) = \frac{F(x_{1\xi_1}^{(1)}, x_{2\xi_1}^{(1)})}{1-u_1''}. \tag{8.7.14}$$

Это означает, что для фиксированных величин $(x_{1*}^{(1)}, x_{2*}^{(1)})$ и, следовательно, для фиксированного значения u_1'', остающиеся $n-1$ элементов исходной выборки ведут себя как выборка O_{n-1} из $n-1$ элементов совокупности, имеющей функцию распределения

$$F^{(1)}(x_1, x_2) = \frac{F(x_1, x_2)}{1-u_1''}. \tag{8.7.15}$$

Эта функция распределения получается приписыванием нулевой вероятности для $B_2^{(1)}$ и нормировкой на единицу части исходного распределения совокупности, соответствующей $\overline{B}_2^{(1)}$.

Теперь мы повторим описанный выше процесс. Пусть $(x_{1*}^{(2)}, x_{2*}^{(2)})$ — элемент выборки, который доставляет $h_2(x_1, x_2)$ наименьшее значение среди $n-1$ элементов $O_{n-1}^{(1)}$, и пусть $(x_{1\xi_2}^{(2)}, x_{2\xi_2}^{(2)};\ \xi_2 = 1, \ldots, n-2)$, обозначенные через $O_{n-2}^{(2)}$, суть $n-2$ элементов из $O_{n-2}^{(1)}$, полученных выбрасыванием $(x_{1*}^{(2)}, x_{2*}^{(2)})$. Элемент вероятности $(x_{1*}^{(2)}, x_{2*}^{(2)})$ и $(x_{1\xi_2}^{(2)}, x_{2\xi_2}^{(2)};\ \xi_2 = 1, \ldots, n-2)$; равен

$$(n-1)\,dF^{(1)}(x_{1*}^{(2)}, x_{2*}^{(2)}) \prod_{\xi_2=1}^{n-2} dF^{(1)}(x_{1\xi_2}^{(2)}, x_{2\xi_2}^{(2)}). \tag{8.7.16}$$

Пусть $B_2^{(2)}$ — подмножество $\overline{B}_2^{(1)}$, для которого $h_2(x_1, x_2) < h_2(x_{1*}^{(2)}, x_{2*}^{(2)})$, и пусть $u_2'' = \int\limits_{B_2^{(2)}} dF^{(1)}(x_1, x_2)$ — *условная доля*, связанная с $B_2^{(2)}$ и определяемая $F^{(1)}(x_1, x_2)$. В качестве частного случая **8.7.2** получаем, что u_2'' имеет элемент вероятности

$$(n-1)(1-u_2'')^{n-2}\, du_2'', \tag{8.7.17}$$

а $(x_{1\xi_2}^{(2)}, x_{2\xi_2}^{(2)} \mid x_{1*}^{(2)}, x_{2*}^{(2)};\ \xi_2 = 1, \ldots, n-2)$ — элемент вероятности

$$\prod_{\xi_2=1}^{n-2} dF^{(2)}(x_{1\xi_2}^{(2)}, x_{2\xi_2}^{(2)}), \tag{8.7.18}$$

где

$$F^{(2)}(x_{1\xi_2}^{(2)}, x_{2\xi_2}^{(2)}) = \frac{F^{(1)}(x_{1\xi_2}^{(2)}, x_{2\xi_2}^{(2)})}{1-u_2''}. \tag{8.7.18a}$$

Таким образом, при фиксированных величинах $(x_{1*}^{(1)}, x_{2*}^{(1)})$ и $(x_{1*}^{(2)}, x_{2*}^{(2)})$ элементы $O_{n-2}^{(1)}$ ведут себя как элементы выборки из совокупности, имеющей к. ф. р.

$$F^{(2)}(x_1, x_2) = \frac{F'(x_1, x_2)}{1-u_2''}. \tag{8.7.19}$$

Распределение u_1'', u_2'' и элементов $O_{n-2}^{(2)}$ равно тогда произведению (8.7.11), (8.7.17) и (8.7.18).

Продолжая описанный выше процесс, мы получаем последовательность долей $u_1'', \ldots, u_n''$, имеющую распределение с э. в.

$$n!\,(1-u_1'')^{n-1} \ldots (1-u_n'')^{n-n}\, du_1'' \ldots du_n''. \tag{8.7.20}$$

Выражая их в терминах долей u_ξ^*, $\xi = 1, \ldots, n$, для которых справедливо **8.7.9**, получаем

$$\begin{aligned} u_1'' &= u_1^*, \\ u_2'' &= \frac{u_2^*}{1-u_1^*}, \\ &\vdots \\ u_n'' &= \frac{u_n^*}{1-u_1^*-\ldots-u_{n-1}^*}. \end{aligned} \tag{8.7.21}$$

Применяя это преобразование к (8.7.20), получаем

$$n!\; du_1^* \ldots du_n^*. \tag{8.7.22}$$

Но это есть не что иное, как элемент вероятности n-мерного распределения Дирихле $D(1, \ldots, 1;\ 1)$, что и доказывает утверждение **8.7.9**.

Дальнейшие результаты в теории выборочного метода для долей были получены Фрезером (1951, 1953), Фрезером и Гаттменом (1956), и Кемперманом (1956). В частности, они показали, как получать статистически эквивалентные блоки, используя упорядочивающие функции, связанные с упорядочивающими функциями, которые применялись для определения выборочных блоков. Расширение понятий выборочных блоков и долей в случае разрывных функций распределения рассматривалось Тьюки (1948).

8.8. Порядковые статистики в выборках из конечных совокупностей

Пусть π_N — конечная совокупность, x-величины элементов которой различны, т. е. $x_{o1} < \ldots < x_{oN}$. Пусть $(x_{(1)}, \ldots, x_{(n)})$ — порядковые статистики выборки объема n из π_N. Мы рассмотрим теорию выборочного метода для порядковой статистики $x_{(k)}$ k-го порядка. Из комбинаторных соображений очевидно, что

$$p(x_{(k)} = x_{ot}) = \frac{\binom{t-1}{k-1}\binom{N-t}{n-k}}{\binom{N}{n}} = p_{N,n,k}(t), \tag{8.8.1}$$

где $t = k, k+1, \ldots, N-k+k$. Таким образом, мы можем рассматривать $p_{N,n,k}(t)$ либо как (i) ф. в. случайной величины $x(k)$, ее точки сосредоточения массы есть x_{ot}, $t = k, k+1, \ldots, N-n+k$, либо как (ii) ф. в. случайной величины t. Последняя же есть *ранг* x-величины в π_N, который совпадает с порядковой статистикой k-порядка в выборке. Проще иметь дело со случайной величиной t, чем со случайной величиной $x_{(k)}$, так как точки сосредоточения массы t есть следующие друг за другом целые числа $k, k+1, \ldots, N-n+k$, в то время как для $x_{(k)}$ они сосредоточены в $x_{o,k}$, $x_{o,k+1}, \ldots, x_{o,N-n+k}$.

Факториальные моменты t довольно легко найти. А именно, используя обозначения (3.3.15), получаем

$$\mathscr{E}((t+r-1)^{[r]}) = \sum_{t=k}^{N-n+k} (t+r-1)^{[r]} p_{N,n,k}(t) =$$

$$= \frac{(k+r-1)^{[r]}\binom{N+r}{n+r}}{\binom{N}{n}} \sum_{t=k}^{N-n+k} \frac{\binom{t+r-1}{k+r-1}\binom{N-t}{n-k}}{\binom{N+r}{n+r}}. \tag{8.8.2}$$

Но сумма в правой части равна 1. Поэтому

$$\mathscr{E}((t+r-1)^{[r]}) = \frac{(k+r-1)^{[r]}\binom{N+r}{n+r}}{\binom{N}{n}}. \tag{8.8.3}$$

Выбирая $r=1$ и $r=2$, находим среднее значение и дисперсию

$$\mathscr{E}(t) = \frac{k(N+1)}{(n+1)}, \qquad \sigma^2(t) = \frac{k(N+1)(N-n)(n-k+1)}{(n+1)^2(n+2)}. \tag{8.8.4}$$

В итоге

8.8.1. *Если* π_N — *конечная совокупность, x-величины элементов которой все различны,* $x_{01} < \ldots < x_{0N}$, *и если* $(x_{(1)}, \ldots, x_{(n)})$ — *порядковые статистики выборки из* π_N, *то ф. в.* x_k *дается* (8.8.1), *точки сосредоточения массы при этом есть* x_{0t}, $t = k, k+1, \ldots, N-n+k$. *Значение* $\mathfrak{E}((t+r-1)^{[r]})$ *дается выражением* (8.8.3).

Если мы рассмотрим бесконечную последовательность случайных величин

$$u_{k,N} = \frac{t}{N}, \quad N = k, k+1, \ldots, \tag{8.8.5}$$

то $\lim\limits_{N\to\infty} u_{k,N}$ есть доля x-величин из бесконечной совокупности, которые превышают $x_{(k)}$. Интуитивно очевидно, что $\lim\limits_{N\to\infty} u_{k,N}$ соответствует $F(x_{(k)})$ для случая выборки из бесконечной совокупности, имеющей к. ф. р. $F(x)$. В согласии с **8.7.2** $F(x)$ должно быть бета-распределением $\mathrm{Be}(k, n-k+1)$. Чтобы убедиться в справедливости этого утверждения, рассмотрим

$$\lim_{N\to\infty} \mathfrak{E}\left(\frac{t}{N}\right)^r = \lim_{N\to\infty} \mathfrak{E}\left(\frac{(t+r-1)^{[r]}}{N^r}\right) = \frac{\Gamma(n+1)\,\Gamma(k+r)}{\Gamma(k)\,\Gamma(n+r+1)}. \tag{8.8.6}$$

Но это есть k-й момент случайной величины, имеющей бета-распределение $\mathrm{Be}(k, n-k+1)$. А из **5.5.3** следует, что последовательность случайных величин (8.8.5) сходится по распределению к случайной величине с этим бета-распределением. Поэтому

8.8.2. *Пусть* π_N, $N = n, n+1, \ldots$ — *последовательность конечных совокупностей таких, что x-величины элементов в* π_N *все различные,* $x_{01} < \ldots < x_{0N}$. *Пусть* $(x_{(1)}, \ldots, x_{(n)})$ — *порядковые статистики выборки объема n из* π_N *и пусть* $u_{k,N}$ — *число элементов в* π_N, *x-величины которых превышают* $x_{(k)}$. *Последовательность случайных величин* $u_{k,N}$, $N = n, n+1, \ldots$, *сходится по распределению к случайной величине, имеющей бета-распределение* $\mathrm{Be}(k, n-k+1)$.

ЗАДАЧИ

8.1. *Пусть* $(x_1, \ldots, x_n)$ — случайная выборка из совокупности, распределение которой имеет конечные моменты $\mu_2, \ldots, \mu_{2r}$ относительно среднего значения распределения, и пусть

$$m_r = \frac{1}{n}\sum_{\xi=1}^{n} (x_\xi - \bar{x})^r.$$

Показать, что

$$\sigma^2(m_r) = \frac{1}{n}\left[\mu_{2r} - \mu_r^2 + r\mu_{r-1}(r\mu_2\mu_{r-1} - 2\mu_{r+1})\right] + O\left(\frac{1}{n^2}\right).$$

8.2. Пусть $\bar{x}_1$, $\bar{x}_2$ — выборочные средние значения и s_1^2, s_2^2 — выборочные дисперсии двух независимых случайных выборок объема n_1 и n_2 из $N(\mu_1, \sigma^2)$ и $N(\mu_2, \sigma^2)$ соответственно. Показать, что при этих условиях

$$\frac{[(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)]}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\left[\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\right]^{1/2}}}$$

имеет распределение Стьюдента $S(n_1+n_2-2)$.

8.3. Пусть $(x_{11}, \ldots, x_{1n})$ и $(x_{21}, \ldots, x_{2n})$ — независимые выборки из $N(\mu_1, \sigma_1^2)$ и $N(\mu_2, \sigma_2^2)$ соответственно и пусть $\bar{x}_1$, $\bar{x}_2$ — выборочные средние значения, s_1^2, s_2^2 — выборочные дисперсии и $r=\frac{1}{(n-1)s_1 s_2}\sum_{\xi=1}^{n}(x_{1\xi}-\bar{x}_1)\times$ $\times(x_{2\xi}-\bar{x}_2)$ — коэффициент корреляции между двумя выборками. Показать, что

$$\frac{[(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)]\sqrt{n}}{[s_1^2+s_2^2-2rs_1s_2]^{1/2}}$$

имеет распределение Стьюдента $S(n-1)$.

8.4. Предположим, что $\bar{x}_1$ и $\bar{x}_2$ — средние значения независимых выборок объема n_1 и n_2 соответственно из распределений, имеющих одинаковые дисперсии, равные σ^2. Пусть $\bar{x}$ — среднее значение двух выборок, объединенных в одну выборку объема n_1+n_2. Показать, что дисперсия $\bar{x}-\bar{x}_1$ равна

$$\frac{\sigma^2 n_2}{[n_1(n_1+n_2)]}.$$

8.5. Пусть x — случайная величина, имеющая распределение $N(\mu, \sigma^2)$, и $F(x)$ — функция распределения этой нормальной величины. Показать, что коэффициент корреляции между случайными величинами x и $F(x)$ равен $\sqrt{\frac{3}{\pi}}$.

8.6. Пусть (x, y) — пара случайных величин, имеющая некоторое двумерное нормальное распределение с коэффициентом корреляции ρ, и пусть $F(y)$ — функция распределения y. Показать, что коэффициент корреляции между случайными величинами x и $F(y)$ равен $\sqrt{\frac{3}{\pi}}\rho$.

8.7. Используя **8.2.2** при $r=2$ и $g(x_1, x_2)\equiv x_1x_2$, построить $Q_{[2]}(x_1, \ldots$ $\ldots, x_n)$ для выборки объема n из распределения, имеющего конечные первый и второй моменты μ_1', μ_2', найдя тем самым несмещенную оценку для $(\mu_1')^2$. Найдите дисперсию этой оценки.

8.8. Пусть $(x_1, \ldots, x_n)$ — простая случайная выборка из конечной популяции π_N объема N, имеющей дисперсию σ^2. Показать, что для любых (вещественных) постоянных $c_1, \ldots, c_n$ дисперсия $c_1x_1+c_2x_2+\ldots+c_nx_n$ есть

$$\left[\left(c_1^2+\ldots+c_n^2\right)-\frac{1}{N}\left(c_1+\ldots+c_n\right)^2\right]\sigma^2.$$

Требуется показать также, что дисперсия разности между средними значениями двух случайных выборок объемов n_1 и n_2, $n_1+n_2<N$, взятых без возвращения из π_N, равна $\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$.

8.9. Допустим, что выборки объемов $n_1, \ldots, n_k$ — независимые выборки из одного нормального распределения $N(\mu, \sigma^2)$. Пусть $\bar{x}_1, \ldots, \bar{x}_k$ — средние значения, а $s_1^2, \ldots, s_k^2$ — дисперсии этих выборок.

Пусть

$$u=\frac{1}{\sigma^2}\left[(n_1-1)\,s_1^2+\ldots+(n_k-1)\,s_k^2\right],$$

$$v=\frac{1}{\sigma^2}\left[n_1(\bar{x}_1-\bar{x})^2+\ldots+n_k(\bar{x}_k-\bar{x})^2\right],$$

$$w=\frac{1}{\sigma^2}\left[n(\bar{x}-\mu)^2\right],$$

где $\bar{x}$ — среднее значение простой выборки, составленной из всех выборок, так что $n=n_1+\ldots+n_k$. Требуется показать с помощью теоремы Кочрена или с помощью аппарата характеристических функций, что u, v и w — независимые величины, имеющие χ^2-распределения $C(n_1+\ldots+n_k-k)$, $C(k-1)$ и $C(1)$ соответственно.

8.10. Пусть $(y_1,\ldots,y_n)$ — независимые случайные величины из $N(\mu+\beta x_1;\sigma_2)$, $\ldots$, $N(\mu+\beta x_n;\sigma^2)$ соответственно, где $\mu,\beta,x_1,\ldots,x_n$ постоянные. Пусть $\hat{\mu}$ и $\hat{\beta}$ — значения μ и β, которые минимизируют сумму квадратов

$$\sum_{\xi=1}^{n}(y_\xi-\mu-\beta x_\xi)^2,$$

и пусть

$$v=\frac{1}{\sigma^2}\sum_{\xi=1}^{n}(y_\xi-\hat{\mu}-\hat{\beta}\,x_\xi)^2.$$

Требуется показать, что $(\hat{\mu},\ \hat{\beta})$ имеет двумерное нормальное распределение $N(\{\mu,\ \beta\};\ \|\sigma_{ij}\|)$, где

$$\|\sigma_{ij}\|=\begin{Vmatrix}\dfrac{\left(\sum x_\xi^2\right)\sigma^2}{na_n} & -\dfrac{\bar{x}\sigma^2}{a_n}\\ -\dfrac{\bar{x}\sigma^2}{a_n} & \dfrac{\sigma^2}{a_n}\end{Vmatrix}$$

и $a_n=\sum\limits_{\xi}(x_\xi-\bar{x})^2$, и что v имеет χ^2-распределение $C(n-2)$. Далее показать, что $(\hat{\mu},\ \hat{\beta})$ и v независимы.

8.11. Пусть $(x_1,\ldots,\ x_n)$ — выборка из $N(\mu,\ \sigma^2)$. Показать, что характеристическая функция

$$u=\frac{1}{\sigma^2}\sum_{\xi=1}^{n}(x_\xi-\mu+\delta)^2$$

равна

$$e^{-\frac{1}{2}\beta^2}\sum_{r=0}^{\infty}\frac{1}{r!}\left(\frac{\beta^2}{2}\right)^r(1-2it)^{-\frac{1}{2}n-r},$$

где $\beta^2=n\delta^2/\sigma^2$, и, таким образом, что ф. п. в. u равна

$$f(u)=e^{-\frac{1}{2}\beta^2}\sum_{r=0}^{\infty}\frac{1}{r!}\left(\frac{\beta^2}{2}\right)^r f_{n+2r}(u),$$

где $f_{n+2r}(u)$ — ф. п. в. для χ^2 с $(n+2r)$ степенями свободы. (Распределение u впервые было получено Фишером (1928b) и называется *нецентральным χ^2-распределением* с n степенями свободы и параметром β^2.)

8.12. Пусть n, $\bar{x}$, s^2 — объем, среднее значение и дисперсия выборки из $N(\mu, \sigma^2)$. Показать, что распределение

$$t = \frac{(\bar{x} - \mu + \delta)\sqrt{n}}{s}$$

имеет ф. п. в.

$$f(t) = \frac{e^{-\frac{1}{2}\gamma^2}\Gamma\left(\frac{1}{2}n\right)}{\sqrt{\pi(n-1)}\,\Gamma\left(\frac{1}{2}n - \frac{1}{2}\right)}\left(1 + \frac{t^2}{n-1}\right)^{-\frac{1}{2}n} g(t),$$

где

$$g(t) = \sum_{r=0}^{\infty}\left(\frac{\sqrt{2}\gamma t}{\sqrt{n-1}}\right)^r\left(1 + \frac{t^2}{n-1}\right)^{-\frac{1}{2}r}\frac{\Gamma\left[\frac{1}{2}(n+r)\right]}{n!\,\Gamma\left(\frac{1}{2}n\right)}$$

и $\gamma = \sqrt{n}\,\delta/\sigma$ (это распределение было получено Тэнгом (1938) и называется *нецентральным распределением Стьюдента* с $(n-1)$ степенями свободы и параметром γ).

8.13. Найти вид формул (8.6.34) и (8.6.35) в случае конечной совокупности, если число элементов в совокупности строк, столбцов и слоев равно N_1, N_2 и N_3 соответственно.

8.14. Пусть $(x_1, \ldots, x_n)$ — выборка из равномерного распределения $R\left(\frac{1}{2}\theta, \theta\right)$. Показать, что выборочное распределение $u = \max(x_1, \ldots, x_n)$ имеет ф. п. в. nu^{n-1}/θ^n на $(0, \theta)$ и 0 вне $(0, \theta)$.

8.15. Пусть из равномерного распределения $R\left(\frac{1}{2}, 1\right)$ взята выборка объема $(2n+1)$. Показать, что медиана этой выборки имеет бета-распределение

$$\mathrm{Be}\,(n+1,\ n+1).$$

8.16. Пусть $(x_1, \ldots, x_n)$ — выборка из равномерного распределения $R(\mu, \omega)$. Показать, что п. ф. р. размаха выборки r равна

$$\frac{n(n-1)}{\omega}\left(\frac{r}{\omega}\right)^{n-2}\left(1 - \frac{r}{\omega}\right)$$

для $r \in (0, \omega)$ и 0 для $r \notin (0, \omega)$.

8.17. Пусть $(x_{(1)}, \ldots, x_{(n)})$ — порядковые статистики выборки объема n из совокупности, имеющей непрерывную к. ф. р. $F(x)$. Показать, что ковариационная матрица $(F(x_{(k_1)}), F(x_{(k_2)}))$, где $1 \leqslant k_1 < k_2 \leqslant n$, равна

$$\left\| \begin{matrix} \dfrac{p_1(1-p_1)}{(n+2)} & \dfrac{p_1(1-p_2)}{(n+2)} \\ \dfrac{p_1(1-p_2)}{(n+2)} & \dfrac{p_2(1-p_2)}{(n+2)} \end{matrix} \right\|,$$

где $p_1 = \dfrac{k_1}{n+1}$, $p_2 = \dfrac{k_2}{n+1}$.

8.18. Пусть $(x_{(1)}, \ldots, x_{(n)})$ — порядковые статистики выборки из совокупности, имеющей непрерывную к. ф. р. $F(x)$. Показать, что среднее значение размаха $R = (x_{(n)} - x_{(1)})$ равно

$$\mathscr{E}(R) = \int_{-\infty}^{+\infty}\{1 - F^n(x) - [1 - F(x)]^n\}\,dx.$$

Этот результат получен Типпетом (1925).

8.19. *(Продолжение.)* Показать, что к. ф. р. R равна

$$n\int_{-\infty}^{+\infty}\{F(x+R)-F(x)\}^{n-1}\,dF(x).$$

8.20. Пусть $x_{(1)}$ и $x_{(n)}$ — наименьшие и наибольшие порядковые статистики выборки объема n из равномерного распределения $R(\mu,\ \omega)$. И пусть $r=\left[\frac{1}{2}\ (x_{(n)}+x_{(1)})-\mu\ \right]/(x_{(n)}-x_{(1)})$. Показать, что элемент вероятности для r равен

$$g(r)\,dr=(n-1)\,(1+2\,|\,r\,|\,)^{-n}\,dr$$

на интервале $(-\infty,\ +\infty)$ (Карлтон (1946)).

8.21. Пусть $x_{(1)}$, $x_{(n)}$ — наименьшая и наибольшая порядковые статистики выборки объема n из непрерывной к. ф. р. $F(x)$. Показать, что случайная величина $2n\sqrt{F(x_{(1)})\,(1-F(x_{(n)}))}$ имеет предельное распределение при $n\to\infty$ со средним значением $\frac{1}{2}\pi$ и дисперсией $4-\frac{\pi^2}{4}$ (Элфвинг (1947)).

8.22. Пусть $(x_{(1)},\ldots,x_{(n)})$ — порядковые статистики выборки из совокупности, имеющей п. ф. р. $f(x)$, где

$$f(x)=\begin{cases}\frac{1}{\sigma}e^{-(x-\mu)/\sigma}, & x>\mu,\\ 0, & x\leqslant\mu,\end{cases}$$

и пусть

$$w=\frac{1}{n-1}\sum_{\xi=2}^{n}(x_{(\xi)}-x_{(1)}).$$

Показать, что п. ф. р. случайной величины t

$$t=(x_{(1)}-\mu)/w$$

равна

$$g(t)=\begin{cases}n\,[1+nt/(n-1)\,]^{-n}, & t>0,\\ 0, & t<0\end{cases}$$

(Гаттмен (1960)).

8.23. Пусть $(x_1,\ldots,x_n)$ — n-мерная случайная величина, имеющая п. ф. р. $f(x_1,\ldots,x_n)$, симметричную относительно $x_1,\ldots,x_n$. Показать, что порядковые статистики $(x_{(1)},\ldots,x_{(n)})$ из $(x_1,\ldots,x_n)$ имеют ф. п. в. $n!\,f(x_{(1)},\ldots,x_{(n)})$ в области, где $-\infty<x_{(1)}<\ldots<x_{(n)}<+\infty$, и 0 вне ее.

8.24. Пусть $(x_1,\ldots,x_n)$ — выборка из равномерного распределения $R\left(\frac{1}{2},\ 1\right)$ и пусть

$$y=(x_1\ldots x_n)^{\frac{1}{n}}.$$

Показать что ф. п. в. y равна

$$\frac{n^n y^{n-1}}{(n-1)\,!}\,(-\log y)^{n-1}$$

на $(0,\ 1)$ и 0 вне $(0,\ 1)$.

8.25. Пусть имеется m независимых выборок объема n из равномерного распределения $R\left(\frac{1}{2},\ 1\right)$ и пусть $u=\prod_{i=1}^{m}z_i$, где $z_1,\ldots,z_m$ — наибольшие порядковые статистики в m выборках соответственно. Показать, что ф. п. в. u равна

$$\frac{n^m}{(m-1)!}\,u^{n-1}\,(-\log u)^{m-1}$$

на $(0, 1)$ и 0 вне $(0, 1)$. Результат получен Райдером (1955).

8.26. Пусть $(x_1, \ldots, x_n)$ — выборка из гамма-распределения $G(\mu)$ и пусть $h(x_1, \ldots, x_n)$ — случайная величина такая, что $h(x_1, \ldots, x_n) = h(cx_1, \ldots, cx_n)$ для всех $c > 0$. Показать, что $(x_1 + \ldots + x_n)$ и $h(x_1, \ldots, x_n)$ независимы (Питмэн (1937a)).

8.27. Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности, имеющей ф. п. в. $f(x)$, и пусть отношение $(c_1x_1 + \ldots + c_nx_n) / (x_1 + x_n)$ и сумма $(x_1 + \ldots + x_n)$ независимы для любого набора постоянных $c_1, \ldots, c_n$, не равных 0 одновременно. Показать, что $f(x)$ есть ф. п. в. для гамма-распределения (Лага (1954)).

8.28. Пусть $(x_{(1)}, \ldots, x_{(n)})$ — порядковые статистики выборки объема n из совокупности, имеющей непрерывную к. ф. р. с конечным средним μ и дисперсией σ^2. Показать, используя неравенство Шварца, что

$$\mathscr{E}(x_{(n)}) \leqslant \mu + \frac{(n-1)\,\sigma}{\sqrt{2n-1}}.$$

Результат, незначительно отличающийся от результата Хартли и Дэвида (1954).

8.29. Пусть $(x_{(1)}, \ldots, x_{(n)})$ — порядковые статистики выборки объема n из непрерывной функции распределения $F(x)$. Показать, что случайные переменные

$$y_i = \left[\frac{F(x_{(i)})}{F(x_{(i+1)})}\right]^i, \qquad i = 1, \ldots, n-1,$$

независимы и распределены согласно $R\left(\frac{1}{2}, 1\right)$ (Реньи (1953)).

8.30. Пусть $(x_1, \ldots, x_n)$ — выборка из равномерного распределения $R\left(\frac{1}{2}\,\omega,\ \omega + \delta\right)$, где $\omega > 0$ и $0 < \frac{1}{2}\,\delta < \omega$. Пусть I_ξ — случайный интервал $\left(x_i - \frac{1}{2}\,\delta,\ x_i + \frac{1}{2}\,\delta\right)$ и пусть E есть событие $\bigcup\limits_{\xi=1}^{n} I_\xi$ и I — интервал $(0, \omega)$. Показать, что

$$\mathscr{E}(E \cap I) = \omega\left[1 - \left(1 - \frac{1}{\omega + \delta}\right)^n\right]$$

(Роббинс (1944a)).

8.31. Пусть n точек взяты случайно на отрезке длины L. Показать, что вероятность события, состоящего в том, что не найдется двух точек, лежащих на расстоянии, меньшем d, равна

$$(L - (n-1)\,d)^n / L^n,$$

если $0 < d < L / (n-1)$ (Парзен (1960)).

8.32. Пусть $(x_1, \ldots, x_n)$ — выборка объема n из совокупности, имеющей среднее значение μ и дисперсию σ^2, и пусть d — выборочное среднее отклонение относительно μ, определяемое согласно

$$d = \frac{1}{n}\sum_{\xi=1}^{n} |x_\xi - \mu|.$$

Показать, что

$$\mathscr{E}(d) = \sqrt{2/\pi}\,\sigma, \qquad \sigma^2(d) = \frac{\sigma^2}{n}\left(1 - \frac{2}{\pi}\right).$$

8.33. Пусть x — случайная величина с ф. п. в. $f(x)$, имеющая среднее значение μ и дисперсию σ^2. Пусть $\varphi(t)$ — характеристическая функция x, а $\bar{x}$ и s^2 — выборочные среднее и дисперсия выборки объема n из $f(x)$. Показать, что если $\bar{x}$ и s^2 независимы для любого n, то $\varphi(t)$ удовлетворяет дифференциальному уравнению

$$\varphi\frac{d^2\varphi}{dt^2}-\left(\frac{d\varphi}{dt}\right)^2+\sigma^2\varphi^2=0.$$

Решением этого уравнения при граничных условиях $\varphi(0)=1$ и $\varphi'(0)=i\mu$ является

$$\varphi(t)=e^{i\mu t-\frac{1}{2}\sigma^2t^2},$$

и, таким образом, $f(x)$ является ф. п. в. для нормального распределения $N(\mu,\sigma^2)$ (Лукач (1942)).

8.34. Пусть $x_{(k)}$ — порядковая статистика k-го порядка из выборки объема m с непрерывной к. ф. р. $F(x)$. И пусть задана вторая независимая выборка объема n с той же самой к. ф. р. $F(x)$. Обозначим через y случайную величину, равную числу x во второй выборке, не превышающему $x_{(k)}$. Показать, что ф. в. этой случайной величины y есть

$$\frac{\binom{m}{k}\binom{n}{y}}{\binom{m+n}{k+y}}\,\frac{k}{k+y},\qquad y=0,\,1,\,\ldots,\,n,$$

а также что r-й факториальный момент y равен

$$\mathscr{E}(y^{[r]})=\frac{m!\;n!\;(k+r-1)!}{(k-1)!\,(m+r)!\,(n-r)!},\qquad r\leqslant n$$

(Эпстейн (1954)).

8.35. (*Продолжение.*) Пусть вторая выборка составлена из элементов, выбираемых по одному до тех пор, пока y элементов выборки меньше, чем $x_{(k)}$. Показать, что объем n этой второй выборки есть случайная величина, имеющая ф. в.

$$\frac{\binom{n-1}{y-1}\binom{m-1}{k-1}}{\binom{m+n-1}{y+k-1}}\,\frac{m}{(m+n)},\qquad n=y,\;y+1,\ldots,$$

а также что

$$\mathscr{E}\,[(n-y)^{[r]}]=\frac{(y+r-1)!\;(k-r-1)!\;(m-k+r)!}{(y-1)!\;(k-1)!\;(m-k)!},\qquad r\leqslant k-1$$

(Уилкс (1959b)).

8.36. Пусть конечная совокупность π_N состоит из N предметов, занумерованных через $1,2,\ldots,N$. И пусть из этой конечной совокупности π_N взята выборка объема n без возвращения. Обозначим через x наибольшее число (из $1,2,\ldots,N$), оказавшееся среди номеров предметов в этой выборке Показать, что ф. в. случайной величины x есть

$$\binom{x-1}{n-1}\Big/\binom{N}{n}\qquad x=n,\ldots,N,$$

и что

$$\mathscr{E}(x)=\frac{n}{n+1}(N+1),\qquad \sigma^2(x)=\frac{n(N-n)(N+1)}{(n+1)^2(n+2)}.$$

8.37. (*Продолжение.*) Пусть из конечной совокупности π_N взята без возвращения выборка объема $2n+1$. И пусть y — медиана этой выборки. Показать, что ф. в. y равна

$$\binom{y-1}{n}\binom{N-y}{n}\Big/\binom{N}{2n+1}, \qquad y=n+1, \ldots, N-n,$$

а также что

$$\mathscr{E}(y)=\frac{N+1}{2}, \qquad \sigma^2(y)=\frac{(N-2n-1)(N+1)}{8n+12}.$$

8.38. (*Продолжение.*) Пусть s и t — порядковые статистики порядка k_1 и k_2 соответственно ($k_1<k_2$) из выборки объема n, взятой из π_N. Показать, что ф. в. (s, t) есть

$$\binom{s-1}{k_1-1}\binom{t-s-1}{k_2-k_1-1}\binom{N-t}{n-k_2}\Big/\binom{N}{n}.$$

Показать также, что

$$\mathscr{E}(s)=k_1\left(\frac{N+1}{n+1}\right),$$

$$\sigma^2(s)=\frac{k_1(N-n)(N+1)(n-k_1+1)}{(n+1)^2(n+2)},$$

$$\operatorname{cov}(s, t)=\frac{k_1(n-k_2+1)(N-n)(n+1)}{(n+1)^2(n+2)}.$$

8.39. Пусть $p(x)$ и $q(x)$ — два дискретных распределения, имеющих одинаковые точки сосредоточения масс $x_1, \ldots, x_k$. И пусть O_m и O_n — независимые выборки объема m и n из $p(x)$ и $q(x)$ соответственно. Пусть $(m_1, \ldots, m_k)$ — число компонент O_m, которые оказываются в точках $x_1, \ldots, x_k$, при этом $m_1+\ldots+m_k=m$. Аналогично определяется $(n_1, \ldots, n_k)$. Если $r_{m,n}$ — коэффициент корреляции между $(\sqrt{m_1/m}, \ldots, \sqrt{m_k/m})$ и $(\sqrt{n_1/n}, \ldots, \sqrt{n_k/n})$, то показать, что при $p(x)\equiv q(x)$ и для произвольного $\lambda>0$

$$P\left(r_{m,n}>1-\frac{\lambda}{2}\right)\geqslant 1-\frac{k-1}{\lambda}\left(\frac{1}{\sqrt{m}}+\frac{1}{\sqrt{n}}\right)^2.$$

Этот результат принадлежит Матусита (1957).

8.40. (*Продолжение.*) Если $\sum\limits_{i=1}^{k}\left(\sqrt{p(x_i)}-\sqrt{q(x_i)}\right)^2\geqslant\delta^2$, где $\delta>\lambda>0$, то показать, что

$$P\left(r_{m,n}<1-\frac{\lambda}{2}\right)\geqslant 1-\frac{k-1}{(\delta-\lambda)^2}\left(\frac{1}{\sqrt{m}}+\frac{1}{\sqrt{n}}\right)^2.$$

Этот результат также принадлежит Матусита (1957).

8.41. Показать, что к. ф. р. величины z для распределения, описанного в § 8.3 (а), есть

$$F_n(z)=\frac{1}{n!}\left[z^n-\binom{n}{1}(z-1)^n+\binom{n}{2}(z-2)^n-\ldots+(-1)^k\binom{n}{k}(z-k)^n\right],$$

$$k<z\leqslant k+1, \quad k=0, 1, \ldots, n-1,$$

и

$$F_n(z)=\begin{cases}0 & \text{для } z\leqslant 0,\\ 1 & \text{для } z>n.\end{cases}$$

8.42. Пусть $(x_{(1)}, \dots, x_{(n)})$ — порядковые статистики выборки объема n из равномерного распределения $R\left(\frac{1}{2}, 1\right)$. И пусть $(u_1, \dots, u_n)$ — n-мерная случайная величина, обозначающая сегменты $(x_{(1)}, x_{(2)} - x_{(1)}, \dots, x_{(n)} - x_{(n-1)})$, а $v = \max(u_1, \dots, u_n)$.

Показать, что к. ф. р. случайной величины v есть

$$F_n(v) = \left[1 - \binom{n}{1}(1-v)^n + \binom{n}{2}(1-2v)^n - \dots + (-1)^k \binom{n}{k}(1-kv)^n\right],$$

$$\frac{1}{k+1} < u \leqslant \frac{1}{k}, \qquad k = n-1,\ n-2, \dots, 1$$

и

$$F_n(v) = \begin{cases} 0 & \text{для } u \leqslant 0, \\ 1 & \text{для } u > 1. \end{cases}$$

(Результат предыдущей задачи полезен для решения данной задачи.)

8.43. (*Продолжение.*) ***Распределение наибольшего сегмента, порожденного n случайными точками в интервале*** (0,1). Порядковые статистики $(x_{(1)}, \dots, x_{(n)})$ делят интервал на $(n+1)$ непересекающихся сегментов $u_1, \dots, u_{n+1}$, при этом, конечно, $u_1 + \dots + u_{n+1} = 1$. Пусть $\omega = \max(u_1, \dots, u_{n+1})$. Показать, что к. ф. р. ω есть

$$F_n(\omega) = \left[1 - \binom{n+1}{1}(1-\omega)^n + \binom{n+1}{2}(1-2\omega)^n - \dots \right.$$
$$\left. \dots + (-1)^k \binom{n+1}{k}(1-k\omega)^n\right],$$

$$\frac{1}{k+1} < \omega \leqslant \frac{1}{k}, \qquad k = n,\ n-1, \dots, 1$$

и

$$F_n(\omega) = \begin{cases} 0 & \text{для } \omega \leqslant \dfrac{1}{n+1}, \\ 1 & \text{для } \omega > 1. \end{cases}$$

8.44. Пусть $(x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)})$ — порядковые статистики выборки объема 4 из $N(0, 1)$ и $y = \frac{1}{2}\left(x_{(3)} + x_{(4)}\right) - \frac{1}{2}\left(x_{(1)} + x_{(2)}\right)$. Показать, что к. ф. р. $G(y)$ случайной величины y есть

$$G(y) = \begin{cases} 8\left[\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^y e^{-\frac{1}{2}x^2}\,dx\right]^3, & y > 0, \\ 0 & ,\ y \geqslant 0 \end{cases}$$

(Уолш (1946)).

8.45. Пусть $(x_{(1)}, x_{(2)})$ — порядковые статистики выборки объема 2 из $N(\mu, \sigma^2)$. Показать, что средние значения этих порядковых статистик суть

$$(\mu - \sigma/\sqrt{\pi},\ \mu + \sigma/\sqrt{\pi}).$$

ГЛАВА 9

АСИМПТОТИЧЕСКАЯ ТЕОРИЯ ВЫБОРОЧНОГО МЕТОДА ДЛЯ БОЛЬШИХ ВЫБОРОК

Теория выборочного метода, изложенная в главе 8, была основана на рассмотрении конечных выборок объема n. Сейчас мы будем выяснять, какие результаты можно получить относительно выборочных распределений, если устремить n к бесконечности. Пусть $(x_1, x_2, \ldots)$ — стохастический процесс, элементы (компоненты) которого взаимно независимы. Любое множество n таких элементов есть выборка объема n из совокупности, имеющей функцию распределения $F(x)$. Этот стохастический процесс иногда называют *простым случайным выбором из бесконечной совокупности или случайным выбором из вероятностного распределения.* В этой главе мы рассмотрим некоторые предельные теоремы и результаты, которые полезны при аппроксимации распределений различных функций от выборок в случае больших выборок.

9.1. Сходимость по вероятности выборочного среднего значения

Один из простейших результатов относительно выборочного среднего значения дается следующей теоремой Хинчина (1929).

9.1.1. *Если $(x_1, \ldots, x_n)$ — выборка из совокупности с к. ф. р. $F(x)$, для которой $\mathscr{E}(x)$ имеет конечное значение μ, то выборочное среднее значение $\bar{x}$ сходится по вероятности к μ при $n \to \infty$.*

Для доказательства заметим сначала, что если среднее значение μ существует, то, как следует из (5.1.10), характеристическая функция $\varphi(x)$, соответствующая $F(x)$, может быть записана в виде

$$\varphi(t) = 1 + i\mu t + o(t). \tag{9.1.1}$$

Используя тот факт, что на основании **8.3.2** характеристическая функция $\bar{x}$ равна $\left[\varphi\left(\frac{t}{n}\right)\right]^n$, находим

$$\left[\varphi\left(\frac{t}{n}\right)\right]^n = \left[1 + \frac{i\mu t}{n} + o\left(\frac{t}{n}\right)\right]^n, \tag{9.1.2}$$

где $n \cdot o\left(\frac{t}{n}\right)$ стремится к нулю при $n \to \infty$ для всех t. Переходя к пределу при $n \to \infty$, получаем

$$\lim_{n\to\infty}\left[\varphi\left(\frac{t}{n}\right)\right]^n = e^{i\mu t}, \tag{9.1.3}$$

что есть не что иное, как характеристическая функция вырожденного распределения $\varepsilon(x-\mu)$, определенного в (2.3.4). Применяя **5.4.1**, получаем тогда, что при $n\to\infty$ распределение $\bar{x}$ сходится к $\varepsilon(x-\mu)$, а это эквивалентно утверждению о том, что $\bar{x}$ сходится по вероятности к μ при $n\to\infty$, что и завершает доказательство **9.1.1.**

Если в теореме **9.1.1** предположить, что случайная величина x обладает как конечным средним μ, так и дисперсией σ^2, то можно довольно просто доказать, используя неравенство Чебышева (3.3.5), что $\bar{x}$ сходится по вероятности к μ. Как мы знаем из **8.2.1**, $\mu(\bar{x})=\mu$ и $\sigma^2_{\bar{x}}=\frac{\sigma^2}{n}$. Применяя (3.3.5), мы можем сказать, что для любого $\varepsilon>0$ справедливо соотношение

$$P(|\bar{x}-\mu|\geqslant\varepsilon)=P\left(|\bar{x}-\mu|\geqslant\left(\frac{\varepsilon}{\sigma_{\bar{x}}}\right)\cdot\sigma_{\bar{x}}\right)\leqslant\frac{\sigma^2_{\bar{x}}}{\varepsilon^2}=\frac{\sigma^2}{n\varepsilon^2}. \tag{9.1.4}$$

Огсюда ясно, что

$$\lim_{n\to\infty} P(|\bar{x}-\mu|\geqslant\varepsilon)=0, \tag{9.1.5}$$

т. е. $\bar{x}$ сходится по вероятности к μ.

Обобщение **9.1.1** на k-мерный случай очевидно. Его формулировку и доказательство мы оставляем читателю.

Теорему **9.1.1** иногда называют *ослабленным законом больших чисел*. Ее основное утверждение состоит в том, что при $n\to\infty$ распределение $\bar{x}$ сосредоточивается в сколь угодно малой окрестности точки $\bar{x}=\mu$. Условия, которые гарантируют такое поведение распределения $\bar{x}$, — существование (конечность) среднего значения.

Распределение Коши. Примером распределения, которое на первый взгляд имеет ожидаемые свойства, а на самом деле для его выборочного среднего значения не имеет места обсуждавшееся выше свойство сходимости по вероятности, служит *распределение Коши* (1853) с плотностью функции распределения

$$f(x)=\frac{\theta_2}{\pi\left[\theta_2^2+(x-\theta_1)^2\right]}. \tag{9.1.6}$$

Область изменения x — вещественная прямая $(-\infty,+\infty)$, а θ_1, θ_2 — вещественные числа и $\theta_2>0$.

Можно показать, что характеристическая функция $\varphi(t)$, соответствующая (9.1.6), равна

$$\varphi(t)=e^{\theta_1 it-\theta_2|t|}. \tag{9.1.7}$$

Следовательно, характеристическая функция $\bar{x}$ для выборки объема n —

$$\left[\varphi\left(\frac{t}{n}\right)\right]^n=\left[e^{\theta_1 it/n-\theta_2|t|/n}\right]^n=\varphi(t). \tag{9.1.8}$$

Другими словами, $\bar{x}$ имеет ту же самую характеристическую функцию, что и x. А это означает, что выборочное распределение $\bar{x}$ при произвольном n такое же, как и распределение x в совокупности. Читатель может заметить, что $f(x)$ — симметрическая функция относительно $x=\theta_1$, однако θ_1 *не есть среднее значение* распределения. В этом случае среднего значения не существует. Не существуют также и моменты более высоких порядков. Величина θ_1 имеет смысл медианы и центра симметрии распределения, а θ_2 — межквартильного расстояния, т. е. расстояния между $\underline{x}_{0,25}$ и $\underline{x}_{0,75}$.

9.2. Предельное распределение выборочных сумм и средних значений

(а) Одномерный случай. Если у распределения совокупности существует как среднее значение, так и дисперсия, то мы можем многое сказать о характере поведения распределения $\bar{x}$ при $n \to \infty$. Фундаментальной теоремой в этой области является теорема Линдеберга (1922), которая может быть сформулирована следующим образом:

9.2.1. *Если z и $\bar{x}$ — соответственно сумма и среднее значение выборки объема n из распределения с конечной дисперсией σ^2 и средним значением μ, то*

$$\lim_{n\to\infty} P\left[\frac{z-n\mu}{\sqrt{n}\,\sigma}<y\right]=\lim_{n\to\infty} P\left[\frac{\sqrt{n}(\bar{x}-\mu)}{\sigma}<y\right]=\frac{1}{\sqrt{2\pi}}\int\limits_{\infty}^{y} e^{-\frac{1}{2}u^2}\,du. \tag{9.2.1}$$

Если (9.2.1) имеет место, то принято говорить, что *z распределено асимптотически нормально* $N(n\mu, \sigma^2)$ для больших n (или $\bar{x}$ распределено асимптотически нормально $N\left(\mu, \frac{\sigma^2}{n}\right)$ для больших n).

Случайную величину $(z-n\mu)/\sqrt{n}\sigma$ можно записать в виде $\sqrt{n}(\bar{x}-\mu)/\sigma$. Будем пользоваться этой последней записью. Пусть $\varphi_n(t)$ — характеристическая функция $(z-n\mu)/\sqrt{n}\sigma$. Мы имеем

$$\varphi_n(t)=\mathscr{E}\left[e^{it(z-n\mu)/\sqrt{n}\,\sigma}\right]=\mathscr{E}\left[\exp\left(it\sum_{\xi=1}^{n}(x_\xi-\mu)/\sqrt{n}\sigma\right)\right]=[\varphi(t)]^n, \tag{9.2.2}$$

где $\varphi(t)$ — характеристическая функция $(x-\mu)/\sqrt{n}\sigma$. Из (5.1.10) получаем

$$\varphi(t)=1+i\mathscr{E}\left(\frac{x-\mu}{\sqrt{n}\sigma}\right)t-\frac{1}{2}\mathscr{E}\left(\frac{x-\mu}{\sqrt{n}\sigma}\right)^2 t^2+o\left(\frac{t^2}{n}\right), \tag{9.2.3}$$

где $n\cdot o\left(\frac{t^2}{n}\right)\to 0$ при $n\to\infty$ для любого $t\neq 0$. Если $t=0$, то $o\left(\frac{t^2}{n}\right)=0$.

Следовательно, для любого t

$$\lim_{n\to\infty}\varphi_n(t)=\lim_{n\to\infty}\left[1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right]^n=e^{-t^2/2}. \tag{9.2.4}$$

Но, как мы знаем из **7.2.1**, $e^{-t^2/2}$ — характеристическая функция нормального распределения $N(0, 1)$. Поэтому из **5.4.1** следует, что функция распределения $(z-n\mu)/\sqrt{n}\sigma$ при $n\to\infty$ сходится к функции распределения $N(0, 1)$, что эквивалентно утверждению, сформулированному в **9.2.1**. Теорема **9.2.1** доказана.

Важное следствие **9.2.1** — *теорема Муавра—Лапласа.*

9.2.1a. *Если z — случайная величина, имеющая биномиальное распределение* Bi (n, p), *то z асимптотически распределена согласно $N(np, npq)$.*

Из частного случая теоремы **8.3.3a** при $m=1$ следует, что если выборка объема n взята из биномиального распределения Bi $(1, p)$, то выборочная сумма z имеет биномиальное распределение Bi (n, p). Применяя тогда **9.2.1**, мы получаем **9.2.1a**. Результат **9.2.1a** был сформулирован Муавром (1718), однако был точно доказан только век спустя Лапласом (1814), Гаусс (1809b) пришел к теореме **9.2.1** из эвристических соображений, связанных с теорией ошибок.

(b) Центральная предельная теорема. Теорема **9.2.1** представляет собой частный случай гораздо более общего результата — *центральной предельной теоремы*, которая утверждает, что при определенных условиях

$$\lim_{n\to\infty} P\left[\frac{\sum\limits_{\xi=1}^{n}(x_\xi-\mu_\xi)}{\sqrt{\sum\limits_{\xi=1}^{n}\sigma_\xi^2}}<y\right]=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{y}e^{-\frac{1}{2}u^2}\,du, \tag{9.2.5}$$

где $(x_1, x_2, \ldots)$ — последовательность независимых случайных величин со средними значениями $(\mu_1, \mu_2, \ldots)$ и дисперсиями $(\sigma_1^2, \sigma_2^2, \ldots)$ соответственно. Различные варианты условий, при которых имеет место утверждение теоремы, были получены Чебышевым (1890), Феллером (1935), Леви (1935), Линдебергом (1922), Ляпуновым (1900, 1901), Марковым (1900) и другими. Исчерпывающее и глубокое изложение центральной предельной теоремы и связанных с ней проблем было дано Гнеденко и Колмогоровым (1954). Современный вариант центральной предельной теоремы в общем случае может быть сформулирован следующим образом.

9.2.2. *Пусть $(x_1, x_2, \ldots)$ — последовательность случайных величин, имеющих функции распределения $(F_1(x), F_2(x), \ldots)$, средние значения $(\mu_1, \mu_2, \ldots)$ и дисперсии $(\sigma_1^2, \sigma_2^2, \ldots)$. Пусть $z_n=x_1+\ldots+x_n$, $\zeta_n=\mu_1+\ldots+\mu_n$, $\tau_n^2=\sigma_1^2+\ldots+\sigma_n^2$ и $\tau_n^{*2}=\sigma_1^{*2}+\ldots+\sigma_n^{*2}$, где при произвольном $\varepsilon>0$*

$$\sigma_\xi^{*2}=\int\limits_{\mu_\xi-\varepsilon\tau_n}^{\mu_\xi+\varepsilon\tau_n}(x-\mu_\xi)^2\,dF_\xi(x), \qquad \xi=1, \ldots, n. \tag{9.2.6}$$

Необходимые и достаточные условия для того, чтобы

$$\lim_{n\to\infty} P\left(\frac{z_n-\zeta_n}{\tau_n}<y\right)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{y}e^{-\frac{1}{2}u^2}\,du, \tag{9.2.7}$$

состоят в выполнении соотношений

$$\lim_{n\to\infty} \frac{\tau_n^{*2}}{\tau_n^2} = 1 \tag{9.2.8}$$

и $\lim\limits_{n\to\infty} \tau_n^2 = \infty$ *для любого* $\varepsilon > 0$.

Доказательство теоремы **9.2.2** достаточно длинное и здесь будет опущено. Достаточность условия (9.2.8) была получена Линдебергом (1922), а необходимость его — Феллером (1935).

Читатель без больших затруднений сможет вывести, что если $(x_1, x_2, \ldots)$ — последовательность независимых случайных величин, имеющих все одно и то же распределение с дисперсией σ^2, то условия (9.2.8) будут выполнены.

(c) k-мерный случай. Аналог **9.2.1** в k-мерном случае может быть сформулирован следующим образом:

9.2.3. Допустим, что $(x_{1\xi}, \ldots, x_{k\xi};\ \xi = 1, \ldots, n)$ — *выборка объема n из k-мерного распределения, имеющего конечные средние значения μ_i, $i = 1, \ldots, k$, и положительно определенную ковариационную матрицу $\|\sigma_{ij}\|$, $i, j = 1, \ldots, k$. Пусть $(z_1, \ldots, z_k)$ — выборочные суммы и $(\bar{x}_1, \ldots, \bar{x}_k)$ — выборочные средние значения, определенные в* § 8.1. *Тогда $(z_1, \ldots, z_k)$* и *$(\bar{x}_1, \ldots, \bar{x}_k)$ в качестве своих асимптотических распределений имеют k-мерные распределения $N(\{\eta\mu_i\}, \|n\sigma_{ij}\|)$ и $N\left(\{\mu_i\}, \left\|\frac{\sigma_{ij}}{n}\right\|\right)$ соответственно, т.е.*

$$\lim_{n\to\infty} P\left(\frac{(z_i - n\mu_i)}{\sqrt{n}} < y_i, \quad i = 1, \ldots, k\right) =$$
$$= \lim_{n\to\infty} P\left((\bar{x}_i - \mu_i)\sqrt{n} < y_i, \quad i = 1, \ldots, k\right) =$$
$$= \frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{\frac{1}{2}k}} \int\limits_{-\infty}^{y_k} \cdots \int\limits_{-\infty}^{y_1} \exp\left(-\frac{1}{2}\sum_{i,j=1}^{k} \sigma^{ij} u_i u_j\right) du_1 \ldots du_k. \tag{9.2.9}$$

Метод доказательства **9.2.3** является естественным обобщением метода доказательства **9.2.1**. Главная идея состоит в том, чтобы показать, что для характеристической функции $\varphi_n(t_1, \ldots, t_k)$ k-мерной случайной величины справедливо соотношение

$$\lim_{n\to\infty} \varphi_n(t_1, \ldots, t_k) = \exp\left(-\frac{1}{2}\sum_{i,j=1}^{k} \sigma_{ij} t_i t_j\right). \tag{9.2.10}$$

Это соотношение с помощью k-мерного аналога **5.4.1** приводит к (9.2.9). Детали доказательства (9.2.10) достаточно ясны, и мы оставляем их читателю.

Одним из наиболее важных приложений теоремы **9.2.3** является обобщение теоремы Муавра **9.2.1a** на k-мерный случай:

9.2.3а. *Если* $(x_{1\xi},\ldots,x_{k\xi};\ \xi=1,\ldots,n)$ — *выборка объема* n *из мультиномиального распределения* $M(1;\ p_1,\ldots,p_k)$, *то выборочные суммы* $(z_1,\ldots,z_k)$ *имеют при больших* n *в качестве асимптотических распределений* $N(\{np_i\},\ \|n(p_i\delta_{ij}-p_ip_j)\|)$, *где* δ_{ij} — *символ Кронекера.*

Доказательство этого утверждения как следствия теоремы **9.2.3** не представляет труда и сводится к проверке того, что для плотности функции распределения мультиномиального распределения $M(1;\ p_1,\ldots$ $\ldots,p_k)$ (см. (6.3.3))

$$\mathscr{E}(x_i)=p_i,\quad \sigma^2(x_i,\ x_j)=\sigma_{ij}=p_i\delta_{ij}-p_ip_j. \tag{9.2.11}$$

9.3. Асимптотическое распределение функций от выборочных средних значений

(а) Общий случай. Часто бывают важными сведения относительно асимптотического распределения некоторых функций от выборочного среднего значения, назовем эти функции $g(\bar{x})$. Полезным при рассмотрении этих проблем является следующий результат:

9.3.1. *Допустим, что* $(x_1,\ldots,x_n)$ — *выборка из распределения с конечным средним значением* μ *и дисперсией* σ^2. *Пусть* $g(x)$ — *функция, которая имеет в некоторой окрестности точки* $x=\mu$ *первую производную такую, что* $g'(\mu)\neq 0$. *Тогда* $g(\bar{x})$ *асимптотически распределена при больших* n *в соответствии с* $N(g(\mu),$ $[\sigma g'(\mu)]^2/n)$.

Докажем этот результат. Пусть $V(\mu)$ — окрестность точки $x=\mu$, для которой $g'(x)$ существует для всех значений x из $V(\mu)$. Так как функция распределения совокупности, из которой взята выборка, имеет конечное среднее значение μ и конечную дисперсию σ^2, то, как следует из **4.6.1**, для любого $\varepsilon>0$ существует такое n_ε, что при $n>n_\varepsilon$

$$P(\bar{x}\in V(\mu)\ \text{для всех}\ n>n_\varepsilon)>1-\varepsilon. \tag{9.3.1}$$

Для любой точки $x=\bar{x}$ из $V(\mu)$ мы можем написать

$$g(\bar{x})=g(\mu)+g'(x^*)\cdot(\bar{x}-\mu), \tag{9.3.2}$$

где x^* — случайная величина такая, что $|x^*-\mu|\leqslant|\bar{x}-\mu|$. Соотношение (9.3.2) можно представить и в следующем виде:

$$\big(g(\bar{x})-g(\mu)\big)\sqrt{n}=g'(x^*)(\bar{x}-\mu)\sqrt{n}. \tag{9.3.3}$$

Пусть теперь случайная величина, стоящая слева в (9.3.3), обозначена через u_n, а в правой части — через v_n. Вероятность того, что (9.3.3) имеет место, превышает $1-\varepsilon$ для всех $n>n_\varepsilon$. Так как ε произвольно, то последовательности случайных величин $(u_1,\ u_2,\ldots)$ и $(v_1,\ v_2,\ldots)$ образуют стохастический процесс такой, что $(u_1,\ u_2,\ldots)$ и $(v_1,\ v_2,\ldots)$ сходятся друг к другу по распределению, если один из них сходится по распределению. Так как σ^2 конечно, то, как мы знаем из **9.2.1**, последовательность случайных величин $\sqrt{n}(\bar{x}-\mu)$, $n=1, 2, \ldots$, сходится по распределению к нормальному закону $N(0, \sigma^2)$. Поскольку $g'(x)$ существует во всех точках $V(\mu)$, то она непрерывна в $V(\mu)$ и, в частности, при $x=\mu$. Из **4.3.7** следует тогда, что $g'(x^*)$ сходится по вероятности к $g'(\mu)$.

Поэтому

$$\lim_{n\to\infty} P(v_n < w) = P(g'(\mu)\, s < w), \tag{9.3.4}$$

где s имеет распределение $N(0, \sigma^2)$, что приводит к тому, что

$$\lim_{n\to\infty} P(u_n < w) = P(t < w), \tag{9.3.5}$$

где t имеет в свою очередь распределение $N(0, (\sigma g'(\mu))^2)$. Но (9.3.5) эквивалентно утверждению, что асимптотическое распределение $g(\bar{x})$ при больших n есть как раз $N(g(\mu), [\sigma g'(\mu)]^2/n)$. Таким образом, теорема **9.3.1** доказана.

Обобщение **9.3.1** на k-мерный случай может быть сформулировано следующим образом:

9.3.1a. *Допустим, что* $(x_{1\xi}, \ldots, x_{k\xi}$, $\xi = 1, \ldots, n)$ — *выборка из k-мерного распределения с конечными средними значениями* $\{\mu_i\}$ *и положительно определенной ковариационной матрицей* $\|\sigma_{ij}\|$, $i, j = 1, \ldots, k$. *Пусть* $g(x_1, \ldots, x_k)$ — *функция, имеющая первые производные* $\frac{\partial g}{\partial x_i} = g_i$, $i = 1, \ldots, k$, *во всех точках некоторой окрестности точки* $(\mu_1, \ldots, \mu_k)$ *и* $g_i^0 = g_i(\mu_1, \ldots, \mu_k)$. *Тогда, если по крайней мере одна из величин* g_i^0 *не равна нулю, то* $g(\bar{x}_1, \ldots, \bar{x}_k)$ *имеет для больших n асимптотическое распределение*

$$N\left(g(\mu_1, \ldots, \mu_k),\ \frac{1}{n}\sum_{i,\,j=1}^{k} \sigma_{ij} g_i^0 g_j^0\right).$$

(b) Квадратичные функции выборочных средних значений и сумм. Сначала заметим, что соотношение (9.2.9) означает, что имеется последовательность независимых k-мерных случайных величин

$$\left(\frac{z_1 - n\mu_1}{\sqrt{n}}, \ldots, \frac{z_k - n\mu_k}{\sqrt{n}}\right), \qquad n = 1, 2, \ldots, \tag{9.3.6}$$

или, что эквивалентно, последовательность k-мерных независимых случайных величин

$$((\bar{x}_1 - \mu_1)\sqrt{n}, \ldots, (\bar{x}_k - \mu_k)\sqrt{n}), \qquad n = 1, 2, \ldots, \tag{9.3.7}$$

которые сходятся по распределению к k-мерному нормальному распределению $N(\{0\}, \|\sigma_{ij}\|)$. Удобно для дальнейшего обозначить к. ф. р. случайной величины (9.3.6) (или (9.3.7)) при фиксированном n через $F_n(y_1, \ldots, y_k)$, а к. ф. р. $N(\{0\}, \|\sigma_{ij}\|)$, вид которой определен (9.2.9), через $\Phi(y_1, \ldots, y_k)$.

В **7.8.2** было показано, что если $(x_1, \ldots, x_k)$ имеет распределение $N(\{\mu_i\}, \|\sigma_{ij}\|)$, то $\sum_{i,j=1}^{k} \sigma^{ij}(x_i - \mu_i)\cdot(x_j - \mu_j)$ имеет χ^2-распределение $C(k)$.

Рассмотрим квадратичную форму

$$Q_n = \sum_{i,j=1}^{k} \sigma^{ij}\left(\frac{z_i - n\mu_i}{\sqrt{n}}\right)\left(\frac{z_j - n\mu_j}{\sqrt{n}}\right) = n\sum_{i,j=1}^{k} \sigma^{ij}(\bar{x}_i - \mu_i)(\bar{x}_j - \mu_j). \tag{9.3.8}$$

Последовательность случайных величин Q_n, $n = 1, 2, \ldots$, сходится по распределению к χ^2-распределению $C(k)$ при $n \to \infty$. Действительно, на основании **4.3.6** следует, что так как стохастический процесс (9.3.6) (или (9.3.7)) сходится по распределению к k-мерному нормальному закону $N(\{0\}, \|\sigma_{ij}\|)$, то последовательность $Q_1, Q_2, \ldots$ сходится по распределению к χ^2-распределению $C(k)$.

Таким образом, приходим к следующему результату:

9.3.2. *Если* $(x_{1\xi}, \ldots, x_{k\xi}, \xi = 1, \ldots, n)$ — *выборка из k-мерного распределения с конечными средними значениями* μ_i, $i = 1, \ldots, k$, *и конечной положительно определенной ковариационной матрицей* $\|\sigma_{ij}\|$, *то* Q_n, *определенные согласно* (9.3.8), *имеют при больших n в качестве асимптотического распределения χ^2-распределение* C_k, т. е.

$$\lim_{n\to\infty} P(Q_n \leqslant y) = \frac{1}{2\Gamma\left(\frac{1}{2}k\right)}\int_0^y \left(\frac{u}{2}\right)^{\frac{1}{2}k-1} e^{-\frac{1}{2}u}\, du. \tag{9.3.9}$$

Необходимо заметить, что Q_n — функция от $\bar{x}_1, \ldots, \bar{x}_k$, для которой нарушено условие теоремы **9.3.1a** о том, что по крайней мере одна из ее первых производных, вычисленных в $(\mu_1, \ldots, \mu_k)$, не равна нулю. Однако Q_n при $n \to \infty$ имеет предельное распределение $C(k)$, которое обладает хорошими свойствами.

(c) Критерий согласия χ^2 Пирсона. Мы имеем важный частный случай **9.3.2**, когда k-мерная совокупность, из которой берется выборка, подчиняется мультиномиальному распределению $M(1; p_1, \ldots, p_k)$. Средние значения и ковариационная матрица при этом даются выражениями (9.2.11). Тогда, как можно показать,

$$\|\sigma^{ij}\| = \|p_i\delta_{ij} - p_ip_j\|^{-1} = \left\|\frac{\delta_{ij}}{p_i} + \frac{1}{p_{k+1}}\right\|. \tag{9.3.10}$$

Подставляя $\|\delta_{ij}/p_i + {}^1/p_{k+1}\|$ вместо σ^{ij} и p_i вместо μ_i в (9.3.8), находим, что Q_n могут быть записаны следующим образом:

$$Q_n = \sum_{i=1}^{k+1} \frac{(z_i - np_i)^2}{np_i}, \qquad \text{где } z_{k+1} = \sum_{\xi=1}^{n} x_{k+1\,\xi} = n - \sum_{i=1}^{k} z_i. \tag{9.3.11}$$

Вспоминая на основании **8.3.3b**, что если выборка объема n взята из мультиномиального распределения $M(1;\ p_1,\ \ldots,\ p_k)$, то выборочные суммы имеют распределение $M(n;\ p_1,\ \ldots,\ p_k)$, получаем следствие из **9.3.2**:

9.3.2a. *Если* $(z_1,\ \ldots,\ z_k)$ — *k-мерная случайная величина, имеющая мультиномиальное распределение* $M(n;\ p_1,\ \ldots,\ p_k)$, *то*

$$Q_n = \sum_{i=1}^{k+1} \frac{(z_i - np_i)^2}{np_i} \qquad (9.3.12)$$

асимптотически при $n \to \infty$ *имеет* χ^2*-распределение* $C(k)$.

Заметим, что (9.3.12) есть сумма квадратов разностей между выборочными «частотами» z_i и их средними значениями, взвешенных с помощью этих средних значений. Эта величина как раз и используется в хорошо известном χ^2-критерии согласия, впервые введенном К. Пирсоном (1900). Можно смотреть на эту величину как на показатель «коллективного отклонения» z_i от соответствующих средних значений. Вероятность Q_n для достаточно большой выборки можно вычислить из ее асимптотического распределения, т. е. из χ^2-распределения. Дальнейшее рассмотрение этой проблемы приводит нас к теории испытания статистических гипотез, которая рассматривается в главе 13.

9.4. Асимптотическое разложение распределения выборочной суммы

Теорема **9.2.1** устанавливает предельную форму распределения $(z - n\mu)/\sqrt{n}\,\sigma$ при $n \to \infty$. Одна из проблем, возникающих при этом, состоит в определении для достаточно больших n лучшей аппроксимации распределения $(z - n\mu)/\sqrt{n}\,\sigma$, чем при замене его распределением $N(0,\ 1)$. Мы будем сейчас исследовать эту проблему.

Допустим, что центральные моменты $\mu_1,\ \ldots,\ \mu_r$, $r > 2$, совокупности, имеющей к. ф. р. $F(x)$, существуют и конечны. Тогда, если $\varphi(t)$ есть характеристическая функция $(x - \mu)/\sqrt{n}\,\sigma$, то для нее имеем

$$\varphi(t) = 1 - \frac{t^2}{2n} + \sum_{j=3}^{r} \frac{(it)^j \alpha_j}{j!\,(\sqrt{n})^j} + o\left(\left(\frac{t}{\sqrt{n}}\right)^r\right), \qquad (9.4.1)$$

где $\alpha_j = \mu_j/\sigma^j$, а $n^{\frac{1}{2}r} \cdot o((t/\sqrt{n})^r)$ стремится к нулю при $n \to \infty$ для любого $t \neq 0$. Но характеристическая функция $(z - n\mu)/\sqrt{n}\,\sigma$, обозначим ее через $\varphi_n(t)$, дается тогда выражением

$$\varphi_n(t) = [\varphi(t)]^n. \qquad (9.4.2)$$

Беря логарифм, находим

$$\log \varphi_n(t) = -\frac{t^2}{2} + n\sum_{j=3}^{r} \frac{(\mathrm{i}t)^j (k_j^*)}{j!\,(\sqrt{n})^j} + n\cdot o\left(\left(\frac{t}{\sqrt{n}}\right)^r\right), \tag{9.4.3}$$

где k^*_j — семиинварианты распределения величин $(x-\mu)/\sigma$ в популяции. Поэтому

$$\varphi_n(t) = e^{-\frac{1}{2}t^2} \exp\left[n\sum_{j=3}^{r} \frac{(\mathrm{i}t)^j (k_j^*)}{j!\,(\sqrt{n})^j} + n\cdot o\left(\left(\frac{t}{\sqrt{n}}\right)^r\right)\right], \tag{9.4.4}$$

что можно записать в виде

$$\varphi_n(t) = e^{-\frac{1}{2}t^2}\left[1 + \sum_{j=1}^{r-2} \frac{u_j(\mathrm{i}t)}{(\sqrt{n})^j} + n\cdot o\left(\left(\frac{t}{\sqrt{n}}\right)^r\right)\right], \tag{9.4.5}$$

где $u_j(\mathrm{i}t)$ — полиномы степени $3j$ от $(\mathrm{i}t)$. Коэффициенты полиномов суть функции k^*_j, но они не зависят от n. Наименьшая степень $\mathrm{i}t$ в $u_j(\mathrm{i}t)$ равна $j+2$.

Положим

$$F_n(x) = P\left(\frac{(z-n\mu)}{\sqrt{n}\,\sigma} \leqslant x\right) \tag{9.4.6}$$

и выберем в (5.1.14) $x' = (x+y)/2$, $\delta = (x-y)/2$. Тогда, если x и y — точки непрерывности $F_n(x)$, то

$$F_n(x) - F_n(y) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{\sin\left(\frac{x-y}{2}\right)t}{t}\, e^{-\mathrm{i}t\left[\frac{1}{2}(x+y)\right]}\, \varphi_n(t)\,dt. \tag{9.4.7}$$

Подставляя выражение для $\varphi_n(t)$ из (9.4.5) и сделав ряд упрощений, получаем

$$F_n(x) - F_n(y) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{(e^{-\mathrm{i}tx} - e^{-\mathrm{i}ty})}{2(-\mathrm{i}t)}\,,$$

$$e^{-\frac{1}{2}t^2}\left[1 + \sum_{j=1}^{r-2} \frac{u_j(\mathrm{i}t)}{(\sqrt{n})^j} + n\cdot o\left(\left(\frac{t}{\sqrt{n}}\right)^r\right)\right]dt. \tag{9.4.8}$$

Нетрудно показать, что

$$\frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{(e^{-\mathrm{i}tx} - e^{-\mathrm{i}ty})}{2(-\mathrm{i}t)}\, e^{-\frac{1}{2}t^2}\,dt = \Phi(x) - \Phi(y), \tag{9.4.9}$$

где $\Phi(x)$ — функция распределения $N(0, 1)$. Все остальные члены в (9.4.7) имеют следующую форму, с точностью до постоянного множителя:

$$\frac{1}{\pi}\int_{-\infty}^{\infty}(-\mathrm{i}t)^j(e^{-\mathrm{i}tx}-e^{-\mathrm{i}ty})\,e^{-\frac{1}{2}t^2}\,dt,$$

что дает

$$\frac{1}{\pi}\cdot\frac{d^j}{dx^j}\int_{-\infty}^{+\infty}e^{-\mathrm{i}tx-\frac{1}{2}t^2}\,dt-\frac{1}{2\pi}\frac{d^j}{dy^j}\int_{-\infty}^{+\infty}e^{-\mathrm{i}ty-\frac{1}{2}t^2}\,dt=\Phi^{(j+1)}(x)-\Phi^{(j+1)}(y), \tag{9.4.10}$$

где

$$\Phi^{(j+1)}(x)=\frac{d^{j+1}}{dx^{j+1}}\Phi(x)=\frac{d^j}{dx^j}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right). \tag{9.4.11}$$

Таким образом, для (9.4.7) находим

$$F_n(x)-F_n(y)=\Phi(x)-\Phi(y)+\sum_{j=1}^{r-2}\frac{[u_j^*(x)-u_j^*(y)]}{\sqrt{n^j}}+o\left(\frac{1}{\sqrt{n^{r-2}}}\right), \tag{9.4.12}$$

где $u_j^*(x)$ и $u_j^*(y)$ — функции, которые получаются заменой $(\mathrm{i}t)^p$ в $u_j(\mathrm{i}t)$ на $\Phi^{p+1}(x)$ и $\Phi^{p+1}(y)$ соответственно.

Если мы в (9.4.12) положим $y\to-\infty$, то получим следующее асимптотическое разложение $F_n(x)$:

$$F_n(x)=\Phi(x)+\sum_{j=1}^{r-2}\frac{u_j^*(x)}{\sqrt{n^j}}+o\left(\frac{1}{\sqrt{n^{r-2}}}\right). \tag{9.4.13}$$

Если $F_n(x)$ имеет плотность функции распределения $f_n(x)$, то, беря производную от (9.4.13) по x, находим, что

$$f_n(x)=\Phi^{(1)}(x)+\sum_{j=1}^{r-2}\frac{u_j^{*\prime}(x)}{\sqrt{n^j}}+o\left(\frac{1}{\sqrt{n^{r-2}}}\right), \tag{9.4.14}$$

где $u_j^{*\prime}(x)$ — первая производная $u_j^*(x)$.

На самом деле, даже если $F_n(x)$ не имеет плотности (т. е. $F_n(x)$ может быть дискретной функцией распределения), (9.4.14) все равно будет хорошей аппроксимацией. В этом случае правая часть (9.4.14) дает представление такой плотности функции распределения, которая, если ее проинтегрировать от $-\infty$ до x, хорошо аппроксимирует $F_n(x)$.

Мы не будем выписывать наиболее общего вида функции $u_j^*(x)$. Однако интересно выписать выражение для правых частей (9.4.13) и (9.4.14) с точностью до членов порядка $n^{-\frac{3}{2}}$. Оказывается, что

$$F_n(x)=\Phi(x)-\frac{1}{\sqrt{n}}\left(\frac{\alpha_3}{3!}\right)\Phi^{(3)}(x)+\frac{1}{n}\left[\frac{1}{4!}(\alpha_4-3)\Phi^{(4)}(x)+\right.$$
$$\left.+\frac{10}{6!}\alpha_3^2\Phi^{(6)}(x)\right]-\frac{1}{n^{\frac{3}{2}}}\left[\frac{1}{5!}(\alpha_5-10\alpha_3)\Phi^{(5)}(x)+\frac{35}{7!}\alpha_3(\alpha_4-3)\Phi^{(7)}(x)+\right.$$
$$\left.+\frac{280}{9!}\alpha_3^3\Phi^{(9)}(x)\right]+o\left(\frac{1}{n^{3/2}}\right) \qquad (9.4.15)$$

и

$$f_n(x)=\Phi^{(1)}(x)-\frac{1}{\sqrt{n}}\left(\frac{\alpha_3}{3!}\right)\Phi^{(4)}(x)+\frac{1}{n}\left[\frac{1}{4!}(\alpha_4-3)\Phi^{(5)}(x)+\right.$$
$$\left.+\frac{10}{6!}\alpha_3^2\Phi^{(7)}(x)\right]-\frac{1}{n^{3/2}}\left[\frac{1}{5!}(\alpha_5-10\alpha_3)\Phi^{(6)}(x)+\frac{35}{7!}\alpha_3(\alpha_4-3)\Phi^{(8)}(x)+\right.$$
$$\left.+\frac{280}{9!}\alpha_3^3\Phi^{(10)}(x)\right]+o\left(\frac{1}{n^{3/2}}\right). \qquad (9.4.16)$$

Мы можем суммировать эти результаты, которые были впервые получены Эджвортом (1905), следующим образом:

9.4.1. *Если* $(x_1,\ldots,x_n)$ — *выборка из распределения, имеющего конечные моменты* $\mu_1,\ldots,\mu_r\,(r>2)$, *то функция распределения* $F_n(x)$ *величины* $(z-n\mu)/\sqrt{n}\sigma$ *может быть разложена согласно* (9.4.13). *С точностью до членов порядка* $n^{-3/2}$ $F_n(x)$ *дается выражением**) (9.4.15).

Величины $\alpha_3\,(=\mu_3/\sigma^3)$ и $\alpha_4-3\,(=\mu_4/\sigma^4-3)$, обычно обозначаемые через γ_1 и γ_2, часто называются *асимметрией и эксцессом* распределения, имеющего среднее значение μ, дисперсию σ^2 и третий и четвертый центральные моменты μ_3 и μ_4. Эти две постоянные играют большую роль для характеристики точности, с которой функция распределения $F_n(x)$ может быть аппроксимирована функцией распределения $\Phi(x)$ нормального распределения $N(0,\ 1)$. Из рассмотрения формулы (9.4.15) можно заметить, что в общем случае $F_n(x)$ аппроксимируется $\Phi(x)$ с точностью до членов порядка $\frac{1}{\sqrt{n}}$. Однако сформулированное ниже следствие **9.4.1** дает условия, при которых аппроксимация справедлива с точностью до членов более высокого порядка.

9.4.1а. *Если асимметрия распределения, из которого взята выборка* $(x_1,\ldots,x_n)$, *равна нулю, то* $F_n(x)$ *аппроксимируется с помощью* $|\Phi(x)|$ *до членов порядка* $\frac{1}{n}$. *Если равны нулю и асимметрия и*

*) На самом деле теорема 9.4.1 (и ее следствие 9.4.1а) имеет место при условии, что $\lim\limits_{|t|\to\infty}\sup|f(t)|<1$, где $f(t)$ — характеристическая функция. Это условие заведомо выполняется, если функция распределения F_n имеет абсолютно непрерывную компоненту. (*Прим. перев.*)

эксцесс, то $F_n(x)$ аппроксимируется с помощью $\Phi(x)$ с точностью до членов порядка $\frac{1}{\sqrt{n^3}}$.

Необходимо подчеркнуть, что пионером в исследовании порядка точности приближения распределения $\bar{x}$ в случае больших выборок нормальным распределением был Ляпунов (1901). Крамер (1937) показал, что остаточный член в (9.4.13) и (9.4.14) того же порядка, что и первый отброшенный член. Эссеен (1944) провел новые исследования точности подобных асимптотических разложений. Асимптотические разложения по степеням $\frac{1}{\sqrt{n}}$ для отличных от выборочного среднего значения статистик были исследованы Крамером (1937), Сюй (1945a, 1945b), Чжуном (1946) и другими. Обстоятельная статья относительно асимптотической аппроксимации распределений с довольно обширной библиографией была опубликована Уоллесом (1958).

9.5. Предельные распределения линейных функций в случае больших выборок из больших конечных совокупностей

Выше мы исследовали предельные распределения сумм и средних значений выборок из бесконечных популяций. Теперь рассмотрим предельные распределения линейных функций от элементов выборок из конечных совокупностей, когда объем выборки и объем совокупности бесконечно растут. Общая теорема, имеющая основное значение при исследовании этой проблемы, была получена Вальдом и Вольфовитцем (1944) и может быть сформулирована следующим образом:

9.5.1. *Пусть $(a_{N1}, \ldots, a_{NN})$ и $(x_{N1}, \ldots, x_{NN})$, $N = 1, 2, \ldots$, — два набора последовательностей вещественных чисел таких, что для $r = 3, 4, \ldots$ и больших N*

$$\frac{m_{r,N}(a)}{[m_{2,N}(a)]^{r/2}} = O(1) \tag{9.5.1}$$

и

$$\frac{m_{r,N}(x)}{[m_{2,N}(x)]^{r/2}} = O(1), \tag{9.5.2}$$

где

$$m_{r,N}(a) = \frac{1}{N}\sum_{i=1}^{N}(a_{Ni} - \bar{a}_N)^r, \qquad \bar{a}_N = \frac{1}{N}\sum_{i=1}^{N} a_{Ni}, \tag{9.5.3}$$

и аналогичным образом определены $m_{r,N}(x)$ и $\bar{x}_N$. Пусть для любого значения N $(x_1, \ldots, x_N)$ — векторная случайная величина, *выборочное пространство которой есть множество всех $N!$ перестановок в $(x_{N1}, \ldots, x_{NN})$, и*

$$L_N = \sum_{i=1}^{N} a_{Ni} x_i. \tag{9.5.4}$$

Тогда

$$\mathscr{E}(L_N) = N\bar{a}_N \bar{x}_N \tag{9.5.5}$$

и

$$\sigma^2(L_N) = \frac{N^2}{N-1} m_{2,N}(a)\, m_{2,N}(x). \tag{9.5.6}$$

Более того,

$$\lim_{N\to\infty} P\left(\frac{L_N - \mathscr{E}(L_N)}{\sigma(L_N)} \leqslant z\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}u^2}\, du. \tag{9.5.7}$$

Доказательство **9.5.1** довольно прямолинейное, но громоздкое. Процедура доказательства начинается с введения множества величин

$$a'_{Ni} = \frac{a_{Ni} - \bar{a}_N}{\sqrt{m_{2,N}(a)}}, \qquad x'_{Ni} = \frac{x_{Ni} - \bar{x}_N}{\sqrt{m_{2,N}(x)}}.$$

Тогда множества последовательностей $(a'_{N1}, \ldots, a'_{NN})$ и $(x'_{N1}, \ldots, x'_{NN})$, $N = 1, 2, \ldots$, удовлетворяют условиям, аналогичным (9.5.1) и (9.5.2). Если $(x'_1, \ldots, x'_N)$ — такая же перестановка в $(x'_{N1}, \ldots, x'_{NN})$, что и $(x_1, \ldots, x_N)$ в $(x_{N1}, \ldots, x_{NN})$, то, положив $L'_N = \sum_{i=1}^{N} a'_{Ni} x'_i$, получаем $\mathscr{E}(L'_N) = 0$ и

$$\frac{L'_N}{\sigma(L'_N)} = \frac{L_N - \mathscr{E}(L_N)}{\sigma(L_N)}. \tag{9.5.8}$$

Опуская детальные вычисления моментов, которые были проделаны Вальдом и Вольфовитцем (1944), находим, что для любого положительного целого числа k

$$\begin{aligned} \mathscr{E}(L'^s_N) &= \frac{(2k)!}{2^k k!} N^k + o(N^k), \qquad & s &= 2k, \\ &= o(N^k), & s &= 2k+1 \end{aligned} \tag{9.5.9}$$

и, следовательно,

$$\begin{aligned} \lim_{N\to\infty} \mathscr{E}\left(\frac{L'_N}{\sigma(L'_N)}\right)^s = \lim_{N\to\infty} \mathscr{E}\left(\frac{L_N - \mathscr{E}(L_N)}{\sigma(L_N)}\right)^s &= \frac{(2k)!}{2^k k!}, \qquad & s &= 2k, \\ &= 0, & s &= 2k+1. \end{aligned} \tag{9.5.10}$$

Тогда при $N \to \infty$ s-й момент $[L_N - \mathscr{E}(L_N)]/\sigma(L_N)$ сходится к s-му моменту случайной величины с распределением $N(0, 1)$. Поэтому, как следует из **5.5.3**, предельное распределение $[L_N - \mathscr{E}(L_N)]/\sigma(L_N)$ при $N \to \infty$ есть $N(0, 1)$.

Если в определении величины L_N согласно (9.5.4) мы выберем $a_{Ni}=\frac{1}{n}$, $i=1, \ldots, n$, и $a_{Ni}=0$, $i=n+1, \ldots, N$, то L_N будет иметь смысл среднего значения выборки объема n из конечной совокупности π_N, чьи N элементов имеют x-значения $x_{N1}, \ldots, x_{NN}$. Более того, $\mathscr{E}(L_N)=\mu_N$, где $\mu_N=\frac{1}{N}\sum_{i=1}^{N} x_{Ni}$ и $\sigma^2(L_N)=\left(\frac{1}{n}-\frac{1}{N}\right)\sigma_N^2$, где μ_N и σ_N^2 в свою очередь — среднее значение и дисперсия π_N, определенные в § 8.5. В этом случае (9.5.1) сводится к условию о том, чтобы при n, $N\to\infty$ предел

$$\frac{\left(\frac{N}{n}-1\right)^{r-1}-(-1)^{r-1}}{\left(\frac{N}{n}\right)\left(\frac{N}{n}-1\right)^{\frac{1}{2}r-1}} \tag{9.5.11}$$

был конечен для $r\geqslant 3$. Можно показать, что это имеет место, если $\lim\limits_{N,\,n\to\infty}\frac{N}{n}=C$, где $+1<C<+\infty$. Таким образом, получаем следствие **9.5.1**.

9.5.1a. *Допустим, что* $\{\pi_N,\ N=1, 2, \ldots\}$ — *последовательность конечных совокупностей таких, что* π_N *имеют средние значения* μ_N *и дисперсии* σ_N^2, *и пусть* $\bar{x}_n$ — *среднее значение случайной выборки объема* n *из* π_N. *Тогда, если* $(x_{N1}, \ldots, x_{NN},\ N=1, 2, \ldots)$ *удовлетворяет при больших* N *условию* (9.5.2) *и если* N, $n\to\infty$, *так что* $\lim\frac{N}{n}=C$, *где* $+1<C<+\infty$, *то*

$$\lim P\left(\frac{(\bar{x}_n-\mu_N)}{\left[\left(\frac{1}{n}-\frac{1}{N}\right)\sigma_N^2\right]^{1/2}}\leqslant z\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{1}{2}u^2}\,du. \tag{9.5.12}$$

9.6. Асимптотические распределения, связанные с порядковыми статистиками

(а) Предельные распределения для сумм долей. В **8.7.2** было показано, что если $(x_{(1)}, \ldots, x_{(n)})$ — вариационный ряд выборки с непрерывной к. ф. р. $F(x)$, то случайная величина $y_{(k)}=F(x_{(k)})$ имеет к. ф. р.

$$D_n(y_{(k)})=\frac{\Gamma(n+1)}{\Gamma(k)\,\Gamma(n-k+1)}\int_0^{y_{(k)}} x^{k-1}(1-x)^{n-k}\,dx. \tag{9.6.1}$$

Необходимо помнить, что, как следует из **8.7.6**, (9.6.1) определяет функцию распределения суммы некоторых k из $n+1$ долей $F(x_{(1)})$, $F(x_{(2)})-F(x_{(1)})$, $\ldots$, $1-F(x_{(n)})$. Величина $F(x_{(s+k)})-F(x_{(s)})$ есть как раз такая сумма, когда $0<s<n-k$. Рассмотрим теперь слу-

чайную величину $w_k = ny_{(k)}$. Если мы обозначим через $H_n(w_k)$ к. ф. р. w_k, то будем иметь

$$H_n(w_k) = D_n\left(\frac{w_k}{n}\right) = \int_0^{w_k} h_n(y)\,dy, \tag{9.6.2}$$

где

$$h_n(y) = \frac{\Gamma(n+1)}{n^k \Gamma(k)\,\Gamma(n-k+1)}\, y^{k-1}\left(1-\frac{y}{n}\right)^{n-k}.$$

Очевидно, что при $n \to \infty$ последовательность $h_n(y)$, $n = k$, $k+1, \ldots$, стремится к функции $\frac{1}{\Gamma(k)} y^{k-1} e^{-y}$ равномерно в интервале $(0, w_k)$ и, следовательно,

$$\lim_{n\to\infty} H_n(w_k) = \frac{1}{\Gamma(k)} \int_0^{w_k} y^{k-1} e^{-y}\,dy. \tag{9.6.3}$$

Таким образом, получаем следующий результат.

9.6.1. *Если $(x_{(1)}, \ldots, x_{(n)})$ — вариационный ряд выборки из совокупности с непрерывной к. ф. р. $F(x)$, то при фиксированном k $nF(x_{(k)})$, $n = k, k+1, \ldots$, есть последовательность случайных величин, сходящаяся по распределению к гамма-распределению $G(k)$.*

Необходимо заметить, что **9.6.1** имеет место и в том случае, если $F(x_{(k)})$ заменить на сумму некоторых k долей $F(x_{(1)})$, $F(x_{(2)}) - F(x_{(1)}), \ldots$, $1 - F(x_{(n)})$. Например, $1 - F(x_{n-k+1})$ или $F(x_{(s+k)}) - F(x_{(s)})$. Утверждение остается также справедливым, если заменить $F(x_{(k)})$ некоторыми k долями, определенными для выборки объема n из совокупности, которая имеет непрерывную многомерную к. ф. р. (см. § 8.7 (с)).

Теорема **9.6.1** может быть обобщена на случай двух и более сумм от фиксированного числа долей. В случае двух таких сумм аналогичный результат может быть сформулирован следующим образом.

9.6.2. *Допустим, что $(x_{(1)}, \ldots, x_{(n)})$ — порядковые статистики выборки из непрерывной к. ф. р. $F(x)$. Тогда, если k_1 и k_2 — фиксированные целые числа, то последовательность пар случайных величин $(nF(x_{(k_1)}), nF(x_{(k_1+k_2)}))$, $n = m, m+1, \ldots$, где $m \geqslant k_1 + k_2$, сходится по распределению к распределению, плотность которого*

$$f(w_1, w_2) = \frac{1}{\Gamma(k_1)\,\Gamma(k_2)}\, w_1^{k_1-1}(w_2 - w_1)^{k_2-1} e^{-w_2}$$

в области $0 < w_1 < w_2 < \infty$ и $f(w_1, w_2) = 0$ вне этой области плоскости (w_1, w_2).

Доказательство этого утверждения аналогично доказательству **9.6.1**, и мы оставляем его читателю. Необходимо особо отметить, что **9.6.2** имеет место, если $F(x_{(k_1)})$ заменить на сумму некоторых k_1 долей и $F(x_{(k_1+k_2)})$ — суммой некоторых $k_1 + k_2$ долей, которые включают в себя первые k_1 долей.

Рассмотрим теперь предельное распределение $y_{(k)}$ при $n \to \infty$ не для фиксированного k, а для увеличивающейся последовательности величин k и n такой, что $\frac{k}{n} = p_n = p + O\left(\frac{1}{n}\right)$. Рассмотрим последовательность случайных величин $v_n = (y_{(np_n)} - p)\sqrt{n}$, $n = m, m+1, \ldots$; $m \geqslant np_n$, где $y_{(np_n)} = F(x_{(np_n)})$. Обозначая к. ф. р. v_n через $H_n^*(v)$ имеем

$$H_n^*(v) = D_n\left(p + \frac{v}{\sqrt{n}}\right), \tag{9.6.4}$$

где $D_n(y_{(np_n)})$ определено согласно (9.6.1). В свою очередь $D_n(p + v/\sqrt{n})$ может быть представлено в виде

$$\int\limits_{-p\sqrt{n}}^{v} h_n^*(z)\,dz, \tag{9.6.5}$$

где

$$\begin{aligned} h_n^*(z) &= \frac{\Gamma(n+1)}{\sqrt{n}\,\Gamma(k)\,\Gamma(n-k+1)}\left(p + \frac{z}{\sqrt{n}}\right)^{k-1}\left(1 - p - \frac{z}{\sqrt{n}}\right)^{n-k} = \\ &= \frac{\Gamma(n+1)\,(p + z/\sqrt{n})^{-1}}{\sqrt{n}\,\Gamma(p_n n)\,\Gamma((1-p_n)\,n+1)}\,[p^{p_n}(1-p)^{1-p_n}]^n \times \\ &\qquad \times\left[\left(1 + \frac{1}{p\sqrt{n}}\right)^{p_n}\left(1 - \frac{z}{(1-p)\sqrt{n}}\right)^{1-p_n}\right]^n. \end{aligned} \tag{9.6.6}$$

Так как $p_n = p + O\left(\frac{1}{n}\right)$, то справедливо соотношение

$$\left(1 + \frac{z}{p\sqrt{n}}\right)^{p_n}\left(1 - \frac{z}{(1-p)\sqrt{n}}\right)^{1-p_n} = \left(1 - \frac{z^2}{2np(1-p)} + \varphi(z, n)\right), \tag{9.6.7}$$

где $\varphi(z, n)$ таково, что $\lim\limits_{n\to\infty} n\varphi(z, n) = 0$ для любого значения z. Используя формулу Стирлинга (7.6.27), для $\Gamma(g)$ при больших g находим, что

$$\begin{aligned} \frac{\Gamma(n+1)}{\sqrt{n}\,\Gamma(p_n n)\,\Gamma((1-p_n)\,n+1)}\,[p^{p_n}(1-p)^{1-p_n}]^n &= \\ &= \sqrt{\frac{p}{2\pi(1-p)}} + O\left(\frac{1}{n}\right). \end{aligned} \tag{9.6.8}$$

В итоге получаем последовательность

$$h_n^*(z) = \left(p + \frac{z}{\sqrt{n}}\right)^{-1}\left(\sqrt{\frac{p}{2\pi(1-p)}} + O\left(\frac{1}{n}\right)\right)\left(1 - \frac{z^2}{2np(1-p)} + \varphi(z, n)\right)^n, \tag{9.6.9}$$

которая равномерно сходится при $n \to \infty$ к функции

$$h^*(z) = \frac{1}{\sqrt{2\pi p(1-p)}}\,e^{-z^2/2p(1-p)} \tag{9.6.10}$$

в любом интервале $(-K, v)$. Для любого $\varepsilon > 0$ мы можем выбрать K так, что

$$\int_{-\infty}^{-K} h^*(z)\,dz < \frac{\varepsilon}{2}. \tag{9.6.11}$$

Более того, существует $n_0 > K^2/p^2$ такое, что для любого v и $n > n_0$

$$\left| \int_{-K}^{v} h_n^*(z)\,dz - \int_{-K}^{v} h^*(z)\,dz \right| < \frac{\varepsilon}{2}. \tag{9.6.12}$$

Поэтому для $n > n_0$

$$\left| \int_{-\infty}^{v} h^*(z)\,dz - \int_{-p\sqrt{n}}^{v} h_n^*(z)\,dz \right| < \varepsilon \tag{9.6.13}$$

и, следовательно,

$$\lim_{n\to\infty} P\left(\frac{[F(x_{(np_n)}) - p]\sqrt{n}}{\sqrt{p(1-p)}} < w\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{w} e^{-\frac{1}{2}u^2}\,du. \tag{9.6.14}$$

Подводя итоги, получаем следующий результат:

9.6.3. *Если $(x_{(1)}, \ldots, x_{(n)})$ — порядковые статистики выборки из непрерывной к. ф. р. $F(x)$ и np_n — целое число такое, что $p_n = p + O\left(\frac{1}{n}\right)$, где $0 < p < 1$, то для больших n случайная величина $F(x_{(np_n)})$ асимптотически распределена согласно $N\left(p, \frac{p(1-p)}{n}\right)$.*

Можно заметить, что **9.6.3** остается справедливым, если заменить $F(x_{(np_n)})$ на сумму некоторых np_n долей в случае одного или более измерений.

Утверждение **9.6.3** может без особых трудностей быть обобщено на случай нескольких сумм долей. Достаточно сформулировать результат для случая двух сумм, а его доказательство мы оставляем читателю.

9.6.4. *Если $(x_{(1)}, \ldots, x_{(n)})$ — вариационный ряд выборки из совокупности с непрерывной к. ф. р. $F(x)$ и если np_{1n} и np_{2n} — целые числа такие, что $p_{1n} = p_1 + O\left(\frac{1}{n}\right)$, $p_{2n} = p_2 + O\left(\frac{1}{n}\right)$, где $0 < p_1 < p_2 < 1$, то для больших n двумерная случайная величина $F(x_{(np_{1n})})$, $F(x_{(np_{2n})})$ имеет $N\left(\{p_i\}; \left\|\frac{\sigma_{ij}}{n}\right\|\right)$ в качестве своего асимптотического распределения, где $\sigma_{11} = p_1(1-p_1)$, $\sigma_{12} = \sigma_{21} = p_1(1-p_2)$ и $\sigma_{22} = p_2(1-p_2)$.*

(b) Предельные распределения порядковых статистик. Предельные распределения, полученные в **9.6.1**, **9.6.2**, **9.6.3** и **9.6.4**, имеют непосредственное отношение к свойствам распределений сумм долей и только косвенно характеризуют свойства распределений самих порядковых статистик в случае больших выборок. Поскольку распределение совокупности $F(x)$ по предположению непрерывно, можно получить некоторые утверждения относительно предельных распределений самих порядковых статистик. Например, пусть в теореме **9.6.1** $F^{-1}(y)$ — функция, обратная к $F(x)$. Естественно рассмотреть последовательность значений n, для которых $F^{-1}\left(\frac{k}{n}\right)$ имеет единственную обратную. Так как $F(x)$ непрерывна и определена для всех значений x, то, очевидно, существует бесконечная последовательность таких значений n, назовем их $n_1, n_2, \ldots$ Тогда, как следует из **9.6.1**, для любого фиксированного k

$$\lim_{i\to\infty} P\left(x_{(k)} \leqslant F^{-1}\left(\frac{u}{n_i}\right)\right) = \frac{1}{\Gamma(k)} \int_0^u y^{k-1} e^{-y}\, dy \tag{9.6.15}$$

и, следовательно, при больших n

$$P(x_{(k)} \leqslant v) \cong \frac{1}{\Gamma(k)} \int_0^{nF(v)} y^{k-1} e^{-y}\, dy. \tag{9.6.16}$$

Для $x_{(n-k'+1)}$ при фиксированном k' могут быть получены формулы, аналогичные (9.6.15) и (9.6.16). Асимптотические результаты этого типа, особенно для $k=1$ (наименьшая порядковая статистика) и для $k'=1$ (наибольшая порядковая статистика), были детально рассмотрены Доддом (1923), Фишером и Типпетом (1928), Фреше (1927), Гамбелем (1935, 1958) и Смирновым (1935).

На основании **9.6.3** мы можем написать

$$\lim_{i\to\infty} P\left(x_{(n_i p_{n_i})} \leqslant F^{-1}\left(p + \frac{w\sqrt{p(1-p)}}{\sqrt{n_i}}\right)\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{w} e^{-\frac{1}{2}y^2}\, dy \tag{9.6.17}$$

и для больших n

$$P(x_{(np_n)} \leqslant v) \cong \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{T} e^{-\frac{1}{2}y^2}\, dy, \tag{9.6.18}$$

где

$$T = \frac{(F(v)-p)\sqrt{n}}{\sqrt{p(1-p)}}.$$

Формулы (9.6.16) и (9.6.18) без особых трудностей могут быть обобщены на случай двух и более порядковых статистик. В двумерном случае результаты непосредственно следуют из **9.6.2** и **9.6.4**.

Из (9.6.14) с очевидностью вытекает, что при $n\to\infty$ $F(x_{(np_n)})$ сходится по вероятности к постоянной p. Допустим теперь, что уравнение $F(x)=p$ имеет единственное решение, т. е. существует единственная квантиль $\underline{x}_p$. Тогда, так как функция $F(x)$ непрерывна, то $x_{(np_n)}$ сходится по вероятности к $\underline{x}_p$. Таким образом получаем следующий результат.

9.6.5. *Если к предположениям, сформулированным в* **9.6.3**, *добавить предположение о том, что существует единственная квантиль* $\underline{x}_p$, *то при* $n \to \infty$ $x_{(np_n)}$ *сходится по вероятности к* $\underline{x}_p$.

Аналогичный факт имеет место в случае двух порядковых статистик $x_{(np_{1n})}$ и $x_{(np_{2n})}$ и вообще для любого числа аналогичных порядковых статистик.

Допустим теперь, что в некоторой окрестности $V(\underline{x}_p)$ точки $x = \underline{x}_p$ функция $F(x)$ имеет производную $f(x)$, причем $f(\underline{x}_p) > 0$. Тогда существует единственная p-я квантиль $\underline{x}_p$ и, следовательно, на основании **9.6.5** $x_{(np_n)}$ сходится по вероятности к $\underline{x}_p$. Если $x_{(np_n)}$ — некоторая точка в $V(\underline{x}_p)$, то мы можем написать

$$F(x_{(np_n)}) = p + f(x^*)\,(x_{(np_n)} - \underline{x}_p), \tag{9.6.19}$$

где x^* — случайная величина такая, что $|x^* - \underline{x}_p| < |x_{(np_n)} - \underline{x}_p|$. Но так как $x_{(np_n)}$ сходится по вероятности к квантиле $\underline{x}_p$ при $n \to \infty$, то для произвольного $\varepsilon > 0$ мы можем выбрать такое n_ε, что

$$P(F(x_{(np_n)}) = p + f(x^*)(x_{(np_n)} - \underline{x}_p) \quad \text{для всех } n > n_\varepsilon) > 1 - \varepsilon. \tag{9.6.20}$$

Отсюда следует, что последовательности

$$\frac{(F(x_{(np_n)}) - p)\sqrt{n}}{\sqrt{p(1-p)}}, \quad n = m,\ m+1,\ \ldots;\ m \geqslant np_n, \tag{9.6.21}$$

и

$$\frac{f(x^*)(x_{(np_n)} - \underline{x}_p)\sqrt{n}}{\sqrt{p(1-p)}}, \quad n = m,\ m+1,\ \ldots;\ m \geqslant np_n, \tag{9.6.22}$$

образуют стохастический процесс такой, что они обе сходятся по распределению к нормальному закону $N(0, 1)$.

Тот факт, что $F(x)$ имеет производную $f(x)$ в $V(\underline{x}_p)$, означает, что $f(x)$ непрерывна в $V(\underline{x}_p)$ и, следовательно, при $x = \underline{x}_p$. Поэтому на основании **4.3.5** $f(x_{(np_n)})$ сходится по вероятности к постоянной $f(\underline{x}_p)$ и на основании **4.3.7** $f(x^*)$ также сходится по вероятности к $f(\underline{x}_p)$ при $n \to \infty$. Наконец, на основании **4.3.3** можно видеть, что последовательность случайных величин (9.6.22) и последовательность

$$\frac{f(\underline{x}_p)(x_{(np_n)} - \underline{x}_p)\sqrt{n}}{\sqrt{p(1-p)}}, \quad n = m,\ m+1,\ \ldots;\ m \geqslant np_n, \tag{9.6.22a}$$

составляют стохастический процесс такой, что эти две последовательности сходятся вместе по распределению к $N(0, 1)$.

Подводя итог, получаем следующий результат.

9.6.6. *Если дополнительно к предположениям, сформулированным* в **9.6.3**, *мы добавим предположение о том, что $F(x)$ имеет производную $f(x)$ в некоторой окрестности $V(\underline{x}_p)$ точки $x=\underline{x}_p$ такую, что $f(\underline{x}_p)>0$, тогда для больших n $x_{(np_n)}$ асимптотически распределено согласно* $N\left(\underline{x}_p, \dfrac{p(1-p)}{nf^2(\underline{x}_p)}\right)$.

Пример. Интересный частный случай **9.6.6** имеет место при $p=\frac{1}{2}$. В этом случае $x_{(np_n)}$ является выборочной медианой, и мы видим, что ее асимптотическое распределение есть $N\left(\underline{x}_{0,5}, \dfrac{1}{4nf^2(\underline{x}_{0,5})}\right)$. Если распределение популяции есть $N(\mu, \sigma^2)$, то выборочная медиана при больших n в качестве асимптотического распределения имеет $N\left(\mu, \dfrac{\pi\sigma^2}{2n}\right)$.

Утверждение, аналогичное **9.6.6**, имеет место, конечно, и для совместного распределения двух и более порядковых статистик. Так, в случае двух порядковых статистик $x_{(np_1n)}$ и $x_{(np_2n)}$ мы можем сказать, что

9.6.7. *Если дополнительно к предположениям теоремы* **9.6.4** *мы добавим предположение о том, что $F(x)$ имеет производную $f(x)$ в окрестности каждой из точек $x=\underline{x}_{p_1}$ и $x=\underline{x}_{p_2}$, $0<p_1<p_2<1$, причем $f(\underline{x}_{p_1})>0$ и $f(\underline{x}_{p_2})>0$, тогда при больших n случайная величина $(x_{(np_1n)}, x_{(np_2n)})$ асимптотически распределена согласно $N\left(\{x_{p_i}\}, \left\|\dfrac{\sigma^*_{ij}}{n}\right\|\right)$, $i, j=1, 2$, где*

$$\sigma^*_{11}=\frac{p_1(1-p_1)}{f^2(\underline{x}_{p_1})}, \quad \sigma^*_{12}=\frac{p_1(1-p_2)}{f(\underline{x}_{p_1})f(\underline{x}_{p_2})}, \quad \sigma^*_{22}=\frac{p_2(1-p_2)}{f^2(\underline{x}_{p_2})}.$$

Теоремы **9.6.6** и **9.6.7** были установлены Смирновым (1935), хотя формулы для дисперсий и ковариации $x_{(np_1n)}$ и $x_{(np_2n)}$ были впервые получены К. Пирсоном (1920). Мостеллер (1946) обобщил теорему **9.6.7** на случай любого числа порядковых статистик.

ЗАДАЧИ

9.1. Сформулировать и доказать k-мерный аналог **9.1.1**.

9.2. Пусть $(x_{1\xi}, \ldots, x_{k\xi};\ \xi=1, \ldots, n)$ — выборка из k-мерного распределения, имеющего конечные средние значения $(\mu_1, \ldots, \mu_k)$ и конечную ковариационную матрицу $\|\sigma_{ij}\|$, и пусть $(\bar{x}_1, \ldots, \bar{x}_k)$ — вектор средних значений выборки. Показать, что для произвольного $\delta_i>0$, $i=1, \ldots, k$,

$$P(|\bar{x}_i-\mu_i|<\delta_i, \quad i=1, \ldots, k)\geqslant 1-\frac{1}{n}\left(\frac{\sigma_{11}}{\delta_1^2}+\ldots+\frac{\sigma_{kk}}{\delta_k^2}\right)$$

и, следовательно, что $(\bar{x}_1, \ldots, \bar{x}_k)$ сходятся по вероятности к $(\mu_1, \ldots, \mu_k)$.

9.3. (*Продолжение.*) Показать, что для произвольного $\delta^2 > 0$

$$P\left(\sum_{i,\,j=1}^{k} \sigma^{ij}(\bar{x}_i-\mu_i)(\bar{x}_j-\mu_j) < \delta^2\right) \geqslant 1-\frac{k}{n\delta^2},$$

где $\|\sigma^{ij}\| = \|\sigma_{ij}\|^{-1}$, и, следовательно, что $(\bar{x}_1, \ldots, \bar{x}_k)$ сходится по вероятности к $(\mu_1, \ldots, \mu_k)$.

9.4. Доказать 9.2.3.

9.5. Пусть $(x_1, \ldots, x_n)$ — выборка из распределения Пуассона $\mathrm{Po}(\mu)$. Показать, что для больших n $2\sqrt{\bar{x}}$ имеет асимптотическое распределение $N\left(2\sqrt{\mu}, \frac{1}{n}\right)$. Показать, что этот же результат имеет место, если $(x_1, \ldots$ $\ldots, x_n)$ — выборка из гамма-распределения $G(\mu)$.

9.6. Пусть $(x_1, \ldots, x_n)$ — выборка из биномиального распределения $\mathrm{Bi}(1, p)$. Показать, что для больших n асимптотическое распределение для $\sin^{-1}(2\bar{x}-1)$ есть

$$N\left(\sin^{-1}(2p-1), \frac{1}{n}\right).$$

9.7. Пусть $(x_1, \ldots, x_n)$ — выборка из распределения времени ожидания, имеющего функцию вероятности

$$qp^{x-1}, \qquad x=1, 2, \ldots,$$

где $0<p<1$, $p+q=1$. Показать, что для больших n асимптотическое распределение $\log\left[\bar{x}\left(1+\sqrt{1-\frac{1}{\bar{x}}}\right)-\frac{1}{2}\right]$ есть

$$N\left(\log\left[\frac{1+\sqrt{p}}{q}-\frac{1}{2}\right], \frac{1}{n}\right).$$

9.8. Пусть $x_1, \ldots, x_n$ — выборка из равномерного распределения $R\left(\frac{1}{2}\theta, \theta\right)$. Показать, что для больших n асимптотическое распределение для $\sqrt{12}\log(2\bar{x})$ есть

$$N\left(\sqrt{12}\log\theta, \frac{1}{n}\right).$$

9.9. Пусть $(x_1, \ldots, x_n)$ — выборка из распределения, имеющего конечное среднее значение μ и конечную дисперсию σ^2. Показать, что выборочная дисперсия s^2 сходится по вероятности к σ^2 и что для больших n асимптотическое распределение $\sqrt{n}(\bar{x}-\mu)/s$ есть $N(0, 1)$.

9.10. Пусть n_1, $\bar{x}_1$, s_1^2 — объем, среднее значение и дисперсия выборки из распределения, имеющего среднее значение μ_1 и дисперсию σ_1^2, а n_2, $\bar{x}_2$, s_2^2 — объем, среднее значение и дисперсия независимой выборки из распределения, имеющего среднее значение μ_2 и дисперсию σ_2^2. Показать, что предельное при $n_1, n_2 \to \infty$ распределение величины

$$\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}}$$

есть $N(0, 1)$.

9.11. Пусть $\tilde{x}$ — медиана выборки объема n из распределения, имеющего непрерывную к. ф. р. $F(x)$. Показать, что для больших n асимптотическое распределение $F(\tilde{x})$ есть $N\left(\frac{1}{2}, \frac{1}{4n}\right)$.

9.12. Пусть $(x_{(1)}, \ldots, x_{(n)})$ — порядковые статистики из выборки с распределением, имеющим непрерывную к. ф. р. $F(x)$. Показать, что при $n \to \infty$ распределение

$$n\left(1 - \int_{x_{(1)}}^{x_{(n)}} dF(x)\right)$$

есть гамма-распределение $G(2)$.

9.13. Пусть $(x_1, \ldots, x_{2n+1})$ — выборка из распределения, имеющего плотность функции распределения $\lambda e^{-\lambda x}$, $x > 0$, $\lambda > 0$. Показать, что при больших n медиана $\tilde{x}$ выборки в качестве асимптотического распределения имеет

$$N(\log 2/\lambda,\ 1/(2\lambda^2 n)).$$

9.14. Пусть $\bar{x}$ и $\tilde{x}$ — среднее значение и медиана выборки объема n из совокупности, имеющей распределение $N(\mu, \sigma^2)$. Показать, что для больших n асимптотическое распределение $(\bar{x}, \tilde{x})$ есть

$$N\left(\mu, \mu; \left\| \frac{\sigma_{ij}}{n} \right\|\right),$$

где

$$\left\| \frac{\sigma_{ij}}{n} \right\| = \left\| \begin{matrix} \dfrac{\sigma^2}{n} & \dfrac{\sigma^2}{n} \\ \dfrac{\sigma^2}{n} & \dfrac{\pi\sigma^2}{2n} \end{matrix} \right\|.$$

9.15. *(Преобразование Фишера* (1925a) *коэффициента корреляции.)* Известно, что распределение выборочного коэффициента корреляции r в выборках из двумерного нормального распределения, коэффициент корреляции которого ρ, при больших n переходит в асимптотическое распределение, равное $N(\rho, (1-\rho^2)^2/n)$. Показать, что асимптотическое распределение $\left(\frac{1+r}{1-r}\right)$ при больших n есть

$$N\left(\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{n}\right).$$

9.16. Доказать **9.6.2.**

9.17. Доказать **9.6.4.**

ГЛАВА 10

ЛИНЕЙНОЕ СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ

10.1. Вводные замечания

В главах 8 и 9 мы изложили ряд результатов теории выборочного метода, т. е. вероятностной теории некоторых функций от элементов выборки из совокупности $F(x)$. В задачах прикладной статистики функция $F(x)$ обычно неизвестна и главная цель выборочного метода состоит в получении информации, на основании которой можно высказать некоторые утверждения или сделать выводы относительно $F(x)$ или некоторых ее свойств. Эти выводы делаются в терминах некоторых функций от элементов (компонент) выборки и формулируются в виде вероятностных утверждений. Условия, налагаемые на $F(x)$ до получения каких-либо выборочных данных, могут быть совершенно различными, начиная от самых общих предположений о том, что $F(x)$ удовлетворяет лишь основным ограничениям, которые согласно **2.2.1** налагаются на к. ф. р., до предположений, которые полностью определяют всю $F(x)$ для всех x в R_1. Вообще говоря, условия, налагаемые на $F(x)$ в реальных ситуациях, лежат между этими крайними предположениями. Например, можно предположить, что $F(x)$ имеет конечные, но неизвестные среднее значение μ и дисперсию σ^2, причем задача состоит в нахождении *оценок* для μ и σ^2 как функций от элементов случайной выборки из $F(x)$. Так, в § 8.2 было показано, что выборочное среднее значение $\bar{x}$ и выборочная дисперсия s^2 выборки из бесконечной совокупности являются *несмещенными оценками* для μ и σ^2, т. е. $\mathscr{E}(\bar{x})=\mu$ и $\mathscr{E}(s^2)=\sigma^2$. В более общем случае задача состоит в том, чтобы, используя информацию, содержащуюся в выборке, выяснить, какими свойствами обладает функция $F(x)$, помимо тех, которые оговорены в предположении.

Во многих задачах прикладной статистики обычно достаточно найти на основании выборок несмещенные оценки *параметров* распределений совокупностей таких, как средние значения, дисперсии и коэффициенты регрессии. Класс относительно простых оценок таких параметров, известных как *линейные оценки*, представляют собой линейные функции элементов выборки. Для его построения не требуется более сильных предположений, чем конечность моментов первого и второго порядков компонент выборки. Существуют, конечно, и другие классы оценок, которые будут рассмотрены в главах 11 и 12. Оценка *параметра* совокупности θ, вообще говоря, есть наблюдаемая случайная величина, определяемая по выборке, т. е. заданная функция элементов выборки, которая используется вместо неизвестного истинного значения параметра, подлежащего оцениванию. При конструировании оценки на основании выборки для параметра θ важно выбрать ее такой, чтобы ее распределение было сконцентрировано, как это только возможно, около *истинного значения* θ_0, характеризующего распределение совокупности, из которой взята выборка. Как мы в дальнейшем увидим, очень естественной мерой концентрации оценок в случае линейных оценок является дисперсия оценки. В настоящей главе мы уделим основное внимание теории линейных оценок, методам построения этих оценок и применениям их для некоторых важных статистических задач. Аналогичное обсуждение для других классов оценок будет изложено в главах 11 и 12.

Как уже было отмечено в §§ 8.2 и 8.5, выборочное среднее значение, которое, конечно, есть линейная функция от элементов выборки, является несмещенной оценкой для среднего значения совокупности как в случае конечной, так и бесконечной совокупностей. Однако выборочное среднее значение есть только одна из возможных линейных оценок для среднего значения совокупности, которые мы можем построить. Например, если $(x_1, \ldots, x_n)$ — выборка из совокупности (конечной или бесконечной) со средним значением μ, то $a_1x_1 + \ldots + a_nx_n$, очевидно, является несмещенной оценкой для μ, если $a_1, \ldots, a_n$ — заданные постоянные, такие, что $a_1 + \ldots + a_n = 1$. Оценки такого типа называются *несмещенными линейными оценками*. Если среди всех известных линейных несмещенных оценок для μ выбрана какая-то определенная оценка, то необходимо указать критерий, на основании которого этот выбор был сделан. Один из широко применяемых критериев состоит в выборе той оценки, которая обладает наименьшей возможной дисперсией. Такая оценка называется *линейной оценкой с минимальной дисперсией*.

Аналогично, если $(x_1, \ldots, x_n)$ — выборка из конечной или бесконечной совокупности, которая имеет среднее значение μ и дисперсию σ^2, то любая квадратичная форма $\sum\limits_{\xi, \eta} a_{\xi\eta}(x_\xi - \bar{x})(x_\eta - \bar{x})$, среднее значение которой равно σ^2, есть *несмещенная квадратичная оценка* для σ^2. Если существует единственная несмещенная квадратичная оценка для σ^2, имеющая наименьшую возможную дисперсию, то она называется *квадратичной оценкой с наименьшей дисперсией* для σ^2. Необходимо при этом отметить, что любая квадратичная оценка есть *линейная функция* от квадратичных членов, и поэтому бóльшая часть результатов теории линейных оценок, полученных в настоящей главе, может быть применена к квадратичным оценкам.

Линейные оценки с наименьшей дисперсией обычно сами имеют дисперсию, которую можно несмещенно оценить с помощью квадратичных оценок.

В настоящей главе мы будем иметь дело с теорией линейного оценивания и ее применениями для оценивания средних значений и дисперсий распределения совокупности. Для этих задач в рамках теории линейного оценивания не требуется каких-либо предположений относительно распределения случайных величин, более сильных, чем конечность средних значений и элементов ковариационной матрицы этих случайных величин. Случайные величины, лежащие в основе квадратичных оценок, являются квадратичными формами компонент выборки.

Иногда удобно использовать обозначение $\mathscr{E}^{-1}(\theta)$ для несмещенной оценки θ. Таким образом, если T — наблюдаемая случайная величина, которая является несмещенной оценкой параметра θ, т. е. если

$$\mathscr{E}T=\theta,$$

то мы можем написать

$$\mathscr{E}^{-1}(\theta)=T.$$

Простая, но фундаментальная теорема теории линейного оценивания, которая будет часто использоваться при исследовании линейных и квадратичных оценок, формулируется следующим образом.

10.1.1. *Допустим, что* $(x_1, \ldots, x_n)$ — *случайные величины, средние значения которых равны*

$$\mathscr{E}x_i=\sum a_{ij}\theta_j, \qquad i=1, \ldots, k,$$

где $\theta_1, \ldots, \theta_k$ — *неизвестные параметры и* $\|a_{ij}\|$ — *невырожденная матрица с известными значениями элементов (т. е. они не зависят от параметров* $\theta_1, \ldots, \theta_k$*). Тогда*

$$\mathscr{E}^{-1}(\theta_i)=\sum_j a^{ij}x_j, \qquad i=1, \ldots, k,$$

являются несмещенными оценками для $\theta_1, \ldots, \theta_k$ *соответственно, где* $\|a^{ij}\|=\|a_{ij}\|^{-1}$.

Доказательство этой теоремы мы оставляем читателю.

10.2. Оценки с наименьшей дисперсией для среднего значения и дисперсии совокупности

(a) Линейная оценка с минимальной дисперсией для среднего значения совокупности. Допустим, что $(x_1, \ldots, x_n)$ — выборка из распределения, имеющего среднее значение μ и дисперсию σ^2. Как мы уже видели в § 8.2, выборочное среднее значение является несмещенной оценкой для μ. Мы покажем сейчас, что $\bar{x}$ является линейной оценкой с наименьшей дисперсией для μ. Пусть $\mathscr{E}^{-1}(\mu)$ — некоторая несмещенная линейная оценка для μ, т. е.

$$\mathscr{E}^{-1}(\mu)=a_1x_1+\ldots+a_nx_n, \tag{10.2.1}$$

где
$$a_1 + \ldots + a_n = 1. \tag{10.2.2}$$

Для дисперсии $\mathcal{E}^{-1}(\mu)$ на основании **3.6.1a** имеем
$$\sigma^2(\mathcal{E}^{-1}(\mu)) = (a_1^2 + \ldots + a_n^2) \cdot \sigma^2. \tag{10.2.3}$$

Можно видеть, что $\sigma^2(\mathcal{E}^{-1}(\mu))$ имеет единственный минимум, который соответствует
$$a_1 = \ldots = a_n = \frac{1}{n}. \tag{10.2.4}$$

Но при таком выборе коэффициентов a_k $\mathcal{E}^{-1}(\mu)$ совпадает с $\bar{x}$. Таким образом,

10.2.1. *Если $(x_1, \ldots, x_k)$ — выборка объема n из распределения, имеющего среднее значение μ и дисперсию σ^2, то $\bar{x}$ является линейной оценкой для μ с наименьшей дисперсией.*

Можно показать, что это же утверждение справедливо, если $(x_1, \ldots, x_n)$ — выборка из конечной совокупности. Доказательство этого факта мы оставляем читателю.

(b) Квадратичная оценка с наименьшей дисперсией для дисперсии совокупности. При исследовании проблемы квадратичных оценок удобно рассмотреть сначала следующую теорему, содержащую общие результаты теории несмещенных оценок, полученную Халмошем (1946).

10.2.2. *Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности с к. ф. р. $F(x)$ и $g(x_1, \ldots, x_n)$ — некоторая статистика, имеющая среднее значение θ и конечную дисперсию. Пусть $(i_1, \ldots, i_n)$ — i-я во множестве (соответствующим образом занумерованном) всех $n!$ перестановок целых чисел $1, \ldots, n$ и пусть $g_i(x_1, \ldots, x_n) = g(x_{i_1}, \ldots, x_{i_n})$. Если $\bar{g} = \frac{1}{n!} \sum_{i=1}^{n!} g_i(x_1, \ldots, x_n)$, тогда $E\bar{g} = \theta$ и дисперсия $\bar{g}$ меньше, чем дисперсия $g(x_1, \ldots, x_n)$, если $g(x_1, \ldots, x_n)$ не является симметрической относительно $x_1, \ldots, x_n$ функцией. Если $g(x_1, \ldots, x_n)$ — симметрическая, то $\bar{g}$ совпадает с $g(x_1, \ldots \ldots, x_n)$.*

Для доказательства **10.2.2** сначала заметим, что так как $(x_1, \ldots, x_n)$ — случайная выборка из $F(x)$, то $Eg_i = \theta$, $i = 1, \ldots, n!$. Поэтому $E\bar{g} = \theta$. Дисперсия $\bar{g}$ дается выражением
$$\sigma^2(\bar{g}) = \mathcal{E}\left[\frac{1}{n!} \sum_{i=1}^{n!} g_i(x_1, \ldots, x_n)\right]^2 - \theta^2. \tag{10.2.5}$$

Но из неравенства Шварца следует, что
$$\mathcal{E}\left[\frac{1}{n!} \sum_i g_i(x_1, \ldots, x_n)\right]^2 \leqslant \mathcal{E}\left[\sum_i \left(\frac{1}{n!}\right)^2\right] \cdot \mathcal{E}\left[\sum_i g_i^2(x_1, \ldots, x_n)\right], \tag{10.2.6}$$

причем знак равенства имеет место тогда и только тогда, если

$$g_i(x_1, \ldots, x_n) = \frac{C}{n!}, \qquad i = 1, \ldots, n!, \tag{10.2.7}$$

для всех точек $(x_1, \ldots, x_n)$ в выборочном пространстве R_n (за исключением множества меры нуль), где $C = C(x_1, \ldots, x_n)$ — симметрическая функция наблюдений. Это условие фактически означает, что $g(x_1, \ldots, x_n)$ — симметрическая функция относительно $(x_1, \ldots, x_n)$. Отсюда утверждение **10.2.2** сразу же следует. Следствие **10.2.2a**, которое мы ниже сформулируем, утверждает, что при определенных, довольно слабых ограничениях выборочная дисперсия есть квадратичная оценка с наименьшей дисперсией для дисперсии σ^2 распределения совокупности. Доказательство этого следствия вполне аналогично доказательству **10.2.2** и требует небольших обобщений, которые мы оставляем читателю.

10.2.2a. *Если $(x_1, \ldots, x_n)$ — выборка из совокупности $F(x)$, имеющей среднее значение μ и дисперсию σ^2, а также конечные третий и четвертый моменты, и если $\{a_{\xi\eta}\}$ — некоторый набор постоянных, для которых*

$$Q = \sum_{\xi, \eta = 1}^{n} a_{\xi\eta}(x_\xi - \bar{x})(x_\eta - \bar{x})$$

имеет среднее значение σ^2, то значения величин $\{a_{\xi\eta}\}$, для которых Q имеет минимальную дисперсию, определены выражениями

$$a_{\xi\xi} = \frac{1}{(n-1)}, \quad \xi = 1, \ldots, n; \quad a_{\xi\eta} = 0, \quad \xi \neq \eta.$$

В этом случае Q сводится к выборочной дисперсии s^2.

Читатель заметит, что **10.2.2** имеет место, если предположение о том, что $(x_1, \ldots, x_n)$ есть выборка из $F(x)$, заменить на предположение о том, что $(x_1, \ldots, x_n)$ — векторная случайная величина, функция распределения которой симметрична относительно $x_1, \ldots, x_n$. На основании этого обобщения **10.2.2** можно увидеть, что справедлива следующая версия **10.2.2a**, утверждающая, что при выборе из конечной совокупности выборочная дисперсия s^2 является несмещенной оценкой с минимальной дисперсией для дисперсии σ^2-совокупности.

10.2.2b. *Если $(x_1, \ldots, x_n)$ — выборка из конечной совокупности, которая имеет дисперсию σ^2, и если $\{a_{\xi\eta}\}$ — набор постоянных, для которых*

$$Q = \sum_{\xi, \eta = 1}^{n} a_{\xi\eta}(x_\xi - \bar{x})(x_\eta - \bar{x})$$

имеет среднее значение σ^2, то величины $\{a_{\xi\eta}\}$, для которых Q имеет минимальную дисперсию, определяются согласно **10.2.2a**, *а Q в этом случае сводится к выборочной дисперсии s^2.*

(c) Интервальные оценки для μ и σ^2 в случае нормального распределения. Если сделать предположение о том, что $(x_1, \ldots, x_n)$ — выборка из нормального распределения $N(\mu, \sigma^2)$, то в этом специальном случае мы можем получить так называемые *интервальные оценки* μ и σ^2. Этот случай заслуживает особого внимания. Как мы знаем из **8.4.3**, $\frac{\sqrt{n}(\bar{x}-\mu)}{s}$ имеет распределение Стьюдента $S(n-1)$ и, следовательно,

$$P\left\{t_1 < \frac{\sqrt{n}\,(\bar{x}-\mu)}{s} < t_2\right\} = \int_{t_1}^{t_2} f_{n-1}(t)\,dt = \gamma, \tag{10.2.8}$$

где $f_{n-1}(t)$ задается выражением (7.8.4), а t_1 и t_2 выбраны так, что интеграл в (10.2.8) имеет величину γ. Но утверждение

$$P\left\{t_1 < \frac{\sqrt{n}\,(\bar{x}-\mu)}{s} < t_2\right\} = \gamma$$

эквивалентно утверждению

$$P\left\{\bar{x} - t_2\frac{s}{\sqrt{n}} < \mu < \bar{x} - t_1\frac{s}{\sqrt{n}}\right\} = \gamma. \tag{10.2.9}$$

Таким образом, $\left(\bar{x} - t_2\frac{s}{\sqrt{n}},\ \bar{x} - t_1\frac{s}{\sqrt{n}}\right)$ является наблюдаемым случайным интервалом, причем вероятность того, что этот интервал включает точку μ, равна γ. Этот интервал называется 100γ-процентным *доверительным интервалом* для параметра μ, а γ называется *доверительным коэффициентом*. Этот пример указывает идею метода оценивания параметра с помощью (наблюдаемых) случайных интервалов, при заданной вероятности попадания истинного значения параметра в этот интервал. Такие интервалы часто называются *интервальными оценками*. Мы отложим дальнейшее обсуждение интервального оценивания при более общих условиях до глав 11 и 12.

В частном случае, рассмотренном выше, длина интервала равна $\frac{(t_2-t_1)\,s}{\sqrt{n}}$ и ее среднее значение минимально, если, при заданном γ, t_1 и t_2 выбраны так, что $t_2 = -t_1 = t_{n-1,\,\gamma}$, где $t_{n-1,\,\gamma}$ удовлетворяет условию

$$\int_{-t_{n-1,\,\gamma}}^{t_{n-1,\,\gamma}} f_{n-1}(t)\,dt = \gamma. \tag{10.2.10}$$

В этом случае 100γ-процентный доверительный интервал для μ представляет собой интервал

$$\bar{x} \pm t_{n-1,\,\gamma}\frac{s}{\sqrt{n}}, \tag{10.2.11}$$

центрированный относительно $\bar{x}$.

Аналогично, если $(x_1, \ldots, x_n)$ — выборка из совокупности с распределением $N(\mu, \sigma^2)$, то мы можем построить интервальную оценку для σ^2, используя тот факт, что $\frac{(n-1)s^2}{\sigma^2}$ имеет χ^2-распределение $C(n-1)$ (см. 8.4.2). На основании этого

$$P\left\{\chi_1^2 < \frac{(n-1)s^2}{\sigma^2} < \chi_2^2\right\} = \int_{\chi_1^2}^{\chi_2^2} dF_{n-1}(\chi^2) = \gamma, \qquad (10.2.12)$$

где $dF_{n-1}(\chi^2)$ дается выражением (7.8.1). Но (10.2.12) эквивалентно тому, что

$$P\left\{\frac{(n-1)s^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_1^2}\right\} = \gamma,$$

отсюда очевидно, что $((n-1)s^2/\chi_2^2, (n-1)s^2/\chi_1^2)$ есть 100γ-процентный доверительный интервал для σ^2. Вообще говоря, существует много способов выбора χ_1^2 и χ_2^2, для которых имеет место (10.2.12). На практике обычно это делают так, чтобы

$$\int_0^{\chi_1^2} dF_{n-1}(\chi^2) = \int_{\chi_2^2}^{\infty} dF_{n-1}(\chi^2) = \frac{1-\gamma}{2}, \qquad (10.2.13)$$

хотя доверительный интервал с наименьшей средней длиной получается, если выбрать $\frac{1}{\chi_1^2} - \frac{1}{\chi_2^2}$ минимальным при условии (10.2.12).

10.3. Оценивание параметров в линейном регрессионном анализе

(а) Оценки коэффициентов регрессии. Рассмотрим теперь обобщение теоремы **10.2.1**, связанное с задачами оценивания в регрессионном анализе, планировании экспериментов и др. Пусть $y_1, \ldots, y_n$, $n > k$, — независимые случайные величины, дисперсии которых одинаковы и равны σ^2, а средние значения задаются *функцией регрессии*

$$\mathscr{E}(y_\xi) = \beta_1 x_{1\xi} + \ldots + \beta_k x_{k\xi}, \qquad (10.3.1)$$

$\xi = 1, \ldots, n$, где $(x_{1\xi}, \ldots, x_{k\xi})$ — известные (вещественные) векторы, $\beta_1, \ldots, \beta_k$ — неизвестные (вещественные) постоянные, называемые *коэффициентами регрессии* и подлежащие оцениванию. Параметр σ^2, обычно неизвестный, называют *остаточной дисперсией*. Векторы $(x_{1\xi}, \ldots, x_{k\xi})$, $\xi = 1, \ldots, n$, можно рассматривать как n частных значений вектора фиксированных переменных $(x_1, \ldots, x_k)$. Обычно полагают $x_{1\xi} = 1$, $\xi = 1, \ldots, n$, но мы здесь не будем вводить это предположение. Покажем, что при некоторых слабых условиях существует как оценка $\beta_1, \ldots, \beta_k$ с минимальной дисперсией, так и довольно простая несмещенная оценка параметра σ^2.

Обратимся сначала к оцениванию параметров $\beta_1, \ldots, \beta_k$. Пусть $\mathscr{E}^{-1}(\beta_i)$ — произвольная линейная несмещенная оценка β_i, т. е.

$$\mathscr{E}^{-1}(\beta_i) = \sum_\xi c_{i\xi} y_\xi, \qquad i = 1, \ldots, k. \tag{10.3.2}$$

Тогда

$$\mathscr{E}(\mathscr{E}^{-1}(\beta_i)) = \sum_\xi \sum_j \beta_i c_{i\xi} x_{j\xi}, \tag{10.3.3}$$

откуда ясно, что $c_{i\xi}$ должны удовлетворять условию

$$\sum_\xi c_{i\xi} x_{j\xi} = \delta_{ij}, \tag{10.3.4}$$

где δ_{ij} — символ Кронекера. Дисперсия $\mathscr{E}^{-1}(\beta_i)$ равна

$$\sigma^2(\mathscr{E}^{-1}(\beta_i)) = \sum_\xi c_{i\xi}^2 \cdot \sigma^2. \tag{10.3.5}$$

Минимизируя $\sigma^2(\mathscr{E}^{-1}(\beta_i))$ относительно $c_{i\xi}$ при условии (10.3.4), найдем, что $c_{i\xi}$ должны иметь вид

$$c_{i\xi} = \sum_j \lambda_{ij} x_{j\xi}, \qquad i = 1, \ldots, k;\ \xi = 1, \ldots, n, \tag{10.3.6}$$

где λ_{ij} определяются уравнениями

$$\sum_{j'} \lambda_{ij'} a_{j'j} = \delta_{ij} \tag{10.3.7}$$

и где

$$a_{j'j} = \sum_\xi x_{j'\xi} x_{j\xi}. \tag{10.3.8}$$

Если матрица $\| a_{ij} \|$ неособая, что имеет место тогда и только тогда, когда векторы $(x_{1\xi}, \ldots, x_{k\xi})$, $\xi = 1, \ldots, n$, линейно независимы, то решением системы (10.3.7) будет

$$\| \lambda_{ij} \| = \| a_{ij} \|^{-1} = \| a^{ij} \|. \tag{10.3.9}$$

Следовательно, оценка с минимальной дисперсией для β_i, которую мы будем обозначать b_i, равна

$$b_i = \sum_j a^{ij} a_{j0}, \tag{10.3.10}$$

где случайные величины a_{j0}, $j = 1, \ldots, k$, определяются равенством

$$a_{j0} = \sum_\xi x_{j\xi} y_\xi = a_{0j}. \tag{10.3.11}$$

Чтобы найти дисперсию b_i, подставим значения λ_{ij} из (10.3.9) в (10.3.6) и полученные значения $c_{i\xi}$ в (10.3.5). Это дает

$$\sigma^2(b_i)=\sum_{j,\,j'=1}^{k} a^{ij}a^{ij'}a_{j'j}\cdot\sigma^2=\sigma^2\sum_j a^{ij}\delta_{ij}=a^{ii}\cdot\sigma^2. \qquad (10.3.12)$$

Аналогичным образом найдем, что матрица ковариаций оценок b_i, $i=1, \ldots, k$, равна

$$\| a^{ij}\sigma^2 \|. \qquad (10.3.13)$$

В итоге мы приходим к теореме Маркова (1900).

10.3.1. *Пусть y_ξ, $\xi=1, \ldots, n$, — независимые случайные величины со средними значениями $\sum_{i=1}^{k}\beta_i x_{i\xi}$, $\xi=1, \ldots, n>k$, и одинаковыми дисперсиями σ^2, $a(x_{1\xi}, \ldots, x_{k\xi})$, $\xi=1, \ldots, n$, — известные линейно независимые векторы. Тогда линейные оценки b_i, $i=1, \ldots, k$, с минимальной дисперсией коэффициентов регрессии β_i даются формулой* (10.3.10). *Матрица ковариаций b_i равна $\| a^{ij}\sigma^2\|$, где $a_{ij}=\sum_\xi x_{i\xi}x_{j\xi}$, $i, j=1, \ldots, k$, и $\| a^{ij}\|=\| a_{ij}\|^{-1}$.*

Оценки наименьших квадратов для $\beta_1, \ldots, \beta_k$ определяются как значения $\beta_1, \ldots, \beta_k$, минимизирующие сумму квадратов

$$Q=\sum_\xi (y_\xi-\beta_1 x_{1\xi}-\ldots-\beta_k x_{k\xi})^2. \qquad (10.3.14)$$

Если матрица $\| a_{ij}\|$ неособая, то оценки наименьших квадратов совпадают с b_i. Это утверждение доказывается непосредственно, и мы оставляем доказательство читателю. Имеем, таким образом, следующую теорему о линейных оценках коэффициентов регрессии.

10.3.2. *В условиях теоремы* **10.3.1** *линейные оценки коэффициентов регрессии с минимальной дисперсией совпадают с оценками наименьших квадратов.*

Подход к оцениванию параметров β_i, основанный на минимизации дисперсии, принадлежит Маркову (1900), в то время как метод наименьших квадратов появился значительно раньше, у Гаусса (1809а). Комбинация теорем **10.3.1** и **10.3.2** обычно называется *теоремой Гаусса — Маркова.*

Во многих задачах регрессии $x_{1\xi}=1$, $\xi=1, \ldots, n$, т. е. средние значения величин y_ξ предполагаются равными $\beta_1+\beta_2 x_{2\xi}+\ldots+\beta_k x_{k\xi}$. Применяя теорему **10.3.2**, найдем, что в этом случае линейная оценка b_i параметра β_i с минимальной дисперсией имеет вид

$$\begin{aligned} b_1&=\bar{y}-b_2\bar{x}_2-\ldots-b_k\bar{x}_k,\\ b_{i'}&=\sum_{j'=2}^{k} A^{i'j'}\cdot A_{j'0}, \quad i'=2, \ldots, k, \end{aligned} \qquad (10.3.15)$$

где

$$\bar{y}=\frac{1}{n}\sum_{\xi} y_{\xi}, \quad \bar{x}_{i'}=\frac{1}{n}\sum_{\xi} x_{i'\xi}, \qquad i'=2, \ldots, k,$$

и

$$A_{i'j'}=\sum_{\xi}(x_{i'\xi}-\bar{x}_{i'})(x_{j'\xi}-\bar{x}_{j'}), \qquad i', j'=2, \ldots, k,$$

$$A_{j'0}=\sum_{\xi}(x_{j'\xi}-\bar{x}_{j'})(y_{\xi}-\bar{y}), \tag{10.3.16}$$

$$\|A^{i'j'}\|=\|A_{i'j'}\|^{-1}. \tag{10.3.17}$$

Полагая в выражении для a_{1i} $x_{1\xi}=1$, $\xi=1, \ldots, n$, получим с помощью (10.3.13), что матрица ковариаций $b_{i'}$, $i'=2, \ldots, k$, равна

$$\|A^{i'j'}\sigma^2\|, \qquad i', j'=2, \ldots, k, \tag{10.3.18}$$

а

$$\sigma^2(b_1)=\left[\frac{1}{n}+\sum_{i', j'=2}^{k} A^{i'j'}\bar{x}_{i'}\bar{x}_{j'}\right]\sigma^2, \tag{10.3.19}$$

$$\sigma(b_1, b_{i'})=\left[\sum_{j'=2}^{k} A^{i'j'}\bar{x}_{j'}\right]\sigma^2.$$

Замечания о теореме Гаусса — Маркова и задача о взвешивании. Предположим, что величины $\beta_1, \ldots, \beta_k$ из (10.3.1) представляют собой «истинные» веса k предметов $o_1, \ldots, o_k$. Рассмотрим взвешивание различных комбинаций этих предметов на точных весах (представляющих собой две чашки для взвешивания, левую и правую, и шкалу, причем показание шкалы равно весу предметов на левой чашке минус вес предметов на правой). Пусть $x_{i\xi}=+1$, если предмет o_i положили на левую чашку, $x_{i\xi}=-1$, если o_i положили на правую, и $x_{i\xi}=0$, если o_i не участвует в ξ-м взвешивании. Тогда $\beta_1 x_{1\xi}+\ldots+\beta_k x_{k\xi}$ есть «истинное» показание шкалы при ξ-м взвешивании, если распределение предметов задается числами $x_{1\xi}, \ldots, x_{k\xi}$. Для N различных взвешиваний получается *матрица планирования взвешиваний* $\|x_{i\xi}\|$, $i=1, \ldots, k$, $\xi=1, \ldots, N$, элементы которой суть -1, 0, $+1$.

Если шкала не имеет смещения, то «фактическое» показание ее y_ξ при ξ-м взвешивании можно рассматривать как случайную величину со средним значением $\beta_1 x_{1\xi}+\ldots+\beta_k x_{k\xi}$. Если предположить $y_1, \ldots, y_N$ независимыми случайными величинами с одинаковыми дисперсиями σ^2, то оценки минимальной дисперсии $b_1, \ldots, b_k$ «истинных» весов $\beta_1, \ldots, \beta_k$ предметов $o_1, \ldots, o_k$ определятся формулой (10.3.10), а матрица ковариаций этих оценок — равенством (10.3.13). Отметим, что k является наименьшим значением N, для которого вообще можно построить матрицу планирования взвешиваний, позволяющую получить оценки для всех $\beta_1, \ldots, \beta_k$. Задача нахождения таких матриц планирования взвешиваний, которые приводят к «наилучшим» в различных смыслах оценкам для $\beta_1, \ldots, \beta_k$, подробно исследовалась Хотеллингом (1944), Кишеном (1945), Мудом (1946) и другими.

(b) Оценка остаточной дисперсии σ^2. Рассмотрим задачу построения несмещенной квадратичной оценки остаточной дисперсии σ^2. Пусть

$$z_\xi = y_\xi - \sum_{i=1}^{k} \beta_i x_{i\xi}, \tag{10.3.20}$$

$$\tilde{y}_\xi = \sum_{i=1}^{k} b_i x_{i\xi}, \tag{10.3.21}$$

$$S = \sum_{\xi=1}^{n} z_\xi^2, \qquad S_1 = \sum_{\xi=1}^{n} (y_\xi - \tilde{y}_\xi)^2,$$

$$S_2 = \sum_{i,\, j=1}^{k} a_{ij} (b_i - \beta_i)(b_j - \beta_j). \tag{10.3.22}$$

Имеем

$$S = S_1 + S_2. \tag{10.3.23}$$

Но

$$\mathfrak{E}S = n\sigma^2 \tag{10.3.24}$$

и, обращаясь к матрице ковариаций (10.3.13), находим, что

$$\mathfrak{E}S_2 = \sum_{i,\, j=1}^{k} a^{ij} a_{ij} \sigma^2 = k\sigma^2. \tag{10.3.25}$$

Так как $\mathfrak{E}S = \mathfrak{E}S_1 + \mathfrak{E}S_2$, то

$$\mathfrak{E}S_1 = (n-k)\sigma^2. \tag{10.3.26}$$

Величина S_1 свободна от параметров $\beta_1, \ldots, \beta_k$ и σ^2 и, следовательно, может служить для оценивания. Итак,

10.3.3. *В условиях теоремы* **10.3.1** $\dfrac{S_1}{(n-k)}$ *является несмещенной оценкой* σ^2.

Средние значения S, S_1 и S_2 равны соответственно $n\sigma^2$, $(n-k)\sigma^2$ и $k\sigma^2$. Числа n, $n-k$, k называются *степенями свободы* S, S_1, S_2.

Преобразуя выражение (10.3.22) для S_1, получим

$$S_1 = a_{00} - \sum_{i,\, j} a_{ij} b_i b_j = a_{00} - \sum_{i,\, j} a^{ij} a_{i0} a_{j0}, \tag{10.3.27}$$

т. е. S_1 можно представить в сравнительно простой форме:

$$S_1 = \frac{|a_{i_0 j_0}|}{|a_{ij}|}, \tag{10.3.27a}$$

где $a_{00} = \sum_\xi y_\xi^2$ и

$$|a_{i_0 j_0}| = \begin{vmatrix} a_{00} & a_{01} & \ldots & a_{0k} \\ a_{10} & a_{11} & \ldots & a_{1k} \\ \cdot & \cdot & \cdot & \cdot \\ a_{k0} & a_{k1} & \ldots & a_{kk} \end{vmatrix}. \tag{10.3.28}$$

Эквивалентность (10.3.27) и (10.3.27а) легко доказать с помощью окаймляющего разложения определителя (10.3.28) по первому столбцу и первой строке [см., например, Бохер (1929)].

Если некоторые из коэффициентов регрессии, например β_{k_1+1}, β_{k_1+2}, ..., β_k, известны, то в предыдущих параграфах можно заменить y_ξ на y'_ξ, где

$$y'_\xi = y_\xi - \beta_{k_1+1} x_{k_1+1,\,\xi} - \ldots - \beta_k x_{k\xi},$$

и затем повторить весь анализ, заменив k на k_1. При этом получатся очевидные модификации теорем **10.3.1**, **10.3.2** и **10.3.3**.

(с) Распределение оценок в теории нормальной регрессии. Сделаем следующее допущение, что y_ξ, $\xi = 1, \ldots, n$, независимы и нормальны $N(\beta_1 x_{1\xi} + \ldots + \beta_k x_{k\xi}, \sigma^2)$. Мы получим результаты, весьма важные в прикладной статистике. Пусть случайные величины z_ξ определяются согласно (10.3.20). Тогда элемент вероятности z_ξ, $\xi = 1, \ldots$ ..., n, равен

$$dF(z_1, \ldots, z_n) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_\xi z_\xi^2\right) dz_1 \ldots dz_n. \quad (10.3.29)$$

Из (10.3.22) найдем, что

$$S_1 = \sum_\xi \left[z_\xi - \sum_{i,\,j=1}^{k} a^{ij}\left(\sum_{\eta=1}^{n} x_{j\eta} z_\eta\right) x_{i\xi}\right]^2, \quad (10.3.30)$$

$$S_2 = \sum_{\xi,\,\eta} \sum_{i,\,j} a^{ij} x_{i\xi} x_{j\eta} z_\xi z_\eta,$$

так что S_1 и S_2 суть квадратичные формы переменных $z_1, \ldots, z_n$, матрицы которых имеют ранги $n-k$ и k. Поскольку S — квадратичная форма от $z_1, \ldots, z_n$ с матрицей ранга n, то по теореме Кочрена **8.4.4** $\frac{S_1}{\sigma^2}$ и $\frac{S_2}{\sigma^2}$ независимы и имеют χ^2-распределения $C(n-k)$ и $C(k)$ соответственно. Далее, из (10.3.10) и (10.3.11) имеем

$$b_i = \sum_\xi \sum_j a^{ij} x_{j\xi} y_\xi, \qquad i = 1, \ldots, k, \quad (10.3.31)$$

причем средние значения b_i равны β_i, а их матрица ковариаций задается (10.3.13). Следовательно, по теореме **7.4.4**, если матрица $\| a_{ij} \|$ неособая, то величины b_i имеют k-мерное нормальное распределение $N(\{\beta_i\}, \| a^{ij}\sigma^2 \|)$. При нашем предположении о нормальности $y_1, \ldots, y_n S_1$ и $(b_1, \ldots, b_k)$ оказываются независимыми. Это можно доказать вычислением характеристической функции $\mathscr{E}[\exp(\mathrm{i}tS_1 + \mathrm{i}(t_1 b_1 + \ldots$ $\ldots + t_k b_k))]$ случайного вектора $(S_1, b_1, \ldots, b_k)$, которая равна

$$(1 - 2\sigma^2 \mathrm{i}t)^{-\frac{1}{2}(n-k)} \exp\left\{-\frac{\sigma^2}{2}\sum_{i,\,j} a^{ij} t_i t_j + \mathrm{i}\sum_i \beta_i t_i\right\}. \quad (10.3.32)$$

Итак, мы имеем следующую важную теорему в теории нормальной регрессии.

10.3.4. *Если случайные величины* y_ξ, $\xi=1, \ldots, n$, *независимы и нормальны* $N(\beta_1 x_{1\xi}+\ldots+\beta_k x_{k\xi}, \sigma^2)$, $\xi=1, \ldots, n$, *матрица* $\| a_{ij} \|$, $i, j=1, \ldots, k$, *определенная в* (10.3.8), *неособая, то*

(i) *Определенные в* (10.3.10) *величины* b_i *являются линейными несмещенными оценками коэффициентов регрессии* β_i, $i=1, \ldots, k$, *и имеют* k*-мерное нормальное распределение* $N(\{\beta_i\}, \| a^{ij}\sigma^2 \|)$;

(ii) S_1 *и* $(b_1, \ldots, b_k)$ *независимы*;

(iii) $\frac{S_1}{\sigma^2}$ *и* $\frac{S_2}{\sigma^2}$ *независимы и имеют* χ^2*-распределения* $C(n-k)$ *и* $C(k)$.

10.4. Оценивание параметров в теории нормальной регрессии посредством интервалов и эллипсоидов

Так как в условиях теоремы **10.3.4** S_1 и $(b_1, \ldots, b_k)$ независимы, то S_1 и любая заданная величина b_i также независимы. Но b_i имеет распределение $N(b_i, a^{ii}\sigma^2)$, а $\frac{S_1}{\sigma^2}$ — χ^2-распределение $C(n-k)$. Следовательно, по теореме **7.8.3** величина

$$\frac{(b_i-\beta_i)\sqrt{n-k}}{\sqrt{a^{ii}S_1}} \tag{10.4.1}$$

имеет распределение Стьюдента $S(n-k)$, и с помощью рассуждений, аналогичных использованным при построении доверительного интервала (10.2.1), устанавливаем, что

$$b_i \pm t_{n-k,\gamma}\sqrt{\frac{a^{ii}S_1}{n-k}} \tag{10.4.2}$$

является 100γ-процентным доверительным интервалом для β_i, где $t_{n-k,\gamma}$ удовлетворяет условию (10.2.10) с заменой $n-1$ на $n-k$.

Если построить доверительные интервалы (10.4.2) для $i=1, \ldots, k$, то станет ясно, что среднее значение числа интервалов, накрывающих соответствующие значения β_i, равно γ_k. Но это мало что говорит о вероятности одновременного накрытия доверительными интервалами соответствующих параметров (за исключением случая независимости $b_1, \ldots, b_k$, что имеет место для диагональных матриц $\| a^{ij} \|$). Возникает вопрос: можно ли построить некоторую простую случайную область R_γ в k-мерном пространстве значений параметров, накрывающую параметрическую точку $(\beta_1, \ldots, \beta_k)$ с вероятностью γ? Такую область можно легко построить, исходя из того, что $\frac{S_1}{\sigma^2}$ и $\frac{S_2}{\sigma^2}$ независимы и имеют χ^2-распределения $C(n-k)$ и $C(k)$. Действительно, по теореме **7.8.5** величина

$$\frac{(n-k)S_2}{kS_1} \tag{10.4.3}$$

имеет распределение Снедекора $S(k,\ n-k)$. Подставляя выражение для S_2 из (10.3.22), получим

$$P\left\{\frac{n-k}{kS_1}\sum_{i,\ j=1}^{k} a_{ij}(b_i-\beta_i)(b_j-\beta_j)<F_{k,n-k,\gamma}\right\}=\\ =\int_0^{F_{k,\ n-k,\ \gamma}} dF_{k,\ n-k}(F), \tag{10.4.4}$$

где $dF_{k,\ n-k}(F)$ — элемент вероятности распределения Снедекора $S(k,\ n-k)$, указанный в (7.8.9). Но (10.4.4) можно записать в виде

$$P((\beta_1,\ \ldots,\ \beta_k)\in R_\gamma)=\gamma, \tag{10.4.5}$$

где R_γ — случайный эллипсоид в пространстве значений $(\beta_1,\ \ldots,\ \beta_k)$ с центром в точке $(b_1,\ \ldots,\ b_k)$, задаваемый уравнением

$$\sum_{i,\ j=1}^{k} a_{ij}(\beta_i-b_i)(\beta_j-b_j)=\frac{kS_1F_{k,\ n-k,\ \gamma}}{n-k}. \tag{10.4.6}$$

Этот эллипсоид является 100γ-*процентной доверительной областью* для $(\beta_1,\ \ldots,\ \beta_k)$. Таким же образом можно построить оценку посредством эллипсоида для части параметров $\beta_1,\ \ldots,\ \beta_k$.

Понятие доверительного эллипсоида было введено Хотеллингом (1931) в связи с предложенным им обобщенным распределением Стьюдента, которое рассматривается в главе 18.

10.5. Одновременное оценивание посредством доверительных интервалов: множественное сравнение

Выше мы видели, как использовать вектор выборочных коэффициентов регрессии $(b_1,\ \ldots,\ b_k)$ для построения доверительного эллипсоида для вектора $(\beta_1,\ \ldots,\ \beta_k)$ коэффициентов регрессии совокупности. Однако в некоторых задачах, в частности в задачах дисперсионного анализа, рассматриваемых в последующих параграфах, желательно строить доверительные интервалы, одновременно накрывающие несколько линейных комбинаций (с известными коэффициентами), составленных из компонент параметра. Эту задачу рассматривали Данкен (1952), Двасс (1959), Рой и Боуз (1953), Рой (1954), Шеффе (1953), Тьюки (1953) и другие.

(а) Неравенство для вероятности одновременного накрытия параметров доверительными интервалами. Получим прежде всего неравенство для вероятности одновременного накрытия h параметров соответствующими доверительными интервалами. Обозначим через $\mu_1,\ \ldots,\ \mu_h$ неизвестные параметры и через $(\underline{\mu}_1,\ \bar{\mu}_1),\ \ldots,\ (\underline{\mu}_h,\ \bar{\mu}_h)$ — случайные интервалы, каждый из которых накрывает свое значение

$\mu_1, \ldots, \mu_h$ с вероятностью $1-\frac{(1-\gamma)}{h}$. Пусть E_i — событие, состоящее в накрытии интервалом $(\underline{\mu}_i, \overline{\mu}_i)$ параметра μ_i, а $\overline{E}_i$ — дополнение к E_i, $i=1, \ldots, h$. Тогда

$$P(\overline{E}_i)=\frac{1}{h}(1-\gamma), \qquad i=1, \ldots, h.$$

Вероятность совместного осуществления событий $E_1, \ldots, E_h$ равна $P(E_1 \cap \ldots \cap E_h)$. Но

$$P(E_1 \cap \ldots \cap E_h)=1-P(\overline{E}_1 \cup \ldots \cup \overline{E}_h) \tag{10.5.1}$$

и

$$P(\overline{E}_1 \cup \ldots \cup \overline{E}_h) \leqslant P(\overline{E}_1)+\ldots+P(\overline{E}_h). \tag{10.5.2}$$

Тогда

$$P(E_1 \cap \ldots \cap E_h) \geqslant 1-[P(\overline{E}_1)+\ldots+P(\overline{E}_h)]. \tag{10.5.3}$$

Следовательно,

$$P(E_1 \cap \ldots \cap E_h) \geqslant \gamma. \tag{10.5.4}$$

Мы получили следующий результат Тьюки (1953):

10.5.1. *Пусть $\mu_1, \ldots, \mu_h$ — неизвестные параметры, а $(\underline{\mu}_1, \overline{\mu}_1), \ldots$ $\ldots, (\underline{\mu}_h, \overline{\mu}_h)$ — $100\left[1-\frac{1}{h}(1-\gamma)\right]$-процентные доверительные интервалы для $\mu_1, \ldots, \mu_h$ соответственно. Тогда с вероятностью, не меньшей γ, эти интервалы одновременно накрывают соответствующие параметры.*

Рассмотрим теперь одновременное накрытие большим числом доверительных интервалов при некоторых специальных условиях, выполненных в задачах модели I дисперсионного анализа.

(b) Метод Шеффе. Основной результат Шеффе (1953) может быть сформулирован в следующем виде.

10.5.2. *Пусть $(u_1, \ldots, u_k)$ — k-мерная случайная величина, имеющая распределение $N(\{\mu_i\}, \|a_{ij}\sigma^2\|)$, $i, j=1, \ldots, k$, где $\|a_{ij}\|$ — известная неособая симметричная матрица, σ^2 неизвестно. Пусть, далее, $\frac{v}{\sigma^2}$ — не зависящая от $(u_1, \ldots, u_k)$ случайная величина с χ^2-распределением с m степенями свободы, а $F_{k,m,\gamma}$ — 100γ-процентная точка распределения Снедекора $S(k, m)$ и $\delta=\sqrt{\frac{kvF_{k,m,\gamma}}{m}}$. Если $\mathscr{C}$ — множество всех вещественных векторов $(c_1, \ldots, c_k)$, кроме нулевого, то с вероятностью γ одновременно для всех векторов из $\mathscr{C}$ выполняется неравенство*

$$\sum_i c_iu_i-\delta\sqrt{\sum_{i,j} a_{ij}c_ic_j} \leqslant \sum_i c_i\mu_i \leqslant \sum_i c_iu_i+\delta\sqrt{\sum_{i,j} a_{ij}c_ic_j}. \tag{10.5.5}$$

Отметим, что в тривиальном случае равенства нулю всех $c_1, \ldots, c_k$ неравенство (10.5.5) выполнено с вероятностью 1.

Для доказательства теоремы **(10.5.2)** заметим, что $\frac{1}{\sigma^2}\sum_{i,j} a^{ij} \times$ $\times (u_i - \mu_i)(u_j - \mu_j)$ и $\frac{v}{\sigma^2}$ являются независимыми случайными величинами, имеющими χ^2-распределения с k и m степенями свободы соответственно. Поэтому

$$\frac{m}{kv}\sum_{i,j} a^{ij}(u_i - \mu_i)(u_j - \mu_j)$$

имеет распределение Снедекора $S(k, m)$. Следовательно, если $F_{k,m,\gamma}$ — 100γ-процентная точка этого распределения, то

$$P\left\{\sum_{i,j} a^{ij}(u_i - \mu_i)(u_j - \mu_j) < \delta^2\right\} = \gamma, \tag{10.5.6}$$

где

$$\delta^2 = \frac{kv}{m} F_{k,m,\gamma}.$$

В дальнейшем удобно пользоваться геометрической терминологией. Множество точек $(\mu_1, \ldots, \mu_k)$, для которых

$$\sum_{i,j} a^{ij}(\mu_i - u_i)(\mu_j - u_j) < \delta^2, \tag{10.5.7}$$

представляет собой внутренность 100γ-процентного доверительного эллипсоида с центром в точке $(u_1, \ldots, u_k)$ для параметра $(\mu_1, \ldots, \mu_k)$. Рассмотрим теперь множество точек $(\mu_1, \ldots, \mu_k)$, лежащих между всеми парами $(k-1)$-мерных параллельных гиперплоскостей, касательных к эллипсоиду (10.5.7). Это множество совпадает с внутренностью эллипсоида (10.5.7). Осталось показать только, что при любом выборе $(c_1, \ldots, c_k)$ в $\mathscr{C}$ параллельные гиперплоскости, задаваемые уравнениями

$$\sum_i c_i\mu_i = \sum_i c_i u_i \pm \delta \sqrt{\sum_{i,j} a_{ij}c_i c_j}, \tag{10.5.8}$$

касательны к эллипсоиду

$$\sum_{i,j} a^{ij}(\mu_i - u_i)(\mu_j - u_j) = \delta^2. \tag{10.5.9}$$

Очевидно, что всякая точка $(\mu_1, \ldots, \mu_k)$, лежащая между гиперплоскостями (10.5.8), удовлетворяет неравенству (10.5.5). Положим $\mu_i - u_i = y_i$. Тогда (10.5.9) запишется в виде

$$\sum_{i,j} a^{ij} y_i y_j = \delta^2, \tag{10.5.10}$$

а уравнение произвольной гиперплоскости в пространстве $(y_1, \ldots, y_k)$ как

$$\sum_i c_i y_i = d. \tag{10.5.11}$$

Мы должны найти два значения d, для которых гиперплоскость (10.5.11) касается эллипсоида (10.5.10). Вводя множитель Лагранжа λ, мы сводим задачу к нахождению в пространстве значений $(y_1, \ldots, y_k)$ стационарных точек функции

$$\Phi = \frac{1}{2}\lambda\left(\delta^2 - \sum_{i,j} a^{ij} y_i y_j\right) + \sum_i c_i y_i. \tag{10.5.12}$$

Дифференцируя по y_j, получим

$$-\lambda \sum_i a^{ij} y_i + c_j = 0$$

или

$$y_i = \frac{1}{\lambda} \sum_j a_{ij} c_j. \tag{10.5.13}$$

Подставляя это в (10.5.10), будем иметь

$$\lambda = \pm \frac{1}{\delta} \sqrt{\sum_{i,j} a_{ij} c_i c_j}. \tag{10.5.14}$$

Из (10.5.14), (10.5.13) и (10.5.11) находим

$$d = \pm \delta \sqrt{\sum_{i,j} a_{ij} c_i c_j}.$$

Подставляя эти значения d в (10.5.11) и вспоминая, что $y_i = \mu_i - u_i$, убеждаемся, что (10.5.8) представляют собой уравнения касательных гиперплоскостей для заданных $(c_1, \ldots, c_k)$.

Отметим, наконец, что если взять только конечное число N векторов из $\mathscr{C}$, то множество тех точек $(\mu_1, \ldots, \mu_k)$, которые лежат между *всеми* N парами гиперплоскостей, отвечающих выбранным N векторам, образует случайный многогранник G_k, описанный около эллипсоида (10.5.9). Поскольку G_k содержит эллипсоид, любое конечное число N неравенств, отвечающих N векторам из $\mathscr{C}$, выполняется одновременно с вероятностью, не меньшей γ. Например, если положить $c_i = 1$, $c_j = -1$, а все остальные компоненты равными нулю, $i > j = 1, \ldots, k$, получим подмножество из $N = \frac{k(k-1)}{2}$ векторов $\mathscr{C}$. Следовательно, с вероятностью, не меньшей γ, одновременно выполняются $\frac{k(k-1)}{2}$ неравенств

$$(u_i - u_j) - \delta\sqrt{a_{ii} + a_{jj} - 2a_{ij}} \leqslant \mu_i - \mu_j \leqslant$$
$$\leqslant (u_i - u_j) + \delta\sqrt{a_{ii} + a_{jj} - 2a_{ij}}, \qquad i > j = 1, \ldots, k. \tag{10.5.5a}$$

(c) Метод Тьюки. Результат, аналогичный (10.5.5а), но использующий в качестве оценки σ «стьюдентизованный размах», был получен Тьюки (1953). Пусть $(z_1, \ldots, z_k)$ — выборка объема k из совокупности $N(\mu, d^2\sigma^2)$, где d^2 — известная положительная постоянная, и $\frac{v}{\sigma^2}$ — случайная величина, не зависящая от $(z_1, \ldots, z_k)$ и имеющая χ^2-распределение с m степенями свободы. Обозначим через R размах выборки $(z_1, \ldots, z_k)$, т. е. $R = \max(z_1, \ldots, z_k) - \min(z_1, \ldots, z_k)$. Величина $\frac{\sqrt{m}R}{\sqrt{vd}}$ называется *стьюдентизованным размахом* $R_{k,m}$. Для заданного коэффициента доверия γ пусть $R_{k,m,\gamma}$ будет верхней 100γ-процентной точкой распределения $R_{k,m}$, т. е.

$$P(R_{k,m} \leqslant R_{k,m,\gamma}) = \gamma. \qquad (10.5.15)$$

Результат Тьюки (1953) формулируется следующим образом.

10.5.3. *Пусть $(u_1, \ldots, u_k)$ — k-мерная случайная величина с распределением $N(\{\mu_i\}, \|a_{ij}\sigma^2\|)$, где $a_{ii} = a^2$, $a_{ij} = a_{ji} = \rho a^2$, $i \neq j = 1, \ldots, k$, причем величина ρ известна.*

Пусть, далее, $\frac{v}{\sigma^2}$ — случайная величина, не зависящая от $(u_1, \ldots, u_k)$ и имеющая χ^2-распределение с m степенями свободы. Если $\mathscr{C}_0$ — множество вещественных векторов $(c_1, \ldots, c_k)$, для которых $\sum_i c_i = 0$, исключая нулевой вектор, то с вероятностью γ неравенства

$$\sum_i c_i u_i - D \leqslant \sum_i c_i \mu_i \leqslant \sum_i c_i u_i + D \qquad (10.5.16)$$

выполняются одновременно для всех векторов из $\mathscr{C}_0$, где

$$D = R_{k,m,\gamma} \sqrt{\frac{a^2 v(1-\rho)}{m}} \cdot \frac{1}{2} \sum_i |c_i|. \qquad (10.5.17)$$

Заметим, что в тривиальном случае равенства всех $c_1, \ldots, c_k$ нулю неравенство (10.5.16) выполняется с вероятностью 1. Для доказательства **10.5.3** положим $u_i - \mu_i = y_i$, $i = 1, \ldots, k$. Тогда $(y_1, \ldots, y_k)$ будет иметь распределение $N(\{0\}, \|a_{ij}\sigma^2\|)$, где a_{ij} определены в теореме **10.5.3.** Пусть $w = y_1 + \ldots + y_k$ и $z_i = y_i + hw$, где постоянная h удовлетворяет квадратному уравнению

$$h^2 k[1 + (k-1)\rho] + 2h[1 + (k-1)\rho] + \rho = 0 \qquad (10.5.18)$$

$\left(\text{причем } -\frac{1}{(k-1)} < \rho < 1,\right.$ так как матрица $\|a_{ij}\|$ положительно опре- $\left.\text{деленная}\right)$. Тогда $z_1, \ldots, z_k$ — независимые величины, одинаково распределенные по закону $N(0, \sigma^2 a^2(1-\rho))$. Рассмотрим неравенства

$$|z_i - z_j| \leqslant H, \qquad i, j = 1, \ldots, k. \qquad (10.5.19)$$

Эти неравенства выполняются совместно тогда и только тогда, когда размах R удовлетворяет неравенству

$$R \leqslant H. \qquad (10.5.20)$$

Вспоминая, что v/σ^2 не зависит от $(z_1, \ldots, z_k)$ и имеет χ^2-распределение с m степенями свободы, и рассматривая $(z_1, \ldots, z_k)$ как выборку объема k из совокупности $N(0, \sigma^2 a^2 (1-\rho))$, получаем, что $R\sqrt{\frac{m}{a^2 v(1-\rho)}}$ является стьюдентизованным размахом $R_{k,m}$. Пусть теперь $(c_1, \ldots, c_k)$ — произвольный вектор из $\mathscr{C}_0$. Если обозначить через $\sum_i{}'$ суммирование по всем значениям i, для которых $c_i > 0$, и через $\sum_i{}''$ — суммирование по всем значениям i, для которых $-c_i > 0$, то будем иметь

$$\sum_i{}' c_i = \sum_j{}'' (-c_j) = \frac{1}{2}\sum_i |c_i| = K. \tag{10.5.21}$$

Далее,

$$\begin{aligned}\sum_i c_i z_i = \sum_i{}' c_i z_i - \sum_j{}'' (-c_j) z_j = \\ = \frac{1}{K}\left[\sum_i{}' \sum_j{}'' c_i(-c_j) z_i - \sum_i{}' \sum_j{}'' c_i(-c_j) z_j\right] = \\ = \frac{1}{K}\sum_i{}' \sum_j{}'' c_i(-c_j)(z_i - z_j).\end{aligned}$$

Если $|z_i - z_j| \leqslant H$, $i, j = 1, \ldots, k$, то

$$\begin{aligned}\left|\sum_i c_i z_i\right| = \frac{1}{K}\left|\sum_i{}' \sum_j{}'' c_i(-c_j)(z_i - z_j)\right| \leqslant \\ \leqslant \frac{1}{K}\sum_i{}' \sum_j{}'' c_i(-c_j)|z_i - z_j| \leqslant H\cdot K.\end{aligned}$$

Следовательно,

$$\left|\sum_i c_i z_i\right| \leqslant H\cdot\left(\frac{1}{2}\sum_i |c_i|\right) \tag{10.5.22}$$

для всех векторов из $\mathscr{C}_0$. Но (10.5.22) выполнено для всех векторов из $\mathscr{C}_0$ тогда и только тогда, когда выполнено (10.5.19), которое в свою очередь выполняется одновременно с (10.5.20). Следовательно,

$$P\left(\left|\sum_i c_i z_i\right| \leqslant H\frac{1}{2}\sum_i |c_i| \text{ для всех } (c_1, \ldots, c_k) \in \mathscr{C}_0\right) = P(R \leqslant H). \tag{10.5.23}$$

Выше мы получили, что

$$R_{k,m} = \sqrt{\frac{R^2 m}{a^2 v(1-\rho)}} \tag{10.5.24}$$

представляет собой стьюдентизованный размах. Если мы выберем

$$H = R_{k,m,\gamma}\sqrt{\frac{a^2 v(1-\rho)}{m}}, \tag{10.5.25}$$

где $R_{k,m,\gamma}$ определена в (10.5.15), то получим

$$P\left\{R \leqslant R_{k,m,\gamma} \sqrt{\frac{a^2 v (1-\rho)}{m}}\right\} = \gamma, \tag{10.5.26}$$

и, следовательно,

$$P\left(\left|\sum_i c_i u_i\right| \leqslant D \text{ для всех } (c_1, \ldots, c_k) \in \mathscr{C}_0\right) = \gamma, \tag{10.5.27}$$

где D определена в (10.5.17). Но $\sum\limits_i c_i z_i = \sum\limits_i c_i (y_i + hw) = \sum\limits_i c_i (u_i - \mu_i)$, и поэтому

$$P\left(\left|\sum_i c_i u_i - \sum c_i \mu_i\right| \leqslant D \text{ для всех } (c_1, \ldots, c_k) \in \mathscr{C}_0\right) = \gamma. \tag{10.5.28}$$

Это как раз эквивалентно тому, что с вероятностью γ неравенства (10.5.16) выполняются одновременно для всех $(c_1, \ldots, c_k) \in \mathscr{C}_0$. Доказательство теоремы **10.5.3** закончено.

Следует отметить, что с вероятностью, не меньшей γ, неравенства (10.5.16) выполняются одновременно для любого конечного множества векторов из $\mathscr{C}_0$. В частности, с вероятностью, не меньшей γ, неравенства

$$(u_i - u_j) - R_{k,m,\gamma} \sqrt{\frac{a^2 v (1-\rho)}{m}} \leqslant \mu_i - \mu_j \leqslant (u_i - u_j) + \\ + R_{k,m,\gamma} \sqrt{\frac{a^2 v (1-\rho)}{m}}$$

выполняются одновременно для всех $i > j = 1, \ldots, k$. Это, конечно, эквивалентно тому, что

$$\left((u_i - u_j) \pm R_{k,m,\gamma} \sqrt{\frac{a^2 v (1-\rho)}{m}}\right)$$

являются доверительными интервалами для $(\mu_i - \mu_j)$ соответственно при всех $i > j = 1, \ldots, k$, одновременно накрывающими соответствующие разности $\mu_i - \mu_j$, $i > j = 1, \ldots, k$, с вероятностью, не меньшей γ.

Двасс (1959) рассмотрел обобщение задачи об одновременном накрытии интервалами, содержащее результаты Шеффе и Тьюки в качестве частных случаев. Следует отметить, что основная идея об одновременном накрытии доверительными интервалами в случае доверительной области для линии регрессии принадлежит Уоркингу и Хотеллингу (1929).

10.6. Линейный регрессионный анализ в предположении нормальности и планирование эксперимента

В этом параграфе регрессионный анализ, изложенный в § 10.3, будет применяться к довольно простому вероятностному описанию планирования эксперимента, известному как *Модель* I. Мы рассмотрим только три простых плана: *двухфакторный*, *трехфакторный*, *латинские квадраты*. Читатель, интересующийся более подробным рассмотрением этих планов, а также статистическим анализом других планов, должен обратиться к книгам Фишера (1935а), Кочрена и Кокса (1957), Грэйбилла (1961), Кемпсорна (1952), Манна (1949), Шеффе (1959). Таблицы планов эксперимента составлены Фишером и Иэйтсом (1938) и Китагавой и Митоме (1953). Путеводитель по литературе, посвященной планированию эксперимента, имеется в книге Гринвуда и Хартли (1961).

Линейный регрессионный анализ применяется также к так называемому *анализу поверхностей реакции*. Введением в этот круг вопросов могут служить работы Бокса (1952, 1954), Бокса и Хантера (1957), Бокса и Уилсона (1961).

(а) Полный двухфакторный план эксперимента. В эксперименте такого рода предполагается, что имеется множество rs независимых случайных величин $\{x_{\xi\eta};\ \xi=1, \ldots, r,\ \eta=1, \ldots, s\}$ с одинаковыми дисперсиями σ^2 и средними значениями

$$\mathscr{E}(x_{\xi\eta})=\mu+\mu_{\xi\cdot}+\mu_{\cdot\eta}, \tag{10.6.1}$$

где

$$\sum_{\xi=1}^{r}\mu_{\xi\cdot}=\sum_{\eta=1}^{s}\mu_{\cdot\eta}=0. \tag{10.6.2}$$

Из (10.6.1) видно, что для любых ξ и η $\mathscr{E}(x_{\xi\eta})$ оказывается при условии (10.6.2) линейной функцией $(r+s-1)$ коэффициентов регрессии μ, $\mu_{\xi\cdot}$, $\mu_{\cdot\eta}$ вида (10.3.1), причем все фиксированные переменные имеют одно из значений 0, 1, —1.

Такую схему удобно представить себе прямоугольной таблицей из r *строк* и s *столбцов*, причем r строк ассоциированы с r фиксированными *уровнями* $R_1, \ldots, R_r$ *фактора* R, а s столбцов — с s фиксированными *уровнями* $C_1, \ldots, C_s$ *фактора* C. Случайная величина $x_{\xi\eta}$ описывает *реакцию*, вызываемую комбинацией (R_ξ, C_η) факторов R и C; она называется *реагирующей* случайной величиной.

Замечание. Для иллюстрации этих идей рассмотрим следующий пример. Пусть имеется r операторов $R_1, \ldots, R_r$ и s машин $C_1, \ldots, C_s$, $s \geqslant r$, производящих изделия некоторого вида. Если каждый оператор может работать на каждой машине в течение рабочего дня, а $x_{\xi\eta}$ обозначает производительность оператора R_ξ на машине C_η, то мы имеем двухфакторный эксперимент, который может протекать в течение s рабочих дней.

Коэффициент регрессии μ в (10.6.1) называется *общим средним значением реакции*, $\mu_{\xi\cdot}$ — *дифференциальным эффектом*, вызываемым R_ξ, $\mu_{\cdot\eta}$ — *дифференциальным эффектом*, вызываемым $C_{\cdot\eta}$. Мы хотим найти линейные оценки с наименьшей дисперсией для μ, $\mu_{\xi\cdot}$, $\mu_{\cdot\eta}$, дисперсии этих оценок, а также оценку для σ^2.

Для множества случайных величин $\{x_{\xi\eta}\}$ определим средние $\bar{x}_{..}$, $\bar{x}_{\xi.}$, $\bar{x}_{.\eta}$ точно так же, как в (8.6.7). Далее, положим

$$\begin{aligned} &m=\bar{x}_{..}, \\ &m_{\xi.}=\bar{x}_{\xi.}-\bar{x}_{..},\ \xi=1,\ \ldots,\ r, \\ &m_{.\eta}=\bar{x}_{.\eta}-\bar{x}_{..},\ \eta=1,\ \ldots,\ s, \end{aligned} \tag{10.6.3}$$

и

$$\begin{aligned} &S=\sum_{\xi,\,\eta}(x_{\xi\eta}-\mu-\mu_{\xi.}-\mu_{.\eta})^2, \\ &S_{..}=\sum_{\xi,\,\eta}(x_{\xi\eta}-m-m_{\xi.}-m_{.\eta})^2,\quad S_{.0}(\mu_{\xi.})=\sum_{\xi,\,\eta}(m_{\xi.}-\mu_{\xi.})^2, \\ &S_{0.}(\mu_{.\eta})=\sum_{\xi,\,\eta}(m_{.\eta}-\mu_{.\eta})^2,\quad S_{00}(\mu)=\sum_{\xi,\,\eta}(m-\mu)^2. \end{aligned} \tag{10.6.4}$$

С помощью элементарных выкладок выводим, что

$$S=S_{..}+S_{.0}(\mu_{\xi.})+S_{0.}(\mu_{.\eta})+S_{00}(\mu). \tag{10.6.5}$$

Заметим, что $S_{..}$ совпадает с $S_{..}$ из (8.6.9), а $S_{.0}(0)=S_{.0}$, $S_{0.}(0)=S_{0.}$, где $S_{.0}$ и $S_{0.}$ также из формулы (8.6.9). Отметим также, что $S_{..}+S_{.0}+S_{0.}(0)=S_T$ из (8.6.9).

Из теоремы **10.3.1** мы знаем, что линейные оценки μ, $\mu_{\xi.}$ и $\mu_{.\eta}$ с минимальной дисперсией суть значения μ, $\mu_{\xi.}$ и $\mu_{.\eta}$, минимизирующие S в (10.6.4). Применяя **10.3.1** и **10.3.2**, убеждаемся, что процесс минимизации приводит соответственно к m, $m_{\xi.}$ и $m_{.\eta}$ в качестве линейных оценок для μ, $\mu_{\xi.}$ и $\mu_{.\eta}$ с наименьшей дисперсией. Из **10.3.3** следует, что $\dfrac{S_{..}}{(r-1)(s-1)}$ — несмещенная оценка σ^2.

Предоставляем читателю проверить, что матрица ковариаций оценок m, $m_{\xi.}$ и $m_{.\eta}$ имеет следующие элементы:

$$\begin{gathered} \sigma^2(m)=\frac{\sigma^2}{rs},\quad \sigma(m,\ m_{\xi.})=\sigma(m,\ m_{.\eta})=0, \\ \sigma^2(m_{\xi.})=\frac{r-1}{rs}\sigma^2,\quad \sigma^2(m_{.\eta})=\frac{s-1}{rs}\sigma^2, \\ \sigma(m_{\xi.},\ m_{\xi'.})=\sigma(m_{.\eta},\ m_{.\eta'})=-\frac{\sigma^2}{rs},\quad \xi\neq\xi',\ \eta\neq\eta', \\ \sigma(m_{\xi.},\ m_{.\eta})=0. \end{gathered} \tag{10.6.6}$$

Заменив в формулах для каждого элемента σ^2 на $\dfrac{S_{..}}{(r-1)(s-1)}$, получим несмещенные оценки элементов этой матрицы. Отметим, что ковариация любых двух оценок из различных множеств m, $\{m_{\alpha.},\ m_{.\beta}\}$ равна нулю.

Итак, получен следующий результат.

10.6.1. *Пусть* $\{x_{\xi\eta};\ \xi=1,\ \ldots,\ r;\ \eta=1,\ \ldots,\ s\}$ — *независимые случайные величины с одинаковой дисперсией* σ^2 *и средними* $\mathscr{E}x_{\xi\eta}=$

$=\mu+\mu_{\xi\cdot}+\mu_{\cdot\eta}$, $\sum_{\xi}\mu_{\xi\cdot}=\sum_{\eta}\mu_{\cdot\eta}=0$. *Линейные оценки для параметров* μ, $\mu_{\xi\cdot}$, $\mu_{\cdot\eta}$ *с наименьшей дисперсией равны* m, $m_{\xi\cdot}$, $m_{\cdot\eta}$ *соответственно, а* $\frac{S_{..}}{(r-1)(s-1)}$ *служит несмещенной оценкой* σ^2, *где* $S_{..}$ *определена в* (10.6.4). *Элементы матрицы ковариаций величин* m, $m_{\xi\cdot}$, $m_{\cdot\eta}$ *приведены в* (10.6.6).

Если сделать дополнительное предположение, что величины $x_{\xi\eta}$ имеют нормальные распределения $N(\mu+\mu_{\xi\cdot}+\mu_{\cdot\eta}, \sigma^2)$, то аналогично теореме **10.3.4** можно доказать следующий результат.

10.6.2. *Пусть* $\{x_{\xi\eta};\ \xi=1, \ldots, r;\ \eta=1, \ldots, s\}$ — *независимые случайные величины, распределенные по законам* $N(\mu+\mu_{\xi\cdot}+\mu_{\cdot\eta}, \sigma^2)$, *где* $\sum_{\xi}\mu_{\xi\cdot}=\sum_{\eta}\mu_{\cdot\eta}=0$. *Тогда* m, $\{m_{\xi\cdot}\}$, $\{m_{\cdot\eta}\}$ *представляют собой три независимых множества нормально распределенных случайных величин со средними* μ, $\{\mu_{\xi\cdot}\}$, $\{\mu_{\cdot\eta}\}$ *соответственно и матрицей ковариаций* (10.6.6). *Распределения величин* $\{m_{\xi\cdot}\}$ *и* $\{m_{\cdot\eta}\}$ — *вырожденные и сосредоточены на поверхностях* $\sum_{\xi}m_{\xi\cdot}=0$, $\sum_{\eta}m_{\cdot\eta}=0$. *Наконец,* $\frac{S_{..}}{\sigma^2}$, $\frac{S_{.0}(\mu_{\xi\cdot})}{\sigma^2}$, $\frac{S_{0.}(\mu_{\cdot\eta})}{\sigma^2}$, $\frac{S_{00}(\mu)}{\sigma^2}$ *являются независимыми случайными величинами, имеющими* χ^2*-распределения* $C((r-1)(s-1))$, $C(r-1)$, $C(s-1)$, $C(1)$ *соответственно.*

С помощью этой теоремы можно строить доверительные интервалы для параметров μ; $\mu_{1\cdot}, \ldots, \mu_{r\cdot}$; $\mu_{\cdot 1}, \ldots, \mu_{\cdot s}$. Например, 100 γ-процентные доверительные интервалы для μ, $\mu_{\xi\cdot}$ и $\mu_{\cdot\eta}$ имеют вид

$$\begin{gathered}
m\pm t_{(r-1)(s-1),\gamma}\sqrt{\frac{S_{..}}{rs(r-1)(s-1)}},\\
m_{\xi\cdot}\pm t_{(r-1)(s-1),\gamma}\sqrt{\frac{S_{..}}{rs(s-1)}},\\
m_{\cdot\eta}\pm t_{(r-1)(s-1),\gamma}\sqrt{\frac{S_{..}}{rs(r-1)}}
\end{gathered}\tag{10.6.7}$$

соответственно. Аналогично можно получить доверительные интервалы для разностей $\mu_{\xi\cdot}-\mu_{\xi'\cdot}$, $\xi\neq\xi'$, или для других линейных функций. С помощью теорем **10.5.2** или **10.5.3** можно строить доверительные интервалы для $\mu_{\xi\cdot}-\mu_{\xi'\cdot}$, $\xi>\xi'=1, \ldots, r$ (или для $\mu_{\cdot\eta}-\mu_{\cdot\eta'}$, $\eta>\eta'=1, \ldots, s$), одновременно накрывающие соответствующие разности с вероятностью, не меньшей γ.

Для одновременного оценивания двух и более параметров μ, $\mu_{1\cdot}, \ldots, \mu_{r\cdot}$, $\mu_{\cdot 1}, \ldots, \mu_{\cdot s}$ можно строить доверительные области, аналогичные (10.4.6). Особенно важен случай одновременного оценивания всех $\mu_{\xi\cdot}$ (или $\mu_{\cdot\eta}$). Согласно теореме **10.6.2** $\frac{S_{.0}(\mu_{\xi\cdot})}{\sigma^2}$ и $\frac{S_{..}}{\sigma^2}$ независимы и имеют χ^2-распределения $C(r-1)$ и $C((r-1)(s-1))$; из **7.8.5** тогда следует, что $\frac{(s-1)\,S_{.0}(\mu_{\xi.})}{S_{..}}$ имеет распределение Снедекора $S((r-1),$ $(r-1)(s-1))$. Пользуясь обозначением из (10.4.6), будем иметь

$$P\left(\frac{(s-1)\,S_{.0}(\mu_{\xi.})}{S_{..}} < F_{r-1,\,(r-1)(s-1),\,\gamma}\right) = \gamma. \tag{10.6.8}$$

Но это можно записать в виде

$$P((\mu_{1.}, \ldots, \mu_{r.}) \in R_\gamma) = \gamma, \tag{10.6.9}$$

где R_γ — $(r-1)$-мерная 100γ-процентная *доверительная сфера* для $(\mu_{1.}, \ldots, \mu_{r.})$, задаваемая уравнением

$$\sum_{\xi,\,\eta} (\mu_{\xi.} - m_{\xi.})^2 = \frac{S_{..}}{s-1} F_{r-1,\,(r-1)(s-1),\,\gamma} \tag{10.6.10}$$

при условии $\sum\limits_{\xi,\,\eta} (\mu_{\xi.} - m_{\xi.}) = 0$. Если мы специально интересуемся возможностью обращения всех $\mu_{\xi.}$ в нуль, то следует проверить, содержит ли сфера (10.6.10) точку $(0, \ldots, 0)$. Это сводится к проверке неравенства

$$\frac{(s-1)\,S_{.0}(0)}{S_{..}} < F_{r-1,\,(r-1)(s-1),\,\gamma}. \tag{10.6.11}$$

Если (10.6.11) выполнено, то говорят, что значения $\{x_{\xi\eta}\}$ *согласуются со статистической гипотезой* о равенстве всех $\mu_{\xi.}$ нулю на 100γ-процентном *доверительном уровне*. При выполнении (10.6.11) говорят также, что $m_{\xi.}$ (оценки для $\mu_{\xi.}$) *значимо не отличаются от нуля при* $100(1-\gamma)$-*процентном уровне значимости*. Существенно, что фигурирующая в (10.6.11) случайная величина $\frac{(s-1)\,S_{.0}(0)}{S_{..}}$ наблюдаема.

Аналогичным образом проверяется, что значения $\{x_{\xi\eta}\}$ согласуются со статистической гипотезой о равенстве нулю всех параметров $\mu_{.\eta}$.

Обычно составляющие полного двухфакторного плана Модели I в предположении нормальности собирают в *таблицу дисперсионного анализа*, помня, что $S_{.0}(0) = S_{.0}$ и $S_{0.}(0) = S_{0.}$, как показано в таблице 10.1.

Первое отношение Снедекора $F_{.0}$ служит для проверки гипотезы о том, что все $\mu_{\xi.}$ равны нулю, а второе — для проверки гипотезы о равенстве нулю всех $\mu_{.\eta}$.

Расположение составляющих дисперсионного анализа в форме таблицы 10.1 предложено Фишером (1925а), который разработал также теорию и приложения Модели I дисперсионного анализа.

(b) Полный трехфакторный план эксперимента. Идеи § 10.6 (а) непосредственно переносятся на планы более высоких порядков, т. е. на экспериментальные схемы с тремя и более факторами. Здесь мы имеем множество из rst независимых случайных величин $\{x_{\xi\eta\zeta};\ \xi =$

Таблица 10.1

Модель I дисперсионного анализа
Таблица полного двухфакторного плана эксперимента

Источник дисперсии	Степени свободы (с. с)	Сумма квадратов (с. к.)	Среднее суммы квадратов (с. с. к.)	Отношение Снедекора
Строки	$r-1$	$S_{\cdot 0}$	$\dfrac{S_{\cdot 0}}{r-1}$	$\dfrac{(s-1)\,S_{\cdot 0}}{S_{\cdot\cdot}}=F_{\cdot 0}$
Столбцы	$s-1$	$S_{0\cdot}$	$\dfrac{S_{0\cdot}}{s-1}$	$\dfrac{(r-1)\,S_{0\cdot}}{S_{\cdot\cdot}}=F_{0\cdot}$
Остаток (ошибка)	$(r-1)(s-1)$	$S_{\cdot\cdot}$	$\dfrac{S_{\cdot\cdot}}{(r-1)(s-1)}$	
Сумма	$rs-1$	S_T		

$=1, \ldots, r;\ \eta=1, \ldots, s;\ \zeta=1, \ldots, t\}$ с одинаковыми дисперсиями σ^2 и средними значениями

$$\mathscr{E} x_{\xi\eta\zeta}=\mu+\mu_{\xi\cdot\cdot}+\mu_{\cdot\eta\cdot}+\mu_{\cdot\cdot\zeta}+\mu_{\xi\eta\cdot}+\mu_{\xi\cdot\zeta}+\mu_{\cdot\eta\zeta}, \tag{10.6.12}$$

где

$$\sum_{\xi}\mu_{\xi\cdot\cdot}=\sum_{\eta}\mu_{\cdot\eta\cdot}=\sum_{\zeta}\mu_{\cdot\cdot\zeta}=0;\quad \sum_{\xi}\mu_{\xi\eta\cdot}=\sum_{\eta}\mu_{\xi\eta\cdot}=0,$$

$$\sum_{\xi}\mu_{\xi\cdot\zeta}=\sum_{\zeta}\mu_{\xi\cdot\zeta}=0;\quad \sum_{\eta}\mu_{\cdot\eta\zeta}=\sum_{\zeta}\mu_{\cdot\eta\zeta}=0.$$

Такую схему можно представить себе в виде r *строк*, s *столбцов* и t *слоев*, где r строк ассоциированы с r заданными уровнями $R_1, \ldots, R_r$ фактора R, s столбцов — с s уровнями $C_1, \ldots, C_s$ фактора C, а t слоев — с t заданными уровнями $L_1, \ldots, L_t$ фактора L. Постоянная $\mu_{\xi\cdot\cdot}$ представляет собой *дифференциальный эффект*, вызываемый уровнем R_ξ, аналогично интерпретируются $\mu_{\cdot\eta\cdot}$ и $\mu_{\cdot\cdot\zeta}$; $\mu_{\xi\eta\cdot}$ есть дифференциальный эффект, вызываемый сочетанием (R_ξ, C_η), или *взаимодействие* между R_ξ и C_η с аналогичной интерпретацией $\mu_{\xi\cdot\zeta}$ и $\mu_{\cdot\eta\zeta}$. Характеристики $\bar{x}_{\cdot\cdot\cdot}$, $\bar{x}_{\xi\cdot\cdot}$, $\bar{x}_{\cdot\eta\cdot}$, $\bar{x}_{\cdot\cdot\zeta}$, $\bar{x}_{\xi\eta\cdot}$, $\bar{x}_{\xi\cdot\zeta}$ и $\bar{x}_{\cdot\eta\zeta}$ определим вполне аналогично (8.6.7). Далее определим

$$\begin{gathered} m=\bar{x}_{\cdot\cdot\cdot}, \\ m_{\xi\cdot\cdot}=\bar{x}_{\xi\cdot\cdot}-\bar{x}_{\cdot\cdot\cdot}, \text{ и аналогично } m_{\cdot\eta\cdot} \text{ и } m_{\cdot\cdot\xi}, \\ m_{\xi\eta\cdot}=\bar{x}_{\xi\eta\cdot}-\bar{x}_{\xi\cdot\cdot}-\bar{x}_{\cdot\eta\cdot}+\bar{x}_{\cdot\cdot\cdot}, \text{ и аналогично } m_{\xi\cdot\zeta} \text{ и } m_{\cdot\eta\zeta}. \end{gathered} \tag{10.6.13}$$

Определим также

$$S=\sum_{\xi,\eta,\zeta}(x_{\xi\eta\zeta}-\mu-\mu_{\xi\cdot\cdot}-\mu_{\cdot\eta\cdot}-\mu_{\cdot\cdot\zeta}-\mu_{\xi\eta\cdot}-\mu_{\xi\cdot\zeta}-\mu_{\cdot\eta\zeta})^2,$$

$$S_{\cdot\cdot\cdot}=\sum_{\xi,\eta,\zeta}(x_{\xi\eta\zeta}-m-m_{\xi\cdot\cdot}-m_{\cdot\eta\cdot}-m_{\cdot\cdot\zeta}-m_{\xi\eta\cdot}-m_{\xi\cdot\zeta}-m_{\cdot\eta\zeta})^2$$

и

$$
\begin{gathered}
S_{\cdot\cdot 0}(\mu_{\xi\eta\cdot}) = \sum_{\xi,\,\eta,\,\zeta} (m_{\xi\eta\cdot} - \mu_{\xi\eta\cdot})^2 \text{ и аналогично } S_{\cdot 0\cdot}(\mu_{\xi\cdot\zeta}),\ S_{0\cdot\cdot}(\mu_{\cdot\eta\zeta}),\\
S_{\cdot 00}(\mu_{\xi\cdot\cdot}) = \sum_{\xi,\,\eta,\,\zeta} (m_{\xi\cdot\cdot} - \mu_{\xi\cdot\cdot})^2 \text{ и аналогично } S_{0\cdot 0}(\mu_{\cdot\eta\cdot}),\ S_{00\cdot}(\mu_{\cdot\cdot\zeta}),\\
S_{000}(\mu) = \sum_{\xi,\,\eta,\,\zeta} (m - \mu)^2.
\end{gathered} \tag{10.6.14}
$$

Отметим, что $S_{\cdots}$ совпадает с $S_{\cdots}$ из (8.6.31), а $S_{\cdot\cdot 0}(0) = S_{\cdot\cdot 0}$, $S_{\cdot 00}(0) = S_{\cdot 00}$ и т. д., где $S_{\cdot\cdot 0}$, $S_{\cdot 00}$ и т. д. определены в (8.6.31).

Имеем следующее разложение для S:

$$
\begin{aligned}
S = S_{\cdots} + S_{\cdot\cdot 0}(\mu_{\xi\eta\cdot}) + S_{\cdot 0\cdot}(\mu_{\xi\cdot\eta}) + S_{0\cdot\cdot}(\mu_{\cdot\eta\zeta}) + S_{\cdot 00}(\mu_{\xi\cdot\cdot}) +\\
+ S_{0\cdot 0}(\mu_{\cdot\eta\cdot}) + S_{00\cdot}(\mu_{\cdot\cdot\zeta}) + S_{000}(\mu).
\end{aligned} \tag{10.6.15}
$$

Из **10.3.1** получаем, что линейными оценками для μ, $\mu_{\xi\cdot\cdot}$, $\mu_{\cdot\eta\cdot}$, $\mu_{\cdot\cdot\zeta}$, $\mu_{\xi\eta\cdot}$, $\mu_{\xi\cdot\zeta}$, $\mu_{\cdot\eta\zeta}$ с минимальной дисперсией являются m, $m_{\xi\cdot\cdot}$, $m_{\cdot\eta\cdot}$, $m_{\cdot\cdot\zeta}$, $m_{\xi\eta\cdot}$, $m_{\xi\cdot\zeta}$, $m_{\cdot\eta\zeta}$ соответственно. Из **10.3.3** следует, что $\dfrac{S_{\cdots}}{[(r-1)(s-1)(t-1)]}$ — несмещенная оценка σ^2.

Для описания матрицы ковариаций оценок $m, \ldots, m_{\cdot\eta\zeta}$ достаточно привести только следующие ее ненулевые элементы, так как остальные ненулевые элементы этой матрицы получаются по соображениям симметрии:

$$
\begin{gathered}
\sigma^2(m) = \frac{\sigma^2}{rst}, \quad \sigma^2(m_{\xi\cdot\cdot}) = \frac{(r-1)\sigma^2}{rst},\\
\sigma(m_{\xi\cdot\cdot}, m_{\xi'\cdot\cdot}) = -\frac{\sigma^2}{rst}, \quad \xi \neq \xi'; \quad \sigma^2(m_{\xi\eta\cdot}) = \frac{(r-1)(s-1)\sigma^2}{rst},\\
\sigma(m_{\xi\eta\cdot}, m_{\xi\eta'\cdot}) = -\frac{(r-1)\sigma^2}{rst}, \quad \eta \neq \eta',\\
\sigma(m_{\xi\eta\cdot}, m_{\xi'\eta'\cdot}) = -\frac{\sigma^2}{rst}, \qquad \xi \neq \xi',\ \eta \neq \eta'.
\end{gathered} \tag{10.6.16}
$$

Ковариация любых двух величин из различных множеств $\{m\}$, $\{m_{\xi\cdot\cdot}\}$, $\{m_{\cdot\eta\cdot}\}$, $\{m_{\cdot\cdot\zeta}\}$, $\{m_{\xi\cdot\zeta}\}$, $\{m_{\cdot\eta\zeta}\}$ равна нулю.

Несмещенные оценки дисперсий и ковариаций из (10.6.16) получаются заменой в соответствующих формулах σ^2 на $\dfrac{S_{\cdots}}{[(r-1)(s-1)(t-1)]}$.

Формулировку теорем, аналогичных **10.6.1** и **10.6.2** для случая трехфакторного эксперимента, предоставляем читателю.

В связи с обобщением теоремы **10.6.2** на трехфакторный план эксперимента отметим, что если случайные величины $\{x_{\xi\eta\zeta}\}$ независимы и распределены по законам

$$
N(\mu + \mu_{\xi\cdot\cdot} + \mu_{\cdot\eta\cdot} + \mu_{\cdot\cdot\zeta} + \mu_{\xi\eta\cdot} + \mu_{\xi\cdot\zeta} + \mu_{\cdot\eta\zeta},\ \sigma^2), \tag{10.6.17}
$$

то величины

$$\frac{S_{\dots}}{\sigma^2},\ \frac{S_{\cdot\cdot 0}(\mu_{\xi\eta\cdot})}{\sigma^2},\ \frac{S_{\cdot 0\cdot}(\mu_{\xi\cdot\zeta})}{\sigma^2},\ \frac{S_{0\cdot\cdot}(\mu_{\cdot\eta\zeta})}{\sigma^2},$$
$$\frac{S_{\cdot 00}(\mu_{\xi\cdot\cdot})}{\sigma^2},\ \frac{S_{0\cdot 0}(\mu_{\cdot\eta\cdot})}{\sigma^2},\ \frac{S_{00\cdot}(\mu_{\cdot\cdot\zeta})}{\sigma^2},\ \frac{S_{000}(\mu)}{\sigma^2} \tag{10.6.18}$$

независимы и имеют χ^2-распределения с количеством степеней свободы

$$(r-1)(s-1)(t-1),\ (r-1)(s-1),\ (r-1)(t-1),$$
$$(s-1)(t-1),\ (r-1),\ (s-1),\ (t-1),\ 1 \tag{10.6.19}$$

соответственно. Для каждого из параметров $\mu, \dots, \mu_{\cdot\eta\zeta}$ отдельно можно строить доверительные интервалы, аналогичные (10.6.7), а также аналогичные (10.6.10) доверительные области для одновременного оценивания нескольких параметров из $\mu, \dots, \mu_{\cdot\eta\zeta}$. Используя теоремы **10.5.2** или **10.5.3**, можно строить доверительные интервалы для разностей

$$\{\mu_{\xi\cdot\cdot}-\mu_{\xi'\cdot\cdot};\ \xi>\xi'=1, \dots, r\},\ \{\mu_{\xi\eta\cdot}-\mu_{\xi'\eta\cdot};\ \xi>\xi'=1, \dots, r\},$$

η фиксировано, одновременно накрывающие соответствующие разности с вероятностью, не меньшей заданной.

Предоставляем читателю построение таблицы, аналогичной таблице 10.1, для полного трехфакторного плана.

(с) Схема латинского квадрата. Описанный выше полный трехфакторный план эксперимента требует наблюдения такого числа случайных величин, которое в некоторых приложениях недопустимо. Уменьшение количества наблюдений может быть достигнуто с помощью сбалансированных неполных трехфакторных планов. Простейший из них — и единственный, который будет рассмотрен, — план латинского квадрата, выбирающий сбалансированную неполную матрицу $\{x_{\xi\eta\zeta}\}^*$ реагирующих случайных величин из трехмерной матрицы $\{x_{\xi\eta\zeta}\}$, состоящей из r строк, r столбцов и r слоев.

В плане латинского квадрата рассматриваются три фактора R, C и L, каждый из которых имеет r уровней. Можно ассоциировать R_ξ с ξ-й *строкой*, C_η с η-м *столбцом* и L_ζ с ζ-м *слоем* трехмерной прямоугольной таблицы. Каждой ячейке этой таблицы соответствует случайная величина $x_{\xi\eta\zeta}$. Случайные величины $\{x_{\xi\eta\zeta}\}^*$ выбираются так, чтобы каждая комбинация значений (ξ, η) участвовала только один раз, точно так же по одному разу должны входить комбинации (ξ, ζ) и (η, ζ) (см. § 8.6 (d)). Обозначим выбранное множество из r^2 значений (ξ, η, ζ) через G^*.

Далее предполагается, что выбранные r^2 случайных величин $\{x_{\xi\eta\zeta}\}^*$ независимы и имеют нормальные распределения $N(Ex_{\xi\eta\zeta}, \sigma^2)$, причем для $(\xi, \eta, \zeta) \in G^*$

$$Ex_{\xi\eta\zeta}=\mu+\mu_{\xi\cdot\cdot}+\mu_{\cdot\eta\cdot}+\mu_{\cdot\cdot\zeta}, \tag{10.6.20}$$

где
$$\sum_{\xi}\mu_{\xi\cdot\cdot}=\sum_{\eta}\mu_{\cdot\eta\cdot}=\sum_{\zeta}\mu_{\cdot\cdot\zeta}=0. \qquad (10.6.21)$$

Сравнивая это с (10.6.12), мы видим, что предполагается равенство нулю всех попарных взаимодействий $\mu_{\xi\eta\cdot}$, $\mu_{\xi\cdot\zeta}$, $\mu_{\cdot\eta\zeta}$. План латинского квадрата дает относительно простые линейные оценки с минимальной дисперсией параметров из (10.6.20). Определим $\bar{x}^*_{\cdot\cdot\cdot}$, $\bar{x}^*_{\xi\cdot\cdot}$, $\bar{x}^*_{\cdot\eta\cdot}$, $\bar{x}^*_{\cdot\cdot\zeta}$ аналогично (8.6.40), а m^*, $m^*_{\xi\cdot\cdot}$, $m^*_{\cdot\eta\cdot}$, $m^*_{\cdot\cdot\zeta}$ определим через $\{x_{\xi\eta\zeta}\}^*$ точно так же, как m, $m_{\xi\cdot\cdot}$, $m_{\cdot\eta\cdot}$, $m_{\cdot\cdot\zeta}$ были определены в (10.6.4) через $\{x_{\xi\eta\zeta}\}$.

Заменим в (10.6.4) m, $m_{\xi\cdot\cdot}$, $m_{\cdot\eta\cdot}$, $m_{\cdot\cdot\zeta}$ на m^*, $m^*_{\xi\cdot\cdot}$, $m^*_{\cdot\eta\cdot}$, $m^*_{\cdot\cdot\zeta}$ соответственно и будем суммировать по $(\xi, \eta, \zeta)\in G^*$; полученные величины обозначим через $S^*_{\cdot 00}(\mu_{\xi\cdot\cdot})$, $S^*_{0\cdot 0}(\mu_{\cdot\eta\cdot})$, $S^*_{00\cdot}(\mu_{\cdot\cdot\zeta})$, $S^*_{000}(\mu)$. Положим
$$\begin{aligned} S^* &= \sum_{\xi,\eta,\zeta}{}^{*}(x_{\xi\eta\zeta}-\mu-\mu_{\xi\cdot\cdot}-\mu_{\cdot\eta\cdot}-\mu_{\cdot\cdot\zeta})^2,\\ S^*_{\cdot\cdot\cdot} &= \sum_{\xi,\eta,\zeta}{}^{*}(x_{\xi\eta\zeta}-m^*-m^*_{\xi\cdot\cdot}-m^*_{\cdot\eta\cdot}-m^*_{\cdot\cdot\zeta})^2, \end{aligned} \qquad (10.6.22)$$

где, как и в (8.6.40), $\sum\limits_{\xi,\eta,\zeta}^{*}$ означает суммирование по всем $(\xi, \eta, \zeta)\in \in G^*$.

Получаем теперь следующее разложение для S^*:
$$S^*=S^*_{\cdot\cdot\cdot}+S^*_{\cdot 00}(\mu_{\xi\cdot\cdot})+S^*_{0\cdot 0}(\mu_{\cdot\eta\cdot})+S^*_{00\cdot}(\mu_{\cdot\cdot\zeta})+S^*_{000}(\mu), \qquad (10.6.23)$$
где компоненты справа имеют соответственно $(r-1)(r-2)$, $(r-1)$, $(r-1)$, $(r-1)$, 1 степеней свободы.

Применяя **10.3.2**, находим, что m^*, $m^*_{\xi\cdot\cdot}$, $m^*_{\cdot\eta\cdot}$, $m^*_{\cdot\cdot\zeta}$ являются линейными оценками параметров μ, $\mu_{\xi\cdot\cdot}$, $\mu_{\cdot\eta\cdot}$, $\mu_{\cdot\cdot\zeta}$ с минимальной дисперсией, а $\dfrac{S^*_{\cdot\cdot\cdot}}{(r-1)(r-2)}$ несмещенно оценивает σ^2.

Ненулевые элементы матрицы ковариаций оценок m^*, $m^*_{\xi\cdot\cdot}$, $m^*_{\cdot\eta\cdot}$, $m^*_{\cdot\cdot\zeta}$ равны
$$\begin{gathered} \sigma^2(m^*)=\frac{\sigma^2}{r^2},\quad \sigma^2(m^*_{\xi\cdot\cdot})=\sigma^2(m^*_{\cdot\eta\cdot})=\sigma^2(m^*_{\cdot\cdot\zeta})=\frac{r-1}{r^2}\sigma^2, \qquad (10.6.24)\\ \sigma(m^*_{\xi\cdot\cdot}, m^*_{\xi'\cdot\cdot})=\sigma(m^*_{\cdot\eta\cdot}, m^*_{\cdot\eta'\cdot})=\sigma(m^*_{\cdot\cdot\zeta}, m^*_{\cdot\cdot\zeta'})=-\frac{\sigma^2}{r^2},\\ \xi\neq\xi',\quad \eta\neq\eta',\quad \zeta\neq\zeta'. \end{gathered}$$

Несмещенные оценки этих дисперсий получаются заменой σ^2 на $\dfrac{S^*_{\cdot\cdot\cdot}}{(r-1)(r-2)}$.

Формулирование утверждений, аналогичных **10.6.1** и **10.6 2**, для плана латинского квадрата предоставляется читателю, так же как и построение таблицы, аналогичной таблице 10.1.

10.7. Оценка компонент дисперсии по наблюдениям над линейными комбинациями

Во многих статистических задачах, встречающихся в планировании эксперимента, теории ошибок, психологическом факторном анализе, ситуация такова, что наблюдаемыми оказываются только линейные комбинации, но не сами случайные величины. С таким примером мы встречались в (8.6.3), где наблюдаемая случайная величина $x(u, v)$ представлялась линейной функцией от μ и ненаблюдаемых случайных величин $\varepsilon_{u\cdot}$, $\varepsilon_{\cdot v}$, ε_{uv}.

В таких ситуациях основной задачей является построение оценок для дисперсий (ненаблюдаемых) компонент по наблюдениям над линейными комбинациями этих компонент.

Пусть наблюдаемые случайные величины x_ξ, $\xi = 1, \ldots, n$, имеют вид

$$x_\xi = \mu + \sum_{g=1}^{m} a_{g\xi}\varepsilon_{g\xi}, \qquad \xi = 1, \ldots, n \geqslant m, \tag{10.7.1}$$

где

$$\begin{gathered} \mathscr{E}(\varepsilon_{g\xi}) = 0, \quad \mathscr{E}(\varepsilon_{g\xi}^2) = \sigma_g^2, \qquad g = 1, \ldots, m;\ \xi = 1, \ldots, n, \\ \mathscr{E}(\varepsilon_{g\xi}\varepsilon_{g\xi'}) = \mathscr{E}(\varepsilon_{g\xi}\varepsilon_{g'\xi}) = 0, \qquad g \neq g',\ \xi \neq \xi', \end{gathered} \tag{10.7.2}$$

$a_{g\xi}$ — известные постоянные, причем ранг матрицы $\|a_{g\xi}\|$ равен m. Не предполагается, что $\varepsilon_{g\xi}$ наблюдаемы, так что они могут быть функциями неизвестных параметров.

Ясно, что выборочное среднее $\bar{x}$ является несмещенной оценкой μ. Дисперсия ее равна

$$\sigma^2(\bar{x}) = \frac{1}{n^2}\sum_g\sum_\xi a_{g\xi}^2\sigma_g^2. \tag{10.7.3}$$

Далее, если s^2 — выборочная дисперсия, то

$$\mathscr{E}(s^2) = \frac{1}{n}\sum_g\sum_\xi a_{g\xi}^2\sigma_g^2. \tag{10.7.4}$$

Заметим, что s^2 является несмещенной оценкой для $\frac{1}{n}\sum_g\sum_\xi a_{g\xi}^2\sigma_g^2$; следовательно, $\frac{s^2}{n}$ будет несмещенной оценкой для $\sigma^2(\bar{x})$. Однако мы желаем оценить дисперсии $\sigma_1^2, \ldots, \sigma_m^2$ компонент по выборке $(x_1, \ldots, x_n)$. Для этого рассмотрим полную сумму квадратов

$$S_T = \sum_\xi (x_\xi - \bar{x})^2. \tag{10.7.5}$$

Справа стоит квадратичная форма от случайных величин $(x_1 - \bar{x}), \ldots, (x_n - \bar{x})$, которую можно представить в виде квадратичной формы от $(x_1 - \mu), \ldots, (x_n - \mu)$ с матрицей ранга $(n-1)$. Если $n - 1 \geqslant m$, то с помощью элементарной теории положительно полуопределенных квадратичных форм можно показать, что S_T (несколькими способами) представляется в виде

$$S_T = S_1 + \ldots + S_m, \tag{10.7.6}$$

где $S_1, \ldots, S_m$ — положительно полуопределенные квадратичные формы от $(x_1 - \bar{x}), \ldots, (x_n - \bar{x})$, которые можно представить в виде квадратичных форм от $(x_1 - \mu), \ldots, (x_n - \mu)$ с матрицами рангов $n_1, \ldots, n_m$ (все $n_h \geqslant 1$) соответственно, $n_1 + \ldots + n_m = n - 1$. Тогда для $h = 1, \ldots, m$

$$S_h = \sum_{\xi, \eta} A_{\xi\eta}^{(h)} (x_\xi - \mu)(x_\eta - \mu), \tag{10.7.7}$$

где $\| A_{\xi\eta}^{(h)} \|$ — симметричная матрица ранга n_h, элементы которой — известные числа. Подставляя значения $(x_\xi - \mu)$ из (10.7.1) в (10.7.7) и беря средние значения, получим

$$\mathscr{E}(S_h) = \sum_g B_{hg} \cdot \sigma_g^2, \tag{10.7.8}$$

где

$$B_{hg} = \sum_\xi A_{\xi\xi}^{(h)} a_{g\xi}^2. \tag{10.7.9}$$

Из свойства матрицы $\| A_{g\xi}^{(h)} \|$ и наших предположений о $\| a_{g\xi} \|$ следует, что $\| B_{hg} \|$ — неособая матрица. Применяя **10.1.1**, получаем из (10.7.8) следующие несмещенные оценки для $\sigma_1^2, \ldots, \sigma_m^2$:

$$\mathscr{E}^{-1}(\sigma_g^2) = \sum_h B^{gh} \cdot S_h. \tag{10.7.10}$$

Несмещенная оценка для $\sigma^2(\bar{x})$ получается заменой σ_g^2 в (10.7.3) на $\mathscr{E}^{-1}(\sigma_g^2)$, $g = 1, \ldots, m$. Суммируем теперь все полученное.

10.7.1. *Пусть* x_ξ, $\xi = 1, \ldots, n \geqslant m$, — *наблюдаемые случайные величины вида*

$$x_\xi = \mu + \sum_{g=1}^m a_{g\xi} \varepsilon_{g\xi},$$

где $\varepsilon_{g\xi}$ — *случайные величины, средние значения и ковариации которых равны нулю,* $\mathscr{E}(\varepsilon_{g\xi}^2) = \sigma_g^2$, *причем* μ *и* σ_g^2 — *неизвестные параметры,* $a_{g\xi}$ — *известные числа, и ранг матрицы* $\| a_{g\xi} \|$ *равен* m. *Если* $\bar{x}$ — *среднее значение выборки* $(x_1, \ldots, x_n)$, *а* $S_T, S_1, \ldots, S_m$ — *квадратичные формы величин* $(x_1 - \bar{x}), \ldots, (x_n - \bar{x})$, *определенные в* (10.7.5) *и* (10.7.6), *то* $\bar{x}$ — *несмещенная оценка для* μ, *а*

$$\mathscr{E}^{-1}(\sigma_g^2) = \sum_k B^{gh} \cdot S_h$$

являются несмещенными оценками для σ_g^2, $g = 1, \ldots, m$, *где*

$$\| B^{gh} \| = \Big\| \sum_\xi A_{\xi\xi}^{(h)} a_{g\xi}^2 \Big\|^{-1},$$

$\| A_{\xi\eta}^{(h)} \|$ — *матрица квадратичной формы* S_h.

Во многих задачах, где требуется оценить компоненту дисперсии, матрица $\|a_{g\xi}\|$ имеет специальную форму, из которой непосредственно получается разложение (10.7.6) для S_T.

В следующем параграфе мы рассмотрим некоторые приложения сформулированных выше принципов к планам эксперимента.

10.8. Оценки компонент дисперсии в планах эксперимента

В этом параграфе идеи оценивания компонент дисперсии мы применим к некоторым простым планам эксперимента. Будут рассмотрены только полный двухфакторный, сбалансированный неполный двухфакторный, полный трехфакторный планы и план латинского квадрата с точки зрения анализа компонент дисперсии. Читатель, интересующийся предметом более подробно, должен обратиться к книгам Кочрена и Кокса (1957), Грэйбилла (1961), Кемпсорна (1952) и Шеффе (1959). Мы рассмотрим только задачу построения несмещенных оценок компонент дисперсии, входящих в планы эксперимента, и не будем заниматься дисперсиями этих оценок. Последние изучались в статьях Хука (1956b), Шеффе (1959) и Тьюки (1956b, 1957a, 1957b). Доверительные интервалы для компонент дисперсии рассматривались Балмером (1957), Моригути (1954), Шеффе (1959) и другими.

Вероятностное описание планов эксперимента, основанных на наблюдаемых линейных функциях от случайных величин, иногда называют Моделью II в отличие от Модели I, базирующейся на теории нормальной регрессии и рассмотренной в § 10.6. Термины «Модель I» и «Модель II» для двух схем дисперсионного анализа введены Эйзенхартом (1947).

(а) Полный двухфакторный план эксперимента. Основные положения выборочного метода, требуемые для описания модели I полного двухфакторного плана эксперимента, содержатся в изложенной в § 8.6 (а) теории матричных (второго порядка) выборок. Мы будем пользоваться здесь обозначениями § 8.6 (а).

Поясним сначала основное различие между подходами к полному двухфакторному плану эксперимента, основанными на модели I и модели II. В модели I полного двухфакторного плана эксперимента, рассмотренной в § 10.4 (а), предполагалось, что уровни $R_1, \ldots, R_r$ фактора R и уровни $C_1, \ldots, C_s$ фактора C фиксированы для всех экспериментов, описываемых реагирующими случайными величинами $x_{\xi\eta}$. Но в некоторых ситуациях уровни фактора R или уровни фактора C, или те и другие вместе, представляют собой выборки из совокупностей уровней, которые могут быть конечными или бесконечными. В таких задачах схема регрессионного анализа (Модель I) не подходит. Мы рассмотрим случай, когда совокупности уровней бесконечны. При подходе, основанном на Модели II, мы предположим, что для любого уровня u фактора R и любого уровня v фактора C реагирующая случайная величина $x(u, v)$ имеет форму (8.6.3). Ее дисперсия дана в (8.6.4) или, в более коротких обозначениях, в (8.6.4а), т. е.

$$\sigma^2 = \sigma^2_{\cdot 0} + \sigma^2_{0\cdot} + \sigma^2_{\cdot\cdot}, \tag{10.8.1}$$

где $\sigma^2_{\cdot 0}$ — компонента *строки* (фактора R), $\sigma^2_{0\cdot}$ — компонента *столбца* (фактора C), $\sigma^2_{\cdot\cdot}$ — *остаточная компонента.*

Если r уровней фактора R и s уровней фактора C выбираются случайно, те и другие из бесконечной совокупности, то реагирующие случайные величины образуют матричную выборку $\{x_{\xi\eta};\ \xi = 1, \ldots, r;\ \eta = 1, \ldots, s\}$, определенную в § 8.6 (а), где $x_{\xi\eta}$ независимы и представляются в форме (8.6.8), т. е.

$$x_{\xi\eta} = \mu + \varepsilon_{\xi\cdot} + \varepsilon_{\cdot\eta} + \varepsilon_{\xi\eta}. \tag{10.8.2}$$

Здесь μ — среднее значение совокупности, $\varepsilon_{\xi\cdot}$, $\varepsilon_{\cdot\eta}$, $\varepsilon_{\xi\eta}$ — случайные величины с нулевыми средними и нулевыми ковариациями.

Дисперсия σ^2 величины $x_{\xi\eta}$ задана в (10.8.1). Основная задача анализа Модели I состоит в оценивании $\sigma^2_{\cdot 0}$, $\sigma^2_{0\cdot}$ и $\sigma^2_{\cdot\cdot}$ по выборке $\{x_{\xi\eta}\}$.

Для этого возьмем $\bar{x}_{\cdot\cdot}$, $\bar{x}_{\xi\cdot}$ и $\bar{x}_{\cdot\eta}$ из (8.6.7) и $S_{\cdot 0}$, $S_{0\cdot}$, $S_{\cdot\cdot}$ из (8.6.9). В условиях **8.6.2** согласно (8.6.11)

$$\begin{aligned} \mathscr{E}(S_{\cdot 0}) &= (r-1)(s\sigma^2_{\cdot 0} + \sigma^2_{\cdot\cdot}), \\ \mathscr{E}(S_{0\cdot}) &= (s-1)(r\sigma^2_{0\cdot} + \sigma^2_{\cdot\cdot}), \\ \mathscr{E}(S_{\cdot\cdot}) &= (r-1)(s-1)\sigma^2_{\cdot\cdot}. \end{aligned} \tag{10.8.3}$$

Применяя **10.1.1** (или **10.7.1**), получим следующие несмещенные оценки для $\sigma^2_{\cdot 0}$, $\sigma^2_{0\cdot}$ и $\sigma^2_{\cdot\cdot}$:

$$\begin{aligned} \mathscr{E}^{-1}(\sigma^2_{\cdot 0}) &= \frac{1}{s(r-1)}\left[S_{\cdot 0} - \frac{S_{\cdot\cdot}}{s-1}\right], \\ \mathscr{E}^{-1}(\sigma^2_{0\cdot}) &= \frac{1}{r(s-1)}\left[S_{0\cdot} - \frac{S_{\cdot\cdot}}{r-1}\right], \\ \mathscr{E}^{-1}(\sigma^2_{\cdot\cdot}) &= \frac{S_{\cdot\cdot}}{(r-1)(s-1)}. \end{aligned} \tag{10.8.4}$$

Оценкой для μ будет $\bar{x}_{\cdot\cdot}$, ее дисперсия равна, как показано в (8.6.12), $\frac{\sigma^2_{\cdot 0}}{r} + \frac{\sigma^2_{0\cdot}}{s} + \frac{\sigma^2_{\cdot\cdot}}{rs}$. Несмещенная оценка этой дисперсии получится при замене $\sigma^2_{\cdot 0}$, $\sigma^2_{0\cdot}$ и $\sigma^2_{\cdot\cdot}$ на их оценки из (10.8.4).

Обычно составляющие Модели II собирают в таблице дисперсионного анализа по образцу таблицы 10.2.

Наконец, заметим, что в случае конечных совокупностей уровней факторов R и C оценки для компонент дисперсии, т. е. в этом случае для $\sigma^2_{f\cdot 0}$, $\sigma^2_{f0\cdot}$ и $\sigma^2_{f\cdot\cdot}$, получаются заменой уравнений (10.8.3) на (8.6.22) и применением **10.7.1**. Таблица, аналогичная 10.2, может быть, конечно, построена и в этом случае.

(b) Сбалансированный неполный двухфакторный план эксперимента. Выборочный метод, входящий в описание Модели II сбалансированного неполного двухфакторного плана, дан в § 8.6 (d)

Таблица 10.2

Модель II дисперсионного анализа
Таблица полного двухфакторного плана эксперимента

Источник дисперсии	Степени свободы (с. с.)	Сумма квадратов (с. к.)	Среднее суммы квадратов (с. с. к.)	Среднее значение с. с. к.
Строки	$r-1$	$S_{\cdot 0}$	$\dfrac{S_{\cdot 0}}{r-1}$	$s\sigma_{\cdot 0}^2+\sigma_{\cdot\cdot}^2$
Столбцы	$s-1$	$S_{0\cdot}$	$\dfrac{S_{0\cdot}}{s-1}$	$r\sigma_{0\cdot}^2+\sigma_{\cdot\cdot}^2$
Остаток (ошибка)	$(r-1)(s-1)$	$S_{\cdot\cdot}$	$\dfrac{S_{\cdot\cdot}}{(r-1)(s-1)}$	$\sigma_{\cdot\cdot}^2$
Сумма	$rs-1$	S_T		

Этот план выбирает подвыборку $\{x_{\xi\eta}\}^*$ объема $n=rs'=s'r$ из реагирующих случайных величин матричной выборки $\{x_{\xi\eta}\}$, как это сделано в § 8.6 (d). Каждый элемент в $\{x_{\xi\eta}\}^*$ является линейной комбинацией (10.8.2) случайных величин. Дисперсия $x_{\xi\eta}\in\{x_{\xi\eta}\}^*$ равна σ^2 и представляется в виде суммы трех компонент (10.8.1).

Основная задача анализа Модели II сбалансированного неполного двухфакторного плана состоит в построении несмещенных оценок для μ и компонент дисперсии $\sigma_{\cdot 0}^2$, $\sigma_{0\cdot}^2$, $\sigma_{\cdot\cdot}^2$ по данным подвыборки $\{x_{\xi\eta}\}^*$.

Среднее значение $\bar{x}_{\cdot\cdot}^*$ выборки $\{x_{\xi\eta}\}^*$ является несмещенной оценкой μ с дисперсией

$$\sigma^2(\bar{x}_{\cdot\cdot}^*)=\frac{\sigma_{\cdot\cdot}^2}{n}+\frac{\sigma_{\cdot 0}^2}{r}+\frac{\sigma_{0\cdot}^2}{s}, \tag{10.8.5}$$

как было показано в (8.6.39).

Заметим, что три последних уравнения (8.6.39) относительно $\sigma_{\cdot 0}^2$, $\sigma_{0\cdot}^2$, $\sigma_{\cdot\cdot}^2$ имеют при $r'>1$, $s'>1$ единственное решение. Применяя теперь **10.1.1** (или **10.7.1**), получим следующие оценки для $\sigma_{\cdot 0}^2$, $\sigma_{0\cdot}^2$ и $\sigma_{\cdot\cdot}^2$:

$$\mathcal{E}^{-1}(\sigma_{\cdot 0}^2)=\frac{S_{\cdot 0}^*+S_E^*}{n-s}-\mathcal{E}^{-1}(\sigma_{\cdot\cdot}^2),$$

$$\mathcal{E}^{-1}(\sigma_{0\cdot}^2)=\frac{S_{0\cdot}^*+S_E^*}{n-r}-\mathcal{E}^{-1}(\sigma_{\cdot\cdot}^2), \tag{10.8.6}$$

$$\mathcal{E}^{-1}(\sigma_{\cdot\cdot}^2)=\frac{1}{K}[AS_{0\cdot}^*+BS_{\cdot 0}^*+CS_E^*],$$

где

$$A=\frac{r-r'}{n-r},\quad B=\frac{s-s'}{n-s},\quad C=A+B-1,$$

$$K=n-r-s+1.$$

Несмещенная оценка для $\sigma^2(\bar{x}^*_{..})$ получится после подстановки $\mathscr{E}^{-1}(\sigma^2_{\cdot 0})$, $\mathscr{E}^{-1}(\sigma^2_{0\cdot})$ и $\mathscr{E}^{-1}(\sigma^2_{..})$ вместо $\sigma^2_{\cdot 0}$, $\sigma^2_{0\cdot}$ и $\sigma^2_{..}$ соответственно в (10.8.5).

Предоставляем читателю в качестве упражнения построить таблицу неполного двухфакторного плана в Модели II дисперсионного анализа, аналогичную таблице 10.2.

(c) Полный трехфакторный план эксперимента. Теория матричных (третьего порядка) выборок, лежащая в основе оценивания компонент дисперсии в полном трехфакторном плане, изложена в § 8.6 (c). Мы используем обозначения этого параграфа. С полным трехфакторным планом эксперимента ассоциируется матричная выборка $\{x_{\xi\eta\zeta}\}$, образованная реагирующими случайными величинами. Мы предполагаем, что r уровней фактора R, s уровней фактора C и t уровней фактора L сами суть выборки из бесконечных совокупностей. Тогда каждый элемент $x_{\xi\eta\zeta}$ выборки $\{x_{\xi\eta\zeta}\}$ является в соответствии с (8.6.30) суммой случайных величин с нулевыми средними и ковариациями:

$$x_{\xi\eta\zeta}=\mu+\varepsilon_{\xi\cdot\cdot}+\varepsilon_{\cdot\eta\cdot}+\varepsilon_{\cdot\cdot\zeta}+\varepsilon_{\xi\eta\cdot}+\varepsilon_{\xi\cdot\zeta}+\varepsilon_{\cdot\eta\zeta}+\varepsilon_{\xi\eta\zeta}, \tag{10.8.7}$$

а дисперсия $\sigma^2(x_{\xi\eta\zeta})$ — суммой семи компонент, согласно (8.6.27) или (8.6.28):

$$\sigma^2=\sigma^2_{\cdot 00}+\sigma^2_{0\cdot 0}+\sigma^2_{00\cdot}+\sigma^2_{\cdot\cdot 0}+\sigma^2_{\cdot 0\cdot}+\sigma^2_{0\cdot\cdot}+\sigma^2_{\cdot\cdot\cdot}. \tag{10.8.8}$$

Применяя теперь **10.1.1** (или **10.7.1**), мы с помощью величин $S_{\cdot 00}$, $S_{0\cdot 0}$, $S_{00\cdot}$, $S_{\cdot\cdot 0}$, $S_{\cdot 0\cdot}$, $S_{0\cdot\cdot}$, $S_{\cdot\cdot\cdot}$, определенных в (8.6.31), построим оценки компонент дисперсии:

$$\begin{aligned}
\mathscr{E}^{-1}(\sigma^2_{\cdot\cdot\cdot})&=\frac{S_{\cdot\cdot\cdot}}{(r-1)(s-1)(t-1)},\\
\mathscr{E}^{-1}(\sigma^2_{0\cdot\cdot})&=\frac{S_{0\cdot\cdot}}{r(s-1)(t-1)}-\frac{1}{r}\mathscr{E}^{-1}(\sigma^2_{\cdot\cdot\cdot}),\\
\mathscr{E}^{-1}(\sigma^2_{\cdot 0\cdot})&=\frac{S_{\cdot 0\cdot}}{s(r-1)(t-1)}-\frac{1}{s}\mathscr{E}^{-1}(\sigma^2_{\cdot\cdot\cdot}),\\
\mathscr{E}^{-1}(\sigma^2_{\cdot\cdot 0})&=\frac{S_{\cdot\cdot 0}}{t(r-1)(s-1)}-\frac{1}{t}\mathscr{E}^{-1}(\sigma^2_{\cdot\cdot\cdot}),\\
\mathscr{E}^{-1}(\sigma^2_{00\cdot})&=\frac{S_{00\cdot}}{rs(t-1)}-\frac{1}{s}\mathscr{E}^{-1}(\sigma^2_{0\cdot\cdot})-\frac{1}{r}\mathscr{E}^{-1}(\sigma^2_{\cdot 0\cdot})-\frac{1}{rs}\mathscr{E}^{-1}(\sigma^2_{\cdot\cdot\cdot}),\\
\mathscr{E}^{-1}(\sigma^2_{0\cdot 0})&=\frac{S_{0\cdot 0}}{rt(s-1)}-\frac{1}{t}\mathscr{E}^{-1}(\sigma^2_{0\cdot\cdot})-\frac{1}{r}\mathscr{E}^{-1}(\sigma^2_{\cdot\cdot 0})-\frac{1}{rt}\mathscr{E}^{-1}(\sigma^2_{\cdot\cdot\cdot}),\\
\mathscr{E}^{-1}(\sigma^2_{\cdot 00})&=\frac{S_{\cdot 00}}{st(r-1)}-\frac{1}{t}\mathscr{E}^{-1}(\sigma^2_{\cdot 0\cdot})-\frac{1}{s}\mathscr{E}^{-1}(\sigma^2_{\cdot\cdot 0})-\frac{1}{st}\mathscr{E}^{-1}(\sigma^2_{\cdot\cdot\cdot}).
\end{aligned} \tag{10.8.9}$$

Несмещенной оценкой μ будет $\bar{x}_{\cdot\cdot\cdot}$, дисперсия этой оценки дается формулой (8.6.35). Несмещенная оценка для $\sigma^2(\bar{x}_{\cdot\cdot\cdot})$ получается заменой в (8.6.35) компонент дисперсии их оценками (10.8.9).

Рекомендуем читателю построить таблицу полного трехфакторного плана в Модели II дисперсионного анализа, аналогичную таблице 10.2.

Отметим также, что оценивание компонент дисперсии в полном трехфакторном плане в случае, когда совокупности уровней факторов R, C, L конечны, производится вполне аналогично изложенному. Подробности оставляем читателю.

(d) План латинского квадрата. Как отмечалось в §§ 8.6 (d), 10.6 (c), план латинского квадрата представляет собой специальный случай сбалансированной неполной матричной (третьего порядка) выборки, именно случай $s=t=r$ и $s'=t'=r'=1$. Определенная в § 8.6 (d) выборка $\{x_{\xi\eta\zeta}\}^*$ из реагирующих случайных величин для плана латинского квадрата такова, что каждая величина $x_{\xi\eta\zeta}\in\{x_{\xi\eta\zeta}\}^*$ имеет вид (10.8.7), а ее дисперсия представляется в виде суммы (10.8.8) семи компонент. Но если полную сумму квадратов S_T^* разложить на суммы квадратов $S^*_{\cdot 00}$, $S^*_{0\cdot 0}$, $S^*_{00\cdot}$ и S^*_E в соответствии с (8.6.40), то, как можно увидеть из (8.6.41), оценки существуют только для следующих компонент из (10.8.8):

$$\sigma_E^2,\ \sigma^2_{\cdot 00},\ \sigma^2_{0\cdot 0},\ \sigma^2_{00\cdot},\tag{10.8.10}$$

где

$$\sigma_E^2=\sigma^2_{\cdots}+\sigma^2_{\cdot\cdot 0}+\sigma^2_{\cdot 0\cdot}+\sigma^2_{0\cdots}$$

Другими словами, мы можем оценить индивидуально только компоненты строк, столбцов и слоев и сумму остальных четырех компонент. Эти оценки получаются применением **10.7.1** к (8.6.41), что дает

$$\begin{aligned}
&\mathscr{E}^{-1}(\sigma^2_{\cdot 00})=\frac{S^*_{\cdot 00}}{r-1}-\mathscr{E}^{-1}(\sigma_E^2),\\
&\mathscr{E}^{-1}(\sigma^2_{0\cdot 0})=\frac{S^*_{0\cdot 0}}{r-1}-\mathscr{E}^{-1}(\sigma_E^2),\\
&\mathscr{E}^{-1}(\sigma^2_{00\cdot})=\frac{S^*_{00\cdot}}{r-1}-\mathscr{E}^{-1}(\sigma_E^2),\\
&\mathscr{E}^{-1}(\sigma_E^2)=\frac{S^*_{\cdot 00}+S^*_{0\cdot 0}+S^*_{00\cdot}}{2r-1}-\frac{S^*_E}{(2r-1)(r-1)}.
\end{aligned}\tag{10.8.11}$$

Построение таблицы плана латинского квадрата в Модели II дисперсионного анализа предоставляется читателю в качестве упражнения.

10.9. Линейные оценки для средних значений расслоенных совокупностей

(a) Определения и обозначения. Во многих практических задачах, связанных с выбором из конечной совокупности π_N, ситуация такова, что π_N разлагается на m дизъюнктных *слоев* $\pi_{N_1}, \ldots, \pi_{N_m}$, где $N=N_1+\ldots+N_m$, среднее и дисперсия слоя π_{N_g} равны μ_g и σ_g^2, $g=1, \ldots, m$. Если среднее и дисперсия π_N равны μ и σ^2, то легко проверить, что

$$\mu=\sum_g p_g\mu_g\tag{10.9.1}$$

и

$$\sigma^2 = \sigma_W^2 + \sigma_B^2, \tag{10.9.2}$$

где

$$\sigma_W^2 = \sum_g \frac{N_g - 1}{N-1} \sigma_g^2, \qquad \sigma_B^2 = \frac{N}{N-1} \sum_g p_g (\mu_g - \mu)^2 \tag{10.9.3}$$

и

$$p_g = \frac{N_g}{N}, \qquad g = 1, \ldots, m. \tag{10.9.4}$$

Отметим, что σ^2 оказывается суммой двух компонент: *внутрислойной* компоненты σ_W^2 и *междуслойной* компоненты σ_B^2.

Если $O_{n_1}, \ldots, O_{n_m}$ — независимые выборки объемов $n_1, \ldots, n_m$ из слоев $\pi_{N_1}, \ldots, \pi_{N_m}$ соответственно, то будем обозначать через $\bar{x}_1, \ldots, \bar{x}_m$ средние, а через $s_1^2, \ldots, s_m^2$ — дисперсии выборок $O_{n_1}, \ldots, O_{n_m}$. Взятые вместе $O_{n_1}, \ldots, O_{n_m}$ образуют *расслоенную выборку* объема n из π_N.

Основная задача здесь — нахождение наилучших линейных оценок для μ, зависящих от средних значений выборок из слоев, при различной информации относительно N_g, μ_g, σ_g^2, $g = 1, \ldots, m$. На самом деле мы рассмотрим две постановки этой задачи:

(i) известны объемы слоев $N_1, \ldots, N_m$;

(ii) известны объемы слоев и дисперсии слоев $\sigma_1^2, \ldots, \sigma_m^2$.

Если совокупность π_N нерасслоенная, то мы уже видели в § 10.2, что линейной оценкой μ с минимальной дисперсией по данным простой случайной выборки O_n из π_N будет $\bar{x}$, причем дисперсия $\bar{x}$ равна $\left(\frac{1}{n} - \frac{1}{N}\right)\sigma^2$. Эта дисперсия является, так сказать, «стандартом», с которым можно сравнивать дисперсии оценок, полученных в условиях (i) или (ii).

Расслоенные выборки широко используются при выборочном обследовании в управлении, коммерческой деятельности, промышленности; имеется обширная литература, посвященная этому предмету. Здесь мы дадим только введение в математическую теорию такого рода выборок. Дополнительные сведения можно найти в книгах Кочрена (1953), Хансена, Гурвица и Мэдоу (1953), Стефана и Маккарти (1958), Сукхатмэ (1954), Иэйтса (1949).

(b) Линейная оценка для среднего значения совокупности по данным общей расслоенной выборки. В этом случае предполагается только, что объемы слоев $N_1, \ldots, N_m$ известны. Пусть $(x_{g1}, \ldots, x_{gn_g})$ — выборка O_{n_g} объема $n_g \geqslant 2$, $g = 1, \ldots, m$. Множество этих выборок из слоев называется *общей расслоенной выборкой* из π_N. Пусть L — линейная функция всех элементов выборки, т. е.

$$L = \sum_{g=1}^{m} \sum_{\xi=1}^{n_g} c_{g\xi} x_{g\xi}. \tag{10.9.5}$$

Мы хотим определить $c_{g\xi}$ так, чтобы L была несмещенной оценкой μ с минимальной дисперсией. Запишем условие несмещенности:

$$\mathscr{E}(L)=\sum_{g=1}^{m}\sum_{\xi=1}^{n_g}c_{g\xi}\mu_g=\sum_g p_g\mu_g,$$

т. е.

$$\sum_{\xi=1}^{n_g}c_{g\xi}=p_g, \qquad g=1, \ldots, m. \tag{10.9.6}$$

Поскольку выборки из различных слоев независимы, дисперсия L равна

$$\sigma^2(L)=\sum_{g=1}^{m}\left[\sigma_g^2\left(\sum_{\xi=1}^{n_g}c_{g\xi}^2\right)-\frac{\sigma_g^2}{N_g}\left(\sum_{\xi=1}^{n_g}c_{g\xi}\right)^2\right]. \tag{10.9.7}$$

Легко проверить, что значения $c_{g\xi}$, удовлетворяющие (10.9.6) и минимизирующие (10.9.7), равны

$$c_{g\xi}=\frac{p_g}{n_g}, \qquad \xi=1, \ldots, n_g;\ g=1, \ldots, m. \tag{10.9.8}$$

Подставим эти значения в (10.9.5) и обозначим полученную оценку для μ через $\mathscr{E}_s^{-1}(\mu)$. Тогда

$$\mathscr{E}_s^{-1}(\mu)=p_1\bar{x}_1+\ldots+p_m\bar{x}_m. \tag{10.9.9}$$

Дисперсия $\mathscr{E}_s^{-1}(\mu)$ равна

$$\sigma^2(\mathscr{E}_s^{-1}(\mu))=\sum_g p_g^2\left(\frac{1}{n_g}-\frac{1}{N_g}\right)\sigma_g^2, \tag{10.9.10}$$

откуда непосредственно получаем, что

$$\sum_g p_g^2\left(\frac{1}{n_g}-\frac{1}{N_g}\right)s_g^2 \tag{10.9.11}$$

является несмещенной оценкой для $\sigma^2(\mathscr{E}_s^{-1}(\mu))$.

Итак,

10.9.1. *Пусть π_N — конечная совокупность с дизъюнктными слоями $\pi_{N_1}, \ldots, \pi_{N_m}$ известных объемов; среднее и дисперсия слоя π_{N_g} равны μ_g, σ_g^2, $g=1, \ldots, m$. Если $O_{n_1}, \ldots, O_{n_m}$ — общая расслоенная выборка из π_N и $\bar{x}_g$ — среднее и s_g^2 — дисперсия выборки O_{n_g}, $g=1, \ldots, m$, то оценка $\mathscr{E}_s^{-1}(\mu)$, определенная в* (10.9.9), *— линейная несмещенная оценка среднего μ совокупности π_N с минимальной дисперсией.*

Дисперсия $\mathscr{E}_s^{-1}(\mu)$ указана в (10.9.10), *а несмещенная оценка этой дисперсии — в* (10.9.11).

Отметим, что в предыдущих рассуждениях объемы $n_1 \geqslant 2, \ldots$ $\ldots, n_m \geqslant 2$ выборок из слоев задавались произвольно.

В специальном случае, когда объемы $n_1, \ldots, n_m$ выборок из слоев пропорциональны объемам слоев, т. е. $n_g = p_g \cdot n$, $g = 1, \ldots, m$, имеем *пропорциональную расслоенную выборку* объема n из π_N (на практике, конечно, $p_g \cdot n$ округляется до ближайшего целого). При таком выборе n_g, $g = 1, \ldots, m$, будем обозначать $\mathcal{E}_s^{-1}(\mu)$ через $\mathcal{E}_{sp}^{-1}(\mu)$. Дисперсия $\mathcal{E}_{sp}^{-1}(\mu)$ равна

$$\sigma^2(\mathcal{E}_{sp}^{-1}(\mu)) = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_g p_g \sigma_g^2, \tag{10.9.10a}$$

а

$$\left(\frac{1}{n} - \frac{1}{N}\right) \sum_g p_g s_g^2 \tag{10.9.11a}$$

служит несмещенной оценкой для $\sigma^2(\mathcal{E}_{sp}^{-1}(\mu))$.

Эффективность $\mathcal{E}_{sp}^{-1}(\mu)$ *относительно* $\bar{x}$ — среднего значения простой случайной выборки из π_N — как оценки μ определяется отношением

$$\operatorname{eff}(\mathcal{E}_{sp}^{-1}(\mu)) = \frac{\sigma^2(\bar{x})}{\sigma^2(\mathcal{E}_{sp}^{-1}(\mu))} = 1 + K, \tag{10.9.12}$$

при этом

$$K = \frac{\sigma_B^2}{\sigma_W^2} + O\left(\frac{1}{N}\right). \tag{10.9.13}$$

Из (10.9.12) видно, что преимущество пропорциональной расслоенной выборки над простой выборкой при линейном оценивании среднего значения совокупности π_N зависит от отношения $\frac{\sigma_B^2}{\sigma_W^2}$. Если $\sigma_B^2 = 0$, т. е. средние значения всех слоев одинаковы, то пропорциональная расслоенная выборка не дает выигрыша по сравнению с простой случайной выборкой. Подытожим все сказанное.

10.9.2. *Линейная оценка* $\mathcal{E}_{sp}^{-1}(\mu)$, *построенная на основе пропорциональной расслоенной выборки с* $n_g = p_g \cdot n$, $g = 1, \ldots, m$, *имеет дисперсию* (10.9.10a). *Несмещенной оценкой этой дисперсии служит* (10.9.11a). *Эффективность* $\mathcal{E}_{sp}^{-1}(\mu)$ *относительно* $\bar{x}$ — *среднего значения случайной выборки объема* n *из совокупности* π_N — *как оценки среднего совокупности* π_N *дается формулой* (10.9.12).

(c) Линейная оценка с минимальной дисперсией среднего значения совокупности, когда объемы и дисперсии слоев известны. Если известны не только $N_1, \ldots, N_m$, но и $\sigma_1, \ldots, \sigma_m$, то можно подобрать объемы выборок из слоев «лучшие», чем при пропорциональном выборе из слоев. Возникающая здесь задача равносильна минимизации выражения (10.9.10) по $n_1, \ldots, n_m$ при условии $n_1 + \ldots + n_m = n$. Предположим на время, что $n_1, \ldots, n_m$ изменяются непрерывно. Тогда легко проверить, что требуемые значения $n_1, \ldots, n_m$ суть

$$n_g = b_g \cdot n, \qquad g = 1, \ldots, m, \tag{10.9.14}$$

где

$$b_g = \frac{p_g \sigma_g}{\sum_g p_g \sigma_g}. \tag{10.9.15}$$

Для этого оптимального выбора $n_1, \ldots, n_m$ будем обозначать $\mathscr{E}_s^{-1}(\mu)$ через $\mathscr{E}_{s0}^{-1}(\mu)$. Тогда после упрощений имеем

$$\sigma^2(\mathscr{E}_{s0}^{-1}(\mu)) = \frac{1}{n}\left(\sum_g p_g \sigma_g\right)^2 - \frac{1}{N}\sum_g p_g \sigma_g^2. \tag{10.9.10b}$$

(На практике, конечно, в качестве n_g выбираются ближайшие к nb_g целые числа.)

Можно показать, что

$$\sigma^2(\mathscr{E}_{s0}^{-1}(\mu)) \leqslant \sigma^2(\mathscr{E}_{sp}^{-1}(\mu)), \tag{10.9.16}$$

причем равенство имеет место тогда и только тогда, когда $\sigma_1 = \ldots = \sigma_m$. Действительно, сравнивая (10.9.10) и (10.9.10a), видим, что (10.9.16) эквивалентно неравенству

$$\left(\sum_g p_g \sigma_g\right)^2 \leqslant \sum_g p_g \sigma_g^2,$$

т. е. неравенству

$$\left(\sum_g p_g \sigma_g\right)^2 \leqslant \left(\sum_g p_g\right) \cdot \left(\sum_g p_g \sigma_g^2\right), \tag{10.9.17}$$

так как $\sum_g p_g = 1$ и все p_g положительны.

Но (10.9.17) — частный случай хорошо известного неравенства Шварца; в нем знак равенства достигается только при $\sigma_1 = \ldots = \sigma_m$.

Расслоенная выборка объема n из совокупности π_N, для которой n_g определяются согласно (10.9.14), называется иногда *оптимальной расслоенной выборкой* объема n из π_N.

Итак, получен следующий результат Неймана (1934):

10.9.3. *Пусть π_N — конечная совокупность с дизъюнктными слоями $\pi_{N_1}, \ldots, \pi_{N_m}$, объемы которых $N_1, \ldots, N_m$ и дисперсии $\sigma_1^2, \ldots, \sigma_m^2$ известны. Тогда расслоенная выборка объема n из π_N, приводящая к линейной оценке $\mathscr{E}_{s0}^{-1}(\mu)$ с минимальной дисперсией среднего значения совокупности π_N, определяется объемами n_g выборок из слоев, задаваемыми формулой* (10.9.14). *Дисперсия оценки $\mathscr{E}_{s0}^{-1}(\mu)$ указана в* (10.9.10b). *$\mathscr{E}_{s0}^{-1}(\mu)$ и $\mathscr{E}_{sp}^{-1}(\mu)$ одинаково эффективны как оценки среднего совокупности π_N тогда и только тогда, когда дисперсии слоев равны.*

(d) Линейная оценка с минимальной дисперсией среднего значения совокупности при фиксированной суммарной стоимости. На практике часто встречаются ситуации, когда стоимости выбора из различных слоев значительно разнятся между собой. В этих условиях желательно подобрать объемы выборок из различных слоев так, чтобы минимизировать выражение (10.9.10) для $\sigma^2(\mathcal{E}_s^{-1}(\mu))$ при фиксированном значении суммарной стоимости получения выборки. Более подробно, пусть c_g — стоимость одного наблюдения в выборке объема n_g из слоя π_{N_g}. Стоимость C получения всей выборки равна

$$C = c_1 n_1 + \ldots + c_m n_m. \qquad (10.9.18)$$

Легко находим, что минимум $\sigma^2(\mathcal{E}_s^{-1}(\mu))$ при условии (10.9.18) достигается при

$$n_g = CB_g, \qquad g = 1, \ldots, m, \qquad (10.9.19)$$

где

$$B_g = \frac{\sqrt{c_g}\cdot p_g \sigma_g}{c_g \sum_g (\sqrt{c_g} p_g \sigma_g)}. \qquad (10.9.20)$$

Отметим, что описанный подход предполагает известными объемы $N_1, \ldots, N_m$ и дисперсии $\sigma_1^2, \ldots, \sigma_m^2$ слоев и стоимости $c_1, \ldots, c_m$.

При выборе значений n_g согласно (10.9.19) будем обозначать $\mathcal{E}_s^{-1}(\mu)$ через $\mathcal{E}_{sc}^{-1}(\mu)$. Тогда

$$\sigma^2(\mathcal{E}_{sc}^{-1}(\mu)) = \frac{1}{C}\left(\sum_g \sqrt{c_g} p_g \sigma_g\right)^2 - \frac{1}{N}\sum_g p_g \sigma_g^2. \qquad (10.9.21)$$

(е) Обобщение результатов на расслоенные выборки из бесконечных совокупностей. Теоремы **10.9.1**, **10.9.2**, **10.9.3** могут быть очевидным образом перенесены на расслоенные выборки из бесконечной совокупности при условии, что p_g и σ_g^2 сходятся к положительным значениям, а μ_g — к конечному пределу, когда $N_1 \to \infty, \ldots, N_m \to \infty$, $g = 1, \ldots, m$. Эти обобщения предоставляются читателю в качестве упражнения.

10.10. Линейная оценка среднего значения расслоенной совокупности при двухступенчатом выборе

(а) Двухступенчатый выбор. Если совокупность π_N состоит из большого числа слоев, то выбор из каждого слоя может оказаться неосуществимым по экономическим соображениям. В этом случае можно воспользоваться *двухступенчатой* выборочной процедурой, при которой сначала с помощью случайного механизма определяются слои, из которых будет производиться выбор, а затем из каждого из этих слоев извлекается случайная выборка. Данные выборки используются для получения линейной оценки для μ, дисперсии линейной оценки и оценки для этой дисперсии.

Для простоты рассмотрим случай, когда объем каждого слоя кратен некоторому целому числу v, т. е.

$$N_g = M_g \cdot v, \qquad g = 1, \ldots, m. \qquad (10.10.1)$$

Будем называть *v выборочной единицей*. Тогда g-й слой можно рассматривать состоящим из M_g выборочных единиц. Общее число выборочных единиц в совокупности π_N равно, следовательно, $M_1+\ldots+M_m=M$ и

$$N=Mv. \tag{10.10.2}$$

Рассмотрим теперь случайный процесс, предложенный Уилксом (1960b) для определения числа выборочных единиц, которое следует извлечь без возвращения из каждого слоя совокупности, при условии, что суммарное число извлеченных выборочных единиц должно равняться u. Предположим этот процесс таким, что каждая из M выборочных единиц имеет одну и ту же вероятность оказаться извлеченной. (На практике это осуществляется с помощью случайных чисел.) Пусть δ_g — случайная величина, равная числу выборочных единиц, подлежащих извлечению из слоя π_{N_g}. Тогда $(\delta_1, \ldots, \delta_m)$ будет $(m-1)$-мерной случайной величиной, имеющей $(m-1)$-мерное гипергеометрическое распределение с функцией вероятности

$$p(\delta_1, \ldots, \delta_m)=\frac{\binom{M_1}{\delta_1}\cdots\binom{M_m}{\delta_m}}{\binom{M}{u}}, \tag{10.10.3}$$

где $\delta_1+\ldots+\delta_m=u$.

Таким образом, в результате первого шага мы определяем, что выборка объема $\delta_g v$ должна быть извлечена из слоя π_{N_g}, $g=1, \ldots, m$.

Второй шаг состоит в извлечении выборок объемов $\delta_1 v, \ldots, \delta_m v$ из $\pi_{N_1}, \ldots, \pi_{N_m}$ соответственно. Эта процедура дает *двухступенчатую выборку объема* uv из совокупности π_N, которую можно рассматривать как расслоенную выборку, когда слои, из которых осуществляется выбор, определяются с помощью случайного механизма. На практике, когда число слоев велико, многие среди $\delta_1, \ldots, \delta_m$ равны нулю, так что многие слои не участвуют в выборке. Общее число слоев, из которых извлекаются выборки, равно числу не равных нулю среди $\delta_1, \ldots, \delta_m$.

(b) Линейная оценка среднего совокупности. Пусть $\mathscr{E}_2^{-1}(\mu)$ — следующая линейная оценка для μ:

$$\mathscr{E}_2^{-1}(\mu)=\frac{(\delta_1 v)\,\bar{x}_1+\ldots+(\delta_m v)\,\bar{x}_m}{uv},$$

т. е.

$$\mathscr{E}_2^{-1}(\mu)=\frac{\delta_1\bar{x}_1+\ldots+\delta_m\bar{x}_m}{u}. \tag{10.10.4}$$

Имеем

$$\mathscr{E}(\mathscr{E}_2^{-1}(\mu))=\sum_g \mathscr{E}\left(\frac{\delta_g\bar{x}_g}{u}\right). \tag{10.10.5}$$

В соответствии с **3.7.2** для $g=1, \ldots, m$

$$\mathscr{E}\left(\frac{\delta_g\bar{x}_g}{u}\right)=\frac{1}{u}\,\mathscr{E}_{(\delta_g)}\left[\delta_g\mathscr{E}_{(\bar{x}_g)}(\bar{x}_g\mid\delta_g)\right], \tag{10.10.6}$$

где $\mathcal{E}_{(\bar{x}_g)}(\bar{x}_g \mid \delta_g)$ — условное математическое ожидание величины $\bar{x}_g$ относительно δ_g, а $\mathcal{E}_{(\delta_g)}[\ ]$ — (безусловное) математическое ожидание по δ_g. Но

$$\mathcal{E}_{(\bar{x}_g)}(\bar{x}_g \mid \delta_g) = \mu_g,$$

и из (6.1.8) следует, что

$$\mathcal{E}_{(\delta_g)}(\delta_g) = p_g \cdot u.$$

Поэтому

$$\mathcal{E}\,(\mathcal{E}_2^{-1}(\mu)) = \sum_g p_g \mu_g = \mu, \qquad (10.10.7)$$

так что $\mathcal{E}_2^{-1}(\mu)$ оказывается несмещенной оценкой μ.

Дисперсия $\mathcal{E}_2^{-1}(\mu)$ равна

$$\sigma^2(\mathcal{E}_2^{-1}(\mu)) = \mathcal{E}\left[\sum_g \left(\frac{\delta_g \bar{x}_g}{u} - p_g \mu_g\right)\right]^2 =$$

$$= \mathcal{E}\left[\sum_g \frac{\delta_g}{u}(\bar{x}_g - \mu_g) + \sum_g \left(\frac{\delta_g}{u} - p_g\right)\mu_g\right]. \qquad (10.10.8)$$

После некоторых преобразований получим

$$\sigma^2(\mathcal{E}_2^{-1}(\mu)) = \frac{N - uv}{u\,(N - v)}\left[\frac{N-1}{N}\left(\frac{\sigma_W^2}{v} + \sigma_B^2\right) - \frac{v-1}{Nv}\sum_g \sigma_g^2\right]. \qquad (10.10.9)$$

Обращаясь к (10.9.3), мы видим, что $\sigma^2(\mathcal{E}_2^{-1}(\mu))$ является линейной функцией величин

$$\sum_g \sigma_g^2, \quad \sum_g p_g \sigma_g^2, \quad \sum_g p_g(\mu_g - \mu)^2. \qquad (10.10.10)$$

Образуем функции G_1, G_2, G_3 от элементов двухступенчатой выборки, $v \geqslant 2$:

$$G_1 = \sum_g \delta_g s_g^2, \quad G_2 = \sum_g \frac{\delta_g}{N_g} s_g^2, \quad G_3 = \sum_g \delta_g (\bar{x}_g - \mathcal{E}_2^{-1}(\mu))^2. \qquad (10.10.11)$$

Можно проверить, что $\mathcal{E}(G_1)$, $\mathcal{E}(G_2)$ и $\mathcal{E}(G_3)$ являются линейными функциями величин (10.10.10), так что существует линейная функция от G_1, G_2, G_3, несмещенно оценивающая $\sigma^2(\mathcal{E}_2^{-1}(\mu))$.

Фактически же несмещенная оценка $\sigma^2(\mathcal{E}_2^{-1}(\mu))$ получается при замене σ_W^2, σ_B^2 и $\sum_g \sigma_g^2$ в (10.10.9) их несмещенными оценками:

$$\mathcal{E}^{-1}(\sigma_W^2) = \frac{N}{u\,(N-1)}(G_1 - G_2),$$

$$\mathcal{E}^{-1}(\sigma_B^2) = \frac{Nv}{uv\,(N-1) + N} \cdot \left[\frac{N-2}{uv\,(N-1)}(G_1 - G_2) - \right.$$
$$\left. - \frac{Nu + 2v}{u^2 v} G_2 + G_3\right], \qquad (10.10.12)$$

$$\mathcal{E}^{-1}\left(\sum_g \sigma_g^2\right) = \frac{N}{u} G_2.$$

Суммируем полученные результаты в виде теоремы

10.10.1. *Пусть π_N — конечная совокупность с дизъюнктными слоями $\pi_{N_1}, \ldots, \pi_{N_m}$, где $N_g = M_g v$, $g = 1, \ldots, m$ и v — заданное целое (выборочная единица). Если из π_N извлечена двухступенчатая выборка из u выборочных единиц, то $\mathfrak{G}_2^{-1}(\mu)$ будет несмещенной линейной оценкой для среднего значения совокупности π_N. Дисперсия $\mathfrak{G}_2^{-1}(\mu)$ указана в* (10.10.9), *причем правая часть* (10.10.9), *с заменой σ_W^2, σ_B^2, $\sum\limits_g \sigma_g^2$ их несмещенными оценками* (10.10.12), *оказывается несмещенной оценкой для $\sigma^2(\mathfrak{G}_2^{-1}(\mu))$, когда $v \geqslant 2$.*

(c) Случай бесконечной совокупности. Если слои представляют собой бесконечные совокупности такие, что p_g и σ_g^2 сходятся к положительным значениям, а μ_g — к конечному пределу, когда $N_g \to \infty$, $g = 1, \ldots, m$, то $(m-1)$-мерная случайная величина $(\delta_1, \ldots, \delta_m)$, $\delta_1 + \ldots + \delta_m = u$, имеет мультиномиальное распределение с функцией вероятности

$$p(\delta_1, \ldots, \delta_m) = \frac{u!}{\delta_1! \ldots \delta_m!} p_1^{\delta_1} \ldots p_m^{\delta_m}. \tag{10.10.13}$$

Дисперсия $\mathfrak{G}_2^{-1}(\mu)$ в этом случае равна

$$\sigma^2(\mathfrak{G}_2^{-1}(\mu)) = \frac{\sigma_{W'}^2}{uv} + \frac{\sigma_{B'}^2}{u}, \tag{10.10.14}$$

где

$$\sigma_{W'}^2 = \sum_g p_g \sigma_g^2, \qquad \sigma_{B'}^2 = \sum_g p_g (\mu_g - \mu)^2.$$

При этом

$$\frac{n+v-1}{un(n+1)} G_1 + \frac{v^2}{n(n+1)} G_3 - \frac{v}{un(n+1)} \sum_g \frac{\delta_g}{p_g} s_g^2 \tag{10.10.15}$$

оказывается несмещенной оценкой для $\sigma^2(\mathfrak{G}_2^{-1}(\mu))$, $n = uv$. Заметим, что $\sigma_{B'}^2 = \sigma_B^2 + O\left(\frac{1}{N}\right)$ и $\sigma_{W'}^2 = \sigma_W^2 + O\left(\frac{1}{N}\right)$.

(d) Линейная оценка с минимальной дисперсией среднего значения совокупности, когда суммарная стоимость фиксирована. Предположим, что объем выборки uv фиксирован, $uv = n$, и запишем $\sigma^2(\mathfrak{G}_2^{-1}(\mu))$ в виде

$$\sigma^2(\mathfrak{G}_2^{-1}(\mu)) = \frac{N-n}{n}\left(\frac{A+NB}{N-v} - B\right), \tag{10.10.16}$$

где

$$A = \sigma_{W'}^2, \qquad B = \sigma_{B'}^2 - \frac{1}{N}\sum_g \sigma_g^2. \tag{10.10.17}$$

Из (10.10.16) видно, что, поскольку v должно быть положительным целым числом, значение v, минимизирующее $\sigma^2(\mathfrak{G}_2^{-1}(\mu))$, есть $v = 1$, конечно, при условии $A + NB \geqslant 0$ и n слоях. Но тогда двухступенчатая выборка из π_N сводится к простой случайной выборке объема n из π_N. В частности, заметим, что в этом случае дисперсия

$\mathscr{E}_2^{-1}(\mu)$, задаваемая формулой (10.10.9), приводится к $\left(\frac{1}{n}-\frac{1}{N}\right)\sigma^2$, где σ^2 определена в (10.9.2).

Эти рассуждения показывают, что в рассматриваемой ситуации выбор значений u и v не может основываться на критерии минимума $\sigma^2(\mathscr{E}_2^{-1}(\mu))$ при фиксированном значении uv, так как этот критерий приводит к $v=1$. В реальной обстановке критерием для выбора u и v может служить такой: минимум $\sigma^2(\mathscr{E}_2^{-1}(\mu))$ при фиксированной суммарной стоимости. Простейшее предположение здесь состоит в том, что стоимость одной выборочной единицы равна C_1, а стоимость одного элемента выборки — C_2. Тогда суммарная стоимость C будет

$$C=C_1u+C_2uv. \tag{10.10.18}$$

Если $\sigma^2(\mathscr{E}_2^{-1}(\mu))$ из (10.10.9) минимизируется при условии (10.10.18), то, допуская непрерывное изменение u и v, найдем, что минимизирующее значение v — обозначим его $\tilde{v}$ — равно

$$\tilde{v}=\frac{N}{1+\sqrt{\left(1+\frac{H}{G}N\right)\left(1+\frac{B}{A}N\right)}}, \tag{10.10.19}$$

где

$$G=N\frac{C_1}{C},\quad H=N\frac{C_2}{C}-1, \tag{10.10.20}$$

а A и B определены в (10.10.17). Соответствующее значение u, которое обозначим $\tilde{u}$, получится подстановкой $\tilde{v}$ вместо v в (10.10.18) и последующим решением уравнения относительно u. Если в рассматриваемой ситуации ввести обозначение $\mathscr{E}_{2c}^{-1}(\mu)$ для оценки $\mathscr{E}_2^{-1}(\mu)$, то $\sigma^2(\mathscr{E}_{2c}^{-1}(\mu))$ определится, конечно, формулой (10.10.9) с заменой u и v на $\tilde{u}$ и $\tilde{v}$. Аналогично, если значения $\tilde{u}$ и $\tilde{v}$ подставить вместо u и v в (10.10.12), получим несмещенную оценку для $\sigma^2(\mathscr{E}_{2c}^{-1}(\mu))$. При больших N

$$\tilde{v}=\frac{\sigma_{W'}}{\sigma_{B'}}\sqrt{\frac{C_1}{C_2}}+O\left(\frac{1}{N}\right), \tag{10.10.21}$$

$$\tilde{u}=\frac{C}{C_1+\frac{\sigma_{W'}}{\sigma_{B'}}\sqrt{C_1C_2}}+O\left(\frac{1}{N}\right).$$

ЗАДАЧИ

10.1. Доказать **10.2.1** в случае, когда $\bar{x}$ — среднее значение выборки объема n из конечной совокупности объема N.

10.2. Пусть $(x_1, \dots, x_n)$ — n-мерная случайная величина, среднее значение x_i равно μ, дисперсия x_i — σ^2, ковариация каждой пары x_i и x_j равна $\rho\sigma^2$. Доказать, что линейной оценкой μ с минимальной дисперсией является $\bar{x}=\frac{x_1+\dots+x_n}{n}$.

10.3. Показать, что если $\bar{x}$ и $\bar{x}'$ — средние значения независимых выборок объемов n и n' из совокупностей со средними μ и μ' и дисперсиями σ^2

и σ'^2 соответственно, то $\bar{x}-\bar{x}'$ будет линейной оценкой $\mu-\mu'$ с минимальной дисперсией.

10.4. Пусть $T_1, \ldots, T_k$ — несмещенные оценки параметра θ с (неособой) матрицей ковариаций $\|\sigma_{ij}\|$, для которой $\|\sigma^{ij}\|$ — обратная; $l_1, \ldots, l_k$ — постоянные, сумма которых равна 1. Доказать, что в классе оценок параметра θ вида $\sum\limits_{i=1}^{k} l_i T_i$ минимум дисперсии достигается при $l_i = \sum\limits_j \sigma^{ij} \Big/ \sum\limits_{i,j} \sigma^{ij}$, $i=1, \ldots, k$, и этот минимум равен $1 \Big/ \sum\limits_{i,j} \sigma^{ij}$.

10.5. Доказать, что утверждения **10.2.2** остаются справедливыми в случае, когда $(x_1, \ldots, x_n)$ — n-мерная случайная величина, к. ф. р. которой $F(x_1, \ldots, x_n)$ симметрична относительно $x_1, \ldots, x_n$.

10.6. Пусть n_1, $\bar{x}_1$, s_1^2 — объем, среднее значение и дисперсия выборки из совокупности $N(\mu_1, \sigma^2)$, а n_2, $\bar{x}_2$, s_2^2 — те же характеристики для независимой выборки из $N(\mu_2, \sigma^2)$. Показать, что

$$(\bar{x}_1-\bar{x}_2) \pm t_{n_1+n_2-2,\,\gamma} \cdot s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

являются 100γ-процентными доверительными пределами для $\mu_1-\mu_2$, где

$$s^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2},$$

а $t_{n_1+n_2-2}$ и γ определены в (10.2.10).

10.7. Доказать **10.3.2.**

10.8. Показать, что если в 10.3.1 единичные векторы $(x_{11}, \ldots, x_{1n}), \ldots \ldots, (x_{k1}, \ldots, x_{kn})$ выбраны попарно ортогональными, то линейные оценки $\sum\limits_{\xi} x_{i\xi} y_{\xi}$, $i=1, \ldots, k$, с минимальной дисперсией для $\beta_1, \ldots, \beta_k$ соответственно, имеют одну и ту же дисперсию σ^2, а все ковариации равны нулю.

10.9. Обращаясь к **10.3.4**, показать, что

$$\sum_{i,j=1}^{k_1} a_{ij}^{*}(\beta_i-b_i)(\beta_j-b_j) = \frac{k_1 S_1 F_{k_1,\,n-k,\,\gamma}}{n-k}$$

служит 100γ-процентной доверительной областью для $(\beta_1, \ldots, \beta_{k_1})$, $k_1 \leqslant k$, где

$$\|a_{ij}^{*}\| = \left\| \begin{matrix} a^{11} & \ldots & a^{1k_1} \\ \ldots & \ldots & \ldots \\ a^{k_1 1} & \ldots & a^{k_1 k_1} \end{matrix} \right\|^{-1},$$

значение $F_{k_1,\,n-k,\,\gamma}$ выбрано так, чтобы

$$\int_0^{F_{k_1,\,n-k,\,\gamma}} dF_{k_1,\,n-k}(F) = \gamma.$$

Здесь $dF_{k_1,\,n-k}(F)$ — элемент вероятности распределения Снедекора $S(k_1,\,n-k)$.

10.10. Доказать, что две гиперплоскости в R_n, параллельные гиперплоскости

$$\sum_{\xi=1}^{n} b_\xi x_\xi = k$$

и касательные к сфере радиуса r с центром в точке $(a_1, \ldots, a_n)$, задаются уравнениями

$$\sum_{\xi=1}^{n} b_\xi (x_\xi - a_\xi) = \pm r \sqrt{\sum_{\xi=1}^{n} b_\xi^2}.$$

10.11. Показать, что с вероятностью γ неравенства

$$\sum_i c_i b_i - \delta \sqrt{\sum_{i,j} a_{ij} c_i c_j} < \sum_i c_i \beta_i < \sum_i c_i b_i + \delta \sqrt{\sum_{i,j} a_{ij} c_i c_j}$$

для всех вещественных векторов $(c_1, \ldots, c_k)$, отличных от нулевого, выполняются одновременно, где

$$\delta^2 = \frac{kS_1}{(n-k)} F_{k,\, n-k,\, \gamma},$$

b_i, β_i, S_1 и $F_{k,\, n-k,\, \gamma}$ определены в теореме **10.3.4** и (10.4.4).

10.12. Пусть $(u_1, \ldots, u_k)$ имеет распределение $N(\{\mu_i\}; \|\sigma_{ij}\|)$, вектор $(\mu_1, \ldots, \mu_k)$ неизвестен, а матрица $\|\sigma_{ij}\|$ известна, $\chi^2_{k,\gamma}$ — 100γ-процентная точка χ^2-распределения с k степенями свободы. Доказать, что с вероятностью γ неравенства

$$\sum_i c_i u_i - \chi_{k,\gamma} \sqrt{\sum_{i,j} \sigma_{ij} c_i c_j} < \sum_i c_i u_i < \sum_i c_i u_i + \chi_{k,\gamma} \sqrt{\sum_{i,j} \sigma_{ij} c_i c_j}$$

выполняются одновременно для всех вещественных векторов $(c_1, \ldots, c_k)$, отличных от нулевого.

10.13. Вывести формулы (10.6.5) и (10.6.6).

10.14. В задаче **8.20** показать, как получить 100γ-процентный доверительный интервал для μ.

10.15. Вывести (10.6.15) и (10.6.16).

10.16. Сформулировать аналоги **10.6.1**, **10.6.2** и построить таблицу дисперсионного анализа в Модели I для полного трехфакторного плана эксперимента.

10.17. Сформулировать аналоги теорем **10.6.1** и **10.6.2** и построить таблицу дисперсионного анализа в Модели I для плана латинского квадрата.

10.18. Построить таблицу дисперсионного анализа в Модели II для неполного сбалансированного двухфакторного плана эксперимента, описанного в § 10.8 (b).

10.19. Проверить, что оценки (10.8.6) компонент дисперсии для неполного сбалансированного двухфакторного плана эксперимента переходят в оценки (10.8.4) для полного двухфакторного плана, если $s' = s$ и $r' = r$.

10.20. Написать оценки, соответствующие (10.8.4), и аналог таблицы 10.2 для случая конечных совокупностей из N_1 уровней фактора R и N_2 уровней фактора C.

10.21. Построить таблицу дисперсионного анализа в Модели II для полного трехфакторного плана эксперимента.

10.22. Построить таблицу дисперсионного анализа в Модели II для плана латинского квадрата.

10.23. Пусть в полном двухфакторном плане эксперимента в Модели I $\mu_{.\eta} = 0$, $\eta = 1, \ldots, s$, тогда план можно описать как s повторений полного однофакторного плана. Показать, что в этом случае $\dfrac{S^*_{..}}{r(s-1)}$ является несмещенной оценкой σ^2, где $S^*_{..} = S_{..} + S_0$. (0). Предполагая нормальность распределения, как в **10.6.2**, показать, что 100γ-процентными доверительными интервалами для μ и $\mu_{\xi\cdot}$ в этом случае будут

$$m \pm t_{r(s-1),\gamma} \sqrt{\frac{S^*_{..}}{r^2 s(s-1)}},$$

$$m \pm t_{r(s-1),\gamma} \sqrt{\frac{S^*_{..}(r-1)}{r^2 s(s-1)}}.$$

Доказать далее, что $(r-1)$-мерная сфера

$$\sum_{\xi=1}^{r} (\mu_{\xi\cdot} - m_{\xi\cdot})^2 = \frac{s(r-1)}{r(s-1)} S^{**}_{..} \cdot F_{(r-1),\, r(s-1),\, \gamma}$$

при условии, что $\sum_{\xi=1}^{r} (\mu_{\xi\cdot} - m_{\xi\cdot}) = 0$ накрывает $(\mu_{1\cdot}, \ldots, \mu_{r\cdot})$ с вероятностью γ.

10.24. Пусть в полном трехфакторном плане эксперимента в Модели I все $\mu_{..\zeta}$, $\mu_{\cdot\eta\zeta}$, $\mu_{\xi\cdot\zeta}$ равны нулю, тогда имеем t повторений полного двухфакторного плана из r строк и s столбцов. Показать, что в этом случае $\frac{S^*_{...}}{[rs(t-1)]}$ будет несмещенной оценкой σ^2, где $S^*_{...} = S_{...} + S_{0..}(0) + S_{.0.}(0) + S_{00.}(0)$. Показать также, что величины

$$\frac{1}{\sigma^2} S^*_{...}, \quad \frac{1}{\sigma^2} S_{..0}(\mu_{\xi\eta\cdot}), \quad \frac{1}{\sigma^2} S_{.00}(\mu_{\xi..}), \quad \frac{1}{\sigma^2} S_{0.0}(\mu_{\cdot\eta\cdot})$$

независимы и имеют χ^2-распределения с $rs(t-1)$, $(r-1)(s-1)$, $(r-1)$, $(s-1)$ степенями свободы соответственно. Построить таблицу дисперсионного анализа в Модели I для проверки гипотез: (i) что все $\mu_{\xi\eta\cdot}$ равны нулю; (ii) что все $\mu_{\xi..}$ равны нулю, (iii) что все $\mu_{\cdot\eta\cdot}$ равны нулю.

10.25. Обращаясь к § 10.6 (b), доказать, что с вероятностью, не меньшей γ, доверительные интервалы

$$(m_{\xi\cdot} - m_{\xi'\cdot}) \pm \delta \sqrt{\frac{2(r-2)}{rs}}$$

одновременно накрывают $(\mu_{\xi\cdot} - \mu_{\xi'\cdot})$, $\xi \neq \xi' = 1, \ldots, r$ соответственно, где

$$\delta^2 = \frac{S_{..}}{s-1} F_{r-1,\,(r-1)(s-1),\,\gamma},$$

а $F_{r-1,\,(r-1)(s-1),\,\gamma}$ — 100γ-процентная точка распределения Снедекора $S(r-1, (r-1)(s-1))$.

10.26. Пусть $\bar{x}_1, \ldots, \bar{x}_k$ и $s_1^2, \ldots, s_k^2$ — средние значения и дисперсии независимых выборок объемов $n_1, \ldots, n_k$ из совокупностей $N(\mu_1, \sigma^2), \ldots, N(\mu_k, \sigma^2)$ соответственно. Доказать, что с вероятностью, не меньшей γ, доверительные интервалы $(\bar{x}_i - \bar{x}_j) \pm \delta \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$ одновременно накрывают соответствующие значения $(\mu_i - \mu_j)$, $i > j = 1, \ldots, k$, где

$$\delta^2 = \frac{k}{n-k} [(n_1 - 1) s_1^2 + \ldots + (n_k - 1) s_k^2] F_{k,\, n-k,\, \gamma},$$

а $F_{k,\, n-k,\, \gamma}$ — 100 γ-процентная точка распределения Снедекора $S(k, n-k)$.

10.27. В обозначениях §§ 10.5 (c), 10.6 (a) показать, что с вероятностью, не меньшей γ, неравенства

$$(\bar{x}_{\xi\cdot}-\bar{x}_{\xi'\cdot})-D<(\mu_{\xi\cdot}-\mu_{\xi'\cdot})<(\bar{x}_{\xi\cdot}-\bar{x}_{\xi'\cdot})+D$$

выполняются одновременно для всех $\xi\neq\xi'=1, \ldots, r-1$, где

$$D=R_{r-2,\,(r-1)(s-1),\,\gamma}\cdot\sqrt{\frac{(r-2)\,S_{..}}{rs\,(r-1)\,(s-1)}}.$$

10.28. В обозначениях §§ 10.5 (c), 10.6 (b) показать, что при фиксированном η с вероятностью, не меньшей γ, выполняются одновременно неравенства

$$(m_{\xi\eta\cdot}-m_{\xi'\eta\cdot})-D<(\mu_{\xi\eta\cdot}-\mu_{\xi'\eta\cdot})<(m_{\xi\eta\cdot}-m_{\xi'\eta\cdot})+D$$

для всех $\xi\neq\xi'=1, \ldots, r$, где

$$D=R_{r-2,\,(r-1)(s-1)(t-1),\,\gamma}\cdot\sqrt{\frac{(r-2)\,S_{..}}{rst\,(r-1)\,(t-1)}}.$$

10.29. В ситуации § 10.9 (a) предположим, что объемы, средние значения и дисперсии слоев неизвестны и используется следующая схема двойной выборки. Сначала извлекается простая случайная выборка объема r из π_N, в которой оказывается $r_1, \ldots, r_m$ элементов из слоев $\pi_{N_1}, \ldots, \pi_{N_m}$ соответственно. Затем извлекается расслоенная выборка объема n из π_N так, что $n_1=\frac{r_1}{r}n, \ldots, n_m=\frac{r_m}{r}n$ элементов извлекаются наудачу из слоев $\pi_{N_1}, \ldots, \pi_{N_m}$ соответственно. Пусть

$$\bar{x}'=\sum_g \frac{r_g}{r}\bar{x}_g,$$

где $\bar{x}_g$ — среднее значение выборки объема n_g из π_{N_g}. Показать, что $\bar{x}'$ — несмещенная оценка среднего значения μ совокупности π_N и дисперсия $\bar{x}'$ равна

$$\sigma^2(\bar{x}')=\frac{r\,(N-n-1)+n}{rnN}\sigma^2_W+\frac{N-r}{rN}\sigma^2_B-\frac{n-r}{rnN\,(N-1)}\sum_{g=1}^{m}\sigma^2_g.$$

10.30. (*Продолжение*). Пусть N достаточно велико, так что можно пренебречь членами порядка $\frac{1}{N}$, тогда

$$\sigma^2(\bar{x}')\cong\frac{\sigma^2_{W'}}{n}+\frac{\sigma^2_{B'}}{r},$$

где $\sigma^2_{W'}$, $\sigma^2_{B'}$ определены в (10.10.14). Предположим, что стоимость извлечения одного элемента первой выборки c_1, а одного элемента второй c_2. При фиксированной суммарной стоимости $c=c_1r+c_2n$ определить выбор r и n, минимизирующий $\sigma^2(\bar{x}')$.

10.31. Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности со средним значением μ и дисперсией σ^2, а u — произвольная линейная несмещенная оценка μ. Показать, что коэффициент корреляции между $\bar{x}$ и u равен $\frac{\sigma(\bar{x})}{\sigma(u)}$.

10.32. Пусть L_1 и L_2 — две линейные несмещенные оценки среднего значения совокупности, построенные по одной и той же выборке. Предположим, что дисперсии L_1 и L_2 равны σ_1^2 и σ_2^2, а коэффициент корреляции между L_1 и L_2 равен ρ. Определить постоянные $c_1 > 0$, $c_2 > 0$, $c_1 + c_2 = 1$ так, чтобы $c_1L_1 + c_2L_2$ имела минимальную дисперсию.

10.33. Пусть n, $\bar{x}$, s^2 — объем, среднее значение и дисперсия выборки из нормальной совокупности. Показать, что с вероятностью γ результат следующего независимого наблюдения над этой же совокупностью попадет в интервал $\bar{x} \pm t_{n-1,\,\gamma}\, s \sqrt{\dfrac{n+1}{n}}$, где $t_{n-1,\,\gamma}$ определена в (10.2.10).

10.34. *Поправки Шеппарда для моментов.* Пусть $x_1, \ldots, x_n$ — выборка из совокупности с п. в. $f(x)$; h — случайная величина, не зависимая с $(x_1, \ldots, x_n)$ и имеющая равномерное распределение $R(0, \delta)$. Образуем интервалы $I_\alpha = \left(h + \delta\alpha - \frac{1}{2}\delta,\ h + \delta\alpha + \frac{1}{2}\delta\right)$, $\alpha = \ldots, -2, -1, 0, +1, \ldots$ и пусть n_α — число элементов выборки, попавших в I_α. Величина

$$M'_{r,\,\delta} = \frac{1}{n}\sum_\alpha n_\alpha (h + \alpha\delta)^r$$

равняется r-му выборочному моменту, если элементы выборки округлены до ближайшего кратного δ, причем начало h выбирается «случайно». Показать, что характеристическая функция $\varphi(t)$ величины $M'_{r,\,\delta}$ равна

$$\varphi(t) = \frac{1}{\delta}\int\limits_{-\frac{1}{2}\delta}^{+\frac{1}{2}\delta}\left[\sum_\alpha p_\alpha e^{\frac{it(h+\alpha\delta)^r}{n}}\right]^n dh,$$

где $p_\alpha = \int\limits_{I_\alpha} f(x)\,dx$. Пользуясь этой характеристической функцией, показать, что

$$\mathcal{E}(M'_{1,\,\delta}) = \mu'_1; \qquad \mathcal{E}(M'_{2,\,\delta}) = \mu'_2 + \frac{\delta^2}{12}, \qquad \mathcal{E}(M'_{3,\,\delta}) = \mu'_3 + \frac{\delta^2\mu'_1}{4},$$

и вообще

$$\mathcal{E}(M'_{r,\,\delta}) = \frac{1}{\delta(r+1)}\cdot\left[\mathcal{E}\left(x + \frac{1}{2}\delta\right)^{r+1} - \mathcal{E}\left(x - \frac{1}{2}\delta\right)^{r+1}\right],$$

где $\mu'_1, \mu'_2, \ldots$ — моменты случайной величины x и

$$\mathcal{E}\left(x \pm \frac{1}{2}\delta\right)^{r+1} = \int\limits_{-\infty}^{+\infty}\left(x \pm \frac{1}{2}\delta\right)^{r+1} f(x)\,dx.$$

Следовательно, $M'_{1,\,\delta}$, $M'_{2,\,\delta} - \dfrac{\delta^2}{12}$, $M'_{3,\,\delta} - M'_{1,\,\delta}\dfrac{\delta^2}{4}$, ... будут несмещенными оценками для $\mu'_1, \mu'_2, \mu'_3, \ldots$ [Шеппард (1898)].

10.35. Пусть функция $f(x_1, \ldots, x_k)$ такова, что $0 < f(x_1, \ldots, x_k) < a_{k+1}$ в k-мерном параллелепипеде $I_k = \{(x_1, \ldots, x_k)\colon\ 0 \leqslant x_i \leqslant a_i,\ i = 1, \ldots, k\}$.

Тогда интеграл

$$I=\int\limits_0^{a_1}\ldots\int\limits_0^{a_k} f(x_1, \ldots, x_k)\,dx_1 \ldots dx_k$$

существует. Предположим, что $x_{1\xi}, \ldots, x_{(k+1)\xi}$ — независимые случайные величины с равномерными распределениями $R\left(\frac{1}{2}a_1, a_1\right), \ldots, R\left(\frac{1}{2}a_{k+1}, a_{k+1}\right)$, $\xi=1, \ldots, n$. Образуем случайную величину

$$z_\xi=\begin{cases}1, \text{ если } x_{(k+1)\xi}\leqslant f(x_{1\xi}, \ldots, x_{k\xi}),\\ 0, \text{ если } x_{(k+1)\xi}> f(x_{1\xi}, \ldots, x_{k\xi}),\end{cases}$$

$\xi=1, \ldots, n$, и положим

$$\bar{z}=\frac{1}{n}\sum_{\xi=1}^{n} z_\xi.$$

Показать, что $(a_1 \ldots a_{k+1})\,\bar{z}$ служит несмещенной оценкой I с дисперсией, не превосходящей

$$\frac{(a_1 \ldots a_{k+1})^2}{4n}.$$

ГЛАВА 11

НЕПАРАМЕТРИЧЕСКОЕ СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ

11.1. Вводные замечания

В главе 10 мы рассматривали задачи построения линейных и квадратичных оценок таких параметров конечных и бесконечных совокупностей как средние дисперсии, ковариации и коэффициенты регрессии. Как мы видели, процесс линейного оценивания, на котором основываются эти оценки, является сравнительно простой процедурой, требующей лишь мягких условий на входящие в оценку случайные переменные, а именно нужна конечность средних и матрицы ковариаций этих переменных.

В этой главе мы будем рассматривать другой класс довольно простых оценок для квантилей к. ф. р. изучаемых совокупностей. Как мы обнаружим, эти оценки не зависят от функциональной формы к. ф. р. совокупности, из которой извлечена выборка. Оценки этого типа иногда называют *непараметрическими оценками*, чтобы противопоставить их *параметрическим оценкам*, обсуждаемым в главе 12. Параметрические оценки, как мы увидим в главе 12, встречаются в задаче оценивания неизвестных параметров к. ф. р., имеющей специальную функциональную форму и зависящей от этих параметров.

Основные случайные переменные в непараметрическом оценивании — это определяемые выборкой порядковые статистики (вариационный ряд) и определяемые ими доли. Выборочная теория вариационного ряда и долей обсуждалась в § 8.7, и некоторые ее результаты используются в последующем изложении.

11.2. Доверительные интервалы для квантилей

(а) Случай малых выборок. Пусть мы имеем бесконечную совокупность с непрерывной к. ф. р. $F(x)$. Сначала мы будем рассматривать задачу построения доверительных интервалов для данной квантили $\underline{x}_p$ по вариационному ряду выборки из $F(x)$.

Более подробно, пусть $(x_1, \ldots, x_n)$ — выборка из $F(x)$ и $(x_{(1)}, \ldots$ $\ldots, x_{(n)})$ — вариационный ряд (порядковые статистики) этой выборки, определенный в § 8.7. Пусть $x_{(k_1)}$ и $x_{(k_1+k_2)}$ — две любые порядковые статистики. Мы знаем из **8.7.3**, что случайные переменные $u = F(x_{(k_1)})$, $v = F(x_{(k_1+k_2)})$, равные суммам долей, подчиняются двумерному распределению Дирихле $D^*(k_1, k_2; n-k_1-k_2+1)$, следовательно, в. э. случайной величины (u, v) равен

$$f(u, v)\,du\,dv = \\ = \frac{\Gamma(n+1)}{\Gamma(k_1)\Gamma(k_2)\Gamma(n-k_1-k_2+1)} u^{k_1-1}(v-u)^{k_2-1}(1-v)^{n-k_1-k_2}\,du\,dv \tag{11.2.1}$$

при $0<u<v<1$ и равен 0 при других значениях u и v. Рассмотрим теперь неравенство

$$F(x_{(k_1)})<p<F(x_{(k_1+k_2)}). \tag{11.2.2}$$

Так как $F(x)$ непрерывна, (11.2.2) выполняется тогда и только тогда, когда

$$x_{(k_1)}<\underline{x}_p<x_{(k_1+k_2)}. \tag{11.2.3}$$

Следовательно, вероятность выполнения (11.2.3) равна вероятности выполнения (11.2.2). Поэтому можно написать

$$P(x_{(k_1)}<\underline{x}_p<x_{(k_1+k_2)})=\int_p^1\int_0^p f(u, v)\,du\,dv. \tag{11.2.4}$$

Но

$$\int_p^1\int_0^p f(u, v)\,du\,dv=\int_0^p\int_u^1 f(u, v)\,dv\,du-\int_0^p\int_0^v f(u, v)\,du\,dv. \tag{11.2.5}$$

Делая замену переменных $u=r$, $v=1-s(1-r)$ в первом интеграле правой части (11.2.5) и $u=rs$, $v=s$ во втором, находим

$$P(x_{(k_1)}<\underline{x}_p<x_{(k_1+k_2)})=I_p(k_1, n-k_1+1)- \\ -I_p(k_1+k_2, n-k_1-k_2+1), \tag{11.2.6}$$

где $I_p(\nu_1, \nu_2)$ — неполная бета-функция

$$I_p(\nu_1, \nu_2)=\frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)}\int_0^p x^{\nu_1-1}(1-x)^{\nu_2-1}\,dx. \tag{11.2.7}$$

Как следует, в частности, отметить, вероятность того, что интервал $(x_{(k_1)}, x_{(k_1+k_2)})$ содержит квантиль $\underline{x}_p$, не зависит от $F(x)$, поэтому этот интервал является доверительным интервалом для $\underline{x}_p$ при доверительной вероятности, равной правой части (11.2.6).

Результат можно сформулировать следующим образом.

11.2.1. *Если $(x_1, \ldots, x_n)$ — выборка из распределения с непрерывной к. ф. р. $F(x)$ и если $x_{(k_1)}$ и $x_{(k_1+k_2)}$ — k_1-я и (k_1+k_2)-я порядковые статистики этой выборки, то $(x_{(k_1)}, x_{(k_1+k_2)})$ — доверительный интервал для квантили $\underline{x}_p$ при доверительной вероятности $I_p(k_1, n-k_1+1)-I_p(k_1+k_2, n-k_1-k_2+1)$.*

Нужно заметить, что, поскольку k_1 и k_2 — (положительные) целые числа, доверительные интервалы для $\underline{x}_p$ можно найти, используя вариационный ряд, лишь для таких доверительных вероятностей γ, которые принадлежат области значений правой части (11.2.6). Можно, конечно, построить доверительные интервалы, для которых доверительные вероятности не менее γ для определенных областей значений γ. Более того, две выделенные порядковые статистики произвольны: (11.2.6) справедливо для *любых* двух членов вариационного ряда $x_{(k_1)}$ и $x_{(k_1+k_2)}$. Обычно для данного γ выбираются две порядковые статистики с наиболее близкими номерами, т. е. выбираются k_1 и k_2 так, чтобы k_2 было как можно меньше. Например, при построении доверительных интервалов для медианы $\underline{x}_{0,5}$ обычно выбирают наибольшее значение k, для которого

$$P(x_{(k)} < \underline{x}_{0,5} < x_{(n-k+1)}) \geqslant \gamma. \tag{11.2.8}$$

Наир (1940) табулировал величины k, для которых выполняется (11.2.8) при $n=6, 7, \ldots, 81$ и $\gamma=0{,}95$ и $\gamma=0{,}99$. Точное значение вероятности $P(x_{(k)} < \underline{x}_{0,5} < x_{(n-k+1)})$ равно $1-2I_{0,5}(k, n-k+1)$, как независимо обнаружили Томпсон (1936) и Сэйвер (1937).

(b) Случай больших выборок. Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности с непрерывной к. ф. р. $F(x)$ и $\underline{x}_p$ — квантиль, отвечающая уровню p. Если n_1 — число компонент выборки, меньших $\underline{x}_p$, то n_1 — случайная величина, распределенная по биномиальному закону $\mathrm{Bi}(n, p)$. Мы знаем из **9.2.1a**, что для больших n статистика n_1 распределена асимптотически по закону $N(np, np(1-p))$. Следовательно, для данной доверительной вероятности γ мы имеем

$$\lim_{n\to\infty} P\left(-y_\gamma < \frac{n_1-np}{\sqrt{np(1-p)}} < +y_\gamma\right) = \gamma, \tag{11.2.9}$$

где

$$\frac{1}{\sqrt{2\pi}} \int_{-y_\gamma}^{y_\gamma} e^{-\frac{1}{2}y^2}\, dy = \gamma. \tag{11.2.10}$$

Таким образом, для больших n приближенный 100γ-процентный доверительный интервал $(\underline{p}_\gamma, \bar{p}_\gamma)$ для p образован множеством всех значений p, удовлетворяющих неравенству в (11.2.9) для фиксированных n_1, n и y_γ, т. е. $\underline{p}_\gamma$ и $\bar{p}_\gamma$ являются двумя решениями уравнения

$$\frac{(n_1-np)^2}{np(1-p)} = y_\gamma^2 \tag{11.2.11}$$

относительно p.

Поэтому для больших n

$$P(\underline{p}_\gamma < p < \bar{p}_\gamma) \cong \gamma, \tag{11.2.12}$$

что эквивалентно равенству

$$P(x_{[n\underline{p}_\gamma]} < \underline{x}_p < x_{[n\bar{p}_\gamma]}) \cong \gamma. \tag{11.2.13}$$

Итак, две порядковые статистики $(x_{([n\underline{p}_\gamma])}, x_{([n\bar{p}_\gamma])})$ образуют приближенный 100γ-процентный доверительный интервал для квантили $\underline{x}_p$, где $[n\underline{p}_\gamma]$ и $[n\bar{p}_\gamma]$ — ближайшие целые соответственно к $n\underline{p}_\gamma$ и $n\bar{p}_\gamma$.

11.3. Доверительные интервалы для интервалов квантилей

Пусть $\underline{x}_{p_1} < \underline{x}_{p_2}$ — любые две квантили, $x_{(k_1)}$ и $x_{(k_1+k_2)}$ — две порядковые статистики выборки объема n. Тогда аналогично тому, как мы вывели (11.2.4), мы найдем

$$P(x_{(k_1)} < \underline{x}_{p_1} < \underline{x}_{p_2} < x_{(k_1+k_2)}) = \int_0^{p_1} \int_{p_2}^{1} f(u, v)\, du\, dv; \qquad (11.3.1)$$

левую часть можно переписать в виде

$$P((\underline{x}_{p_1}, \underline{x}_{p_2}) \subset (x_{(k_1)}, x_{(k_1+k_2)})) =$$
$$= \frac{n!}{k_2!} \sum_{i=0}^{k_1-1} \frac{(-1)^i p_1^{i+k_1}}{i!\,(n-k_2-i)!} I_{p_1}(n-k_1-k_2+1,\ k_1-i). \qquad (11.3.2)$$

Если для данного γ существуют такие величины k_1, k_2 и n, что вероятность, выраженная равенством (11.3.2), принимает значение γ, то мы будем говорить, что $(x_{(k_1)}, x_{(k_1+k_2)})$ — 100γ-*процентный внешний доверительный интервал* для интервала квантилей $(\underline{x}_{p_1}, \underline{x}_{p_2})$.

Аналогично 100γ-*процентный внутренний доверительный интервал* $(x_{(k_1)}, x_{(k_1+k_2)})$ для $(\underline{x}_{p_1}, \underline{x}_{p_2})$ определяется, исходя из соотношения

$$P(\underline{x}_{p_1} < x_{(k_1)} < x_{(k_1+k_2)} < \underline{x}_{p_2}) = \int_{p_1}^{p_2} \int_{p_1}^{v} f(u, v)\, du\, dv, \qquad (11.3.3)$$

где k_1, k_2 и n имеют значения, при которых вероятность (11.3.3) равна γ и $f(u, v)$ определена по (11.2.1).

В случае внешнего (внутреннего) доверительного интервала за «наилучшие» k_1 и k_2 принимаются такие, при которых k_2 минимально (максимально). Например, если $p_1 = p$, $p_2 = 1 - p$, $2p < 1$, можно показать, что «наилучшему» в указанном смысле выбору k_1 и k_2 как для внутренних, так и для внешних доверительных интервалов отвечают статистики $x_{(c)}$ и $x_{(n-c+1)}$.

11.4. Доверительные интервалы квантилей конечных совокупностей

Пусть π_N — конечная совокупность с N различными значениями $x_{01} < \ldots < x_{0N}$. В § 8.8 было показано, что ф. в. k-й порядковой статистики $x_{(k)}$ выборки объема n из π_N равна

$$P_{N,n,k}(t) = \frac{\binom{t-1}{k-1}\binom{N-t}{n-k}}{\binom{N}{n}}, \qquad (11.4.1)$$

причем точки сосредоточения вероятностей для величин $x_{(k)}$ равны x_{0t}, $t=k, k+1, \ldots, N-n+k$. Для любого фиксированного значения t, скажем t',

$$P(x_{(k)} \leqslant x_{0t'}) = \sum_{t=k}^{t'} p_{N,n,k}(t). \tag{11.4.2}$$

Если для фиксированных N, n, t' и $\gamma > 0$ существует наибольшее k, скажем k', такое, что

$$\sum_{t=k'}^{t'} p_{N,n,k'}(t) \geqslant \gamma, \tag{11.4.3}$$

мы будем рассматривать $x_{(k')}$ как «наилучший» нижний 100γ-процентный доверительный предел для $x_{0t'}$. Мы можем рассматривать $x_{0t'}$ как квантиль совокупности π_N, отвечающую уровню t'/N. Как можно показать, такие доверительные пределы существуют для фиксированных величин, за исключением неинтересных малых значений N, n, t' и $1-\gamma$.

Подобным же образом «наилучший» верхний 100γ-процентный доверительный предел для $x_{0t'}$ получается выбором наименьшего k, скажем k'', для которого

$$\sum_{t=t'}^{N-n+k''} p_{N,n,k''}(t) \geqslant \gamma. \tag{11.4.4}$$

Можно также выписать «наилучший» 100γ-процентный доверительный интервал для $x_{0t'}$, т. е. одновременно верхний и нижний доверительные пределы, однако при этом ф. в. оказывается более громоздкой, чем ф. в. $p_{N,n,k}(t)$, и мы не будем ее выписывать.

Можно показать, что если в совокупности π_N элементы x принимают различные значения $x_{01} < \ldots < x_{0N}$, а $x_{(k_1)}$ и $x_{(k_1+k_2)}$ — указанные порядковые статистики выборки объема n из π_N, то

$$\lim_{n\to\infty} P(x_{(k_1)} < x_{0\,(Np)} < x_{(k_1+k_2)}) = \\ = I_p(k_1, n-k_1+1) - I_p(k_1+k_2, n-k_1-k_2+1). \tag{11.4.5}$$

Таким образом, для больших N интервал $(x_{(k_1)}, x_{(k_1+k_2)})$ является доверительным для $x_{0\,(Np)}$ с доверительной вероятностью, примерно равной правой части (11.4.5).

11.5. Толерантные пределы

(а) Случай малых выборок. Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности с непрерывной к. ф. р. $F(x)$ и $L_1(x_1, \ldots, x_n) < L_2(x_1, \ldots, x_n)$ — две любые симметричные функции наблюдений $(x_1, \ldots, x_n)$, такие, что распределение величины $F(L_2) - F(L_1)$ не зависит от $F(x)$ и для $0<\beta<1$

$$P[(F(L_2)-F(L_1))\geqslant\beta]=\gamma. \tag{11.5.1}$$

Тогда L_1 и L_2 будем называть 100β-*процентными независимыми от распределения толерантными пределами на уровне* γ. Это понятие ввел Шьюарт (1931).

Как было показано Уилксом (1941, 1942), порядковые статистики можно использовать для независимых от распределения толерантных пределов, если $F(x)$ непрерывна. Чтобы увидеть это, допустим, что $(x_{(1)}, \ldots, x_{(n)})$ — вариационный ряд выборки из неизвестного распределения с к. ф. р. $F(x)$. Для любых двух порядковых статистик $x_{(k_1)}$, $x_{(k_1+k_2)}$ вероятность, сосредоточенная в интервале $(x_{(k_1)}, x_{(k_1+k_2)})$, сосчитанная по данному распределению, является случайной величиной, которую мы обозначим U_{k_2}. Из § 8.7 (b) видно, что U_{k_2} равна сумме k_2 долей и что

$$U_{k_2}=F(x_{(k_1+k_2)})-F(x_{(k_1)}). \tag{11.5.2}$$

Как следует из **8.7.6**, U_{k_2} имеет бета-распределение $\mathrm{Be}(k_2, n-k_2+1)$, конечно, не зависящее от к. ф. р. $F(x)$ совокупности. Теперь допустим, что для данных значений $\beta>0$ и $\gamma>0$ существуют n, k_1 и k_2 такие, что

$$P(U_{k_2}\geqslant\beta)=\gamma. \tag{11.5.3}$$

Тогда $x_{(k_1)}$ и $x_{(k_1+k_2)}$ — 100β-процентные независимые от распределения толерантные пределы на уровне γ. Заметим, что U_{k_2} — частный случай выражения $F(L_2)-F(L_1)$ в (11.5.1). Роббинс (1944 b) показал, что, если $F(x)$ абсолютно непрерывна, т. е. имеет производную, то L_1 и L_2 — необходимо члены вариационного ряда.

Пусть $k_1=c$ и $k_2=n-2c+1$, что делает интервал симметричным относительно вариационного ряда. Если использовать обозначение неполной бета-функции, то (11.5.3) сведется к

$$1-I_\beta(n-2c+1,\ 2c)=\gamma. \tag{11.5.4}$$

Для фиксированного c может не существовать такого объема выборки n, чтобы это равенство выполнялось точно. Однако, так как сосредоточенная в интервале $(1-\varepsilon, 1)$ вероятность, рассчитанная по бета-распределению $\mathrm{Be}(n-2c+1, 2c)$, для любого $\varepsilon>0$ может быть сделана как угодно близкой к 1, если взять достаточно большое n, существует *наименьшее* n, для которого

$$1-I_\beta(n-2c+1,\ 2c)\geqslant\gamma. \tag{11.5.5}$$

Например, если мы выберем $\beta=0{,}99$, $\gamma=0{,}95$ и $c=1$ (т. е. $k_1=1$, $k_2=n-1$), найдем $n=473$. Мёрфи (1948) табулировал и представил в графической форме значения β, для которых (11.5.3) выполняется при $n=1\,(1)\,10\,(10)\,100\,(100)\,500$, $\gamma=0{,}90$, $\gamma=0{,}95$, $\gamma=0{,}99$ и

$n-k_2+1=n=1\,(1)\,6\,(2)\,10\,(5)\,30\,(10)\,60\,(20)\,100$. Сомервилл (1958) расширил таблицу Мёрфи.

Понятие независимых от распределения толерантных пределов для одномерных распределений с непрерывными к. ф. р. обобщается на случай многомерных распределений с непрерывными к. ф. р. Пусть мы имеем выборку объема n из бесконечной совокупности с непрерывной k-мерной к. ф. р. $F(x_1, \ldots, x_k)$. Пусть k-мерное выборочное пространство $(x_1, \ldots, x_k)$ разбито на $(n+1)$ не имеющих общих точек и заполняющих все пространство выборочных блоков $B_k^{(1)}, \ldots, B_k^{(n+1)}$ с помощью упорядочивающих функций, как это описано в § 8.7 (с). Возьмем любое правило выбора некоторых r из этих блоков и обозначим их объединение T_r. Пусть сумма долей для выбранных блоков будет U_r, т. е.

$$U_r = \int\limits_{T_r} dF(x_1, \ldots, x_k). \tag{11.5.6}$$

Тогда U_r — случайная величина, и из **8.7.9** следует, что U_r имеет бета-распределение $\mathrm{Be}(r, n-r+1)$. Если для данных $\beta > 0$ и $\gamma > 0$

$$P(U_r \geqslant \beta) = \gamma, \tag{11.5.7}$$

то U_r можно назвать 100β-процентной независимой от распределения *толерантной областью* на уровне γ. Если для любого фиксированного положительного целого c мы выберем $r=n-2c+1$, то (11.5.7) сводится к (11.5.4), к тому же уравнению, которое мы имели для одномерных толерантных пределов.

(b) Случай конечных совокупностей. Если t' выбрано так, что $t'/N = 1-\beta$, то (11.4.2) можно переписать в виде

$$P\left(x_{(k')} \leqslant x_{0\,(N(1-\beta))}\right) \geqslant \gamma. \tag{11.5.8}$$

В этом случае можно рассматривать $x_{(k')}$ как 100 β-*процентный нижний толерантный предел* на уровне γ, т. е. с вероятностью, не меньшей γ, доля элементов π_N, значения x которых превышают $x_{(k')}$, равна β.

Аналогично, если $t' = \beta N$, мы можем переписать (11.4.4)

$$P\left(x_{(k'')} \geqslant x_{0\,(N\beta)}\right) \geqslant \gamma, \tag{11.5.9}$$

т. е. $x_{(k'')}$ — 100 β-процентный верхний толерантный предел на уровне γ.

Ясно, как определить 100 β-*процентный толерантный интервал* на уровне γ.

Если для большей совокупности π_N, элементы которой принимают различные значения x, мы обозначим $U_{k_2,N}$ часть величин x, лежащих в интервале $(x_{(k_1)}, x_{(k_1+k_2)})$, можно показать, что предельное распределение $U_{k_2,N}$ при $N \to \infty$ — $\mathrm{Be}(k_2, n-k_2+1)$. Следовательно, при больших N указанный интервал является приближенным 100 β-процентным толерантным интервалом на уровне γ.

11.6. Односторонние границы для непрерывной функции распределения

Одной из основных проблем применения выборочной теории вариационного ряда к статистическим выводам представляется построение функций $F_n^+(x)$ и $F_n^-(x)$ по выборке из совокупности с непрерывной к. ф. р. $F(x)$ таких, что для данного γ

$$\begin{aligned} &P(F_n^+(x) \geqslant F(x) \text{ для всех } x) = \gamma, \\ &P(F_n^-(x) \leqslant F(x) \text{ для всех } x) = \gamma. \end{aligned} \tag{11.6.1}$$

Такие функции $F_n^+(x)$ и $F_n^-(x)$ будем называть соответственно *верхней и нижней* $100\,\gamma$*-процентной доверительной границей* для $F(x)$.

Асимптотическое решение этой проблемы для больших n было первоначально получено Смирновым (1939a). Позднее Бирнбаум и Тинджи (1951) получили сравнительно простое точное выражение для вероятностей (11.6.1), которое мы и приведем*).

Пусть $(x_{(1)}, \ldots, x_{(n)})$ — вариационный ряд выборки из бесконечной совокупности с к. ф. р. $F(x)$. Эмпирическая к. ф. р. $F_n(x)$ строится по вариационному ряду следующим образом:

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \dfrac{\xi-1}{n}, & x_{(\xi-1)} \leqslant x < x_{(\xi)}, \quad \xi = 2, \ldots, n, \\ 1, & x \geqslant x_{(n)}. \end{cases} \tag{11.6.2}$$

Для $0 < d < 1$ и любого x пусть

$$\begin{aligned} F_n^+(x, d) &= \min\,[F_n(x) + d;\ 1], \\ F_n^-(x, d) &= \max\,[F_n(x) - d;\ 0] \end{aligned} \tag{11.6.3}$$

и

$$\begin{aligned} D_n^+(d) &= \sup_x\,(F_n^+(x, d) - F(x)), \\ D_n^-(d) &= \inf_x\,(F_n^-(x, d) - F(x)). \end{aligned} \tag{11.6.4}$$

Заметим, что $P(D_n^+(d) > 0)$ — вероятность того, что график $F_n(x) + d$ нигде не имеет общих точек с графиком $F(x)$; аналогично $P(D_n^-(d) < 0)$ — вероятность того, что график $F_n(x) - d$ не имеет общих точек с графиком $F(x)$.

Основной результат Бирнбаума и Тинджи можно сформулировать так:

11.6.1. *Пусть* $(x_{(1)}, \ldots, x_{(n)})$ — *вариационный ряд выборки из совокупности с непрерывной к. ф. р.* $F(x)$, $F_n^+(x, d)$ *и* $F_n^-(x, d)$ —

*) Точное выражение (11.6.5) значительно раньше нашел Н. В. Смирнов (см. УМН **10** (1944), стр. 179—206). (*Прим. перев.*)

построены из эмпирической к. ф. р. $F_n(x)$ *по формуле* (11.6.3). *Тогда* $P(D_n^+(d)>0)=P(D_n^-(d)<0)=P_n(d)$, *где*

$$P_n(d)=1-d\sum_{i=1}^{[n(1-d)]}\binom{n}{i}\left(1-d-\frac{i}{n}\right)^{n-i}\left(d+\frac{i}{n}\right)^{i-1} \qquad (11.6.5)$$

и $[n(1-d)]$ — *целая часть числа* $n(1-d)$.

Чтобы доказать **11.6.1**, перейдем сначала к новым случайным переменным $y_{(1)}, \ldots, y_{(n)}$ по формуле

$$y_{(\xi)}=F(x_{(\xi)}), \qquad \xi=1, \ldots, n. \qquad (11.6.6)$$

Тогда мы знаем из (8.7.2), что $y_{(1)}, \ldots, y_{(n)}$ имеют в. э., равный

$$n!\,dy_{(1)}\ldots dy_{(n)} \qquad (11.6.7)$$

в области $0<y_{(1)}<\ldots<y_{(n)}<1$ и 0 при других значениях переменных.

Сначала рассмотрим $P(D_n^+(d)>0)$. Очевидно, что вариационный ряд $x_{(1)}, \ldots, x_{(n)}$ будет удовлетворять неравенству $D_n^+(d)>0$ тогда и только тогда, когда вариационный ряд $y_{(1)}, \ldots, y_{(n)}$ удовлетворяет следующим неравенствам:

$$\begin{aligned} &y_{(\xi-1)}<y_{(\xi)}<\frac{\xi-1}{n}+d, && \xi=1, \ldots, k+1,\\ &y_{(\xi-1)}<y_{(\xi)}<1, && \xi=k+2, \ldots, n, \end{aligned} \qquad (11.6.8)$$

где k — целая часть числа $n(1-d)$ и $y_{(0)}=0$. Аналогично $D_n^-(d)<0$ удовлетворяется в том и только том случае, когда

$$\begin{aligned} &\frac{\xi}{n}-d<y_{(\xi)}<y_{(\xi+1)}, && \xi=n-k, \ldots, n,\\ &0<y_{(\xi)}<y_{(\xi+1)}, && \xi=1, \ldots, n-k+1, \end{aligned} \qquad (11.6.9)$$

где $y_{(n+1)}=1$. Но если сделать замену переменных $y_{(\xi)}=1-y'_{(n-\xi+1)}$, $\xi=1, \ldots, n$, то случайная величина $(y'_{(1)}, \ldots, y'_{(n)})$ будет иметь в. э. (11.6.7) и неравенства (11.6.9) сведутся к (11.6.8). Следовательно, $P(D_n^+(d)>0)=P(D_n^-(d)<0)$. Общее значение этих вероятностей, обозначим его $P_n(d)$, получается интегрированием в. э. (11.6.7) по области, определяемой неравенствами (11.6.8) и принадлежащей области $0<y_{(1)}<\ldots<y_{(n)}<1$, т. е.

$$P_n(d)=n!\,G_n(k, d), \qquad (11.6.10)$$

где

$$G_n(k, d)=\int_0^d\int_{y_{(1)}}^{\frac{1}{n}+d}\ldots\int_{y_{(k)}}^{\frac{k}{n}+d}\int_{y_{(k+1)}}^{1}\ldots\int_{y_{(n-1)}}^{1}dy_{(n)}\ldots dy_{(1)}. \qquad (11.6.11)$$

Интегрируя по $y_{(n)}, \ldots, y_{(k+2)}$, находим

$$G_n(k, d) = \int_0^d \int_{y_{(1)}}^{\frac{1}{n}+d} \cdots \int_{y_{(k)}}^{\frac{k}{n}+d} \frac{(1-y_{(k+1)})^{n-k-1}}{(n-k-1)!} dy_{(k+1)} \ldots dy_{(1)}. \quad (11.6.12)$$

Отсюда значение $G_n(k, d)$ можно найти по индукции. Так, интегрируя по $y_{(k+2)}$ в $G_n(k+1, d)$, найдем

$$G_n(k+1, d) = G_n(k, d) - \frac{\left(1-d-\frac{k+1}{n}\right)^{n-k-1}}{(n-k-1)!} H_n(k, d), \quad (11.6.13)$$

где

$$H_n(k, d) = \int_0^d \int_{y_{(1)}}^{\frac{1}{n}+d} \cdots \int_{y_{(k)}}^{\frac{k}{n}+d} dy_{(k+1)} \ldots dy_{(1)}. \quad (11.6.14)$$

Опуская детали математической индукции, находим

$$H_n(k, d) = -\frac{d}{(k+1)!} \sum_{i=1}^{k+1} (-1)^i \binom{k+1}{i} \left(d + \frac{k+1-i}{n}\right)^k, \quad (11.6.15)$$

что можно записать в виде

$$H_n(k, d) = -\frac{d}{(k+1)!} \frac{\partial^k}{\partial t^k} \left\{ e^{\left(d+\frac{k+1}{n}\right)t} \left[\left(1-e^{-\frac{t}{n}}\right)^{k+1} - 1 \right] \right\}_{t=0}. \quad (11.6.16)$$

Но

$$\left\{ \frac{\partial^k}{\partial t^k} \left[e^{\left(\frac{d}{k+1}+\frac{1}{n}\right)t} - e^{\frac{d}{k+1}t} \right]^{k+1} \right\}_{t=0} = 0. \quad (11.6.17)$$

Поэтому

$$H_n(k, d) = \frac{d}{(k+1)!} \left\{ \frac{\partial^k}{\partial t^k} e^{\left(d+\frac{k+1}{n}\right)t} \right\}_{t=0} = \frac{d}{(k+1)!} \left(d + \frac{k+1}{n}\right)^k. \quad (11.6.18)$$

Подставляя значение $H_n(k, d)$ в (11.6.13) и замечая, что

$$G_n(0, d) = \frac{1-(1-d)^n}{n!}, \quad (11.6.19)$$

находим по индукции значение $G_n(k, d)$. Подставляя $G_n(k, d)$ в (11.6.10) и замечая, что k — целая часть числа $n(1-d)$, мы завершаем доказательство **11.6.1.**

Бирнбаум и Тинджи табулировали значения d, для которых $P_n(d) = \gamma$ при $\gamma = 0{,}90$; $0{,}95$; $0{,}99$; $0{,}999$ и $n = 5, 8, 10, 20, 40, 50$. Для значений n, превышающих 50, асимптотическая аппроксимация Смирнова (11.6.22) дает хорошее приближение для d.

Теперь рассмотрим предел $P_n(d)$ при $n \to \infty$. Положив $d = \frac{\lambda}{\sqrt{n}}$, можем переписать (11.6.5) в следующем виде:

$$P_n\left(\frac{\lambda}{\sqrt{n}}\right) = 1 - \lambda \left\{ \sum_{i=1}^{[n-\sqrt{n}\lambda]} \sqrt{n}\binom{n}{i}\left(1 - \frac{\lambda}{\sqrt{n}} - \frac{i}{n}\right)^{n-i} \left(\frac{\lambda}{\sqrt{n}} + \frac{i}{n}\right)^{i-1} \Delta y \right\}, \tag{11.6.20}$$

где $\Delta y = \frac{1}{n}$. Применяя приближение Стирлинга для больших факториалов (7.6.29), можно убедиться в том, что сумма в фигурных скобках сходится к интегралу

$$\frac{1}{\sqrt{2\pi}} \int_0^1 \frac{1}{\sqrt{y(1-y)}} e^{-\frac{\lambda^2}{2y(1-y)}} dy \tag{11.6.21}$$

при $n \to \infty$, который после интегрирования дает

$$\frac{1}{\lambda} e^{-2\lambda^2}.$$

Таким образом, имеем следующий результат:

11.6.2. *В условиях* **11.6.1**

$$\lim_{n \to \infty} P_n\left(\frac{\lambda}{\sqrt{n}}\right) = 1 - e^{-2\lambda^2}. \tag{11.6.22}$$

Этот результат был первоначально установлен Смирновым (1939a), но более прямой вывод его из (11.6.5) дал Демпстер (1955).

11.7. Доверительные полосы для непрерывной функции распределения

Задача нахождения двусторонних границ, или *доверительной полосы*, для $F(x)$, т. е. определения вероятности

$$P(D_n < d), \tag{11.7.1}$$

где

$$D_n = \sup_x |F_n(x) - F(x)|,$$

для произвольного d более трудна, чем задача установления односторонней границы, обсуждавшаяся в предыдущем параграфе. Колмогоров (1933b) дал асимптотическое решение проблемы для больших n и рекуррентные формулы, по которым можно сосчитать величину (11.7.1) для конечных n. Вальд и Вольфовитц (1939) дали решение для конечных n в детерминантной форме. Позднее Масси (1950) нашел довольно простое решение проблемы, если n конечно и d кратно $\frac{1}{n}$. Его решение имеет вид рекуррентных формул, которые мы здесь и приведем. Результат Масси можно сформулировать так.

11.7.1. *Пусть* $(x_{(1)}, \ldots, x_{(n)})$ — *вариационный ряд выборки объема* n *из совокупности с непрерывной к. ф. р.* $F(x)$ *и*

$$D_n = \sup_x |F_n(x) - F(x)|, \tag{11.7.2}$$

где $F_n(x)$ *определяется по* (11.6.2). *Тогда*

$$P\left(D_n < \frac{k}{n}\right) = \frac{n!}{n^n} U(k, n), \qquad k = 1, \ldots, n-1, \tag{11.7.3}$$

где $U(j, m+1)$, $j = 1, \ldots, 2k-1$; $m = 0, 1, \ldots, n-1$, *удовлетворяют системе уравнений*

$$U(j, m+1) = \sum_{i=1}^{j+1} \frac{U(i, m)}{(j+1-i)!} \tag{11.7.4}$$

с граничными условиями

$$\begin{aligned} U(i, m) &= 0, \qquad i \geqslant m+k, \\ U(i, 0) &= 0, \qquad i = 1, \ldots, k-1, \\ U(k, 0) &= 1. \end{aligned} \tag{11.7.5}$$

Докажем этот результат. Пусть $(x_1, \ldots, x_n)$ — выборка из распределения с непрерывной к. ф. р. $F(x)$. Для удобства произведем замену переменной $y = F(x)$. Случайная величина y имеет равномерное распределение $R\left(\frac{1}{2}, 1\right)$ на интервале (0,1). Пусть $G(y)$ — к. ф. р. величины y и $G_n(y)$ — эмпирическая к. ф. р., построенная по вариационному ряду $(y_{(1)}, \ldots, y_{(n)})$. Само собой разумеется, что

$$P\left(D_n < \frac{k}{n}\right) = P\left(\sup_y |G_n(y) - G(y)| < \frac{k}{n}\right). \tag{11.7.6}$$

Разделим интервал (0,1) на n равных интервалов $I_\xi = \left(\frac{\xi-1}{n}, \frac{\xi}{n}\right)$, $\xi = 1, \ldots, n$. Пусть случайная величина $(r_1, \ldots, r_n)$ (вырожденная, удовлетворяющая условию $r_1 + \ldots + r_n = n$) указывает, сколько элементов выборки $(y_1, \ldots, y_n)$ попали соответственно в $I_1, \ldots, I_n$. Величина $(r_1, \ldots, r_n)$ имеет, очевидно, $(n-1)$-мерное полиномиальное распределение $M\left(n; \frac{1}{n}, \ldots, \frac{1}{n}\right)$ с ф. в.

$$p(r_1, \ldots, r_n) = \frac{n!}{r_1! \ldots r_n! \, n^n} \tag{11.7.7}$$

и однозначно определяет $G_n(y)$. Поэтому $P\left(D_n < \frac{k}{n}\right)$ находится суммированием $p(r_1, \ldots, r_n)$ по всем точкам выборочного пространства $(r_1, \ldots, r_n)$, для которых $|G_n(y) - G(y)| < \frac{k}{n}$ для всех y. Если двигаться слева направо по графику $G_n(y)$, лежащему полностью внутри полосы E, для которой $|G_n(y) - G(y)| < \frac{k}{n}$ при $y = \frac{1}{n}, \frac{2}{n}, \ldots, \frac{m+1}{n}$, то график должен проходить через одну из точек $\left(\frac{m}{n}, \frac{m-k+i}{n}\right)$, $i=1, \ldots, 2k-1$. Назовем эти точки $A(i, m)$, $i=1, \ldots, 2k-1$. Пусть

$$U(i, m)=\sum_{(i)} \frac{1}{r_1! \ldots r_m!}, \qquad (11.7.8)$$

где $\sum_{(i)}$ означает суммирование по всему множеству значений $(r_1, \ldots, r_n)$, для которых график $G_n(y)$ приходит в $A(i, m)$, оставаясь внутри полосы E. Тогда, так как $G_n(y)$ не убывает, график может достичь $A(j, m+1)$, только пройдя через одну из точек $A(1, m), A(2, m), \ldots$ $\ldots, A(j+1, m)$, причем r_{m+1} примет соответственно значения j, $j-1, \ldots, 1, 0$. Следовательно, мы должны иметь

$$U(j, m+1)=\frac{U(1, m)}{j!}+\frac{U(2, m)}{(j-1)!}+\ldots+\frac{U(j+1, m)}{0!} \qquad (11.7.9)$$

для $j=1, 2, \ldots, 2k-1$ и $m=0, 1, \ldots, n-1$, причем, конечно, $U(i, m)=0$, если $i \geqslant m+k$. Более того, очевидно, что $U(i, m)$ удовлетворяет следующим граничным условиям:

$$\begin{aligned} &U(i, 0)=0, \qquad i=1, \ldots, k-1, \\ &U(k, 0)=1. \end{aligned} \qquad (11.7.10)$$

Если решить систему разностных уравнений (11.7.9) для $U(k, n)$ с граничными условиями (11.7.8), мы получим для $P\left(D_n<\frac{k}{n}\right)$ формулу (11.7.3), что и доказывает **11.7.1.**

Значения $P\left(D_n<\frac{k}{n}\right)$ для $n=5(5)80$ и $k=1(1)9$ были табулированы Масси, и по этим значениям с помощью интерполяции он рассчитал величины $P\left(D_n<\frac{\lambda}{\sqrt{n}}\right)$ для $n=10(10)80$ и $\lambda=0{,}9(0{,}1)1{,}40$.

Асимптотический результат Колмогорова (1933b), упомянутый ранее, найден в условиях **11.7.1** и имеет вид

$$\lim_{n \to \infty} P\left(D_n<\frac{\lambda}{\sqrt{n}}\right)=\sum_{i=-\infty}^{\infty}(-1)^i e^{-2i^2\lambda^2}. \qquad (11.7.11)$$

Доказательство Колмогорова сложно. Более простое доказательство найдено Феллером (1948). Дуб (1949) и Донскер (1952) провели доказательство с применением теории гауссовских случайных процессов. Дарлинг (1957) опубликовал обзорную статью по данной и связанной с ней проблемам с большой библиографией.

Заметим, что из (11.7.11) следует для произвольного $\varepsilon>0$

$$\lim_{n \to \infty} P(D_n<\varepsilon)=1, \qquad (11.7.12)$$

т. е. $F_n(x)$ сходится по вероятности к $F(x)$ равномерно по x при $n \to \infty$.

ЗАДАЧИ

11.1. Пусть E — множество точек, удовлетворяющих условию $u<p$, в выборочном пространстве (u, v) с в. э. (11.2.1), F — множество точек, для которых $v>p$. Вывести (11.2.6) из основного вероятностного закона $P(E \cup F) = P(E)+P(F)-P(E \cap F)$.

11.2. Показать, что вероятность (11.2.8) равна

$$P(x_{(k)} < \underline{x}_{0,5} < x_{(n-k+1)}) = 1-2I_{0,5}(n-k+1,\ k).$$

11.3. Пусть $(x_{(1)}, \ldots, x_{(n)})$ — вариационный ряд выборки из распределения с непрерывной к. ф. р. $F(x)$. Показать, что с вероятностью $1-n\beta^{n-1}+(n-1)\beta^n$

$$F(x_{(n)})-F(x_{(1)}) > \beta$$

(т. е. сумма долей, содержащихся между крайними членами вариационного ряда, превосходит β с вероятностью $1-n\beta^{n-1}+(n-1)\beta^n$).

11.4. (*Продолжение*). Показать, что для любых положительных δ_1 и δ_2, удовлетворяющих условию $\delta_1+\delta_2<1$, вероятность совместного выполнения двух неравенств

$$F(x_{(1)})<\delta_1, \quad 1-F(x_{(n)})<\delta_2$$

равна

$$1-(1-\delta_1)^n-(1-\delta_2)^n+(1-\delta_1-\delta_2)^n.$$

11.5. (*Продолжение*). Показать, что для любого фиксированного целого $k<\frac{n}{2}$ максимум вероятности $P(x_{(k+r)} < \underline{x}_{0,5} < x_{(n-k+r+1)})$ по всем возможным значениям r (положительным и отрицательным целым числам и нулю) достигается при $r=0$, причем $F(\underline{x}_{0,5})=0{,}5$.

11.6. (*Продолжение*). Показать, что

$$P(x_{(k)} < \underline{x}_p) = I_p(k,\ n-k+1),$$

где $F(\underline{x}_p)=p$.

11.7. (*Продолжение*). Показать, что интервал $(x_{(k)}, x_{(n-k+1)})$ можно рассматривать как доверительный при доверительной вероятности

$$\gamma_m = \sum_{t=k}^{n-k} p_m(t)$$

для медианы новой независимой выборки объема $2m+1$ из того же распределения. Здесь

$$p_m(t) = \frac{(m+1)}{(m+t+1)} \frac{\binom{2m+1}{m}\binom{n}{t}}{\binom{2m+n+1}{m+t+1}}.$$

Показать, что предел γ_m при $m\to\infty$ равен $1-2I_{0,5}(n-k+1,\ k)$.

11.8. В (11.2.6) показать, что при $p=0{,}5$

$$P(x_{(k_1)} < \underline{x}_{0,5} < x_{(k_1+k_2)}) = \left(\frac{1}{2}\right)^n \sum_{\xi=k_1}^{k_1+k_2-1} \binom{n}{\xi}.$$

11.9. Если $(x_{(1)}, \ldots, x_{(n)})$ — вариационный ряд выборки объема n из конечной совокупности объема $2m+1$, элементы которой принимают различные значения x, то показать, что $(x_{(k)}, \ldots, x_{(n-k+1)})$ — доверительный интервал для медианы совокупности при доверительной вероятности γ_m, совпадающей с аналогичной величиной предшествующей задачи.

11.10. Если выборка объема n извлечена из совокупности с непрерывной к. ф. р., показать, что с вероятностью

$$\frac{n(n-1)}{(m+n)} \frac{\binom{m}{t}}{\binom{m+n-1}{m-t+1}}$$

t величин x в другой независимой выборке объема m окажется между крайними членами вариационного ряда первой выборки.

11.11. (*Продолжение*). Пусть $U_{1,n,m}$ — доля x во второй выборке, лежащих между крайними членами первой. Показать, что предельным распределением $U_{1,n,m}$ при $m \to \infty$ является бета-распределение Be $(n-1, 2)$. Вообще, если $(x)_{k_1}$ и $x_{(k_1+k_2)}$ — две порядковые статистики первой выборки и $U_{k_1, k_1+k_2, m}$ — доля x второй выборки, лежащих в интервале $(x_{(k_1)}, x_{(k_1+k_2)})$, показать, что предельным распределением $U_{k_1, k_1+k_2, m}$ является бета-распределение Be $(k_2, n-k_2+1)$.

11.12. Пусть выборка объема n из совокупности с непрерывной двумерной к. ф. р. представлена n точками в плоскости xy. Проведем вертикали через две точки с наименьшей и наибольшей координатой x. Горизонтали проведем через две точки из оставшихся $(n-2)$ точек с наименьшей и наибольшей координатой y. Рассмотреть долю U совокупности, заключенной в прямоугольнике, ограниченном этими четырьмя линиями. Показать, что с вероятностью

$$1-4\binom{n}{4}\beta^{n-3}\left[\frac{1}{n-3}-3\cdot\frac{\beta}{n-2}+3\cdot\frac{\beta^2}{n-1}-\frac{\beta^3}{n}\right]$$

доля U превосходит β.

11.13. Определив U_{n+1-k}, как в параграфе 11.5 (a), показать, что для фиксированного k, $n \to \infty$, $\beta \to 1$, так что $n(1-\beta) \to \lambda$,

$$\lim P(U_{n+1-k} > \beta) = 1 - e^{-\lambda}\left[1+\lambda+\frac{\lambda^2}{2}+\ldots+\frac{\lambda^k}{k!}\right].$$

11.14. Проверить (11.4.5).

11.15. Проверить, что $H_n(k, d)$ выражается по (11.6.18).

11.16. Показать, что выражение в фигурных скобках в (11.6.20) сходится к интегралу (11.6.21) при $n \to \infty$.

11.17. Пусть $\bar{x}$ и s^2 — среднее и дисперсия выборки объема n из $N(\mu, \sigma^2)$ и $G(x)$ — к. ф. р. $N(\mu, \sigma^2)$. Пусть (L_1, L_2) — интервал $\bar{x} \pm \lambda s$, $\lambda > 0$. Показать, что $G(L_2) - G(L_1)$ как доля имеет распределение, зависящее только от λ, и что для $\lambda = t_{n-1,\gamma}\sqrt{\frac{n+1}{n}}$, где $t_{n-1,\gamma}$ определяется по (10.2.10), $\mathscr{E}(G(L_2) - G(L_1)) = \gamma$. [См. Уилкс (1941) и Вальд и Вольфовитц (1946).]

ГЛАВА 12

ПАРАМЕТРИЧЕСКОЕ СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ

Во многих задачах статистического оценивания приходится иметь дело с выборкой из совокупности с к. ф. р., имеющей конкретную функциональную форму $F(x;\ \theta)$, где θ — неизвестный (вещественный) параметр, истинное значение которого θ_0 нужно оценить по элементам выборки $(x_1, \ldots, x_n)$. Истинное значение θ_0 — точка из *параметрического пространства* Ω значений θ, где Ω — открытый интервал или область эвклидова пространства или же все пространство. Параметрическое пространство Ω иногда называют множеством *допустимых значений* θ. Нет необходимости говорить, что как x, так и θ могут быть многомерными.

Имеются два важных типа параметрических оценок, именно: *точечные оценки* и *интервальные оценки*. При точечном оценивании оценкой для истинного значения θ_0 параметра θ в $F(x;\ \theta)$ служит наблюдаемая случайная величина $\tilde{\theta}(x_1, \ldots, x_n)$, функция элементов выборки $(x_1, \ldots, x_n)$, распределение которой в некотором смысле концентрируется вокруг истинного значения θ_0 параметра θ. Как и при линейном оценивании, дисперсия точечной оценки часто дает разумный критерий для измерения ее концентрации.

При интервальном оценивании мы строим две наблюдаемые случайные величины $\underline{\theta}(x_1, \ldots, x_n)$, $\bar{\theta}(x_1, \ldots, x_n)$, обычно сокращенно записываемые $(\underline{\theta}, \bar{\theta})$, где $\underline{\theta} < \bar{\theta}$, такие, что с заданной вероятностью случайный интервал $(\underline{\theta}, \bar{\theta})$ содержит θ_0 и является в некотором смысле кратчайшим. Эти идеи обобщаются, конечно, на случай многомерного x или θ. Вместо того чтобы говорить «оценка $\tilde{\theta}$ для истинного значения θ_0 параметра θ», мы обычно будем говорить «оценка $\tilde{\theta}$ для θ».

Линейные оценки, рассмотренные в главе 10, дают примеры точечных оценок. Конкретные примеры интервальных оценок даны в §§ 10.2 (с) 10.4, 10.5, 10.6 и 11.2 (а) для выборок из нормального распределения.

В этой главе мы будем развивать идеи точечного и интервального оценивания в более общей постановке параметрического оценивания для выборок из бесконечных совокупностей и затем изложим некоторые основные результаты для конечных выборок, а также некоторые асимптотические результаты для больших выборок.

12.1. Дифференцирование параметрических функций распределения

(а) Случай одномерного параметра. При параметрическом статистическом оценивании, проверке параметрических статистических гипотез и в близких задачах нам понадобятся некоторые свойства производных функций распределения, зависящих от параметра θ. Будет полезно обсудить здесь кратко эти вопросы.

Пусть x — случайная величина с к. ф. р. $F(x;\ \theta)$, где θ — (действительный) одномерный параметр, принимающий значения из *параметрического пространства* Ω. Для удобства будем считать x одномерной случайной величиной, хотя будет видно, что все полученные результаты справедливы при небольших изменениях в обозначениях и для k-мерной случайной величины x.

Итак, для различных точек $\theta_1, \theta_2, \ldots$ из Ω мы имеем соответствующий набор к. ф. р. $F(x;\ \theta_1)$, $F(x;\ \theta_2), \ldots$ Особенно интересен открытый интервал Ω_0 из Ω, содержащий точку θ_0 — *истинное значение* θ. Рассмотрим тождество

$$\int_{-\infty}^{\infty} dF(x;\ \theta) = 1. \tag{12.1.1}$$

Если формально продифференцировать обе части (12.1.1) дважды, мы получим

$$\frac{d}{d\theta}\int_{-\infty}^{\infty} dF(x;\ \theta) = \int_{-\infty}^{\infty}\left[\frac{\partial}{\partial\theta}\log dF(x;\ \theta)\right] dF(x;\ \theta) = 0 \tag{12.1.2}$$

и

$$\frac{d^2}{d\theta^2}\int_{-\infty}^{\infty} dF(x;\ \theta) = \int_{-\infty}^{\infty}\left[\frac{\partial^2}{\partial\theta^2}\log dF(x;\ \theta)\right] dF(x;\ \theta) + \\ + \int_{-\infty}^{\infty}\left[\frac{\partial}{\partial\theta}\log dF(x;\ \theta)\right]^2 dF(x;\ \theta) = 0. \tag{12.1.3}$$

Для того чтобы исследовать (12.1.2) и (12.1.3) подробнее, и для дальнейших ссылок введем обозначения:

$$S(x;\ \theta) = \frac{\partial}{\partial\theta}\log dF(x;\ \theta), \tag{12.1.4}$$

$$S'(x;\ \theta) = \frac{\partial}{\partial\theta} S(x;\ \theta), \tag{12.1.5}$$

$$H(\theta,\ \theta') = \int_{-\infty}^{\infty}\log dF(x;\ \theta')\, dF(x;\ \theta), \tag{12.1.6}$$

$$A(\theta,\ \theta') = \int_{-\infty}^{\infty} S(x;\ \theta')\, dF(x;\ \theta), \tag{12.1.7}$$

$$B^2(\theta,\ \theta') = \int_{-\infty}^{\infty} [S(x;\ \theta')]^2\, dF(x;\ \theta), \tag{12.1.8}$$

$$D(\theta,\ \theta') = \int_{-\infty}^{\infty} S'(x;\ \theta')\, dF(x;\ \theta), \tag{12.1.9}$$

где θ — любая точка из Ω_0, а (θ, θ') — любая точка из прямого произведения $\Omega_0 \times \Omega_0$.

Рассмотрим сначала $S(x; \theta)$ и $S'(x; \theta)$. Предположим, что $F(x; \theta)$ имеет первую производную по θ в каждой точке θ из Ω_0 для всех x, за исключением, может быть, множества, имеющего вероятность нуль; очевидно, что $S(x; \theta)$ определяется как

$$S(x'; \theta) = \lim_{x \to x'} \frac{\frac{\partial}{\partial \theta}[F(x'; \theta) - F(x; \theta)]}{[F(x'; \theta) - F(x; \theta)]}, \tag{12.1.10}$$

где $x < x'$, при условии, что написанный предел существует. Мы будем предполагать, что предел существует для всех θ из Ω_0 и всех x из R_1 и что он не равен нулю на множестве значений x, имеющем положительную вероятность. $S'(x; \theta)$ определяется аналогичным образом. Если для $\theta = \theta'$ случайная величина x имеет к. ф. р. $F(x; \theta')$, то $S(x; \theta)$ и $S'(x; \theta)$ являются случайными величинами для $\theta \in \Omega_0$.

Заметим, что для дискретной случайной величины x с ф. в. $p(x; \theta)$

$$S(x; \theta) = \frac{\partial}{\partial \theta} \log p(x; \theta) \tag{12.1.10a}$$

и для непрерывной случайной переменной x с ф. п. в. $f(x; \theta)$

$$S(x; \theta) = \frac{\partial}{\partial \theta} \log f(x; \theta). \tag{12.1.10b}$$

Аналогичные утверждения о $S'(x; \theta)$ справедливы для дискретной и непрерывной случайных величин.

Рассмотрим теперь $H(\theta, \theta')$. Разобьем ось x на дизъюнктные интервалы $I_\alpha = (x_\alpha, x_{\alpha+1}]$, $\alpha = \ldots, -1, 0, 1, \ldots$ Тогда

$$P(x \in I_\alpha \mid \theta) = F(x_{\alpha+1}; \theta) - F(x_\alpha; \theta).$$

Аналогично выражается $P(x \in I_\alpha \mid \theta')$. Пусть $\Delta = \max\limits_\alpha \{I_\alpha\}$ и

$$H_\Delta(\theta, \theta') = \sum_{\alpha=-\infty}^{\infty} \log P(x \in I_\alpha \mid \theta') P(x \in I_\alpha \mid \theta) =$$

$$= \log \left\{ \prod_{\alpha=-\infty}^{\infty} [P(x \in I_\alpha \mid \theta')]^{P(x \in I_\alpha \mid \theta)} \right\}. \tag{12.1.11}$$

Ясно, что каждое слагаемое в верхней строчке (12.1.11) отрицательно и, следовательно, $H_\Delta(\theta, \theta')$ отрицательно. Чтобы избежать слагаемых, равных $-\infty$, достаточно предположить, что не существует такого множества E в пространстве x, для которого $P(x \in E \mid \theta)$ и $P(x \in E \mid \theta')$ не были бы одновременно равны нулю или положительны. Говорят, что два распределения $F(x; \theta)$ и $F(x; \theta')$, удовлетворяющих этому условию, *абсолютно непрерывны друг относительно друга.* Это допущение несколько сильнее, чем здесь требуется, так как было бы достаточно абсолютной непрерывности $F(x; \theta')$ относительно $F(x; \theta)$, т. е. допущения, что не существует множества E, для которого

$P(x \in E \mid \theta') = 0$, а $P(x \in E \mid \theta) > 0$, но это обобщение устраняется симметрией $F(x;\ \theta)$ и $F(x;\ \theta')$ в нашем случае.

Воспользовавшись тем фактом, что для любых двух наборов положительных чисел $p_1, \ldots, p_r$ и $q_1, \ldots, q_r$

$$p_1^{q_1} \ldots p_r^{q_r} \leqslant (p_1 + \ldots + p_r)^{q_1 + \ldots + q_r}, \qquad (12.1.12)$$

мы увидим, что положительная величина в фигурных скобках в (12.1.11) не может возрастать при последовательном дроблении интервалов $\{I_\alpha\}$. Следовательно, если существует некоторый набор $\{I_\alpha\}$ такой, что $H_\Delta(\theta, \theta')$ конечно, то конечная отрицательная величина $H_\Delta(\theta, \theta')$ не может убывать при $\Delta \to 0$ и, следовательно, $\lim\limits_{\Delta \to 0} H_\Delta(\theta, \theta')$ существует для всех точек (θ, θ') из $\Omega_0 \times \Omega_0$ и неположителен. Как можно показать, этот предел не зависит от способа разбиения оси x при условии, конечно, что $\Delta \to 0$.

Функция $H(\theta, \theta')$ играет большую роль в связи с теорией информации, энтропией в статистической механике, оптимальным оцениванием и оптимальными статистическими тестами в статистических выводах.

Если для данной функции $g(x;\ \theta')$, измеримой относительно $F(x;\ \theta)$, существует неотрицательная функция $h(x)$, измеримая и имеющая конечное среднее значение относительно $F(x;\ \theta)$ и такая, что

$$|g(x;\ \theta')| < h(x)$$

для $(\theta, \theta') \in \Omega_0 \times \Omega_0$, мы будем говорить, что $g(x;\ \theta')$ *мажорируется* интегрируемой функцией $h(x)$.

Чтобы (12.1.2) было справедливо, достаточно, чтобы $\frac{\partial}{\partial \theta'} \log dF(x;\ \theta')$ мажорировалась некоторой интегрируемой функцией $h_1(x)$. Аналогично (12.1.3) справедливо, если $\frac{\partial^2}{\partial \theta'^2} \log dF(x;\ \theta')$ и $\left[\frac{\partial}{\partial \theta'} \log dF(x;\ \theta')\right]^2$ мажорируются интегрируемыми функциями $h_2(x)$ и $h_3(x)$. По общим теоремам о достаточности условий такого рода отсылаем читателя к книгам Макшейна (1944), Макшейна и Ботса (1959) и Сакса (1937) по теории интегрирования и действительной переменной.

Выражая (12.1.2) и (12.1.3) в терминах средних значений, мы будем говорить, что $F(x;\ \theta)$ *регулярна в смысле ее первой производной по* θ *в* Ω_0, *если*

$$\mathscr{E}(S(x;\ \theta)) = \frac{\partial}{\partial \theta} \int\limits_{-\infty}^{\infty} dF(x;\ \theta) = 0, \qquad (12.1.2a)$$

и регулярна в смысле второй производной по θ *в* Ω_0, *если*

$$\mathscr{E}(S'(x;\ \theta)) + \mathscr{E}(S(x;\ \theta))^2 = \frac{\partial^2}{\partial \theta^2} \int\limits_{-\infty}^{\infty} dF(x;\ \theta) = 0. \qquad (12.1.3a)$$

Возвратимся теперь к $H(\theta, \theta')$, определенной по (12.1.6). Если $F(x;\ \theta)$ регулярна относительно ее первых двух производных, можно показать, что $H(\theta, \theta')$ имеет первые частные производные по θ и θ' в $\Omega_0 \times \Omega_0$. Более того, $H(\theta, \theta')$ как функция θ' при фиксированном θ имеет максимум при $\theta = \theta'$.

(b) Случай векторного параметра. Пусть $F(x; \theta)$ — к. ф. р. и θ — r-мерный параметр с (функционально независимыми) компонентами $(\theta_1, \ldots, \theta_r)$ из параметрического пространства Ω_r. Для удобства мы оставим x одномерным; тривиальные изменения обозначений показывают, что результаты остаются справедливыми, если x — k-мерная величина. Достаточные условия, при которых мы можем дифференцировать

$$\int_{-\infty}^{\infty} dF(x; \theta)$$

под интегралом один или больше раз по одной или более компонент θ, могут быть выведены без труда. Рассматривая только две типичные производные, мы получаем аналогично равенствам (12.1.2) и (12.1.3)

$$\frac{\partial}{\partial\theta_p} \int_{-\infty}^{\infty} dF(x; \theta) = \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial\theta_p} \log dF(x; \theta)\right] dF(x; \theta) = 0, \tag{12.1.13}$$

$$\frac{\partial^2}{\partial\theta_p\,\partial\theta_q} \int_{-\infty}^{\infty} dF(x; \theta) = \int_{-\infty}^{\infty} \left[\frac{\partial^2}{\partial\theta_p\,\partial\theta_q} \log dF(x; \theta)\right] dF(x; \theta) +$$
$$+ \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial\theta_p} \log dF(x; \theta)\right]\left[\frac{\partial}{\partial\theta_q} \log dF(x; \theta)\right] dF(x; \theta) = 0. \tag{12.1.14}$$

Мы заинтересованы в том, чтобы тождества (12.1.13) и (12.1.14) имели место для $p, q = 1, \ldots, r$ и всех точек r-мерного (эвклидова) открытого интервала Ω_{r0}.

Для удобства изложения и дальнейших ссылок выпишем выражения, аналогичные (12.1.4)—(12.1.9):

$$S_p(x; 0) = \frac{\partial}{\partial\theta_p} \log dF(x; \theta), \tag{12.1.15}$$

$$S_{pq}(x; \theta) = \frac{\partial^2}{\partial\theta_p\,\partial\theta_q} \log dF(x; \theta), \tag{12.1.16}$$

$$H(\theta, \theta') = \int_{-\infty}^{\infty} \log dF(x; \theta')\, dF(x; \theta), \tag{12.1.17}$$

$$A(\theta, \theta') = \int_{-\infty}^{\infty} S_p(x; \theta')\, dF(x; \theta), \tag{12.1.18}$$

$$B_{pq}(\theta, \theta') = \int_{-\infty}^{\infty} S_p(x; \theta')\, S_q(x; \theta')\, dF(x; \theta), \tag{12.1.19}$$

$$D_{pq}(\theta, \theta') = \int_{-\infty}^{\infty} S_{pq}(x; \theta')\, dF(x; \theta), \tag{12.1.20}$$

где $p, q = 1, \ldots, r$, θ — точка из Ω_{r0}. Так как компоненты θ по предположению функционально независимы, то компоненты случайной переменной $(S_p(x;\ \theta)$, $p = 1, \ldots, r)$ линейно независимы и, следовательно, матрица $\|B_{pq}(\theta,\ \theta')\|$ оказывается положительно определенной для $(\theta,\ \theta')$ из $\Omega_{r0} \times \Omega_{r0}$. В свете обсуждения случая одномерного параметра эти функции не требуют дополнительных пояснений.

Из одномерного случая очевидно, что для справедливости (12.1.13) достаточно, чтобы $\frac{\partial}{\partial\theta'_p} \log dF(x;\ \theta')$ мажорировались интегрируемыми функциями $h_{1p}(x)$, $p = 1, \ldots, r$. Аналогично для справедливости (12.1.14) достаточно, чтобы

$$\frac{\partial^2}{\partial\theta'_p\, \partial\theta'_q} \log dF(x;\ \theta')$$

и

$$\left[\frac{\partial}{\partial\theta'_p} \log dF(x;\ \theta')\right]\left[\frac{\partial}{\partial\theta'_q} \log dF(x;\ \theta')\right]$$

соответственно мажорировались интегрируемыми функциями $h_{2pq}(x)$ и $h_{3pq}(x)$, $p, q = 1, \ldots, r$.

Выражая (12.1.13) и (12.1.14) в терминах средних значений, мы будем говорить, что $F(x;\ \theta)$ *регулярна в смысле ее первых частных производных по* θ *в* Ω_{r0}, *если*

$$\mathscr{E}(S_p(x;\ \theta)) = \frac{\partial}{\partial\theta_p} \int_{-\infty}^{\infty} dF(x;\ \theta) = 0, \tag{12.1.13a}$$

$p = 1, \ldots, r$, *и регулярна в смысле ее вторых частных производных по* θ, *если имеют место следующие равенства*:

$$\mathscr{E}(S_p(x;\ \theta)\, S_q(x;\ \theta)) + \mathscr{E}(S_{pq}(x;\ \theta)) = \frac{\partial^2}{\partial\theta_p\, \partial\theta_q} \int_{-\infty}^{\infty} dF(x; \theta) = 0. \tag{12.1.14a}$$

(c) Замечания, касающиеся обобщения на векторные случайные величины. Предыдущее обсуждение относилось к случаю одномерной случайной величины. С небольшими изменениями оно может быть обобщено на векторный случай. Основные изменения касаются определений $S(x;\ \theta)$ и $H(\theta,\ \theta')$.

Пусть x — векторная величина $(x_1, \ldots, x_k)$. В этом случае простая разность $[F(x';\ \theta) - F(x;\ \theta)]$ на интервале $(x,\ x']$, входящая в (12.1.10), заменяется разностью функции $F(x_1, \ldots, x_k;\ \theta)$ на k-мерном интервале $(x_1, \ldots, x_k;\ x'_1, \ldots, x'_k]$. Предел, определяющий $\frac{\partial}{\partial\theta} \log dF(x_1, \ldots, x_k;\ \theta)$, по предположению существует и обозначается $S(x_1, \ldots, x_k;\ \theta)$.

Аналогично разность $F(x_{\alpha+1};\ \theta) - F(x_\alpha;\ \theta)$ функции $F(x;\ \theta)$ на интервале $(x_\alpha;\ x_{\alpha+1}]$, появляющаяся в (12.1.11), заменяется k-й разностью функции $F(x_1, \ldots, x_k; \theta)$ на k-мерном интервале $(x_{1,\alpha}, \ldots, x_{k,\alpha}; x_{1,\alpha+1}, \ldots, x_{k,\alpha+1}]$, причем $\Delta = \max_{\alpha} \{\Delta_\alpha\}$, где Δ_α — наибольшая составляющая k-мерного интервала.

Остальные изменения в обозначениях очевидны.

12.2. Точечное оценивание

(а) Определения. Пусть $(x_1, \ldots, x_n)$ — n-мерная случайная величина с к. ф. р. $F_n(x_1, \ldots, x_n; \theta)$, где θ — одномерный действительный параметр с параметрическим пространством Ω.

Пусть $\tilde{\theta}(x_1, \ldots, x_n)$, или кратко $\tilde{\theta}$, — функция от $(x_1, \ldots, x_n)$ и, таким образом, является случайной величиной. Если полученное (наблюденное) значение $\tilde{\theta}$, соответствующее полученному (наблюденному) значению $(x_1, \ldots, x_n)$, используется вместо θ_0, *истинного* значения θ, то случайная переменная $\tilde{\theta}$ называется *точечной оценкой* или просто оценкой для θ_0. Так используют $\tilde{\theta}$, конечно, только в том случае, когда величина θ_0 неизвестна. Если при $\theta = \theta_0$ $\mathscr{E}(\tilde{\theta}) = \theta_0$, что мы будем короче записывать в виде $\mathscr{E}(\tilde{\theta} \mid \theta_0) = \theta_0$, то $\tilde{\theta}$ называется *несмещенной оценкой* для θ_0. Было бы точнее сказать, что $\tilde{\theta}$ — *оценка для* θ_0, *несмещенная в среднем.* Если бы $\tilde{\theta}$ была статистикой с к. ф. р. $W(\tilde{\theta}; \theta_0)$, непрерывной при $\tilde{\theta} = \theta_0$ и такой, что $W(\theta_0; \theta_0) = \frac{1}{2}$, мы сказали бы, что $\tilde{\theta}$ — *оценка для* θ_0, *несмещенная по медиане.* Ниже, если не оговорено противное, несмещенная оценка будет пониматься как несмещенная в среднем.

Если оценка $\tilde{\theta}$ сходится по вероятности к θ_0 при $n \to \infty$, она называется *состоятельной* оценкой для θ_0.

Если $\tilde{\theta}$ — несмещенная оценка для θ_0, имеющая конечную дисперсию, и при этом не существует никакой другой несмещенной оценки с меньшей дисперсией, то $\tilde{\theta}$ называется *эффективной оценкой* для θ_0.

Если $\tilde{\theta}$ — такая статистика, что для любой другой статистики $\breve{\theta}$ условная случайная величина $\breve{\theta} \mid \tilde{\theta}$ имеет распределение, не зависящее от θ_0, то $\tilde{\theta}$ называется *достаточной оценкой* для θ_0.

Понятия состоятельности, эффективности и достаточности введены Фишером (1922, 1925b).

Для простоты случайного выбора, т. е. когда $(x_1, \ldots, x_n)$ — случайная выборка объема n из совокупности с к. ф. р. $F(x; \theta)$, понятия несмещенности, состоятельности, достаточности и эффективности имеют особое значение. В этом случае

$$F_n(x_1, \ldots, x_n; \theta) = \prod_{\xi=1}^{n} F(x_\xi; \theta). \qquad (12.2.1)$$

Для данной выборки $(x_1, \ldots, x_n)$ величина $dF_n = \prod_{\xi=1}^{n} dF(x_\xi; \theta)$ называется *элементом правдоподобия* θ для $(x_1, \ldots, x_n)$.

Бо́льшая часть материала этой главы относится к простому случайному выбору.

(b) Нижняя граница дисперсии оценки. Пусть $(x_1, \ldots, x_n)$ — n-мерная случайная переменная с к. ф. р. $F_n(x_1, \ldots, x_n; \theta)$. Если $\tilde{\theta}(x_1, \ldots, x_n)$ — несмещенная оценка для θ, то

$$\int_{R_n} (\tilde{\theta} - \theta)\, dF_n = 0. \tag{12.2.2}$$

Если $F_n(x_1, \ldots, x_n)$ регулярна относительно ее первой производной по θ в некотором (открытом) интервале Ω_0, содержащем θ_0 — истинное значение θ, — то мы можем дифференцировать (12.2.2) под интегралом и получить

$$\int_{R_n} (\tilde{\theta} - \theta)\, S_n(x_1, \ldots, x_n; \theta)\, dF_n = 1, \tag{12.2.3}$$

где

$$S_n(x_1, \ldots, x_n; \theta) = \frac{\partial \log dF_n}{\partial \theta}. \tag{12.2.4}$$

Применяя неравенство Шварца к (12.2.3), получаем

$$1 = \left[\int_{R_n} (\tilde{\theta} - \theta)\, S_n\, dF_n\right]^2 \leqslant \int_{R_n} (\tilde{\theta} - \theta)^2\, dF_n \int_{R_n} S_n^2\, dF_n.$$

Так как

$$\sigma^2(\tilde{\theta}) = \int_{R_n} (\tilde{\theta} - \theta)^2\, dF_n, \tag{12.2.5}$$

находим

$$\sigma^2(\tilde{\theta}) \geqslant \frac{1}{\mathscr{E}(S_n^2)}, \tag{12.2.6}$$

причем равенство имеет место в том и только том случае, когда

$$K[\tilde{\theta}(x_1, \ldots, x_n) - \theta] \equiv S_n(x_1, \ldots, x_n; \theta) \tag{12.2.7}$$

в R_n с вероятностью 1, где K может зависеть от θ, но не зависит от $(x_1, \ldots, x_n)$.

Если равенство имеет место, то $\tilde{\theta}$ — *эффективная оценка* для θ; в этом случае (12.2.7) дает форму, которую имеет $\tilde{\theta}$.

Если обозначить $\sigma^2(\tilde{\theta})$ при $\theta = \theta_0$ через $\sigma^2(\tilde{\theta} \mid \theta_0)$, положив $\theta = \theta_0$ в (12.2.5), и взять соответствующее выражение для $\mathscr{E}(S_n^2 \mid \theta_0)$, то

$$\sigma^2(\tilde{\theta} \mid \theta_0) \geqslant \frac{1}{\mathscr{E}(S_n^2 \mid \theta_0)}.$$

Эффективная оценка для θ обычно обозначается $\hat{\theta}$.

Таким образом, мы получили следующий результат:

12.2.1. *Пусть* $(x_1, \ldots, x_n)$ — *случайная величина с к. ф. р.* $F_n(x_1, \ldots, x_n; \theta)$, *регулярной относительно ее первой производной по* θ *в* Ω_0, *где* θ — *одномерный параметр. Если* $\tilde{\theta}(x_1, \ldots, x_n)$ — *несмещенная оценка для* θ_0, *то* $\sigma^2(\tilde{\theta} \mid \theta_0) \geqslant \frac{1}{\mathscr{E}(S_n^2 \mid \theta_0)}$, *где* S_n *дается выражением* (12.2.4), *причем равенство имеет место в том и только том случае, когда справедливо* (12.2.7) *с вероятностью* 1 *при* $\theta = \theta_0$. *Если эффективная оценка* $\hat{\theta}$ *для* θ_0 *существует, то ее дисперсия равна* $\frac{1}{\mathscr{E}(S_n^2 \mid \theta_0)}$.

Нижние границы для $\sigma^2(\tilde{\theta} \mid \theta_0)$ без предположения регулярности получили Чэпмен и Роббинс (1951) и Кифер (1952).

Если существует эффективная оценка $\hat{\theta}$ для θ_0 и если $\tilde{\theta}$ — любая другая несмещенная оценка для θ_0, то эффективность $\tilde{\theta}$ для оценивания θ_0 определяется отношением

$$\operatorname{eff}(\tilde{\theta} \mid \theta_0) = \frac{\sigma^2(\hat{\theta} \mid \theta_0)}{\sigma^2(\tilde{\theta} \mid \theta_0)}. \tag{12.2.8}$$

(c) Случайный выбор. В этом важном случае имеет место (12.2.1) и

$$S_n(x_1, \ldots, x_n; \theta) = \sum_{\xi=1}^{n} S(x_\xi; \theta), \tag{12.2.9}$$

где

$$S(x; \theta) = \frac{\partial}{\partial \theta} \log dF(x; \theta). \tag{12.2.10}$$

Сумма $\sum_{\xi=1}^{n} S(x_\xi; \theta)$ называется *информантом* для θ, основанным на $(x_1, \ldots, x_n)$.

Кроме того,

$$\mathscr{E}(S_n^2) = n\mathscr{E}(S^2) = nB^2(\theta, \theta), \tag{12.2.11}$$

где $B^2(\theta, \theta)$ определяется по (12.1.8). Поэтому (12.2.6) сводится к

$$\sigma^2(\tilde{\theta}) \geqslant \frac{1}{nB^2(\theta, \theta)}. \tag{12.2.12}$$

Этот результат первоначально был получен Фишером (1922). Позднее он был установлен Крамером (1946), Дюге (1937) и Рао (1945).

Равенство в (12.2.12) имеет место в том и только том случае, когда

$$K(\tilde{\theta} - \theta) \equiv \sum_{\xi=1}^{n} S(x_\xi; \theta) \tag{12.2.13}$$

на R_n с вероятностью 1, причем K не зависит от $(x_1, \ldots, x_n)$.

Таким образом, если существует эффективная оценка $\hat{\theta}$, то ею является статистика $\tilde{\theta}$, удовлетворяющая (12.2.13), и ее дисперсия дается равенством

$$\sigma^2(\hat{\theta}) = \frac{1}{nB^2(\theta, \theta)}. \tag{12.2.14}$$

Итак, мы имеем следующее важное следствие **12.2.1**:

12.2.1a. *Если* $(x_1, \ldots, x_n)$ — *выборка из совокупности с к. ф. р.* $F(x; \theta)$, *регулярной относительно ее первой производной по* θ *в* Ω_0, *и если* $\tilde{\theta}$ — *несмещенная оценка для* θ_0, *то* $\sigma^2(\tilde{\theta} \mid \theta_0) \geqslant \frac{1}{nB^2(\theta_0, \theta_0)}$, *где* $B^2(\theta, \theta)$ *дается формулой* (12.1.8). *Равенство имеет место в том и только том случае, когда* $\tilde{\theta}$ *удовлетворяет* (12.2.13) *и является эффективной оценкой для* θ_0 *с дисперсией* $\frac{1}{nB^2(\theta_0, \theta_0)}$.

Если $\tilde{\theta}$ — произвольная несмещенная оценка для θ и если существует эффективная оценка $\hat{\theta}$, то эффективность $\tilde{\theta}$ при оценивании θ_0 определяется равенством

$$\operatorname{eff}(\tilde{\theta} \mid \theta_0) = \frac{\sigma^2(\hat{\theta} \mid \theta_0)}{\sigma^2(\tilde{\theta} \mid \theta_0)} = \frac{1}{\sigma^2(\sqrt{n}\tilde{\theta} \mid \theta_0)\, B^2(\theta_0, \theta_0)}. \tag{12.2.15}$$

Фишер (1922) назвал величину $nB^2(\theta_0, \theta_0)$, обратную дисперсии эффективной оценки $\hat{\theta}$, *количеством информации*, содержащимся в выборке относительно θ_0, а $B^2(\theta_0, \theta_0)$ — количеством информации о θ_0 на одно наблюдение.

С этой точки зрения эффективность оценки $\tilde{\theta}$ для θ можно рассматривать как отношение информации, содержащейся в $\tilde{\theta}$ при оценивании θ_0, к информации, содержащейся в эффективной оценке.

Пример. Пусть $(x_1, \ldots, x_n)$ — выборка из распределения Пуассона $\mathrm{Po}(\mu_0)$, определенного в **8.3.3с.** Мы имеем

$$S_n(x_1, \ldots, x_n; \mu) = \frac{\partial}{\partial \mu} \log \prod_{\xi=1}^{n} dF(x_\xi; \mu) = \frac{n}{\mu}\left[\frac{1}{n}\sum_{\xi=1}^{n} x_\xi - \mu\right],$$

что в соответствии с (12.2.13) показывает, что $\frac{1}{n}\sum_{\xi=1}^{n} x_\xi$, т. е. выборочное среднее, — эффективная оценка для μ_0, истинного значения μ, для любого заданного объема выборки. Обозначив $\frac{1}{n}\sum_{\xi=1}^{n} x_\xi$ через $\hat{\mu}(x_1, \ldots, x_n)$ или $\hat{\mu}$, мы найдем дисперсию $\hat{\mu}$, применяя (12.2.14), т. е.

$$\sigma^2(\hat{\mu} \mid \mu_0) = \frac{1}{n \sum\limits_{x=0}^{\infty} \left(\frac{x}{\mu_0} - 1\right)^2 \frac{\mu_0^x e^{-\mu_0}}{x!}} = \frac{\mu_0}{n},$$

откуда

$$\sigma^2(\sqrt{n}\hat{\mu} \mid \mu_0) = \mu_0,$$

и из **12.2.1a** следует, что невозможно найти несмещенную оценку $\tilde{\mu}(x_1, \ldots, x_n)$ для μ_0, для которой $\sigma^2(\sqrt{n}\tilde{\mu} \mid \mu_0) < \mu_0$.

(d) Нижняя граница дисперсии смещенной оценки. Крамер (1946) и Дюге (1937) обобщили **12.2.1**, рассмотрев *смещенную* оценку $\tilde{\theta}(x_1, \ldots, x_n)$ для θ. В этом случае мы замещаем $(\tilde{\theta} - \theta)$ в (12.2.2) выражением

$$[\tilde{\theta} - \theta - b_n(\theta)], \tag{12.2.16}$$

где $b_n(\theta)$ — *смещение* оценки $\tilde{\theta}$, и повторяем рассуждения, приведшие к **12.2.1a.** В (12.2.3) мы заменяем $(\tilde{\theta} - \theta)$ на $(\tilde{\theta} - \theta - b_n(\theta))$ и 1 в правой части на $1 + b'_n(\theta)$, где $b'_n(\theta) = \frac{d}{d\theta} b_n(\theta)$. Вместо (12.2.12) получаем

$$\sigma^2(\tilde{\theta}) \geqslant \frac{[1 + b'_n(\theta)]^2}{nB^2(\theta, \theta)}, \tag{12.2.17}$$

причем равенство имеет место в том и только том случае, когда

$$K[\tilde{\theta} - \theta - b_n(\theta)] \equiv \sum_{\xi=1}^{n} S(x_\xi; \theta) \tag{12.2.18}$$

на R_n с вероятностью 1, причем K не зависит от $(x_1, \ldots, x_n)$.

(e) Свойство достаточных оценок. Пусть $(x_1, \ldots, x_n)$ — случайная переменная с ф. п. в. $f_n(x_1, \ldots, x_n; \theta)$, представимой в виде

$$f_n(x_1, \ldots, x_n; \theta) = v(\tilde{\theta}; \theta)\, u(x_1, \ldots, x_n \mid \tilde{\theta}), \tag{12.2.19}$$

где $v(\tilde{\theta}; \theta)$ — ф. п. в. $\tilde{\theta}$ и $u(x_1, \ldots, x_n \mid \tilde{\theta})$ — ф. п. в. условной случайной переменной $(x_1, \ldots, x_n \mid \tilde{\theta})$, не зависящая от θ. Тогда $\tilde{\theta}$ — достаточная статистика для θ, так как если $\breve{\theta}$ — любая другая статистика, не зависящая от $\tilde{\theta}$, то распределение условной случайной переменной $\breve{\theta} \mid \tilde{\theta}$ полностью определяется функцией $u(x_1, \ldots, x_n \mid \tilde{\theta})$.

Обратно, пусть $\tilde{\theta}$ — достаточная статистика с ф. п. в. $v(\tilde{\theta}; \theta)$. Пусть $y_2, \ldots, y_n$ — любые другие $(n-1)$ статистики такие, что $\tilde{\theta}, y_2, \ldots, y_n$ имеют ф. п. в.

$$g(\tilde{\theta}, y_2, \ldots, y_n; \theta) = f_n(x_1, \ldots, x_n; \theta) \frac{1}{|J|}, \tag{12.2.20}$$

где J — якобиан $(\tilde{\theta}, y_2, \ldots, y_n)$ относительно $(x_1, \ldots, x_n)$. [Условия, при которых имеет место (12.2.20), см. в § 2.8 (d).]

Пусть $h(y_2, \ldots, y_n \mid \tilde{\theta}; \theta)$ — ф. п. в. условной случайной переменной $y_2, \ldots, y_n \mid \tilde{\theta}$. Тогда

$$h(y_2, \ldots, y_n \mid \tilde{\theta}; \theta) = \frac{g(\tilde{\theta}, y_2, \ldots, y_n; \theta)}{v(\tilde{\theta}; \theta)}, \tag{12.2.21}$$

где $v(\tilde{\theta};\ \theta)\neq 0$ определяется равенством

$$v(\tilde{\theta};\ \theta)=\int\limits_{R_n} g(\tilde{\theta},\ y_2,\ \ldots,\ y_n;\ \theta)\,dy_2\ \ldots\ dy_n. \qquad (12.2.22)$$

Если $h(y_2, \ldots, y_n \mid \tilde{\theta};\ \theta)$ не зависит от θ (обозначим тогда эту функцию $h^*(y_2, \ldots, y_n \mid \tilde{\theta})$, то $\tilde{\theta}$ — достаточная оценка для θ, и мы имеем

$$g(\tilde{\theta},\ y_2,\ \ldots,\ y_n;\ \theta)=v(\tilde{\theta};\ \theta)\,h^*(y_2,\ \ldots,\ y_n \mid \tilde{\theta}). \qquad (12.2.23)$$

Используя (12.2.20), получаем

$$f_n(x_1,\ \ldots,\ x_n;\ \theta)=v(\tilde{\theta};\ \theta)\,h^*(y_2,\ \ldots,\ y_n \mid \tilde{\theta})\,|J|. \qquad (12.2.24)$$

Заметим, что $h^*(y_2, \ldots, y_n \mid \tilde{\theta})$ зависит от $(\tilde{\theta}, y_2, \ldots, y_n)$ и, следовательно, от $(x_1, \ldots, x_n)$, но не зависит от θ. Таким образом, всякая статистика, зависящая от $(x_1, \ldots, x_n)$ через $(y_2, \ldots, y_n)$, но не зависящая от θ, имеет распределение, не зависящее от θ.

Результат можно сформулировать следующим образом.

12.2.2. *Пусть $(x_1, \ldots, x_n)$ — случайная переменная, имеющая ф. п. в. $f_n(x_1, \ldots, x_n; \theta)$. Для того чтобы статистика $\tilde{\theta}(x_1, \ldots, x_n)$ была достаточной для θ, необходимо и достаточно, чтобы*

$$f_n(x_1,\ \ldots,\ x_n;\ \theta)=v(\tilde{\theta};\ \theta)\,w(x_1,\ \ldots,\ x_n), \qquad (12.2.25)$$

где $v(\theta;\ \theta)$ — ф. п. в. $\tilde{\theta}$ и $w(x_1, \ldots, x_n)$ — функция $(x_1, \ldots, x_n)$, не зависящая от θ.

Фишер (1922) первый отметил достаточность *критерия факторизации* (12.2.25). Нейман (1935) показал, что он является также необходимым.

В частном важном случае простого случайного выбора справедливо следствие **12.2.2**, в котором $f_n(x_1, \ldots, x_n; \theta)$ принимает форму $\prod\limits_{\xi=1}^{n} f(x_\xi;\ \theta)$.

Аналогично можно показать, что если $(x_1, \ldots, x_n)$ — n-мерная дискретная случайная переменная с ф. в. $p_n(x_1, \ldots, x_n; \theta)$, то необходимым и достаточным условием для того, чтобы статистика $\tilde{\theta}$ была достаточной для θ, является представление $p_n(x_1, \ldots, x_n; \theta)$ в виде

$$p_n(x_1,\ \ldots,\ x_n;\ \theta)=v(\tilde{\theta};\ \theta)\,w(x_1,\ \ldots,\ x_n), \qquad (12.2.26)$$

где $v(\tilde{\theta};\ \theta)$ — ф. в. $\tilde{\theta}$, а $w(x_1, \ldots, x_n)$ зависит от $(x_1, \ldots, x_n)$, но не зависит от θ.

Для простого случайного выбора из совокупности с ф. в. $p(x;\ \theta)$, $p_n(x_1, \ldots, x_n;\ \theta)$ принимает форму $\prod\limits_{\xi=1}^{n} p(x_\xi;\ \theta)$.

Наконец, заметим, что **12.2.2** обобщается на случай, когда θ — вектор $(\theta_1, \ldots, \theta_r)$, $r \leqslant n$. Необходимое и достаточное условие для того, чтобы $(\tilde{\theta}_1, \ldots, \tilde{\theta}_r)$ был *вектором достаточных статистик* для $(\theta_1, \ldots, \theta_r)$, состоит в представлении

$$f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_r) = v(\tilde{\theta}_1, \ldots, \tilde{\theta}_r; \theta_1, \ldots, \theta_r)\, w(x_1, \ldots, x_n), \tag{12.2.27}$$

где $w(x_1, \ldots, x_n)$ зависит от $(x_1, \ldots, x_n)$, но не зависит от θ.

Аналогичный критерий факторизации имеет место, конечно, и в случае, когда θ — вектор $(\theta_1, \ldots, \theta_r)$, а $(x_1, \ldots, x_n)$ — дискретная случайная переменная с ф. в. $p(x_1, \ldots, x_n; \theta_1, \ldots, \theta_r)$.

Для случайной переменной $(x_1, \ldots, x_n)$, имеющей распределение общего вида, обобщение критерия факторизации, сформулированного в **12.2.2**, и его распространение на случай векторного параметра сделали Халмош и Сэвидж (1949), применив теорему Радона — Никодима. Еще более абстрактное изложение необходимых и достаточных условий характеризации достаточных статистик дал Бахадур (1954).

Пример. Пусть $(x_1, \ldots, x_n)$ — выборка из распределения Пуассона $\mathrm{Po}(\mu_0)$. Для произвольного μ из $(0, \infty)$ имеем

$$dF(x; \mu) = \frac{\mu^x e^{-\mu}}{x!}$$

и

$$\prod_{\xi=1}^{n} dF(x_\xi; \mu) = \frac{\mu^{\sum_\xi x_\xi} e^{-n\mu}}{\prod_\xi x_\xi!},$$

что можно записать в виде

$$\left[\frac{\mu^{\sum_\xi x_\xi} e^{-n\mu}}{(\sum_\xi x_\xi)!}\right]\left\{\frac{(\sum_\xi x_\xi)!}{\prod_\xi x_\xi!}\right\}.$$

Обозначив $\frac{1}{n}\sum_\xi x_\xi$ через $\tilde{\mu}(x_1, \ldots, x_n)$, замечаем по **8.3.3с**, что выражение в квадратных скобках — ф. в. $\tilde{\mu}$, т. е.

$$dV(\tilde{\mu}; \mu) = \frac{\mu^{n\tilde{\mu}} e^{-n\mu}}{(n\tilde{\mu})!},$$

тогда как выражение в фигурных скобках фактически ф. в. условной случайной переменной $(x_1, \ldots, x_n \mid \tilde{\mu})$, т. е.

$$dU(x_1, \ldots, x_n \mid \tilde{\mu}) = \frac{(n\tilde{\mu})!}{\prod_\xi x_\xi!}.$$

Очевидно, если $\tilde{\mu}^*$ — любая другая оценка, то ф. в. $p(\tilde{\mu}^* \mid \tilde{\mu})$ условной случайной переменной $\tilde{\mu}^* \mid \mu$ получается суммированием $\frac{(n\tilde{\mu})!}{\prod_\xi x_\xi!}$ по положительным и нулевым значениям x_ξ, связанным двумя условиями $\tilde{\mu}^*(x_1, \ldots, x_n) = \tilde{\mu}^*$ и $\tilde{\mu}(x_1, \ldots, x_n) = \tilde{\mu}$. Ясно, что $p(\tilde{\mu}^* \mid \tilde{\mu})$ не зависит от μ.

Следовательно, $\frac{1}{n}\sum\limits_{\xi} x_\xi$ — достаточная оценка для μ_0.

Важным свойством достаточной оценки является следующее: если имеется исходная несмещенная оценка для параметра, не являющаяся функцией достаточной оценки, то можно найти несмещенную оценку, зависящую от достаточной и имеющую меньшую дисперсию, чем исходная. Точнее это утверждение формулируется в следующей теореме Блекуэла (1947) и Рао (1949).

12.2.3. *Пусть $\tilde{\theta}$ — достаточная оценка для θ_0 и $\breve{\theta}$ — любая несмещенная оценка для θ_0. Пусть $\mathscr{E}(\breve{\theta} \mid \tilde{\theta}) = h(\tilde{\theta})$. Тогда $h(\tilde{\theta})$ — несмещенная оценка для θ_0, дисперсия которой не превышает дисперсию $\breve{\theta}$.*

Докажем **12.2.3.** Пусть $G(\tilde{\theta}, \breve{\theta}; \theta_0)$ — к. ф. р. $(\tilde{\theta}, \breve{\theta})$, $V(\tilde{\theta}; \theta_0)$ — к. ф. р. $\tilde{\theta}$ и $U(\breve{\theta} \mid \tilde{\theta})$ — к. ф. р. условной случайной переменной $\breve{\theta} \mid \tilde{\theta}$. Тогда

$$h(\tilde{\theta}) = \int_{-\infty}^{\infty} \breve{\theta}\, dU(\breve{\theta} \mid \tilde{\theta}), \tag{12.2.28}$$

откуда видно, что $h(\tilde{\theta})$ — несмещенная оценка для θ_0.

Для дисперсии $\breve{\theta}$ имеем

$$\sigma^2(\breve{\theta}) = \mathscr{E}(\breve{\theta} - \theta_0)^2 = \mathscr{E}[(h(\tilde{\theta}) - \theta_0) + (\breve{\theta} - h(\tilde{\theta}))]^2 =$$
$$= \sigma^2(h(\tilde{\theta})) + \mathscr{E}(\breve{\theta} - h(\tilde{\theta}))^2 + 2\mathscr{E}[(h(\tilde{\theta}) - \theta_0)(\breve{\theta} - h(\tilde{\theta}))].$$

Но

$$\mathscr{E}[(h(\tilde{\theta}) - \theta_0)(\breve{\theta} - h(\tilde{\theta}))] =$$
$$= \int_{-\infty}^{\infty} [h(\tilde{\theta}) - \theta_0]\left\{\int_{-\infty}^{\infty} (\breve{\theta} - h(\tilde{\theta}))\, dU(\breve{\theta} \mid \tilde{\theta})\right\} dV(\tilde{\theta}; \theta_0) = 0,$$

так как выражение в фигурных скобках равно нулю. Ввиду неравенства $\mathscr{E}(\breve{\theta} - h(\tilde{\theta}))^2 \geqslant 0$ имеем

$$\sigma^2(\breve{\theta}) \geqslant \sigma^2(h(\tilde{\theta})), \tag{12.2.29}$$

что и завершает доказательство **12.2.3.**

12.3. Точечное оценивание при больших выборках

(а) Асимптотическое распределение информанта. Эффективная оценка параметра существует только в некоторых частных случаях к. ф. р. $F(x, \theta)$. При этом величина $S_n(x_1, x_2, \ldots, x_n; \theta)$, представленная (12.2.8), принимает специальный вид (12.2.13), и поэтому, по существу, эффективная оценка $\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ является решением уравнения

$$S_n(x_1, \ldots, x_n; \theta) = 0 \tag{12.3.1}$$

относительно θ.

Предположим же теперь, что $S_n(x_1, \ldots, x_n; \theta)$ не имеет специального вида (12.2.13), но уравнение (12.3.1) относительно θ имеет решение, которое мы, не вызывая путаницы, можем также назвать $\hat{\theta}_n$.

Какими свойствами обладает это решение, рассматриваемое как оценка параметра θ_0? Ответ на этот вопрос состоит в том, что при определенных условиях, которые мы оговорим специально, $\hat{\theta}_n(x_1, \ldots, x_n)$ является эффективной оценкой θ_0 в асимптотическом смысле для больших выборок.

Прежде всего сформулируем следующий результат.

12.3.1. *Пусть* $(x_1, \ldots, x_n)$ — *выборка из совокупности с к. ф. р.* $F(x; \theta_0)$, *где* $F(x; \theta)$ *регулярна с первой производной по* θ *в области* Ω_0. *Тогда, если величина* $B^2(\theta; \theta)$, *определенная в* (12.1.8.), *существует и конечна, то* $S_n(x_1, \ldots, x_n; \theta)$ *асимптотически имеет распределение* $N(0, nB^2(\theta_0;\theta_0))$.

По существу, это утверждение является следствием **9.2.1,** ибо если обозначить через y случайную величину $S(x; \theta)$, то y будет иметь к. ф. р.

$$G(y) = \int_{E_y} dF(x, \theta),$$

где E_y — множество точек на оси x, для которых $S(x, \theta) \leqslant y$. Так как $F(x, \theta)$ регулярна в смысле первой производной по θ, очевидно, что $\mathscr{E}(y) = 0$. Таким образом, $S_n(x_1, \ldots, x_n; \theta)$ есть среднее арифметическое выборки объема n из совокупности со средним 0 и конечной дисперсией $B^2(\theta, \theta)$.

Применяя **9.2.1,** получаем **12.3.1.**

(b) Сходимость оценок максимального правдоподобия. Очевидно, что при условиях, указанных в **12.3.1,** случайная величина $\frac{1}{n} S_n(x_1, \ldots, x_n; \theta)$ сходится по вероятности к нулю. Это означает, что если мы положим

$$S_n(x_1, \ldots, x_n; \theta) = 0, \tag{12.3.2}$$

то получим последовательность наборов корней $\{\hat{\theta}(x_1, \ldots, x_n)\}$, $n = 1, 2, \ldots$ Каждый набор содержит по крайней мере один корень такой, что их последовательность сходится с вероятностью 1 к истинному значению параметра θ. Мы покажем, что это имеет место при некоторых условиях. Предположим, что $S(x, \theta)$ непрерывна как функция θ в области Ω_0 для всех значений x из R_1, исключая, может быть, множество вероятностной меры 0. Кроме того, предположим, что $F(x, \theta)$ регулярна с первой производной по θ в области Ω_0, откуда следует, что функция $A(\theta_0, \theta)$, определенная в (12.1.7), непрерывна и строго убывает по θ в некотором подынтервале Ω_0' интервала Ω_0, который содержит θ_0.

Обращаясь к (12.2.7.) и доказательству теоремы **12.3.1**, видим, что $\frac{1}{n} S_n(x_1, \ldots, x_n; \theta)$ есть среднее выборки объема n из совокупности с математическим ожиданием $A(\theta_0, \theta)$, если θ_0 — истинное значение параметра θ. Поэтому согласно теореме **4.6.1** $\frac{1}{n} S_n(x_1, \ldots, x_n; \theta)$ сходится почти наверное к $A(\theta_0, \theta)$. Не ограничивая общности, можем считать, что Ω_0' представляет собой интервал $(\theta_0 - \delta, \theta_0 + \delta)$, $\delta > 0$. Таким образом, $A(\theta_0, \theta)$ монотонно убывает по θ на этом интервале, и так как $A(\theta_0, \theta_0) = 0$, мы имеем $A(\theta_0, \theta_0 - \delta) > 0$ и $A(\theta_0, \theta_0 + \delta) < 0$. Поэтому существует $n(\delta, \varepsilon)$ такое, что с вероятностью, большей $1 - \varepsilon$ для $n > n(\delta, \varepsilon)$, если θ_0 — истинное значение параметра θ, имеют место следующие два неравенства:

$$\begin{aligned} &\frac{1}{n} S_n(x_1, \ldots, x_n; \theta) > 0, \quad \text{если } \theta = \theta_0 - \delta, \\ &\frac{1}{n} S_n(x_1, \ldots, x_n; \theta) < 0, \quad \text{если } \theta = \theta_0 + \delta, \end{aligned} \tag{12.3.3}$$

так как $S(x, \theta)$ непрерывна по θ на интервале $(\theta_0 - \delta, \theta_0 + \delta)$ для почти всех x из R_1; аналогичное утверждение имеет место для

$$\frac{1}{n} \sum_{\xi=1}^{n} S(x_\xi, \theta), \text{ т. е. для } \frac{1}{n} S_n(x_1, \ldots, x_n; \theta).$$

Поэтому, если θ_0 — истинное значение параметра θ, то

$$P\left(\frac{1}{n} S_n(x_1, \ldots, x_n; \theta) = 0 \text{ для некоторого } \theta \right.$$
$$\left. \text{из интервала } (\theta_0 \pm \delta) \text{ для всех } n > n(\delta, \varepsilon) \mid \theta_0 \right) > 1 - \varepsilon, \tag{12.3.4}$$

что эквивалентно существованию последовательности корней (12.3.2), сходящейся почти наверное к θ_0. В частности, если (12.3.2) имеет единственное решение $\theta(x_1, \ldots, x_n)$ для $n = n_0, n_0 + 1, \ldots$ при некотором целом n_0, то последовательность $\hat{\theta}(x_1, \ldots, x_n)$, $n = n_0, n_0 + 1, \ldots$ сходится почти наверное к θ_0.

Сформулируем полученный результат.

12.3.2. *Предположим, что* $(x_1, \ldots, x_n)$ — *выборка из совокупности с к. ф. р.* $F(x; \theta_0)$, *где* $F(x; \theta)$ *регулярна с первой производной по* θ *в области* Ω_0. *Пусть* $S(x; \theta)$ *непрерывна как функция* θ *для почти всех* x *из* R_1. *Тогда существует последовательность решений уравнения* (12.3.2), *которая сходится почти наверное к* θ_0. *В частности, если* (12.3.2) *имеет единственное решение* $\hat{\theta}(x_1, \ldots, x_n)$ *для* $n >$ *некоторого* n_0, *то последовательность* $\hat{\theta}(x_1, \ldots, x_n)$, $n = n_0, n_0 + 1, \ldots$, *сходится почти наверное к* θ_0.

В соответствии с тем, что $S_n(x_1, \ldots, x_n;\ \theta)$ есть производная по θ логарифма элемента правдоподобия, именно $\sum_{\xi=1}^{n} \log dF(x_\xi;\ \theta)$, и что $\hat{\theta}(x_1, \ldots, x_n)$ есть то значение θ, для которого $S_n(x_1, \ldots, x_n;\ \theta)$ равно 0, оценка $\hat{\theta}$, если она единственна и максимизирует элемент правдоподобия, называется оценкой максимального правдоподобия для θ_0. Этот термин ввел Фишер (1922). Вальд (1949 b) показал, что решение (12.3.2), которое доставляет элементу правдоподобия абсолютный максимум, является состоятельной оценкой θ_0. Дальнейшие детальные исследования состоятельности оценок максимального правдоподобия и связанных с этим задач принадлежат Баранкину и Гёрланду (1951), Хузурбазару (1945), Ле Каму (1956), Вальду (1948, 1949 b) и другим.

(с) Асимптотическое распределение оценок максимального правдоподобия. Требование регулярности функции $F(x;\ \theta)$ в смысле первой производной по θ является вполне достаточным, чтобы сделать следующее заключение об асимптотическом распределении $\hat{\theta}$ при больших n:

12.3.3. *Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности с к. ф. р. $F(x;\ \theta_0)$, где $F(x;\ \theta)$ регулярна в смысле первой и второй производных по θ в области Ω_0. Если оценка максимального правдоподобия единственна для $n > n_0$ и является (измеримой) случайной величиной относительно вероятностной меры $\prod_{\xi=1}^{n} F(x_\xi;\ \theta)$, как указано в* (12.2.1), *тогда $\hat{\theta}$ асимптотически нормальна* $N(\theta_0;\ 1/[nB^2(\theta_0, \theta_0)])$.

Поскольку $S(x;\ \theta)$ имеет производную по θ всюду в Ω_0 для почти всех x из R_1, аналогичным свойством обладает и $S(x_1, \ldots, x_n;\ \theta)$ для почти всех точек $(x_1, \ldots, x_n)$ из R_1.

Если оценка $\hat{\theta}$ единственна для n, бо́льших некоторого n_0, то из **12.3.2** следует, что $\hat{\theta}$ сходится почти наверное к θ_0 при $n \to \infty$. Более того, если $\hat{\theta}(x_1, \ldots, x_n)$ является случайной величиной, то для произвольных $\delta > 0$ и $\varepsilon > 0$ существует такое $n(\delta,\ \varepsilon,\ n_0)$ и такое множество E_n из R_n, определяемое условием $|\theta_0 - \hat{\theta}(x_1, \ldots, x_n)| < \delta$, что

$$P((x_1, \ldots, x_n) \in E_n \text{ для всех } n > n(\delta,\ \varepsilon,\ n_0) \mid \theta_0) > 1 - \varepsilon. \qquad (12.3.5)$$

Для точек множества E_n

$$\frac{1}{\sqrt{n}} S_n(x_1, \ldots, x_n;\ \theta_0) = \frac{1}{\sqrt{n}} S_n(x_1, \ldots, x_n;\ \hat{\theta}) + \\ + \frac{1}{\sqrt{n}} S_n'(x_1, \ldots, x_n;\ \theta^*)(\theta_0 - \hat{\theta}), \qquad (12.3.6)$$

где $\theta^*(x_1, \ldots, x_n)$ — случайная величина, удовлетворяющая условию

$$|\theta_0 - \theta^*| \leqslant |\theta_0 - \hat{\theta}|, \qquad (12.3.7)$$

и где

$$S_n'(x_1, \ldots, x_n;\ \theta^*) = \sum_{\xi=1}^{n} S'(x_\xi;\ \theta^*), \tag{12.3.8}$$

$$S'(x;\ \theta) = \frac{d}{d\theta} S(x;\ \theta). \tag{12.3.9}$$

Но (12.3.5) и (12.3.6) вместе эквивалентны тому, что

$$P\left(\frac{1}{\sqrt{n}} S_n(x_1, \ldots, x_n;\ \theta_0) = \frac{1}{\sqrt{n}} S_n(x_1, \ldots, x_n;\ \hat\theta) + \right.$$
$$\left. + \frac{1}{n} S_n'(x_1, \ldots, x_n;\ \theta^*)\right)[\sqrt{n}(\theta_0 - \hat\theta)] \text{ для всех } n > n(\delta, \varepsilon, n_0) \mid \theta_0\Big) >$$
$$> 1 - \varepsilon. \tag{12.3.10}$$

Так как $\hat\theta$ по определению удовлетворяет (12.3.2), (12.3.10) дает

$$P\left(\frac{1}{\sqrt{n}} S_n(x_1, \ldots, x_n;\ \theta^*) = \frac{1}{n} S_n'(x_1, \ldots, x_n;\ \theta^*)[\sqrt{n}(\theta_0 - \hat\theta)] \text{ для всех}\right.$$
$$\left. n > n(\delta, \varepsilon, n_0) \mid \theta_0\right) > 1 - \varepsilon, \tag{12.3.11}$$

иначе говоря, последовательности случайных величин

$$\frac{1}{\sqrt{n}} S_n(x_1, \ldots, x_n;\ \theta_0),\ n = 1, 2, \ldots, \tag{12.3.12}$$

и

$$\frac{1}{n} S_n'(x_1, \ldots, x_n;\ \theta_n^*)\ [\sqrt{n}(\theta_0 - \hat\theta)],\ n = 1, 2, \ldots, \tag{12.3.13}$$

таковы, что если одна из них сходится по распределению, то и другая также сходится по распределению. Далее, величина

$$\frac{1}{n} S_n'(x_1, \ldots, x_n;\ \theta) = \frac{1}{n} \sum_{\xi=1}^{n} S'(x_\xi;\ \theta_0) \tag{12.3.14}$$

является средним выборки объема n из совокупности с математическим ожиданием

$$\int_{-\infty}^{+\infty} S'(x;\ \theta_0)\, dF(x;\ \theta_0) = -B^2(\theta_0;\ \theta_0). \tag{12.3.15}$$

Следовательно, согласно **9.1.1** выражение в левой части равенства (12.3.14) сходится по вероятности к (12.3.15). Теперь согласно теореме **4.3.7**, так как $\hat\theta$ сходится по вероятности к θ_0, θ^* также сходится по вероятности к θ_0. Поэтому, применяя **4.3.8**, мы заключаем, что $\frac{1}{n}\ S_n(x_1, \ldots, x_n;\ \theta^*)$ сходится по вероятности к выражению в левой части (12.3.15).

Наконец, согласно **4.3.6** последовательности случайных величин (12.3.12) и (12.3.13) и последовательность

$$\sqrt{n}(\hat\theta - \theta_0) B^2(\theta_0;\ \theta_0),\ \ n = 1, 2, \ldots, \tag{12.3.16}$$

сходятся по распределению, если хотя бы одна из них сходится по распределению. Но мы знаем из **2.3.1**, что последовательность (12.3.12) сходится по распределению к

$$N(0,\ B^2(\theta_0;\ \theta_0)). \tag{12.3.17}$$

Отсюда для больших n $\hat\theta$ имеет асимптотическое распределение

$$N\left(\theta_0,\ \frac{1}{nB^2(\theta_0;\ \theta_0)}\right), \tag{12.3.18}$$

и, таким образом, теорема **12.3.3** доказана.

Асимптотическое распределение $\hat\theta$ в виде (12.3.18) было впервые установлено Фишером (1922).

(d) Асимптотическая эффективность оценок максимального правдоподобия. Предположим, что $\tilde\theta(x_1, \ldots, x_n)$ — смещенная оценка для θ_0 со смещением $b_n(\theta_0)$, определенным в (12.2.16). Если $\tilde\theta$ — состоятельная оценка θ_0, то ее смещение $b_n(\theta_0)$ стремится к 0 при $n \to \infty$.

Мы видим, что при любом данном n нижняя граница дисперсии $\sigma^2(\tilde\theta_n \mid \theta_0)$ дается (12.2.17). Отсюда

$$\sigma^2(\sqrt{n}\,\tilde\theta \mid \theta_0) = \mathscr{E}([\sqrt{n}\,(\tilde\theta - \theta_0 - b_n(\theta_0))]^2 \mid \theta_0) \geqslant \frac{[1 + b_n'(\theta_0)]^2}{B^2(\theta_0;\ \theta_0)}. \tag{12.3.19}$$

Так как это неравенство имеет место при любом n, предполагая, что $b_n'(\theta_0) \to 0$, при $n \to \infty$, мы имеем

$$\lim_{n\to\infty} \sigma^2(\sqrt{n}\,\tilde\theta \mid \theta_0) \geqslant \lim_{n\to\infty}\left(\frac{[1 + b_n'(\theta_0)]^2}{B^2(\theta_0;\ \theta_0)}\right) = \frac{1}{B^2(\theta_0;\ \theta_0)}, \tag{12.3.20}$$

т. е.

$$\lim_{n\to\infty} \sigma^2(\sqrt{n}\,\tilde\theta \mid \theta_0) \geqslant \frac{1}{B^2(\theta_0;\ \theta_0)}. \tag{12.3.21}$$

Отсюда получаем следующий результат:

12.3.5. *Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности с к. ф. р. $F(x;\ \theta_0)$, где $F(x;\ \theta)$ регулярна в смысле производной по θ и $B^2(\theta;\ \theta)$ существует при всех θ из Ω_0. Если $\tilde\theta$ — состоятельная оценка θ_0 и предел производной по θ смещения оценки $\tilde\theta$ равен 0 при $n \to \infty$, то предел дисперсии $\sqrt{n}\,\tilde\theta$ при $n \to \infty$ не может быть меньше $1/B^2(\theta_0;\ \theta_0)$.*

Важный частный случай состоятельной оценки θ представляет оценка максимального правдоподобия при условиях теоремы **12.3.3.** Из **12.2.3** мы также имеем

$$\lim_{n\to\infty} \sigma^2(\sqrt{n}\,\tilde\theta \mid \theta_0) = \frac{1}{B^2(\theta_0;\ \theta_0)}. \tag{12.3.22}$$

Предельная или асимптотическая эффективность оценки $\hat{\theta}$ при $n \to \infty$ определяется как

$$\operatorname{leff}(\hat{\theta} \mid \theta_0) = \lim_{n\to\infty} \frac{\sigma^2(\sqrt{n}\,\hat{\theta} \mid \theta_0)}{\sigma^2(\sqrt{n}\,\tilde{\theta} \mid \theta_0)} = \lim_{n\to\infty}\left[\frac{1}{\sigma^2(\sqrt{n}\,\tilde{\theta} \mid \theta_0)\, B^2(\theta_0;\, \theta_0)}\right], \qquad (12.3.23)$$

откуда ясно, что

$$\operatorname{leff}(\hat{\theta} \mid \theta_0) = 1. \qquad (12.3.24)$$

Любая оценка $\tilde{\theta}$ параметра θ_0, для которой $\operatorname{leff}(\tilde{\theta} \mid \theta_0) = 1$, называется асимптотически эффективной оценкой θ_0. Оценка $\hat{\theta}$ является таковой. Следовательно,

12.3.6. *При условиях теоремы* **12.3.3** *оценка максимального правдоподобия $\hat{\theta}$ для θ_0 асимптотически эффективна.*

Читателю следует отметить разницу в определениях эффективности (12.2.15) и асимптотической эффективности (12.3.23). Первое определение имеет смысл для любого n, но только когда существует эффективная оценка. Последнее представляет собой свойство, которое определяется как предел для $n \to \infty$ и имеет смысл, когда существует асимптотически эффективная оценка.

Асимптотическую эффективность состоятельной оценки θ можно также рассматривать как обобщение для больших выборок фишеровского понятия количества информации в $\tilde{\theta}$ относительно θ_0.

Дальнейшее детальное изучение асимптотических свойств оценок максимального правдоподобия принадлежит Бахадуру (1958), Баранкину и Гёрланду (1951), Крафту и Ле Каму (1956), Рао (1949), Вальду (1948) и др.

Пример. Чтобы проиллюстрировать асимптотическую эффективность оценки, предположим, что $(x_1, \ldots, x_n)$ — выборка из нормальной совокупности $N(\mu, \sigma^2)$. Пусть $\tilde{\mu}$ — медиана и $\hat{\mu}$ — среднее выборки. Из теоремы **9.6.6** для $p = \frac{1}{2}$ следует, что асимптотическим распределением для $\tilde{\mu}$ при больших n является $N(\mu, \pi\sigma^2/2n)$. Читатель может проверить, что при любом n существует эффективная оценка $\hat{\mu}$ для μ и что $\hat{\mu}$ — выборочное среднее. Более того, из **8.3.3** мы знаем, что выборочное среднее распределено согласно $N(\mu, \sigma^2/n)$.

Поэтому, используя (12.3.23), получаем, что асимптотическая эффективность выборочной медианы $\tilde{\mu}$ при оценивании истинного значения μ равна

$$\operatorname{leff}(\tilde{\mu} \mid \mu_0) = \frac{\sigma^2}{\left(\frac{\pi}{2}\right)\sigma^2} = 0{,}637.$$

Это означает, что при больших n среднее выборки объема $\frac{2}{\pi}\, n$ будет оценивать истинное значение среднего μ нормальной совокупности $N(\mu, \sigma^2)$ с той же точностью, что и медиана выборки объема n, независимо от того, каковы значения μ и σ^2.

(е) Построение функции оценки максимального правдоподобия, имеющей асимптотическую дисперсию $1/n$. Возвращаясь к теореме **12.3.3**, замечаем, что при соответствующих условиях регулярности, наложенных на функцию $F(x;\ \theta)$, асимптотическая дисперсия оценки максимального правдоподобия $\hat{\theta}$ при больших n равна величине $\frac{1}{nB^2(\theta_0;\ \theta_0)}$. Большой теоретический и практический интерес представляет вопрос, можно ли использовать в качестве нового параметра ζ некоторую функцию от θ, скажем $\zeta(\theta)$ такую, что асимптотическая дисперсия оценки $\zeta(\hat{\theta})$ равнялась бы просто $\frac{1}{n}$. При некоторых условиях функции такого сорта могут быть найдены. Для этого мы поступим следующим образом. Пусть $\zeta(\theta)$ — функция от θ, имеющая производную $\zeta'(\theta)$ и однозначную обратную функцию $\theta(\zeta)$ в области Ω_0; пусть $\theta(\zeta)$ также имеет производную $\theta'(\zeta)$ по ζ. Обозначим через ζ_0 истинное значение ζ, т. е. $\theta(\zeta_0)=\theta_0$. Согласно (12.1.8) мы можем написать

$$B^2(\theta,\ \theta)=\int_{-\infty}^{+\infty}(S(x;\ \theta))^2\,dF(x;\ \theta). \tag{12.3.25}$$

Так как

$$S(x,\ \theta)=\frac{\partial}{\partial\theta}\log dF(x;\ \theta), \tag{12.3.26}$$

имеем

$$S(x;\theta(\zeta))=\frac{\partial}{\partial\zeta}\log dF(x;\ \theta(\zeta))\,\zeta'(\theta). \tag{12.3.27}$$

Подставляя полученное в (12.3.25), будем иметь

$$B^2(\theta(\zeta);\ \theta(\zeta))=[\zeta'(\theta)]^2\int_{-\infty}^{+\infty}\left[\frac{\partial}{\partial\zeta}\log dF(x;\ \theta(\zeta))\right]^2 dF(x;\ \theta(\zeta)). \tag{12.3.28}$$

Требование, чтобы оценка максимального правдоподобия $\hat{\zeta}$ (для ζ_0) имела асимптотическую дисперсию $\frac{1}{n}$, эквивалентно требованию, чтобы интеграл в правой части (12.3.28) был равен 1. Отсюда функция $\zeta(\theta)$ должна удовлетворять дифференциальному уравнению

$$\left(\frac{d\theta}{d\zeta}\right)^2=\frac{1}{B^2(\theta(\zeta),\ \theta(\zeta))} \tag{12.3.29}$$

или

$$\zeta(\theta)=\int_{\theta_0}^{\theta}B(\theta,\ \theta)\,d\theta+\zeta_0,\qquad B(\theta,\ \theta)=+\sqrt{B^2(\theta,\ \theta)}. \tag{12.3.30}$$

Суммируя эти факты, мы получаем следующий результат:

12.3.7. *Пусть $(x_1,\ \ldots,\ x_n)$ — выборка из совокупности с к. ф. р. $F(x;\ \theta_0)$, где $F(x;\ \theta)$ регулярна в смысле первой и второй производных по θ. Если $\hat{\theta}$ — оценка максимального правдоподобия для θ_0,* *тогда функция* $\zeta(\hat{\theta})$, *задаваемая* (12.3.30), *имеющая производную* $\zeta'(\theta)$ *и однозначную обратную* $\theta(\zeta)$ *в области* Ω_0, *асимптотически нормальна* $N\left(\zeta_0, \frac{1}{n}\right)$.

Пример. Пусть $(x_1, \ldots, x_n)$ — выборка из распределения Пуассона Ро (μ_0) с к. ф. р.

$$p(x, \mu_0) = \frac{\mu_0^x e^{-\mu_0}}{x!}, \qquad x = 0, 1, \ldots$$

Мы хотим найти функцию от μ (обозначим ее через $\zeta(\mu)$) такую, что $\zeta(\hat{\mu})$ асимптотически при $n \to \infty$ имеет распределение $N\left(\zeta(\mu_0), \frac{1}{n}\right)$. Из примера § 12.2 (с) мы знаем, что оценкой максимального правдоподобия $\hat{\mu}$ для параметра μ_0 служит выборочное среднее $\bar{x}$. С помощью формулы (12.3.25) можно проверить, что

$$B^2(\mu, \mu) = \sum_{x=0}^{\infty} \left(\frac{x}{\mu} - 1\right)^2 p(x, \mu) = \frac{1}{\mu}$$

(мы положили $dF(x, \mu) = p(x, \mu)$).

Применяя (12.3.30) и выбирая $\zeta_0 = \sqrt{2\mu_0}$, находим

$$\zeta(\mu) = 2\sqrt{\mu}.$$

Отсюда согласно теореме **12.3.7** мы заключаем, что если $(x_1, \ldots, x_n)$ — выборка из популяции с распределением Пуассона Ро (μ_0), то $2\sqrt{\bar{x}}$ асимптотически при $n \to \infty$ имеет нормальное распределение $N\left(2\sqrt{\mu_0}, \frac{1}{n}\right)$.

12.4. Оценивание с помощью интервалов

(а) Определение доверительного интервала. Предположим, что $(x_1, \ldots, x_n)$ — случайная величина с к. ф. р. $F_n(x_1, \ldots, x_n; \theta)$. Пусть $\underline{\theta}(x_1, \ldots, x_n)$, $\bar{\theta}(x_1, \ldots, x_n)$ — две функции выборочных значений (случайные величины) такие, что $\underline{\theta} < \bar{\theta}$. Если $\underline{\theta}$ и $\bar{\theta}$ могут быть выбраны так, что при заданном γ

$$P(\underline{\theta} < \theta < \bar{\theta}) = \gamma, \tag{12.4.1}$$

где $P(\underline{\theta} < \theta < \bar{\theta} \mid \theta)$ означает вероятность события, написанного в скобках при $(x_1, \ldots, x_n)$, подчиненных к. ф. р. $F_n(x_1, \ldots, x_n; \theta)$, то $(\underline{\theta}, \bar{\theta})$ называется 100γ-процентным доверительным интервалом для θ; $\underline{\theta}$ и $\bar{\theta}$ называются соответственно нижней и верхней доверительными границами для параметра θ, а γ — доверительным коэффициентом. Заметим, что $(\underline{\theta}, \bar{\theta})$ — двумерная случайная величина такая, что вероятность того, что интервал $(\underline{\theta}, \bar{\theta})$ содержит истинное значение параметра θ распределения $F_n(x_1, \ldots, x_n; \theta)$, равна γ. Для случая выбора из нормальной совокупности примеры доверительных интервалов уже были даны в §§ 10.2 (с), 10.4, 10.5 и 10.6. Математический аппарат доверительных интервалов был впервые введен Лапласом (1814). Лаплас занимался задачей распределения значения p в биномиальном распределении (6.2.2) по наблюденным значениям случайной величины x и рассматривал доверительный интервал фиксированным, а значение p случайным. Процедура Лапласа была заново открыта Уилсоном (1927), который дал правильную интерпретацию доверительного интервала как случайного. Современная терминология и теория доверительных интервалов принадлежит Нейману (1937).

(b) Процедура построения доверительных интервалов по выборкам из совокупностей с непрерывными к. ф. р. Процедура, с помощью которой в некоторых случаях может быть построен доверительный интервал по выборке с непрерывной к. ф. р., состоит в следующем:

12.4.1. *Предположим, что* $(x_1, \ldots, x_n)$ — *выборка из распределения с непрерывной к. ф. р.* $F(x; \theta)$. *Пусть функция* $g(x_1, \ldots, x_n; \theta)$ 1) *определена в каждой точке* θ *интервала* (θ_1, θ_2), *содержащего* θ_0, *и в каждой точке выборочного пространства* R_n, *за исключением, может быть, лишь множества вероятностной меры* 0, 2) *непрерывна и монотонно убывает или возрастает по* θ *и* 3) *имеет к. ф. р., не зависящую от* θ. *Если* (g_1, g_2) — *интервал, для которого* $P(g_1 < g < g_2 \mid \theta) = \gamma$, *то при истинном значении параметра* θ, *равном* θ_0, *решения* $\underline{\theta}$ *и* $\bar{\theta}$ (*где* $\underline{\theta} < \bar{\theta}$) *уравнений* $g(x_1, \ldots, x_n; \theta) = g_1, g_2$ *существуют, и* $(\underline{\theta}, \bar{\theta})$ — 100γ-*процентный доверительный интервал для* θ_0.

Для доказательства **12.4.1** заметим, что, так как к. ф. р. случайной величины $g(x_1, \ldots, x_n; \theta)$ не зависит от θ, мы можем по заданным θ и γ, $0 < \gamma < 1$, выбрать g_1 и g_2 (конечно, многими способами) так, чтобы если θ_0 — истинное значение θ, то

$$P(g_1 < g(x_1, \ldots, x_n; \theta) < g_2 \mid \theta_0) = \gamma. \tag{12.4.2}$$

Так как $g(x_1, \ldots, x_n; \theta)$ непрерывно и монотонно убывает (возрастает) по θ на интервале (θ_1, θ_2), (12.4.2), очевидно, эквивалентно утверждению

$$P(\underline{\theta} < \theta_0 < \bar{\theta} \mid \theta_0) = \gamma, \tag{12.4.3}$$

где $\underline{\theta}$, $\bar{\theta}$, $\underline{\theta} < \bar{\theta}$ — решения уравнений $g(x_1, \ldots, x_n; \theta) = g_1, g_2$ и, конечно, случайные величины.

Встает вопрос, всегда ли можно найти функцию $g(x_1, \ldots, x_n; \theta)$, к. ф. р. которой не зависела бы от θ, если известно, что $F(x; \theta)$ непрерывна по θ. Это можно сделать всегда. Действительно, при простом случайном выборе форма, принимаемая функцией $F_n(x_1, \ldots, x_n; \theta)$, именно $\prod_{\xi=1}^{n} F(x_\xi; \theta)$, и является требуемой функцией, если $F(x; \theta)$ — непрерывная монотонно возрастающая или убывающая функция от θ для всех точек пространства изменения x, за исключением, быть может, множества вероятностной меры 0, ибо случайные величины $F(x_\xi, 0)$, $\xi = 1, 2, \ldots, n$, независимы и, как указано в § 8.7(а), имеют прямоугольное распределение $R\left(\frac{1}{2}, 1\right)$. Кроме того, $-\log F(x_\xi; \theta)$ имеет Г-распределение $G(1)$; $-\sum_{\xi=1}^{n} \log F(x_\xi; \theta)$ также имеет распределение $G(n)$ согласно известному свойству Г-распределений. Поэтому

$$P\left(-\log b_2 < -\sum_{\xi=1}^{n} \log F(x_\xi; \theta) < -\log b_1 \mid \theta\right) = \\ = \frac{1}{\Gamma(n)} \int_{-\log b_2}^{-\log b_1} y^{n-1} e^{-y}\, dy. \quad (12.4.4)$$

Если b_1 и b_2 выбраны так, что интеграл в правой части (12.4.4) равен γ, то мы имеем

$$P\left(b_1 < \prod_{\xi=1}^{n} F(x_\xi; \theta) < b_2 \mid \theta\right) = \gamma. \quad (12.4.5)$$

Если теперь функция $F(x; \theta)$ непрерывна и монотонно убывает (или возрастает) по θ при всех значениях x из R_1 (исключая, может быть, множество нулевой вероятности), то такова же и функция $\prod_{\xi=1}^{n} F(x_\xi; \theta)$. Поэтому неравенства в скобках в (12.4.5) могут быть обращены и записаны в виде

$$P(\underline{\theta} < \theta_0 < \bar{\theta} \mid \theta_0) = \gamma, \quad (12.4.6)$$

давая, таким образом, 100γ-процентный доверительный интервал для θ_0.

Читателю следует заметить, что если предположение о монотонности функции $g(x_1, \ldots, x_n; \theta)$ в **12.4.1** отброшено, то (12.4.2) все же имеет место, но обращение неравенств в (12.4.2) дает вместо доверительного интервала $(\underline{\theta}, \bar{\theta})$ случайное множество E значений (θ_1, θ_2), зависящее от $(x_1, \ldots, x_n)$. В этом случае (12.4.3) заменилось бы на $P(\theta \in E \mid \theta) = \gamma$. Следует настоятельно подчеркнуть, что случайным является множество E, а не θ. С точки зрения оценивания параметра доверительный интервал является более привлекательной и более полезной оценкой, чем случайное множество E. Обобщение идеи интервального оценивания параметра θ в виде оценивания с помощью множества будет нам полезно в § 17.7, где мы будем заниматься задачей обобщения интервального оценивания на многомерный параметр θ.

(с) Доверительные интервалы по выборкам из дискретных распределений. В том случае, когда $F(x; \theta)$ служит к. ф. р. дискретной случайной величины, теорема **12.4.1**, конечно, не применима, т. е. нельзя найти доверительного интервала, имеющего коэффициент доверия, в точности равный γ, если γ произвольно. Однако при некоторых предположениях можно построить доверительные интервалы, имеющие коэффициент доверия, не меньший чем γ, хотя здесь ситуация менее интересна, чем в случае непрерывной случайной величины. Сформулируем результат более точно:

12.4.2. *Предположим, что* $(x_1, \ldots, x_n)$ — *выборка из совокупности с функцией распределения* $F(x, \theta)$, *где* x — *дискретная случайная величина и* θ — *параметр, область изменения которого — интервал* (θ_1, θ_2). *Пусть* $\tilde{\theta}$ — *оценка для* θ, *определенная в каждой точке пространства* R_n, *несущей положительную массу вероятности, и принимающая значения из* θ_1, θ_2. *Обозначим через* $V(\tilde{\theta}; \theta)$ *к. ф. р. оценки* $\tilde{\theta}$ *и* $V^*(\theta^*; \theta) = 1 - V(\tilde{\theta}; \theta)$. *Пусть, кроме того,* $V(\tilde{\theta}, \theta)$ *непрерывна и убывает по* θ *в каждой точке* $\tilde{\theta}$ *сосредоточения массы, так что* $\lim\limits_{\theta \to \theta_1} V(\tilde{\theta}; \theta) = 1$, $\lim\limits_{\theta \to \theta_2} V(\tilde{\theta}, \theta) = 0$ *для всех* $\tilde{\theta} \in (\theta_1, \theta_2)$. *Пусть* $\underline{\theta}$ *и* $\bar{\theta}$ — *значения* θ, *для которых* $V(\tilde{\theta}, \theta) = \gamma_1$ *и* $V^*(\tilde{\theta}; \theta) = \gamma_1^*$, *где* γ_1 *и* γ_2 *неотрицательны и* $0 < \gamma = 1 - \gamma_1 - \gamma_1^* < 1$. *Тогда* $(\underline{\theta}, \bar{\theta})$ — *доверительный интервал для* θ *с коэффициентом доверия* $\geqslant \gamma$.

Доказательство **12.4.2.** Пусть θ_1 обозначает наибольшее значение θ, для которого $V_1^*(\tilde{\theta}, \theta_0) \leqslant \gamma_1$, и θ_2 — наименьшее значение θ, для которого $V_1^*(\tilde{\theta}, \theta_0) \leqslant \gamma_1^*$. Тогда

$$P(\theta_1 < \tilde{\theta} < \theta_2) \geqslant \gamma, \tag{12.4.7}$$

где $\gamma = 1 - \gamma_1 - \gamma_1^*$. Но так как $V(\tilde{\theta}; \theta)$ монотонно убывает по θ и не возрастает по $\tilde{\theta}$, очевидно, что $\theta_1 < \tilde{\theta} < \theta_2$ тогда и только тогда, когда $V(\tilde{\theta}; \theta_0) > \gamma_1$ и $V^*(\tilde{\theta}; \theta_0) > \gamma_1^*$, т. е. тогда и только тогда, когда $\underline{\theta} \leqslant \theta_0 \leqslant \bar{\theta}_0$. Поэтому

$$P(\underline{\theta} \leqslant \theta_0 \leqslant \bar{\theta} \mid \theta_0) \geqslant \gamma \tag{12.4.8}$$

и, следовательно, $(\underline{\theta}, \bar{\theta})$ — доверительный интервал для θ_0 с коэффициентом доверия $\geqslant \gamma$.

Пример. В качестве иллюстрации рассмотрим выборку $(x_1, \ldots, x_n)$ из совокупности с биномиальным распределением $\mathrm{Bi}(1, p_0)$. Плотность этого распределения равна

$$f(x; p_0) = p_0^x (1 - p_0)^{1-x}.$$

Точки $x = 0, 1$ несут всю вероятностную массу, и p_0 — некоторое число из $(0, 1)$. Выборочное среднее $\bar{x}$ имеет биномиальное распределение $\mathrm{Bi}(1, p)$ на своем выборочном пространстве $0, \frac{1}{n}, \ldots, \frac{n}{n}$ и является состоятельной оценкой для p_0. Кроме того, если совокупность, из которой производится выбор, имеет биномиальное распределение $\mathrm{Bi}(1, p)$, к. ф. р. для $\bar{x}$ определяется как

$$V(\bar{x};\ p) = \sum_{i=0}^{n\bar{x}} \binom{n}{i} p^i (1-p)^{n-i},$$

и $V^*(\bar{x};\ p)$ равна

$$V(\bar{x};\ p) = 1 - V(\bar{x};\ p).$$

Теперь $V(\bar{x};\ p)$ монотонно убывает по p и $V^*(\bar{x};\ p)$ монотонно возрастает по p для $0 < n\bar{x} < p$, ибо $(d/dp)\,V^*(\bar{x};\ p) > 0$ и $(d/dp)\,V(\bar{x},\ p) < 0$. Далее, $V(\bar{x};\ 0) = 1$ и $V(\bar{x};\ 1) = 0$ для всех $x \in (0,\ 1)$. Отсюда, применяя теорему **12.4.2**, получаем, что если $\underline{p}$ и $\bar{p}$ — соответственно решения уравнений $V(\bar{x};\ p_0) = \gamma_1$ и $V^*(\bar{x};\ p_0) = \gamma_1^*$ и $\gamma = 1 - \gamma_1 - \gamma_1^*$ (p_0 — истинное значение параметра p популяции), то

$$P(\underline{p} \leqslant p_0 \leqslant \bar{p} \mid p_0) \geqslant \gamma,$$

т. е. $(\underline{p},\ \bar{p})$ — доверительный интервал для p_0 с коэффициентом доверия $\geqslant \gamma$.

Доверительный интервал такого типа для p_0 был построен, представлен графически и опубликован Клоппером и Пирсоном (1934) для случая, когда $\gamma_1 = \gamma_1^* = 0{,}05,\ 0{,}025$, и для различных значений n от 10 до 1000.

Процедура, аналогичная только что приведенной, была применена Гарвудом (1936) и Рикером (1937) к задаче построения доверительных интервалов для параметра μ распределения Пуассона.

Описанная в **12.4.2** процедура может быть обобщена на случай, когда параметрическое пространство состоит из дискретного множества точек или является каким-либо другим множеством интервала $(\theta_1,\ \theta_2)$. Такая ситуация возникает, например, при построении доверительного интервала для p в случае, когда случайная величина x имеет гипергеометрическое распределение $H(N,\ n;\ p)$, определенное в § 6.1(a), где N и n известны. Здесь параметрическое пространство Ω для p состоит из дискретного набора точек $0,\ 1/N,\ 2/N,\ \ldots,\ 1$. Доверительные интервалы для p были построены Чжуном и Де Лёэри (1950) для $N = 500,\ 2500,\ 10\,000$, $n = 0{,}05,\ 0{,}10,\ (0{,}10),\ 0{,}90$ и для $\gamma = 0{,}90,\ 0{,}95,\ 0{,}99$.

(d) Замечания относительно фидуциальных вероятностей и фидуциальных интервалов. Понятие фидуциального интервала было введено Фишером (1935b). Оно тесно связано с понятием доверительного интервала. Предположим, что параметрическое пространство θ представляет собой интервал $(\theta_1,\ \theta_2)$ и s — статистика, выборочное пространство которой — интервал $(a,\ b)$. Пусть $v(s,\ \theta)$ — плотность и $V(s;\ \theta)$ — к. ф. р. статистики s. Если $V^*(\theta;\ s) = 1 - V(s;\ \theta)$ непрерывна и строго монотонно возрастает по θ при каждом s из $(a,\ b)$, так что $\lim\limits_{\theta \to \theta_1} V^*(\theta;\ s) = 0$ и $\lim\limits_{\theta \to \theta_1} V^*(\theta;\ s) = 1$ при каждом s из $(a,\ b)$, то $G(\theta,\ s)$ как функция θ на интервале $(\theta_1,\ \theta_2)$ имеет все формальные свойства к. ф. р. при каждом s из $(a,\ b)$ и называется фидуциальной к. ф. р. для θ, основанной на статистике s. Таким образом, если θ' и θ'', $\theta' < \theta''$, выбраны так, что $V^*(\theta';\ s) = \gamma_1$ и $V^*(\theta'';\ s) = \gamma_2$, где

$\gamma_2 - \gamma_1 = \gamma$, $0 < \gamma < 1$, мы можем написать следующее фидуциально-вероятностное утверждение:

$$\operatorname{fid} P(\theta' < \theta < \theta'' \mid s) = \gamma, \tag{12.4.9}$$

так что (θ', θ'') будет называться 100γ-процентным фидуциальным интервалом для θ. Если θ_0 — истинное значение θ и s', s'' — значения s, удовлетворяющие соответственно уравнениям $V(s;\ \theta_0) = 1 - \gamma_2$ и $V(s;\ \theta_0) = 1 - \gamma_1$, где $\gamma_2 - \gamma_1 = \gamma$, то мы видим, что $P(s' < s < s'' \mid \theta_0) = \gamma$. Но θ_0 лежит в интервале (θ', θ'') только в том случае, если s лежит в интервале (θ', θ''). Таким образом, (θ', θ'') есть также 100γ-процентный доверительный интервал для θ_0.

Фидуциальное распределение и фидуциальные интервалы порождаются еще одной процедурой, именно с помощью центральных функций (pivotal functions). Для статистики s и рассматриваемого параметра θ определим функцию $g(s;\ \theta)$, которая строго монотонно возрастает по θ для каждого s и имеет то же свойство, как функция s при каждом θ. Пусть $g(s;\ \theta)$ имеет элемент вероятности $h(g)\,dg$, который зависит от θ только через g. Если $g(s,\ \theta)$ имеет первые частные производные по s и θ, то $h(g)(\partial g/\partial s)\,ds$ есть элемент вероятности случайной величины s и $h(g)(\partial g/\partial\theta)\,d\theta$ — элемент фидуциальной вероятности параметра θ. Таким образом, пусть g_1 и g_2 $(g_1 < g_2)$ — такие числа, что

$$P(g_1 < g(s;\ \theta) < g_2 \mid \theta) = \gamma. \tag{12.4.10}$$

Пусть s_1 и s_2 $(s_1 < s_2)$ — значения s, для которых соответственно $g(s;\ \theta) = g_1$, g_2, тогда как θ_1 и θ_2 $(\theta_1 < \theta_2)$ — значения θ, для которых соответственно $g(s;\ \theta) = g_1$, g_2. Тогда $\theta \in (\theta_1, \theta_2)$ тогда и только тогда, когда $s \in (s_1, s_2)$. Поэтому

$$\int_{\theta_1}^{\theta_2} h(g)\frac{\partial g}{\partial\theta}\,d\theta = \int_{s_1}^{s_2} h(g)\frac{\partial g}{\partial s}\,ds,$$

где левая часть есть фидуциальная вероятность $\operatorname{fid} P(\theta_1 < \theta < \theta_2)$, тогда как правая часть есть вероятность $P(s_1 < s < s_2)$. Аналогичные рассмотрения можно провести, когда $g(s;\ \theta)$ строго монотонно возрастает по s и убывает по θ и наоборот, а также когда $g(s;\ \theta)$ убывает по обеим переменным.

Замечание. В обеих приведенных процедурах, порождающих фидуциальные распределения, фидуциальные и доверительные интервалы эквивалентны.

Существуют, однако, ситуации, когда фидуциальные интервалы не эквивалентны доверительным интервалам, и это обстоятельство вызвало спор относительно того, какой смысл следует придавать фидуциальному интервалу. Этот спор сосредоточился вокруг смысла формального фидуциального распределения разности $\mu_1 - \mu_2$ средних двух нормальных распределений $N(\mu_1, \sigma_1^2)$ и $N(\mu_2, \sigma_2^2)$, когда в рассмотрение вводятся выборочные средние и дисперсии для каждого распределения. Более точно предположим, что n_1, $\bar{x}_1$, s_1^2 и n_2, $\bar{x}_2$, s_2^2 — объем, среднее и дисперсия двух независимых выборок из $N(\mu_1, \sigma_1^2)$ и $N(\mu_2, \sigma_2^2)$ соответственно. Тогда

$$t_1 = \sqrt{n_1}(\bar{x}_1 - \mu_1)/s_1, \quad t_2 = \sqrt{n_2}(\bar{x}_2 - \mu_2)/s_2 \tag{12.4.11}$$

независимы и имеют распределения Стьюдента с $(n_1 - 1)$ и $(n_2 - 1)$ степенями свободы соответственно. Если $f_m(t)\,dt$ — элемент вероятности стьюдентова отношения t с m степенями свободы, то элемент вероятности величин t_1 и t_2 будет $f_{n_1-1}(t_1) f_{n_2-1}(t_2)\,dt_1\,dt_2$. Используя величины t_1 и t_2 из (12.4.11) в качестве центральных, получаем фидуциальный элемент вероятности (μ_1, μ_2) при заданных n_1, $\bar{x}_1$, s_1^2, n_2, $\bar{x}_2$, s_2^2:

$$f_{n_1-1}(\sqrt{n_1}(\bar{x}_1 - \mu_1)/s_1) f_{n_2-1}(\sqrt{n_2}(\bar{x}_2 - \mu_2)/s_2)\,[\sqrt{n_1 n_2}/s_1 s_2]\,d\mu_1\,d\mu_2. \tag{12.4.12}$$

Из (12.4.12) можно формально найти фидуциальное распределение $(\mu_1 - \mu_2)$, а из этого распределения — фидуциальные интервалы для $(\mu_1 - \mu_2)$. Но можно показать, что не существует функции $g(x_1, x_2, s_1^2, s_2^2, \mu_1 - \mu_2)$, удовлетворяющей некоторым условиям регулярности, распределение которой зависело бы только от n_1, n_2 и g, из которой можно было бы получить доверительные интервалы для $(\mu_1 - \mu_2)$. Таким образом, в этом примере фидуциальные и доверительные интервалы, по-видимому, различны. Возникает вопрос, как здесь интерпретировать фидуциальные интервалы. Эта задача известна под названием проблемы Беренса — Фишера [см. Фишер (1935 b)]. Читателю, заинтересованному в детальном изучении проблемы Беренса — Фишера, следует познакомиться с работой Тьюки (1957c) и ссылками в конце этой работы.

12.5. Интервальное оценивание при помощи больших выборок

(а) Интервальные оценки, связанные с функцией правдоподобия. Асимптотическая теория интервального оценивания по большим выборкам проистекает непосредственно из теории точечного оценивания по большим выборкам, развитой в § 12.3. Асимптотически эффективные точечные оценки, как мы увидим, порождают для больших выборок асимптотически наикратчайшие доверительные интервалы.

Предположим, что $(x_1, \ldots, x_n)$ — выборка из распределения с к. ф. р. $F(x;\,\theta_0)$. Нам будет удобно называть функцию

$$h_n(x_1, \ldots, x_n;\,\theta) = \frac{S_n(x_1, \ldots, x_n;\,\theta)}{nB(\theta_0, \theta)} \tag{12.5.1}$$

функцией правдоподобия. Здесь $B(\theta_0, \theta) = +\sqrt{B^2(\theta_0\theta)}$.

При выполнении условий теоремы **12.3.1**, как мы знаем, $\sqrt{n} \times h_n(x_1, \ldots, x_n;\,\theta)$ имеет в качестве предельного распределения $N(0, 1)$, когда истинное значение параметра θ равно θ_0. Если мы потребуем выполнения более сильных условий теоремы **12.3.3**, то, так как вероятность того, что при всех $n > n_\varepsilon$

$$\sqrt{n}\,h_n(x_1, \ldots, x_n;\,\theta) = h'_n(x_1, \ldots, x_n;\,\theta^*)\,[\sqrt{n}\,(\theta_0 - \hat{\theta})]$$

превышает $1 - \varepsilon$ для точек из множества E_n, определенного в (12.3.5), где $h'_n(x_1, \ldots, x_n;\,\theta) = \frac{\partial}{\partial\theta} h_n(x_1, \ldots, x_n;\,\theta)$ и $\hat{\theta}$, и θ^* определены в (12.3.7), две последовательности случайных величин

$$\sqrt{n}\,h_n(x_1, \ldots, x_n;\,\theta_0), \quad n = 1, 2, \ldots, \tag{12.5.2}$$

и
$$h_n'(x_1, \ldots, x_n; \theta^*)[\sqrt{n}(\theta_0 - \hat{\theta})] \qquad (12.5.3)$$

сходятся по распределению к нормальному закону $N(0, 1)$. Кроме того, если $(\partial/\partial\theta) B^2(\theta_0, \theta)$ существует и ограничена на Ω_0 и $\theta = \theta_0$, то
$$h_n'(x_1, \ldots, x_n; \theta^*) \quad \text{и} \quad h_n'(x_1, \ldots, x_n; \hat{\theta}) \qquad (12.5.4)$$

сходятся по вероятности к $B(\theta_0, \theta_0)$ при $n \to \infty$.

Мы можем написать
$$\lim P[-\lambda_\gamma < h_n'(x_1, \ldots, x_n; \hat{\theta})[\sqrt{n}(\theta_0 - \hat{\theta})] < +\lambda_\gamma \mid \theta_0] = \gamma,$$

где $\Phi(\lambda_\gamma) - \Phi(-\lambda_\gamma) = \gamma$. Отсюда видно, что
$$\lim_{n\to\infty} P\left[\hat{\theta} - \frac{\lambda_\gamma}{\sqrt{n}\, h_n'(x_1, \ldots, x_n; \hat{\theta})} < \theta_0 < \hat{\theta} + \frac{\lambda_\gamma}{\sqrt{n}\, h_n'(x_1, \ldots, x_n; \theta)} \,\middle|\, \theta_0\right] = \gamma. \qquad (12.5.5)$$

Мы будем говорить, что интервал, имеющий конечные точки $\hat{\theta} \pm \dfrac{\lambda_\gamma}{\sqrt{n}\, h_n'(x_1, \ldots, x_n; \hat{\theta})}$, является асимптотически 100γ-процентным доверительным интервалом для θ_0 при большом n. Его длина равна $\dfrac{2\lambda_\gamma}{\sqrt{n}\, h_n'(x_1, \ldots, x_n; \hat{\theta})}$.

(b) Интервальные оценки, полученные по произвольным оценкам. Мы покажем, что предел отношения квадрата длины только что полученного интервала к квадрату длины аналогичного интервала, полученного при помощи произвольной оценки $g_n(x_1, \ldots, x_n; \theta)$, удовлетворяющей некоторым условиям, не может превосходить 1. Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности с к. ф. р. $F(x, \theta)$, которая регулярна вместе с двумя производными по θ, и пусть через F_n обозначена величина $\prod_{\xi=1}^{n} F(x_\xi; \theta)$ и F_{n_0} — значение F_n при $\theta = \theta_0$.

Пусть
$$g_n(x_1, \ldots, x_n; \theta) \qquad (12.5.6)$$

— случайная величина такая, что

i) $\int\limits_{R_n} [\sqrt{n}\, g_n(x_1, \ldots, x_n; \theta)]^j dF_n = 0, 1; \qquad j = 1, 2, \quad (12.5.7)$

для θ из Ω_0;

ii) $\sqrt{n}\, g_n(x_1, \ldots, x_n; \theta_0)$ имеет в пределе при $n \to \infty$ распределение $N(0, 1)$, если θ_0 — истинное значение θ;

iii) $g_n(x_1, \ldots, x_n; \theta)$ имеет непрерывную производную по θ, $g_n'(x_1, \ldots, x_n; \theta)$, для θ из Ω и для всех точек выборочного пространства R_n, за исключением, может быть, множества вероятностной меры 0;

iv) если θ_0 — истинное значение θ, то $g_n'(x_1, \ldots, x_n; \theta)$ сходится по вероятности к $B^*(\theta_0, \theta)$ равномерно по θ из Ω, где $B^*(\theta_0, \theta)$ ограничена на Ω_0 и $B^*(\theta_0, \theta_0) \neq 0$;

v) если

$$A_n^*(\theta, \theta') = \int\limits_{R_n} g_n(x_1, \ldots, x_n; \theta')\, dF_n$$

и

$$B_n^*(\theta, \theta') = \int\limits_{R_n} g_n'(x_1, \ldots, x_n; \theta')\, dF_n,$$

то

$$\begin{aligned} &\frac{\partial}{\partial\theta} A_n^*(\theta, \theta') = B_n^*(\theta, \theta') \quad \text{для } (\theta, \theta') \in \Omega_0 \times \Omega_0, \\ &A_n^*(\theta, \theta) = 0 \qquad \text{для } \theta \in \Omega_0, \\ &\lim_{n\to\infty} B_n^*(\theta, \theta) = B^*(\theta, \theta) \quad \text{для } \theta \in \Omega_0. \end{aligned} \tag{12.5.8}$$

Любая функция, имеющая перечисленные свойства, будет называться *регулярной оценивающей функцией* для θ_0. Следует заметить, что функция $h_n(x_1, \ldots, x_n; \theta)$, определенная в (12.5.1), является регулярной оценивающей функцией для θ_0. Следуя пути, намеченному при доказательстве теоремы **12.3.2**, можно проверить, что если для $n \geqslant$ некоторого n_0 уравнение $g_n(x_1, \ldots, x_n; \theta) = 0$ имеет единственный корень, то уравнения

$$g_n(x_1, \ldots, x_n; \theta) = 0, \qquad n = n_0, n_0 + 1, \ldots, \tag{12.5.9}$$

имеют последовательность корней

$$\tilde{\theta}(x_1, \ldots, x_n), \qquad n = n_0, n_0 + 1, \ldots, \tag{12.5.10}$$

которая сходится по вероятности к θ_0. Кроме того, две последовательности случайных величин

$$\sqrt{n}\, g_n(x_1, \ldots, x_n; \theta_0), \qquad n = n_0, n_0 + 1, \ldots, \tag{12.5.11}$$

$$g_n'(x_1, \ldots, x_n; \tilde{\theta}^*)\,[\sqrt{n}\,(\theta_0 - \tilde{\theta})], \quad n = n_0, n_0 + 1, \ldots, \tag{12.5.12}$$

где $\tilde{\theta}^*$ удовлетворяет неравенству $|\theta_0 - \tilde{\theta}^*| \leqslant |\theta_0 - \tilde{\theta}|$, сходятся по распределению к $N(0, 1)$. Но из теоремы **4.3.8** мы видим, что

$$g_n'(x_1, \ldots, x_n; \tilde{\theta}^*) \text{ и } g_n'(x_1, \ldots, x_n; \tilde{\theta}) \tag{12.5.13}$$

сходятся по вероятности к $B^*(\theta_0, \theta_0)$ при $n \to \infty$. Поэтому, как и в (12.5.5), мы можем утверждать:

$$\lim_{n\to\infty} P\left[\tilde{\theta} - \frac{\lambda_\gamma}{\sqrt{n}\, g_n'(x_1, \ldots, x_n; \tilde{\theta})} < \theta_0 < \tilde{\theta} + \frac{\lambda_\gamma}{\sqrt{n}\, g_n'(x_1, \ldots, x_n; \tilde{\theta})}\right] = \gamma. \tag{12.5.14}$$

Таким образом, $\tilde{\theta} \pm \dfrac{\lambda_\gamma}{\sqrt{n}\, g_n'(x_1, \ldots, x_n; \tilde{\theta})}$ — конечные точки асимптотического 100γ-процентного доверительного интервала для θ_0 при большом n, длина которого равна $\dfrac{2\lambda_\gamma}{\sqrt{n}\, g_n'(x_1, \ldots, x_n; \tilde{\theta})}$.

(c) Асимптотически наикратчайшие доверительные интервалы. Отношение r_n квадрата длины доверительного интервала из (12.5.5) к соответствующей величине из (12.5.14) равно

$$r_n = c_n^2 d_n^2, \tag{12.5.15}$$

где

$$c_n = \frac{g_n'(x_1, \ldots, x_n; \theta_0)}{h_n'(x_1, \ldots, x_n; \hat{\theta})} \tag{12.5.16}$$

и

$$d_n = \frac{g_n'(x_1, \ldots, x_n; \tilde{\theta})}{g_n'(x_1, \ldots, x_n; \theta_0)}. \tag{12.5.17}$$

Здесь c_n и d_n — случайные величины, но из **4.3.7** следует, что числитель и знаменатель d_n^2 сходятся по вероятности к $B^{*2}(\theta_0, \theta_0)$ и, следовательно, d_n^2 сходится по вероятности к 1. То же самое можно сказать и о c_n^2. Отсюда r_n сходится по вероятности к величине

$$\frac{B^{*2}(\theta_0, \theta_0)}{B^2(\theta_0, \theta_0)}, \tag{12.5.18}$$

и мы хотим показать, что эта величина не может быть больше 1.

Так как (12.5.7) имеет место для всех n, учитывая (12.5.8), мы можем дифференцировать (12.5.7) для $j=1$ по θ под знаком интеграла. При $\theta = \theta_0$

$$\int\limits_{R_n} g_n'(x_1, \ldots, x_n; \theta)\, dF_{n0} + \int\limits_{R_n} \sqrt{n}\, g_n(x_1, \ldots, x_n; \theta_0)\left[\frac{1}{\sqrt{n}} S_n(x_1, \ldots, x_n; \theta)\right] dF_{n0} = 0. \tag{12.5.19}$$

Следовательно, квадраты этих интегралов равны. Но квадрат первого интеграла есть $B_n^{*2}(\theta_0, \theta_0)$. Применяя неравенство Шварца ко второму интегралу и учитывая тот факт, что интеграл в (12.5.7) для $j=2$ равен 1, мы получим

$$B_n^{*2}(\theta_0, \theta_0) \leqslant \int\limits_{R_n} \left[\frac{1}{\sqrt{n}} S_n(x_1, \ldots, x_n; \theta_0)\right]^2 dF_{n0}. \tag{12.5.20}$$

Но правая часть сходится к $B^2(\theta_0, \theta_0)$, не зависящей от n. Отсюда

$$B_n^{*2}(\theta_0, \theta_0) \leqslant B^2(\theta_0, \theta_0),$$

и, переходя к пределу при $n \to \infty$, мы имеем

$$\frac{B^{*2}(\theta_0, \theta_0)}{B^2(\theta_0, \theta_0)} \leqslant 1. \tag{12.5.21}$$

Таким образом, отношение квадратов длин доверительных интервалов сходится по вероятности к числу, которое $\leqslant 1$. Мы будем говорить, что доверительные интервалы, определенные в (12.5.5), являются асимптотически при больших n наикратчайшими доверительными интервалами для θ_0.

Полученный результат сформулируем в виде теоремы

12.5.1. *Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности с к.ф.р. $F(x; \theta_0)$, где $F(x; \theta)$ регулярна в смысле первой и второй производных по θ в области Ω_0. Тогда, если $g_n(x_1, \ldots, x_n; \theta)$ — какая-либо регулярная оценивающая функция для θ_0, то асимптотические при больших n 100γ-процентные доверительные пределы для θ_0 даются формулой* (12.5.14). *Кроме того, не существует никакой регулярной оценивающей функции для θ_0, которая давала бы асимптотически более короткий 100γ-процентный доверительный интервал для θ, чем функция правдоподобия $h_n(x_1, \ldots \ldots, x_n; \theta)$, определенная формулой* (12.5.5). *Таким образом, асимптотически наикратчайший доверительный интервал дается формулой* (12.5.5).

Эквивалентность задачи нахождения асимптотически наикратчайших доверительных интервалов и асимптотически эффективной оценки становится очевидной, если вспомнить, что асимптотическая эффективность $\tilde{\theta}$ в (12.5.10), вычисленная в формуле (12.3.23), есть просто отношение (12.5.21). Таким образом, квадратный корень из асимптотической эффективности оценки $\tilde{\theta}$ параметра θ_0 (см. (12.5.10)) равен в пределе (по вероятности) отношению длины асимптотического 100γ-процентного доверительного интервала, построенного по функции правдоподобия $h_n(x_1, \ldots, x_n; \theta)$, к длине доверительного интервала, построенного по $g_n(x_1, \ldots, x_n; \theta)$.

Задача об асимптотически наикратчайших доверительных интервалах для случая, когда $g_n(x_1, \ldots, x_n; \theta)$ имеет вид $\sum_{\xi=1}^{n} g(x_\xi; \theta)$, была рассмотрена Уилксом (1938 b) и в более общем случае Вальдом (1942).

12.6. Многомерное точечное оценивание

(а) Вводные замечания. Результаты, полученные в §§ 12.2—12.5, относятся к выборкам с к. ф. р. $F(x, \theta)$, в которых θ — одномерный параметр. Как указано в начале § 12.1, результаты этих параграфов остаются верными с незначительными изменениями в определениях и тогда, когда мы имеем выборку из k-мерной совокупности с к. ф. р. $F(x_1, \ldots, x_k; \theta)$, где θ — одномерный параметр.

В том случае, если θ — r-мерный параметр с компонентами $\theta_1, \ldots, \theta_r$, основные результаты, которые мы получили, имеют r-мерный вариант, который может быть не сразу очевиден читателю. Мы сформулируем несколько r-мерных результатов, иногда опуская в доказательствах некоторые подробности. Читатель, достаточно хорошо знакомый с уже рассмотренными результатами для одномерного случая, не встретит затруднений в восстановлении этих деталей.

На протяжении этого параграфа мы будем рассматривать простой случайный выбор из совокупности с к. ф. р. $F(x; \theta)$, где θ — r-мерный параметр, x для удобства считается одномерным. Результаты с небольшими изменениями могут быть обобщены на случай многомерного x. Компоненты θ мы будем обозначать через $(\theta_1, \ldots, \theta_r)$. Истинное значение $(\theta_{10}, \ldots, \theta_{r0})$ параметра θ будет обозначаться через θ_0. По аналогии с одномерным интервалом Ω_0 мы будем иметь (открытый) r-мерный прямоугольник Ω_{r0}, содержащий θ_0.

Оценка $(\tilde{\theta}_1, \ldots, \tilde{\theta}_r)$ будет обозначаться через $\tilde{\theta}$. Несмещенной оценкой $\tilde{\theta}$ для θ является такая оценка, для которой $\tilde{\theta}_p$ — несмещенная оценка θ_{p0}, $p = 1, \ldots, r$. Набор достаточных статистик для оценивания θ был определен в § 12.2 (e).

(b) Эффективность многомерной оценки. Теперь мы покажем, как обобщить понятие эффективности на случай многомерных оценок. При этом мы будем пользоваться обозначениями § 12.1 (b). Пусть

$$\begin{aligned} S_{pn}(x_1, \ldots, x_n; \theta) &= \frac{\partial}{\partial\theta_p} \log dF_n(x_1, \ldots, x_n; \theta) = \sum_{\xi=1}^{n} S_p(x_\xi; \theta), \\ S_{pqn}(x_1, \ldots, x_n; \theta) &= \frac{\partial}{\partial\theta_q} S_{pn}(x_1, \ldots, x_n; \theta). \end{aligned} \tag{12.6.1}$$

Если $\tilde{\theta}$ — несмещенная оценка некоторого θ из Ω_{r0} и c'_p, $p = 1, \ldots$ $\ldots, r$, — произвольные константы, то можно написать

$$\int\limits_{R_n} \sum_{p=1}^{r} (\tilde{\theta}_p - \theta_p)\, c'_p dF_n = 0, \tag{12.6.2}$$

что, конечно, также означает, что $\sum_{p=1}^{r} c'_p \tilde{\theta}_p$ — несмещенная оценка для $\sum_{p=1}^{r} c'_p \theta_p$. Будем предполагать, что компоненты $\tilde{\theta}_p$ линейно независимы.

Если $F(x; \theta)$ регулярна вместе со всеми первыми производными по θ, мы можем продифференцировать (12.6.2) по θ_q. Получаем

$$\int\limits_{R_n} \sum_{p=1}^{r} (\tilde{\theta}_p - \theta_p)\, c'_p S_{qn}(x_1, \ldots, x_n; \theta)\, dF_n = c'_q \tag{12.6.3}$$

для $q=1, \ldots, r$. Умножим обе части (12.6.3) на c_q, просуммируем по q, далее возведем в квадрат обе части получившегося равенства и применим неравенство Шварца к квадрату интеграла в левой части.

Таким образом, имеем для θ из Ω_{r0}

$$\Big(\sum_1^r c_q c_q'\Big)^2 \leqslant \Big\{\int_{R_n} \Big[\sum_1^r (\tilde\theta_p - \theta_p) c_p'\Big]^2 dF_n\Big\} \times$$
$$\times \Big\{\int_{R_n} \Big[\sum_1^r S_{qn}(x_1, \ldots, x_n; \theta) c_q\Big]^2 dF_n\Big\}. \quad (12.6.4)$$

Но величина в первых фигурных скобках справа равна $\sigma^2\Big(\sum\limits_1^r c_p'\tilde\theta_p \,|\, \theta\Big)$, а во вторых — $n\sigma^2\Big[\sum\limits_1^r S_p(x, \theta) c_p \,|\, \theta\Big]$. Поэтому при $\theta = \theta_0$ (12.6.4) сводится к неравенству

$$\sigma^2\Big(\sum_1^r c_p'\tilde\theta_p \,|\, \theta_0\Big) \geqslant \frac{\Big(\sum\limits_1^r c_p c_p'\Big)^2}{n\sigma^2\Big[\sum\limits_{p=1}^r S_p(x; \theta) c_p \,|\, \theta_0\Big]}, \quad (12.6.5)$$

что дает точную верхнюю границу дисперсии несмещенной оценки $\sum\limits_1^r c_p'\tilde\theta_p$ для $\sum\limits_1^r c_p'\theta_{p0}$.

Как мы увидим, в (12.6.4) знак равенства имеет место тогда и только тогда, когда для всех точек x пространства R_n, за исключением, может быть, множества вероятностной меры 0, существует линейная зависимость вида

$$K(\theta) \sum_p (\tilde\theta_p - \theta_p) c_p' \equiv \sum_p S_{pn}(x_1, \ldots, x_n; \theta) c_p, \quad (12.6.6)$$

где $K(\theta)$ не зависит от $(x_1, \ldots, x_n)$ и в каждом из наборов чисел $c_1', \ldots, c_r'$ и $c_1, \ldots, c_r$ имеется хотя бы одно отличное от 0. Если существует несмещенная оценка θ_0 такая, что (12.6.6) имеет место для любого θ из Ω_{r0}, то мы будем называть ее эффективной оценкой и обозначать через $\hat\theta$.

Обозначим матрицу ковариаций величин $\sqrt{n}\,\tilde\theta_p$, $p=1, \ldots, r$, в точке $\theta=\theta_0$ через $\|B_{pq}'^{(n)}\|$, матрицу ковариаций величин $S_q/\sqrt{n}$ в точке $\theta=\theta_0$ через $\|B_{pq}\|$, как и в § 12.1 (b). Ковариация $\sqrt{n}\,\tilde\theta_p$ и $S_q/\sqrt{n}$ равна δ_{pq}, где δ_{pq} равно 0, если $p \neq q$, и 1, если $p=q$. Неравенство (12.6.4) мы теперь можем переписать в виде

$$\Big(\sum_{p,q} B_{pq}'^{(n)} c_p' c_q'\Big)\Big(\sum_{p,q} B_{pq} c_p c_q\Big) \geqslant \Big(\sum_{p,q} \delta_{pq} c_p c_q'\Big)^2. \quad (12.6.7)$$

Здесь знак равенства имеет место только в случае, если выполнено (12.6.6), если, конечно, в наборах чисел $c_1', \ldots, c_r'$ и $c_1, \ldots, c_r$ имеется хотя бы по одному отличному от 0.

Теперь знак $>$ в (12.6.7) имеет место только, когда матрица ковариаций случайных величин $\sqrt{n}\,\tilde{\theta}_1, \ldots, \sqrt{n}\,\tilde{\theta}_r$, $S_1/\sqrt{n}, \ldots, S_r/\sqrt{n}$ положительно определена, откуда следует, что

$$|B_{pq}'^{(n)}|\,|B_{pq}| > |\delta_{pq}| = 1. \tag{12.6.8}$$

Знак равенства в (12.6.7) имеет место только тогда, когда матрица ковариаций тех же случайных величин становится положительно полуопределенной после учета линейной зависимости (12.6.8), т. е.

$$|B_{pq}'^{(n)}|\,|B_{pq}| = 1. \tag{12.6.9}$$

Иначе говоря, знак неравенства (равенства) в (12.6.7) имеет место тогда и только тогда, когда имеет место знак неравенства (равенства) в соотношении

$$|B_{pq}'^{(n)}|\,|B_{pq}| \geqslant 1. \tag{12.6.10}$$

Так как компоненты $\tilde{\theta}_1, \ldots, \tilde{\theta}_r$ линейно независимы, из теоремы **3.5.2** следует, что матрица $\|B_{pq}'^{(n)}\|$ положительно определена и, следовательно, $|B_{pq}'^{(n)}| > 0$. По аналогичным причинам $|B_{pq}| > 0$.

Эти факты наводят на мысль, что, как обобщение (12.2.15), мы можем определить эффективность несмещенной оценки как отношение

$$\operatorname{eff}(\tilde{\theta} \mid \theta_0) = \frac{1}{|B_{pq}'^{(n)}|\,|B_{pq}|}. \tag{12.6.11}$$

Таким образом, $\operatorname{eff}(\tilde{\theta} \mid \theta_0) = 1$ для эффективной оценки $\hat{\theta}$.

Сформулируем полученный результат в виде теоремы

12.6.1. *Предположим, что* $(x_1, \ldots, x_n)$ — *выборка из совокупности с к. ф. р.* $F(x;\ \theta_0)$, *где* θ_0 — r*-мерный параметр. Пусть функция* $F(x, \theta)$ *регулярна в смысле первой и второй производных по* θ *в* Ω_{r0}, $\|B_{pq}(\theta, \theta)\|$ — *положительно определенная матрица при всех* θ *из* Ω_{r0}. *Если* $\tilde{\theta}$ — *несмещенная оценка для* θ_0 *с линейно независимыми компонентами, матрица ковариаций которой* $\|B_{pq}'^{(n)}\|$ *существует и положительно определена, то для любых двух наборов констант* $c_1, \ldots, c_r$ *и* $c_1', \ldots, c_r'$, *содержащих хотя бы по одному ненулевому элементу, имеют место неравенства* (12.6.5) *и* (12.6.10). *Кроме того, в обоих случаях знак равенства имеет место тогда и только тогда, когда выполнено* (12.6.6).

12.7. Многомерное точечное оценивание по большим выборкам

(а) Асимптотическое распределение информанта. Информант r-мерного вектора θ представляет собой по определению (12.6.1) r-мерный вектор с компонентами $S_{pn}(x_1, \ldots, x_n;\ \theta)$, $p = 1, \ldots, r$, При условиях, аналогичных приведенным в теореме **12.3.1**, при большом n эти компоненты подчиняются r-мерному нормальному закону. Именно

12.7.1. *Предположим, что* $(x_1, \ldots, x_n)$ — *выборка из совокупности с функцией распределения* $F(x, \theta_0)$, *где* θ_0 — r*-мерный параметр. Пусть* $F(x; \theta)$ *регулярна в смысле первых производных по* θ *в* Ω_{r0}. *Тогда, если матрица* $\| B'_{pq} \|, p, q = 1, 2, \ldots, r$, *положительно определена при всех* θ *из* Ω_{r0}, *вектор* $(S_{pn}(x_1, \ldots, x_n; \theta_p), p = 1, \ldots, r)$ *асимптотически при больших* n *распределен в соответствии с* n*-мерным распределением* $N(\{0\}, \| nB_{pq} \|)$, *где* $B_{pq} = B_{pq}(\theta_0, \theta_0)$.

Доказательство этого утверждения аналогично доказательству теоремы **12.3.1** для одномерного случая, и мы предоставляем читателю проделать его в качестве упражнения.

(b) Сходимость оценки максимального правдоподобия. Как и в одномерном случае, если не существует несмещенной оценки $\tilde{\theta}$ для θ_0, компоненты которой удовлетворяли бы (12.6.6) при каждом n, можно показать, что при некоторых условиях существует последовательность решений системы уравнений

$$S_p(x_1, \ldots, x_n; \theta) = 0, \; p = 1, \ldots, r, \tag{12.7.1}$$

которая сходится к θ_0 с вероятностью 1. Более точно, мы имеем следующее обобщение теоремы **12.3.2** на n-мерный случай:

12.7.2. *Предположим, что* $(x_1, \ldots, x_n)$ — *выборка из распределения с к. ф. р.* $F(x; \theta_0)$, *где* $\theta_0 = (\theta_{10}, \ldots, \theta_{r0})$ — r*-мерный параметр и функция* $F(x; \theta)$ *регулярна в смысле первых производных по* θ *в* Ω_{r0}. *Пусть* $S_p(x; \theta)$, $p = 1, \ldots, r$, — *непрерывные функции* θ *в области* Ω_{r0} *при всех значениях* x, *за исключением, может быть, множества вероятностной меры* 0. *Тогда существует последовательность решений системы* (12.7.1), *которая сходится почти наверное к* $(\theta_{10}, \ldots, \theta_{r0})$. *Если для* $n \geqslant$ *некоторого* n_0 *решение представляет собой единственный вектор* $(\hat{\theta}_1, \ldots, \hat{\theta}_r)$, *то последовательность этих векторов сходится почти наверное к* $(\theta_{10}, \ldots, \theta_{r0})$ *при* $n \to \infty$.

Доказательство **12.7.2** полностью аналогично **12.3.2**, и мы его опускаем.

(c) Асимптотическое распределение оценки максимального правдоподобия. Сформулируем r-мерный аналог теоремы **12.3.3**.

12.7.3. *Если* $(x_1, \ldots, x_n)$ — *выборка из совокупности с к. ф. р.* $F(x; \theta_0)$, *где* θ_0 — r*-мерный параметр,* $F(x; \theta)$ *регулярна в смысле первых и вторых смешанных производных по* θ *для* θ *из* Ω_{r0} *и оценка максимального правдоподобия* $(\hat{\theta}_1, \ldots, \hat{\theta}_r)$, *удовлетворяющая* (12.7.1), *единственна для* $n \geqslant$ *некоторого* n_0 *и является измеримой относительно вероятностной меры* $\prod_{\xi=1}^{n} F(x_\xi; \theta)$, *то асимп-*

тотически при больших n она имеет r-мерное нормальное распределение $N(\{\theta_{p0}\}, \|nB_{pq}\|^{-1})$, где $\|B_{pq}\| = \|B_{pq}(\theta_0, \theta_0)\|$.

Доказательство аналогично **12.3.3**, и мы его опускаем.

(d) Асимптотическая эффективность оценки максимального правдоподобия. При r-мерном оценивании $\tilde\theta$ является состоятельной оценкой для θ_0, если каждая компонента $\tilde\theta$ является состоятельной оценкой соответствующей компоненты θ_0.

Асимптотическая эффективность состоятельной оценки $\tilde\theta$ для θ_0 определяется как

$$\operatorname{leff}(\tilde\theta\,|\,\theta_0) = \frac{1}{\lim\limits_{n\to\infty} |B_{pq}'^{(n)}|\,|B_{pq}|}. \qquad (12.7.2)$$

Можно проверить, что если оценка $\hat\theta$ асимптотически нормальна, как установлено в **12.7.3**, то $\hat\theta$ является состоятельной оценкой для θ_0. Линейная функция $\sum\limits_p c_p'\hat\theta_p$ является состоятельной оценкой для $\sum\limits_p c_p'\theta_{p0}$ и

$$\lim_{n\to\infty} \sigma^2\Big(\sqrt{n}\sum_p c_p'\hat\theta_p\Big) = \sum_{p,q} B^{pq}c_p'c_q', \qquad (12.7.3)$$

где $\|B^{pq}\| = \|B_{pq}\|^{-1}$. Таким образом, если $\|B_{pq}'^{(n)}\|$ — матрица ковариаций вектора $(\sqrt{n}\hat\theta_1, \ldots, \sqrt{n}\hat\theta_r)$, то

$$\lim_{n\to\infty} |B_{pq}'^{(n)}| = |B^{pq}| \qquad (12.7.4)$$

и, следовательно,

$$\operatorname{leff}(\hat\theta\,|\,\theta_0) = \frac{1}{|B^{pq}|\,|B_{pq}|} = 1. \qquad (12.7.5)$$

12.7.4. *При условиях теоремы* **12.7.3** *оценка максимального правдоподобия $\hat\theta$ является асимптотически эффективной оценкой для θ_0.*

Как и в одномерном случае, первые исследования по эффективному оцениванию многомерного параметра принадлежат Фишеру (1922). Более поздние работы в этой области, использующие современные математические методы, принадлежат Баранкину и Гёрланду (1951).

12.8. Многомерные доверительные области

Идея, лежащая в основе интервального оценивания одномерного параметра (см. § 12.4), состоит в том, что при фиксированном доверительном коэффициенте γ ищутся такие случайные величины $\underline{\theta} < \overline{\theta}$, что

$$P(\underline{\theta} < \theta < \overline{\theta}\,|\,\theta) = \gamma.$$

Метод построения $\underline{\theta}$ и $\bar{\theta}$ при некоторых условиях дается теоремой **12.4.1**. Как мы вскользь заметили в § 12.4 (b), если в теореме **12.4.1** откинуть требование монотонности функции $g_n(x_1, \ldots, x_n; \theta)$ по θ на интервале Ω_0, то мы получим случайное множество $E_1(x_1, \ldots, x_n)$ из R_1, которое коротко может быть записано как E_1, состоящее из всех вещественных y, для которых

$$g_1 < g_n(x_1, \ldots, x_n; y) < g_2, \tag{12.8.1}$$

и не зависящее от θ_0, так что

$$P(g_1 < g_n(x_1, \ldots, x_n; \theta_0) < g_2 \mid \theta_0) = P(\theta_0 \in E_1 \mid \theta_0) = \gamma. \tag{12.8.2}$$

Иными словами, вероятность того, что E_1 содержит θ_0, равна γ. E_1 — борелевское множество, ибо $g_n(x_1, \ldots, x_n; \theta)$ — случайная величина для всех значений θ из Ω. Независимость E_1 от θ следует из того, что к. ф. р. $g_n(x_1, \ldots, x_n; \theta)$ не зависит от θ. Надо заметить, что E_1 не обязательно состоит только из точек параметрического пространства Ω; E_1 состоит из точек эвклидова пространства R_1, в которое вложено Ω. Вот почему мы предлагаем в (12.8.1) писать y вместо θ.

Ввиду сказанного (12.8.2) представляет собой расширение идеи интервального оценивания. С помощью такого подхода мы можем обобщить основной результат, касающийся интервального оценивания, на случай r-мерного параметра следующим образом.

12.8.1. *Предположим, что $(x_1, \ldots, x_n)$ — выборка из совокупности с к. ф. р. $F(x; \theta_0)$, где θ_0 — r-мерный параметр. Пусть $g_n(x_1, \ldots x_n; \theta)$ — случайная величина, определенная таким образом, что при всех θ из параметрического пространства Ω и в каждой выборочной точке пространства R_n, за исключением, может быть, множества вероятностной меры 0, ее к. ф. р. не зависит от θ, если $(x_1, \ldots, x_n)$ имеют в качестве к. ф. р. $F(x; \theta)$. Пусть $E_r(x_1, \ldots, x_n) = E_r$ — множество точек $y = (y_1, \ldots, y_r)$ из R_r, для которых $g_1 < g_n(x_1, \ldots, x_n; y) < g_2$, где g_1 и g_2 выбраны так, что*

$$P(g_1 < g_n(x_1, \ldots, x_n; \theta) < g_2 \mid \theta) = \gamma, \tag{12.8.3}$$

где γ не зависит от θ. Тогда E_r — r-мерное случайное множество такое, что для истинной параметрической точки θ_0 мы имеем

$$P(\theta_0 \in E_r \mid \theta_0) = \gamma. \tag{12.8.4}$$

Доказательство этого утверждения очевидно и опускается. Мы назовем множество E_r 100γ-процентной доверительной областью для θ_0. Пример доверительной области для r-мерного параметра был дан в § 10.4.

Более общо, если $g_n(x_1, \ldots, x_n; \theta)$ — *вектор-функция* с компонентами $g_{p'n}(x_1, \ldots, x_n; \theta)$, $p' = 1, \ldots, r'$, к. ф. р. которой не зависит от θ, когда популяционная к. ф. р. есть $F(x, \theta)$, то для любого множества $E^*_{r'}$ из $R_{r'}$, для которого

$$P((g_{1n}, \ldots, g_{r'n}) \in E^*_{r'}) = \gamma, \tag{12.8.5}$$

множество E_r, состоящее из всех точек $y = (y_1, \ldots, y_r)$ из R_r, для которых

$$(g_{1n}(x_1, \ldots, x_n; y), \ldots, g_{r'n}(x_1, \ldots, x_n; y)) \in E^*_{r'}, \tag{12.8.6}$$

представляет собой 100γ-процентную доверительную область для θ_0 т. е.

$$P(\theta_0 \in E_r) = \gamma. \tag{12.8.7}$$

Обычно наиболее интересными, полезными и изящными являются функции $g_n(x_1, \ldots, x_n; \theta)$, которые дают выпуклые или по крайней мере просто связные доверительные области (случайные множества) E_r, ибо такие множества, по-видимому, легче описывать и использовать. При $r = r' = 1$ E_1 будет таковым, если $g_n(x_1, \ldots, x_n; \theta)$ непрерывна и монотонна по θ, как показано в **12.4.1**. Для более общих значений r и r' ситуация более сложная.

Но в дополнение к требованию выпуклости или связности наша интуиция требует, чтобы область E_r была в каком-то смысле наименьшей. При конечном n это, однако, сложная проблема. Для больших n существует удовлетворительное асимптотическое решение ее при некоторых ограничениях, которые мы рассмотрим далее.

Пример. Для иллюстрации рассмотренных идей предположим, что $(x_1, \ldots, x_n)$ — выборка из $N(\mu_0, \sigma_0^2)$. Тогда из **8.4.1** мы знаем, что функция

$$g(x_1, \ldots, x_n; \mu_0, \sigma_0) = \frac{(n-1)s^2 + n(\bar{x} - \mu_0)^2}{\sigma_0^2}$$

имеет χ^2-распределение $C(n)$. Поэтому для данного γ мы можем найти χ^2_γ такое, что

$$P(g < \chi^2_\gamma) = \gamma.$$

Обозначим через E_2 область в $(y_1 y_2)$-полупространстве $y_2 > 0$, лежащую между двумя ветвями гиперболы

$$\frac{y_2^2}{n} - \frac{(y_1 - \bar{x})^2}{\chi^2_\gamma} = \frac{(n-1)s^2}{n\chi^2_\gamma}.$$

Мы видим, что если (μ_0, σ_0) — истинная параметрическая точка, то

$$P((\mu_0, \sigma_0) \in E_2) = \gamma.$$

Теперь рассмотрим двумерную вектор-функцию

$$g_{p'}(x_1, \ldots, x_n; \mu_0, \sigma_0), \quad p' = 1, 2,$$

где

$$g_1 = \frac{n(\bar{x} - \mu_0)^2}{\sigma_0^2}, \quad g_2 = \frac{n-1}{\sigma_0^2} s^2.$$

Из теоремы **8.4.2** следует, что g_1 и g_2 независимы и имеют χ^2-распределения $C(1)$ и $C(n-1)$ соответственно. Если мы выберем χ_1^2 и χ_2^2 так, что

$$P(g_1 < \chi_1^2, \quad g_2 > \chi_2^2) = \gamma,$$

и определим E_2^* как область в $(y_1 y_2)$-плоскости, для которой

$$\frac{n(y_1 - \bar{x})^2}{y_2} < \chi_1^2, \qquad \frac{\sqrt{(n-1)s}}{y_2} > \chi_2,$$

где $y_2 > 0$, то будем иметь

$$P((\mu_0, \sigma_0) \in E_2^*) = \gamma.$$

Таким образом, E_2 и E_2^* — примеры доверительных областей, определенных оценивающими вектор-функциями, имеющими одну и две компоненты соответственно. Интуитивно ясно, что E_2^* является более удовлетворительной, чем E_2, ибо E_2^* ограничено и выпукло, в то время как E_2 не обладает ни одним из этих свойств. Если в качестве $g(x_1, \ldots, x_n; \mu_0, \sigma_0)$ выбрать плотность распределения вектора $(\bar{x}, s^2)$, то мы получим область E_2^{**}, которая еще лучше, чем E_2^*, с точки зрения критерия «малости», но E_2^{**} имеет более сложный вид, чем E_2^*, и мы не будем заниматься ее детальным описанием. При большом n существует оптимальный метод получения асимптотически наименьшей области для оценивания (μ_0, σ_0). Общая задача об асимптотически наименьших доверительных областях рассматривается нами в следующем параграфе.

12.9. Асимптотически наименьшие доверительные области при больших выборках

Теперь мы переходим к рассмотрению r-мерного обобщения теоремы **12.5.1**.

Из теорем **12.7.1** и **9.3.2** следует, что если истинное значение равно θ_0, то последовательность случайных величин

$$U_n(x_1, \ldots, x_n; \theta_0) = n \sum_{p, q=1}^{r} B^{pq} h_{pn}(\theta_0) h_{qn}(\theta_0), \qquad n = 1, 2, \ldots, \tag{12.9.1}$$

где $h_{pn}(\theta) = \frac{1}{n} S_{pn}(x_1, \ldots, x_n; \theta)$, как указано в (12.6.1), сходится по распределению к χ^2-распределению $C(r)$. При более сильных условиях теоремы **12.7.3** мы можем применить теорему о среднем значении из дифференциального исчисления к U_n и написать

$$U_n^* \equiv U_n, \text{ если значение } \hat{\theta} \text{ принадлежит } \Omega_{r0},$$

и

$$U_n^* \equiv 0, \quad \text{если значение } \hat{\theta} \text{ не принадлежит } \Omega_{r0},$$

где

$$U_n^*(x_1, \ldots, x_n; \theta_0) = \sum_{p, q} B_{pq}^{(n)} [\sqrt{n}(\theta_{p0} - \hat{\theta}_p)][\sqrt{n}(\theta_{q0} - \hat{\theta}_q)] \tag{12.9.2}$$

для всех точек пространства R_n, исключая, может быть, множество вероятностной меры 0, где

$$B_{pq}^{(n)} = \left[\sum_{p',\, q'=1}^{r} B^{p'q'} \left(\frac{\partial h_{p'n}}{\partial \theta_p} \right) \left(\frac{\partial h_{q'n}}{\partial \theta_q} \right) \right]_{\theta = \hat{\theta}^*} \tag{12.9.3}$$

и

$$| \theta_{p0} - \hat{\theta}_p^* | < | \theta_{p0} - \hat{\theta}_p |, \qquad p = 1, \ldots, r.$$

Можно проверить, что из условий **12.7.3** и теорем **4.3.7** и **4.3.8** следует, что $B_{pq}^{(n)}$ сходится по вероятности к

$$\sum_{p',\, q'=1}^{r} B^{p'q'} B_{p'p} B_{q'q} \; (= B_{pq}) \quad \text{при } n \to \infty$$

и, следовательно, U_n и U_n^*, $n = 1, 2, \ldots,$ — последовательности случайных величин, сходящиеся одновременно по распределению к $C(r)$. Теперь выберем χ_γ^2 так, чтобы

$$\gamma = \lim_{n \to \infty} P(U_n < \chi_\gamma^2) = \lim_{n \to \infty} P(U_n^* < \chi_\gamma^2), \tag{12.9.4}$$

и пусть $E_r(x_1, \ldots, x_n)$ и $E_r^*(x_1, \ldots, x_n)$ — множества точек $y = (y_1, \ldots, y_r)$ из R_r такие, что

$$U_n(x_1, \ldots, x_n; y) < \chi_\gamma^2 \tag{12.9.5}$$

и

$$U_n^*(x_1, \ldots, x_n; y) < \chi_\gamma^2 \tag{12.9.6}$$

соответственно. Поэтому, если θ_0 — истинное значение θ, то

$$\lim_{n \to \infty} P(\theta_0 \in E_r) = \lim_{n \to \infty} P(\theta_0 \in E_r^*) = \gamma, \tag{12.9.7}$$

т. е. E_r и E_r^* — асимптотически эквивалентные доверительные области для оценки θ_0. Но E_r^* представляет собой эллипсоид в R_r с центром в точке θ_0, объем которого равен

$$\frac{K(r)\, \chi_\gamma}{\sqrt{n\, | B_{pq}^{(n)} |}}, \tag{12.9.8}$$

где $K(r)$ — константа, зависящая только от r. Теперь пусть $g_n(x_1, \ldots, x_n; \theta)$ — векторная случайная величина с линейно независимыми компонентами $g_{pn}(x_1, \ldots, x_n; \theta)$, $p = 1, 2, \ldots, r$, и первыми производными

$$g_{pqn}(x_1, \ldots, x_n; \theta) = \frac{\partial}{\partial \theta_q} g_{pn}(x_1, \ldots, x_n; \theta), \quad p, q = 1, \ldots, r. \tag{12.9.9}$$

Обозначим $g_{pn}(x_1, \ldots, x_n; \theta)$ и $g_{pqn}(x_1, \ldots, x_n; \theta)$ через $g_p(\theta)$ и $g_{pq}(\theta)$. Предположим, что $g_p(\theta)$, $p = 1, \ldots, r$, и $g_{pq}(\theta)$, $p, q = 1, 2, \ldots, r$, удовлетворяют условиям (iii) — (v) из (12.5.7) для $g_n(x_1, \ldots, x_n; \theta)$ и $g_n'(x_1, \ldots, x_n; \theta)$. Обозначим через $A_{pn}^*(\theta_0, \theta)$,

$p=1, 2, \ldots, r$, и $B^*_{pqn}(\theta_0, \theta)$, $p, q=1, 2, \ldots, r$, функции, соответствующие A^*_n и B^*_n. При надлежащей замене имеет место (12.5.8). Пусть $B^*_{pq}(\theta, \theta)=\lim\limits_{n\to\infty} B^*_{pqn}(\theta, \theta)$ и $B^*_{pq}=B^*_{pq}(\theta_0, \theta)$. Вместо (i) в условиях (12.5.7) мы будем иметь для $j=1$

$$\int\limits_{R_n} \sqrt{n}\, g_p(\theta)\, dF_n=0 \qquad (12.9.10)$$

и для $j=2$

$$\int\limits_{R_n} [\sqrt{n}\, g_p(\theta)]\, [\sqrt{n}\, g_q(\theta)]\, dF_n=C_{pqn}(\theta). \qquad (12.9.11)$$

В соответствии с (12.5.7) (ii) предположим, что если $\theta=\theta_0$ в распределении популяции, то $(\sqrt{n}\, g_p(\theta_0),\ p=1, \ldots, r)$ имеет асимптотическое распределение $N(\{0\}, \|C_{pq}\|)$, где

$$C_{pq}=\lim_{n\to\infty} C_{pqn}(\theta_0), \qquad (12.9.12)$$

и $\|C_{pq}\|$ — положительно определенная матрица.

Вектор-функция $g(x_1, \ldots, x_n; \theta)$, удовлетворяющая условиям предыдущего абзаца, будет называться регулярной оценивающей вектор-функцией для θ_0. Читатель может заметить, что вектор-функция правдоподобия $h_p(\theta)=(1/n)\, S_{pn}(x_1, \ldots, x_n; \theta)$ является регулярной оценивающей вектор-функцией. Для такой оценивающей функции существует последовательность единственных вектор-решений $(\tilde\theta_1, \ldots, \tilde\theta_r)$ для $n\geqslant$ некоторого n_0 уравнений

$$g_{pn}(x_1, \ldots, x_n; \theta)=0,\ p=1, \ldots, r,\ n=n_0, n_0+1, \ldots, \qquad (12.9.13)$$

и если θ_0 — истинное значение θ, то последовательность случайных величин $[\sqrt{n}(\tilde\theta_1-\theta_{10}), \ldots, \sqrt{n}(\tilde\theta_r-\theta_{r0})]$, $n=n_0$, $n_0+1, \ldots$ имеет в пределе r-мерное распределение $N(\{0\}, \|D^{pq}\|)$, где

$$D_{pq}=\sum_{p', q'=1}^{r} C^{p'q'} B^*_{p'p} B^*_{q'q}. \qquad (12.9.14)$$

Из теоремы **9.3.2** следует, что если θ_0 — истинное значение θ, то $V_n(x_1, \ldots, x_n; \theta_0)$, определенное как

$$V_n=n \sum_{p, q=1}^{r} C^{pq} g_p(\theta_0)\, g_q(\theta_0), \qquad (12.9.15)$$

имеет в качестве предельного распределения при $n\to\infty$ χ^2-распределение $C(r)$.

Кроме того, тем же путем, каким мы доказали, что если θ_0 — истинное значение θ, то предельное распределение каждой из случайных величин U^*_n и U_n при $n\to\infty$ есть $C(r)$, можно показать, что V_n и связанная с ней случайная величина $V_n^*(x_1, \ldots, x_n; \theta)$, определяемая как

$$V_n^* = \sum_{p,q=1}^{r} D_{pq}^{(n)} [\sqrt{n}(\tilde{\theta}_p - \theta_{p0})] [\sqrt{n}(\tilde{\theta}_q - \theta_{q0})], \tag{12.9.16}$$

где

$$D_{pq}^{(n)} = \sum_{p',q'=1}^{r} C^{p'q'} g_{p'p}(\tilde{\theta}^*) g_{q'q}(\tilde{\theta}^*) \tag{12.9.17}$$

и

$$|\theta_{p0} - \tilde{\theta}_p^*| < |\theta_{p0} - \tilde{\theta}_p|, \qquad p = 1, \ldots, r,$$

сходятся по распределению к $C(r)$ при $n \to \infty$.

Множества $E_r'(x_1, \ldots, x_n)$ и $E_r^{*\prime}(x_1, \ldots, x_n)$, т. е. множества точек $y = (y_1, \ldots, y_r)$ из R_r, для которых

$$V_n(x_1, \ldots, x_n, y) < \chi_\gamma^2 \tag{12.9.18}$$

и

$$V_n^*(x_1, \ldots, x_n, y) < \chi_\gamma^2 \tag{12.9.19}$$

соответственно, где значение χ_γ^2 выбрано так же, как и в (12.9.4), являются асимптотически эквивалентными множествами R_r, т. е. пределы $P(E_r')$ и $P(E_r^{*\prime})$ при $n \to \infty$ равны. Но $E_r^{*\prime}$ является эллипсоидом с центром θ_0 с объемом

$$\frac{K(r)\chi_\gamma}{\sqrt{n\,|D_{pq}^{(n)}|}}. \tag{12.9.20}$$

Теперь отношение r_n объема E_r^*, данного в (12.9.8), к объему $E_r^{*\prime}$ из (12.9.20) сходится (по вероятности) при $n \to \infty$ к числу

$$r = \frac{\sqrt{|D_{pq}|}}{\sqrt{B_{pq}}}, \tag{12.9.21}$$

но, подставляя выражение для D_{pq}, данное в (12.9.14), мы находим

$$r = \frac{|B_{pq}^*|}{\sqrt{|C_{pq}|\,|B_{pq}|}}. \tag{12.9.22}$$

Предположения, которые мы сделали относительно $g_p(\theta)$, позволяют нам дифференцировать (12.9.10) по θ_q под знаком интеграла. Имеем

$$\int_{R_n} g_{pq}(\theta)\, dF_n + \int_{R_n} g_p(\theta) S_q(\theta)\, dF_n = 0, \tag{12.9.23}$$

где $S_q(\theta)$ обозначает $S_{qn}(x_1, \ldots, x_n; \theta)$.

Из (12.9.23) очевидно, что при $\theta = \theta_0$

$$\int_{R_n} g_{pq}(\theta_0)\, dF_n = -\operatorname{cov}\left(\sqrt{n}\, g_p(\theta_0), \frac{S_q(\theta_0)}{\sqrt{n}}\right), \tag{12.9.24}$$

но левая часть (12.9.24) равна $B_{pqn}^*(\theta_0, \theta_0)$. Поэтому правая часть также равна $B^*_{pqn}(\theta_0, \theta_0)$ (или коротко B^*_{pqn}), и последовательность $2r$-мерных случайных величин

$$\left(\sqrt{n}\, g_1(\theta_0), \ \ldots, \ \sqrt{n}\, g_r(\theta_0), \ \frac{1}{\sqrt{n}} S_1(\theta_0), \ldots, \ \frac{1}{\sqrt{n}} S_r(\theta_0)\right),$$
$$n = 1, 2, \ldots,$$

сходится по вероятности к $2r$-мерной случайной величине, имеющей матрицу ковариаций

$$\left\| \begin{matrix} C_{pq} & B^*_{pq} \\ B^*_{pq} & B_{pq} \end{matrix} \right\|. \tag{12.9.25}$$

Для матрицы ковариаций (12.9.25) мы имеем

$$|C_{pq}|\,|B_{pq}| \geqslant |B^*_{pq}|^2. \tag{12.9.26}$$

Поэтому предел (по вероятности) отношения объемов доверительных областей, данный в (12.9.22), не может превосходить 1.

Так как E_r и E'_r асимптотически эквивалентны соответственно E^*_r и E^{**}_r, то не существует регулярной оценивающей вектор-функции $g(x_1, \ldots, x_n; \theta)$, для которой 100γ-процентная доверительная область для θ_0, обозначенная в (12.9.18) через $E'_r(x_1, \ldots, x_n)$, асимптотически меньше, чем 100γ-процентная доверительная область для θ_0, полученная аналогично заменой $g_p(\theta)$ на $h_p(\theta)$.

Подведем итог предшествующему обсуждению и выводам в форме следующего r-мерного обобщения теоремы **12.5.1.**

12.9.1. *Предположим, что $(x_1, \ldots, x_n)$ — выборка из совокупности с к. ф. р. $F(x; \theta_0)$, где θ — r-мерный параметр и $F(x, \theta)$ регулярна в смысле всех первых и вторых производных по θ в Ω_{r0}. Тогда, если $g_{pn}(x_1, \ldots, x_n; \theta)$, $p = 1, 2, \ldots, r$, — какая-либо регулярная вектор-функция для θ_0, то асимптотическая 100γ-процентная доверительная область для θ_0 представляет собой множество E'_r, определенное точками $y = (y_1, \ldots, y_r)$, удовлетворяющими* (12.9.18). *Кроме того, не существует регулярной оценивающей вектор-функции, которая давала бы асимптотически меньшую 100γ-процентную доверительную область для θ_0, чем область, полученная при помощи вектор-функции правдоподобия $h_{pn}(x_1, \ldots, x_n; \theta)$, $p = 1, \ldots, r$, определенной в* (12.9.1). *Таким образом, асимптотически наименьшая 100γ-процентная доверительная область E_r состоит из точек $y = (y_1, \ldots, y_r)$, удовлетворяющих* (12.9.5).

Задача об асимптотически наименьших доверительных областях в случае многомерных параметров рассматривалась Бартлеттом (1953), Бил (1960) и Уилксом и Дэйли (1939).

Читателю, заинтересованному в более детальном рассмотрении, следует обратиться к этим статьям.

Пример. В качестве примера получения асимптотически наименьшей 100γ-процентной доверительной области рассмотрим выборку из мультиномиальной популяции с плотностью, в которую входит k-мерный параметр:

$$dF(s; \theta) = \theta_1^{s_1} \ldots \theta_{k+1}^{s_{k+1}},$$

где случайная величина $(s_1, \ldots, s_{k+1})$ может принимать одно и только одно из $k+1$ значений $(1, 0 \ldots 0), \ldots, (0, 0 \ldots 1)$, где $\theta_1 > 0, \ldots, \theta_{k+1} > 0$ и $\theta_1 + \ldots + \theta_{k+1} = 1$. Для удобства мы можем оценивать при помощи доверительной области k параметров $\theta_1, \ldots, \theta_k$. Предположим, что $(s_{1\xi}, \ldots, s_{k+1\xi}, \xi = 1, \ldots, n)$ — выборка из этого распределения. Для $h_p(\theta)$ мы имеем

$$h_p(\theta) = \frac{1}{n}\sum_{\xi=1}^{n} \frac{\partial \log dF(s_\xi;\ \theta)}{\partial \theta_p} = \frac{1}{n}\sum_{\xi=1}^{n}\left(\frac{s_{p\xi}}{\theta_p} - \frac{s_{k+1\xi}}{\theta_{k+1}}\right),$$

где $\theta_{k+1} = 1 - \theta_1 - \ldots - \theta_k$. Полагая $\sum_{\xi=1}^{n} s_{p\xi} = x_p$, мы можем написать

$$h_p(\theta) = \frac{x_p}{n\theta_p} - \frac{x_{k+1}}{n\theta_{k+1}}, \qquad p = 1, \ldots, k,$$

где $x_{k+1} = n - x_1 - \ldots - x_k$. Кроме того,

$$B_{pq} = \mathscr{E}\left[\left(\frac{\partial \log dF(s;\ \theta)}{\partial \theta_p}\right)\left(\frac{\partial \log dF(s;\ \theta)}{\partial \theta_q}\right)\right] =$$

$$= \mathscr{E}\left[\left(\frac{s_p}{\theta_p} - \frac{s_{k+1}}{\theta_{k+1}}\right)\left(\frac{s_q}{\theta_q} - \frac{s_{k+1}}{\theta_{k+1}}\right)\right] = \frac{\delta_{pq}}{\theta_p} + \frac{1}{\theta_{k+1}}, \qquad p, q = 1, \ldots, k,$$

де δ_{pq} — символ Кронекера, т. е. δ_{pq} равно 1, если $p = q$, и $\delta_{pq} = 0$, если $p \neq q$. Легко видеть, что элемент B^{pq} обратной матрицы равен $B^{pq} = \delta_{pq}\theta_p - \theta_p\theta_q$. Поэтому квадратичная форма U_n в (12.9.1) после некоторых алгебраических преобразований приводится к виду

$$U_n = n \sum_{p,\ q=1}^{k} B^{pq} h_p(\theta)\, h_q(\theta) = \sum_{p=1}^{k+1} \frac{(x_p - n\theta_p)^2}{n\theta_p},$$

что является классической χ^2-функцией, предложенной К. Пирсоном (1900). Таким образом, при больших n 100γ-процентный асимптотически наименьший доверительный интервал для оценивания k-мерного параметра $\theta = (\theta_1, \ldots, \theta_k)$ представляет собой множество E_k вещественных положительных векторов $x = (y_1, \ldots, y_k)$, для которых

$$\sum_{p=1}^{k+1} \frac{(x_p - ny_p)^2}{ny_p} < \chi^2_\gamma,$$

где $y_{k+1} = 1 - y_1 - \ldots - y_k$ и χ^2_γ — 100γ-процентная точка χ^2-распределения $C(k)$.

ЗАДАЧИ

12.1. Показать, что если $\theta' \neq \theta$, то два распределения, имеющие плотности

$$f_1(x;\ \theta) = \begin{cases} \dfrac{1}{\theta}, & 0 < x \leqslant \theta, \\ 0 & \text{в противном случае}, \end{cases}$$

$$f_2(x;\ \theta') = \begin{cases} \dfrac{1}{\theta'}, & 0 < x \leqslant \theta', \\ 0 & \text{в противном случае}, \end{cases}$$

не являются абсолютно непрерывными относительно друг друга.

12.2. Пусть x — случайная величина, имеющая к. ф. р.

$$F(x;\ \theta)=\begin{cases} 0 & , \quad x\leqslant 0, \\ 1-e^{-\theta x}, & x>0, \end{cases}$$

где θ — любое число из $(\theta, +\infty)$. Показать, что $\mathscr{E}(S(x;\ \theta))=0$, и найти $\sigma^2(S(x;\ \theta))$. Определить функцию $H(\theta, \theta')$ и показать, что при фиксированном θ она имеет максимум относительно θ' при $\theta'=\theta$.

12.3. Описать случайную величину с параметрической плотностью

$$p(x;\ \theta)=(1-\theta)\,\theta^{x-1}, \qquad x=1,\ 2,\ \dots,$$

где θ — любое число из $(0,1)$. Показать, что $\mathscr{E}(S(x;\ \theta))=0$, и найти $\sigma^2(S(x;\ \theta))$.

12.4. Если $(x_1, \dots, x_n)$ — выборка из распределения, имеющего k моментов $\mathscr{E}(x^k)$, то $\frac{1}{n}\sum\limits_{\xi=1}^{n} x_\xi^k$ является состоятельной оценкой для $\mathscr{E}(x^k)$. Показать.

12.5. Если $\bar{x}$ — среднее выборки из биномиального распределения $\mathrm{Bi}(1, p)$, то $\bar{x}$ — достаточная оценка параметра p. Показать.

12.6. Предположим, что $(x_1, \dots, x_n)$ — выборка из совокупности с плотностью

$$\theta e^{-\theta x}, \quad x>0,$$

где θ — положительный параметр. Показать, что выборочное среднее $\bar{x}$ — достаточная оценка для θ и что $(n-1)\,\bar{x}/n$ — единственная несмещенная оценка θ, зависящая от $\bar{x}$.

12.7. (*Продолжение*). Показать, что $1/\bar{x}$ является оценкой максимального правдоподобия для θ и имеет бóльшую дисперсию, чем несмещенная оценка $(n-1)\,\bar{x}/n$.

12.8. Показать, что если $\bar{x}$ и s^2 — среднее и дисперсия выборки объема $n>1$ из нормального распределения $N(\mu, \sigma^2)$, то двумерная случайная величина $(\bar{x}, s^2)$ является достаточной для оценивания вектора параметров (μ, σ^2).

12.9. Показать, что если $\hat{\theta}_1$ и $\hat{\theta}_2$ — две достаточные оценки для параметра θ, то коэффициент корреляции между $\hat{\theta}_1$ и $\hat{\theta}_2$ равен 1.

12.10. Показать, что если $\hat{\theta}$ — достаточная оценка для θ, $\tilde{\theta}$ — какая-либо несмещенная оценка θ и σ^2 и $k\sigma^2$, $k>1$ — дисперсии $\hat{\theta}$ и $\tilde{\theta}$ соответственно, то коэффициент корреляции между $\hat{\theta}$ и $\tilde{\theta}$ равен $\frac{1}{\sqrt{k}}$.

12.11 (*Продолжение*). Показать, что если $\hat{\theta}$ — эффективная оценка θ в то время как $\tilde{\theta}_1$ и $\tilde{\theta}_2$ — не эффективные, но несмещенные оценки с дисперсиями, равными $k\sigma^2$, где $k>1$, то коэффициент корреляции между $\tilde{\theta}_1$ и $\tilde{\theta}_2$ равен по крайней мере $(2-k)/k$.

12.12. Предположим, что $(x_1, \dots, x_n)$ — k-мерная случайная величина, имеющая плотность $f(x_1, \dots, x_k;\ \theta)$, $\theta\in\Omega_0$. Пусть $g(x_1, \dots, x_n)$ — несмещенная оценка $\psi(\theta)$. Показать, что если f регулярна вместе с первой производной по θ в Ω_0 и $\psi(\theta)$ имеет производную в Ω_0, то

$$\sigma^2(g)\geqslant\left(\frac{\partial\psi}{\partial\theta}\right)^2 \Big/ \mathscr{E}\left(\frac{\partial\log f}{\partial\theta}\right)^2,$$

и знак равенства имеет место тогда и только тогда, когда $\partial\log f/\partial\theta\equiv\equiv c\,(g\psi(\theta))$ с вероятностью 1, где c не зависит от $(x_1, \dots, x_n)$.

12.13. Показать, что если $(x_{(1)}, \dots, x_{(n)})$ — вариационный ряд, составленный из выборки из прямоугольного распределения $R(\theta/2, \theta)$, то $x_{(n)}$ является достаточной статистикой для θ и $(n-1)\,\bar{x}/n$ — эффективная оценка θ. Показать, что $(x_{(n)}, x_{(n)}/\sqrt[n]{1-\gamma})$ — 100γ-процентный доверительный интервал для θ.

12.14. Показать, что если $(x_{(1)}, \ldots, x_{(n)})$ — вариационный ряд, составленный из выборки с прямоугольным распределением $R((\theta_1+\theta_2)/2, \theta_2-\theta_1)$, $\theta_2>\theta_1>0$, двумерная случайная величина $(x_{(1)}, x_{(n)})$ является достаточной оценкой для (θ_1, θ_2). Показать, что

$$\frac{nx_{(1)}}{n-1}-\frac{x_{(n)}}{n-1}$$

и

$$\frac{nx_{(n)}}{n-1}-\frac{x_{(1)}}{n-1}$$

— оценки для θ_1 и θ_2 соответственно, имеющие минимальную дисперсию. Показать также, что

$$\frac{x_{(1)}-x_{(n)}}{2} \quad \text{и} \quad \frac{n+1}{n-1}(x_{(n)}-x_{(1)})$$

— оценки средней точки $(\theta_1+\theta_2)/2$ и размаха $\theta_2-\theta_1$, имеющие минимальную дисперсию.

12.15. Показать, что дисперсия любой несмещенной оценки для σ^2 по выборке объема n из $N(\mu, \sigma^2)$ не меньше $2\sigma^4/n$. Вывести отсюда, что эффективность выборочной дисперсии s^2 (как несмещенной оценки) при оценивании σ^2 равна $(n-1)/n$.

12.16. Показать, что если $(x_1, \ldots, x_n)$ — выборка из нормального распределения $N(\mu, \sigma^2)$, где μ известно, то

$$\frac{1}{n}\sqrt{\pi/2}\sum_{\xi=1}^{n}|x_\xi-\mu|$$

является несмещенной оценкой для σ, эффективность которой равна $1/(\pi-2)$.

12.17. Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности с распределением Коши

$$\frac{1}{\pi[1+(x-\mu)^2]}.$$

Показать, что предельное значение эффективности выборочной медианы для оценки μ при $n\to\infty$ равно $8/\pi^2$.

12.18. Пусть $(x_1, \ldots, x_n)$ — выборка из распределения с плотностью

$$\frac{\lambda^{k+1}x^k e^{-\lambda x}}{\Gamma(k+1)}, \qquad x>0,$$

где $\lambda>0$ и k - известная константа. Показать, что оценкой максимального правдоподобия $\hat{\lambda}$ для λ является величина $(k+1)/x$. Показать, что эта оценка смещена, но состоятельна и ее асимптотическое распределение при больших n — $N(\lambda, \lambda^2/[n(k+1)])$.

12.19. (*Продолжение*). Показать, что $\sqrt{k+1}\log\hat{\lambda}$ есть функция $\hat{\lambda}$, асимптотическое распределение которой при больших n нормально с дисперсией $\frac{1}{n}$.

12.20. Пусть $(x_1, \ldots, x_n)$ — выборка из биномиального распределения $\mathrm{Bi}(1, p)$. Показать, что оценкой максимального правдоподобия $\hat{p}$ для параметра p является $\bar{x}$ и ее асимптотическое распределение для больших выборок нормально $N(p, [p(1-p)]/n)$. Показать, что $\sin^{-1}(2\hat{p}-1)$ имеет асимптотически нормальное распределение с дисперсией $\frac{1}{n}$ при больших выборках.

12.21. Пусть $\bar{x}$ — среднее выборки объема n из биномиального распределения $\mathrm{Bi}(1, p)$. Показать, что асимптотически наикратчайший 100γ-процентный интервал при больших n задается двумя значениями p, которые удовлетворяют уравнению

$$\frac{(\bar{x}-p)\sqrt{n}}{\sqrt{p(1-p)}} = \pm y_\gamma,$$

где $P(-y_\gamma < u < +y_\gamma) = \gamma$, u — случайная величина с распределением $N(0, 1)$.

12.22. Пусть $\bar{x}$ — среднее выборки объема n из распределения Пуассона $\mathrm{Po}(\mu)$. Показать, что асимптотически наикратчайший 100γ-процентный доверительный интервал для μ при больших n задается двумя значениями μ, которые удовлетворяют уравнению

$$\frac{(\bar{x}-\mu)\sqrt{n}}{\sqrt{\mu}} = \pm y_\gamma,$$

где y_γ определяется, как и в предыдущем упражнении.

12.23. Пусть $(x_{1\xi}, \ldots, x_{k\xi}, \xi = 1, \ldots, n)$ — выборка объема n из мультиномиального распределения $M(1; p_1, \ldots, p_k)$. Показать, что оценкой максимального правдоподобия для $(p_1, \ldots, p_k)$ будет $(\bar{x}_1, \ldots, \bar{x}_k)$, где $\bar{x}_i = \frac{1}{n}\sum_{\xi=1}^{n} x_{i\xi}$ с матрицей ковариаций

$$\left\| \frac{1}{n}(\delta_{ij}p_i - p_i p_j) \right\|,$$

где $\delta_{ij} = 1$, если $i = j$, и $\delta_{ij} = 0$, если $i \neq j$.

12.24. Пусть $(x_{1\xi}, \ldots, x_{k\xi}, \xi = 1, 2, \ldots, n)$ — выборка объема n из совокупности с k-мерным нормальным распределением $N(\{\mu_i\}, \|\sigma_{ij}\|)$. Показать, что оценкой максимального правдоподобия для вектора $(\mu_1, \ldots, \mu_k)$ будет вектор выборочных средних компонент $(\bar{x}_1, \ldots, \bar{x}_k)$ и оценкой максимального правдоподобия матрицы $\|\sigma_{ij}\|$ —

$$\left\| \frac{n-1}{n} s_{ij} \right\|,$$

где

$$s_{ij} = \frac{1}{n-1}\sum_{\xi=1}^{n}(\bar{x}_{i\xi} - \bar{x}_i)(\bar{x}_{j\xi} - \bar{x}_j).$$

12.25. Предположим, что $(y_1, \ldots, y_n)$ — независимые случайные величины, имеющие нормальные распределения

$$N(\beta_1 x_{11} + \ldots + \beta_k x_{k1}, \sigma^2), \ldots, N(\beta_1 x_{1n} + \ldots + \beta_k x_{kn}, \sigma^2)$$

соответственно, где $x_{11}, \ldots, x_{k1}, \ldots, x_{1n}, \ldots, x_{kn}$ постоянны. Пусть

$$a_{pq} = \sum_{\xi=1}^{n} x_{p\xi}x_{q\xi}, \qquad p, q = 1, \ldots, k,$$

$$a_{p0} = \sum_{\xi=1}^{n} x_{p\xi}y_\xi, \qquad p = 1, \ldots, k.$$

Показать, что оценкой максимального правдоподобия для $(\beta_1, \ldots, \beta_k)$ является

$$\left(\sum_{p=1}^{k} a_{p0}a^{p1}, \ldots, \sum_{p=1}^{k} a_{p0}\, a^{pk}\right),$$

где $\| a^{pq} \| = \| a_{pq} \|^{-1}$, $p, q = 1, \ldots, k$, и $\| a_{pq} \|$ — неособая матрица. Показать также, что оценкой максимального правдоподобия для σ^2 является

$$\frac{1}{n} | a_{p_0q_0} | / | a_{pq} |, \; p_0, q_0 = 0, 1, \ldots, k, \; p, q = 1, \ldots, k,$$

где $a_{00} = \sum_{\xi=1}^{n} y_\xi^2$.

12.26. Предположим, что $(x_1, \ldots, x_n)$ имеет плотность $f_n(x_1, \ldots, x_n; \theta)$ и $\tilde{\theta}$ — несмещенная оценка θ. Показать, что для любого $\delta > 0$

$$\sigma^2(\tilde{\theta} \mid \theta_0) \geqslant 1/[\mathscr{E}(g^2/\theta_0)],$$

где

$$g = \frac{f_n(x_1, \ldots, x_n, \theta_0 + \delta) - f_n(x_1, \ldots, x_n; \theta_0)}{\delta f_n(x_1, \ldots, x_n; \theta_0)},$$

и, следовательно, что

$$\sigma^2(\tilde{\theta} \mid \theta_0) \geqslant 1/[\inf_{\delta} \mathscr{E}(g^2 \mid \theta_0)].$$

Этот результат принадлежит Чэпмену и Роббинсу (1951).

12.27. Пусть $(x_1, \ldots, x_n)$ — случайная величина, имеющая плотность распределения $f_n(x_1, \ldots, x_n; \theta)$ такую, что имеет место (12.2.25), т. е. существует достаточная статистика $\tilde{\theta}$ для θ, и пусть $f_n(x_1, \ldots, x_n; \theta)$ и $v(\tilde{\theta}, \theta)$ имеют вторые частные производные $\partial^2 f_n/\partial x_\xi \partial\theta$ и $\partial^2 v/\partial x_\xi \partial\theta$ в каждой точке $R_n \times \Omega$ (за исключением, может быть, множества вероятностной меры 0), где R_n — выборочное пространство и Ω — параметрическое пространство (открытый интервал на вещественной оси).

Показать, что $f_n(x_1, \ldots, x_n; \theta)$ должна иметь вид $\exp[k_1(\theta)\, g(\tilde{\theta}) + k_2(\theta) + h_n(x_1, \ldots, x_n)]$, где $k_1(\theta)$ и $k_2(\theta)$ не зависят от $(x_1, \ldots, x_n)$ $h_n(x_1, \ldots, x_n)$ не зависит от θ и $g(\tilde{\theta})$ зависит от $\tilde{\theta}$ и не зависит от θ [Купмэн (1936) и Питмэн (1936)].

12.28. (*Продолжение*). *Функции от достаточных оценок как оценки с минимальной дисперсией для их средних значений* [Рао (1949)]. Пусть $u(g(\tilde{\theta}))$ будет функцией $\tilde{\theta}$ со средним значением $q(\theta)$. Пусть $v(g(\theta))$ — другая функция со средним значением $q(\theta)$. Тогда, если $W(g(\tilde{\theta})) = u(g(\tilde{\theta})) - v(g(\tilde{\theta}))$, то $\int_{R_n} W(g) \exp[k_1(\theta)\, g + k_1(\theta) + h_n]\, dx_1, \ldots, dx_n = 0$ для всех $\theta \in \Omega$. Показать, что если это выражение можно повторно дифференцировать по θ под знаком интеграла и $W(g)$ может быть представлена в виде ряда Тейлора, то $\mathscr{E}(W(g))^2 = 0$ для всех $\theta \in \Omega$, откуда следует, что $W(g) = 0$ с вероятностью 1 для всех $\theta \in \Omega$. Следовательно, $u(g(\tilde{\theta}))$ отличается от $v(g(\tilde{\theta}))$ разве что на множестве вероятностной меры 0, что означает, что при сделанных предположениях $u(g(\tilde{\theta}))$ — единственная несмещенная оценка своего среднего значения. Заметим, что мы делаем вывод, что $W(g) = 0$ с вероятностью 1 из того факта, что $\mathscr{E}(W(g))^2 = 0$ для всех $\theta \in \Omega$. Более общо: если y — случайная величина с к. ф. р. $F(y; \theta)$ такая, что из равенства $\mathscr{E}(y \mid \theta) = 0$ для всех $\theta \in \Omega$ следует, что $y = 0$ с вероятностью 1, то говорят, что семейство распределений $\{F(y; \theta);\; \theta \in \Omega\}$ является полным.

ГЛАВА 13

ПРОВЕРКА ПАРАМЕТРИЧЕСКИХ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

13.1. Предварительные замечания и определения

Эта глава посвящена главным образом основным принципам теории проверки параметрических статистических гипотез, построенной Нейманом и Пирсоном (1928, 1933), и связям этой теории с оценками максимального правдоподобия в больших выборках. Идеи статистической проверки гипотез значительно развиты в последние несколько лет; но мы не будем здесь подробно рассматривать эти обобщения. Читатель, интересующийся ими, найдет блестящее изложение предмета в книге Лемана (1959).

В этом параграфе сформулированы некоторые основные понятия и результаты для малых выборок; в следующих параграфах рассматривается теория больших выборок. Эта асимптотическая теория тесно примыкает к теории оценивания параметров в больших выборках, изложенной в главе 12.

Пусть $(x_1, \ldots, x_n)$ — n-мерный случайный вектор с ф. р. $F_n(x_1, \ldots, x_n; \theta)$, где θ — r-мерный параметр, принадлежащий множеству Ω, которое здесь будем считать открытым подмножеством эвклидова пространства R_r. Наиболее важен тот случай, когда $(x_1, \ldots, x_n)$ — повторная выборка из совокупности с к. ф. р. $F(x; \theta)$. В этом случае $F_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} F(x_i, \theta)$. Бо́льшая часть этой главы посвящена схеме повторной выборки.

Пусть ω — подмножество Ω. Если θ_0 — истинное значение параметра в совокупности, то *статистической гипотезой* $\mathscr{H} = \mathscr{H}(\omega; \Omega)$ называется утверждение: $\theta_0 \in \omega$ против альтернативы: $\theta_0 \in \Omega \setminus \omega$. Говорят, что гипотеза $\mathscr{H}$ *истинна*, если $\theta_0 \in \omega$, и *ложна*, если $\theta_0 \in \Omega \setminus \omega$. Утверждение $\theta_0 \in \omega$ называют иногда *нулевой гипотезой*. При отсутствии недоразумений, мы будем писать θ вместо θ_0. Если ω содержит одну точку θ, гипотеза $\mathscr{H}$ называется *простой*; в противном случае $\mathscr{H}$ называют *сложной* гипотезой. Решение принять или отвергнуть гипотезу $\mathscr{H}$ выносится на основании значения вектора $(x_1, \ldots, x_n)$. Пусть W_n — множество в выборочном пространстве R_n, не зависящее от θ и такое, что $\mathscr{H}$ отвергается, когда $(x_1, \ldots, x_n) \in W_n$, и принимается в противном случае. Здесь возможны ошибки двух типов: *ошибка 1-го рода* происходит при отклонении гипотезы $\mathscr{H}$, когда она истинна, *ошибка 2-го рода* происходит при принятии $\mathcal{H}$, когда она ложна. Выбор множества W_n и отклонение гипотезы $\mathcal{H}$ при попадании $(x_1, \ldots, x_n)$ в W_n называют *статистическим критерием* (*тестом*) для проверки гипотезы $\mathcal{H}$ или, короче, *критерием* гипотезы $\mathcal{H}$; W_n называется *критическим множеством* или *критической областью* теста. Без каких-либо недоразумений можно для сокращения называть множество W_n *критерием* гипотезы $\mathcal{H}$. Величина *)

$$P(W_n \mid \theta) = \int\limits_{W_n} dF_n \tag{13.1.1}$$

называется *мощностью* критерия W_n; она представляет собой функцию θ со значениями в интервале $[0,1]$. Функция $P(\overline{W}_n \mid \theta) = 1 - P(W_n \mid \theta)$ параметра θ называется *оперативной характеристикой* критерия W_n. Вероятности ошибок 1-го и 2-го рода равны соответственно

$$P(W_n \mid \theta), \quad \theta \in \omega, \tag{13.1.2}$$

и

$$1 - P(W_n \mid \theta), \quad \theta \in \Omega \setminus \omega. \tag{13.1.3}$$

Эти вероятности часто называют *рисками ошибок 1-го и 2-го рода*. Конечно, тест будет приемлемым, если он удовлетворяет некоторым условиям. Тест W_n был бы идеальным, если бы вероятности ошибок 1-го и 2-го рода, определенные в (13.1.2) и (13.1.3), равнялись нулю. Это требование, очевидно, слишком сильное, чтобы надеяться на возможность удовлетворить ему в общем случае; приходится ограничиться меньшим.

Желательным свойством критерия W_n является *несмещенность*, которая для заданного *уровня значимости* α, $0 < \alpha < 1$, определяется условиями:

$$P(W_n \mid \theta) \leqslant \alpha, \quad \theta \in \omega, \tag{13.1.4}$$

и

$$P(W_n \mid \theta) > \alpha, \quad \theta \in \Omega \setminus \omega. \tag{13.1.5}$$

Эти условия означают, что вероятность ошибки 1-го рода не превосходит α и в то же время вероятность отвергнуть $\mathcal{H}$, когда она ложна (т. е. при $\theta \in \Omega \setminus \omega$), превышает α. Если тест W_n таков, что $P(W_n \mid \theta) \leqslant \alpha$ для $\theta \in \omega$, то говорят, что он имеет объем α, и иногда подчеркивают это обозначением $W_{n,\alpha}$. Там, где это не приведет к неясностям, мы будем в дальнейшем обозначать через W_α критерий объема α. Тест, не обладающий свойством несмещенности, называется *смещенным*.

*) Некоторые авторы предпочитают вводить *характеристическую функцию* φ_{W_n} *множества* W_n, $\varphi_{W_n} = 1$ во всех точках W_n и 0 в остальных точках; $P(W_n \mid \theta) = E(\varphi_{W_n} \mid \theta)$.

Если для критерия W_α в (13.1.4) имеет место знак равенства при всех $\theta \in \omega$, то область W_α называется *подобной пространству выборок*, или просто *подобной критической областью*.

Если

$$\lim_{n\to\infty} P(W_\alpha \mid \theta) = 1, \quad \theta \in \Omega \setminus \omega,$$

то W_α называется *состоятельным критерием* объема α гипотезы $\mathscr{H}(\omega;\ \Omega)$. Этот термин введен Вальдом и Вольфовитцем (1940).

Пусть W_α и W_α^* — два критерия объема α гипотезы $\mathscr{H}(\omega;\ \Omega)$ таких, что

$$P(W_\alpha^* \mid \theta) \leqslant P(W_\alpha \mid \theta), \qquad \theta \in \omega, \tag{13.1.6}$$

и

$$P(W_\alpha^* \mid \theta) > P(W_\alpha \mid \theta), \qquad \theta \in \Omega \setminus \omega. \tag{13.1.7}$$

Тогда критерий W_α^* равномерно мощнее критерия W_α; следует предпочесть W_α^* перед W_α. Если (13.1.6) и (13.1.7) выполняются для всякого (борелевского) подмножества $W_\alpha \subset R_n$, то W_α^* называется *равномерно наиболее мощным критерием* для проверки гипотезы $\mathscr{H}$.

Пример. Чтобы проиллюстрировать некоторые из изложенных выше идей, рассмотрим выборку $(x_1, \ldots, x_n)$ из нормальной совокупности $N(\mu, \sigma^2)$. Пусть $\mathscr{H}(\omega;\ \Omega)$ — сложная гипотеза, где Ω — множество точек (μ, σ^2) с $\sigma^2 > 0$ (т. е. эвклидова полуплоскость), а ω — подмножество Ω, в точках которого $\mu = \mu_0$ (другими словами, ω — полупрямая: $\mu = \mu_0$, $\sigma^2 > 0$).

Согласно теореме **8.4.3**, если $\mathscr{H}$ истинна, то $\sqrt{n}\,(\bar{x} - \mu)/s$ имеет распределение Стьюдента $S(n-1)$. Пусть W_α — множество тех точек R_n, в которых

$$\left| \frac{(\bar{x} - \mu_0)\sqrt{n}}{s} \right| > t_\alpha,$$

где

$$1 - \int_{-t_\alpha}^{+t_\alpha} f_{n-1}(t)\,dt = \alpha$$

и $f_{n-1}(t)$ — плотность распределения Стьюдента $S(n-1)$, задаваемая формулой (7.8.4). Тогда

$$P(W_\alpha \mid (\mu, \sigma^2) \in \omega) = P(|t| > t_\alpha) = \alpha,$$

и следовательно, тест W_α гипотезы $\mathscr{H}(\omega;\ \Omega)$ имеет объем α и является подобным выборочному пространству. Кроме того, для всякой точки $(\mu_1, \sigma_1^2) \in \Omega \setminus \omega$

$$P(W_\alpha \mid (\mu_1, \sigma_1^2)) = P\left(\left| \frac{(\bar{x} - \mu_0)\sqrt{n}}{s} \right| > t_\alpha \mid (\mu_1, \sigma_1^2) \right) = P\left(\left| t + \frac{\delta}{\sqrt{u}} \right| > t_\alpha \right),$$

где $\delta = (\mu_1 - \mu_0)\sqrt{n(n-1)}/\sigma_1$, а t и u — случайные величины, для которых элемент вероятности в полуплоскости $u \geqslant 0$ имеет вид

$$f(t, u)\,dt\,du = \frac{\left(\frac{u}{2}\right)^{\frac{1}{2}(n-3)}}{2\sqrt{\pi(n-1)}\,\Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{1}{2}(1 + t^2/(n-1))\,u}\,dt\,du.$$

Если учесть, что для всех $u \in (0, +\infty)$

$$\int\limits_{-t_\alpha - \frac{\delta}{\sqrt{u}}}^{+t_\alpha - \frac{\delta}{\sqrt{u}}} f(t, u)\, dt < \int\limits_{-t_\alpha}^{+t_\alpha} f(t, u)\, dt,$$

то получим

$$P\left(\left|t + \frac{\delta}{\sqrt{u}}\right| > t_\alpha\right) > P(|t| > t_\alpha) = \alpha, \quad \delta \neq 0,$$

что доказывает несмещенность критерия W_α.

К сожалению, несмещенные и равномерно наиболее мощные критерии по выборкам конечного объема существуют только в некоторых специальных случаях. Для простой гипотезы имеется достаточно общее решение, когда $\Omega \setminus \omega$ содержит ровно одну точку, и имеется решение при определенных условиях, если $\Omega \setminus \omega$ состоит более чем из одной точки. Случай простой гипотезы будет рассмотрен в § 13.2. Для сложных гипотез решения получены только в специальных ситуациях, в частности, когда W_α подобен пространству выборок. Но равномерно наиболее мощные несмещенные тесты существуют в асимптотическом смысле в больших выборках для некоторых сложных гипотез при условии, что оценки максимального правдоподобия параметров существуют и асимптотически нормальны. Мы встретимся с этой задачей в §§ 13.3—13.4.

Историческое замечание. При разработке теории выборочного контроля для приема или браковки партий изделий массового производства Додж и Ромиг из лабораторий компании «Белл Телефон» около 1925 г. ввели понятия *риска производства* и *риска потребителя*. Их первые опубликованные результаты появились в 1929 г. [См. Додж и Ромиг (1929, 1959).] Эти понятия были, по существу, предшественниками рисков ошибок 1-го и 2-го рода, определенных в (13.1.2) и (13.1.3). Именно, пусть θ — доля дефектных изделий в *партии* из N изделий, а t — доля дефектных изделий в выборке объема n из этой партии. При заданном значении θ t будет случайной величиной с к. ф. р. $F_n(t; \theta)$ и со значениями $0, \frac{1}{n}, \ldots, \frac{r}{n}, r \leqslant n$, в интервале $[0, 1]$. Множество значений параметра θ состоит из точек $0, \frac{1}{N}, \ldots, \frac{N}{N}$ интервала $[0, 1]$. Заданное значение параметра θ_0 определяет *недопустимую* долю дефектных изделий в партии как значение $\theta > \theta_0$, что соответствует $\omega = (\theta_0, 1)$. *Допустимая* доля дефектных изделий (*зона предпочтения*) определяется значением $\theta \leqslant \theta_0$, так что $\Omega \setminus \omega = (0, \theta_0)$. Выберем теперь значение t_α и критическое множество $W_\alpha = [t_\alpha, 1]$, так что при $t \in W_\alpha$ партия бракуется и $P(t \in W_\alpha \mid \theta \in \omega) \leqslant \alpha$. Величина $P(t \in \overline{W}_\alpha \mid \theta \in \omega)$ называется *риском производства*; это — вероятность того, что контролер забракует партию при допустимой доле θ дефектных изделий в этой партии. Значение t_α выбирается так, чтобы риск производства не превышал α и приблизительно был равен α при $\theta = \theta_0$. Обычно на практике полагают α равным 0,05 или 0,10. Величина $P(t \in \overline{W}_\alpha \mid \theta \in \Omega \setminus \omega)$ называется *риском потребителя*, она представляет собой вероятность того, что контролер примет партию, когда она содержит недопустимую долю дефектных изделий θ. На практике, если N достаточно велико, всегда можно выбрать n так, чтобы сделать риск потребителя сколь угодно малым для любого фиксированного значения доли дефектных изделий, превосходящего θ_0. Обычно объем выборки n подбирают так, чтобы риск потребителя β имел заданное значение (как правило, 0,05 или 0,10) при некотором $\theta=\theta_1$, немного превосходящем θ_0. Интервал $(\theta_0, 1)$ недопустимых значений θ разбивается на два множества: *зону безразличия* (θ_0, θ_1) и *зону отклонения* $(\theta_1, 1)$. Вероятность $P(t \in \overline{W}_\alpha \mid \theta)$ принять партию при доле дефектных изделий θ для заданного n будет функцией θ, ее обозначают $L(\theta)$ и называют ***оперативной характеристикой*** выборочного плана, определенного заданием объема выборки n и критической доли дефектных изделий t_α. Оперативная характеристика удовлетворяет условиям $L(\theta_0)=1-\alpha$ и $L(\theta_1)=\beta$.

Вторым важным понятием, введенным Доджем и Ромигом (1929), было понятие ***предела среднего выходного качества***. Он определяется как максимум по всем θ среднего значения доли дефектных изделий в партии при условии, что в забракованных партиях все дефектные изделия заменены заведомо годными. При этом, конечно, предполагается неразрушительный характер проверки изделий, т. е. в процессе проверки изделие не разрушается.

Понятия ошибок 1-го и 2-го рода давно уже играют фундаментальную роль при отправлении правосудия; роль рисков ошибок 1-го и 2-го рода здесь играют соответственно риски осуждения невиновного и оправдания виновного.

13.2. Критерий простой гипотезы

(а) Случай двух альтернатив. Случай простой гипотезы $\mathcal{H}$ может быть описан указанием точки θ_0 в ω и точки θ_1 в $\Omega \setminus \omega$. Если вектор $(x_1, \ldots, x_n)$ имеет Ф. п. в. $f_n(x_1, \ldots, x_n; \theta)$ для $\theta=\theta_0$ и $\theta=\theta_1$, то при формулируемых ниже условиях существует несмещенный и наиболее мощный тест гипотезы $\mathcal{H}$ $(\theta_0; \theta_0 \cup \theta_1)$. Эта задача была впервые рассмотрена Нейманом и Пирсоном (1933). Следующий принадлежащий им результат дает наиболее мощный несмещенный критерий для $\mathcal{H}$.

13.2.1. *Пусть $(x_1, \ldots, x_n)$ — случайный вектор с ф. п. в. $f_n(x_1, \ldots, x_n; \theta)$ и параметрическое множество Ω состоит из двух различных точек θ_0 и θ_1. Если непрерывные распределения, задаваемые плотностями $f_n(x_1, \ldots, x_n; \theta_0)$ и $f_n(x_1, \ldots, x_n; \theta_1)$, взаимно абсолютно непрерывны, то для любого α, $0<\alpha<1$, можно выбрать $c_\alpha>0$ так, чтобы*

$$P(W_\alpha \mid \theta_0)=\alpha, \tag{13.2.1}$$

где W_α — множество тех точек выборочного пространства, для которых

$$f_n(x_1, \ldots, x_n; \theta_1) \geqslant c_\alpha f_n(x_1, \ldots, x_n; \theta_0). \tag{13.2.2}$$

При этом W_α будет несмещенным и равномерно наиболее мощным критерием гипотезы $\mathcal{H}$ $(\theta_0; \theta_0 \cup \theta_1)$.

Так как распределения с плотностями $f_n(x_1, \ldots, x_n; \theta_0)$ и $f_n(x_1, \ldots, x_n; \theta_1)$ взаимно абсолютно непрерывны, то для любого множества $E^* \subset R_n$ $P(E^* \mid \theta_0)$ и $P(E^* \mid \theta_1)$ одновременно положительны или равны нулю. Поэтому, если $E^*(c)$ — множество тех точек R_n, в которых $f_n(x_1, \ldots, x_n; \theta_1) \geqslant cf_n(x_1, \ldots, x_n; \theta_0)$, то

$$g(c) = P(E^*(c) \mid \theta_0)$$

будет непрерывной монотонно убывающей функцией c и $g(0) = 1$, $g(+\infty) = 0$. Следовательно, для любого $\alpha > 0$ существует такое значение $c = c_\alpha$, что если положить $W_\alpha = E^*(c_\alpha)$, то это множество W_α будет удовлетворять (13.2.1).

Пусть теперь W_α^* — любой другой тест гипотезы $\mathscr{H}$ объема α, так что

$$P(W_\alpha^* \mid \theta_0) = \alpha. \qquad (13.2.3)$$

Имеем

$$P(W_\alpha \setminus (W_\alpha \cap W_\alpha^*) \mid \theta_0) = P(W_\alpha^* \setminus (W_\alpha \cap W_\alpha^*) \mid \theta_0). \qquad (13.2.4)$$

Но во всякой точке множества W_α выполняется условие $f_n(x_1, \ldots$ $\ldots, x_n; \theta_1) \geqslant c_\alpha f_n(x_1, \ldots, x_n; \theta_0)$, так что

$$P(W_\alpha \setminus (W_\alpha \cap W_\alpha^*) \mid \theta_1) \geqslant c_\alpha P(W_\alpha \setminus (W_\alpha \cap W_\alpha^*) \mid \theta_0). \qquad (13.2.5)$$

Для всякой точки, не принадлежащей W_α, $f_n(x_1, \ldots, x_n; \theta_1) <$ $< c_\alpha f_n(x_1, \ldots, x_n; \theta_0)$ и, следовательно,

$$c_\alpha P(W_\alpha^* \setminus (W_\alpha \cap W_\alpha^*) \mid \theta_0) \geqslant P(W_\alpha^* \setminus (W_\alpha \cap W_\alpha^*) \mid \theta_1). \qquad (13.2.6)$$

Используя (13.2.4), (13.2.5) и (13.2.6), получим

$$P(W_\alpha \setminus (W_\alpha \cap W_\alpha^*) \mid \theta_1) \geqslant P(W_\alpha^* \setminus (W_\alpha \cap W_\alpha^*) \mid \theta_1). \qquad (13.2.7)$$

Множества $W_\alpha \setminus (W_\alpha \cap W_\alpha^*)$ и $W_\alpha \cap W_\alpha^*$ не пересекаются, так же как множества $W_\alpha^* \setminus (W_\alpha \cap W_\alpha^*)$ и $W_\alpha \cap W_\alpha^*$. Поэтому, добавляя величину $P(W_\alpha \cap W_\alpha^* \mid \theta_1)$ к обеим частям неравенства (13.2.7), будем иметь

$$P(W_\alpha \mid \theta_1) \geqslant P(W_\alpha^* \mid \theta_1). \qquad (13.2.8)$$

Полученное неравенство как раз означает, что W_α — более мощный критерий гипотезы $\mathscr{H}$, чем W_α^*. Но так как в качестве W_α^* можно брать любой тест объема α, то W_α — наиболее мощный тест гипотезы $\mathscr{H}$.

Очевидно, тест W_α несмещенный, если в (13.2.2) $c_\alpha \geqslant 1$. Если $c_\alpha < 1$, то найдется такое подмножество $W'_\alpha \subset W_\alpha$, что

$$f_n(x_1, \ldots, x_n; \theta_1) \geqslant f_n(x_1, \ldots, x_n; \theta_0), \quad (x_1, \ldots, x_n) \in W'_\alpha. \qquad (13.2.9)$$

Действительно, в противном случае во всем R_n было бы

$$0 \leqslant f_n(x_1, \ldots, x_n; \theta_1) < f_n(x_1, \ldots, x_n; \theta_0),$$

что противоречит соотношению

$$\int_{R_n} f_n(x_1, \ldots, x_n; \theta_0)\, dx_1 \ldots dx_n = \int_{R_n} f_n(x_1, \ldots, x_n; \theta_1)\, dx_1 \ldots dx_n = 1.$$

Далее,

$$P(W'_\alpha\,|\,\theta_1)-P(W'_\alpha\,|\,\theta_0)=P(R_n\setminus W'_\alpha\,|\,\theta_0)-P(R_n\setminus W'_\alpha\,|\,\theta_1)\geqslant$$
$$\geqslant P(W_\alpha\setminus W'_\alpha\,|\,\theta_0)-P(W_\alpha\setminus W'_\alpha\,|\,\theta_1), \quad (13.2.10)$$

так что

$$P(W_\alpha\,|\,\theta_1)\geqslant P(W_\alpha\,|\,\theta_0)=\alpha, \quad (13.2.11)$$

т. е. W_α — несмещенный критерий.

Теорему, аналогичную **13.2.1**, можно доказать и в случае, когда вектор $(x_1, \ldots, x_n)$ имеет дискретное распределение вероятностей $p_n(x_1, \ldots, x_n; \theta)$, а распределения $p_n(x_1, \ldots, x_n; \theta_0)$ и $p_n(x_1, \ldots \ldots, x_n; \theta_1)$ взаимно абсолютно непрерывны. В этом случае существует наименьшее c_α, для которого $P(W_\alpha\,|\,\theta_0)\leqslant\alpha$.

Наконец, отметим, что если $(x_1, \ldots, x_n)$ — случайная выборка из совокупности с п. в. $f(x; \theta)$, то в **13.2.1** надо просто заменить $f_n(x_1, \ldots, x_n)$ на $\prod_{i=1}^{n} f(x_i; \theta)$. Аналогичное замечание сохраняет силу для выборки из совокупности с дискретным распределением.

(b) Случай более общей простой гипотезы. Для некоторых специальных плотностей $f_n(x_1, \ldots, x_n; \theta)$ семейство множеств, определяемых формулой (13.2.2), когда c пробегает все положительные значения, а альтернативой служит $\theta=\theta_1$, оказывается одним и тем же для всех $\theta_1\in\Omega$, $\theta_1\neq\theta_0$. В этом случае W_α есть равномерно наиболее мощный несмещенный тест гипотезы $\mathscr{H}(\theta_0; \Omega)$.

В частности, если существует достаточная статистика $\tilde\theta$ с п. в. $v(\tilde\theta; \theta)$ такой, что семейство множеств в R_n, задаваемых условием $v(\tilde\theta; \theta_1)\geqslant cv(\tilde\theta; \theta_0)$ при всевозможных положительных значениях c, оказывается одним и тем же при любом выборе $\theta_1\neq\theta_0$ из Ω, то для заданного α можно найти равномерно наиболее мощный несмещенный тест W_α гипотезы $\mathscr{H}(\theta_0; \Omega)$, который, очевидно, определяется достаточной статистикой $\tilde\theta$. Но даже этот подход дает мало надежды на успех, за исключением лишь случая одномерного параметра θ.

Приведем пример существования равномерно наиболее мощного несмещенного теста.

Пример. Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности $N(\mu, 1)$. Параметрическим множеством для μ пусть служит вся прямая R_1. Рассмотрим гипотезу

$$\mathscr{H}(\mu_0; \mu_0\cup\mu_1).$$

Используя (13.2.2), после некоторых преобразований находим, что W_α — множество тех точек R_n, для которых

$$e^{n(\mu_1-\mu_0)\bar x}\geqslant c'_\alpha,$$

где $c'_\alpha>0$. Поэтому W_α определяется значением достаточной статистики $\bar x$. Поскольку $\bar x$ имеет распределение $N\left(\mu, \frac{1}{n}\right)$, то

(i) если $\mu_1 > \mu_0$, то W_α — множество точек R_n, для которых $\bar{x} > z_\alpha$, где z_α определяется из уравнения

$$1 - \Phi(\sqrt{n}(z_\alpha - \mu_0)) = \alpha,$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx;$$

(ii) если $\mu_1 < \mu_0$, то W_α — множество точек R_n, для которых $\bar{x} < z'_\alpha$, где

$$\Phi(\sqrt{n}(\mu_0 - z'_\alpha)) = \alpha.$$

В случае (i) тест, задаваемый критическим множеством $\bar{x} > z_\alpha$, оказывается равномерно наиболее мощным для гипотезы $\mathscr{H}(\mu_0; \mu \geqslant \mu_0)$, т. е. для проверки нулевой гипотезы $\mu = \mu_0$ против любой альтернативы $\mu = \mu_1$ при $\mu_1 > \mu_0$. Так как

$$1 - \Phi(\sqrt{n}(z_\alpha - \mu_1)) > 1 - \Phi(\sqrt{n}(z_\alpha - \mu_0)),$$

если $\mu_1 > \mu_0$, то очевидно, что этот критерий несмещенный, когда Ω содержит все значения μ, $\mu \geqslant \mu_0$.

Аналогичные замечания можно сделать в случае (ii).

(c) Вальдовская редукция сложной гипотезы с одной точкой в множестве $\Omega \setminus \omega$ к простой гипотезе. В задаче построения теста гипотезы $\mathscr{H}(\omega; \omega \cup \theta_1)$ Вальд (1939) предложил использовать замену семейства плотностей $f_n(x_1, \ldots, x_n; \theta)$ для $\theta \in \omega$ плотностью $f_G(x_1, \ldots, x_n)$, получающейся усреднением $f_n(x_1, \ldots, x_n; \theta)$ по некоторому *априорному* распределению $G(\theta)$ на ω. Положим

$$f_G(x_1, \ldots, x_n) = \int_\omega f_n(x_1, \ldots, x_n; \theta)\, dG(\theta), \tag{13.2.12}$$

тогда $f_G(x_1, \ldots, x_n)$ будет плотностью некоторого распределения. Рассмотрим задачу проверки гипотезы $\mathscr{H}_G$ о том, что Ф. п. в. есть $f_G(x_1, \ldots, x_n)$, против альтернативы, согласно которой Ф. п. в. равна $f_n(x_1, \ldots, x_n; \theta_1)$.

Если $f_G(x_1, \ldots, x_n)$ удовлетворяет условиям, наложенным на $f_n(x_1, \ldots, x_n; \theta)$ в теореме **13.2.1**, то эта теорема даст наиболее мощный критерий $W_{G\alpha}$ объема α для гипотезы $\mathscr{H}_G$. Вальдом (1939) доказана следующая теорема.

13.2.2. *Если существует такое распределение $G(\theta)$ на ω, что наиболее мощный тест $W_{G\alpha}$ объема α для гипотезы $\mathscr{H}_G$ имеет в то же время объем α для гипотезы $\mathscr{H}(\omega; \omega \cup \theta_1)$, то*

$$\int_{\omega_0} dG(\theta) = 0, \tag{13.2.13}$$

где ω_0 — *подмножество ω, в точках которого*

$$\int_{W_{G\alpha}} f_n(x_1, \ldots, x_n; \theta)\, dx_1 \ldots dx_n \leqslant \alpha. \tag{13.2.14}$$

Тест $W_{G\alpha}$ будет наиболее мощным тестом гипотезы $\mathscr{H}(\omega; \omega \cup \theta_1)$.

Для доказательства **13.2.2** заметим, что если $W_{G\alpha}$ имеет объем α для гипотезы $\mathscr{H}(\omega;\ \omega \bigcup \theta_1)$, то

$$\int\limits_{W_{G\alpha}} f_n(x_1, \ldots, x_n; \theta)\, dx_1 \ldots dx_n \leqslant \alpha, \qquad \theta \in \omega, \qquad (13.2.15)$$

и, следовательно,

$$\int\limits_{\omega} \int\limits_{W_{G\alpha}} f_n(x_1, \ldots, x_n;\ \theta)\, dx_1 \ldots dx_n\, dG(\theta) \leqslant \alpha \int\limits_{\omega} dG(\theta). \qquad (13.2.16)$$

Если изменить здесь порядок интегрирования, то станет ясно, что для того, чтобы имело место равенство

$$\int\limits_{W_{G\alpha}} f_G(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = \alpha,$$

другими словами, для того чтобы критерий $W_{G\alpha}$ имел объем α, должно выполняться (13.2.13). Если $W_{G\alpha}$ — наиболее мощный тест гипотезы $\mathscr{H}_G$, то

$$\int\limits_{W_{G\alpha}} f_n(x_1, \ldots, x_n;\ \theta_1)\, dx_1 \ldots dx_n \leqslant$$
$$\leqslant \int\limits_{W'_{G\alpha}} f_n(x_1, \ldots, x_n;\ \theta_1)\, dx_1 \ldots dx_n, \qquad (13.2.17)$$

где $W'_{G\alpha}$ — произвольный тест объема α гипотезы $\mathscr{H}_G$. Но (13.2.17) в точности является условием того, что $W_{G\alpha}$ — наиболее мощный критерий гипотезы $\mathscr{H}(\omega;\ \omega \bigcup \theta_1)$.

Следует заметить, что если бы $W_{G\alpha}$ был подобным тестом, т. е. в (13.2.15) имел место знак равенства при $\theta \in \omega$, то (13.2.13) не являлось бы необходимым условием. Но в этом случае, конечно, свойство подобия, по существу, решает задачу определения критической области.

13.3. Критерий отношения правдоподобия

Критерии сложных гипотез, обладающие оптимальными свойствами в выборках конечного объема, в различных специальных задачах были получены с помощью принадлежащего Нейману и Пирсону (1928, 1933) принципа *отношения правдоподобия*. Этот принцип служит естественным обобщением критерия (13.2.2) на случай сложной гипотезы.

Критерии отношения правдоподобия для больших выборок обладают оптимальными асимптотическими свойствами при тех же условиях, при которых оценки максимального правдоподобия асимптотически нормальны.

С другой стороны, более общие подходы к задаче проверки сложной гипотезы для выборок конечного объема развивались Леманом (1950, 1959), Леманом и Шеффе (1950), Леманом и Стейном (1948) и другими.

Мы рассмотрим подробно только критерии отношения правдоподобия, особенно их асимптотические свойства, при больших значениях n. Критерии отношения правдоподобия при больших n играют такую же фундаментальную роль в статистике, что и рассмотренная в главе 12 статистическая теория оценивания параметров в больших выборках.

Асимптотические распределения критериев отношения правдоподобия сложных гипотез впервые были получены при нулевой гипотезе Уилксом (1938a). Позже Вальд (1941a, 1941b) предпринял детальное изучение этих критериев, включая асимптотические свойства смещения и мощности.

Многие из результатов, содержащихся в следующих параграфах, несколько менее общи, чем результаты, изложенные в фундаментальных статьях Вальда. Но зато используемые здесь методы, по-видимому, проще методов Вальда.

(a) Определение критерия отношения правдоподобия. Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности $F(x; \theta)$, где θ — r-мерный параметр. Условимся считать x одномерным, хотя перенесение результатов на многомерный случай потребует только небольших изменений в обозначениях. Элемент правдоподобия есть

$$dF_n(x_1, \ldots, x_n; \theta) = \prod_{\xi=1}^{n} dF(x_\xi; \theta). \tag{13.3.1}$$

Рассмотрим гипотезу

$$\mathscr{H}(\omega; \Omega) \tag{13.3.2}$$

и положим

$$(dF_n)_\omega = \sup_{\theta \in \omega} dF_n(x_1, \ldots, x_n; \theta), \tag{13.3.3}$$

$$(dF_n)_\Omega = \sup_{\theta \in \Omega} dF_n(x_1, \ldots, x_n; \theta). \tag{13.3.4}$$

В большинстве обычных приложений $(dF_n)_\omega$ и $(dF_n)_\Omega$ равны максимальным значениям dF_n при $\theta \in \omega$ и $\theta \in \Omega$, получаемым с помощью обычного дифференцирования по θ.

Критерий отношения правдоподобия гипотезы $\mathscr{H}(\omega; \Omega) = \mathscr{H}$ определяется статистикой

$$\lambda_{\mathscr{H}} = \frac{(dF_n)_\omega}{(dF_n)_\Omega}. \tag{13.3.5}$$

Значения $\lambda_{\mathscr{H}}$ лежат в интервале $[0, 1]$. Критическое множество $W_\alpha \subset R_n$ для проверки гипотезы $\mathscr{H}$ имеет вид $\lambda_{\mathscr{H}} < c_\alpha$, где c_α — такая постоянная, что

$$\int_{W_\alpha} dF_n \leqslant \alpha, \quad \theta \in \omega. \tag{13.3.6}$$

Тест вида $\lambda_{\mathscr{H}} < c$ есть, по существу, обобщение критерия (13.2.2) на случай, когда ω и Ω содержат каждое более одной точки.

В дополнение к этому $\lambda_{\mathscr{H}}$ имеет интуитивную привлекательность в качестве теста. В самом деле, сравнивая «правдоподобие» одного значения θ против другого при заданной выборке $(x_1, \ldots, x_n)$, мы интуитивно склонны отдать предпочтение тому, которое доставляет большее значение элементу правдоподобия. Поэтому, если нельзя заметно увеличить значение элемента правдоподобия за счет расширения области значений θ от ω до всего параметрического пространства Ω, то интуиция подсказывает нам, что «наиболее вероятное» значение θ принадлежит ω, т. е. гипотеза $\mathscr{H}$ верна. Мы увидим, что для больших выборок эти интуитивные соображения могут быть строго обоснованы при довольно общих условиях.

Пример. Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности $N(\mu, \sigma^2)$. Определим критерий отношения правдоподобия гипотезы $\mathscr{H}$, для которой

$$\Omega \text{ — полуплоскость } -\infty < \mu < +\infty,\ \sigma^2 > 0,$$

$$\omega \text{ — полупрямая } \mu = \mu_0,\ \sigma^2 > 0.$$

Имеем

$$dF_n = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}n} \exp\left[-\frac{1}{2\sigma^2}\sum_{\xi=1}^{n}(x_\xi - \mu)^2\right] dx_1 \ldots dx_n.$$

Максимизируя dF_n по $(\mu, \sigma^2) \in \Omega$ и по $(\mu, \sigma^2) \in \omega$, беря затем в соответствии с (13.3.5) отношение этих максимумов, получим

$$\lambda_{\mathscr{H}} = \left(\frac{S_\Omega}{S_\omega}\right)^{\frac{1}{2}n},$$

где S_Ω — минимум по μ суммы $\sum_{\xi=1}^{n}(x_\xi - \mu)^2$, т. е.

$$S_\Omega = \sum_{\xi=1}^{n}(x_\xi - \bar{x})^2,$$

а

$$S_\omega = \sum_{\xi=1}^{n}(x_\xi - \mu_0)^2.$$

Отметим, что величина $\lambda_{\mathscr{H}}$ может быть записана в виде

$$\lambda_{\mathscr{H}} = \left(1 + \frac{t^2}{n-1}\right)^{-\frac{1}{2}n},$$

где

$$t = \sqrt{n}\,(\bar{x} - \mu_0)/s,$$

а $\bar{x}$ и s^2 — выборочные среднее и дисперсия, определенные в (8.2.1) и (8.2.6). Случайная величина t является, конечно, дробью Стьюдента, введенной в (8.4.16). Поэтому критическая область W_α объема α критерия отношения правдоподобия имеет вид

$$\lambda_{\mathscr{H}} < \lambda_\alpha,$$

где

$$\lambda_\alpha = \left(1 + \frac{t_\alpha^2}{n-1}\right)^{-\frac{1}{2}n}$$

и t_α определяется так, чтобы

$$\int_{-t_\alpha}^{+t_\alpha} f_{n-1}(t)\, dt = 1 - \alpha.$$

Здесь $f_{n-1}(t)$ — п. в. распределения Стьюдента $S(n-1)$, определенная в (7.8.4). Построенная выше критическая область обладает свойством

$$P(W_\alpha) = \alpha \text{ для } \mu = \mu_0.$$

Итак, критерий отношения правдоподобия $\lambda_{\mathscr{H}}$ гипотезы $\mathscr{H}$ эквивалентен t-критерию Стьюдента.

(b) Критерий отношения правдоподобия в теории нормальной регрессии. Одним из наиболее важных приложений критерия отношения правдоподобия в выборках конечного объема является проверка различных гипотез о параметрах нормального распределения. Предыдущий пример в этом отношении типичен. Во многих таких приложениях удается получить точное выборочное распределение отношения правдоподобия для конечных n. Бо́льшая часть тестов такого рода строится непосредственно, и мы их рассмотрим только в качестве задач в конце этой главы.

Однако заслуживает подробного рассмотрения критерий отношения правдоподобия для проверки гипотезы о равенстве нулю некоторых коэффициентов в схеме нормальной регрессии. Как в § 10.3 (c), будем считать y_ξ, $\xi = 1, \ldots, n$, независимыми случайными величинами, имеющими распределения $N(\beta_1 x_{1\xi} + \ldots + \beta_k x_{k\xi}, \sigma^2)$, $\xi = 1, \ldots, n$, где $(x_{i\xi}, i = 1, \ldots, k)$ — известные линейно независимые векторы (т. е. матрица $\| a_{ij} \|$ в (10.3.8) неособая). Пусть $\mathscr{K}$ — гипотеза, в которой параметрическое множество Ω — $(k+1)$-мерное эвклидово полупространство

$$-\infty < \beta_i < +\infty, \qquad i = 1, \ldots, k, \ \sigma^2 > 0; \tag{13.3.7}$$

ω — подмножество Ω, на котором

$$\beta_{k'+1} = \ldots = \beta_k = 0, \qquad k' < k.$$

Определенную таким образом гипотезу $\mathscr{K}$ иногда называют *общей линейной гипотезой* теории нормальной регрессии. Функция правдоподобия выборки равна

$$dF_n = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{\xi=1}^{n}(y_\xi - \beta_1 x_{1\xi} - \ldots - \beta_k x_{k\xi})^2\right] dy_1 \ldots dy_n. \tag{13.3.8}$$

Определяя $(dF_n)_\Omega$ и $(dF_n)_\omega$ обычной операцией дифференцирования, получим

$$\lambda_{\mathscr{H}}=\frac{(dF_n)_\omega}{(dF_n)_\Omega}=\left(\frac{S_\Omega}{S_\omega}\right)^{\frac{1}{2}n}, \tag{13.3.9}$$

где S_Ω — минимум суммы квадратов

$$\sum_{\xi=1}^{n}(y_\xi-\beta_1 x_{1\xi}-\ldots-\beta_k x_{k\xi})^2 \tag{13.3.10}$$

по $\beta_1, \ldots, \beta_k$, а S_ω — минимум

$$\sum_{\xi=1}^{n}(y_\xi-\beta_1 x_{1\xi}-\ldots-\beta_{k'} x_{k'\xi})^2 \tag{13.3.11}$$

по $\beta_1, \ldots, \beta_{k'}$.

Обращаясь к § 10.3 (b), найдем, что S_Ω совпадает с S_1 из (10.3.27), так что

$$S_\Omega=\frac{|a_{i_0 j_0}|}{|a_{ij}|}, \tag{13.3.12}$$

где a_{ij}, $i, j=1, \ldots, k$ и $a_{i_0 j_0}$ определены в (10.3.8), (10.3.11) и (10.3.28); $|a_{ij}|$ — определитель матрицы $\|a_{ij}\|$, а величина $|a_{i_0 j_0}|$ определена в (10.3.28).

Точно так же

$$S_\omega=\frac{|a_{i'_0 j'_0}|}{|a_{i'j'}|}, \tag{13.3.13}$$

где $|a_{i'_0 j'_0}|$ и $|a_{i'j'}|$ аналогичны $|a_{i_0 j_0}|$ и $|a_{ij}|$ с i', $j'=1, \ldots, k'$. С помощью теоремы Кочрэна **8.4.4** можно вывести, что при нулевой гипотезе S_Ω/σ^2 и $(S_\omega-S_\Omega)/\sigma^2$ независимы и имеют соответственно χ^2-распределения $C(n-k)$ и $C(k-k')$. Мы не останавливаемся на деталях, поскольку техника вполне аналогична использованной при доказательстве того, что статистики S_1/σ^2 и S_2/σ^2, определенные в (10.3.30), независимы и имеют соответственно χ^2-распределения $C(n-k)$ и $C(k)$.

Таким образом, отношение правдоподобия $\lambda_{\mathscr{H}}$ из (13.3.9) может быть записано в виде

$$\lambda_{\mathscr{H}}=\left(1+\frac{k-k'}{n-k}F\right)^{-\frac{1}{2}n}, \tag{13.3.14}$$

где

$$F=\frac{(n-k)(S_\omega-S_\Omega)}{(k-k')S_\Omega} \tag{13.3.15}$$

имеет распределение Снедекора $S(k-k', n-k)$, элемент вероятности которого определен в (7.8.9). Поскольку имеется взаимно однозначное соответствие между $\lambda_{\mathcal{H}}$ и F, то критерий отношения правдоподобия гипотезы $\mathcal{H}$ эквивалентен F-критерию. Критическая область $\lambda_{\mathcal{H}} < \lambda_\alpha$ в пространстве значений статистики $\lambda_{\mathcal{H}}$, где λ_α — 100α-процентная квантиль распределения $\lambda_{\mathcal{H}}$, соответствует критической области $F > [(n-k)/(k-k')](\lambda_\alpha^{-\frac{2}{n}} - 1)$ в пространстве значений F. Описанный только что критерий F является наиболее общей формой *критерия Модели* I *дисперсионного анализа.* Дэйли (1940) показал, что этот тест несмещенный.

Итак, мы получили следующий основной результат относительно общей линейной гипотезы в теории нормальной регрессии.

13.3.1. *Пусть* y_ξ, $\xi = 1, \ldots, n$, — *независимые случайные величины, имеющие нормальные распределения* $N(\beta_1 x_{1\xi} + \ldots + \beta_k x_{k\xi}, \sigma^2)$, $\xi = 1, \ldots, n$, *где матрица* $\| a_{ij} \|_{i, j=1, \ldots, k}$, *определенная в* (10.3.8), *неособая. Тогда критерий отношения правдоподобия* $\lambda_{\mathcal{H}}$ *гипотезы* $\mathcal{H}$, *заданной условиями* (13.3.7), *имеет вид* (13.3.4), *где* F *определяется формулой* (13.3.15). *Если гипотеза* $\mathcal{H}$ *верна, то статистика* $[(n-k)/(k-k')](\lambda_{\mathcal{H}}^{-\frac{2}{n}} - 1)$ *имеет распределение Снедекора* $S(k-k', n-k)$.

Общая линейная гипотеза охватывает большее число ситуаций, чем это может показаться на первый взгляд. Пусть, например, мы желаем проверить гипотезу $\mathcal{H}'$: $\beta_{k'+1} = \beta_{k'+1, 0}, \ldots, \beta_k = \beta_{k0}$.

Введем в рассмотрение случайные величины

$$y'_\xi = y_\xi - \beta_{k'+1, 0} x_{k'+1, \xi} - \ldots - \beta_{k0} x_{k\xi}.$$

Структура критерия отношения правдоподобия $\lambda_{\mathcal{H}'}$ гипотезы $\mathcal{H}'$ совпадает со структурой критерия $\lambda_{\mathcal{H}}$, в котором в матрицах $\| a_{i'_0 j'_0} \|$ и $\| a_{i'j'} \|$ y_ξ заменены на y'_ξ. Вывод распределения $\lambda_{\mathcal{H}'}$ при нулевой гипотезе точно такой же, как и для $\lambda_{\mathcal{H}}$.

Можно также проверить гипотезу $\mathcal{H}''$ о том, что между $\beta_1, \ldots, \beta_k$ существуют определенные линейно независимые соотношения вида

$$\sum_{i=1}^{k} c_{iu} \beta_i = \gamma_{u0}, \qquad u = k'+1, \ldots, k,$$

где c_{iu} и γ_{u0} — известные числа. В этом случае можно использовать следующее преобразование коэффициентов регрессии:

$$\beta_1 = \gamma_1, \ldots, \quad \beta_{k'} = \gamma_{k'},$$
$$\sum_{i=1}^{k} c_{iu} \beta_i = \gamma_u, \qquad u = k'+1, \ldots, k,$$

в результате которого $\beta_1 x_{1\xi} + \ldots + \beta_k x_{k\xi}$ перейдет в $\gamma_1 z_{1\xi} + \ldots + \gamma_k z_{k\xi}$, где $z_{1\xi}, \ldots, z_{k\xi}$ — линейные функции от $x_{1\eta}, \ldots, x_{k\eta}$,

$\eta = 1, \ldots, n$, с коэффициентами, зависящими от c_{pu}. Гипотеза $\mathscr{H}''$ сведется тогда к гипотезе $\mathscr{H}'$ рассмотренного в предыдущем абзаце типа; $\gamma_1, \ldots, \gamma_k$ будут играть роль $\beta_1, \ldots, \beta_k$, а $z_{1\xi}, \ldots, z_{k\xi}$ — роль $x_{1\xi}, \ldots, x_{k\xi}$, $\xi = 1, \ldots, n$. Выборочная теория теста $\lambda_{\mathscr{H}''}$, когда $\mathscr{H}''$ верна, точно такая же, как $\lambda_{\mathscr{H}}$.

13.4. Асимптотическое распределение отношения правдоподобия в больших выборках

Мы перейдем теперь к задачам в больших выборках. Прежде всего будет рассмотрен случай простой гипотезы

$$\mathscr{H}(\theta_0;\ \Omega_0), \tag{13.4.1}$$

которая всюду в этом параграфе будет обозначаться просто через $\mathscr{H}$, где θ — одномерный параметр, параметрическое множество, Ω — интервал на вещественной прямой R_1, содержащий некоторый (открытый) интервал Ω_0 и $\theta_0 \in \Omega_0$. Мы увидим, что в асимптотической выборочной теории отношения правдоподобия существенную роль играет только обозначенная через Ω_0 часть Ω. Предположим, что оценка максимального правдоподобия $\hat{\theta}_n$ параметра при объеме выборки $n \to \infty$ сходится по вероятности к θ_0. Соответствующая последовательность тестов отношения правдоподобия есть

$$\lambda_{\mathscr{H}} = \frac{\prod\limits_{\xi=1}^{n} dF(x_\xi;\ \theta_0)}{\prod\limits_{\xi=1}^{n} dF(x_\xi;\ \hat{\theta}_n)}. \tag{13.4.2}$$

Покажем, что при сформулированных ниже условиях распределение

$$-2\log\lambda_{\mathscr{H}}, \qquad n = 1, 2, \ldots,$$

сходится при нулевой гипотезе к χ^2-распределению $C(1)$.

Предположим, что $F(x;\ \theta)$ регулярна в смысле первой и второй производных по θ, $\theta \in \Omega_0$ (см. § 12.1). При этих условиях функцию

$$H(\theta_0,\ \theta) = \int\limits_{-\infty}^{+\infty} \log dF(x;\ \theta)\, dF(x;\ \theta_0) \tag{13.4.3}$$

можно дважды дифференцировать по θ под знаком интеграла и при этом

$$\begin{aligned} &H'(\theta_0,\ \theta_0) = 0,\\ &H''(\theta_0,\ \theta_0) = -B^2(\theta_0,\ \theta_0), \end{aligned} \tag{13.4.4}$$

где величина $B^2(\theta_0,\ \theta_0)$ определена в (12.1.8).

Следовательно, функция $H(\theta_0, \theta)$ имеет относительный максимум $H(\theta_0, \theta_0)$ при $\theta = \theta_0$. Функция $H(\theta_0, \theta)$ играет основную роль в развитой Калбэком (1959) статистической теории информации. Вообще, если $F_1(x)$, $F_2(x)$ и $G(x)$ — три взаимно абсолютно непрерывные ф. р., то *среднее количество информации* или *интеграл информации* Калбэка для различения между $F_1(x)$ и $F_2(x)$ по наблюдению над $G(x)$ определяется как $I(1;\, 2) = \int\limits_{-\infty}^{+\infty} \log \frac{dF_1(x)}{dF_2(x)}\, dG(x)$. Таким образом,

$$H(\theta_0, \theta_0) - H(\theta_0, \theta_1) = \int\limits_{-\infty}^{+\infty} \log \frac{dF(x;\, \theta_0)}{dF(x;\, \theta_1)}\, dF(x;\, \theta_0)$$

представляет собой, по Калбэку, интеграл информации для различения $F(x;\, \theta_0)$ и $F(x;\, \theta_1)$ по наблюдению над $F(x;\, \theta_0)$.

Замечание. Частный случай функции $H(\theta, \theta)$ впервые появился в статистической механике. Пусть x — точка, компоненты которой представляют собой координаты и моменты отдельной молекулы газа. Пространство точек x называется *фазовым пространством* системы. Если $F(x;\, t)$ — ф. р. точки x в системе в момент времени t, то интеграл $H(t,\, t)$, определяемый формулой (13.4.3) для $F(x;\, t)$, будет H-функцией Больцмана (1910), введенной им для приближения $-\log P$ (с точностью до аддитивной постоянной), где

$$P = \frac{n!}{n_1!\, n_2! \ldots n_N!} \left(\frac{1}{N}\right)^n$$

— вероятность того, что из n молекул системы n_i находятся в объеме E_i, $i = 1, \ldots, N$, в момент t. Объемы $E_1, \ldots, E_N$ имеют одинаковые вероятности $\frac{1}{N}$, не пересекаются и вместе составляют все фазовое пространство. $H(t,\, t)$ измеряет, по существу, отклонение системы от «наиболее вероятного» или «равновесного» состояния.

Функции типа $H(\theta_0, \theta)$ играют также важную роль в построенной Шенноном (1948) теории передачи сообщений.

Напомним сказанное в § 12.3 (с): величина $B^2(\theta_0, \theta_0)$ введена Фишером под названием *количество информации относительно* θ_0, *содержащееся в наблюдении над* $F(x;\, \theta_0)$. Из (13.4.4) замечаем, что фишеровское информационное количество совпадает со второй производной *интеграла информации* Калбэка $[H(\theta_0, \theta_0) - H(\theta_0, \theta)]$ при $\theta = \theta_0$. Другими словами, $B^2(\theta_0, \theta_0)$ измеряет кривизну функции $H(\theta_0, \theta)$ в точке $\theta = \theta_0$ ее максимального значения.

Суммируем все сказанное.

13.4.1. *Если $F(x;\, \theta)$ регулярна в смысле первой и второй производных по θ при $\theta \in \Omega_0$ то $H(\theta_0, \theta)$ имеет относительный максимум $H(\theta_0, \theta_0)$ при $\theta = \theta_0$. Вторая производная интеграла информации Калбэка $[H(\theta_0, \theta_0) - H(\theta_0, \theta)]$ при $\theta = \theta_0$ равна $B^2(\theta_0, \theta_0)$, количеству информации Фишера относительно θ_0, содержащемуся в наблюдении над $F(x;\, \theta_0)$.*

Если $(x_1, \ldots, x_n)$ — выборка из совокупности $F(x;\ \theta_0)$, то

$$\frac{1}{n}\sum_{\xi=1}^{n} \log dF(x_\xi;\ \hat{\theta}) \tag{13.4.5}$$

будет состоятельной оценкой величины $H(\theta_0,\ \theta_0)$.

Действительно, из теоремы **9.1.1** заключаем, что

$$\frac{1}{n}\sum_{\xi=1}^{n} \log dF(x_\xi;\ \theta_0) \tag{13.4.6}$$

сходится по вероятности к $H(\theta_0,\ \theta_0)$. Если $\hat{\theta}(x_1, \ldots, x_n)$, $n=1, 2, \ldots$, сходится по вероятности к θ_0 (а это имеет место в условиях теоремы **12.3.2**), то из **4.3.8** получаем, что

$$\frac{1}{n}\sum_{\xi=1}^{n} \log dF(x_\xi,\ \hat{\theta})$$

сходится по вероятности к $H(\theta_0,\ \theta_0)$.

Мы доказали, таким образом, следующее:

13.4.2. *Если $H(\theta_0,\ \theta_0)$ существует и при $n \to \infty$ $\hat{\theta}$ сходится по вероятности к θ_0, то случайные величины*

$$\frac{1}{n}\sum_{\xi=1}^{n} \log dF(x_\xi;\ \theta_0)$$

и

$$\frac{1}{n}\sum_{\xi=1}^{n} \log dF(x_\xi;\ \hat{\theta})$$

обе сходятся по вероятности к $H(\theta_0,\ \theta_0)$.

Если же $\hat{\theta}$ имеет асимптотически нормальное распределение, то справедлива следующая теорема об асимптотическом распределении $-2\log\lambda_{\mathscr{H}}$ при нулевой гипотезе.

13.4.3. *Если $\hat{\theta}$ распределена асимптотически нормально в соответствии с теоремой* **12.3.3**, *то*

$$\lim_{n\to\infty} P(-2\log\lambda_{\mathscr{H}} < \chi^2 \mid \theta_0) = \frac{1}{\sqrt{2\pi}}\int_0^{\chi^2} u^{-\frac{1}{2}} e^{-\frac{1}{2}u}\, du, \tag{13.4.7}$$

т. е. если гипотеза $\mathscr{H}$ верна, то распределение $-2\log\lambda_{\mathscr{H}}$ сходится при $n\to\infty$ к χ^2-распределению $C(1)$.

Чтобы доказать **13.4.3**, обратимся к условиям теоремы **12.3.3**. Так как $\hat{\theta}$ сходится почти наверное к θ_0 при $n\to\infty$, то по любому $\varepsilon > 0$

можно указать n_ε так, чтобы при $n > n_\varepsilon$ с вероятностью, не меньшей $1-\varepsilon$, выполнялось соотношение

$$\sum_{\xi=1}^{n} \log dF(x_\xi;\ \theta_0) = \sum_{\xi=1}^{n} \log dF(x_\xi;\ \hat\theta) + \\ + \frac{1}{2}\sum_{\xi=1}^{n}\left[\frac{\partial^2}{\partial\theta^2}\log dF(x_\xi;\ \theta)\right]_{\theta=\theta^*}\cdot(\theta_0-\hat\theta)^2, \tag{13.4.8}$$

где θ^* — случайная величина с условием $|\theta_0-\theta^*| < |\theta_0-\hat\theta|$. Но (13.4.8) может быть записано в виде

$$-2\sum_{\xi=1}^{n}\log\frac{dF(x_\xi;\ \theta_0)}{dF(x_\xi;\ \hat\theta)} = \\ = -\frac{1}{n}\sum_{\xi=1}^{n}\left[\frac{\partial^2}{\partial\theta^2}\log dF(x_\xi;\ \theta)\right]_{\theta=\theta^*}[\sqrt{n}(\theta_0-\hat\theta)]^2. \tag{13.4.9}$$

Мы видим, что левая часть (13.4.9) есть $-2\log\lambda_{\mathcal{H}}$. Поскольку (13.4.9) с вероятностью, не меньшей $1-\varepsilon$, имеет место для всех $n > n_\varepsilon$, то левая и правая части (13.4.9) распределены асимптотически одинаково. Так как

$$-\frac{1}{n}\sum_{\xi=1}^{n}\left[\frac{\partial^2}{\partial\theta^2}\log dF(x_\xi;\ \theta)\right]_{\theta=\theta^*} \tag{13.4.10}$$

сходится по вероятности к $B^2(\theta_0,\ \theta_0)$, а $\sqrt{n}B(\theta_0,\ \theta_0)(\theta_0-\hat\theta)$ распределена асимптотически нормально $N(0,\ 1)$, то распределение $-2\log\lambda_{\mathcal{H}}$ сходится к $C(1)$. Следовательно, теорема **13.4.3** доказана.

13.5. Состоятельность критерия отношения правдоподобия

Пусть W_α — множество в выборочном пространстве R_n, на котором

$$-2\log\lambda_{\mathcal{H}} > \chi^2_\alpha, \tag{13.5.1}$$

где $\mathcal{H} = \mathcal{H}(\theta_0;\ \Omega_0)$, а χ^2_α выбирается из таблиц χ^2-распределения $C(1)$ так, чтобы $P(\chi^2 > \chi^2_\alpha) = \alpha$. Если $F(x;\ \theta)$ регулярна в смысле первой и второй производных по θ в Ω_0, то согласно теореме **13.4.3**

$$\lim_{n\to\infty} P(W_\alpha\,|\,\theta_0) = \alpha. \tag{13.5.2}$$

Пусть теперь $\theta_1 \neq \theta_0$ — произвольная точка в Ω_0. Рассмотрим

$$\lim_{n\to\infty} P(W_\alpha\,|\,\theta_1). \tag{13.5.3}$$

Имеем
$$\lambda_{\mathcal{H}} = \lambda'_{\mathcal{H}} \cdot \nu, \tag{13.5.4}$$
где
$$\lambda'_{\mathcal{H}} = \prod_{\xi=1}^{n} \frac{dF(x_\xi;\ \theta_1)}{dF(x_\xi;\ \theta)}, \qquad \nu = \prod_{\xi=1}^{n} \frac{dF(x_\xi;\ \theta_0)}{dF(x_\xi;\ \theta_1)}.$$
Если положить
$$-2\log\lambda'_{\mathcal{H}} = u_n, \qquad -2\log\nu = nv_n, \tag{13.5.5}$$
то W_α будет множеством в R_n, на котором
$$u_n + nv_n > \chi^2_\alpha. \tag{13.5.6}$$

Если истинным значением θ является θ_1, то по теореме **13.4.3** распределение u_n, $n = 1, 2, \ldots$, сходится к χ^2-распределению $C(1)$. Далее,
$$v_n = 2\left[\frac{1}{n}\sum_{\xi=1}^{n} \log dF(x_\xi;\ \theta_1) - \frac{1}{n}\sum_{\xi=1}^{n} \log dF(x_\xi;\ \theta_0)\right]. \tag{13.5.7}$$
Из **9.1.1** получаем, что $\frac{1}{n}\sum_{\xi=1}^{n} \log dF(x_\xi;\ \theta_1)$ и $\frac{1}{n}\sum_{\xi=1}^{n} \log dF(x_\xi;\ \theta_0)$ сходятся по вероятности к $H(\theta_1,\ \theta_1)$ и $H(\theta_1,\ \theta_0)$ соответственно, когда истинное значение θ есть θ_1. Но из **13.4.1** мы знаем, что $H(\theta_1, \theta_1) > H(\theta_1,\ \theta_0)$. Поэтому последовательность случайных величин v_n, $n = 1, 2, \ldots$, сходится по вероятности к положительному числу $v_0 = 2[H(\theta_1,\ \theta_1) - H(\theta_1,\ \theta_0)]$. Поскольку распределение u_n, $n = 1, 2, \ldots$, сходится к $C(1)$, то
$$\lim_{n\to\infty} P(u_n + nv_n > \chi^2_\alpha) = 1, \tag{13.5.8}$$
или иначе
$$\lim_{n\to\infty} P(W_\alpha \mid \theta = \theta_1) = 1, \tag{13.5.9}$$
т. е. W_α является состоятельным тестом гипотезы $\mathcal{H}(\theta_0;\ \Omega_0)$.

Таким образом, для $\theta \in \Omega_0$ последовательность функций мощности
$$P(W_\alpha \mid \theta), \qquad n = 1, 2, \ldots, \tag{13.5.10}$$
сходится к функции
$$\zeta(\theta) = \begin{cases} \alpha, & \theta = \theta_0, \\ 1, & \theta \neq \theta_0, \end{cases} \tag{13.5.11}$$
что означает состоятельность W_α как теста гипотезы $\mathcal{H}$.

Суммируем полученный результат в виде теоремы

13.5.1. *Если $F(x;\ \theta)$ регулярна в смысле первой и второй производных по θ, то критерий отношения правдоподобия* (13.5.1) *гипотезы $\mathcal{H}(\theta_0;\ \Omega_0)$ состоятелен.*

13.6. Асимптотическая мощность критерия отношения правдоподобия

В предыдущем параграфе было показано, что функция мощности теста отношения правдоподобия гипотезы $\mathscr{H}(\theta_0;\ \Omega_0)$ сходится при $n\to\infty$ к функции $\zeta(\theta)$, определенной в (13.5.11). Покажем теперь, что при определенных условиях функция мощности никакого другого состоятельного теста гипотезы $\mathscr{H}(\theta_0;\ \Omega_0)$ объема α не может сходиться к $\zeta(\theta)$ быстрее, чем для теста, задаваемого статистикой $\lambda_{\mathscr{H}}$.

С этой целью вернемся к введенным в § 12.5 регулярным оценочным функциям $g_n(x_1,\ \ldots,\ x_n;\ \theta)$; свойства этих функций гарантируют, что при нулевой гипотезе случайные величины последовательностей (12.5.11) и (12.5.12) имеют асимптотически нормальное распределение $N(0,\ 1)$. Следовательно, распределение величины

$$ng_n^2(x_1,\ \ldots,\ x_n;\ \theta_0),\qquad n=1,\ 2,\ldots, \tag{13.6.1}$$

сходится при $n\to\infty$ к χ^2-распределению $C(1)$. Пусть критическая область W_α^* имеет вид

$$ng_n^2(x_1,\ \ldots,\ x_n;\ \theta_0)>\chi_\alpha^2, \tag{13.6.2}$$

где χ_α^2 — то же число, что и в (13.5.1). Тогда

$$\lim_{n\to\infty} P(W_\alpha^*\mid\theta_0)=\alpha. \tag{13.6.3}$$

Из свойств $g_n(x_1,\ \ldots,\ x_n;\ \theta)$ как регулярной оценочной функции следует, что если истинное значение θ есть $\theta_1\neq\theta_0$, $\theta_1\in\Omega_0$, то распределения случайных величин

$$ng_n^2(x_1,\ \ldots,\ x_n;\ \theta_0),\qquad n=1,\ 2,\ \ldots, \tag{13.6.4}$$

и

$$w_n^*=\{g_n'(x_1,\ \ldots,\ x_n;\ \theta^*)[\sqrt{n}(\theta_1-\tilde\theta)+\sqrt{n}(\theta_0-\theta_1)]\}^2 \tag{13.6.5}$$

сходятся или нет одновременно. Но

$$w_n^*=u_n^*+nv_n^*, \tag{13.6.6}$$

где

$$u_n^*=[g_n'(x_1,\ \ldots,\ x_n;\tilde\theta^*)\sqrt{n}(\theta_1-\tilde\theta)]^2, \tag{13.6.7}$$

$$v_n^*=\frac{1}{n}[w_n^*-u_n^*].$$

Так как истинным значением θ является θ_1, то распределение u_n^*, $n=1,\ 2,\ \ldots$, сходится к χ^2-распределению $C(1)$, а v_n^*, $n=1,\ 2,\ \ldots$, сходится по вероятности к постоянной v_0^*, где

$$v_0^*=B^{*2}(\theta_1,\ \theta_1)(\theta_0-\theta_1)^2. \tag{13.6.8}$$

При этом $v_0^*>0$, поскольку $\theta_1\neq\theta_0$ и $B^{*2}(\theta_1,\ \theta_1)>0$. Следовательно,

$$\lim_{n\to\infty} P(u_n^*+nv_n^*>\chi_\alpha^2)=1, \tag{13.6.9}$$

что эквивалентно утверждению

$$\lim_{n\to\infty} P(W_\alpha^* \mid \theta=\theta_1)=1. \tag{13.6.10}$$

Соотношения (13.6.3) и (13.6.10) как раз означают, что тест (13.6.2) состоятелен для проверки гипотезы $\mathscr{H}$.

Задача сравнения функций мощности критерия (13.5.1), основанного на отношении правдоподобия, и произвольного критерия W_α^* вида (13.6.2) осложняется тем обстоятельством, что оба эти критерия состоятельны и имеют поэтому одну и ту же предельную функцию мощности $\zeta(\theta)$, определенную в (13.5.11). Чтобы обойти эту трудность, мы введем и используем некоторые новые понятия.

Если W_α — тест гипотезы $\mathscr{H}$, предельная функция мощности которого $\eta(\theta)$ удовлетворяет условиям:

$$\begin{gathered}\eta(\theta_0)=\alpha,\\ \alpha<\eta(\theta)<1,\ \theta\neq\theta_0,\quad \theta\in\Omega_0,\end{gathered} \tag{13.6.11}$$

то будем называть W_α *асимптотически несмещенным тестом* гипотезы $\mathscr{H}(\theta_0;\ \Omega_0)$ с уровнем значимости α. Заметим, что асимптотическая несмещенность теста — свойство более слабое, нежели состоятельность.

Пусть $W_{1\alpha}$ и $W_{2\alpha}$ — два асимптотически несмещенных теста объема α гипотезы $\mathscr{H}$, предельные функции мощности которых суть $\eta_1(\theta)$ и $\eta_2(\theta)$ соответственно.

Если

$$\eta_1(\theta)\geqslant\eta_2(\theta) \tag{13.6.12}$$

и $\eta_1(\theta_0)=\eta_2(\theta_0)=\alpha$, то тест $W_{1\alpha}$ называется *асимптотически более мощным**), чем $W_{2\alpha}$. Если в (13.6.12) для всех $\theta\in\Omega_0$ имеет место знак равенства, то $W_{1\alpha}$ и $W_{2\alpha}$ называются ***эквивалентными асимптотически несмещенными тестами*** объема α.

Вернемся теперь к задаче сравнения мощностей критериев W_α и W_α^*, определенных соответственно неравенствами (13.5.1) и (13.6.2). Рассмотрим тесты $W_{c\alpha}$ и $W_{c\alpha}^*$, задаваемые условиями:

$$u_n+cv_n>\chi_\alpha^2,\qquad u_n^*+cv_n^*>\chi_\alpha^2 \tag{13.6.13}$$

соответственно, где постоянная $c>0$ произвольна. Мы покажем, что при каждом $c>0$ $W_{c\alpha}$ имеет асимптотически бо́льшую мощность, чем $W_{c\alpha}^*$. Употребляя терминологию Вальда (1941a), мы скажем в этом случае, что W_α *асимптотически более точный* тест гипотезы $\mathscr{H}$, чем W_α^*.

---

*) Строго говоря, $W_{1\alpha}$ следовало бы назвать *асимптотически не менее мощным*, чем $W_{2\alpha}$, если только не существует хотя бы одной точки θ, для которой в (13.6.12) имеет место знак строгого неравенства. Но употребление более краткого термина не вызывает недоразумений.

Если $W_{c\alpha}$ и $W^*_{c\alpha}$ асимптотически эквивалентны при каждом $c > 0$, то будем говорить, что W_α и W^*_α асимптотически одинаково точны.

Важный случай асимптотически одинаково точных тестов описывается теоремой

13.6.1. *Пусть* W_α — *тест, определенный в* (13.5.1), *а* W^*_α — *тест, получающийся из* (13.6.2) *заменой* $g_n(x_1, \ldots, x_n; \theta_0)$ *на* $h_n(x_1, \ldots, x_n; \theta_0)$, *где* $h_n(x_1, \ldots, x_n; \theta_0)$ *определяется левой частью* (12.5.1). *Тогда, если* $F(x; \theta)$ *регулярна в смысле первых двух производных по* θ *в* Ω_0, *то тесты* W_α *и* W^*_α *гипотезы* $\mathscr{H}$ *асимптотически одинаково точны.*

В § 13.5 мы, по существу, доказали, что распределение пары (u_n, v_n) случайных величин из (13.6.13) сходится при $n \to \infty$ к вырожденному распределению в плоскости (u, v) которое представляет собой χ^2-распределение $C(1)$ на полупрямой $v = v_0$, $u > 0$, где

$$v_0 = 2\,[H(\theta_1, \theta_1) - H(\theta_1, \theta_0)]. \tag{13.6.14}$$

Аналогично было показано, что распределение пары случайных величин (u^*_n, v^*_n) сходится при $n \to \infty$ к χ^2-распределению $C(1)$ на полупрямой $v = v^*_0$, $u > 0$, где

$$v^*_0 = B^{*2}(\theta_1, \theta_1)(\theta_0 - \theta_1)^2. \tag{13.6.15}$$

Но

$$H(\theta_1, \theta_0) = H(\theta_1, \theta_1) + H'(\theta_1, \theta_1)(\theta_0 - \theta_1) + \\ + \frac{1}{2} H''(\theta_1, \theta^*_1)(\theta_0 - \theta_1)^2, \tag{13.6.16}$$

де $|\theta_0 - \theta^*_1| < |\theta_0 - \theta_1|$. Поскольку $H'(\theta_1, \theta_1) = 0$, имеем

$$2\,[H(\theta_1, \theta_1) - H(\theta_1, \theta_0)] = -H''(\theta_1, \theta^*_1)(\theta_0 - \theta_1)^2. \tag{13.6.17}$$

При этом

$$-H''(\theta_1, \theta_1) = B^2(\theta_1, \theta_1),$$

и так как $F(x; \theta)$ регулярна в смысле первой и второй производных по θ в Ω_0, имеем следующий результат для $\theta = \theta_1$:

$$\frac{B^{*2}(\theta_1, \theta_1)}{B^2(\theta_1, \theta_1)} \leqslant 1, \tag{13.6.18}$$

который аналогичен (12.5.21) при $\theta = \theta_0$. Следовательно, в силу непрерывности $H''(\theta_1, \theta)$ по θ, $\theta \in \Omega_0$, получаем, что для достаточно близких друг к другу θ_0 и θ_1 (13.6.18) будет иметь место с заменой $B^2(\theta_1, \theta_1)$ на $-H''(\theta_1, \theta^*_1)$, откуда в свою очередь следует, что

$$0 < v^*_0 \leqslant v_0. \tag{13.6.19}$$

Рассмотрим теперь функции мощности критериев $W_{c\alpha}$ и $W^*_{c\alpha}$ для фиксированных значений c, т. е.

$$P(u_n + cv_n > \chi^2_\alpha \mid \theta_1) \quad \text{и} \quad P(u^*_n + cv^*_n > \chi^2_\alpha \mid \theta_1) \tag{13.6.20}$$

как функции θ_1. Обозначим пределы этих функций при $n \to \infty$, другими словами, асимптотические функции мощности тестов $W_{c\alpha}$ и $W_{c\alpha}^*$, через

$$\eta_c(\theta_1) \quad \text{и} \quad \eta_c^*(\theta_1). \tag{13.6.21}$$

Так как u_n и u_n^* неотрицательны, а распределения векторов (u_n, v_n) и (u_n^*, v_n^*) сходятся при $n \to \infty$ соответственно к χ^2-распределениям $C(1)$ на полупрямых $v = v_0$, $u > 0$ и $v = v_0^*$, $u > 0$, то

$$\begin{aligned}
&\text{(i)} \quad \eta_c(\theta_1) = \eta_c^*(\theta_1) = 1 \text{ для тех } \theta_1, \text{ для которых } v_0^* \geqslant \frac{\chi_\alpha^2}{c};\\
&\text{(ii)} \; 1 > \eta_c(\theta_1) > \eta_c^*(\theta_1) > \alpha \text{ для тех } \theta_1, \text{ для которых } 0 < v_0^* < \frac{\chi_\alpha^2}{c};\\
&\text{(iii)} \; \eta_c(\theta_0) = \eta_c^*(\theta_0) = \alpha \text{ для } \theta_1 = \theta_0, \text{ т. е. для того } \theta_1,\\
&\qquad \text{для которого } v_0^* = 0.
\end{aligned} \tag{13.6.22}$$

Но из (13.6.8) следует, что три множества, фигурирующих в (13.6.22), состоят из значений θ_1: (i), принадлежащих Ω_0, но лежащих вне некоторого интервала Ω_0', содержащего θ_0; (ii) отличных от θ_0 и лежащих внутри Ω_0'; (iii) одного θ_0.

Таким образом, поскольку асимптотические функции мощности $\eta_c(\theta)$ и $\eta_c^*(\theta)$ удовлетворяют условию (13.6.11) для каждого $c > 0$ и неравенству (13.6.12), то тест $W_{c\alpha}$ имеет асимптотически бóльшую мощность, чем $W_{c\alpha}^*$, и, следовательно, W_α *асимптотически более точный*, чем W_α^*.

Если $v_0^* = v_0$, то (13.6.22) останется в силе, если заменить (ii) на

$$1 > \eta_c(\theta_1) = \eta_c^*(\theta_1) > \alpha$$

для значений θ_1, для которых $0 < v_0^* < \frac{\chi_\alpha^2}{c}$, что означает эквивалентность асимптотически несмещенных тестов $W_{c\alpha}$ и $W_{c\alpha}^*$ *при каждом* $c > 0$. В этом случае W_α и W_α^* асимптотически одинаково точные. Но $v_0^* = v_0$ тогда и только тогда, когда знак равенства имеет место в (13.6.18), что в свою очередь имеет место в том и только том случае, если регулярная оценочная функция $g_n(x_1, \ldots, x_n; \theta)$, используемая в построении W_α^*, заменена на $h_n(x_1, \ldots, x_n; \theta)$. Но из теоремы **13.6.1** мы знаем, что использование $h_n(x_1, \ldots, x_n; \theta)$ в (13.6.2) дает тест W_α^*, асимптотически столь же точный, что и тест W_α из (13.5.1).

Объединим полученные результаты в форме следующей важной теоремы.

13.6.2. *Пусть* $(x_1, \ldots, x_n)$ — *выборка из совокупности* $F(x; \theta)$, *где* $F(x; \theta)$ *регулярна в* Ω_0 *в смысле первой и второй производных по* θ. *Если* W_α — *тест отношения правдоподобия объема* α, *определенный в* (13.5.1), $g_n(x_1, \ldots, x_n; \theta)$ — *любая регулярная оценочная функция для* θ, *определенная в* § 12.5, *а* W_α^* — *тест объема* α, *построенный по* $g_n(x_1, \ldots, x_n; \theta)$ *в соответствии с* (13.6.2), *то* W_α — *асимптотически более точный по сравнению с* W_α^* *тест гипотезы* $\mathscr{H}(\theta_0; \Omega_0)$. *Но если* $g_n(x_1, \ldots, x_n; \theta)$ *в* (13.6.2) *заменена функцией* $h_n(x_1, \ldots, x_n; \theta)$, *определенной в* (12.5.1), *то* W_α *и* W_α^* *будут асимптотически одинаково точными тестами гипотезы* $\mathscr{H}(\theta_0; \Omega_0)$.

13.7. Критерий отношения правдоподобия простой гипотезы

В §§ 13.4, 13.5, 13.6 были изложены основные асимптотические свойства критерия отношения правдоподобия простой гипотезы, в которой параметр θ предполагался одномерным. На самом деле в развитой в этих параграфах теории существенную роль играла часть параметрического множества Ω, именно, открытый интервал Ω_0, содержащий точку θ_0. Использованные там соображения могут быть без большого труда перенесены на случай r-мерного параметра $\theta = (\theta_1, \ldots, \theta_r)$. Достаточно поэтому привести r-мерные аналоги результатов §§ 13.4 — 13.6 с минимумом подробностей. Будем считать случайную величину x с ф. р. $F(x; \theta)$ одномерной, хотя перенесение наших результатов на случай k-мерных случайных величин потребует лишь небольших изменений в обозначениях.

Рассматривается простая гипотеза

$$\mathscr{H}_r = \mathscr{H}(\theta_0, \Omega_r),$$

где $\theta_0 = (\theta_{10}, \ldots, \theta_{r0})$, а Ω_r — произвольное множество в эвклидовом пространстве R_r, содержащее r-мерный открытый интервал Ω_{r0}, $\theta_0 \in \Omega_{r0}$. При изучении асимптотических свойств критерия отношения правдоподобия интерес представляет только Ω_{r0}. Поэтому в асимптотических задачах будем под $\mathscr{H}_r$ понимать гипотезу $\mathscr{H}(\theta_0; \Omega_{r0})$.

Прежде всего удобно сформулировать обобщение теоремы **13.4.1** на случай r-мерного параметра θ.

13.7.1. *Если параметр* θ *r-мерный и* $F(x; \theta)$ *регулярна в* Ω_{r0} *в смысле всех первых и вторых производных по* θ, *то определенная в* (12.1.17) *функция* $H(\theta_0, \theta)$ *имеет максимум* $H(\theta_0, \theta_0)$ *при* $\theta = \theta_0$. *Матрица вторых производных функции* — $H(\theta_0, \theta)$ *в точке* $\theta = \theta_0$ *есть определенная в* (12.1.19) *матрица*

$$\| B_{pq}(\theta_0, \theta_0) \|_{p, q = 1, \ldots, r}.$$

Доказательство теоремы **13.7.1** является непосредственным обобщением доказательства **13.4.1** и опускается. Отметим, что в r-мерном случае $H(\theta_0, \theta_0) - H(\theta_0, \theta_1)$ есть интеграл информации Калбэка для различения $F(x; \theta_0)$ и $F(x; \theta_1)$ по наблюдению над $F(x; \theta_0)$, а $\| B_{pq}(\theta_0, \theta_0) \|$ — *фишеровская матрица информации относительно* θ_0, *содержащейся в наблюдении над* $F(x; \theta_0)$, — совпадает с матрицей вторых производных интеграла информации Калбэка при $\theta_1 = \theta_0$. [См. Калбэк и Лейблер (1951).]

r-мерный вариант теоремы **13.4.2** не требует изменений в формулировке, надо только считать θ, θ_0 и $\hat{\theta}$ r-мерными. Аналог теоремы **13.4.3** формулируется следующим образом.

13.7.2. *Если θ — r-мерный параметр, $F(x;\ \theta)$ регулярна в смысле всех первых и вторых производных по θ, то при нулевой гипотезе*

$$\lim_{n\to\infty} P(-2\log\lambda_{\mathscr{H}_r}<\chi^2)=\frac{1}{2^{\frac{1}{2}r}\Gamma\left(\frac{r}{2}\right)}\int_0^{\chi^2} u^{\frac{1}{2}r-1}e^{-\frac{1}{2}u}\,du, \qquad (13.7.1)$$

т. е. распределение $-2\log\lambda_{\mathscr{H}_r}$ сходится к χ^2-распределению $C(r)$.

Доказательство этой теоремы предоставляется читателю.

Теорема **13.5.1** переносится на случай r-мерного параметра без изменения.

В r-мерном варианте **13.6.1** W_α и W_α^* задаются неравенствами

$$-2\log\lambda_{\mathscr{H}_r}>\chi_\alpha^2 \qquad (13.7.2)$$

и

$$U_n(x_1,\ \ldots,\ x_n;\ \theta_0)>\chi_\alpha^2, \qquad (13.7.3)$$

где $U_n(x_1,\ \ldots,\ x_n;\ \theta_0)$ определена в (12.9.1), а χ_α^2 — 100α-процентная точка χ^2-распределения $C(r)$. Доказательство **13.6.1** непосредственно переносится на r-мерный случай.

Теорема **13.6.2** переносится на случай r-мерного параметра без каких-либо трудностей, и мы предоставим это читателю в качестве упражнения. Аналогом фигурировавшей в теореме **13.6.2** области W_α^* будет область $V_n>\chi_\alpha^2$, где V_n определена в (12.9.15). Место одномерной оценочной функции $h_n(x_1,\ \ldots,\ x_n;\ \theta)$ займет в r-мерном случае вектор-функция $h_{pn}(x_1,\ \ldots,\ x_n;\ \theta)$, $p=1,\ \ldots,\ r$, определенная в (12.9.1):

$$h_{pn}(x_1,\ \ldots,\ x_n;\ \theta)=\frac{1}{n}S_{pn}(x_1,\ \ldots,\ x_n;\ \theta). \qquad (13.7.4)$$

13.8. Критерий отношения правдоподобия сложной гипотезы

В задачах проверки гипотез, включающих несколько параметров, часто встречается случай, когда подпространство $\omega_{r'}$ представляет собой r'-мерное эвклидово сечение Ω_r, т. е. цилиндрическое множество, состоящее из точек вида $(\theta_1,\ \ldots,\ \theta_{r'},\ \theta_{r'+1,\,0},\ \ldots,\ \theta_{r0})$, где $\theta_{r'+1,\,0},\ \ldots$ $\ldots,\ \theta_{r0}$ имеют фиксированные значения. Обозначим эту сложную гипотезу через $\mathscr{H}(\omega_{r'};\ \Omega_r)$ или, короче, через $\mathscr{H}_{r-r'}$. Тест отношения правдоподобия $\lambda_{\mathscr{H}_{r-r'}}$ гипотезы $\mathscr{H}_{r-r'}$ определяется обычным образом посредством (13.3.5). В больших выборках оказывается существенной только часть Ω_r, лежащая внутри некоторого r-мерного открытого интервала Ω_{r0}, содержащего истинную параметрическую точку $(\theta_{10}, \ldots, \theta_{r0})$. Поэтому для больших выборок можно рассматривать $\mathscr{H}_{r-r'}$ как сложную гипотезу $\mathscr{H}(\omega_{r'0}; \Omega_{r0})$, где $\omega_{r'0} = \omega_{r'} \cap \Omega_{r0}$.

Основная теорема [Уилкс (1938а)] о тесте отношения правдоподобия формулируется следующим образом:

13.8.1. *Пусть* $(x_1, \ldots, x_n)$ — *выборка из совокупности* $F(x; \theta)$, *где параметр* θ *r-мерный и* $F(x; \theta)$ *регулярна при* $\theta \in \Omega_{r0}$ *в смысле всех первых и вторых производных по* θ. *Тогда*

$$\lim_{n\to\infty} P(-2\log\lambda_{\mathscr{H}_{r-r'}} < \chi^2 \mid \theta \in \omega_{r'0}) =$$

$$= \frac{1}{2\Gamma\left(\frac{r-r'}{2}\right)} \int_0^{\chi^2} \left(\frac{u}{2}\right)^{\frac{r-r'}{2}-1} e^{-\frac{1}{2}u}\,du, \qquad (13.8.1)$$

т. е. при нулевой гипотезе распределения $-2\log\lambda_{\mathscr{H}_{r-r'}}$, $n = 1, 2, \ldots$, *сходится к* χ^2*-распределению* $C(r-r')$.

Для доказательства этой теоремы определим $\hat\theta_1, \ldots, \hat\theta_r$ как в **12.7.2.** Пусть $\hat\theta_1', \ldots, \hat\theta_{r'}'$ — решение системы уравнений

$$S_{p'n}(x_1, \ldots, x_n; \theta) = 0, \quad p' = 1, \ldots, r', \qquad (13.8.2)$$

относительно $\theta_1, \ldots, \theta_{r'}$, другими словами, $\hat\theta_1', \ldots, \hat\theta_{r'}'$ — оценки максимального правдоподобия $\theta_1, \ldots, \theta_{r'}$, когда оставшиеся компоненты фиксированы на значениях $\theta_{r'+1,0}, \ldots, \theta_{r0}$. В условиях теоремы **12.7.3** $(\hat\theta_1', \ldots, \hat\theta_{r'}')$ распределены асимптотически при $n \to \infty$ по закону

$$N(\{\theta_{p'0}\}; \ \|nB_{p'q'}\|^{-1}), \ p', q' = 1, \ldots, r'. \qquad (13.8.3)$$

Примем для $\theta_0 \in \omega_{r'0}$ следующие обозначения:

$$\begin{gathered}
\frac{1}{\sqrt{n}} S_{pn}(x_1, \ldots, x_n; \theta_0) = \zeta_{np}, \\
\sqrt{n}\,(\theta_{p0} - \hat\theta_p) = \eta_{np}, \\
\sqrt{n}\,(\theta_{p'0} - \hat\theta_{p'}') = \eta_{np'}', \\
\frac{1}{n} S_{pqn}(x_1, \ldots, x_n; \theta) = A_{pq}^{(n)}(\theta), \\
p, q = 1, \ldots, r; \quad p', q' = 1, \ldots, r'.
\end{gathered} \qquad (13.8.4)$$

Ввиду регулярности $F(x; \theta)$ в смысле первых и вторых производных по θ, для любого $\varepsilon > 0$ найдется n_ε такое, что с вероятностью, не меньшей $1-\varepsilon$, четыре нижеследующих равенства имеют место при всех $n>n_\varepsilon$:

$$\zeta_{np}=\sum_{q=1}^{r} A_{pq}^{(n)}(\theta^*)\,\eta_{nq}, \qquad p=1, \ldots, r, \tag{13.8.5}$$

$$\zeta_{np'}=\sum_{q'=1}^{r'} A_{p'q'}^{\prime(n)}(\theta^{*\prime})\,\eta'_{nq'}, \; p'=1, \ldots, r', \tag{13.8.6}$$

$$\sum_{\xi=1}^{n}\log dF(x_\xi;\ \theta_0)=\sum_{\xi=1}^{n}\log dF(x_\xi;\ \hat\theta_1, \ldots, \hat\theta_r)+\frac{1}{2}\sum_{p,\,q=1}^{r} A_{pq}^{(n)}(\theta_1^*)\,\eta_{np}\eta_{nq}, \tag{13.8.7}$$

$$\sum_{\xi=1}^{n}\log dF(x_\xi;\ \theta_0)=\sum_{\xi=1}^{n}\log dF(x_\xi;\ \hat\theta_1, \ldots, \hat\theta'_{r'}, \theta_{r'+1,0}, \ldots, \theta_{r0})+\frac{1}{2}\sum_{p',\,q'=1}^{r'} A_{p'q'}^{(n)}(\theta_1^{*\prime})\,\eta'_{np'}\eta'_{nq'}, \tag{13.8.8}$$

где θ^* и θ_1^* — точки из Ω_{r0}, лежащие между θ_0 и $\hat\theta$, а $\theta^{*\prime}$ и $\theta_1^{*\prime}$ — точки из $\omega_{r'0}$ между θ_0 и $(\hat\theta'_1, \ldots, \hat\theta'_{r'}, \theta_{r'+1,0}, \ldots, \theta_{r0})$.

Далее, η_{nq} и $\eta'_{nq'}$ можно представить в виде

$$\begin{aligned}\eta_{nq}&=\sum_{p=1}^{r} A_{(n)}^{pq}(\theta^*)\,\zeta_{np},\\ \eta'_{nq'}&=\sum_{p'=1}^{r'} A_{(n)}^{\prime p'q'}(\theta^{*\prime})\,\zeta_{np'},\end{aligned} \tag{13.8.9}$$

где $\| A_{(n)}^{pq}(\theta^*)\|=\| A_{pq}^{(n)}(\theta^*)\|^{-1}$ и $\| A_{(n)}^{\prime p'q'}(\theta^{*\prime})\|=\| A_{p'q'}^{\prime(n)}(\theta^{*\prime})\|^{-1}$.

Поскольку левые части равенств (13.8.7) и (13.8.8) одни и те же и

$$-2\log\lambda_{\mathscr{H}_{r-r'}}=-2\left\{\sum_{\xi=1}^{n}\log dF(x_\xi;\ \hat\theta'_1, \ldots, \hat\theta'_{r'}, \theta_{r'+1,0}, \ldots, \theta_{r0})-\sum_{\xi=1}^{n}\log dF(x_\xi;\ \hat\theta_1, \ldots, \hat\theta_r)\right\}, \tag{13.8.10}$$

то при всех $n>n_\varepsilon$ с вероятностью, не меньшей $1-\varepsilon$, будем иметь

$$-2\log\lambda_{\mathscr{H}_{r-r'}}=\sum_{p',\,q'=1}^{r'} A_{p'q'}^{(n)}(\theta_1^{*\prime})\,\eta'_{np'}\eta'_{nq'}-\sum_{p,\,q=1}^{r} A_{pq}^{(n)}(\theta_1^*)\,\eta_{np}\eta_{nq}, \tag{13.8.11}$$

где η_{np} и $\eta'_{np'}$ следует выразить через ζ_{np} по формулам (13.8.9).

Если $\theta_0 \in \omega_{r'0}$ — истинное значение θ, то согласно **12.7.1** распределение величины $(\zeta_{n1}, \ldots, \zeta_{nr})$ сходится при $n \to \infty$ к $N(\{0\}; \|B_{pq}\|)$. Кроме того, обе матрицы $\|A_{pq}^{(n)}(\theta^*)\|$ и $\|A_{pq}^{(n)}(\theta_1^*)\|$ сходятся по вероятности к матрице $\|-B_{pq}\|$, а матрицы $\|A_{p'q'}^{(n)}(\theta^{*\prime})\|$ и $\|A_{p'q'}^{(n)}(\theta_1^{*\prime})\|$ сходятся по вероятности к $\|-B_{p'q'}\|$, когда $n \to \infty$, где B_{pq} определены в **12.7.1**.

Произведя алгебраические преобразования правой части (13.8.11), найдем, что при $n \to \infty$ распределение обеих частей (13.8.11) сходится к распределению случайной величины

$$Q = \sum_{p,q=1}^{r} B^{pq} Y_p Y_q - \sum_{p',q'=1}^{r'} B_{[r']}^{p'q'} Y_{p'} Y_{q'}, \tag{13.8.12}$$

где $\|B_{[r']}^{p'q'}\| = \|B_{p'q'}\|^{-1}$ и $(Y_1, \ldots, Y_r)$ имеет распределение $N(\{0\}, \|B_{pq}\|)$. Можно показать, что при $\theta_0 \in \omega_{r'0}$ Q имеет χ^2-распределение $C(r - r')$. Тем самым доказательство теоремы **13.8.1** закончено.

Наконец, отметим без доказательства, что в условиях теоремы **13.8.1** тест отношения правдоподобия гипотезы $\mathscr{H}_{r-r'}$ состоятелен, т. е.

$$\lim_{n\to\infty} P(-2\log\lambda_{\mathscr{H}_{r-r'}} > \chi^2_\alpha \mid \theta \in \Omega_{r0} \setminus \omega_{r'0}) = 1.$$

Используя упоминавшийся в § 12.9 класс регулярных оценочных функций $ng_{pn}(x_1, \ldots, x_n; \theta)$ вместо функций $S_{pn}(x_1, \ldots, x_n; \theta)$, можно построить класс состоятельных тестов гипотезы $\mathscr{H}_{r-r'}$. Однако ни один из тестов этого класса не будет асимптотически более точным, чем тест отношения правдоподобия, если понятие асимптотической точности достаточно прямолинейно распространить на сложные гипотезы.

ЗАДАЧИ

13.1. Пусть выборка $(x_1, \ldots, x_n)$ извлечена из совокупности с п. в. вида $\theta e^{-\theta x}$, $x > 0$, $\Omega = (0, +\infty)$. С помощью теоремы Неймана — Пирсона определить критическую область W_α для проверки гипотезы $\mathscr{H}(\theta_0; \theta_0 \cup \theta_1)$ с условием

$$P(W_\alpha \mid \theta_0) = \alpha.$$

Получить функцию мощности этого теста, т. е. $P(W_\alpha \mid \theta)$, как функцию θ.

13.2. Пусть n_1, $\bar{x}_1$, s_1^2 — объем, выборочные среднее и дисперсия для выборки из совокупности $N(\mu_1, \sigma^2)$, а n_2, $\bar{x}_2$, s_2^2 — те же величины для выборки из совокупности $N(\mu_2, \sigma^2)$. Рассмотрите сложную гипотезу $\mathscr{H}(\omega; \Omega)$, в которой Ω — полупространство, эвклидова пространства R_3 точек (μ_1, μ_2, σ^2), $\sigma^2 > 0$, ω — подмножество Ω, для которого $\mu_1 = \mu_2$. Покажите, что отношение правдоподобия гипотезы $\mathscr{H}$ эквивалентно стьюдентовскому отношению

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{1/n_1 + 1/n_2}},$$

где

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

и t имеет при нулевой гипотезе распределение Стьюдента $S(n_1+n_2-2)$. Покажите далее, что критерий отношения правдоподобия несмещенный.

13.3. Если предыдущую задачу обобщить на случай k выборок, где n_i, $\bar{x}_i$, s_i^2, $i=1, \ldots, k$, — объемы, выборочные средние и дисперсии k независимых выборок из совокупностей $N(\mu_i, \sigma^2)$, а гипотеза $\mathscr{H}(\omega; \Omega)$ такова, что Ω — полупространство эвклидова пространства R_{k+1} точек $(\mu_1, \ldots, \mu_k, \sigma^2)$, $\sigma^2>0$; ω — подмножество Ω, для которого $\mu_1=\ldots=\mu_k$, то докажите, что отношение правдоподобия гипотезы эквивалентно отношению Снедекора

$$F=\frac{(n-k)\sum\limits_{i=1}^{k} n_i(\bar{x}_i-\bar{x})^2}{k\sum\limits_{i=1}^{k}(n_i-1)s_i^2},$$

где $n=n_1+\ldots+n_k$, $\bar{x}=\frac{1}{n}\sum\limits_{i=1}^{k} n_i\bar{x}_i$, и F имеет при нулевой гипотезе распределение Снедекора $S(k, n-k)$. Докажите несмещенность теста отношения правдоподобия гипотезы $\mathscr{H}(\omega; \Omega)$.

13.4. Пусть s_1^2 и s_2^2 — дисперсии выборок объемов n_1 и n_2 из совокупностей $N(\mu_1, \sigma_1^2)$ и $N(\mu_2, \sigma_2^2)$, а гипотеза $\mathscr{H}(\omega; \Omega)$ такова, что Ω — четвертая часть эвклидова пространства R_4 точек $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$, $\sigma_1^2>0$, $\sigma_2^2>0$; ω — подмножество Ω, для которого $\sigma_1^2=\sigma_2^2$. Показать, что отношение правдоподобия гипотезы $\mathscr{H}(\omega; \Omega)$ эквивалентно отношению Снедекора

$$F=\frac{s_1^2}{s_2^2},$$

которое имеет распределение Снедекора $S(n_1-1, n_2-1)$, когда гипотеза $\mathscr{H}(\omega; \Omega)$ верна.

13.5. Пусть x_1 и x_2 — независимые случайные величины, имеющие биномиальные распределения $\mathrm{Bi}(n_1, p_1)$ и $\mathrm{Bi}(n_2, p_2)$. Предположим, что для гипотезы $\mathscr{H}(\omega; \Omega)$ Ω представляет собой множество всех точек (p_1, p_2) единичного квадрата с вершинами $(0, 0)$, $(0, 1)$, $(1, 0)$ и $(1, 1)$, а ω — подмножество Ω, для которого $p_1=p_2$. Показать, что отношение правдоподобия для гипотезы $\mathscr{H}(\omega; \Omega)$ равно

$$\lambda=\left(\frac{x}{n}\right)^x\left(1-\frac{x}{n}\right)^{n-x}\left(\frac{x_1}{n_1}\right)^{-x_1}\left(\frac{x_2}{n_2}\right)^{-x_2}\left(1-\frac{x_1}{n_1}\right)^{-n_1+x_1}\left(1-\frac{x_2}{n_2}\right)^{-n_2+x_2},$$

где $x=x_1+x_2$ и $n=n_1+n_2$, и предельное распределение $-2\log\lambda$ при n_1, $n_2\to\infty$ есть χ^2-распределение $C(1)$, когда гипотеза $\mathscr{H}(\omega; \Omega)$ верна.

13.6. Обобщить предыдущую задачу и ее решение на случай, когда $x_1, \ldots, x_k$ — независимые случайные величины, имеющие биномиальные распределения $\mathrm{Bi}(n_1, p_1), \ldots, \mathrm{Bi}(n_k, p_k)$.

13.7. *Критерий независимости в $(r\times s)$-таблице сопряженности признаков.* Рассмотрим $(rs-1)$-мерную случайную величину $(n_{ij}, i=1, \ldots, r, j=1, \ldots, s)$, имеющую мультиномиальное распределение

$$\frac{n!}{\prod\limits_{j=1}^{s}\prod\limits_{i=1}^{r} n_{ij}!}\prod_{j=1}^{s}\prod_{i=1}^{r} p_{ij}^{n_{ij}},$$

где $\sum_j \sum_i n_{ij} = n$, $p_{ij} > 0$ и $\sum_j \sum_i p_{ij} = 1$. Пусть в гипотезе $\mathscr{H}$ (ω, Ω) Ω есть множество всех возможных значений p_{ij}, а ω — подмножество Ω, для которого $p_{ij} = p_i q_j$, где $\sum_i p_i = \sum_i q_j = 1$. Показать, что тест отношения правдоподобия гипотезы $\mathscr{H}$ (ω; Ω) задается величиной

$$\lambda = \prod_{i=1}^{r} \left(\frac{n_{i.}}{n}\right)^{n_{i.}} \prod_{j=1}^{s} \left(\frac{n_{.j}}{n}\right)^{n_{.j}} \Big/ \prod_{j=1}^{s} \prod_{i=1}^{r} \left(\frac{n_{ij}}{n}\right)^{n_{ij}},$$

где

$$n_{i.} = \sum_j n_{ij}, \qquad n_{.j} = \sum_i n_{ij},$$

и предельное при $n \to \infty$ распределение $-2 \log \lambda$ при нулевой гипотезе есть χ^2-распределение $C((r-1)(s-1))$.

13.8. (*Продолжение.*) Пусть g определена как

$$g = \sum_{j=1}^{s} \sum_{i=1}^{r} \left(n_{ij} - \frac{n_{i.} n_{.j}}{n}\right)^2 \Big/ \left(\frac{n_{i.} n_{.j}}{n}\right),$$

где

$$n_{i.} = \sum_j n_{ij}, \qquad n_{.j} = \sum_i n_{ij}.$$

Покажите, что при нулевой гипотезе распределения величин g и $-2 \log \lambda$ сходятся при $n \to \infty$ к одному и тому же предельному распределению $C((r-1)(s-1))$.

13.9. *Независимость слоев в* ($r \times s \times t$)-*таблице сопряженности признаков.* Пусть (n_{ijk}; $i = 1, \ldots, r$, $j = 1, \ldots, s$, $k = 1, \ldots, t$) — ($rst - 1$)-мерная случайная величина, имеющая мультиномиальное распределение

$$\frac{n!}{\prod_{k=1}^{t} \prod_{j=1}^{s} \prod_{i=1}^{r} n_{ijk}!} \prod_k \prod_j \prod_i p_{ijk}^{n_{ijk}},$$

где $\sum_k \sum_j \sum_i n_{ijk} = n$, $p_{ijk} > 0$ и

$$\sum_k \sum_j \sum_i p_{ijk} = 1.$$

Пусть Ω — множество всех возможных значений p_{ijk}, а ω — подмножество Ω, для которого $p_{ijk} = p_{ij} \cdot q_k$, $\sum_j \sum_i p_{ij} = \sum_k q_k = 1$. Доказать, что отношение правдоподобия для гипотезы $\mathscr{H}$ (ω; Ω) равно

$$\lambda = \prod_j \prod_i \left(\frac{n_{ij.}}{n}\right)^{n_{ij.}} \prod_k \left(\frac{n_{..k}}{n}\right)^{n_{..k}} \Big/ \prod_k \prod_j \prod_i \left(\frac{n_{ijk}}{n}\right)^{n_{ijk}},$$

где

$$n_{ij.} = \sum_k n_{ijk} \quad \text{и} \quad n_{..k} = \sum_j \sum_i n_{ijk},$$

и при нулевой гипотезе предельное распределение $-2 \log \lambda$, когда $n \to \infty$, есть χ^2-распределение $C((rs-1)(t-1))$.

13.10. Имеем независимые выборки $(x_{11}, \ldots, x_{1n_1})$ и $(x_{21}, \ldots, x_{2n_2})$ из пуассоновских совокупностей $\mathrm{Po}(\mu_1)$ и $\mathrm{Po}(\mu_2)$. Пусть $\mathscr{H}(\omega; \Omega)$ — гипотеза, в которой Ω — множество всех точек (μ_1, μ_2), лежащих в первом квадранте плоскости (μ_1, μ_2), а ω — подмножество Ω, для которого $\mu_1 = \mu_2$. Докажите, что отношение правдоподобия для $\mathscr{H}(\omega; \Omega)$ равно

$$\lambda = \frac{\bar{x}^{n\bar{x}}}{\bar{x}_1^{n_1\bar{x}_1} \cdot \bar{x}_2^{n_2\bar{x}_2}},$$

где $\bar{x}_1$ и $\bar{x}_2$ — выборочные средние,

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2},$$

и $n = n_1 + n_2$. Докажите, что при $n_1 \to \infty$, $n_2 \to \infty$ предельное распределение $-2\log\lambda$ есть χ^2-распределение $C(1)$, когда гипотеза $\mathscr{H}(\omega; \Omega)$ верна. Обобщите этот вывод на случай k выборок.

13.11. Пусть $(k-1)$-мерная случайная величина $(n_1, \ldots, n_k)$ имеет мультиномиальное распределение

$$\frac{n!}{n_1! \ldots n_k!} p_1^{n_1} \ldots p_k^{n_k},$$

где $n_1 + \ldots + n_k = n$, $p_i > 0$ и $p_1 + \ldots + p_k = 1$. Покажите, что отношение правдоподобия для гипотезы $\mathscr{H}(\omega; \Omega)$, в которой Ω — множество всех возможных значений p_i, а ω—множество, состоящее из одной точки $p_1 = p_{10}, \ldots, p_k = p_{k0}$, равно

$$\lambda = \left[\left(\frac{p_{10}}{\hat{p}_1}\right)^{\hat{p}_1} \ldots \left(\frac{p_{k0}}{\hat{p}_k}\right)^{\hat{p}_k}\right]^n,$$

где $\hat{p}_1 = n_1/n, \ldots, \hat{p}_k = n_k/n$, и предельное распределение $-2\log\lambda$ при $n \to \infty$ есть χ^2-распределение $C(k-1)$, если гипотеза $\mathscr{H}(\omega; \Omega)$ верна.

13.12. *(Продолжение).* Показать, что при нулевой гипотезе распределение каждой из величин $-2\log\lambda$, $\sum_{i=1}^{k} \frac{(n_i - np_{i0})^2}{np_{i0}}$ сходится к распределению $C(k-1)$, когда $n \to \infty$.

13.13. Пусть $s_1^2, \ldots, s_k^2$ — дисперсии независимых выборок объемов $n_1, \ldots, n_k$ из совокупностей $N(\mu_1, \sigma_1^2), \ldots, N(\mu_k, \sigma_k^2)$ соответственно; гипотеза $\mathscr{H}(\omega; \Omega)$ задается условиями: Ω — 2^{-k}-я часть эвклидова пространства R_{2k} точек $(\mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2)$, где $\sigma_1^2 > 0, \ldots, \sigma_k^2 > 0$, ω — подмножество Ω, на котором $\sigma_1^2 = \ldots = \sigma_k^2$. Показать, что отношение правдоподобия для гипотезы $\mathscr{H}(\omega; \Omega)$ равно

$$\lambda = \left[\frac{(n_1-1)\,s_1^2}{n_1 s_0^2}\right]^{\frac{1}{2}n_1} \ldots \left[\frac{(n_k-1)\,s_k^2}{n_k s_0^2}\right]^{\frac{1}{2}n_k},$$

где

$$s_0^2 = \frac{(n_1-1)\,s_1^2 + \ldots + (n_k-1)\,s_k^2}{n_1 + \ldots + n_k}.$$

13.14. Пусть s^2 — дисперсия выборки объема n из совокупности $N(\mu, \sigma^2)$, а гипотеза $\mathscr{H}(\omega; \Omega)$ определена следующим образом: Ω — множество точек (μ, σ^2), для которых $0 < \sigma^2 \leqslant \sigma_0^2$; ω — подмножество Ω, на котором $\sigma^2 = \sigma_0^2$.

Доказать, что использование для проверки гипотезы $\mathscr{H}(\omega;\ \Omega)$ отношения правдоподобия эквивалентно использованию отношения

$$\frac{(n-1)\,s^2}{\sigma_0^2},$$

которое имеет при нулевой гипотезе χ^2-распределение $C(n-1)$. Следовательно, критическая область W_α объема α имеет вид

$$\frac{(n-1)\,s^2}{\sigma_0^2} > \chi_\alpha^2,$$

где χ_α^2 — 100α-процентная точка распределения $C(n-1)$.

13.15. *Проверка гипотезы о параллельности линий регрессии.* Пусть $y_{11}, \ldots, y_{1n_1}$ — независимые случайные величины, каждая из которых имеет распределение $N(\beta_{10}+\beta_{11}x_{1\xi_1},\ \sigma^2)$, $\xi_1 = 1, \ldots, n_1$, а $y_{21}, \ldots, y_{2n_2}$ — независимые случайные величины, одинаково распределенные по закону $N(\beta_{20}+\beta_{21}x_{2\xi_2},\ \sigma^2)$, $\xi_2 = 1, \ldots, n_2$. Гипотеза $\mathscr{H}(\omega;\ \Omega)$ задается условиями: Ω — полупространство R_5 точек $(\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \sigma^2)$, на котором $\sigma^2 > 0$; ω — подмножество Ω, для которого $\beta_{11} = \beta_{21}$. Показать, что отношение правдоподобия для $\mathscr{H}(\omega,\ \Omega)$ эквивалентно статистике

$$F = \frac{S_\omega - S_\Omega}{S_\omega/(n_1+n_2-4)},$$

и при нулевой гипотезе F имеет распределение Снедекора $S(1,\ n_1+n_2-4)$. Здесь

$$S_\omega = \sum_{p=1}^{2}\sum_{\xi_p=1}^{n_p}[(y_{p\xi_p}-\bar{y}_p)-\hat{\beta}_p(x_{p\xi_p}-\bar{x}_p]^2,$$

$$\bar{y}_p = \frac{1}{n_p}\sum_{\xi_p=1}^{n_p} y_{p\xi_p},\qquad \bar{x}_p = \frac{1}{n_p}\sum_{\xi_p=1}^{n_p} x_{p\xi_p},\qquad p = 1,\ 2,$$

$$\hat{\beta}_p = a_p/b_p,\quad a_p = \sum_{\xi_p=1}^{n_p}(y_{p\xi_p}-\bar{y}_p)(x_{p\xi_p}-\bar{x}_p),\quad b_p = \sum_{\xi_p=1}^{n_p}(x_{p\xi_p}-\bar{x}_p)^2,$$

$$S_\Omega = \sum_{p=1}^{2}\sum_{\xi_p=1}^{n_p}[(y_{p\xi_p}-\bar{y}_p)-\hat{\beta}(x_{p\xi_p}-\bar{x}_p)]^2,$$

$$\hat{\beta} = (a_1+a_2)/(b_1+b_2).$$

13.16. Пусть $x_{\xi\eta}$, $\xi = 1, \ldots, r$, $\eta = 1, \ldots, s$, — независимые случайные величины, распределенные соответственно по законам $N(\mu+\mu_{\xi\cdot}+\mu_{\cdot\eta},\ \sigma^2)$, где $\sum_{\xi=1}^{r}\mu_{\xi\cdot} = 0$, $\sum_{\eta=1}^{s}\mu_{\cdot\eta} = 0$. Рассмотрим гипотезу $\mathscr{H}(\omega;\ \Omega)$, где Ω — множество вещественных векторов $(\mu, \mu_{1\cdot}, \ldots, \mu_{r\cdot}, \mu_{\cdot 1}, \ldots, \mu_{\cdot s}, \sigma^2)$, где $\sum_{\xi}\mu_{\xi\cdot} = \sum_{\eta}\mu_{\cdot\eta} = 0$, $\sigma^2 > 0$; ω — подмножество Ω, на котором $\mu_{1\cdot} = \ldots = \mu_{r\cdot} = 0$. Доказать, что отношение правдоподобия для $\mathscr{H}$ эквивалентно статистике

$$F = \frac{S_{\cdot 0}/(r-1)}{S_{\cdot\cdot}/[(r-1)(s-1)]},$$

имеющей при нулевой гипотезе распределения Снедекора $S((r-1),(r-1)\times\times(s-1))$. Здесь

$$S_{\cdot 0}=s\sum_{\xi}(\bar{x}_{\xi\cdot}-\bar{x})^2,$$

$$S_{\cdot\cdot}=\sum_{\eta}\sum_{\xi}(x_{\xi\eta}-\bar{x}_{\xi\cdot}-\bar{x}_{\cdot\eta}+\bar{x})^2,$$

$$\bar{x}=\frac{1}{rs}\sum_{\eta}\sum_{\xi}x_{\xi\eta},\quad \bar{x}_{\xi\cdot}=\frac{1}{s}\sum_{\eta}x_{\xi\eta},\quad \bar{x}_{\cdot\eta}=\frac{1}{r}\sum_{\xi}x_{\xi\eta}.$$

13.17. *Критерий равенства отношений вероятностей в модели Льюса (1959) выбора поведения.* В этой модели предполагается, что имеется выбор между альтернативами A_1, A_2, A_3 и проверяется гипотеза о том, что отношение вероятности выбора A_2 к вероятности выбора A_1 постоянно.

Более подробно, пусть (n_{1i}, n_{2i}, n_{3i}), где $n_{1i}+n_{2i}+n_{3i}=n_i$, $i=1, \ldots, k$, — k независимых групп случайных величин, имеющих распределения

$$\frac{n_i!}{n_{1i}!\; n_{2i}!\; n_{3i}!}\,p_i^{n_{1i}}q_i^{n_{2i}}r_i^{n_{3i}},$$

$i=1, \ldots, k$, $p_i>0$, $q_i>0$, $r_i>0$, $p_i+q_i+r_i=1$. Пусть Ω — множество всех положительных векторов

$$(p_1, q_1, r_1, \ldots, p_k, q_k, r_k),\quad p_i+q_i+r_i=1,\quad i=1, \ldots, k,$$

а ω — подмножество Ω, на котором $\dfrac{q_i}{p_i}=t$, $p_i(1+t)+r_i=1$. Показать, что отношение правдоподобия для $\mathscr{H}(\omega;\Omega)$ равно

$$\lambda=\frac{\prod\limits_{i=1}^{k}\{(\hat{p}_i)^{n_{1i}+n_{2i}}(\hat{t})^{n_{2i}}[1-\hat{p}_i(1+\hat{t})]^{n_{3i}}\}}{\prod\limits_{i=1}^{k}[n_{1i}^{n_{1i}}n_{2i}^{n_{2i}}n_{3i}^{n_{3i}}n^{-n}]},$$

где

$$\hat{t}=\sum_i n_{2i}\Big/\sum_i n_{1i},$$

$$\hat{p}_i=\frac{n_{1i}+n_{2i}}{n_i(1+\hat{t})}$$

и при $n_1\to\infty, \ldots, n_k\to\infty$ распределение $-2\log\lambda$ сходится (при нулевой гипотезе) к χ^2-распределению с $k-1$ степенями свободы. Показать также, что при нулевой гипотезе дисперсия $\hat{t}$ в больших выборках приближенно равна

$$t(1+t)\Big/\sum_i n_ip_i.$$

ГЛАВА 14

ПРОВЕРКА НЕПАРАМЕТРИЧЕСКИХ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

В главе 13 был дан краткий обзор теории тестов параметрических статистических гипотез и отдельно рассмотрен критерий отношения правдоподобия в больших выборках. В этой теории класс допустимых к. ф. р. имеет форму $\{F(x;\ \theta):\theta \in \Omega\}$, т. е. является классом специального функционального вида. Функции этого класса находятся в соответствии со значениями вещественного параметра θ из некоторого параметрического множества Ω; при этом и x и θ могут быть многомерными.

В теории тестов непараметрических статистических гипотез класс возможных гипотез является в общем случае классом или подклассом непрерывных ф. р. в зависимости от рассматриваемой конкретной задачи. Фактически еще не существует удовлетворительной общей теории проверки непараметрических статистических гипотез. Поэтому в этой главе теория непараметрических статистических тестов будет изложена на языке более или менее важных задач из рассматриваемой области, а не в общей форме. Читателя, интересующегося дальнейшей литературой по специальным непараметрическим тестам, мы отсылаем к книгам Фрезера (1957) и Кендалла (1953), а также к обзорным статьям Кендалла и Сандрума (1953), Морана, Уайтфилда и Даниелса (1950), Шеффе (1943), Вольфовитца (1949), Уилкса (1948, 1959а). Исчерпывающая библиография составлена Сэвиджем (1953).

14.1. Критерий квантилей

Простейшим типом непараметрического статистического теста является критерий для проверки гипотезы о том, что выборка $(x_1, \ldots, x_n)$ извлечена из совокупности, к. ф. р. $F(x)$ которой имеет заданную квантиль $\underline{x}_p$ порядка p. Напомним, что квантиль $\underline{x}_p$ порядка p определяется формулой

$$F(\underline{x}_p) = p. \qquad (14.1.1)$$

Опишем гипотезу более аккуратно. Пусть $\mathscr{C}_{0p}$ — класс всех к. ф. р., для которых $\underline{x}_{0p}$ есть квантиль порядка p, а класс $\mathscr{C}_p$ допустимых функций распределения состоит из всех к. ф. р. Тогда гипотеза, которую мы хотим проверить, есть $\mathscr{H}(\mathscr{C}_{0p};\ \mathscr{C}_p)$, и согласно этой гипотезе $F(x)$ принадлежит подклассу $\mathscr{C}_{0p}$ класса $\mathscr{C}_p$. Существенно, что $\mathscr{H}(\mathscr{C}_{0p};\ \mathscr{C}_p)$ является *непараметрической сложной статистической* *гипотезой*, и ее не следует путать с параметрической сложной гипотезой $\mathscr{H}(\omega;\ \Omega)$. В последнем случае Ω может быть представлено множеством точек эвклидова пространства, а ω — подмножество Ω, в то время как для гипотезы $\mathscr{H}(\mathscr{C}_{0p};\ \mathscr{C}_p)$ множество допустимых ф. р. $\mathscr{C}_p$ не может быть поставлено во взаимно однозначное непрерывное в обе стороны соответствие с подмножеством эвклидова пространства. Это же следует сказать и о $\mathscr{C}_{0p}$.

Пусть теперь $(x_1, \ldots, x_n)$ — выборка из распределения, принадлежащего классу $\mathscr{C}_{0p}$. Пусть r — число компонент вектора $(x_1, \ldots, x_n)$, значения которых лежат в интервале $(-\infty, \underline{x}_{0p})$; r имеет биномиальное распределение $\mathrm{Bi}(n, p)$. Интуитивно ясно, что разумной критической областью W_α для гипотезы $\mathscr{H}(\mathscr{C}_{0p};\ \mathscr{C}_p)$ будет объединение следующих вух множеств в пространстве значений r:

$$\{0, 1, \ldots, r_{\frac{1}{2}\alpha}\}, \quad \{r'_{\frac{1}{2}\alpha}, \ldots, n\}, \tag{14.1.2}$$

где $r_{\frac{1}{2}\alpha}$ — наибольшее целое число, для которого

$$P(r \leqslant r_{\frac{1}{2}\alpha} \mid F \in \mathscr{C}_{0p}) \leqslant \frac{1}{2}\alpha, \tag{14.1.3a}$$

и $r'_{\frac{1}{2}\alpha}$ — наименьшее целое число, для которого

$$P(r \geqslant r'_{\frac{1}{2}\alpha} \mid F \in \mathscr{C}_{0p}) \leqslant \frac{1}{2}\alpha. \tag{14.1.3b}$$

Из **9.2.1a** следует, что, при больших n, r распределено асимптотически нормально $N(np,\ np(1-p))$, откуда

$$\lim_{n\to\infty} P(W_\alpha \mid F \in \mathscr{C}_{0p}) = \alpha, \tag{14.1.4}$$

если выбрать

$$\begin{aligned} r_{\frac{1}{2}\alpha} &= np - y_{\frac{1}{2}\alpha}\cdot\sqrt{npq} + O(1),\\ r'_{\frac{1}{2}\alpha} &= np + y_{\frac{1}{2}\alpha}\cdot\sqrt{npq} + O(1), \end{aligned} \tag{14.1.5}$$

где $q = 1 - p$, $y_{\frac{1}{2}\alpha} > 0$ и $\Phi(-y_{\frac{1}{2}\alpha}) = \frac{1}{2}\alpha$, $\Phi(x)$ — к. ф. р. закона $N(0, 1)$.

Исследуем теперь состоятельность критерия W_α. Если $F(x)$ — какой-либо элемент из $\mathscr{C}_p - \mathscr{C}_{0p}$, то r будет иметь биномиальное распределение $\mathrm{Bi}(n, p')$, где $p' \neq p$. Так как при $n \to \infty$ $\frac{r}{n}$ сходится по вероятности к p для $F \in \mathscr{C}_{0p}$ (интервал $(r_{\frac{1}{2}\alpha}/n,\ r'_{\frac{1}{2}\alpha}/n)$ стягивается к точке p) и к p' для $F \in \mathscr{C}_p - \mathscr{C}_{0p}$, то ясно, что

$$\lim_{n\to\infty} P(W_\alpha \mid F \in \mathscr{C}_p - \mathscr{C}_{0p}) = 1. \tag{14.1.6}$$

Итак,

14.1.1. *Тест* W_α, *критическая область которого есть объединение двух множеств* (14.1.2), *состоятелен для проверки гипотезы* $\mathscr{H}(\mathscr{C}_{0p}; \mathscr{C}_p)$.

Следует заметить, что тест W_α^L с критической областью $\{0, 1, \ldots, r_\alpha\}$ будет *нижним односторонним тестом*, состоятельным при любой альтернативе из $\mathscr{C}_p$, у которой квантиль порядка p *больше*, чем $\underline{x}_{0p}$. Аналогично тест W_α^U с критической областью $\{r'_\alpha, \ldots, n\}$ будет *верхним односторонним тестом*, состоятельным при любой альтернативе из $\mathscr{C}_p$, у которой квантиль порядка p *меньше*, чем $\underline{x}_{0p}$.

Замечание. При $p = 0{,}5$ тест W_α иногда называют *критерием знаков*; гипотеза $\mathscr{H}(\mathscr{C}_{0p}; \mathscr{C}_p)$ становится в этом случае гипотезой о том, что выборка $(x_1, \ldots, x_n)$ извлечена из совокупности $F(x)$ с медианой $\underline{x}_{0,5}$. Практически важен тот случай, когда компоненты $(x_1, \ldots, x_n)$ сами являются разностями между независимыми парами случайных величин $(u_1, v_1; u_2, v_2; \ldots; u_n, v_n)$, т. е. $x_\xi = u_\xi - v_\xi$, $\xi = 1, \ldots, n$. Медиана $\underline{x}_{0,5}$ в такой схеме обычно равна 0. Более подробное обсуждение критерия знаков см. у Диксона и Муда (1946). Другие тесты, касающиеся медианы, имеются у Уолша (1949).

14.2. Непараметрические простые статистические гипотезы

(а) Предварительные замечания. Одна из основных непараметрических статистических гипотез возникает в следующей задаче. Пусть $(x_1, \ldots, x_n)$ — выборка из совокупности с некоторой непрерывной к. ф. р. $F(x)$. Верно ли, что эта выборка извлечена из совокупности с данной ф. р. $F_0(x)$? Если $\mathscr{C}$ — класс всех непрерывных ф. р., то условимся обозначать эту гипотезу $\mathscr{H}(F_0; \mathscr{C})$ и называть ее *непараметрической простой статистической гипотезой*. Ее не следует смешивать с параметрической простой статистической гипотезой $\mathscr{H}(\theta_0; \Omega)$, определенной в § 13.2. В последнем случае допустимые ф. р. принадлежат множеству $F(x; \theta)$, для которого параметр θ пробегает подмножество Ω эвклидова пространства, в то время как в случае гипотезы $\mathscr{H}(F_0, \mathscr{C})$ множество допустимых к. ф. р. представляет собой класс $\mathscr{C}$ всех непрерывных к. ф. р. и не может быть поставлено во взаимно однозначное и взаимно непрерывное соответствие с подмножеством эвклидова пространства. Если $\mathscr{C}_1$ — подкласс $\mathscr{C}$, то гипотеза $\mathscr{H}(\mathscr{C}_1; \mathscr{C})$ будет называться *непараметрической сложной статистической гипотезой*.

Мы рассмотрим три подхода к задаче построения тестов гипотезы $\mathscr{H}(F_0; \mathscr{C})$. Первый состоит в построении непараметрического теста сложной гипотезы, основанного на пирсоновском критерии хи-квадрат; второй — простой непараметрический подход к к. ф. р. в классе $\mathscr{A}$ абсолютно непрерывных к. ф. р. — приводит к *критерию пустых ящиков*; третий заключается в непараметрическом рассмотрении к. ф. р. в классе непрерывных к. ф. р., основанном на изученных в §§ 11.6—11.7 доверительных границах.

(b) Тест непараметрической сложной гипотезы, основанный на критерии хи-квадрат Пирсона. Пусть $\underline{x}_{i/m}$ — квантиль порядка i/m к. ф. р. $F_0(x)$, т. е. $F_0(\underline{x}_{i/m})=i/m$, $i=1, \ldots, m-1$, и пусть $\mathscr{C}_0$ — подкласс $\mathscr{C}$, состоящий из к. ф. р. с указанными выше квантилями. Положим

$$I_i=(\underline{x}_{(i-1)/m},\ \underline{x}_{i/m}], \qquad i=1, \ldots, m, \ \underline{x}_0=-\infty,\ \underline{x}_1=+\infty.$$

Пусть r_i — число элементов выборки, попавших в I_i. Если выборка извлечена из совокупности, принадлежащей $\mathscr{C}_0$, то частоты $(r_1, \ldots, r_m)$ образуют m-мерную случайную величину, $r_1+\ldots+r_m=n$, имеющую $(m-1)$-мерное мультиномиальное распределение $M(n;\ 1/m, \ldots, 1/m)$,

$$p(r_1, \ldots, r_m)=\frac{n!}{r_1! \ldots r_m!}\left(\frac{1}{m}\right)^n. \tag{14.2.1}$$

Считая m одним и тем же при всех n, мы, с некоторыми жертвами, заменяем задачу проверки гипотезы $\mathscr{H}(F_0;\ \mathscr{C})$ весьма простой задачей проверки непараметрической сложной гипотезы на основе пирсоновского критерия χ^2. Этот критерий, как мы увидим, будет состоятельным для гипотезы $\mathscr{H}(\mathscr{C}_0;\mathscr{C})$, т. е. для проверки гипотезы $F(x)\in\mathscr{C}_0$ против любой альтернативы $F(x)\in\mathscr{C}-\mathscr{C}_0$, и для многих практических задач он является адекватной заменой более тонких непараметрических тестов гипотезы $\mathscr{H}(F_0;\ \mathscr{C})$.

Для фиксированного m положим

$$Q_n=\frac{m}{n}\sum_{i=1}^{m}\left(r_i-\frac{n}{m}\right)^2 \tag{14.2.2}$$

и определим тест $W_{1\alpha}$ критическим множеством

$$Q_n>\chi^2_{\alpha,\,n}, \tag{14.2.3}$$

где $\chi^2_{\alpha,\,n}$ при каждом n выберем так, чтобы сделать объем $W_{1\alpha}$ возможно более близким к α.

Из **9.3.2a** следует, что

$$\lim_{n\to\infty} P(W_{1\alpha}\,|\,F\in\mathscr{C}_0)=\int_{\chi^2_\alpha}^{+\infty} dF_{m-1}(\chi^2)=\alpha, \tag{14.2.4}$$

где $dF_{m-1}(\chi^2)$ — элемент вероятности распределения χ^2 с $(m-1)$ степенями свободы, задаваемый формулой (7.8.1).

Будем изучать теперь мощность $W_{1\alpha}$ как теста гипотезы $\mathscr{H}(\mathscr{C}_0;\mathscr{C})$, т. е. значения $P(W_{1\alpha}\,|\,F\in\mathscr{C}-\mathscr{C}_0)$. Пусть p_i — вероятность попадания в I_i для некоторого распределения $F_1(x)\in\mathscr{C}-\mathscr{C}_0$.

Имеем

$$p_i=\int_{I_i} dF_1(x), \qquad i=1, \ldots, m, \tag{14.2.5}$$

и точка $(p_1, \ldots, p_m)$, конечно, отлична от $(1/m, \ldots, 1/m)$. Из многомерного обобщения **9.1.1** следует, что если выборка $(x_1, \ldots, x_n)$

извлечена из совокупности $F_1(x)$, то $(r_1/n, \ldots, r_m/n)$ сходится при $n \to \infty$ по вероятности к $(p_1, \ldots, p_m)$. Рассмотрим при заданном n множество E_n точек $(x_1, \ldots, x_n)$, для которых

$$\sum_{i=1}^{m} \left(\frac{r_i}{n} - p_i\right)^2 < D^2, \tag{14.2.6}$$

где D^2 — произвольное положительное число, меньшее чем квадрат (эвклидова) расстояния между точками $(p_1, \ldots, p_m)$ и $(1/m, \ldots, 1/m)$. Из того, что $W_{1\alpha}$ состоит из точек $(x_1, \ldots, x_n)$, для которых

$$\sum_{i=1}^{m} \left(\frac{r_i}{n} - \frac{1}{m}\right)^2 > \frac{\chi^2_{\alpha,\, n}}{m \cdot n}, \tag{14.2.7}$$

где $\lim\limits_{n \to \infty} \chi^2_{\alpha,\, n} = \chi^2_\alpha$, получаем, что для n, бо́льших некоторого n_1,

$$E_n \subset W_{1\alpha}.$$

Поэтому для $n > n_1$

$$P(E_n \mid F_1) \leqslant P(W_{1\alpha} \mid F_1). \tag{14.2.8}$$

Но поскольку $(r_1/n, \ldots, r_m/n)$ сходится по вероятности к $(p_1, \ldots, p_m)$ при $n \to \infty$, то по любому $\delta > 0$ можно указать n_2 так, что

$$P(E_n \mid F_1) > 1 - \delta$$

при $n > n_2$. Следовательно, при $n > \max(n_1, n_2)$

$$P(W_\alpha \mid F_1) \geqslant 1 - \delta,$$

т. е.

$$\lim_{n \to \infty} P(W_{1\alpha} \mid F_1) = 1. \tag{14.2.9}$$

Итак, тест $W_{1\alpha}$ состоятелен для проверки гипотезы $\mathscr{H}(\mathscr{C}_0; \mathscr{C})$ о том, что $F(x)$ принадлежит $\mathscr{C}_0$, против любой альтернативы из $\mathscr{C} - \mathscr{C}_0$. Он не будет, конечно, состоятельным для проверки гипотезы $F(x) = F_0(x)$ против альтернативы из $\mathscr{C}_0$. Суммируя все сказанное, будем иметь

14.2.1. *Пусть $\mathscr{C}_0$ — класс непрерывных к. ф. р. [включающий $F_0(x)$], имеющих квантили $\underline{x}_{i/m}$, $i = 1, \ldots, m-1$, а $I_i = (\underline{x}_{(i-1)/m}, \underline{x}_{i/m}]$, $\underline{x}_0 = -\infty$, $\underline{x}_1 = +\infty$, — интервалы, определяемые этими квантилями. Пусть r_i — число элементов выборки $(x_1, \ldots, x_n)$, попавших в интервалы I_i, $i = 1, 2, \ldots, m$, а W_α — тест, задаваемый критическим множеством*

$$\frac{m}{n} \sum_{i=1}^{m} \left(r_i - \frac{n}{m}\right)^2 > \chi^2_{\alpha,\, n},$$

где $\chi^2_{\alpha,\, n}$ выбирается так, чтобы сделать объем теста $W_{1\alpha}$ возможно более близким к α для любой $F \in \mathscr{C}_0$. Тогда для всякой $F \in \mathscr{C}_0$

имеет место (14.2.4). *Тест* $W_{1\alpha}$ *состоятелен для проверки гипотезы* $F\in\mathscr{C}_0$ *против любой альтернативы* $F\in\mathscr{C}-\mathscr{C}_0$. *Но в силу* (14.2.4) *тест* $W_{1\alpha}$ *не является состоятельным для проверки гипотезы* $F=F_0$ *против любой альтернативы из* $\mathscr{C}_0$.

(c) Критерий пустых ящиков. Главный недостаток развитого выше подхода к задаче проверки непараметрической простой гипотезы $\mathscr{H}(F_0;\mathscr{C})$ состоит в том, что $W_{1\alpha}$ оказался несостоятельным тестом гипотезы $F=F_0$ против альтернатив из класса $\mathscr{C}_0$. Этот класс может быть, конечно, сужен за счет выбора больших (но фиксированных при $n\to\infty$) значений m. Тем не менее, пока m остается фиксированным при $n\to\infty$, тест $W_{1\alpha}$ будет сложным непараметрическим статистическим тестом, не отличающим F_0 от любого другого элемента $\mathscr{C}_0$. Возникает вопрос, можно ли построить тест, который, позволяя m расти вместе с n, отличал бы F_0 от «почти всех» альтернатив из $\mathscr{C}_0$. Мы рассмотрим простой критерий, который отличает F_0 от альтернатив из подкласса $\mathscr{A}\subset\mathscr{C}$, содержащего абсолютно непрерывные к. ф. р. $F(x)$ с производными $f(x)$, предполагая, что F_0 также принадлежит к $\mathscr{A}$.

Пусть $s_0, s_1, \ldots, s_n$ — *групповые частоты*, определяемые выборкой $(x_1, \ldots, x_n)$, т. е. s_j есть число интервалов I_i, $i=1, \ldots, m$, содержащих j элементов выборки, $j=0, 1, 2, \ldots, n$. Тогда, если выборка извлечена из какой-либо совокупности, принадлежащей $\mathscr{C}_0$, то распределение (вырожденной) случайной величины $(s_0, s_1, \ldots, s_n)$ легко получается из (14.2.1):

$$p(s_0, s_1, \ldots, s_n)=\frac{m!\,n!}{m^n (0!)^{s_0} (1!)^{s_1} \ldots (n!)^{s_n} s_0!\, s_1! \ldots s_n!}, \tag{14.2.10}$$

где $s_0, s_1, \ldots, s_n$ — неотрицательные целые числа, удовлетворяющие условиям

$$\begin{aligned} s_0+s_1+\ldots+s_n&=m,\\ s_1+2s_2+\ldots+ns_n&=n. \end{aligned} \tag{14.2.11}$$

Употребляя методы, аналогичные использованным при выводе (6.1.4) и (6.1.7), получим следующую формулу для общего факториального момента компонент $(s_0, s_1, \ldots, s_n)$:

$$\mathscr{E}\left(s_0^{[g_0]} s_1^{[g_1]} \ldots s_n^{[g_n]}\right)= \\ =\frac{m!\,n!\,(m-g_0-g_1-\ldots-g_n)^{n-g_1-2g_2-\ldots-ng_n}}{m^n (2!)^{g_2} \ldots (n!)^{g_n} (m-g_0-g_1-\ldots-g_n)!\,(n-g_1-2g_2-\ldots-ng_n)!}, \tag{14.2.12}$$

где неотрицательные целые числа $g_0, \ldots, g_n$ должны удовлетворять условиям $m-g_0-g_1-\ldots-g_n\geqslant 0$ и $n-g_1-2g_2-\ldots-ng_n\geqslant 0$. Формула (14.2.12) позволяет получить средние значения, дисперсии, ковариации и другие моменты величин $s_0, s_1, \ldots, s_n$.

Критерий пустых ящиков, предложенный Дэвид (1950), основан на s_0—числе интервалов среди $I_1, \ldots, I_m$, не содержащих ни одного элемента выборки. Среднее и дисперсия s_0 находятся согласно (14.2.12):

$$\mu(s_0) = m\left(1-\frac{1}{m}\right)^n,$$

$$\sigma^2(s_0) = m(m-1)\left(1-\frac{2}{m}\right)^n + m\left(1-\frac{1}{m}\right)^n - m^2\left(1-\frac{1}{m}\right)^{2n}. \quad (14.2.13)$$

Если внимательно рассмотреть вид формулы (14.2.10), то легко заметить, что величина $p(s_0, s_1, \ldots, s_n)$ равна коэффициенту при $u_0^{s_0}u_1^{s_1}\ldots$ $\ldots u_n^{s_n}v^n$ в разложении функции

$$\frac{n!}{m^n}\left(u_0 + \frac{u_1 v}{1!} + \frac{u_2 v^2}{2!} + \ldots + \frac{u_n v^n}{n!}\right)^m, \quad (14.2.14)$$

откуда ясно, что для вычисления $p(s_0)$ надо положить в (14.2.14) $u_1 = \ldots = u_n = 1$ и взять коэффициент при $u_0^{s_0}v^n$ в разложении функции

$$\frac{n!}{m^n}\left(u_0 + \frac{v}{1!} + \frac{v^2}{2!} + \ldots + \frac{v^n}{n!}\right)^m. \quad (14.2.15)$$

Но коэффициент при $u_0^{s_0}v^n$ в (14.2.15) точно такой же, как в

$$\frac{n!}{m^n}(u_0 + e^v - 1)^m, \quad (14.2.16)$$

который в свою очередь равен коэффициенту при v^n в разложении

$$\varphi(v) = \frac{n!}{m^n}\binom{m}{s_0}(e^v-1)^{m-s_0}. \quad (14.2.17)$$

Коэффициент при v^n в разложении $\varphi(v)$ равен

$$\frac{1}{n!}\left[\frac{d^n\varphi(v)}{dv^n}\right]_{v=0}. \quad (14.2.18)$$

Если продифференцировать $\varphi(v)$ n раз и положить $v=0$, получим

$$p(s_0) = \frac{m!}{m^n s_0!}\sum_{i=0}^{m-s_0}\frac{(-1)^i(m-s_0-i)^n}{i!\,(m-s_0-i)!}, \quad (14.2.19)$$

где $s_0 = k, k+1, \ldots, m-1$, $k = \max(0, m-n)$.

Для больших значений m и n непосредственным анализом можно получить следующий результат, принадлежащий Дэвид (1950).

14.2.2. *Если $(x_1, \ldots, x_n)$ — выборка из совокупности с непрерывной к. ф. р. $F_0(x)$, то при $m\to\infty$, $n\to\infty$ так, что $\frac{n}{m}\to\rho>0$, s_0 асимптотически нормальна* $N(me^{-\rho},\ m[e^{-\rho}-e^{-2\rho}(1+\rho)])$.

Пусть теперь $W_{2\alpha}$ — тест, задаваемый критической областью $s_0 > s_{0\alpha}$, где $s_{0\alpha}$ — наименьшее целое число, для которого $\sum\limits_{s_0=s_{0\alpha}}^{m-1} p(s_0) \leqslant \alpha$. Из **14.2.2** следует, что при больших m и n при $n = \rho m + O(1)$

$$s_{0\alpha} = m\left[e^{-\rho} + \frac{y_\alpha}{\sqrt{m}}\left(e^{-\rho} - e^{-2\rho}(1+\rho)\right)^{\frac{1}{2}} + O\left(\frac{1}{m}\right)\right] \qquad (14.2.20)$$

и при $m \to \infty$, $n \to \infty$, $\frac{n}{m} \to \rho > 0$,

$$\lim_{n\to\infty} P(W_{2\alpha} \mid F_0) = \frac{1}{\sqrt{2\pi}} \int\limits_{y_\alpha}^{\infty} e^{-\frac{1}{2}t^2}\, dt = \alpha. \qquad (14.2.21)$$

Выбор теста $W_{2\alpha}$ разумен, так как интуитивно ясно, что распределение случайной величины s_0 стремится сдвинуться вправо по оси s_0, если выборка извлечена из любой совокупности из $\mathscr{C}$, отличной от $F_0(x)$.

Убедимся формально в разумном выборе теста $W_{2\alpha}$, исследуя его состоятельность для проверки гипотезы $F = F_0$ с абсолютно непрерывной $F_0(x)$ против альтернатив из класса A абсолютно непрерывных к. ф. р. Если $f_0(x)$ — производная $F_0(x)$ и $f_1(x)$ — производная какой-либо ф. р. $F_1(x) \in \mathscr{A}$, отличной от F_0, то f_0 и f_1 отличаются на множестве, имеющем положительную вероятность, вычисленную как при $F_0(x)$, так и при $F_1(x)$. Пусть $\mathscr{A}^*$ — такой подкласс $\mathscr{A}$, что все моменты отношения $f_1(x)/f_0(x)$ существуют по мере $F_0(x)$ и по мере $F_1(x)$, $F_1 \in \mathscr{A}^*$. Тогда $\mathscr{A}^*$ содержит $F_0(x)$. Докажем следующий результат:

14.2.3. *Тест $W_{2\alpha}$ состоятелен для проверки $F_0(x)$ против любой альтернативы $F_1(x) \in \mathscr{A}^*$, $F_1 \neq F_0$. Другими словами,*

$$\lim P(W_{2\alpha} \mid F_1 \in \mathscr{A}^* - F_0) = 1,$$

если $m \to \infty$, $n \to \infty$, $\frac{n}{m} \to \rho > 0$.

Для доказательства **14.2.3** достаточно установить, что если выборка $(x_1, \ldots, x_n)$ извлечена из совокупности $F_1 \in \mathscr{A}^* - F_0$, то s_0/m сходится по вероятности к постоянной, большей чем $e^{-\rho}$, когда $m \to \infty$, $n \to \infty$, так что $\frac{n}{m} \to \rho > 0$.

Пусть z_i — случайная величина, равная 1, если интервал I_i не содержит ни одного элемента выборки, и 0 в противном случае. Тогда

$$s_0 = z_1 + \ldots + z_m.$$

Обозначив $\mathscr{E}(s_0/m \mid F_1 \in \mathscr{A}^* - F_0)$ через $\mathscr{E}(s_0/m \mid F_1)$, будем иметь

$$\mathscr{E}\left(\frac{s_0}{m} \,\middle|\, F_1\right) = \frac{1}{m}\left[\mathscr{E}z_1 + \ldots + \mathscr{E}z_m\right].$$

Но

$$\mathscr{E}\left(\frac{z_i}{m}\right)=\frac{1}{m}(1-p_i)^n, \qquad i=1, \ldots, m,$$

где p_i — вероятность попадания в I_i при распределении $F_1(x)$. Отсюда

$$\mathscr{E}\left(\frac{s_0}{m}\middle|F_1\right)=\frac{1}{m}\sum_{i=1}^{m}(1-p_i)^n=$$

$$=1-\frac{n}{m}+\frac{n(n-1)}{2!m}\sum_i p_i^2-\frac{n(n-1)(n-2)}{3!m}\sum_i p_i^3+\ldots$$

$$\ldots+(-1)^k\frac{n(n-1)\ldots(n-k+1)}{k!m}\sum_i p_i^k+\ldots+\frac{(-1)^n}{m}\sum_i p_i^n. \qquad (14.2.22)$$

Общий член этого выражения может быть записан в виде

$$(-1)^k\frac{1}{k!}\left(\frac{n}{m}\right)\left(\frac{n-1}{m}\right)\ldots\left(\frac{n-k+1}{m}\right)\left\{\sum_{i=2}^{m-1}\left[\left(\frac{p_i}{\Delta_i}\right)\middle/\left(\frac{\left(\frac{1}{m}\right)}{\Delta_i}\right)\right]^k\frac{1}{m}\right\}+\delta_{m,n},$$

где Δ_i — длина интервала I_i, а

$$\delta_{m,n}=(-1)^k\frac{1}{k!}\left(\frac{n}{m}\right)\left(\frac{n-1}{m}\right)\ldots\left(\frac{n-k+1}{m}\right)\left[\left(\frac{p_1}{1/m}\right)^k+\left(\frac{p_m}{1/m}\right)^k\right]\cdot\frac{1}{m}.$$

Переходя к пределу при $m\to\infty$, $n\to\infty$, $\frac{n}{m}\to\rho>0$, и вспоминая предположение о конечности всех моментов отношения $f_1(x)/f_0(x)$ как при $F_0(x)$, так и при $F_1(x)$, найдем, что $\delta_{m,n}\to 0$ и предел выражения (14.2.22) равен

$$\sum_{k=0}^{\infty}(-1)^k\frac{\rho^k}{k!}\int_{-\infty}^{+\infty}[h(x)]^k\,dF_0(x),$$

где $h(x)=f_1(x)/f_0(x)$. Следовательно, при $m\to\infty$, $n\to\infty$, $n/m\to\rho>0$,

$$\lim\mathscr{E}\left(\frac{s_0}{m}\middle|F_1\right)=\int_{-\infty}^{+\infty}e^{-\rho h(x)}\,dF_0(x). \qquad (14.2.23)$$

Используя теперь то обстоятельство, что если $g(x)$ — неотрицательная случайная величина с конечным математическим ожиданием, то

$$\mathscr{E}g(x)\geqslant e^{\mathscr{E}\log g(x)} \qquad (14.2.24)$$

с равенством только в случае, когда $g(x)$ постоянна с вероятностью 1, получим

$$\int_{-\infty}^{+\infty}e^{-\rho h(x)}\,dF_0(x)\geqslant\exp\left[-\int_{-\infty}^{+\infty}\rho h(x)\,dF_0(x)\right]=e^{-\rho}. \qquad (14.2.25)$$

Равенство в (14.2.25) имеет место тогда и только тогда, когда $h(x)\equiv C$ с вероятностью 1. Но отсюда следует $f_0(x)=f_1(x)$ с вероятностью 1 и $F_0(x)=F_1(x)$. Так как $f_0(x)$ и $f_1(x)$ отличаются на множестве положительной вероятности, то

$$\lim \mathscr{E}\left(\frac{s_0}{m}\,\Big|\,F_1\right)>e^{-\rho}. \tag{14.2.26}$$

Для завершения доказательства **14.2.3** мы покажем, что $\sigma^2\left(\frac{s_0}{m}\,\Big|\,F_1\right)\to 0$ при $m\to\infty$ равномерно по n. Если обозначить матрицу ковариаций случайной величины $(z_1, \ldots, z_m)$ через $\|\sigma_{ij}\|$, то

$$\sigma^2\left(\frac{s_0}{m}\,\Big|\,F_1\right)=\frac{1}{m^2}\sum_{i,j=1}^{m}\sigma_{ij}=\frac{1}{m^2}\sum_{i=1}^{m}(1-p_i)^n[1-(1-p_i)^n]+$$
$$+\frac{1}{m^2}\sum_{i\neq j=1}^{m}[(1-p_i-p_j)^n-(1-p_i)^n(1-p_j)^n]. \tag{14.2.27}$$

Так как $(1-p_i-p_j)^n\leqslant(1-p_i)^n(1-p_j)^n$, то

$$\sigma^2\left(\frac{s_0}{m}\,\Big|\,F_1\right)\leqslant\frac{1}{m^2}\sum_{i=1}^{m}(1-p_i)^n[1-(1-p_i)^n].$$

Но $(1-p_i)^n[1-(1-p_i)^n]\leqslant\frac{1}{4}$, $i=1, \ldots, m$, для любого n. Следовательно,

$$\sigma^2\left(\frac{s_0}{m}\,\Big|\,F_1\right)\leqslant\frac{1}{4m},$$

откуда $\sigma^2\left(\frac{s_0}{m}\,\Big|\,F_1\right)\to 0$ при $m\to\infty$ равномерно по n. Из того, что $\lim \mathscr{E}\left(\frac{s_0}{m}\,\Big|\,F_1\right)>\lim \mathscr{E}\left(\frac{s_0}{m}\,\Big|\,F_0\right)=e^{-\rho}$ и $\sigma^2\left(\frac{s_0}{m}\,\Big|\,F_1\right)\to 0$, следует, что

$$\lim P(W_{2\alpha}\,|\,F_1\in A^*-F_0)=1$$

при $m\to\infty$, $n\to\infty$, $\frac{n}{m}\to\rho>0$. Доказательство **14.2.3** закончено.

В действительности Окамото (1952) и Китабатаке (1958) доказали, что если выборка $(x_1, \ldots, x_n)$ извлечена из любой совокупности с к. ф. р. $F_1(x)\in\mathscr{A}^*$, то при $m\to\infty$, $n=\rho m+O(1)$ асимптотическое распределение s_0 есть

$$N\left(m\int_0^1 e^{-\rho h(x)}\,dF_0(x),\ mK^2\right),$$

где

$$K^2=\int_{-\infty}^{+\infty}[e^{-\rho h(x)}-e^{-2\rho h(x)}]\,dF_0(x)-\rho\left[\int_{-\infty}^{+\infty}e^{-\rho h(x)}h(x)\,dF_0(x)\right]^2. \tag{14.2.28}$$

(d) Доверительные границы как непараметрические критерии. В §§ 11.6—11.7 рассматривалась задача оценивания к. ф. р. $F(x)$ по данным выборки $(x_1, \ldots, x_n)$. В этой задаче основным инструментом является определенная в (11.6.2) эмпирическая к. ф. р. $F_n(x)$. Оказывается, что в задаче построения теста гипотезы $\mathscr{H}(F_0; \mathscr{C}_1)$, где $\mathscr{C}_1$ — некоторый подкласс класса всех непрерывных к. ф. р., $F_n(x)$ также играет важную роль.

Андерсон и Дарлинг (1952), Крамер (1928), Кимбел (1947), Колмогоров (1933b), Малмквист (1954), Шерман (1950), Смирнов (1939 b), Мизес (1931) и другие рассматривали тесты, основанные на $F_n(x)$. Мы отсылаем читателя к сравнительному рассмотрению этих тестов у Бирнбаума (1953 b) и Дарлинга (1957). Из всех таких тестов мы изложим только основанный на доверительных границах из §§ 11.6—11.7.

Пусть $W_{3\alpha}$ — подмножество выборочного пространства, на котором неравенство

$$F_0(x) \leqslant F_n(x) + d \tag{14.2.29}$$

нарушается хотя бы при одном x. Если выбрать d так, чтобы $P(W_{3\alpha} \mid F_0) = \alpha$, то $W_{3\alpha}$ определит тест объема α. Исследуем мощность этого теста. Точнее, используя результат Бирнбаума (1953a), мы определим нижнюю грань $P(W_{3\alpha} \mid F_1)$ по отличным от $F_0(x)$ функциям из класса $\mathscr{C}$, удовлетворяющим некоторому слабому условию, формулируемому ниже. Мы также покажем, что при этом условии $W_{3\alpha}$ будет состоятельным тестом гипотезы $F = F_0$ против альтернатив из широкого подкласса класса $\mathscr{C}$.

Из **11.6.1** мы знаем, что, если $(x_1, \ldots, x_n)$ — выборка из совокупности $F_0(x)$, то

$$P(F_0(x) \leqslant F_n(x) + d \text{ для всех } x \mid F_0) = P_n(d), \tag{14.2.30}$$

т. е.

$$P(W_{3\alpha} \mid F_0) = 1 - P_n(d), \tag{14.2.31}$$

где $P_n(d)$ дается формулой (11.6.5). При заданном уровне значимости α и объеме выборки n можно из (11.6.5) найти такое значение d — обозначим его $d'(n, \alpha)$ или просто d', — что $P_n(d') = 1 - \alpha$, т. е.

$$P(W_{3\alpha} \mid F_0) = \alpha. \tag{14.2.32}$$

Пусть теперь $F_1(x)$ — функция распределения из $\mathscr{C}$, отличная от $F_0(x)$. Мы хотим изучить мощность $P(W_{3\alpha} \mid F_1)$ теста $W_{3\alpha}$ при альтернативе $F_1(x)$. Имеем

$$1 - P(W_{3\alpha} \mid F_1) = P(F_0(x) \leqslant F_n(x) + d' \text{ для всех } x \mid F_1). \tag{14.2.33}$$

Но неравенство $F_0(x) \leqslant F_n(x) + d'$ имеет место при всех x тогда и только тогда, когда оно имеет место в точках скачков функции $F_n(x)$, т. е. тогда и только тогда, когда

$$F_0(x_{(\xi)}) \leqslant \frac{\xi - 1}{n} + d', \tag{14.2.34}$$

где $(x_{(1)}, \ldots, x_{(n)})$ — вариационный ряд, построенный по выборке. Но (14.2.34) эквивалентно неравенству

$$x_{(\xi)} \leqslant F_0^{-1}\left(\frac{\xi-1}{n}+d'\right), \qquad \xi=1, \ldots, n, \tag{14.2.35}$$

где $F_0^{-1}(y)$ — функция, обратная к $F_0(x)$. В свою очередь, неравенство (14.2.35) эквивалентно такому:

$$F_1(x_{(\xi)}) \leqslant F_1\left[F_0^{-1}\left(\frac{\xi-1}{n}+d'\right)\right], \qquad \xi=1, \ldots, n. \tag{14.2.36}$$

Положим

$$F_1(F_0^{-1}(z))=G(z) \tag{14.2.37}$$

и сделаем замену переменных $y_{(\xi)}=F_1(x_{(\xi)})$, $\xi=1, \ldots, n$. Согласно (8.7.2) элемент вероятности $(y_{(1)}, \ldots, y_{(n)})$ равен

$$n!\, dy_{(1)} \ldots dy_{(n)}. \tag{14.2.38}$$

Следовательно, вероятность выполнения неравенства (14.2.36) равна

$$n! \int\limits_0^{G(d')} \int\limits_{y_{(1)}}^{G(1/n+d')} \cdots \int\limits_{y_{(n-1)}}^{G((n-1)/n+d')} dy_{(n)}\, dy_{(n-1)} \ldots dy_{(1)}. \tag{14.2.39}$$

Предположим теперь, что

$$\sup_{-\infty<x<+\infty} (F_0(x)-F_1(x))=\delta, \tag{14.2.40}$$

и пусть x' — значение x, для которого

$$F_0(x')-F_1(x')=\delta. \tag{14.2.41}$$

Если k — наибольшее целое, не превосходящее $n(F_0(x')-d')$, то

$$G\left(\frac{\xi-1}{n}+d'\right) \leqslant F_0(x')-\delta, \qquad \xi \leqslant k+1. \tag{14.2.42}$$

Кроме того,

$$G\left(\frac{\xi-1}{n}+d'\right) \leqslant 1, \qquad \xi=k+2, \ldots, n. \tag{14.2.43}$$

Замена верхних пределов интегрирования в (14.2.39) на правые части неравенств (14.2.42) и (14.2.43), очевидно, не уменьшит значения интеграла. Полученный же интеграл может быть вычислен по индукции, его значение равно

$$1-\sum_{\xi=0}^{k}\binom{n}{\xi} Q^{\xi}(1-Q)^{n-\xi}, \tag{14.2.44}$$

где

$$Q=F_0(x')-\delta. \tag{14.2.45}$$

Так как k — наибольшее целое, не превосходящее $n(F_0(x')-d')=$ $=n(Q+d-\delta')$, то имеем следующую нижнюю границу для мощности теста $W_{3\alpha}$:

$$P(W_{3\alpha}\mid F_1)\geqslant\sum_{\xi=0}^{k}\binom{n}{\xi}Q^{\xi}(1-Q)^{n-\xi}. \tag{14.2.46}$$

Сформулируем теперь полученный результат.

14.2.4. *Пусть для выборки объема n $W_{3\alpha}$ будет множеством тех точек выборочного пространства, для которых неравенство* (14.2.29) *не имеет места. Пусть, далее, $F_1(x)$ — элемент класса $\mathscr{C}$, для которого выполнено* (14.2.40). *Тогда нижняя граница мощности $P(W_{3\alpha}\mid F_1)$ теста $W_{3\alpha}$ указана в* (14.2.46).

При больших n выражение, стоящее в правой части (14.2.46), приближенно равно, как это следует из **9.2.1a**,

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y}e^{-\frac{1}{2}t^2}\,dt, \tag{14.2.47}$$

где

$$y=\frac{\sqrt{n}\,(\delta-d')}{\sqrt{Q(1-Q)}}. \tag{14.2.48}$$

При $n\to\infty$ $\lim d'=0$. Следовательно, если $\delta>0$, то

$$\lim P(W_{3\alpha}\mid F_1)=1. \tag{14.2.49}$$

Итак,

14.2.5. *Пусть $\mathscr{C}_1$ — подкласс класса $\mathscr{C}$, для элементов которого в* (14.2.40) $\delta>0$. *Тогда $W_{3\alpha}$ — состоятельный тест непараметрической простой гипотезы $\mathscr{H}(F_0;\ \mathscr{C}_1\cup F_0)$.*

Этот результат, грубо говоря, означает, что если график $F_0(x)$ лежит где-либо выше графика $F_1(x)$, то при $n\to\infty$ вероятность того, что график $F_n(x)+d'$ будет лежать где-либо под графиком $F_0(x)$, стремится к 1, если выборка $(x_1,\ \ldots,\ x_n)$ извлечена из совокупности $F_1(x)$.

Из (11.6.1) можно получить, что тест $W'_{3\alpha}$, критическая область которого представляет собой дополнение к множеству точек выборочного пространства, удовлетворяющих при всех x неравенству

$$F_0(x)\geqslant F_n(x)-d', \tag{14.2.29a}$$

имеет объем α. Если $\mathscr{C}'_1$ — подкласс класса $\mathscr{C}$, для элементов которого

$$\inf_{-\infty<x<+\infty}(F_0(x)-F_1(x))=\delta' \tag{14.2.40a}$$

с $\delta'<0$, то вполне аналогично **14.2.5** можно доказать состоятельность $W'_{3\alpha}$ для гипотезы $\mathscr{H}(F_0;\ \mathscr{C}'_1\cup F_0)$. Формулировку соответствующих вариантов теорем **14.2.4** и **14.2.5** мы предоставляем читателю.

Из теоремы **14.2.5** и из состоятельности теста $W'_{3\alpha}$ против альтернатив из класса $\mathscr{C}'_1$ следует, что тест $W''_{3\alpha}$, критической областью которого является дополнение к множеству точек выборочного пространства, удовлетворяющих при всех x неравенству

$$|F_n(x) - F_0(x)| \leqslant d'', \qquad (14.2.29\text{ b})$$

где d'' выбрана так, чтобы объем $W''_{3\alpha}$ равнялся α, будет состоятельным тестом непараметрической простой гипотезы $\mathscr{H}(F_0; \mathscr{C}_1 \cup \mathscr{C}'_1 \cup F_0)$.

14.3. Задача о двух выборках из непрерывных совокупностей

(a) Предварительные замечания. Пусть $(x_1, \ldots, x_{n_1})$ и $(x'_1, \ldots, x'_{n_2})$ — независимые выборки из совокупностей с непрерывными к. ф. р. $F_1(x)$ и $F_2(x)$ соответственно. Обозначим эти выборки через O_{n_1} и O_{n_2}. Основной проблемой непараметрической статистики является построение на базе информации, доставляемой выборками O_{n_1} и O_{n_2}, критериев гипотезы $F_1(x) \equiv F_2(x)$, состоятельных против альтернатив из различных классов пар $(F_1(x), F_2(x))$, где $F_1(x) \not\equiv F_2(x)$.

Критерии такого типа были предложены Диксоном (1940), Матисеном (1943), Смирновым (1939a), Вальдом и Вольфовитцем (1940), Уилксом (1961) и другими. Некоторые из них мы рассмотрим в этом параграфе.

При изучении критериев гипотезы $F_1(x) \equiv F_2(x)$ мы должны рассмотреть классы альтернатив. Условимся обозначать буквой $\mathfrak{F}$ класс всех пар непрерывных распределений $(F_1(x), F_2(x))$ и буквой $\mathfrak{F}_0$ — подкласс $\mathfrak{F}$, в котором $F_1(x) = F_2(x)$; эту общую к. ф. р. будем обозначать $F(x)$. Наиболее общей формой критерия $F_1(x) \equiv F_2(x)$ является критерий гипотезы $(F_1(x), F_2(x)) \in \mathfrak{F}_0$ против любой альтернативы $(F_1(x), F_2(x)) \in \mathfrak{F} \setminus \mathfrak{F}_0$.

Рассматриваемый в § 14.3 (e) критерий Смирнова (1939a) представляет собой пример такого общего критерия, который является состоятельным. Другие рассматриваемые критерии являются состоятельными для проверки гипотезы $(F_1(x), F_2(x)) \in \mathfrak{F}_0$ только против альтернатив из некоторых подклассов $\mathfrak{F} \setminus \mathfrak{F}_0$. Если $\mathfrak{F}^*$ — один из таких подклассов, то тест гипотезы $(F_1(x), F_2(x)) \in \mathfrak{F}_0$ против альтернатив $(F_1(x), F_2(x)) \in \mathfrak{F}^*$ будем называть тестом гипотезы $\mathscr{H}(\mathfrak{F}_0; \mathfrak{F}_0 \cup \mathfrak{F}^*)$.

Прежде чем рассматривать специальные критерии, удобно обсудить некоторые задачи о двух независимых выборках из одинаковых совокупностей с непрерывной к. ф. р.

(b) Распределение основных случайных величин, определяемых порядковыми статистиками двух выборок. Простой, но фундаментальной теоремой о порядковых статистиках двух выборок из совокупностей с одинаковой непрерывной ф. р. является следующая:

14.3.1. *Пусть O_{n_1} и O_{n_2} — независимые выборки из одинаковых совокупностей с непрерывной к. ф. р., а $(x_{(1)}, \ldots, x_{(n_1)})$ и $(x'_{(1)}, \ldots, x'_{(n_2)})$ — порядковые статистики этих выборок. Если $z_1, \ldots, z_{n_1+n_2}$ — упорядоченное размещение*) порядковых статистик обеих выборок, то*

$$P(z_1 < \ldots < z_{n_1+n_2}) = 1 \Big/ \binom{n_1+n_2}{n_1}. \tag{14.3.1}$$

Пусть $F(x)$ — общая к. ф. р. Чтобы не загромождать изложение, вычислим только вероятность

$$P(x_{(1)} < \ldots < x_{(n_1)} < x'_{(1)} < \ldots < x'_{(n_2)}). \tag{14.3.2}$$

Переобозначим случайные величины в (14.3.2) на $z_1, \ldots, z_{n_1+n_2}$. Если положить $y_1 = F(z_1), \ldots, y_{n_1+n_2} = F(z_{n_1+n_2})$, то согласно (8.7.2) элемент вероятности величин $(y_1, \ldots, y_{n_1+n_2})$ будет равен

$$n_1!\, n_2!\, dy_1 \ldots dy_{n_1+n_2}. \tag{14.3.3}$$

Событие $z_1 < \ldots < z_{n_1+n_2}$ наступает одновременно с $y_1 < \ldots < y_{n_1+n_2}$. Вероятность последнего есть

$$n_1!\, n_2! \int_0^1 \ldots \int_0^{y_3} \int_0^{y_2} dy_1\, dy_2 \ldots dy_{n_1+n_2} = 1 \Big/ \binom{n_1+n_2}{n_1}. \tag{14.3.4}$$

Очевидно, что (14.3.4) имеет место для произвольного упорядоченного размещения порядковых статистик двух выборок, что и доказывает теорему **14.3.1.**

Пусть, как и раньше, O_{n_1} и O_{n_2} — независимые выборки из одинаковых совокупностей с к. ф. р. $F(x)$. Порядковые статистики выборки O_{n_1} определяют выборочные блоки $B_1^{(1)}, \ldots, B_1^{(n_1+1)}$, как это указано в § 8.7. Пусть определяемые $F(x)$ вероятности попадания в эти блоки равны соответственно $u_1, \ldots, u_{n_1+1}$.

Обозначим через $r_1, \ldots, r_{n_1+1}$ количества элементов выборки O_{n_2}, попавших соответственно в $B_1^{(1)}, \ldots, B_1^{(n_1+1)}$, где $r_{n_1+1} = n_2 - r_1 - \ldots - r_{n_1}$. Случайные величины $(r_1, \ldots, r_{n_1+1})$ будем называть частотами блоков, определяемыми выборкой O_{n_1} по выборке O_{n_2}. Получим теперь распределение этих частот.

Пусть $(u_1, \ldots, u_{n_1}, r_1, \ldots, r_{n_1})$ — смешанная (n_1+n_2)-мерная случайная величина, где величины $(u_1, \ldots, u_{n_1})$ непрерывны, а $(r_1, \ldots, r_{n_1})$

*) Другими словами, $(z_1, \ldots, z_{n_1+n_2})$ получается следующим образом: сначала выписываются несколько первых порядковых статистик одной выборки, затем — несколько первых другой выборки, далее несколько первых порядковых статистик из числа оставшихся для первой выборки и т. д. (*Прим. перев.*)

дискретны и условное распределение $(r_1, \ldots, r_{n_1})$ при фиксированных значениях $(u_1, \ldots, u_{n_1})$ есть

$$\frac{n_2!}{r_1! \ldots r_{n_1+1}!} u_1^{r_1} \ldots u_{n_1+1}^{r_{n_1+1}}. \tag{14.3.5}$$

Из теоремы **8.7.4** элемент вероятности величин $(u_1, \ldots, u_{n_1})$ равен

$$n_1!\, du_1 \ldots du_{n_1}. \tag{14.3.6}$$

Поэтому совместное распределение величин $(r_1, \ldots, r_{n_1})$ и $(u_1, \ldots, u_{n_1})$ задается произведением (14.3.5) и (14.3.6). Чтобы получить (безусловное) распределение $(r_1, \ldots, r_{n_1})$, надо проинтегрировать это произведение по симплексу $S_{n_1}: \{(u_1, \ldots, u_{n_1}): u_1 \geqslant 0, \ldots, u_{n_1} \geqslant 0, u_1 + \ldots + u_{n_1} \leqslant 1\}$. Если использовать свойства распределения Дирихле из § 7.7, то интегрирование дает значение

$$1 \Big/ \binom{n_1+n_2}{n_1} \tag{14.3.7}$$

для всех тех наборов $(r_1, \ldots, r_{n_1})$, для которых $r_1 \geqslant 0, \ldots, r_{n_1} \geqslant 0$ и $r_1 + \ldots + r_{n_1} \leqslant n_2$.

Примерно так же, как доказывается (14.3.7), можно доказать, что распределение *любых* t величин $(r_1^*, \ldots, r_t^*)$ из $(r_1, \ldots, r_{n_1})$ есть

$$\binom{n_1+n_2-r_1^*-\ldots-r_t^*-t}{n_1-t} \Big/ \binom{n_1+n_2}{n_1}. \tag{14.3.8}$$

Суммируем полученные результаты.

14.3.2. *Пусть O_{n_1} и O_{n_2} — независимые выборки из одинаковых совокупностей с непрерывной к. ф. р. Если $(r_1, \ldots, r_{n_1+1})$ — частоты блоков, определяемые выборкой O_{n_1} по O_{n_2}, то случайная величина $(r_1, \ldots, r_{n_1+1})$, $r_1 + \ldots + r_{n_1+1} = n_2$ принимает все возможные значения с одинаковой вероятностью $1 \Big/ \binom{n_1+n_2}{n_1}$. Распределение любых t величин $(r_1^*, \ldots, r_t^*)$ из $(r_1, \ldots, r_{n_1+1})$ дается формулой* (14.3.8).

Если обозначить через $(r_1', \ldots, r_{n_2+1}')$ частоты блоков, определяемые O_{n_2} по O_{n_1}, то $(r_1', \ldots, r_{n_2+1}')$ будет с вероятностью

$$1 \Big/ \binom{n_1+n_2}{n_2} \tag{14.3.9}$$

принимать каждое из значений $(r_1', \ldots, r_{n_2+1}')$, удовлетворяющих условиям $r_1' \geqslant 0, \ldots, r_{n_2+1}' \geqslant 0$, $r_1' + \ldots + r_{n_2+1}' = n_1$. Так как выражения (14.3.7) и (14.3.9) одинаковы, то случайные величины $(r_1', \ldots, r_{n_2+1}')$ и $(r_1, \ldots, r_{n_1+1})$ принимают возможные значения в своих выборочных пространствах с одинаковыми постоянными вероятностями.

Отметим, что вероятности отдельных значений случайной величины $(r_1, \ldots, r_{n_1})$ такие же, как и задаваемые формулой (14.3.1) вероятности различных упорядоченных размещений порядковых статистик выборок O_{n_1} и O_{n_2}. Это и неудивительно, поскольку имеется взаимно однозначное соответствие между возможными значениями случайной величины $(r_1, \ldots, r_{n_1})$ и различными упорядоченными размещениями порядковых статистик обеих выборок.

Замечание. Можно рассмотреть и случай, когда выборки O_{n_1} и O_{n_2} извлечены из k-мерных совокупностей с непрерывной к. ф. р. При этом может быть принят любой из методов построения k-мерных выборочных блоков, описанных в § 8.7 (с). Пусть $B_k^{(1)}, \ldots, B_k^{(n_1+1)}$ — блоки, образованные n_1 точками выборки O_{n_1}, а $(r_1, \ldots, r_{n_1+1})$ — количества элементов выборки O_{n_2}, попавших соответственно в $B_k^{(1)}, \ldots, B_k^{(n_1+1)}$. Тогда распределение случайной величины $(r_1, \ldots, r_{n_1+1})$ в точности то же, что и в теореме **14.3.2.**

Пусть теперь s_0 — число блоков из $B_1^{(1)}, \ldots, B_1^{(n_1+1)}$, не содержащих ни одного элемента выборки O_{n_2}, s_1 — число блоков, содержащих ровно один элемент O_{n_2}, и т. д.; величины $s_0, s_1, \ldots, s_{n_2}$ будем называть *групповыми частотами блоков*, определяемыми выборкой O_{n_1} по O_{n_2}. Тогда $(s_0, s_1, \ldots, s_{n_2})$ — многомерная случайная величина, удовлетворяющая условиям:

$$\begin{gathered} s_0+s_1+\ldots+s_{n_2}=n_1+1, \\ s_1+2s_2+\ldots+n_2 s_{n_2}=n_2. \end{gathered} \tag{14.3.10}$$

Поскольку все точки в выборочном пространстве случайной величины $(r_1, \ldots, r_{n_1+1})$ имеют одинаковые вероятности, задача нахождения распределения $(s_0, s_1, \ldots, s_{n_2})$ сводится просто к подсчету числа точек в выборочном пространстве $(r_1, \ldots, r_{n_1+1})$, удовлетворяющих определенным условиям. После некоторого размышления читатель увидит, что число точек выборочного пространства $(r_1, \ldots, r_{n_1+1})$, которые отображаются в заданную точку $(s_0', s_1', \ldots, s_{n_2}')$ выборочного пространства величины $(s_0, s_1, \ldots, s_{n_2})$, равно коэффициенту при $u_0^{s_0'} u_1^{s_1'} \ldots u_{n_2}^{s_{n_2}'} v^{n_2}$ в формальном разложении выражения

$$(u_0+u_1 v+u_2 v^2+\ldots+u_{n_2} v^{n_2})^{n_1+1}. \tag{14.3.11}$$

Разделив этот коэффициент на $\binom{n_1+n_2}{n_1}$, мы получим распределение величины $(s_0, s_1, \ldots, s_{n_2})$:

$$\frac{(n_1+1)!}{s_0!\, s_1! \ldots s_{n_2}!} \Big/ \binom{n_1+n_2}{n_1}, \tag{14.3.12}$$

где $s_0, s_1, \ldots, s_{n_2}$, конечно, удовлетворяют условиям (14.3.10). Чтобы получить распределение некоторых из величин $(s_0, s_1, \ldots, s_{n_2})$, например $(s_0, s_1, \ldots, s_t)$, надо положить в (14.3.11) $u_{t+1}=\ldots=u_{n_2}=$

$=u$ и взять коэффициент при $u_0^{s_0} u_1^{s_1} \ldots u_t^{s_t} u^s v^{n_2}$, где $s_0+s_1+\ldots$ $\ldots+s_t+s=n_1+1$, $s=s_{t+1}+\ldots+s_{n_2}$. Положив $t=0$, получим распределение величины s_0:

$$p(s_0)=\frac{\binom{n_1+1}{s_0}\binom{n_2-1}{n_1-s_0}}{\binom{n_1+n_2}{n_1}}, \tag{14.3.13}$$

где возможные значения s_0 суть n_1-n_2+1, n_1-n_2+2, ..., n_1. Для больших значений t распределение имеет более сложный вид, и мы его не приводим.

Факториальные моменты одной или нескольких величин $(s_0, s_1, \ldots, s_{n_2})$ можно получить из (14.3.12) методами, аналогичными использованным при выводе (6.1.7). В общем виде

$$\mathscr{E}\left(s_0^{[g_0]} s_1^{[g_1]} \ldots s_t^{[g_t]}\right)= =(n_1+1)^{g_0+\ldots+g_t}\binom{n_1+n_2-g_0-2g_1-\ldots-(t+1)g_t}{n_1-g_0-\ldots-g_t}\Bigg/\binom{n_1+n_2}{n_1}, \tag{14.3.14}$$

где $g_0, \ldots, g_t$ — неотрицательные целые числа такие, что $g_0+\ldots+g_t \leqslant n_1$ и $g_1+2g_2+\ldots+tg_t \leqslant n_2$. Из (14.3.14) можно определить средние, дисперсии, ковариации и другие моменты одной или нескольких величин $(s_0, s_1, \ldots, s_{n_2})$.

Сформулируем полученные результаты.

14.3.3. *Если O_{n_1} и O_{n_2} — независимые выборки из одинаковых совокупностей с непрерывной к.ф.р., то распределение групповых частот блоков $(s_0, s_1, \ldots, s_{n_2})$, определяемых выборкой O_{n_1} по O_{n_2}, дается формулой* (14.3.12), *причем компоненты вектора $(s_0, s_1, \ldots, s_{n_2})$ удовлетворяют условиям* (14.3.10). *Общий факториальный момент величины $(s_0, s_1, \ldots, s_t)$ дается формулой* (14.3.14).

В частности, для s_0 имеем

$$\mathscr{E}s_0=n_1(n_1+1)/(n_1+n_2), \tag{14.3.15}$$

$$\sigma^2(s_0)=\frac{n_1^2(n_1^2-1)}{(n_1+n_2)(n_1+n_2-1)}+\frac{n_1(n_1+1)}{n_1+n_2}-\frac{n_1^2(n_1+1)^2}{(n_1+n_2)^2}.$$

Если положить $n_2=\rho n_1+O(1)$, $\rho>0$, то

$$\begin{aligned}\mathscr{E}s_0&=n_1(1/(1+\rho)+O(1/n_1)),\\ \sigma^2(s_0)&=n_1(\rho^2/(1+\rho)^3+O(1/n_1)).\end{aligned} \tag{14.3.16}$$

Сформулируем теперь утверждение, описывающее асимптотическое распределение s_0 в больших выборках.

14.3.4. *Если O_{n_1} и O_{n_2} — независимые выборки из одинаковых совокупностей с непрерывным распределением, n_1 и n_2 стремятся* *к бесконечности так, что* $n_2=\rho n_1+O(1)$, то s_0 *имеет асимптотически нормальное распределение* $N\left(\frac{n_1}{1+\rho}, \frac{n_1\rho^2}{(1+\rho)^3}\right)$.

Доказательство **14.3.4** можно провести, применив к факториалам в выражении (14.3.13) для $p(s_0)$ формулу Стирлинга с

$$s_0=\frac{n_1}{1+\rho}+y\sqrt{\frac{n_1\rho^2}{(1+\rho)^3}}. \tag{14.3.17}$$

Будем иметь

$$\lim_{n_1\to\infty}\sum_{s_0=s_0'(n_1)}^{s_0''(n_1)} p(s_0)=\frac{1}{\sqrt{2\pi}}\int_{y'}^{y''} e^{-\frac{1}{2}t^2}\,dt, \tag{14.3.18}$$

где $s_0'(n_1)$ и $s_0''(n_1)$ — наибольшие целые, содержащиеся в числах, получающихся при подстановке y' и y'' вместо t в правую часть (14.3.17). Мы опускаем детали.

(с) Критерий пустых блоков. Первым из непараметрических критериев в задаче о двух выборках мы рассмотрим простой *критерий пустых блоков*. Этот критерий подсказан предложенным Дэвид критерием пустых ящиков, обсуждавшимся в § 14.2 (с). Его критической областью служит множество $W_{a\alpha}$ значений s_0, $s_0\geqslant s_0(\alpha, n_1, n_2)$, где $s_0(\alpha, n_1, n_2)$ — наименьшее целое число, для которого

$$P(W_{a\alpha}\mid(F_1, F_2)\in\mathfrak{F}_0)\leqslant\alpha. \tag{14.3.19}$$

Так как при $(F_1, F_2)\in\mathfrak{F}_0$, т. е. при $F_1(x)\equiv F_2(x)$, распределение величины s_0 дается формулой (14.3.13), то значения $s_0(\alpha, n_1, n_2)$ при небольших n_1 и n_2 можно найти с помощью этой формулы.

При больших n_1, когда $n_2=\rho n_1+O(1)$, из **14.3.4** следует, что критическое значение $s_0(\alpha, n_1, n_2)$ равно

$$s_0(\alpha, n_1, n_2)=\frac{n_1}{1+\rho}+y_\alpha\sqrt{\frac{n_1\rho^2}{(1+\rho)^3}}+O(1), \tag{14.3.20}$$

где

$$\frac{1}{\sqrt{2\pi}}\int_{y_\alpha}^{\infty} e^{-\frac{1}{2}t^2}\,dt=\alpha. \tag{14.3.21}$$

Согласно **14.3.4** при $n_1\to\infty$, $n_2\to\infty$, так что $n_2/n_1\to\rho>0$,

$$\lim P(W_{a\alpha}\mid(F_1, F_2)\in\mathfrak{F}_0)=\alpha. \tag{14.3.22}$$

Тест $W_{a\alpha}$ разумен, так как интуитивно ясно, что распределение величины s_0 стремится сдвинуться вправо по оси s_0, если (F_1, F_2) не принадлежит $\mathfrak{F}_0$.

Убедимся в разумности выбора $W_{a\alpha}$ формально, исследуя его состоятельность. Пусть $F_1^{-1}(u)$ — функция, обратная к $F_1(x)$, а $\mathfrak{F}^*$ — класс пар непрерывных к. ф. р. $(F_1(x), F_2(x))$ таких, что

(i) $F_2(F_1^{-1}(u))$ имеет производную $g(u)$ для почти всех $u\in(0, 1)$;

(ii) производные функции $F_2(F_1^{-1}(u))$ и функции $F_1(F_1^{-1}(u)) \equiv u$ не равны на множестве положительной меры.

Докажем следующий результат:

14.3.5. *Критерий W_{aa} гипотезы $(F_1, F_2) \in \mathfrak{F}_0$ состоятелен против альтернативы $(F_1, F_2) \in \mathfrak{F}^*$, если только $n_1 \to \infty$, $n_2 \to \infty$ так, что $n_2 = \rho n_1 + O(1)$.*

Для доказательства **14.3.5** достаточно установить, что если $(F_1, F_2) \in \mathfrak{F}^*$, то $s_0/(n_1+1)$ сходится по вероятности к числу, большему, чем $1/(1+\rho)$, когда $n_1 \to \infty$, $n_2 \to \infty$ и $n_2/n_1 \to \rho > 0$. Напомним, что при $(F_1, F_2) \in \mathfrak{F}_0$ $s_0/(n_1+1)$ сходится по вероятности к $1/(1+\rho)$.

Пусть z_ξ — случайная величина, принимающая значение 1, если построенный по выборке O_{n_1} блок $B_1^{(\xi)}$ не содержит ни одного элемента выборки O_{n_2}, и равная 0 в остальных случаях, $\xi = 1, \ldots, n_1+1$.

Тогда

$$s_0 = z_1 + \ldots + z_{n_1+1}. \tag{14.3.23}$$

Вычислим $\mathscr{E}s_0$, считая O_{n_1} и O_{n_2} независимыми выборками из совокупностей с к. ф. р. F_1 и F_2, $(F_1, F_2) \in \mathfrak{F}^*$. Так как

$$\mathscr{E}z_\xi = P(z_\xi = 1),$$

то, обозначив $\mathscr{E}[s_0/(n_1+1) \mid (F_1, F_2) \in \mathfrak{F}^*]$ через $\mathscr{E}[s_0/(n_1+1) \mid \mathfrak{F}^*]$, будем иметь

$$\mathscr{E}\left(\frac{s_0}{n_1+1}\middle|\,\mathfrak{F}^*\right) = \frac{1}{n_1+1}[P(z_1=1) + \ldots + P(z_{n_1+1}=1)]. \tag{14.3.24}$$

Для $\xi = 2, \ldots, n_1$

$$P(z_\xi = 1) = \frac{n_1!}{(\xi-2)!\,(n_1-\xi)!}\int\limits_{S_2}[1 - F_2(x_{(\xi)}) + F_2(x_{(\xi-1)})]^{n_2}[F_1(x_{(\xi-1)})]^{\xi-2}[1-F_1(x_{(\xi)})]^{n_1-\xi} \times dF_1(x_{(\xi-1)})\,dF_1(x_{(\xi)}), \tag{14.3.25}$$

где S_2 — область в эвклидовом пространстве R_2, задаваемая условием $-\infty < x_{(\xi-1)} < x_{(\xi)} < +\infty$.

Выражения для $P(z_1=1)$ и $P(z_{n_1+1}=1)$ несколько проще (14.3.25), но предел каждого из них, поделенного на n_1+1, при $n_1 \to \infty$, $n_2 \to \infty$, $n_2/n_1 \to \rho > 0$ равен 0.

После замены переменных

$$u_1 = F_1(x_{(\xi-1)}), \quad u_2 = F_1(x_{(\xi)})$$

выражение (14.3.25) приведется к виду

$$P(z_\xi = 1) = \frac{n_1!}{(\xi-2)!\,(n_1-\xi)!}\int\limits_{T_2}[1 - G(u_1) + G(u_2)]^{n_2} u_1^{\xi-2}(1-u_2)^{n_1-\xi}\,du_1\,du_2, \tag{14.3.26}$$

где T_2 — треугольник $0 < u_1 < u_2 < 1$, а $G(u) = F_2(F_1^{-1}(u))$. Мы предположили, что $G(u)$ имеет производную $g(u)$. Тогда $g(u)$ будет п. в. на $(0, 1)$.

Имеем

$$\mathfrak{E}\left(\frac{s_0}{n_1+1}\Big|\mathfrak{F}^*\right) = \frac{1}{n_1+1}\sum_{\xi=2}^{n_1} P(z_\xi = 1) + \delta_{n_1, n_2}, \qquad (14.3.27)$$

где $\delta_{n_1, n_2} = (n_1+1)^{-1}[P(z_1 = 1) + P(z_{n_1+1} = 1)]$.

Подставляя выражение (14.3.26) для $P(z_\xi = 1)$ в (14.3.27) и произведя суммирование, получим

$$\mathfrak{E}\left(\frac{s_0}{n_1+1}\Big|\mathfrak{F}^*\right) = \frac{n_1(n_1-1)}{n_1+1}\int_{T_2}[1-u_2+u_1]^{n_1-2}[1-G(u_2)+ G(u_1)]^{n_2}\,du_1\,du_2 + \delta_{n_1, n_2}. \qquad (14.3.28)$$

Сделаем преобразование

$$u_1 = v, \qquad u_2 = v + \frac{y}{n_1}.$$

Тогда (14.3.28) запишется в виде

$$\mathfrak{E}\left(\frac{s_0}{n_1+1}\Big|\mathfrak{F}^*\right) = \frac{n_1(n_1-1)}{n_1(n_1+1)} \times \int_0^1\int_0^{(1-v)n_1}\left[1-\frac{y}{n_1}\right]^{n_1-2}\left[1-\frac{G\left(v+\frac{y}{n_1}\right)-G(v)}{y/n_1}\,\frac{y}{n_1}\right]dy\,dv + \delta_{n_1, n_2}. \qquad (14.3.29)$$

Переходя к пределу при $n_1 \to \infty$, $n_2 \to \infty$ так, что $n_2/n_1 \to \rho > 0$, получим

$$\lim \mathfrak{E}\left(\frac{s_0}{n_1+1}\Big|\mathfrak{F}^*\right) = \int_0^1\int_0^\infty e^{-y[1+\rho g(v)]}\,dy\,dv. \qquad (14.3.30)$$

Проинтегрировав по y, окончательно будем иметь

$$\lim \mathfrak{E}\left(\frac{s_0}{n_1+1}\Big|\mathfrak{F}^*\right) = \int_0^1\frac{dv}{1+\rho g(v)}. \qquad (14.3.31)$$

Мы должны теперь показать, что интеграл в правой части (14.3.31) превосходит $1/(1+\rho)$, если $g(v)$ отличается от единицы на множестве положительной меры. Последнее как раз имеет место в случае $(F_1, F_2) \in \mathfrak{F}^*$, поскольку производные функций $F_2(F_1^{-1}(u))$ и u не равны на множестве положительной меры.

При этом условии согласно неравенству Шварца

$$\int_0^1 \frac{dv}{1+\rho g(v)} \cdot \int_0^1 [1+\rho g(v)]\,dv > \left[\int_0^1 \sqrt{\frac{1}{1+\rho g(v)}} \cdot \sqrt{1+\rho g(v)}\,dv\right]^2. \tag{14.3.32}$$

Но в правой части неравенства стоит 1, а второй интеграл в левой части равен $1+\rho$. Следовательно,

$$\int_0^1 \frac{dv}{1+\rho g(v)} > \frac{1}{1+\rho}, \tag{14.3.33}$$

так что при $n_1 \to \infty$, $n_2 \to \infty$ и $n_2 = \rho n_1 + O(1)$, $\rho > 0$,

$$\lim \mathscr{E}\left(\frac{s_0}{n_1+1}\,\Big|\,\mathfrak{F}^*\right) > \lim \mathscr{E}\left(\frac{s_0}{n_1+1}\,\Big|\,\mathfrak{F}_0\right).$$

Чтобы закончить доказательство состоятельности критерия пустых блоков, достаточно установить, что дисперсия случайной величины $s_0/(n_1+1)$ для $(F_1, F_2) \in \mathfrak{F}^*$ при $n_1 \to \infty$, $n_2 \to \infty$, $n_2 = \rho n_1 + O(1)$ имеет порядок $1/n_1$. Имеем

$$\sigma^2\left(\frac{s_0}{n_1+1}\,\Big|\,\mathfrak{F}^*\right) = \frac{1}{(n_1+1)^2} \sum_{\xi,\eta=1}^{n_1+1} \sigma_{\xi\eta}, \tag{14.3.34}$$

где $\|\sigma_{\xi\eta}\|$ — матрица ковариаций вектора $(z_1, \ldots, z_{n_1+1})$. Ковариация $\sigma_{\xi\eta}$ компонент z_ξ и z_η равна

$$\sigma_{\xi\eta} = P(z_\xi = 1,\ z_\eta = 1) - P(z_\xi = 1)\,P(z_\eta = 1). \tag{14.3.35}$$

Покажем теперь, что $\sigma_{\xi\eta} \leqslant 0$ при всех $\xi \neq \eta$. Согласно (14.3.6) элемент вероятности величины $(u_1, \ldots, u_{n_1})$ равен $n!\,du_1 \ldots du_{n_1}$. Обозначим этот дифференциал через $dH(u)$.

Пусть $p_\xi = F_2(x_{(\xi)}) - F_2(x_{(\xi-1)})$ и аналогично же определим p_η. Так как равенства $u_1 + \ldots + u_\xi = F_1(x_{(\xi)})$, $\xi = 1, \ldots, n_1$, задают взаимно однозначное соответствие между $x_{(\xi)}$ и u_ξ, то ясно, что случайные величины $p_1, \ldots, p_{n_1}$ являются функциями от $u_1, \ldots, u_{n_1}$. При фиксированном значении выборки O_{n_1}

$$P(z_\xi = 1,\ z_\eta = 1 \mid u_1, \ldots, u_{n_1}) = (1 - p_\xi - p_\eta)^{n_2}, \tag{14.3.36}$$

откуда безусловная вероятность равна

$$P(z_\xi = 1,\ z_\eta = 1) = \mathscr{E}(1 - p_\xi - p_\eta)^{n_2}, \tag{14.3.37}$$

где математическое ожидание относится к распределению $(u_1, \ldots, u_{n_1})$. Аналогично

$$P(z_\xi = 1 \mid u_1, \ldots, u_{n_1}) = (1 - p_\xi)^{n_2},$$

$$P(z_\eta = 1 \mid u_1, \ldots, u_{n_1}) = (1 - p_\eta)^{n_2},$$

и для безусловных вероятностей $P(z_\xi=1)$ и $P(z_\eta=1)$ будем иметь:

$$P(z_\xi=1)\,P(z_\eta=1)=\mathfrak{E}(1-p_\xi)^{n_2}\,\mathfrak{E}(1-p_\eta)^{n_2}. \qquad (14.3.38)$$

Так как

$$(1-p_\xi-p_\eta)^{n_2}\leqslant(1-p_\xi)^{n_2}(1-p_\eta)^{n_2}, \qquad (14.3.39)$$

то

$$\mathfrak{E}(1-p_\xi-p_\eta)^{n_2}\leqslant\mathfrak{E}\,[(1-p_\xi)^{n_2}(1-p_\eta)^{n_2}]. \qquad (14.3.40)$$

Но поскольку $0\leqslant(1-p_\xi)^{n_2}(1-p_\eta)^{n_2}\leqslant 1$, то

$$\mathfrak{E}\,[(1-p_\xi)^{n_2}(1-p_\eta)^{n_2}]\leqslant\{\mathfrak{E}\,[(1-p_\xi)^{\frac{1}{2}n_2}(1-p_\eta)^{\frac{1}{2}n_2}]\}^2, \qquad (14.3.41)$$

и по неравенству Шварца

$$\{\mathfrak{E}\,[(1-p_\xi)^{\frac{1}{2}n_2}(1-p_\eta)^{\frac{1}{2}n_2}]\}^2\leqslant\mathfrak{E}(1-p_\xi)^{n_2}\,\mathfrak{E}(1-p_\eta)^{n_2}. \qquad (14.3.42)$$

Объединяя (14.3.40), (14.3.41), (14.3.42), получим

$$\mathfrak{E}(1-p_\xi-p_\eta)^{n_2}\leqslant\mathfrak{E}(1-p_\xi)^{n_2}\,\mathfrak{E}(1-p_\eta)^{n_2}, \qquad (14.3.43)$$

что эквивалентно следующему:

$$\sigma_{\xi\eta}\leqslant 0, \qquad \xi\neq\eta. \qquad (14.3.44)$$

Последнее вместе с (14.3.34) дает

$$\sigma^2\left(\frac{s_0}{n_1+1}\,\Big|\,\mathfrak{F}^*\right)\leqslant\frac{1}{(n_1+1)^2}\sum_{\xi=1}^{n_1+1}\sigma_{\xi\xi}. \qquad (14.3.45)$$

Но

$$\sigma_{\xi\xi}=P(z_\xi=1)-[P(z_\xi=1)]^2=$$
$$=P(z_\xi=1)\,[1-P(z_\xi=1)]\leqslant\frac{1}{4} \qquad (14.3.46)$$

для $\xi=1,\ \dots,\ n_1+1$.

Окончательно получаем

$$\sigma^2\left(\frac{s_0}{n_1+1}\,\Big|\,\mathfrak{F}^*\right)\leqslant\frac{1}{4\,(n_1+1)}, \qquad (14.3.47)$$

откуда $\sigma^2\left(\frac{s_0}{n_1+1}\,\Big|\,\mathfrak{F}^*\right)\to 0$ при $n_1\to\infty$, что и доказывает утверждение **14.3.5.**

Блюм и Уэйсс (1957) рассматривали такие подклассы $\mathfrak{F}^*$, что любая групповая частота блоков s_i оказывается состоятельной для гипотезы $(F_1, F_2)\in\mathfrak{F}_0$ против альтернатив из этих подклассов. Уилкс (1961) рассматривал критерий типа χ^2, построенный по $s_0, s_1, \dots, s_k$ для фиксированного k и больших выборок. Критическая область $W^*_{a\alpha}$ для него определяется условием

$$\sum_{i=0}^{k}\frac{(s_i-na_i)^2}{na_i}+\frac{u^2+v^2}{n\rho^2\,(1+\rho)\,a_k}>\chi^2_\alpha,$$

где $a_i = \rho^i/(1+\rho)^{i+1}$,

$$u = \sum_{i=0}^{k} (s_i - na_i)(i - \rho - k - 1),$$

$$v = \sqrt{\rho(1+\rho)} \sum_{i=0}^{k} (s_i - na_i)$$

и χ^2_α — 100α-процентная точка χ^2-распределения $C(k+1)$. Можно ожидать, что этот критерий несколько мощнее, чем $W_{a\alpha}$, который зависит только от s_0.

Замечание. Тест $W_{a\alpha}$ (или $W^*_{a\alpha}$) может быть обобщен на случай k-мерных непрерывных к. ф. р. В этом случае выборка O_{n_1} порождает k-мерные выборочные блоки $B'_k, \dots, B_k^{(n_1+1)}$, как это указано в § 8.7 (с). Структура этих блоков зависит от принятой системы упорядочения. Число элементов выборки O_{n_2} в блоках является случайной величиной $(r_1, \dots, r_{n_1+1})$, где $r_1 + \dots + r_{n_1+1} = n_2$. Распределение числа пустых блоков s_0 дается формулой (14.3.13), если оба распределения одинаковы. Так же, как в одномерном случае, можно доказать состоятельность критерия $W_{a\alpha}$ в k-мерном случае при определенных условиях на к. ф. р.

(d) Критерий серий. Если каждую из порядковых статистик выборки O_{n_1} обозначить через C, а выборки O_{n_2} — через $\bar{C}$, то каждому упорядоченному размещению порядковых статистик обеих выборок будет отвечать некоторое расположение n_1 букв C и n_2 букв $\bar{C}$. В условиях теоремы **14.3.1** все $\binom{n_1+n_2}{n_1}$ возможных расположений n_1 букв C и n_2 букв $\bar{C}$ равновероятны. Поэтому вся теория серий из § 6.6 применима к сериям, порожденным $\binom{n_1+n_2}{n_1}$ возможными последовательностями букв C и $\bar{C}$. (Заметим, что в § 6.6 n_1+n_2 обозначено через n.)

Вальд и Вольфовитц (1940) предложили использовать в задаче о двух выборках в качестве критерия статистику u — *общее число серий* в последовательности букв C и $\bar{C}$, отвечающей упорядоченному расположению порядковых статистик обеих выборок (см. § 6.6 (a)). Элемент вероятности u при $(F_1(x), F_2(x)) \in \mathfrak{F}_0$ дается формулой (6.6.20), а $\mathcal{E}u$ и $\sigma^2(u)$ — формулой (6.6.19). Отметим, что случайные величины u и s_0 строго отрицательно коррелированы.

Так как при малых значениях u мы склонны подвергнуть сомнению истинность гипотезы $(F_1(x), F_2(x)) \in \mathfrak{F}_0$, то зададим критическую область условием

$$P(u \leqslant u(\alpha, n_1, n_2)) \leqslant \alpha, \tag{14.3.48}$$

где $u(\alpha, n_1, n_2)$ — наибольшее целое, для которого (14.3.48) имеет место. Обозначим этот тест объема α через $W_{b\alpha}$.

При больших значениях n_1 и n_2 величина u распределена асимптотически нормально со средним и дисперсией, определяемыми формулой (6.6.19). Более точно, Вальд и Вольфовитц доказали теорему

14.3.6. *Если* $(F_1(x),\ F_2(x)) \in \mathfrak{F}_0$, *а* $n_1 \to \infty$, $n_2 \to \infty$ *так, что* $\frac{n_2}{n_1} \to \rho > 0$, *то статистика и распределена асимптотически по закону* $N\left(\frac{2n_1\rho}{1+\rho},\ \frac{4\rho^2 n_1}{(1+\rho)^3}\right)$.

Доказательство теоремы **14.3.6** аналогично доказательству **9.6.3** и мы приведем только набросок его.

Пусть

$$y = \frac{\left(u - \frac{2n_1\rho}{1+\rho}\right)(1+\rho)^{\frac{3}{2}}}{2\rho\sqrt{n_1}}.$$

Тогда

$$u = \frac{2\rho\sqrt{n_1}\,(y + \sqrt{n_1}\,(1+\rho))}{\sqrt{(1+\rho)^3}}. \tag{14.3.49}$$

Обозначим через $p_1(u)$ и $p_2(u)$ выражения для $p(u)$ в (6.6.20) для четных и нечетных значений u соответственно.

Тогда при заданных n_1 и числах $t' < t''$ имеем.

$$P(y' < y < y'') = \sum\nolimits_1 p_1\left(\frac{2\rho\sqrt{n_1}\,(y + \sqrt{n_1}\,(1+\rho))}{\sqrt{(1+\rho)^3}}\right) + \\ + \sum\nolimits_2 p_2\left(\frac{2\rho\sqrt{n_1}\,(y + \sqrt{n_1}\,(1+\rho))}{\sqrt{(1+\rho)^3}}\right), \tag{14.3.50}$$

где $\sum_1$ означает суммирование по (дискретным) значениям y между y' и y'', удовлетворяющим условию (14.3.48), для четных u, а $\sum_2$ — аналогичное суммирование для нечетных u.

Рассмотрим теперь первую сумму, которую можно записать в виде

$$\sum\nolimits_1 \left[p_1\left(\frac{2\rho n_1\,(y + \sqrt{n_1}\,(1+\rho))}{\sqrt{(1+\rho)^3}}\right)\sqrt{n_1\rho^2/(1+\rho)^3}\right]\Delta y, \tag{14.3.51}$$

где $\Delta y = \sqrt{(1+\rho)^3/(n_1\rho^2)}$. Применяя формулу Стирлинга (7.6.29) к факториалам в выражении для $p_1(u)$ (т. е. для $p(u)$ при четных u), найдем, что при $n_1 \to \infty$ величина (14.3.51) сходится к

$$\frac{1}{2\sqrt{2\pi}} \int\limits_{y'}^{y''} e^{-\frac{1}{2}t^2}\,dt. \tag{14.3.52}$$

Аналогично устанавливается, что вторая сумма в (14.3.50) сходится к (14.3.52) при $n_1 \to \infty$. Следовательно,

$$\lim_{n_1 \to \infty} P(y' < y < y'') = \frac{1}{\sqrt{2\pi}} \int\limits_{y'}^{y''} e^{-\frac{1}{2}t^2}\,dt \tag{14.3.53}$$

для любых $y' < y''$, что доказывает **14.3.6**.

Если $n_1 \to \infty$, $n_2 \to \infty$ так, что $\lim n_2/n_1 = \rho$, и $(F_1(x), F_2(x)) \in \mathfrak{F}_0$, то из теоремы **14.3.6** получим

$$\lim_{n_1 \to \infty} P(u \leqslant u(\alpha, n_1, \rho n_1)) = \alpha. \tag{14.3.54}$$

Вальд и Вольфовитц доказали состоятельность теста $W_{b\alpha}$ гипотезы $(F_1(x), F_2(x)) \in \mathfrak{F}_0$ против альтернатив из введенного в **14.3.5** класса $\mathfrak{F}^*$.

(е) Критерий Смирнова. Пусть $F_{1n_1}(x)$ и $F_{2n_2}(x)$ — эмпирические к. ф. р., построенные по выборкам O_{n_1} и O_{n_2} в соответствии с (11.6.2). Статистика Смирнова (1939а)

$$D_{n_1,\, n_2} = \sup_x | F_{1n_1}(x) - F_{2n_2}(x) | \tag{14.3.55}$$

сама напрашивается в качестве разумного критерия гипотезы $(F_1(x), F_2(x)) \in \mathfrak{F}_0$.

Пусть $W_{c\alpha}$ — критическая область

$$D_{n_1,\, n_2} > \delta(\alpha, n_1, n_2),$$

где $\delta(\alpha, n_1, n_2)$ — наименьшее целое, для которого

$$P(D_{n_1,\, n_2} > \delta(\alpha, n_1, n_2)) = \alpha \tag{14.3.56}$$

при $(F_1(x), F_2(x)) \in \mathfrak{F}_0$.

Смирнов доказал, что при $(F_1(x), F_2(x)) \in \mathfrak{F}_0$ и $n_1 \to \infty$, $n_2 \to \infty$, $n_2/n_1 \to \rho > 0$

$$\lim P(D_{n_1,n_2} > \lambda \sqrt{1/n_1 + 1/n_2}) = 1 - \sum_{i=-\infty}^{+\infty} (-1)^i e^{-2i^2\lambda^2}. \tag{14.3.57}$$

Следовательно, если положить $\delta(\alpha, n_1, n_2) = \lambda_\alpha \sqrt{1/n_1 + 1/n_2}$, где λ_α удовлетворяет условию

$$\sum_{i=-\infty}^{+\infty} (-1)^i e^{-2i^2\lambda_\alpha^2} = 1 - \alpha, \tag{14.3.58}$$

то при $n_1 \to \infty$, $n_2 \to \infty$, $n_2/n_1 \to \rho > 0$ будем иметь

$$\lim P\left(D_{n_1,n_2} > \lambda_\alpha \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = \alpha. \tag{14.3.59}$$

Критерий $W_{c\alpha}$ состоятелен для проверки гипотезы $(F_1(x), F_2(x)) \in \mathfrak{F}_0$ против альтернатив $(F_1(x), F_2(x)) \in \mathfrak{F} \setminus \mathfrak{F}_0$, т. е. можно показать, что при $F_1(x) \not\equiv F_2(x)$ и $n_1 \to \infty$, $n_2 \to \infty$, $n_2/n_1 \to \rho > 0$

$$\lim P\left(D_{n_1,n_2} > \lambda_\alpha \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 1. \tag{14.3.60}$$

Доказательство этого утверждения мы опускаем.

Отметим, что случайная величина $D_{n_1,\, n_2}$ является функцией от частот блоков $(r_1, \ldots, r_{n_1+1})$; поэтому для любого d $P(D_{n_1,\, n_2} \leqslant d)$

можно сосчитать для малых выборок простым подсчетом числа выборочных точек $(r_1, \ldots, r_{n_1+1})$, для которых $D_{n_1, n_2} \leqslant d$, и последующим делением его на $\binom{n_1+n_2}{n_1}$. Это довольно сложно при $n_1 \neq n_2$. Но при $n_1 = n_2 = n$ $D_{n,n}$ является кратным $1/n$. В этом случае Масси (1951) методом, аналогичным использованному при доказательстве **11.7.1**, установил, что

$$P\left(D_{n,n} \leqslant \frac{k}{n}\right) = \sum_{i=0}^{k} U(i, n) \Big/ \binom{2n}{n}, \tag{14.3.61}$$

где $U(\xi, \eta)$, $\xi = 0, 1, \ldots, 2k-1$; $\eta = 1, \ldots, n$, — число выборочных точек в пространстве значений $(r_1, \ldots, r_{n_1+1})$, для которых $F_{1n}(x_{(\eta)}) = (\xi + \eta - k)/n$ и $|F_{1n}(x) - F_{2n}(x)| \leqslant k/n$ для $x < x_{(\eta)}$ ($x_{(\eta)}$ — η-я порядковая статистика выборки O_{n_1}). $U(\xi, \eta)$ удовлетворяет разностным уравнениям

$$\begin{aligned} U(0, \eta+1) &= U(0, \eta) + U(1, \eta), \\ U(1, \eta+1) &= U(0, \eta) + U(1, \eta) + U(2, \eta), \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ U(2k-2, \eta+1) &= U(0, \eta) + \ldots + U(2k-1, \eta), \\ U(2k-1, \eta+1) &= U(0, \eta) + \ldots + U(2k-1, \eta) \end{aligned} \tag{14.3.62}$$

с граничными условиями

$$\begin{aligned} U(i, 0) &= 0, \qquad i = 1, \ldots, k-1, \\ U(k, 0) &= 0. \end{aligned} \tag{14.3.63}$$

Аналогичные формулы можно получить при $n_1 \neq n_2$; в этом случае D_{n_1, n_2} кратно $1/n_1 n_2$. Масси (1951), используя такие рекуррентные соотношения, составил таблицы вероятностей

$$P\left(D_{n_1, n_2} \leqslant \frac{k}{n_1 n_2}\right), \quad k = 1, \ldots, n_1 n_2,$$

для $n_1 \leqslant n_2 \leqslant 10$.

Весьма простая точная формула для $P(D_{n_1, n_2} \leqslant k/n_1 n_2)$ в случае $n_1 = n_2 = n$ найдена Гнеденко и Королюком (1951).

Чтобы получить результат Гнеденко—Королюка, введем случайные величины

$$D_n^+ = \sup_x (F_{1n}(x) - F_{2n}(x)) \tag{14.3.64}$$

и

$$D_n^- = \inf_x (F_{1n}(x) - F_{2n}(x)). \tag{14.3.64a}$$

Пусть $z_{(1)} < z_{(2)} < \ldots < z_{(2n)}$ — вариационный ряд выборки, полученной объединением выборок O_{n_1} и O_{n_2}, а ζ_t — случайная величина, принимающая значение $1/n$, если $z_{(t)}$ — элемент первой выборки, и значение $-1/n$, если $z_{(t)}$ — элемент второй выборки, $t=1, \ldots, 2n$. Положим

$$s_t=\zeta_1+\ldots+\zeta_t. \tag{14.3.65}$$

Тогда

$$D_n^+=\sup_t s_t, \qquad D_n^-=\inf_t s_t. \tag{14.3.66}$$

Если принять $s_0=0$, то ломаную, соединяющую последовательно точки (t, s_t), $t=0, 1, \ldots, 2n$, можно рассматривать как траекторию в плоскости (t, s), начинающуюся в $(0, 0)$ и оканчивающуюся в $(2n, 0)$. Заметим, что между любыми двумя последовательными целыми значениями t траектория либо поднимается, либо опускается. Имеется всего $\binom{2n}{n}$ возможных траекторий; каждая отвечает определенному расположению порядковых статистик обеих выборок среди $z_{(1)}, \ldots, z_{(2n)}$. При условии $(F_1(x), F_2(x)) \in \mathfrak{F}_0$ все эти траектории равновероятны. Поэтому вычисление $P(D_n^+ \leqslant k/n)$ сводится к определению числа траекторий, лежащих целиком над прямой $s=(k+1)/n$, и делению этого числа на $\binom{2n}{n}$. Точно так же вычисление $P(D_{n,n} \leqslant k/n)$ эквивалентно определению числа траекторий, лежащих между прямыми L^+: $s=(k+1)/n$ и $L^-: s=-(k+1)/n$, и делению его на $\binom{2n}{n}$. Рассмотрим сначала $P(D_n^+ \leqslant k/n)$ и возьмем любую траекторию, не лежащую целиком под L^+. Найдется первая точка A, когда траектория достигает прямой L^+. Рассмотрим зеркальное отображение относительно L^+ участка траектории от A до точки $(2n; 0)$; оно представляет собой ломаную, начинающуюся в A и оканчивающуюся в точке $(2n, 2(k+1)/n)$. Поэтому число траекторий, соединяющих точки $(0, 0)$ и $(2n, 0)$ и достигающих L^+, равно числу ломаных, соединяющих точки $(0, 0)$ и $(2n, 2(k+1)/n)$. Последнее же равно $\binom{2n}{n+k+1}$ — числу расположений $(n+k+1)$ «подъемов» и $(n-k-1)$ «спусков». Число траекторий, нигде не достигающих L^+, равно $\binom{2n}{n}-\binom{2n}{n+k+1}$; следовательно,

$$P\left(D_n^+ \leqslant \frac{k}{n}\right)=1-\frac{\binom{2n}{n+k+1}}{\binom{2n}{n}}. \tag{14.3.67}$$

По соображениям симметрии

$$P\left(D_n^- \geqslant -\frac{k}{n}\right)=P\left(D_n^+ \leqslant \frac{k}{n}\right). \tag{14.3.68}$$

Если положить $\lambda = k/\sqrt{2n}$, то получим

$$P(\sqrt{n/2}\,D_n^+ \leqslant \lambda) = 1 - \frac{\binom{2n}{n+\lambda\sqrt{2n}+1}}{\binom{2n}{n}}. \qquad (14.3.69)$$

Применяя формулу Стирлинга (7.6.29), находим, что

$$\lim_{n\to\infty} \frac{\binom{2n}{n+\lambda\sqrt{2n}+1}}{\binom{2n}{n}} = e^{-2\lambda^2} \qquad (14.3.70)$$

и, следовательно,

$$\lim_{n\to\infty} P(\sqrt{n/2}\,D_n^+ < \lambda) = 1 - e^{-2\lambda^2}. \qquad (14.3.71)$$

Итак,

14.3.7. *Если $F_{1n}(x)$ и $F_{2n}(x)$ — эмпирические к. ф. р., построенные по двум независимым выборкам объема n из одинаковых совокупностей с непрерывной к. ф. р., а D_n^+ и D_n^- определены формулами* (14.3.64) *и* (14.3.64а), *то $P(D_n^+ \leqslant k/n)$ и $P(D_n^- \geqslant -k/n)$ имеют одинаковые значения $P_n^*(k/\sqrt{2}\,n)$, где*

$$P_n^*\left(\frac{k}{\sqrt{2}\,n}\right) = 1 - \frac{\binom{2n}{n+k+1}}{\binom{2n}{n}} \qquad (14.3.72)$$

и

$$\lim_{n\to\infty} P_n^*\left(\frac{\lambda}{\sqrt{n}}\right) = 1 - e^{-2\lambda^2}. \qquad (14.3.73)$$

Читатель должен был заметить, что $P_n(\lambda/\sqrt{n})$, где $P_n(d)$ для одной выборки объема n определена в (11.6.5), имеет при $n\to\infty$ тот же предел, что и $P_n^*(\lambda/\sqrt{n})$.

Переходим теперь к вычислению $P(D_{n,n} \leqslant k/n)$. Пусть S_0^+ — множество траекторий, впервые достигающих прямой L^+, и S_0^- — множество траекторий, которые впервые достигают L^-. Тогда $S_0^+ \cup S_0^-$ состоит из всех траекторий, которые не лежат целиком между L^+ и L^-. Если обозначить через $N(E)$ число элементов в множестве E траекторий, то

$$N(S_0^+ \cup S_0^-) = N(S_0^+) + N(S_0^-) - N(S_0^+ \cap S_0^-), \qquad (14.3.74)$$

где

$$N(S_0^+) = N(S_0^-) = \binom{2n}{n+k+1}. \qquad (14.3.75)$$

Далее, $S_0^+ \cap S_0^-$ является объединением двух множеств S_1^+ и S_1^-, где S_1^+ состоит из всех траекторий, содержащих по крайней мере один участок ломаной, идущий от L^+ к L^-, а S_1^- состоит из всех траекторий, содержащих хотя бы один участок ломаной, идущий от L^- к L^+. Тогда

$$N(S_0^+ \cap S_0^-) = N(S_1^+ \cup S_1^-) = N(S_1^+) + N(S_1^-) - N(S_1^+ \cap S_1^-). \tag{14.3.76}$$

Рассуждения, аналогичные использованным при вычислении $N(S_0^+)$ и $N(S_0^-)$, дают

$$N(S_1^+) = N(S_1^-) = \binom{2n}{n+2(k+1)}. \tag{14.3.77}$$

Если определить S_i^+ (и аналогию S_i^-) как множество траекторий, содержащих по меньшей мере i участков ломаной, соединяющих L^+ и L^-, причем первый участок начинается на $L^+(L^-)$, то можно показать, что

$$N(S_0^+ \cup S_0^-) = 2\left[\binom{2n}{n+k+1} - \binom{2n}{n+2(k+1)} + \ldots \right.$$
$$\left. \ldots + (-1)^{r-1}\binom{2n}{n+r(k-1)}\right], \tag{14.3.78}$$

где $r = [n/(k+1)]$ — наибольшее целое число, не превосходящее $n/(k+1)$.

Так как $N(D_{n,n} \leqslant k/n) = \binom{2n}{n} - N(S_0^+ \cup S_0^-)$, то, деля обе части равенства на $\binom{2n}{n}$, получим

$$P\left(D_{n,n} \leqslant \frac{k}{n}\right) = 1 + 2\sum_{i=1}^{r} (-1)^i \frac{\binom{2n}{n+i(k+1)}}{\binom{2n}{n}} =$$

$$= \sum_{i=-r}^{+r} (-1)^i \frac{\binom{2n}{n+i(k+1)}}{\binom{2n}{n}}. \tag{14.3.79}$$

Исследуем теперь $P(D_{n,n} \leqslant k/n)$ при $n \to \infty$. Положим $\binom{2n}{n+i(k+1)} = g_n(i, k)$ и $k = \lambda\sqrt{2n}$. Для любого $\varepsilon > 0$ можно выбрать целые r' и $n_1(r', \varepsilon)$ так, чтобы

$$\left| 2 \sum_{i=r'+1}^{\infty} (-1)^i e^{-2i^2\lambda^2} \right| < \frac{\varepsilon}{3} \tag{14.3.80}$$

и

$$\left|P\left(D_{n,n}<\sqrt{\frac{2}{n}}\lambda\right)-\sum_{i=-r'}^{+r'}(-1)^i g_n(i,\ \lambda\sqrt{2n})\right|<\frac{\varepsilon}{3} \quad (14.3.81)$$

для $n>n_1(r',\ \varepsilon)$. Применяя к $g_n(i,\ \lambda\sqrt{2n})$, $i=0,\ \pm 1,\ \ldots,\ \pm r'$, формулу Стирлинга, как это уже делалось в (14.3.70), найдем, что существует значение $n_2(r',\ \varepsilon)$ такое, что

$$\left|\sum_{i=-r'}^{+r'}(-1)^i g_n(i,\ \lambda\sqrt{2n})-\sum_{i=-r'}^{+r'}(-1)^i e^{-2i^2\lambda^2}\right|<\frac{\varepsilon}{3} \quad (14.3.82)$$

при $n>n_2(r',\ \varepsilon)$.

Объединяя (14.3.80), (14.3.81) и (14.3.82), будем иметь

$$\left|P\left(\sqrt{\frac{n}{2}}D_{n,n}<\lambda\right)-\sum_{i=-\infty}^{+\infty}(-1)^i e^{-2i^2\lambda^2}\right|<\varepsilon \quad (14.3.83)$$

для $n>\max(n_1,\ n_2)$.

Таким образом, доказана теорема.

14.3.8. *Если $F_{1n}(x)$ и $F_{2n}(x)$ — эмпирические к. ф. р., построенные по двум независимым выборкам объема n из одинаковых совокупностей с непрерывной к. ф. р., то*

$$P\left(D_{n,n}\leqslant\frac{k}{n}\right)=\sum_{i=-r}^{+r}(-1)^i\binom{2n}{n+i(k+1)}\Big/\binom{2n}{n}, \quad (14.3.84)$$

где $r=[n/(k+1)]$, *и*

$$\lim_{n\to\infty}P\left(\sqrt{\frac{n}{2}}D_{n,n}<\lambda\right)=\sum_{i=-\infty}^{+\infty}(-1)^i e^{-2i^2\lambda^2}, \quad (14.3.85)$$

где $D_{n,n}$ определена в (14.3.55).

Читатель, конечно, заметил, что формула (14.3.85) является частным случаем (14.3.57), отвечающим выборкам одинакового объема. Отметим также, что $P(\sqrt{n}D_n<\lambda)$ (статистика D_n определена для одной выборки объема n в (11.7.1)) имеет своим пределом при $n\to\infty$ (14.3.85).

(f) Критерий Манна — Уитни (Уилкоксона). Рассмотрим теперь важный в задаче о двух выборках критерий, предложенный впервые Уилкоксоном (1945) для выборок одинакового объема и распространенный на случай выборок разных объемов Манном и Уитни (1947), которые получили также некоторые асимптотические результаты для больших выборок.

Пусть $O_{n_1}:(x_1,\ \ldots,\ x_{n_1})$ и $O_{n_2}:(x'_1,\ \ldots,\ x'_{n_2})$ — выборки из совокупностей с непрерывными ф. р. $F_1(x)$ и $F_2(x)$ соответственно. Определим множество случайных величин $z_{\xi\eta}$, $\xi=1, \ldots, n_1$; $\eta=1, \ldots, n_2$, следующим образом:

$$z_{\xi\eta}=\begin{cases} 1, & x_\xi < x'_\eta, \\ 0, & x_\xi > x'_\eta \end{cases} \tag{14.3.86}$$

и положим

$$U=\sum_{\eta=1}^{n_2}\sum_{\xi=1}^{n_1} z_{\xi\eta}. \tag{14.3.87}$$

Таким образом, если $(x_1, \ldots, x_{n_1})$ и $(x'_1, \ldots, x'_{n_2})$ совместно упорядочить по возрастанию, то U будет представлять собой общее число тех случаев, когда элементы выборки $(x_1, \ldots, x_{n_1})$ предшествуют элементам выборки $(x'_1, \ldots, x'_{n_2})$. При таком упорядочении элементы каждой выборки приобретают номера от 1 до n_1+n_2. Сумма номеров элементов выборки $(x_1, \ldots, x_{n_1})$ обозначается через T и называется *статистикой Уилкоксона*. Можно показать, что

$$U+T=n_1n_2+\frac{n_1(n_1+1)}{2}. \tag{14.3.88}$$

На самом деле Уилкоксон рассматривал статистику T только для выборок одинакового объема.

Критерий Манна — Уитни является односторонним и основывается на статистике U. Его критическая область W_α объема α представляет собой множество тех точек выборочного пространства, для которых $U=0, 1, \ldots, U_\alpha$, где U_α — наибольшее целое, для которого

$$P(U\leqslant U_\alpha \mid (F_1, F_2)\in \mathfrak{F}_0)\leqslant \alpha. \tag{14.3.89}$$

Вероятности $p_{n_1n_2}(U)$ отдельных значений случайной величины U при $(F_1, F_2)\in\mathfrak{F}_0$ удовлетворяют разностному уравнению

$$p_{n_1n_2}(U)=\frac{n_1}{n_1+n_2}\cdot p_{n_1-1,\,n_2}(U)+\frac{n_2}{n_1+n_2}p_{n_1,\,n_2-1}(U-n_1), \tag{14.3.90}$$

где $p_{0i}(U)=p_{i0}(U)=1$ при $U=0$ и $p_{0i}(U)=p_{i0}(U)=0$ при $U\neq 0$. Используя это разностное уравнение, Манн и Уитни (1947) составили таблицы значений $p_{n_1n_2}(U)$ для $1\leqslant n_1\leqslant n_2\leqslant 8$.

Вычислим среднее и дисперсию U для $(F_1(x), F_2(x))\in\mathfrak{F}$ (а не только для $F_1(x)\equiv F_2(x)$) — это пригодится при исследовании состоятельности теста Манна — Уитни.

Имеем

$$\mathscr{E}U=\mathscr{E}\sum_{\xi,\eta} z_{\xi\eta}=\sum_{\xi,\eta}\mathscr{E}z_{\xi\eta}=n_1n_2a, \tag{14.3.91}$$

где

$$a=\int_{-\infty}^{+\infty} F_1(x)\,dF_2(x). \tag{14.3.92}$$

Далее,

$$\mathscr{E}U^2 = \mathscr{E}\Big(\sum_{\xi,\eta} z_{\xi\eta}\Big)^2 = \sum_{\xi,\eta} \mathscr{E}z_{\xi\eta}^2 + \sum_{\substack{\xi,\eta,\zeta\\ \xi\neq\eta}} \mathscr{E}(z_{\xi\zeta}z_{\eta\zeta}) +$$

$$+ \sum_{\substack{\xi,\eta,\zeta\\ \eta\neq\zeta}} \mathscr{E}(z_{\xi\eta}z_{\xi\zeta}) + \sum_{\substack{\xi\neq\zeta\\ \eta\neq\omega}} \mathscr{E}(z_{\xi\eta}z_{\zeta\omega}) =$$

$$= n_1n_2a + n_1n_2(n_1-1)b + n_1n_2(n_2-1)c + n_1n_2(n_1-1)(n_2-1)a^2, \tag{14.3.93}$$

где

$$b = \int\limits_{-\infty}^{+\infty} (F_1(x))^2\,dF_2(x), \qquad c = \int\limits_{-\infty}^{+\infty} (1-F_2(x))^2\,dF_1(x). \tag{14.3.94}$$

Следовательно,

$$\sigma^2(U) = n_1n_2[a + (n_1-1)b + (n_2-1)c - (n_1+n_2-1)a^2]. \tag{14.3.95}$$

Если $(F_1, F_2) \in \mathfrak{F}_0$, то $a = \frac{1}{2}$, $b = c = \frac{1}{2}$ и

$$\mathscr{E}U = \frac{n_1n_2}{2}, \quad \sigma^2(U) = \frac{n_1n_2(n_1+n_2+1)}{12}. \tag{14.3.96}$$

Манн и Уитни показали, что при $(F_1, F_2) \in \mathfrak{F}_0$ статистика U распределена асимптотически по закону $N(n_1n_2/2,\ n_1n_2(n_1+n_2+1)/12)$ при больших n_1 и n_2. Если рассмотреть отношение U/n_1n_2, то

$$\mathscr{E}\left(\frac{U}{n_1n_2}\,\Big|\,\mathfrak{F}_0\right) = \frac{1}{2}, \quad \sigma^2\left(\frac{U}{n_1n_2}\,\Big|\,\mathfrak{F}_0\right) = \frac{n_1+n_2+1}{12n_1n_2}, \tag{14.3.97}$$

откуда с помощью неравенства Чебышева (3.3.5) можно вывести, что при $n_1 \to \infty$, $n_2 \to \infty$ $\frac{U}{n_1n_2}$ сходится по вероятности к постоянной $\frac{1}{2}$; $\frac{U_\alpha}{n_1n_2}$ также сходится по вероятности к $\frac{1}{2}$.

Исследуем теперь состоятельность критерия U. Пусть $\mathfrak{F}^-$ — класс пар $(F_1(x), F_2(x))$ таких, что

$$a < \frac{1}{2}. \tag{14.3.98}$$

Можно показать, что для любой пары $(F_1(x), F_2(x)) \in \mathfrak{F}^-$ имеем $b < \frac{1}{3}$, $c < \frac{1}{3}$. Среднее и дисперсия статистики $\frac{U}{n_1n_2}$ для любой пары ф. р. из $\mathfrak{F}^-$ равны

$$\mathscr{E}\left(\frac{U}{n_1n_2}\right) = a < \frac{1}{2};$$

$$\sigma^2\left(\frac{U}{n_1n_2}\right) = \frac{a + (n_1-1)b + (n_2-1)c - (n_1+n_2-1)a^2}{n_1n_2}, \tag{14.3.99}$$

откуда ясно, что при $n_1 \to \infty$, $n_2 \to \infty$ $\frac{U}{n_1 n_2}$ сходится по вероятности к постоянной $a < \frac{1}{2}$. Следовательно,

$$\lim_{n_1, n_2 \to \infty} P\left(\frac{U}{n_1 n_2} < \frac{U_\alpha}{n_1 n_2} \middle| (F_1, F_2) \in \mathfrak{F}^-\right) = 1, \qquad (14.3.100)$$

тем самым доказана теорема.

14.3.9. *Критерий Манна — Уитни состоятелен для гипотезы* $\mathscr{H}(\mathfrak{F}_0;\ \mathfrak{F}_0 \cup \mathfrak{F}^-)$, *т. е. для проверки гипотезы* $(F_1, F_2) \in \mathfrak{F}_0$ *против любой альтернативы* $(F_1, F_2) \in \mathfrak{F}^-$.

Случай, когда $(F_1, F_2) \in \mathfrak{F}^+$, $\mathfrak{F}^+$ — класс таких пар, что $a > \frac{1}{2}$, — сводится к рассмотренному перестановкой F_1 и F_2.

Если бы двусторонний тест Манн — Уитни определялся критической областью W'_α, состоящей из тех выборочных точек, для которых $U = 0, 1, \ldots, U_{\frac{1}{2}\alpha}$ и $U = U'_{\frac{1}{2}\alpha}, \ldots, n_1 n_2$, где $U_{\frac{1}{2}\alpha}$ — наибольшее целое число, для которого $P(U \leqslant U_{\frac{1}{2}\alpha} \mid (F_1, F_2) \in \mathfrak{F}_0) \leqslant \frac{1}{2}\alpha$, и $U'_{\frac{1}{2}\alpha}$ — наименьшее целое, для которого $P(U \geqslant U'_{\frac{1}{2}\alpha} \mid (F_1, F_2) \in \mathfrak{F}_0) \leqslant \frac{1}{2}\alpha$, то W'_α оказался бы состоятельным для проверки гипотезы $(F_1, F_2) \in \mathfrak{F}_M$ против любой альтернативы $(F_1, F_2) \in \mathfrak{F} \setminus \mathfrak{F}_M$, где $\mathfrak{F}_M$ — класс всех пар непрерывных к. ф. р., для которых $a = \frac{1}{2}$. Заметим, что $\mathfrak{F}_M$ не совпадает с классом пар одинаковых непрерывных к. ф. р., а $\mathfrak{F}_0 \subset \mathfrak{F}_M \subset \mathfrak{F}$.

14.4. Метод рандомизации

В этом параграфе мы рассмотрим процедуру, известную под названием метода рандомизации для построения тестов некоторых непараметрических гипотез. Имеются два типа критериев рандомизации: *критерии рандомизации компонент и критерии ранговой рандомизации.* Остановимся на них вкратце, отсылая читателя за дальнейшими подробностями к работам Фрезера (1957), Гофдинга (1948а), Лемана (1959), Лемана и Стейна (1949).

(а) Рандомизация компонент. Метод рандомизации компонент предложен Фишером (1926b). Он принципиально прост, но технически сложен в применении к конкретным задачам, кроме случаев малых выборок. В своей простейшей форме на примере одной выборки метод рандомизации компонент состоит в рассмотрении $n!$ точек, получаемых из выборочной точки $(x_1, \ldots, x_n)$ всевозможными перестановками компонент. Если $(x_1, \ldots, x_n)$ имеет плотность $f(x_1, \ldots, x_n)$, то, исключая случай нулевой вероятности, эти $n!$ точек будут различными. Обозначим множество $n!$ точек через S. Если п. в. $f_0(x_1, \ldots, x_n)$ — симметричная функция $x_1, \ldots, x_n$, то для любой перестановки $\xi_1, \ldots, \xi_n$ чисел $1, \ldots, n$

$$p_0(x_{\xi_1}, \ldots, x_{\xi_n} \mid S) = \frac{f_0(x_{\xi_1}, \ldots, x_{\xi_n})}{\sum^* f_0(x_{\xi_1}, \ldots, x_{\xi_n})}, \tag{14.4.1}$$

где $\sum^*$ означает суммирование по всем перестановкам чисел $1, \ldots, n$.

Если $f_1(x_1, \ldots, x_n)$ — альтернативная п. в., то

$$p_1(x_{\xi_1}, \ldots, x_{\xi_n} \mid S) = \frac{f_1(x_{\xi_1}, \ldots, x_{\xi_n})}{\sum^* f_1(x_{\xi_1}, \ldots, x_{\xi_n})}. \tag{14.4.2}$$

Предположим, что W_α — наиболее мощный тест объема α гипотезы $f_0(x_1, \ldots, x_n)$ против $f_1(x_1, \ldots, x_n)$, и пусть для простоты α кратно $1/n!$. Тогда для данной выборочной точки критическая область W_α должна представлять собой множество $n!\alpha$ точек S, для которого $P(W_\alpha \mid f_1)$ максимальна. Такой тест существует и единствен, если значения p_1 в $n!$ точках множества S различны, или, более общо, если существует точка из S, в которой значение p_1 не больше значений p_1 в $(n!\,\alpha - 1)$ точках S и больше значений p_1 во всех остальных точках S.

Заметим, что при проверке гипотезы $f_0(x_1, \ldots, x_n)$ против $f_1(x_1, \ldots, x_n)$, где $f_1(x_1, \ldots, x_n)$ также симметрична по $x_1, \ldots, x_n$, мощность теста W_α равна его объему, так что в этом случае критерий W_α бесполезен.

Выше мы по соображениям простоты обсуждали идеи метода рандомизации компонент для случая одной выборки $(x_1, \ldots, x_n)$. Однако наиболее интересные и важные обобщения этих идей получаются в задачах о двух или нескольких выборках, планирования экспериментов, о многомерных выборках. Но во всех этих задачах встречаются значительные технические трудности при строгом проведении метода рандомизации компонент, кроме случая выборок очень малого объема. По существу, эти трудности состоят в вычислении функции вероятности p_1 при альтернативной плотности f_1 и выборе требуемого множества из $n!\alpha$ точек S для построения W_α. Чтобы преодолеть их, отказываются от строгого применения метода рандомизации компонент и привлекают для построения критической области W_α тестовые функции $g(x_1, \ldots, x_n)$ из параметрической теории проверки гипотез. Исследуют моменты и асимптотические (в больших выборках) свойства распределений этих тестовых функций на пространстве перестановок S компонент выборки, при нулевой гипотезе и в некоторых случаях при альтернативе. Затем критическая область W_α в пространстве S строится из тех точек, в которых значения g лежат в некотором интервале или множестве.

Невозможно рассмотреть все полезные критерии рандомизации компонент. Достаточно привести два примера.

(b) Примеры критериев рандомизации компонент. Самым первым критерием рандомизации компонент был *критерий Фишера — Питмэна* для проверки гипотезы о том, что две выборки извлечены из совокупностей с одинаковыми средними, против альтернативы, согласно которой средние значения не равны. Этот критерий рассматривался впервые Фишером (1926b) для специальной задачи; более общее его исследование проведено Питмэном (1937b).

Пусть $O_{n_1}:(x_1, \ldots, x_{n_1})$ и $O_{n_2}:(x'_1, \ldots, x'_{n_2})$ — две независимые выборки из совокупностей с непрерывными ф. р. Объединим эти выборки в одну и через $y_1, \ldots, y_{n_1+n_2}$ обозначим произвольную перестановку n_1+n_2 компонент объединенной выборки. Пусть $\bar{y}$ и s^2 — выборочные среднее и дисперсия выборки $(y_1, \ldots, y_{n_1})$, а $\bar{y}'$ и s'^2 — те же характеристики выборки $(y_{n_1+1}, \ldots, y_{n_1+n_2})$. В качестве тестовой функции критерия принимается

$$g(y_1, \ldots, y_{n_1+n_2}) = \frac{n_1 n_2 (\bar{y}-\bar{y}')^2}{(n_1+n_2)\left[(n_1-1)s^2+(n_2-1)s'^2\right]+n_1 n_2(\bar{y}-\bar{y}')^2}. \tag{14.4.3}$$

Если известно, что обе выборки извлечены из нормальных совокупностей с одинаковыми дисперсиями, то $g(y_1, \ldots, y_{n_1+n_2})$ имеет вид $t^2/(n_1+n_2-2+t^2)$, где t — дробь Стьюдента для гипотезы о том, что выборки извлечены из одинаковых нормальных совокупностей.

Питмэн вычислил четыре первых момента условного распределения $g(y_1, \ldots, y_{n_1+n_2})$, при заданных $y_1, \ldots, y_{n_1+n_2}$, на множестве всех перестановок элементов $y_1, \ldots, y_{n_1+n_2}$. Так как знаменатель выражения (14.4.3) постоянен на множестве всех перестановок $y_1, \ldots, y_{n_1+n_2}$, достаточно рассматривать только числитель g, который принимает самое большее $\binom{n_1+n_2}{n_1}$ различных значений в соответствии с числом различных способов разбиения множества n_1+n_2 элементов на два подмножества, содержащих n_1 и n_2 элементов. Питмэн показал, что это условное распределение g приближенно есть бета-распределение $\mathrm{Be}\left(\frac{1}{2}, \frac{1}{2}(n_1+n_2-2)\right)$. В том случае, когда обе выборки извлечены из одинаковых нормальных совокупностей, безусловное распределение g есть в точности $\mathrm{Be}\left(\frac{1}{2}, \frac{1}{2}(n_1+n_2-2)\right)$. Критическая область W_α имеет вид $g > g_\alpha$, где $P(g > g_\alpha) = \alpha$; приближением к g_α может служить верхняя 100α-процентная точка распределения $\mathrm{Be}\left(\frac{1}{2}, \frac{1}{2} \times \right.$ $\left. \times (n_1+n_2-2)\right)$. Разность между средними значениями выборок O_{n_1} и O_{n_2} значима при 100α-процентном уровне значимости, если $g(x_1, \ldots$ $\ldots, x_{n_1}, x'_1, \ldots, x'_{n_2}) > g_\alpha$. Вальд и Вольфовитц (1944) рассматривали асимптотическое распределение g при больших значениях n_1 и n_2; данное ими доказательство эквивалентно доказательству того, что выбор в качестве g_α верхней 100α-процентной точки распределения $\mathrm{Be}\left(\frac{1}{2}, \frac{1}{2}(n_1+n_2-2)\right)$ асимптотически правилен при больших n_1+n_2.

В качестве второго примера критерия рандомизации компонент мы рассмотрим стандартный тест в Модели I дисперсионного анализа для проверки гипотезы об отсутствии эффектов по столбцам в таблице из r строк и s столбцов (см. § 10.6) в предположении, что наблюдения в каждой строке образуют выборку объема s из совокупности с непрерывной к. ф. р., а к. ф. р., отвечающие различным строкам, одинаковы с точностью до средних. Если $(x_{\xi\eta},\ \xi=1, \ldots, r;\ \eta=1, \ldots, s)$ — выборка, то рассмотренная Уэлчем (1937) и Питмэном (1938) тестовая функция есть

$$g(\{x_{\xi\eta}\})=\frac{\sum\limits_{\xi,\eta}(\bar{x}_{\cdot\eta}-\bar{x})^2}{\sum\limits_{\xi,\eta}(x_{\xi\eta}-\bar{x}_{\cdot\eta}-\bar{x}_{\xi\cdot}+\bar{x})^2+\sum\limits_{\xi,\eta}(\bar{x}_{\cdot\eta}-\bar{x})^2}, \tag{14.4.4}$$

где $\sum\limits_{\xi,\eta}$ обозначает суммирование по $\xi=1, \ldots, r;\ \eta=1, \ldots, s$, $\bar{x}_{\xi\cdot}$, $\bar{x}_{\cdot\eta}$ — средние значения ξ-й строки и η-го столбца матрицы $\|x_{\xi\eta}\|$, $\bar{x}$ — среднее значение всех rs наблюдений. Питмэн вычислил четыре первых момента g при всевозможных перестановках внутри столбцов и показал, что распределение g удовлетворительно аппроксимируется бета-распределением $\mathrm{Be}\left(\frac{1}{2}(s-1), \frac{1}{2}(r-1)(s-1)\right)$, которое является точным распределением g в нормальной теории, в предположении об отсутствии эффектов по столбцам, как было показано в §10.6 (а). Критическая область W_α состоит из тех точек в пространстве перестановок, для которых $g>g_\alpha$, где $P(g>g_\alpha)=\alpha$. Приближенно за g_α можно взять верхнюю 100α-процентную точку распределения $\mathrm{Be}\left(\frac{1}{2}(s-1), \frac{1}{2}(r-1)(s-1)\right)$.

Арнольд (1958) обобщил критерий рандомизации на случай, когда $x_{\xi\eta}$ — вектор и $s=2$. Критерий тогда эквивалентен отношению T^2 Хотеллинга для двух выборок, введенному и рассмотренному в главе 18. Уэлч (1937) использовал метод рандомизации компонент при исследовании критерия в Модели I дисперсионного анализа для латинских квадратов.

Гофдинг (1952) предпринял изучение мощности критериев рандомизации компонент (в том числе и рассмотренных нами выше) для больших выборок и показал, что эти критерии асимптотически столь же мощны, как и соответствующие параметрические критерии.

(с) Критерии ранговой рандомизации. Метод рандомизации компонент, строго говоря, не является непараметрической процедурой, так как тестовые функции и их распределения, при рандомизации компонент, зависят от значений компонент выборки (а не только от отношений порядка между ними). Однако можно добиться непараметричности, используя критерии ранговой рандомизации. В этих критериях фигурируют только отношения порядка между компонентами выборки, а не сами компоненты. Если $x_1, \ldots, x_n$ — компоненты n-мерной случайной величины с непрерывной к. ф. р. $F(x_1, \ldots, x_n)$, то *рангами* $a_1, \ldots, a_n$ компонент $x_1, \ldots, x_n$ соответственно называются номера (индексы) этих компонент при расположении их в порядке возрастания. Таким образом, a_ξ — индекс ξ-й по величине компоненты среди $(x_1, \ldots, x_n)$. Набор $(a_1, \ldots, a_n)$ является перестановкой чисел $1, \ldots, n$. Обозначим пространство $n!$ перестановок чисел $1, \ldots, n$ через S'. Всякая точка $(x_1, \ldots, x_n)$ выборочного пространства R_n, для которой $x_{\xi_1} < \ldots < x_{\xi_n}$, отображается в точку $(\xi_1, \ldots, \xi_n)$ пространства S'. Если к. ф. р. $(x_1, \ldots, x_n)$ $F_0(x_1, \ldots, x_n)$ симметрична относительно своих аргументов, то

$$p_0(\xi_1, \ldots, \xi_n) = \frac{1}{n!} \tag{14.4.5}$$

для каждой точки S'. Читатель помнит, что согласно (14.4.1) $p_0(x_{\xi_1}, \ldots, x_{\xi_n} \mid S)$ представляет собой условное распределение при данных $(x_1, \ldots, x_n)$. Распределение рангов $p_0(\xi_1, \ldots, \xi_n)$, однако, не является условным распределением. Если $F_1(x_1, \ldots, x_n)$ — произвольная непрерывная n-мерная к. ф. р., то распределение вероятностей $p_1(\xi_1, \ldots, \xi_n)$ задается формулой

$$p_1(\xi_1, \ldots, \xi_n) = \int\limits_{E_n} dF_1(x_1, \ldots, x_n), \tag{14.4.6}$$

где E_n — область $x_{\xi_1} < \ldots < x_{\xi_n}$.

Задача построения из точек S' критической области W'_α для проверки гипотезы F_0 против альтернативы F_1 вполне аналогична задаче построения из точек S критической области W_α для проверки гипотезы f_0 против альтернативы f_1, которая рассматривалась выше. Но критерий W'_α не зависит от наблюденных значений $(x_1, \ldots, x_n)$, как это было для W_α. В этом смысле W'_α является критерием, непараметрическим в более точном смысле слова, чем W_α. Конечно, если и F_0 и F_1 симметричны по своим аргументам, то мощность W'_α равна его объему. Как и в методе рандомизации компонент, имеются существенные технические трудности построения W'_α при строгом следовании принципам ранговой рандомизации. Чтобы добиться успеха в теории критериев ранговой рандомизации, нужно использовать тестовые функции $h(a_1, \ldots, a_n)$, определенные на S', из аналогичных параметрических задач проверки гипотез.

Хотя разработано значительное число тестов ранговой рандомизации, мы рассмотрим только три примера.

(d) Примеры критериев ранговой рандомизации. Критерий Уилкоксона T, определенный в (14.3.88), является критерием ранговой рандомизации для гипотезы $\mathcal{K}(\mathcal{A}; \mathcal{B})$, где $\mathcal{A}$ (нулевая гипотеза) — класс непрерывных и симметричных (n_1+n_2)-мерных к. ф. р. вида $F(x_1) \ldots F(x_{n_1}) F(x'_1) \ldots F(x'_{n_2})$, а $\mathcal{B}$ — класс непрерывных (n_1+n_2)-мерных к. ф. р. вида $F_1(x_1) \ldots F_1(x_{n_1}) F_2(x'_2) \ldots F_2(x'_{n_2})$ с $a < \frac{1}{2}$,

a определена в (14.3.92). Если обозначить $x_1, \ldots, x_{n_1}, x_1', \ldots, x_{n_2}'$ через $y_1, \ldots, y_{n_1+n_2}$ и через $a_1, \ldots, a_{n_1+n_2}$ — ранги компонент $y_1, \ldots$ $\ldots, y_{n_1+n_2}$, то статистика Уилкоксона T есть просто

$$h(a_1, \ldots, a_{n_1+n_2}) = \sum_1^{n_1} a_i. \tag{14.4.7}$$

Другой пример критерия ранговой рандомизации доставляет коэффициент ранговой корреляции. Пусть $(x_1, y_1; x_2, y_2; \ldots; x_n, y_n)$ — выборка объема n из совокупности с непрерывной к. ф. р. $F(x, y)$. Пусть $a_1, \ldots, a_n$ — ранги $x_1, \ldots, x_n$ и $b_1, \ldots, b_n$ — ранги $y_1, \ldots, y_n$. Проверяется гипотеза $\mathscr{H}(\mathscr{A}; \mathscr{B})$, где $\mathscr{B}$ — класс непрерывных $2n$-мерных к. ф. р. вида $F(x_1, y_1), \ldots, F(x_n, y_n)$, а $\mathscr{A}$ (нулевая гипотеза) — подкласс $\mathscr{B}$, для которого $F(x_\xi, y_\xi) = F_1(x_\xi) F_2(y_\xi)$, $\xi = 1, \ldots, n$, и $F_1(x)$, $F_2(y)$ — непрерывные к. ф. р. Для проверки гипотезы $\mathscr{H}(\mathscr{A}; \mathscr{B})$ используется ранговая статистика $g(a_1, \ldots, a_n, b_1, \ldots, b_n)$, представляющая собой обычный коэффициент корреляции R между двумя множествами рангов. Хотеллинг и Пэбст (1936) изучали распределение R при нулевой гипотезе и показали, что распределение $\sqrt{n}R$ при $n \to \infty$ стремится к нормальному распределению $N(0, 1)$. Олдс (1938) составил таблицы точного распределения величины S, связанной с R соотношением

$$R = 1 - \frac{6S}{n^3 - n}. \tag{14.4.8}$$

Если критическая область W_α критерия R берется в виде $R^2 \geqslant R_\alpha^2$, где R_α^2 — наименьшее целое, для которого $P(R^2 \geqslant R_\alpha^2) \leqslant \alpha$, то, как показал Гофдинг (1948b), этот критерий не будет состоятельным для проверки гипотезы $\mathscr{A}$ против альтернатив из $\mathscr{B} \setminus \mathscr{A}$.

Гофдинг (1948b) построил критерий ранговой рандомизации, который состоятелен для проверки определенной выше гипотезы $\mathscr{H}(\mathscr{A}; \mathscr{B})$. Статистика этого критерия имеет вид

$$D = \frac{(n-5)!}{n!}\left[A - 2(n-2)B + (n-2)(n-3)C\right], \tag{14.4.9}$$

где

$$\begin{aligned} A &= \sum_{\xi=1}^{n} a_\xi b_\xi (a_\xi - 1)(b_\xi - 1), \\ B &= \sum_{\xi=1}^{n} (a_\xi - 1)(b_\xi - 1) c_\xi, \\ C &= \sum_{\xi=1}^{n} c_\xi (c_\xi - 1) \end{aligned} \tag{14.4.10}$$

и c_ξ — число выборочных пар (x_ξ, y_ξ), (x_η, y_η), для которых $x_\eta < x_\xi$ и $y_\eta < y_\xi$.

Статистика D является несмещенной оценкой величины

$$\Delta=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[F(x, y)-F(x, \infty)F(\infty, y)]^2\,dF(x, y), \quad (14.4.11)$$

дисперсия которой при нулевой гипотезе равна

$$\sigma^2(D)=\frac{2(n^2+5n-32)}{8100n(n-1)(n-2)(n-3)}. \quad (14.4.12)$$

Гофдинг вычислил $\sigma^2(D)$ также для распределений из $\mathscr{B}\setminus\mathscr{A}$ и доказал, что его тест несмещенный.

Важные тесты ранговой рандомизации для проверки различных гипотез дисперсионного анализа построили Фридман (1937), Кендалл и Смит (1939), Краскел (1952), Уоллис (1939), Уормлитон (1959). Эндрюс (1954) исследовал мощности некоторых из этих критериев в больших выборках.

ЗАДАЧИ

14.1. Пусть $\mathscr{C}_0$ — класс всех непрерывных к. ф. р., у которых квантили порядков $p_1, p_2, \ldots, p_k$ равны соответственно $x_{01}, x_{02}, \ldots, x_{0k}$, $0=p_0<p_1<<\ldots<p_k<p_{k+1}=1$; $\mathscr{C}$ — класс всех непрерывных к. ф. р. Обозначим через $I_1, I_2, \ldots, I_{k+1}$ интервалы $(-\infty, x_{01}]$, $(x_{01}, x_{02}]$, $\ldots$, $(x_{0k}, +\infty)$. Рассмотрим тест W_α, задаваемый критической областью

$$\sum_{i=1}^{k+1}\frac{[n_i-n(p_i-p_{i-1})]^2}{n(p_i-p_{i-1})}>\chi^2_{k,\alpha},$$

где $n_1, \ldots, n_{k+1}$ $(n_1+\ldots+n_{k+1}=n)$ — числа элементов выборки объема n, попавших в интервалы $I_1, \ldots, I_{k+1}$ соответственно. Показать, что W_α — состоятельный критерий объема α гипотезы $\mathscr{H}(\mathscr{C}_0;\ \mathscr{C})$, если $\chi^2_{k,\alpha}$ выбрано так, чтобы $P(\chi^2>\chi^2_{k,\alpha})=\alpha$, где случайная величина χ^2 имеет распределение $C(k)$.

14.2. Пусть $\mathscr{C}_{0,p}$ — класс всех n-мерных к. ф. р. вида $F_1(x_1)F_2(x_2)\ldots F_n(x_n)$, где $F_1(x_{0p})=\ldots=F_n(x_{0p})=p$, а $\mathscr{C}_p$ — класс всех n-мерных к. ф. р. вида $F_1(x_1)F_2(x_2)\ldots F_n(x_n)$, где $F_\xi(x_{0p})=p_\xi$, $\xi=1, \ldots, n$, и $\lim\limits_{n\to\infty}(p_1+\ldots+p_n)/n=p'\neq p$. Обозначим через r число компонент n-мерной случайной величины $(x_1, \ldots, x_n)$, принадлежащих интервалу $(-\infty, x_{0p}]$, и определим W_α формулой (14.1.2). Доказать, что W_α — состоятельный тест объема α гипотезы $\mathscr{H}(\mathscr{C}_{0p};\ \mathscr{C}_p)$.

14.3. Проверить формулу (14.2.12).

14.4. Вывести из (14.2.12), что

$$\lim \mathscr{E}\left(\frac{s_i}{m}\right)=\frac{\rho^i e^{-\rho}}{i!}$$

при $m\to\infty$, $n\to\infty$ и $n/m\to\rho>0$.

14.5. Показать, что определенная в § 14.2 (c) случайная величина s_1 при $m\to\infty$, $n\to\infty$, $n=\rho m+O(1)$, $\rho>0$, асимптотически распределена по закону

$$N(m\rho e^{-\rho},\ m\rho e^{-\rho}(1-e^{-\rho}-\rho e^{-\rho})).$$

14.6. Показать, что коэффициент при $u_0^{s_0}v^n$ в (14.2.15) совпадает с коэффициентом при $u_0^{s_0}v^n$ в (14.2.16).

14.7. Доказать (14.2.24).

14.8. Доказать, что если $(r_1^*, \ldots, r_t^*)$ — любые t из случайных величин $(r_1, \ldots, r_t)$ с распределением (14.3.7), то их распределение дается формулой (14.3.8).

14.9. Проверить (14.3.13).

14.10. Проверить (14.3.14).

14.11. Вывести из (14.3.14), что

$$\lim \mathscr{E}\left(\frac{s_i}{n_1+1}\right)=\frac{\rho^i}{(1+\rho)^{i+1}}$$

при $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to \rho > 0$.

14.12 *(Продолжение).* Показать, что при том же предельном переходе

$$\lim \mathscr{E}\left(\frac{s_i}{n_1+1}' \middle| \mathfrak{F}^*\right)=\int\limits_0^1 \frac{[\rho g(u)]^i}{[1+\rho g(u)^{i+1}]}du.$$

14.13. Пусть в ситуации теоремы **14.3.1** m_i обозначает число элементов выборки O_{n_i}, превосходящих z_k, $i=1, 2$. Доказать, что функция вероятности (m_1, m_2) дается формулой

$$p(m_1, m_2)=\binom{n_1}{m_1}\binom{n_2}{m_2}\Big/\binom{n_1+n_2}{k},$$

где m_1, m_2 — неотрицательные целые числа, для которых $m_1+m_2=n_1+$ $+n_2-a$ [Муд (1954)].

14.14. Пусть $\mathfrak{F}_1$ — класс всех пар непрерывных к. ф. р. вида $(F(x), F(x-\delta))$, а $\mathfrak{F}_0$ — класс всех пар одинаковых непрерывных к. ф. р. Доказать, что критерий Манна — Уитни состоятелен для проверки гипотезы $\mathfrak{F}_0$ против любой альтернативы из $\mathfrak{F}_1$, для которой $\delta > 0$.

14.15 (*Продолжение*). Если $(x_{(1)}, \ldots, x_{(n_1)})$ и $(x'_{(1)}, \ldots, x'_{(n_2)})$ — порядковые статистики независимых выборок объемов n_1 и n_2 из совокупностей с к. ф. р. $F(x)$ и $F(x-\delta)$ соответственно, то докажите, что

$$P(x_{(r)}-x'_{(s)}<\delta)=\sum_{i=1}^{r-1}\left(\frac{s}{s+i}\right)\binom{n_1}{i}\binom{n_2}{s}\Big/\binom{n_1+n_2}{i+s}.$$

Следовательно, при $r<s$ и $r'>s'$ интервал $(x'_{(s')}-x_{(r')}, x'_{(s)}-x_{(r)})$ будет 100γ-процентным доверительным интервалом для δ, где

$$\gamma=1-\sum_{i=1}^{r-1}\left(\frac{s}{s+i}\right)\binom{n_1}{i}\binom{n_2}{s}\Big/\binom{n_1+n_2}{i+s}-\sum_{i=r'}^{n_1}\left(\frac{s'}{s'+i}\right)\binom{n_1}{i}\binom{n_2}{s'}\Big/\binom{n_1+n_2}{i+s'}$$

[Муд (1954)].

14.16. Проверить (14.3.78).

14.17. Проверить (14.3.88).

14.18. Вывести разностное уравнение (14.3.90).

14.19. Проверить (14.3.93).

14.20. Показать, что $\mathscr{E}D=\Delta$, где D определено формулой (14.4.9), а Δ — формулой (14.4.11).

14.21. *Задача Бертрана* (1887) *о баллотировке.* Пусть в урне находится N_1 бюллетеней, поданных за кандидата A, и N_2 баллотировочных бюллетеней за кандидата B, $N_1>N_2$. Предположим, что бюллетени извлекают по одному, и обозначим через y_i разность между числом голосов, набранных кандидатом A, и числом голосов, набранных B, после подсчета i бюллетеней. Если все возможные $\binom{N_1+N_2}{N_1}$ (различных) способов извлечения бюллетеней равновероятны, то докажите (используя методы, аналогичные употреблявшимся в § 14.3(с), что с вероятностью $(N_1-N_2)/(N_1+N_2)$ кандидат A при подсчете голосов был все время впереди B (т. е. все $y_1, \ldots, y_{N_1+N_2}$ положительны).

14.22 *(Продолжение).* Предположим, что $N_1=N_2=N$; тогда $y_{2N}=0$. Рассмотрим $\binom{2N}{N}$ ломаных путей, связывающих точки (0, 0) и ($2N$, 0), причем отрезки этих ломаных соединяют последовательно точки (0, 0), (1, y_1), ..., ($2N$, 0) плоскости (x, y). Отметим, что число отрезков каждой такой ломаной, лежащих над (или под) осью x, четно. Доказать, что число ломаных, у которых $2n$ отрезков лежат над осью x (и $2N-2n$ — под осью x), равно

$$\binom{2N}{N}\Big/(N+1),$$

и, следовательно, вероятность того, что $\frac{n}{N}$-я часть ломаной лежит над осью x, есть

$$P_N(n)=\frac{1}{N+1}, \qquad n=0,\ 1, \ldots, N,$$

если все ломаные равновероятны.

14.23. ***Распределение лидерства при бросании монеты.*** [Чжун и Феллер (1949)]. Обозначим через y_i (случайную) разность между числом выпадений герба и числом выпадений решетки при i первых бросаниях монеты. Рассмотрим 2^{2N} возможных ломаных путей, образованных отрезками, соединяющими последовательно точки (0, 0), (1, y_1), ..., ($2N$, y_{2N}). Показать, что число ломаных, у которых $2n$ отрезков лежат над осью x, есть

$$\binom{2n}{n}\binom{2N-2n}{N-n},$$

и, следовательно, вероятность того, что $\frac{n}{N}$-я часть ломаной расположена над осью x, равна

$$\binom{2n}{n}\binom{2N-2n}{N-n}\Big/ 2^{2N}, \qquad n=0, 1, \ldots, N,$$

в предположении, что все пути равновероятны (т. е. все $2N$ бросаний монеты независимы).

14.24 *(Продолжение).* ***Первый закон арксинуса*** [Чжун и Феллер (1949)]. Используя формулу Стирлинга, показать, что предельная к. ф. р. $G(u)$ случайной величины $\frac{n}{N}$ при $N\to\infty$ равна

$$G(u)=\frac{2}{\pi}\arcsin\sqrt{u}.$$

Заметим, что плотность предельного распределения $q(u)=1/[\pi\sqrt{u(1-u)}]$ имеет U-образную форму, причем значения на концах интервала (0, 1) бесконечны. Это свидетельствует о большей вероятности того, что относительно большая или малая часть ломаной лежит над осью x, по сравнению с вероятностью того, что близкая к $\frac{1}{2}$ часть ломаной лежит над осью x.

ГЛАВА 15

ПОСЛЕДОВАТЕЛЬНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ

15.1. Предварительные замечания

Во всем круге вопросов, рассматривавшихся в главах 8—14, объем выборки предполагался фиксированным. Соответствующие теории включали вычисление распределений некоторых статистик, т. е. функций от элементов выборки при заданном объеме выборки. Исследовалось также предельное поведение этих распределений при стремлении объема выборки к бесконечности. По существу, ни одна среди изложенных в главах 8—14 теорий не рассматривала ситуаций, в которых объем выборки сам является случайной величиной. Но в главах 6—7 мы встретились с *распределениями времени ожидания*, где число испытаний, требуемых для достижения некоторой цели, оказывается случайной величиной. Схемы, приводящие к этим распределениям времени ожидания, представляют собой примеры простых статистических процедур последовательного типа. Данная глава посвящена преимущественно критериям статистических гипотез, хотя приводятся также и некоторые результаты о последовательном оценивании.

Излагаемые результаты имеют первостепенное значение для проверки простых гипотез. Некоторые факты, касающиеся проверки сложных гипотез, можно найти в книге Вальда (1947a), а более поздние результаты — в статьях Барнарда (1952) и Кокса (1952). Дэвид и Краскел (1956) и Раштон (1950, 1952) изучали последовательные t-критерии.

Основная идея последовательного критерия принадлежит Доджу и Ромигу (1929), которые предложили схему *двойной выборки* для вынесения решения: принять или отвергнуть партию из N изделий, содержащую неизвестное число $N\theta$ дефектных изделий, где θ кратно $1/N$, $\theta \in [0, 1]$. В схеме Доджа — Ромига из партии изделий производится выборка объема n_1 и затем принимается одно из трех решений, в зависимости от числа m_1 дефектных изделий в выборке:

i) Принять партию, если $m_1 \leqslant c_1$.

ii) Отвергнуть партию, если $m_1 > c_2$.

iii) Извлечь вторую выборку объема n_2, если $c_1 < m_1 \leqslant c_2$.

Если принято решение iii), то в зависимости от числа m_2 дефектных изделий во второй выборке выносится затем одно из двух решений:

i) Принять партию, если $m_1 + m_2 \leqslant c_2$.

ii) Отвергнуть партию, если $m_1 + m_2 > c_2$.

Двойной выборочный план полностью определяется указанием четырех чисел: n_1, n_2, c_1, c_2. Ясно, что суммарный объем выборки есть случайная величина n, принимающая два значения: n_1 и n_1+n_2. Детальное рассмотрение простого и двойного выборочного планов содержится в книге Доджа и Ромига (1959).

Чтобы показать, что двойной выборочный план вместе с правилом принятия решения, связанным с ним, представляет собой статистический критерий, достаточно просто заметить, что партия отвергается тогда и только тогда, когда значение двумерной случайной величины (m_1, m_2) попадает в критическое множество $W=E_1 \bigcup E_2$. Здесь E_1 — множество точек выборочного пространства случайной величины (m_1, m_2), для которых $m_1>c_2$, а E_2 — множество точек, для которых $c_1<m_1\leqslant c_2$ и $m_1+m_2>c_2$. Вероятность $P(W\mid\theta)$ отвергнуть партию, когда доля дефектных изделий в ней равна θ, может быть без особого труда записана с помощью гипергеометрических распределений. Как функция θ $P(W\mid\theta)$ есть функция мощности критерия W. Если положить

$$L(\theta)=P(\overline{W}\mid\theta),$$

то $L(\theta)$ будет вероятностью принять партию, доля дефектных изделий в которой равна θ. Пользуясь обозначениями и терминологией из *Исторического замечания* в § 13.1, мы скажем, что партия изделий принадлежит зоне предпочтения, если $\theta\leqslant\theta_0$, зоне безразличия, если $\theta_0<\theta<\theta_1$, и зоне отклонения, если $\theta\geqslant\theta_1$. При заданных значениях риска производства α (риск ошибки 1-го рода) для $\theta=\theta_0$ и риска потребителя β (риск ошибки 2-го рода) для $\theta=\theta_1$ имеем

$$L(\theta_0)=1-\alpha, \qquad L(\theta_1)=\beta,$$

причем на практике обычно выбирают α и β в интервале [0, 01; 0, 10]. Кривая оперативной характеристики, т. е. график функции $L(\theta)$, проходит через точки (0, 0), $(\theta_0,\ 1-\alpha)$, $(\theta_1,\ \beta)$ и (1, 0).

Идея последовательной процедуры в более развитой форме принадлежит Бартки (1943), который обобщил понятие двойной выборки до понятия *многоступенчатой выборки* и распространил его на бесконечные партии с неизвестной долей дефектных изделий θ. Другие ранние примеры математического изучения многошаговых процедур включают последовательность выборочных переписей участков, занятых под джутом в Бенгалии, осуществленную под руководством Махаланобиса (1940), предложенные Хотеллингом (1941) последовательные эксперименты для нахождения максимума линии регрессии, методы стохастической аппроксимации Диксона и Муда (1948). Однако значительная часть работы пришлась на долю Вальда (1945), сделавшего главный шаг на пути построения общей теории последовательного анализа, каким мы его знаем в настоящее время. В этой главе мы даем только краткий обзор основных результатов последовательного анализа, принадлежащих в основном Вальду. Читатель, интересующийся дальнейшими деталями, должен обратиться к книге Вальда (1947а).

Более современные исследования процедур последовательного типа имеются в статьях Дворецкого, Кифера и Вольфовитца (1953a, 1953b), Кифера (1948, 1953, 1957), Кифера и Вольфовитца (1952), Роббинса (1952), Роббинса и Монро (1951) и других.

15.2. Основная схема последовательного критерия

(а) Описание событий, встречающихся в последовательном критерии. Изложенная выше схема двойной выборки Доджа — Ромига подсказывает основную структуру общего последовательного критерия. Пусть имеется случайный процесс $(x_1, x_2, \ldots)$, зависящий от вещественного параметра θ; это означает, что каждое конечное множество величин $x_1, x_2, \ldots$ имеет к. ф. р., зависящую от θ, пробегающего параметрическое пространство Ω. Таким образом, для $n = 1, 2, \ldots$ $(x_1, \ldots, x_n)$ имеет к. ф. р. $F(x_1, \ldots, x_n; \theta)$. Рассмотрим сначала простую гипотезу $\mathscr{H}(\theta_0; \Omega)$ (короче, $\mathscr{H}_0$), согласно которой $\theta = \theta_0$.

Пусть $R_1^{(1)}, \ldots, R_1^{(n)}$ — выборочные пространства величин $x_1, \ldots, x_n$ соответственно; обозначим через $R_n = R_1^{(1)} \times \ldots \times R_1^{(n)}$ выборочное пространство величины $(x_1, \ldots, x_n)$. Положим также $R_\infty = R_1^{(1)} \times \times R_1^{(2)} \times \ldots$

Для каждого целого положительного n пусть G_n°, G_n', G_n — дизъюнктные события в R_n такие, что

$$G_n^\circ \cup G_n' \cup G_n = G_{n-1} \times R_1^{(n)}. \tag{15.2.1}$$

Последовательный эксперимент начинается с наблюдения над x_1; принимаются следующие правила остановки или продолжения эксперимента:

$$\begin{array}{lll} \text{Если} & x_1 \in G_1^\circ, & \text{то принять } \mathscr{H}_0 \text{ без дальнейших наблюдений,} \\ & x_1 \in G_1', & \text{то отклонить } \mathscr{H}_0 \text{ без дальнейших наблюдений,} \\ & x_1 \in G_1, & \text{то произвести наблюдение над } x_2. \end{array} \tag{15.2.2}$$

$$\begin{array}{lll} \text{Если} & (x_1, x_2) \in G_2^\circ, & \text{то принять } \mathscr{H}_0 \text{ без дальнейших наблюдений,} \\ & (x_1, x_2) \in G_2', & \text{то отклонить } \mathscr{H}_0 \text{ без дальнейших наблюдений,} \\ & (x_1, x_2) \in G_2, & \text{то произвести наблюдение над } x_3 \end{array} \tag{15.2.3}$$

и, вообще, если

$$\begin{array}{ll} (x_1, \ldots, x_n) \in G_n^\circ, & \text{то принять } \mathscr{H}_0 \text{ без дальнейших наблюдений,} \\ (x_1, \ldots, x_n) \in G_n', & \text{то отклонить } \mathscr{H}_0 \text{ без дальнейших наблюдений,} \\ (x_1, \ldots, x_n) \in G_n, & \text{то произвести наблюдение над } x_{n+1}, \end{array} \tag{15.2.4}$$

$n = 1, 2, \ldots$

Условимся называть $G_1, G_2, \ldots$ последовательностью *событий, продолжающих эксперимент.*

Отметим, что $G_1^\circ, \ldots, G_{n-1}^\circ, G_1', \ldots, G_{n-1}'$ суть цилиндрические множества в R_n; далее, $(2n+1)$ событий $G_1^\circ, \ldots, G_{n-1}^\circ, G_1', \ldots, G_{n-1}'$, G_n°, G_n', G_n дизъюнктны и

$$G_1^\circ \cup \ldots \cup G_n^\circ \cup G_1' \cup \ldots \cup G_n' \cup G_n = R_n. \qquad (15.2.5)$$

При любом целом положительном n множества $G_1^\circ, \ldots, G_n^\circ, G_1', \ldots, G_n'$, G_n цилиндрические в R_∞ и их объединение совпадает с R_∞.

События $G_1^\circ, \ldots, G_n^\circ, G_1', \ldots, G_n', G_n$, $n=1, 2, \ldots$, вместе с заданным формулами (15.2.2), (15.2.3) и (15.2.4) правилом принятия решения определяют последовательный критерий S гипотезы $\mathscr{H}_0$. Критической областью W_α является множество $G_1' \cup G_2' \cup \ldots$, где $G_1', G_2', \ldots$ выбраны так, чтобы $P(W_\alpha \mid \theta_0)=\alpha$. Процедуру построения шаг за шагом последовательного критерия S мы будем называть *последовательным процессом* для S.

Стоит отметить, что всякий тест гипотезы $\mathscr{H}_0$, построенный по выборке фиксированного объема n и задаваемый критической областью W, можно рассматривать как *вырожденный последовательный тест,* в котором $G_1^\circ, \ldots, G_{n-1}^\circ, G_1', \ldots, G_{n-1}', G_n$ суть пустые множества, а $G_n'=W$, $G_n^\circ=\overline{W}$.

(b) Вероятности, связанные с последовательным критерием. Вероятности принять $\mathscr{H}_0$, отклонить $\mathscr{H}_0$, произвести наблюдение над x_{n+1} после наблюдения над x_n равны соответственно

$$P(G_n^\circ \mid \theta), \ P(G_n' \mid \theta), \ P(G_n \mid \theta). \qquad (15.2.6)$$

Мы уже отмечали, что события $G_1^\circ, \ldots, G_n^\circ, G_1', \ldots, G_n', G_n$ дизъюнктны и их вероятности являются функциями параметра θ. Положим

$$\begin{aligned} L_n(\theta) &= P(G_1^\circ \mid \theta)+\ldots+P(G_n^\circ \mid \theta), \\ M_n(\theta) &= P(G_1' \mid \theta)+\ldots+P(G_n' \mid \theta), \\ N_n(\theta) &= P(G_n \mid \theta) \end{aligned} \qquad (15.2.7)$$

и

$$\begin{aligned} L(\theta) &= \lim_{n\to\infty} L_n(\theta)=P(G^\circ \mid \theta), \\ M(\theta) &= \lim_{n\to\infty} M_n(\theta)=P(G' \mid \theta), \\ N(\theta) &= \lim_{n\to\infty} N_n(\theta), \end{aligned} \qquad (15.2.8)$$

где $G^\circ=G_1^\circ \cup G_2^\circ \cup \ldots$ и $G'=G_1' \cup G_2' \cup \ldots$ Эти пределы существуют, поскольку функции $L_n(\theta)$, $M_n(\theta)$, $N_n(\theta)$ неотрицательны, $L_n(\theta)$ и $M_n(\theta)$ неубывающие по $n=1, 2, \ldots$ и $L_n(\theta)+M_n(\theta)+N_n(\theta)=1$. Величины $L(\theta')$ и $M(\theta')$ представляют собой вероятности принятия и отклонения гипотезы $\mathscr{H}_0$, когда истинное значение параметра θ есть θ'. Функция параметра θ $L(\theta)$ называется *оперативной характеристикой* последовательного критерия S. $N(\theta)$ есть вероятность того, что последовательный процесс для S продолжается бесконечно. Равенство $N(\theta)=0$ означает, что этот последовательный процесс с вероятностью 1 оканчивается в конечное число шагов. Очевидно, что

$$L(\theta)+M(\theta)+N(\theta)=1. \tag{15.2.9}$$

Если положить $G_n^* = G_n^\circ \bigcup G_n'$ и

$$p(n\mid\theta)=P(G_n^*\mid\theta), \tag{15.2.10}$$

то $p(n\mid\theta)$ будет вероятностью того, что последовательный процесс для S заканчивается (принятием или отклонением гипотезы $\mathscr{H}_0$) после n наблюдений, когда истинное значение параметра есть θ.

Если последовательный процесс для S с вероятностью 1 длится конечное число шагов, т. е. если $N(\theta)=0$, то среднее число наблюдений до окончания процесса равно

$$\mathscr{E}(n\mid\theta)=\sum_{t=1}^{\infty} tp(t\mid\theta) \tag{15.2.11}$$

и, конечно, является функцией θ. В последовательном анализе $\mathscr{E}(n\mid\theta)$ обычно называют *средним объемом выборки*. Для вырожденного последовательного теста, т. е. теста, построенного по выборке фиксированного объема n, имеем

$$\mathscr{E}(n\mid\theta)=n.$$

Заметим, что поскольку

$$\mathscr{E}(n\mid\theta)\geqslant\sum_{t=1}^{m} tp(t\mid\theta)+mP(G_m\mid\theta),$$

$m=1, 2, \ldots$, то условие $\lim\limits_{m\to\infty} mP(G_m\mid\theta)=0$ необходимо для конечности среднего объема выборки.

Для случая независимых случайных величин с одинаковыми (и конечными) средними значениями имеется следующая общая теорема относительно $\mathscr{E}(n\mid\theta)$, принадлежащая Вальду (1945) и Блекуэлу (1946).

15.2.1. Пусть $(x_1, x_2, \ldots)$ — *последовательность независимых случайных величин, распределения которых зависят от параметра* θ *и имеют одно и то же (конечное) среднее значение* $\mathscr{E}(x\mid\theta)$, *а* S — *последовательный критерий, для которого* $\mathscr{E}(n\mid\theta)$ *конечно. Если* $x_1+\ldots+x_n$ — *сумма случайных величин, наблюдения над которыми предшествовали окончанию последовательного процесса, то*

$$\mathscr{E}(x_1+\ldots+x_n\mid\theta)=\mathscr{E}(x\mid\theta)\,\mathscr{E}(n\mid\theta). \tag{15.2.12}$$

Для доказательства теоремы 15.2.1 заметим, что

$$\mathscr{E}(x_1+\ldots+x_n\mid\theta)=\mathscr{E}x_1+\mathscr{E}(x_2\mid x_1\in G_1)\,P(G_1\mid\theta)+ \\ +\mathscr{E}(x_3\mid(x_1, x_2)\in G_2)\,P(G_2\mid\theta)+\ldots, \tag{15.2.13}$$

где множества $G_1, G_2, \ldots$ определены в (15.2.4). Так как случайные величины $x_1, x_2, \ldots$ взаимно независимы и средние значения этих величин равны $\mathcal{E}(x \mid \theta)$, то (15.2.13) сводится к

$$\mathcal{E}(x_1 + x_2 + \ldots + x_n \mid \theta) = \mathcal{E}(x \mid \theta)\,[1 + P(G_1 \mid \theta) + P(G_2 \mid \theta) + \ldots]. \tag{15.2.14}$$

Пусть $G_n^* = G_n^\circ \cup G_n'$, $n = 1, 2, \ldots$ Поскольку $G_1^*, \ldots, G_n^*, G_n$ при каждом n дизъюнктны в R_∞ и сумма вероятностей этих событий равна 1, а условие $\mathcal{E}(n \mid \theta) < \infty$ влечет за собой $N(\theta) = 0$, то

$$1 = \sum_{t=1}^{\infty} P(G_t^* \mid \theta),$$

$$P(G_t \mid \theta) = \sum_{t'=t+1}^{\infty} P(G_{t'}^* \mid \theta), \qquad t = 1, 2, \ldots \tag{15.2.15}$$

Подставляя значения $P(G_t \mid \theta)$ в (15.2.14), получим

$$\mathcal{E}(x_1 + x_2 + \ldots + x_n \mid \theta) = \mathcal{E}(x \mid \theta) \sum_{t=1}^{\infty} tP(G_t^* \mid \theta), \tag{15.2.16}$$

что как раз совпадает с (15.2.12). Доказательство теоремы **15.2.1** завершено.

(с) Критерии для выбора последовательного теста. При формальном определении последовательного критерия мы не налагали никаких ограничений на множества G_n°, G_n', G_n, кроме естественных: при каждом n G_n°, G_n', G_n должны быть дизъюнктными множествами, удовлетворяющими условию (15.2.1). При построении последовательного критерия S одна из основных задач состоит в выборе множеств G_n°, G_n', G_n. Очевидно, естественным требованием к последовательному критерию S — и мы будем налагать его всегда — является окончание последовательного процесса для S с вероятностью 1 за конечное число шагов, т. е.

$$N(\theta) = 0. \tag{15.2.17}$$

Как и во всех статистических критериях, другое желательное условие фиксирует заранее вероятность ошибки 1-го рода, т. е. величину

$$L(\theta_0) = 1 - \alpha. \tag{15.2.18}$$

При сравнении двух последовательных тестов (как и в случае тестов, построенных по выборкам фиксированного объема) используют также вероятность β ошибки 2-го рода для некоторого значения $\theta = \theta_1$:

$$L(\theta_1) = \beta. \tag{15.2.19}$$

Будем обозначать через $\mathcal{H}_0$ гипотезу $\theta = \theta_0$ и через $\mathcal{H}_1$ — гипотезу $\theta = \theta_1$.

Условия (15.2.17), (15.2.18) и (15.2.19) дают

$$M(\theta_0) = \alpha, \qquad M(\theta_1) = 1 - \beta. \tag{15.2.20}$$

Если $\theta_0 < \theta_1$, то подмножества параметрического пространства $\theta \leqslant \theta_0$, $\theta_0 < \theta < \theta_1$, $\theta \geqslant \theta_1$ называются соответственно *областью принятия, областью безразличия* и *областью отклонения* гипотезы $\mathscr{H}_0$. В случае $\theta_0 > \theta_1$ эти области строятся аналогично.

Типичный вид графика функции $L(\theta)$ при условиях (15.2.17), (15.2.18) и (15.2.19) показан на рис. 15.1.

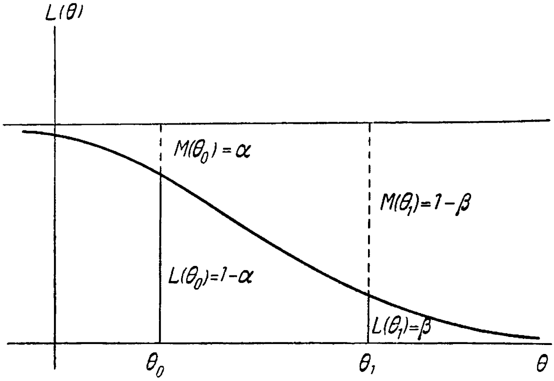

Рис. 15.1.

Последовательный критерий, для которого $L(\theta)$ удовлетворяет условиям (15.2.18) и (15.2.19), называется последовательным критерием *силы* $(\alpha, \theta_0; \beta, \theta_1)$.

Если S_1 и S_2 — два последовательных критерия, оперативные характеристики которых $L_1(\theta)$ и $L_2(\theta)$ таковы, что $L_1(\theta_0) = L_2(\theta_0) = 1 - \alpha$, $L_1(\theta_1) < L_2(\theta_2)$, то говорят, что S_1 *сильнее* S_2 для проверки $\mathscr{H}_0$ против $\mathscr{H}_1$. Если S_1 *сильнее* всех последовательных критериев S_2, то говорят, что S_1 имеет *максимальную силу*.

Если средние объемы выборок $\mathscr{E}_1(n \mid \theta)$ и $\mathscr{E}_2(n \mid \theta)$ двух последовательных критериев S_1 и S_2 одинаковой силы удовлетворяют условию

$$\mathscr{E}_1(n \mid \theta) \leqslant \mathscr{E}_2(n \mid \theta), \quad \theta = \theta_0, \ \theta_1, \tag{15.2.21}$$

и знак равенства не имеет места одновременно для θ_0 и θ_1, то S_1 считается предпочтительнее S_2 для проверки $\mathscr{H}_0$ против $\mathscr{H}_1$.

Если (15.2.21) имеет место для всех критериев S_2, отличных от S_1, то $S_1 = S^*$ называется критерием максимальной эффективности. В § 15.4 мы покажем, что такой последовательный критерий S^* может быть построен с помощью метода, основанного на отношении вероятностей.

15.3. Декартовы последовательные критерии

(а) Описание и свойства критерия. Прежде чем переходить к наилучшему последовательному критерию, полезно уяснить главные идеи на примере простого, хотя и не очень эффективного критерия, основанного на биномиальном распределении времени ожидания (6.5.15). Пусть $(x_1, x_2, \ldots)$ — последовательность независимых случайных величин, имеющих одну и ту же к. ф. р. $F(x;\ \theta)$.

Рассмотрим следующий последовательный критерий S для проверки $\mathscr{H}_0$ против $\mathscr{H}_1$. Пусть G_0°, G_0', G_0 — дизъюнктные *исходные множества*, объединение которых дает всю вещественную ось, и G_n°, G_n', G_n — множества G_0°, G_0', G_0 на $R_1^{(n)}$, $n=1, 2, \ldots$, где $R_1^{(n)}$ — выборочное пространство величины x_n, $n=1, 2, \ldots$ Тогда задаваемые в (15.2.4) дизъюнктные множества $G_1^\circ, \ldots, G_n^\circ, G_1', \ldots, G_n', G_n$ можно строить таким образом:

$$\begin{aligned}
&G_1^i = \{x_1 \in G_1^{(i)}\}, \qquad i = {}^\circ, {}', \\
&G_2^i = \{(x_1, x_2) \in G_{(1)} \times G_{(2)}\}, \qquad i = {}^\circ, {}', \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&G_n^i = \{(x_1, \ldots, x_n) \in G_{(1)} \times \ldots \times G_{(n-1)} \times G_{(n)}^i\}, \qquad i = {}^\circ, {}', \\
&G_n = \{(x_1, \ldots, x_n) \in G_{(1)} \times \ldots \times G_{(n)}\}.
\end{aligned} \tag{15.3.1}$$

Таким образом, последовательный процесс для S состоит в том, чтобы проводить по одному наблюдению в единицу времени над величинами $(x_1, x_2, \ldots)$ до тех пор, пока одна из них не попадет в множество $G_0^\circ \cup G_0'$ вещественной оси. Если результат наблюдения оказался принадлежащим множеству G_0° (G_0'), то принимается гипотеза $\mathscr{H}_0$ ($\mathscr{H}_1$). Условимся называть такой критерий *декартовым последовательным критерием*.

Положим

$$\begin{aligned}
a(\theta) &= \int_{G_0^\circ} dF(x;\ \theta), \\
b(\theta) &= \int_{G_0'} dF(x;\ \theta), \\
c(\theta) &= \int_{G_0} dF(x;\ \theta) = 1 - a(\theta) - b(\theta).
\end{aligned} \tag{15.3.2}$$

Обозначим $a(\theta)$, $b(\theta)$ и $c(\theta)$ просто через a, b, c, и пусть $a(\theta_i) = a_i$, $b(\theta_i) = b_i$, $c(\theta_i) = c_i$, $i = 0, 1$. Тогда

$$P(G_t^\circ \mid \theta) = ac^{t-1},\ P(G_t' \mid \theta) = bc^{t-1},\ P(G_n \mid \theta) = c^n, \quad t = 1, \ldots, n \tag{15.3.3}$$

и

$$\begin{aligned}
L_n(\theta) &= \frac{a(1-c^n)}{1-c}, \\
M_n(\theta) &= \frac{b(1-c^n)}{1-c}, \qquad N_n(\theta) = c^n.
\end{aligned} \tag{15.3.4}$$

Если G_0 выбрано так, что $c(\theta)<1$, то $\lim\limits_{n\to\infty} N_n(\theta)=0$ и

$$L(\theta)=\frac{a}{a+b}, \qquad M(\theta)=\frac{b}{a+b}. \tag{15.3.5}$$

Вероятность того, что последовательный процесс заканчивается наблюдением над x_n, равна

$$p(n\mid\theta)=(a+b)\,c^{n-1}, \tag{15.3.6}$$

и средний объем выборки есть

$$\mathscr{E}(n\mid\theta)=(a+b)\sum_{n=1}^{\infty} nc^{n-1}=\frac{1}{a+b}. \tag{15.3.7}$$

Мы видим, что $p(n\mid\theta)$ — просто специальный случай биномиального распределения времени ожидания (6.5.15) для $k=1$.

Если вероятности ошибок 1-го и 2-го рода должны быть равны α и β при $\theta=\theta_0$ и $\theta=\theta_1$ соответственно, то множества G_0°, G_0' следует выбирать таким образом, чтобы

$$\frac{a_0}{a_0+b_0}=1-\alpha, \qquad \frac{a_1}{a_1+b_1}=\beta, \tag{15.3.8}$$

т. е.

$$\frac{a_0}{b_0}=\frac{1-\alpha}{\alpha}, \qquad \frac{a_1}{b_1}=\frac{\beta}{1-\beta}. \tag{15.3.9}$$

Если для заданных значений α, β и $\mathscr{E}(n\mid\theta_0)$ (равного $1/(a_0+b_0)$) имеется несколько множеств G_0°, G_0', приводящих к последовательным критериям с такими α, β, $\mathscr{E}(n\mid\theta_0)$, то выбор множеств, дающих самый сильный декартов последовательный критерий, осуществляется с помощью теоремы Неймана — Пирсона **13.2.1.** Если $f(x;\ \theta)$ — плотность распределения $F(x;\ \theta)$, то за G_0° надо взять множество тех x, для которых

$$\frac{f(x;\ \theta_1)}{f(x;\ \theta_0)}\leqslant k_0, \tag{15.3.10}$$

а за G_0' — множество

$$\frac{f(x;\ \theta_1)}{f(x;\ \theta_0)}\geqslant k_1, \tag{15.3.11}$$

где $k_1>k_0>0$ выбираются так, чтобы при заданных α, β и $\mathscr{E}(n\mid\theta_0)$ удовлетворить условиям (15.3.8). Всякий иной выбор множеств G_0° и G_0', который приводит к тем же самым значениям вероятности ошибки 1-го рода и $\mathscr{E}(n\mid\theta_0)$, порождает последовательный критерий S, для которого $a_1/(a_1+b_1)>\beta$, т. е. вероятность ошибки 2-го рода будет большей, чем у последовательного критерия с множествами G_0° и G_0', задаваемыми условиями (15.3.10) и (15.3.11).

Доказательство этого утверждения аналогично доказательству теоремы **13.2.1** и предоставляется читателю в качестве упражнения (задача **15.6**).

Замечание. Если в единицу времени производится r наблюдений над случайными величинами последовательности $(x_1, x_2, \ldots)$, то мы получаем *r-кратный декартов последовательный критерий* S. Достаточно рассмотреть только случай, когда области принятия и отклонения гипотезы в выборочном пространстве R_r величин $(x_1, \ldots, x_r)$ определяются отношением правдоподобия. Таким образом, G_0° — это множество тех $(x_1, \ldots, x_r)$, для которых

$$\prod_{i=1}^{r} \frac{f(x_i;\ \theta_1)}{f(x_i;\ \theta_0)} \leqslant k_0, \tag{15.3.12}$$

и G_0' — множество $(x_1, \ldots, x_r)$ с условием

$$\prod_{i=1}^{r} \frac{f(x_i;\ \theta_1)}{f(x_i;\ \theta_0)} \geqslant k_1, \tag{15.3.13}$$

где $k_1 > k_0 > 0$ выбираются так, чтобы удовлетворить условию (15.3.8). Функции $a(\theta)$, $b(\theta)$, $c(\theta)$ по определению равны $P(G_0^\circ \mid \theta)$, $P(G_0' \mid \theta)$, $P(G_0 \mid \theta)$. Вполне аналогично теореме **13.2.1** можно доказать, что такой выбор множеств G_0° и G_0' приводит к декартову последовательному критерию, более сильному, чем критерий, отвечающий любому другому выбору G_0° и G_0' в пространстве $(x_1, \ldots, x_r)$ с теми же значениями $a_0/(a_0 + b_0) = \alpha$ и $\mathfrak{E}(n \mid \theta_0)$ (и даже более сильному, чем s-кратный декартов критерий с $s < r$).

Читателю, конечно, ясно, что за счет подходящего выбора значения r можно добиться любой близости вероятностей ошибок 1-го и 2-го рода r-кратного декартова последовательного критерия к наперед заданным величинам α и β.

(b) Применение к проверке непараметрической гипотезы. Как уже отмечалось, декартов последовательный критерий не является лучшим среди последовательных критериев. Однако, если относительно $F(x;\ \theta)$ известно лишь, что она непрерывна, то декартов тест становится непараметрическим последовательным критерием. Действительно, пусть $F(x;\ \theta_0)$ и $F(x;\ \theta_1)$ — непрерывные к. ф. р., которые мы обозначим $F_0(x)$ и $F_1(x)$. Выберем G_0° и G_0' в виде интервалов $(-\infty, a)$ и $(b, +\infty)$, $a < b$; тогда (15.3.8) можно записать в виде

$$\frac{F_0(a)}{F_0(a) + 1 - F_0(b)} = 1 - \alpha, \qquad \frac{F_1(a)}{F_1(a) + 1 - F_1(b)} = \beta. \tag{15.3.14}$$

Декартов последовательный критерий, отвечающий такому выбору множеств G_0° и G_0', является тестом гипотезы, согласно которой непрерывная к. ф. р. имеет в точках a и b квантили $F_0(a)$ и $1 - \alpha F_0(a)/(1-\alpha)$ соответственно, против альтернативы, в соответствии с которой непрерывная к. ф. р. имеет в точках a и b квантили $F_1(a)$ и $1 - (1-\beta) F_1(a)/\beta$. Средний объем выборки равен $[F_0(a) + 1 - F_0(b)]^{-1}$, если истинным распределением оказывается $F_0(x)$, и $[F_1(a) + 1 - F_1(b)]^{-1}$, если им оказывается $F_1(x)$.

15.4. Последовательный критерий отношения вероятностей

(a) Определение критерия. Рассмотренная в предыдущем параграфе задача выбора множеств G_0°, G_0' и G_0, приводящего к сильнейшему декартову последовательному критерию гипотезы $\mathcal{H}_0$ против

$\mathscr{H}_1$, подсказывает, каков вообще наилучший возможный выбор множеств $G_1^\circ, \ldots, G_n^\circ, G_1', \ldots, G_n', G_n$, $n=1, 2, \ldots$, для последовательного теста. Именно, удовлетворяющие условию (15.2.1) множества G_n°, G_n', G_n в R_n, $n=1, 2, \ldots$, следует выбирать соответственно в виде

$$\frac{Q_{1n}}{Q_{0n}} \leqslant k_0, \quad \frac{Q_{1n}}{Q_{0n}} \geqslant k_1, \qquad k_0 < \frac{Q_{1n}}{Q_{0n}} < k_1, \tag{15.4.1}$$

где

$$Q_{in} = \prod_{t=1}^{n} f(x_t;\ \theta_i), \qquad i=0,\ 1, \tag{15.4.2}$$

а $f(x;\ \theta_0)$ и $f(x;\ \theta_1)$ — взаимно абсолютно непрерывные функции вероятности или плотности распределения вероятностей. В этом параграфе мы рассмотрим только случай плотностей; полученные результаты с тривиальными изменениями справедливы и для функций вероятности.

Критерий, основанный на множествах (15.4.1), был впервые предложен Вальдом (1945), он называется *последовательным критерием отношения вероятностей*. Критерий, следовательно, состоит в принятии $\mathscr{H}_0$, принятии $\mathscr{H}_1$ или в проведении наблюдения над x_{n+1}, после того как сделано наблюдение над x_n, в зависимости от того, удовлетворяет точка $(x_1, \ldots, x_n)$ первому, второму или третьему неравенству из (15.4.1), $n=1, 2, \ldots$ Постоянные k_0 и k_1 фиксируются так, чтобы

$$L(\theta_0) = 1-\alpha, \qquad L(\theta_1) = \beta, \tag{15.4.3}$$

т. е. чтобы вероятности ошибок 1-го и 2-го рода равнялись α и β.

(b) Свойства функции распределения числа наблюдений до окончания последовательного процесса отношения вероятностей. Посмотрим сначала, оканчивается ли с вероятностью 1 за конечное число шагов последовательный процесс отношения вероятностей. Пусть

$$z = \log \frac{f(x;\ \theta_1)}{f(x;\ \theta_0)}, \tag{15.4.4}$$

и обозначим через $H(z;\theta)$ ф. р. z для фиксированного значения параметра $\theta \in \Omega$. Предположим, что при $\theta_0 \neq \theta_1$ z оказывается невырожденной случайной величиной. Положим

$$z_t = \log \frac{f(x_t;\ \theta_1)}{f(x_t;\ \theta_0)}, \qquad t=1,\ 2,\ \ldots \tag{15.4.5}$$

Тогда $(z_1, z_2, \ldots)$ будет случайным процессом, порожденным выборкой из совокупности с функцией распределения $H(z;\ \theta)$. Последовательный процесс заканчивается при первом n, для которого нарушается неравенство

$$\log k_0 < z_1 + \ldots + z_n < \log k_1. \tag{15.4.6}$$

Заметим, что событие G_n состоит в выполнении неравенства (15.4.6).

Пусть для некоторого целого r $n=mr$, $\zeta_1=z_1+\ldots+z_r$, ..., $\zeta_m=z_{(m-1)r+1}+\ldots+z_{mr}$. Тогда $(\zeta_1, \ldots, \zeta_m)$ оказывается выборкой из совокупности с некоторой функцией распределения $J(\zeta;\ \theta)$. Введем событие I_{mr}

$$|\zeta_i|<D, \qquad i=1, \ldots, m, \tag{15.4.7}$$

где $D=\max(|\log k_0|, |\log k_1|)$. Так как $G_{mr}\subset I_{mr}$, то

$$N_{mr}(\theta)=P(G_{mr}\,|\,\theta)\leqslant P(I_{mr}\,|\,\theta)=p^m, \tag{15.4.8}$$

где

$$p=P(|\zeta_1|)<D).$$

Ввиду невырожденности z она имеет положительную дисперсию σ^2. Дисперсия ζ_1 равна $r\sigma^2$ и может быть сделана больше D^2, если выбрать $r>D^2/\sigma^2$. Но в этом случае

$$P(\zeta_1^2<D^2)=p<1.$$

Если перейти в (15.4.8) к пределу при $m\to\infty$, получим

$$\lim_{n\to\infty} P(G_n\,|\,\theta)=\lim_{m\to\infty} P(G_{mr}\,|\,\theta)=N(\theta)=0.$$

Итак,

15.4.1. *Если случайная величина z из* (15.4.4) *невырожденная (т. е. ее дисперсия положительна), то последовательный процесс отношения вероятностей заканчивается с вероятностью* 1 *в конечное число шагов.*

На самом деле верна более сильная теорема Стейна (1946) о числе наблюдений до окончания последовательного процесса.

15.4.2. *Если случайная величина z невырожденная, то все моменты величины n конечны.*

Для доказательства **15.4.2** рассмотрим производящую функцию моментов $\psi(u)$ случайной величины n

$$\psi(u)=\mathcal{E}e^{un}=\sum_{t=1}^{\infty} e^{ut}P(G_t^*\,|\,\theta), \tag{15.4.9}$$

где $G_t^*=G_t^0\cup G_t'$ — событие, состоящее в окончании последовательного процесса после t наблюдений. Если сгруппировать члены, стоящие в правой части (15.4.9), по r штук, получим для $u\geqslant 0$

$$\psi(u)\leqslant e^{ru}P(0<n\leqslant r)+e^{2ru}P(r<n\leqslant 2r)+\ldots\leqslant$$
$$\leqslant e^{ru}P(n>0)+e^{2ru}P(n>r)+\cdots$$

Но если дисперсия z положительна, то в соответствии с (15.4.8)

$$P(n\geqslant mr)\leqslant p^m.$$

Следовательно,

$$\psi(u)\leqslant e^{ru}+e^{2ru}p+e^{3ru}p^2+\ldots=e^{ru}(1-e^{ru}p)^{-1}. \tag{15.4.10}$$

Для $u\leqslant 0$ аналогично получим

$$\psi(u)\leqslant e^{u}(1-e^{ru}p)^{-1}. \tag{15.4.10a}$$

Таким образом, для любого вещественного u с условием $e^{ru}p < 1$ ряд для $\psi(u)$ сходится. Интервал сходимости содержит точку $u = 0$ в качестве внутренней, поэтому k-я производная $\psi(u)$ в нуле существует, что обеспечивает конечность k-го момента величины n, $k = 1$ 2, ...

Отметим также следствие из теоремы **15.2.1**, важное для вычисления $\mathcal{E}(n \mid \theta)$ и для доказательства эффективности последовательного критерия отношения вероятностей

15.4.3. *Если z — невырожденная случайная величина, то*

$$\mathcal{E}(z_1 + z_2 + \ldots + z_n \mid \theta) = \mathcal{E}(z \mid \theta)\, \mathcal{E}(n \mid \theta). \tag{15.4.11}$$

Конечность $\mathcal{E}(n \mid \theta)$ следует из **15.4.2**, после чего непосредственно применяется теорема **15.2.1.**

(c) Определение граничных постоянных k_0 и k_1 для последовательных критериев отношения вероятностей. Рассмотрим теперь задачу определения *граничных* постоянных k_0 и k_1 для последовательного критерия отношения вероятностей силы (α, θ_0; β, θ_1). Точное определение k_0 и k_1 является трудной проблемой, в то время как неравенства и хорошие приближения для k_0 и k_1 получаются сравнительно легко.

Согласно определению последовательного критерия отношения вероятностей во всех точках множества G_n°

$$Q_{1n} \leqslant k_0 Q_{0n}. \tag{15.4.12}$$

Из (15.2.8) следует, что

$$L(\theta) = P(G_1^\circ \mid \theta) + P(G_2^\circ \mid \theta) + \ldots \tag{15.4.13}$$

Если критерий имеет силу (α, θ_0; β, θ_1), то

$$L(\theta_0) = 1 - \alpha, \quad L(\theta_1) = \beta. \tag{15.4.14}$$

Из (15.4.12) выводим, что

$$P(G_n^\circ \mid \theta_1) \leqslant k_0 P(G_n^\circ \mid \theta_0), \qquad n = 1, 2, \ldots \tag{15.4.15}$$

Следовательно,

$$\sum_{n=1}^{\infty} P(G_n^\circ \mid \theta_1) \leqslant k_0 \sum_{n=1}^{\infty} P(G_n^\circ \mid \theta_0),$$

т. е.

$$L(\theta_1) \leqslant k_0 L(\theta_0). \tag{15.4.16}$$

Комбинируя это неравенство с (15.4.14), получим

$$k_0 \geqslant \frac{\beta}{1-\alpha}.$$

Если провести такие же рассуждения для множества G_n', то будем иметь

$$M(\theta_1) \geqslant k_1 M(\theta_0). \tag{15.4.17}$$

Так как $N(\theta)=0$, то $M(\theta)=1-L(\theta)$. Комбинация (15.4.17) и (15.4.14) дает второе неравенство:

$$k_1 \leqslant \frac{1-\beta}{\alpha},$$

Следовательно,

15.4.4. *Постоянные k_0 и k_1 последовательного критерия отношения вероятностей силы $(\alpha, \theta_0; \beta, \theta_1)$ удовлетворяют неравенствам*

$$k_0 \geqslant \frac{\beta}{1-\alpha}, \qquad k_1 \leqslant \frac{1-\beta}{\alpha}. \tag{15.4.18}$$

Исследуем теперь действительную силу последовательного критерия отношения вероятностей при выборе граничных постоянных в виде

$$k_0 = \frac{\beta}{1-\alpha}, \qquad k_1 = \frac{1-\beta}{\alpha}. \tag{15.4.19}$$

Условимся называть $(\alpha, \theta_0; \beta, \theta_1)$ *предполагаемой силой* критерия, а $(\alpha', \theta_0; \beta', \theta_1)$ — его *действительной силой*. Тогда из (15.4.8) будем иметь

$$\frac{\beta}{1-\alpha} \geqslant \frac{\beta'}{1-\alpha'} \quad \text{и} \quad \frac{1-\beta}{\alpha} \leqslant \frac{1-\beta'}{\alpha'}. \tag{15.4.20}$$

Из этих двух неравенств выводим

$$\beta' \leqslant \frac{\beta}{1-\alpha} \quad \text{и} \quad \alpha' \leqslant \frac{\alpha}{1-\beta}, \tag{15.4.21}$$

а также

$$\alpha' + \beta' \leqslant \alpha + \beta. \tag{15.4.22}$$

Последнее неравенство просто означает, что сумма *действительных* вероятностей ошибок 1-го и 2-го рода при выборе k_0 и k_1 согласно (15.4.19) не превосходит суммы *предполагаемых* вероятностей ошибок 1-го и 2-го рода. Таким образом, доказана теорема

15.4.5. *Если постоянные k_0 и k_1 последовательного критерия отношения вероятностей предполагаемой силы $(\alpha, \theta_0; \beta, \theta_1)$ выбраны соответственно равными $\beta/(1-\alpha)$ и $(1-\beta)/\alpha$, то действительная сила полученного критерия есть $(\alpha', \theta_0; \beta', \theta_1)$, где α' и β' удовлетворяют неравенствам* (15.4.21) *и* (15.4.22).

В практических задачах предполагаемые значения α и β редко превосходят 0,10. Неравенства (15.4.21) и (15.4.22) показывают, что разность между предполагаемой и действительной силой в этих случаях незначительна.

(d) Оперативная характеристика последовательного критерия отношения вероятностей. До сих пор мы рассматривали функцию оперативной характеристики $L(\theta)$ только для двух значений θ: θ_0 и θ_1. Точное определение $L(\theta)$ для произвольного значения $\theta \in \Omega$ является трудной задачей. Однако приближенное выражение для $L(\theta)$ можно получить без чрезмерных усилий.

Вначале докажем лемму, принадлежащую Вальду (1945).

15.4.6. *Пусть* $z = \log f(x;\ \theta_1) - \log f(x;\ \theta_0)$, *где* x — *случайная величина с плотностью распределения* $f(x;\ \theta)$, $\theta \in \Omega$. *Если*

(i) $\mathscr{E}e^{zh}$ *существует при всех вещественных* h *и его можно дифференцировать по* h *дважды,*

(ii) $\mathscr{E}z \neq 0$,

(iii) *при некоторых* $\delta_1 > 0$ *и* $0 < \delta_2 < 1$

$$P(e^z > 1 + \delta_1) > 0 \quad \text{и} \quad P(e^z < 1 - \delta_2) > 0,$$

то каждому θ *отвечает значение* $h = h(\theta) \neq 0$ *с условием*

$$\int_{-\infty}^{+\infty} \left(\frac{f(x;\ \theta_1)}{f(x;\ \theta)}\right)^h f(x;\ \theta)\,dx = 1, \tag{15.4.23}$$

т. е.

$$\mathscr{E}e^{zh(\theta)} = 1.$$

Для доказательства **15.4.6** обозначим через E_1 множество тех x, для которых $e^z > 1 + \delta_1$, и через E_2 — множество x, для которых $e^z < 1 - \delta_2$. Тогда при $h > 0$ имеем

$$\mathscr{E}e^{zh} \geqslant \int_{E_1} e^{zh} f(x;\ \theta)\,dx > (1 + \delta_1)^h P(E_1), \tag{15.4.24}$$

а при $h < 0$

$$\mathscr{E}e^{zh} \geqslant \int_{E_2} e^{zh} f(x;\ \theta)\,dx > (1 - \delta_2)^h P(E_2). \tag{15.4.25}$$

Так как $P(E_1)$ и $P(E_2)$ отличны от нуля, то из (15.4.24) и (15.4.25) следует, что

$$\lim_{h \to +\infty} \mathscr{E}e^{zh} = \lim_{h \to -\infty} \mathscr{E}e^{zh} = 0. \tag{15.4.26}$$

Если положить $\mathscr{E}e^{zh} = \psi(h)$, то

$$\psi'(0) = \mathscr{E}z \neq 0 \tag{15.4.27}$$

и

$$\psi''(h) = \mathscr{E}(z^2 e^{zh}) > 0. \tag{15.4.28}$$

Поскольку $\psi(0) = 1$, $\psi'(0) \neq 0$ и $\psi''(0) > 0$, то ясно, что функция $\psi(h)$, зависящая также и от θ, имеет минимум, меньший 1, и уравнение $\psi(h) - 1 = 0$ имеет два корня: $h = 0$ и $h = h(\theta) \neq 0$, что и доказывает лемму **15.4.6.**

Перейдем теперь к построению приближенного выражения для $L(\theta)$. Событие $\mathscr{C}_n^\circ$ определялось первым из неравенств (15.4.1), которое эквивалентно неравенству

$$Q_{*n} \leqslant k_0^h Q_n, \tag{15.4.29}$$

где

$$Q_{*n} = \left(\frac{Q_{1n}}{Q_{0n}}\right)^h Q_n \quad \text{и} \quad Q_n = \prod_{i=1}^{n} f(x_i;\ \theta), \tag{15.4.30}$$

а h — произвольное вещественное число. Q_{*n} можно записать в виде

$$Q_{*n}=\prod_{t=1}^{n}\left[\left(\frac{f(x_t;\ \theta_1)}{f(x_t;\ \theta_0)}\right)^h f(x_t;\ \theta)\right]. \qquad (15.4.31)$$

В условиях леммы **15.4.6** для каждого θ имеется единственное значение $h=h(\theta)\neq 0$ такое, что

$$\int\left(\frac{f(x;\ \theta_1)}{f(x;\ \theta_0)}\right)^{h(\theta)} f(x;\ \theta)\,dx=1. \qquad (15.4.32)$$

Следовательно, функцию

$$f^*(x;\ \theta)=\left(\frac{f(x;\ \theta_1)}{f(x;\ \theta_0)}\right)^{h(\theta)} f(x;\ \theta) \qquad (15.4.33)$$

можно рассматривать как плотность распределения вероятностей.

Пусть $P^*(G_n^\circ\mid\theta)$ — вероятность события G_n°, вычисленная по плотности $f^*(x;\ \theta)$. Определим $L^*(\theta)$ по аналогии с $L(\theta)$ из (15.2.8):

$$L^*(\theta)=\sum_{n=1}^{\infty}P^*(G_n^\circ\mid\theta)=P^*(G^\circ\mid\theta).$$

Рассуждения, аналогичные использованным при выводе (15.4.16), дают

$$L^*(\theta)\leqslant k_0^{h(\theta)}L(\theta). \qquad (15.4.34)$$

Событие G_n' определялось вторым из неравенств (15.4.1), которое эквивалентно такому:

$$Q_{*n}\geqslant k_1^h Q_n. \qquad (15.4.35)$$

Вполне аналогично (15.4.34) получим

$$1-L^*(\theta)\geqslant k_1^{h(\theta)}(1-L(\theta)). \qquad (15.4.36)$$

Если теперь по аналогии с заменой неравенств (15.4.18) равенствами (15.4.19) заменить (15.4.34) и (15.4.36) на равенства, то мы будем иметь следующее приближенное выражение для $L(\theta)$:

$$L(\theta)\approx\frac{1-k_1^{h(\theta)}}{k_0^{h(\theta)}-k_1^{h(\theta)}}. \qquad (15.4.37)$$

Отметим, что поскольку $h(\theta_0)=1$ и $h(\theta_1)=-1$, то подстановка в (15.4.37) приближенных значений k_0 и k_1 из (15.4.19) приводит к точным значениям $L(\theta_0)$ и $L(\theta_1)$, которые определяются формулой (15.4.14) для последовательного критерия отношения вероятностей силы $(\alpha,\ \theta_0;\ \beta,\ \theta_1)$.

Приближение (15.4.37) удовлетворительно для практических целей. Исследование того, насколько оно точно для значений θ, отличных от θ_0 и θ_1, требует утомительного анализа, который мы опускаем. Подробности можно найти в книге Вальда (1947а).

Сформулируем полученный результат.

15.4.7. *В условиях леммы* **15.4.6** *приближение к оперативной характеристике* $L(\theta)$ *последовательного критерия отношения вероятностей силы* $(\alpha,\ \theta_0;\ \beta,\ \theta_1)$ *дается формулой* (15.4.37), *где* $h(\theta)$ *удовлетворяет условию* (15.4.32), *а приближенные значения* k_0 *и* k_1 *даны в* (15.4.19).

(e) Средний объем выборки для последовательного критерия отношения вероятностей. Чтобы получить приближенное выражение для $\mathcal{E}(n \mid \theta)$, можно рассуждать следующим образом. Если $\mathcal{E}(z \mid \theta) \neq 0$, то из (15.4.11) находим, что

$$\mathcal{E}(n \mid \theta) = \frac{\mathcal{E}(z_1 + z_2 + \ldots + z_n \mid \theta)}{\mathcal{E}(z \mid \theta)}. \tag{15.4.38}$$

Как и раньше, пусть $G^\circ = G_1^\circ \cup G_2^\circ \cup \ldots$ и $G' = G_1' \cup G_2' \cup \ldots$ Тогда $P(G^\circ \mid \theta) = L(\theta)$, $P(G' \mid \theta) = 1 - L(\theta)$ и можно написать

$$\begin{aligned} \mathcal{E}(z_1 + z_2 + \ldots + z_n \mid \theta) = \mathcal{E}(z_1 + z_2 + \ldots + z_n \mid G^\circ;\ \theta) L(\theta) + \\ + \mathcal{E}(z_1 + z_2 + \ldots + z_n \mid G';\ \theta)(1 - L(\theta)). \end{aligned} \tag{15.4.39}$$

Но $\mathcal{E}(z_1 + z_2 + \ldots + z_n \mid G^\circ;\ \theta) \leqslant \log k_0$ и $\mathcal{E}(z_1 + z_2 + \ldots + z_n \mid G';\ \theta) \geqslant \log k_1$. Если заменить в этих неравенствах k_0 и k_1 приближенными значениями из (15.4.19), а сами неравенства — равенствами, то (15.4.38) дает приближенное выражение для $\mathcal{E}(n \mid \theta)$:

$$\mathcal{E}(n \mid \theta) \approx \frac{L(\theta) \log \frac{\beta}{1-\alpha} + (1 - L(\theta)) \log \frac{1-\beta}{\alpha}}{\mathcal{E}(z \mid \theta)}. \tag{15.4.40}$$

При $\theta = \theta_0$ и $\theta = \theta_1$ приближенные значения $\mathcal{E}(n \mid \theta)$ равны

$$\mathcal{E}(n \mid \theta_0) \approx \frac{(1-\alpha) \log \frac{\beta}{1-\alpha} + \alpha \log \frac{1-\beta}{\alpha}}{\mathcal{E}(z \mid \theta_0)}, \tag{15.4.41}$$

$$\mathcal{E}(n \mid \theta_1) \approx \frac{\beta \log \frac{\beta}{1-\alpha} + (1-\beta) \log \frac{1-\beta}{\alpha}}{\mathcal{E}(z \mid \theta_1)}.$$

Для практических целей эти приближения удовлетворительны. Оценка погрешности приближенной формулы (15.4.40) требует серьезного анализа, который можно найти у Вальда (1947а).

(f) Эффективность последовательного критерия отношения вероятностей. Следующая доказанная Вальдом (1945) теорема дает нижнюю границу $\mathcal{E}(n \mid \theta)$ в точках θ_0 и θ_1 для любого последовательного критерия.

15.4.8. *Пусть* $(x_1, x_2, \ldots)$ — *последовательность независимых случайных величин с плотностью распределения* $f(x;\ \theta)$, *для которой* $\mathcal{E}(z \mid \theta) \neq 0$, *где* z *определена в* (15.4.4). *Если* S — *любой после-*

довательный критерий силы $(\alpha, \theta_0; \beta, \theta_1)$, *оканчивающийся с вероятностью* 1 *за конечное число шагов, то*

$$\begin{aligned} \mathscr{E}(n \mid \theta_0) &\geqslant \frac{(1-\alpha)\log\frac{\beta}{1-\alpha}+\alpha\log\frac{1-\beta}{\alpha}}{\mathscr{E}(z \mid \theta_0)}, \\ \mathscr{E}(n \mid \theta_1) &\geqslant \frac{\beta\log\frac{\beta}{1-\alpha}+(1-\beta)\log\frac{1-\beta}{\alpha}}{\mathscr{E}(z \mid \theta_1)}. \end{aligned} \tag{15.4.42}$$

Обозначим через G_n°, G_n', G_n множества в выборочном пространстве $(x_1, \ldots, x_n)$, $n = 1, 2, \ldots$, удовлетворяющие условию (15.2.1) и определяющие произвольный последовательный тест S. [Если S — последовательный критерий отношения вероятностей, то множества G_n°, G_n', G_n определяются условием (15.2.1) и тремя неравенствами (15.4.1).]

Пусть $z_1, z_2, \ldots$ определены согласно (15.4.5). Из (15.4.11) для последовательного критерия S получим

$$\mathscr{E}(n \mid \theta) = \frac{\mathscr{E}(z_1+z_2+\ldots+z_n \mid \theta)}{\mathscr{E}(z \mid \theta)}. \tag{15.4.43}$$

Но

$$\begin{aligned} \mathscr{E}(z_1+z_2+\ldots+z_n \mid \theta) = \mathscr{E}(z_1+z_2+\ldots+z_n \mid G^\circ;\ \theta) P(G^\circ \mid \theta) + \\ + \mathscr{E}(z_1+z_2+\ldots+z_n \mid G';\ \theta) P(G' \mid \theta). \end{aligned} \tag{15.4.44}$$

Рассмотрим сначала $\mathscr{E}(n \mid \theta_0)$. Поскольку S имеет силу $(\alpha, \theta_0; \beta, \theta_1)$, то

$$P(G^\circ \mid \theta_0) = 1-\alpha, \qquad P(G' \mid \theta_0) = \alpha. \tag{15.4.45}$$

Используя неравенство

$$\mathscr{E}y \leqslant \log \mathscr{E}e^y, \tag{15.4.46}$$

справедливое для любой случайной величины y, находим, что

$$\begin{aligned} \mathscr{E}(z_1+z_2+\ldots+z_n \mid G^\circ;\ \theta_0) \leqslant \log \mathscr{E}(e^{z_1+z_2+\ldots+z_n} \mid G^\circ;\ \theta_0) = \\ = \log \mathscr{E}\left(\frac{Q_{1n}}{Q_{0n}} \mid G^\circ;\ \theta_0\right), \end{aligned} \tag{15.4.47}$$

где $\mathscr{E}\left(\frac{Q_{1n}}{Q_{0n}} \mid G^\circ;\ \theta_0\right)$ — условное среднее значение, на множестве $G^\circ = G_1^\circ \cup G_2^\circ \cup \ldots$ при $\theta = \theta_0$, произведения $\frac{f(x_1;\ \theta_1)}{f(x_1;\ \theta_0)} \frac{f(x_2;\ \theta_1)}{f(x_2;\ \theta_0)} \cdots$, в котором число множителей определяется числом наблюдений до окончания последовательного процесса. Имеем

$$\mathscr{E}\left(\frac{Q_{1n}}{Q_{0n}} \mid G^\circ;\ \theta_0\right) = P(G^\circ \mid \theta_1)/P(G^\circ \mid \theta_0) = \frac{\beta}{1-\alpha}. \tag{15.4.48}$$

Следовательно,

$$\mathscr{E}(z_1+z_2+\ldots+z_n \mid G^\circ;\ \theta_0) \leqslant \log \frac{\beta}{1-\alpha}, \tag{15.4.49}$$

и аналогично

$$\mathscr{E}(z_1+z_2+\ldots+z_n \mid G';\ \theta_0) \leqslant \log \frac{1-\beta}{\alpha}. \tag{15.4.50}$$

Объединяя (15.4.45), (15.4.49) и (15.4.50) с (15.4.44), получим первое из неравенств (15.4.42). Таким же образом доказывается и второе неравенство.

Дальнейшее изучение нижних границ для $\mathscr{E}(n \mid \theta)$ проводилось Гофдингом (1960) и Кифером и Уэйссом (1957). Андерсон (1960) и Донелли (1957) предложили для случая выборки из нормальной совокупности с известной дисперсией и неизвестным средним некоторые модификации критерия отношения вероятностей с целью уменьшить $\mathscr{E}(n \mid \theta)$.

Для последовательного критерия отношения вероятностей, когда множества G_n°, G_n', G_n определяются условием (15.2.1) и неравенствами (15.4.1), мы видели, что

$$\begin{gathered}\mathscr{E}(z_1+z_2+\ldots+z_n \mid G^\circ;\ \theta) \leqslant \log k_0,\\ \mathscr{E}(z_1+z_2+\ldots+z_n \mid G';\ \theta) \geqslant \log k_1.\end{gathered} \tag{15.4.51}$$

Если после наблюдения над x_n при попадании точки $(x_1, \ldots, x_n)$ в G_n° (т. е. при $z_1+\ldots+z_n \leqslant \log k_0$) мы будем приписывать $z_1+\ldots+z_n$ значение $\log k_0$, а при попадании точки в G_n' (т. е. при $z_1+\ldots+z_n \geqslant \log k_1$) — значение $\log k_1$, то для определенной таким образом суммы $z_1+\ldots+z_n$ в (15.4.49) и (15.4.50) имеют место знаки равенства. В этом случае неравенства (15.4.42) также перейдут в равенства, которые, как это видно из (15.4.41), действительно реализуются на последовательном критерии отношения вероятностей. Следовательно, если бы можно было «исправлять» $z_1+\ldots+z_n$ по указанному выше правилу, то не существовало бы критерия, более эффективного, чем последовательный критерий отношения вероятностей. Эти эвристические соображения об оптимальном характере последовательного критерия отношения вероятностей принадлежат Вальду (1945). Строгое доказательство оптимальности было дано Вальдом и Вольфовитцем (1948).

(g) Усечение последовательного критерия отношения вероятностей. Предположим, что если последовательный процесс для критерия отношения вероятностей не закончился при $n=1, \ldots, N-1$, то принимается следующее правило для окончания процесса после наблюдения над x_N.

Принять $\mathscr{H}_0$, если

$$\log k_0 < z_1+\ldots+z_N \leqslant 0. \tag{15.4.52}$$

Принять $\mathscr{H}_1$, если

$$0 < z_1+\ldots+z_N < \log k_1. \tag{15.4.53}$$

Обозначим этот *усеченный последовательный критерий* через S_N, а через α_N и β_N — связанные с S_N вероятности ошибок 1-го и 2-го рода. Займемся определением верхних границ для α_N и β_N. Пусть $G_{(N)}^\circ$ и $G_{(N)}'$ — области в R_N, в которых тест S_N принимает соответственно $\mathscr{H}_0$ и $\mathscr{H}_1$. Множества $G_{(N)}^\circ$ и $G_{(N)}'$ дизъюнктны и их объединение дает R_N. Условимся рассматривать $G^{\circ}_{(N)}$ и $G'_{(N)}$ как цилиндрические множества в R_∞.

Имеем
$$P(G'_{(N)} \mid \theta_0) = \alpha_N. \tag{15.4.54}$$

Как и раньше, пусть G° и G' обозначают области в R_∞, в которых (неусеченный) последовательный критерий отношения вероятностей S принимает соответственно $\mathscr{H}_0$ и $\mathscr{H}_1$. Напомним, что
$$P(G' \mid \theta_0) = \alpha. \tag{15.4.55}$$

Пусть G'^* — область в R_∞, в которой гипотеза $\mathscr{H}_1$ принимается усеченным критерием и отклоняется неусеченным. Тогда
$$G'_{(N)} \subset G' \cup G'^*. \tag{15.4.56}$$

Если J — множество точек в R_∞, для которых выполнено (15.4.53), то
$$G'^* \subset J$$
и, следовательно,
$$G'_{(N)} \subset G' \cup J. \tag{15.4.57}$$

Выберем теперь N достаточно большим, так чтобы можно было считать сумму $z_1 + \ldots + z_N$ распределенной приближенно нормально. Тогда
$$P(0 < z_1 + \ldots + z_N < \log k_1 \mid \theta_0) = \\ = \Phi(y_0') - \Phi(y_0) + O\left(\frac{1}{\sqrt{N}}\right), \tag{15.4.58}$$
где
$$\begin{aligned} y_0 &= -\sqrt{N}\,\mathscr{E}(z \mid \theta_0)/\sigma(z \mid \theta_0), \\ y_0' &= \sqrt{N}\left[\frac{1}{N}\log k_1 - \mathscr{E}(z \mid \theta_0)\right]\Big/\sigma(z \mid \theta_0). \end{aligned} \tag{15.4.59}$$

$\Phi(y)$ — к. ф. р. закона $N(0, 1)$, а $\mathscr{E}(z \mid \theta_0)$ и $\sigma^2(z \mid \theta_0)$ — среднее значение и дисперсия случайной величины
$$z = \log \frac{f(x;\ \theta_1)}{f(x;\ \theta_0)} \qquad \text{при } \theta = \theta_0.$$

Таким образом, с точностью до членов порядка $1/\sqrt{N}\,\alpha + \Phi(y_0') - \Phi(y_0)$ оказывается верхней границей для α_N. Аналогично находим, что с точностью до членов порядка $1/\sqrt{N}\,\beta + \Phi(y_1') - \Phi(y_1)$ есть верхняя граница для β_N, где
$$\begin{aligned} y_1 &= \sqrt{N}\left[\frac{1}{N}\log k_0 - \mathscr{E}(z \mid \theta_1)\right]\Big/\sigma(z \mid \theta_1), \\ y_1' &= -\sqrt{N}\,\mathscr{E}(z \mid \theta_1)/\sigma(z \mid \theta_1), \end{aligned} \tag{15.4.59a}$$
а $\mathscr{E}(z \mid \theta_1)$ и $\sigma^2(z \mid \theta_1)$ — среднее значение и дисперсия z при $\theta = \theta_1$.

Как мы видели в §15.4 (c), при малых значениях α и β неизвестные k_0 (в выражении для y_1) и k_1 (в выражении для y_0') хорошо приближаются числами $\beta/(1-\alpha)$ и $(1-\beta)/\alpha$.

Суммируя все сказанное, получаем следующий результат Вальда (1945).

15.4.9. *Если случайная величина*

$$z = \log \frac{f(x;\ \theta_1)}{f(x;\ \theta_0)}$$

имеет конечные среднее значение и дисперсию при $\theta = \theta_0$ *и* $\theta = \theta_1$, *то верхняя граница для вероятности* α_N *ошибки 1-го рода усеченного последовательного критерия отношения вероятностей* S_N *равна*

$$\alpha + \Phi(y_0') - \Phi(y_0) + O\left(\frac{1}{\sqrt{N}}\right),$$

где α — *вероятность ошибки 1-го рода для неусеченного последовательного критерия отношения вероятностей* S, *а* y_0 *и* y_0' *даны в* (15.4.59). *Верхняя граница для вероятности* β_N *ошибки 2-го рода есть*

$$\beta + \Phi(y_1') - \Phi(y_1) + O\left(\frac{1}{\sqrt{N}}\right),$$

где β — *вероятность ошибки 2-го рода у* S, *а* y_1 *и* y_1' *определены в* (15.4.59a).

15.5. Применение последовательного критерия отношения вероятностей к биномиальному распределению

Проиллюстрируем результаты § 15.4 на примере выборки из биномиального распределения Bi $(1,\ \theta)$.

В этом случае $(x_1,\ x_2, \ldots)$ — последовательность независимых случайных величин, функция вероятности которых равна

$$f(x;\ \theta) = \theta^x (1-\theta)^{1-x}, \qquad x = 0,\ 1. \tag{15.5.1}$$

Обозначив $\sum_{t=1}^{n} x_t$ через n_1, будем иметь

$$\frac{Q_{1n}}{Q_{0n}} = \left(\frac{\theta_1}{\theta_0}\right)^{n_1} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{n-n_1}. \tag{15.5.2}$$

Определим постоянные a_n и b_n равенствами

$$a_n = \frac{\log \frac{\beta}{1-\alpha} - n \log \frac{1-\theta_1}{1-\theta_0}}{\log \frac{\theta_1 (1-\theta_0)}{\theta_0 (1-\theta_1)}}, \qquad b_n = \frac{\log \frac{1-\beta}{\alpha} - n \log \frac{1-\theta_1}{1-\theta_0}}{\log \frac{\theta_1 (1-\theta_0)}{\theta_0 (1-\theta_1)}}. \tag{15.5.3}$$

Если пользоваться приближенными значениями (15.4.19) для k_0 и k_1, то (15.4.1) перейдут соответственно в неравенства

$$n_1 \leqslant a_n, \quad n_1 \geqslant b_n, \quad a_n < n_1 < b_n. \tag{15.5.4}$$

Так что последовательный процесс продолжается, пока $a_n < n_1 < b_n$; он заканчивается принятием $\mathscr{H}_0$ после наблюдения над x_n, если $n_1 \leqslant a_n$, и заканчивается принятием $\mathscr{H}_1$, если $n_1 \geqslant b_n$.

Подставляя приближенные значения (15.4.19) для k_0 и k_1 в (15.4.37), получим следующее приближенное выражение для оперативной характеристики последовательного критерия:

$$L(\theta) \approx \frac{1 - \left(\frac{1-\beta}{\alpha}\right)^{h(\theta)}}{\left(\frac{\beta}{1-\alpha}\right)^{h(\theta)} - \left(\frac{1-\beta}{\alpha}\right)^{h(\theta)}}, \tag{15.5.5}$$

где функция $h(\theta)$ определяется условием (15.4.32), которое в рассматриваемом случае переходит в

$$\sum_{x=0}^{1} \left[\frac{\theta_1^x (1-\theta_1)^{1-x}}{\theta_0^x (1-\theta_0)^{1-x}}\right]^h \theta^x (1-\theta)^{1-x} = 1. \tag{15.5.6}$$

Упрощая (15.5.6), получим соотношение между θ и h:

$$\theta = \frac{1 - \left(\frac{1-\theta_1}{1-\theta_0}\right)^h}{\left(\frac{\theta_1}{\theta_0}\right)^h - \left(\frac{1-\theta_1}{1-\theta_0}\right)^h}. \tag{15.5.7}$$

Заметим, что $h(\theta_0) = 1$, $h(\theta_1) = -1$, и из (15.5.5) мы видим, что приближенная формула для $L(\theta)$ дает точные значения $L(\theta_0)$ и $L(\theta_1)$.

Обращаясь к (15.4.40) и вычисляя $\mathscr{E}(z \mid \theta)$, получим приближенное значение для среднего объема выборки

$$\mathscr{E}(n \mid \theta) \approx \frac{L(\theta) \log \frac{\beta}{1-\alpha} + (1 - L(\theta)) \log \frac{1-\beta}{\alpha}}{\theta \log \frac{\theta_1}{\theta_0} + (1-\theta) \log \frac{1-\theta_1}{1-\theta_0}}. \tag{15.5.8}$$

Если подставить в (15.5.8) приближенное выражение для $L(\theta)$ из (15.5.5) и выражение (15.5.7) для θ, мы получим выражение для $\mathscr{E}(n \mid \theta)$ в виде функции параметра h. Для $\theta = 0$, θ_0, θ_1 $\mathscr{E}(n \mid \theta)$ вычисляется легко.

Применения вальдовского последовательного критерия отношения вероятностей к выборкам из различных специальных распределений опубликованы Вальдом (1947а) и Статистической исследовательской группой Колумбийского университета (1945, 1947).

15.6. Последовательное оценивание

(а) Общие замечания. Общая теория последовательного оценивания параметров еще не создана. Вальд (1947а) сформулировал общую задачу последовательного интервального оценивания, но не дал ее решения. Он и Стейн (1947) рассматривали последовательную процедуру построения доверительного интервала фиксированной длины с заданным коэффициентом доверия для средного нормального распределения с известной дисперсией. Рассматривались также другие специфические задачи оценивания. Обзор некоторых результатов в этой области имеется у Энскомба (1953).

Основная идея последовательного оценивания параметров состоит в том, чтобы производить наблюдения до тех пор, пока мы не окажемся в состоянии получить оценку с заданной степенью точности (в подходящем смысле), не зависящей от (неизвестного) значения оцениваемого параметра. Степень точности может быть определена различными способами. Можно, например, потребовать, чтобы дисперсия оценки параметра не зависела от самого параметра. Иногда этому требованию можно удовлетворить, используя выборку фиксированного объема. Для больших выборок приближенное решение этой задачи при некоторых условиях регулярности изложено в §12.3 (е).

Другое определение степени точности заключается в требовании фиксированной (не зависящей от результатов выборки) длины доверительного интервала для параметра. В некоторых случаях этого также можно достичь в выборках фиксированного объема. Например, если $\bar{x}$ — среднее значение выборки объема n из нормальной совокупности $N(\mu,\sigma^2)$ с неизвестным μ и известным σ^2, то $\left(\bar{x} - \frac{1}{2}\delta,\ \bar{x} + \frac{1}{2}\delta\right)$ оказывается доверительным интервалом длины δ, накрывающим неизвестное значение μ с вероятностью, не меньшей γ, при условии, что $(n-1)$ — наибольшее целое число, не превосходящее $y_\gamma^2\sigma^2/\delta^2$ (если последнее само не является целым; если же $y_\gamma^2\sigma^2/\delta^2$ — целое, то нужно выбрать $n = y_\gamma^2\sigma^2/\delta^2$), y_γ — корень уравнения $\Phi(y_\gamma) - \Phi(-y_\gamma) = \gamma$, а $\Phi(y)$ — к. ф. р. закона $N(0, 1)$.

(b) Оценка среднего нормального распределения интервалами фиксированной длины по Стейну. Для случая, когда неизвестны μ и σ^2, Стейн (1945) показал, как использовать процедуру двойной выборки для построения доверительного интервала фиксированной длины δ, накрывающего неизвестное значение μ с вероятностью, не меньшей γ. Его результат можно сформулировать следующим образом.

15.6.1. *Пусть* $(x_1, \ldots, x_n)$ — *первая и* $(x_{n+1}, \ldots, x_{n+m})$ — *не зависящая от нее вторая выборки из совокупности* $N(\mu, \sigma^2)$. *Обозначим через* s^2 *выборочную дисперсию первой выборки, через* $\bar{x}$ — *среднее, построенное по обеим выборкам, и положим* $k = [2st_{n-1,\gamma}/\delta^2]$, *где* $t_{n-1,\gamma}$ — *верхняя* $100\left(1 - \frac{1}{2}\gamma\right)$-*процентная точка распределения*

Стьюдента $S(n-1)$. *Если* m *равно* 0 *при* $k-n\leqslant 0$ *и равно наименьшему целому, превосходящему* $k-n$, *при* $k-n>0$, *то* $\left(\bar{x}-\frac{1}{2}\delta,\ \bar{x}+\frac{1}{2}\delta\right)$ *будет доверительным интервалом длины* δ *для* μ *с коэффициентом доверия, не меньшим* γ.

Для доказательства этого результата заметим, что, поскольку $(\bar{x}-\mu)/[\sigma/\sqrt{m+n}]$ имеет распределение $N(0, 1)$, а $(n-1)s^2/\sigma^2$ имеет χ^2-распределение $C(n-1)$ и эти две величины независимы, то по доказанному в §7.8(b)

$$\frac{(\bar{x}-\mu)\sqrt{m+n}}{s}$$

имеет распределение Стьюдента $S(n-1)$. Следовательно,

$$P\left(-t_{n-1,\gamma}<\frac{\bar{x}-\mu}{s}\sqrt{m+n}<t_{n-1,\gamma}\right)=\gamma \tag{15.6.1}$$

или, в эквивалентной форме,

$$P\left(\bar{x}-\frac{st_{n-1,\gamma}}{\sqrt{m+n}}<\mu<\bar{x}+\frac{st_{n-1,\gamma}}{\sqrt{m+n}}\right)=\gamma. \tag{15.6.2}$$

Если $st_{n-1,\gamma}/\sqrt{n}\leqslant\frac{1}{2}\delta$, то $\left(\bar{x}-\frac{1}{2}\delta,\ \bar{x}+\frac{1}{2}\delta\right)$ оказывается доверительным интервалом длины δ для μ с коэффициентом доверия, не меньшим γ, построенным только по первой выборке. Если же $st_{n-1,\gamma}/\sqrt{n}>\frac{1}{2}\delta$, то извлекаем вторую выборку объема m, где m — наименьшее целое, для которого

$$\frac{st_{n-1,\gamma}}{\sqrt{m+n}}\leqslant\frac{1}{2}\delta,$$

т. е. наименьшее целое, превосходящее

$$\left[\frac{2st_{n-1,\gamma}}{\delta}\right]^2-n.$$

В этом случае $\left(\bar{x}-\frac{1}{2}\delta,\ \bar{x}+\frac{1}{2}\delta\right)$ также будет доверительным интервалом для μ длины δ, построенным уже по результатам обеих выборок, имеющим коэффициент доверия, не меньший γ.

ЗАДАЧИ

15.1. Пусть $(x_1, x_2, \ldots)$ — последовательность независимых случайных величин со средними значениями $\mu_1, \mu_2, \ldots$ и $T_t=\mu_1+\ldots+\mu_t$, $t=1, 2, \ldots$ Показать, что если $G_1, G_2, \ldots, G_1^*, G_2^*, \ldots$ определены как в § 15.2, то

$$\mathscr{E}(x_1+x_2+\ldots+x_n)=\sum_{t=0}^{\infty}(T_tp(t)),$$

где

$$p(t) = P(G_t^+).$$

15.2. Пусть $(x_1, x_2, \ldots)$ — последовательность независимых случайных величин с функциями распределения $F_1(x; \theta)$, $F_2(x; \theta)$, ..., а $E^{(i)}$ — такие множества, что

$$\int_{E^{(i)}} dF_i(x; \theta) = p_i(\theta) \leqslant p < 1.$$

Доказать, что если S — произвольный последовательный процесс, для которого продолжающими эксперимент множествами служат G_1, G_2, ...,

$$G_n = E^{(1)} \times E^{(2)} \times \ldots \times E^{(n)},$$

то он оканчивается с вероятностью 1 за конечное число шагов.

15.3. Пусть $(x_1, x_2, \ldots)$ — последовательность независимых случайных величин с плотностью распределения $\theta e^{-\theta x}$, $x > 0$, $\theta > 0$. Надо построить декартов последовательный процесс для проверки гипотезы $\mathscr{H}_0$: $\theta = \theta_0$ против $\mathscr{H}_1$: $\theta = \theta_1$, где $\theta_1 > \theta_0$ и вероятности ошибок 1-го и 2-го рода равны соответственно α и β. Показать, что для декартова критерия отношения вероятностей множества G_0°, G_0' и G_0 будут интервалами (x_2, ∞), $(0, x_1)$ и (x_1, x_2), где x_1, x_2, $x_1 < x_2$, определяются условиями:

$$\frac{e^{-\theta_0 x_2}}{1 - e^{-\theta_0 x_1}} = \frac{1-\alpha}{\alpha}, \qquad \frac{e^{-\theta_1, x_2}}{1 - e^{-\theta_1 x_1}} = \frac{\beta}{1-\beta}.$$

Доказать, что средний объем выборки $\mathscr{E}(n \mid \theta)$ для этого последовательного процесса достигает наибольшего значения при $\theta = \dfrac{\log x_2 - \log x_1}{x_2 - x_1}$.

15.4. (*Продолжение*). Предположим, что для проверки $\mathscr{H}_0$ против $\mathscr{H}_1$ используется r-кратный декартов последовательный процесс. Показать, что множества G_0°, G_0' и G_0 (в r-мерном эвклидовом пространстве), отвечающие критерию отношения вероятностей, имеют вид

$$G_0^\circ = \left\{(x_1, \ldots, x_r) : \sum_1^r x_i \geqslant y_2\right\},$$

$$G_0' = \left\{(x_1, \ldots, x_r) : \sum_1^r x_i \leqslant y_1\right\},$$

$$G_0 = \left\{(x_1, \ldots, x_r) : y_1 < \sum_1^r x_i < y_2\right\},$$

где (y_1, y_2) определяется уравнениями

$$\int_{y_2}^{\infty} f(y; \theta_0)\, dy = \frac{1-\alpha}{\alpha} \int_0^{y_1} f(y; \theta_0)\, dy,$$

$$\int_{y_2}^{\infty} f(y; \theta_1)\, dy = \frac{\beta}{1-\beta} \int_0^{y_1} f(y; \theta_1)\, dy,$$

$$f(y; \theta) = \frac{\theta^r y^{r-1} e^{-\theta y}}{\Gamma(r)}, \quad y > 0.$$

15.5. Рассмотрим следующее усечение декартова последовательного критерия, обсуждавшегося в §15.3 (а). Если последовательный процесс не закончился после n-го шага, то независимо от результата наблюдения будем принимать $\mathcal{H}_0$ с вероятностью $a/(a+b)$ и $\mathcal{H}_1$ с вероятностью $b/(a+b)$. Показать, что тогда

$$L_n(\theta)=\frac{a}{a+b},$$
$$M_n(\theta)=\frac{b}{a+b},$$
$$\mathcal{E}(n\mid\theta)=\frac{1-c^n}{a+b}$$

и вероятности α и β ошибок 1-го и 2-го рода удовлетворяют условию (15.3.8).

15.6. Доказать, что декартов последовательный критерий, для которого G_0° и G_0' определены в (15.3.10) и (15.3.11), сильнее любого другого декартова последовательного критерия с теми же значениями $a_0/(a_0+b_0)=\alpha$ и $\mathcal{E}(n\mid\theta_0)=n_0$.

15.7. Пусть $(x_1, x_2, \ldots)$ — последовательность независимых случайных величин, имеющих нормальное распределение $N(\theta, 1)$; $\mathcal{H}_0$ — гипотеза $\theta=\theta_0$; $\mathcal{H}_1$ — гипотеза $\theta=\theta_1$, $\theta_1>\theta_0$. Используя, при заданных вероятностях α и β ошибок 1-го и 2-го рода приближенные значения (15.4.19) для k_0 и k_1, показать, что последовательный критерий отношения вероятностей для проверки $\mathcal{H}_0$ против $\mathcal{H}_1$ принимает гипотезу $\mathcal{H}_0$ после n-го наблюдения, если впервые на n-ом шаге

$$x_1+\ldots+x_n\leqslant\frac{1}{\theta_1-\theta_0}\log\frac{\beta}{1-\alpha}+n\frac{\theta_1+\theta_2}{2},$$

принимает гипотезу $\mathcal{H}_1$, если впервые на n-м шаге

$$x_1+\ldots+x_n\geqslant\frac{1}{\theta_1-\theta_0}\log\frac{1-\beta}{\alpha}+n\frac{\theta_1+\theta_2}{2},$$

и предписывает провести $(n+1)$-е наблюдение, если $x_1+\ldots+x_n$ лежит между указанными пределами.

Показать также, что

$$L(\theta)\approx\frac{\left(\frac{1-\beta}{\alpha}\right)^h-1}{\left(\frac{1-\beta}{\alpha}\right)^h-\left(\frac{\beta}{1-\alpha}\right)^h},$$

где

$$h=\frac{\theta_1+\theta_0-2\theta}{\theta_1-\theta_0},$$

и что это приближение дает точные значения при $\theta=\theta_0$ и $\theta=\theta_1$. Доказать, что средний объем выборки равен

$$\mathcal{E}(n\mid\theta)\approx-2\,\frac{L(\theta)\log\frac{\beta}{1-\alpha}+(1-L(\theta))\log\frac{1-\beta}{\alpha}}{h(\theta_1-\theta_0)^2}.$$

Перенести эти результаты на случай выборки из совокупности $N(\theta, \sigma^2)$, где σ^2 известно.

15.8. Пусть $(x_1, x_2, \ldots)$ — последовательность независимых случайных величин, имеющих нормальное распределение $N(0, \sigma^2)$; $\mathscr{H}_0$ — гипотеза $\sigma^2=\sigma_0^2$, $\mathscr{H}_1$ — гипотеза $\sigma^2=\sigma_1^2$, $\sigma_1^2>\sigma_0^2$. Используя, при заданных вероятностях α и β ошибок 1-го и 2-го рода, приближенные значения (15.4.19) для k_0 и k_1, показать, что последовательный критерий отношения вероятностей для проверки $\mathscr{H}_0$ против $\mathscr{H}_1$ принимает $\mathscr{H}_0$ после n-го наблюдения, если впервые на n-м шаге

$$x_1^2+\ldots+x_n^2\leqslant\frac{\sigma_0^2\sigma_1^2\left[2\log\frac{\beta}{1-\alpha}+n\log\frac{\sigma_1^2}{\sigma_0^2}\right]}{\sigma_1^2-\sigma_0^2},$$

принимает гипотезу $\mathscr{H}_1$, если впервые на n-м шаге

$$x_1^2+\ldots+x_n^2\geqslant\frac{\sigma_0^2\sigma_1^2\left[2\log\frac{1-\beta}{\alpha}+n\log\frac{\sigma_1^2}{\sigma_0^2}\right]}{\sigma_1^2-\sigma_0^2},$$

и предписывает провести $(n+1)$-е наблюдение, если $x_1^2+\ldots+x_n^2$ лежит между указанными пределами.

15.9. (*Продолжение*). Показать, что приближенное выражение для оперативной характеристики $L(\sigma^2)$ имеет вид

$$L(\sigma^2)\approx\frac{\left(\frac{1-\beta}{\alpha}\right)^h-1}{\left(\frac{1-\beta}{\alpha}\right)^h-\left(\frac{\beta}{1-\alpha}\right)^h},$$

где σ^2 и h связаны соотношением

$$\sigma^2=\frac{\sigma_0^2\sigma_1^2\left[1-\left(\frac{\sigma_0}{\sigma_1}\right)^{2h}\right]}{h(\sigma_1^2-\sigma_0^2)}.$$

Проверить, что это приближение дает точные значения при $\sigma^2=\sigma_0^2$ и $\sigma^2=\sigma_1^2$.

Доказать, что средний объем выборки равен

$$\mathscr{E}(n\mid\sigma^2)\approx\frac{2\sigma_0^2\sigma_1^2\left[L(\sigma^2)\log\frac{\beta}{1-\alpha}+(1-L(\sigma^2))\log\frac{1-\beta}{\alpha}\right]}{(\sigma_1^2-\sigma_0^2)\,\sigma^2+\sigma_0^2\sigma_1^2\log\frac{\sigma_0^2}{\sigma_1^2}}.$$

15.10. Построить последовательный критерий отношения вероятностей для проверки гипотезы $\mathscr{H}_0$, согласно которой совокупность имеет пуассоновское распределение $\mathrm{Po}(\theta_0)$, против альтернативы $\mathscr{H}_1$, согласно которой совокупность имеет распределение Пуассона $\mathrm{Po}(\theta_1)$, $\theta_1>\theta_0$. Найти приближения для функций оперативной характеристики $L(\theta)$ и среднего объема выборки $\mathscr{E}(n\mid\theta)$ этого критерия.

15.11. Положим для биномиального распределения времени ожидания

$$\binom{n-1}{k-1}p^k(1-p)^{n-k},\qquad n=k,\ k+1,\ \ldots,$$

$\tilde{p}=\dfrac{k-1}{n-1}$. Доказать, что $\tilde{p}$ — несмещенная оценка p, и при этом

$$\sigma^2(\tilde{p})=\frac{p^2}{k-2}\left[1-\frac{k-1}{k-3}p+\frac{2(k-1)}{(k-3)(k-4)}p^2-\ldots\right].$$

Следовательно, коэффициент изменчивости $\tilde{p}$ равен

$$\frac{\sigma(\tilde{p})}{\mathscr{E}(\tilde{p})}=\frac{1}{\sqrt{k-2}}(1+O(p))$$

[см. работы Холдэйна (1945) и Гудмена (1953) об аналогичных задачах].

15.12. Пусть s^2 — дисперсия выборки объема n из совокупности $N(\mu, \sigma^2)$. Положим для произвольного $\varepsilon>0$ m равным наименьшему целому с условием $m \geqslant (n-1)s^2/[(n-3)\varepsilon]$. Показать, что, если $\bar{x}'$ — среднее значение независимой (от данной) выборки объема m из совокупности $N(\mu, \sigma^2)$, то $\sigma^2(\bar{x}') \leqslant \varepsilon$.

15.13. Пусть s_1^2 и s_2^2 — дисперсии независимых выборок объемов n_1 и n_2 из совокупностей $N(\mu_1, \sigma_1^2)$ и $N(\mu_2^2, \sigma_2^2)$. Для произвольного $\varepsilon>0$ пусть m_i — наименьшее целое с условием $m_i \geqslant [2(n_i-1)s_i^2]/[(n_i-3)\varepsilon]$, $i=1, 2$. Доказать, что если $\bar{x}_1$ и $\bar{x}_2$ — средние значения независимых (от исходных) выборок объемов m_1 и m_2 из совокупностей $N(\mu_1, \sigma_1^2)$ и $N(\mu_2, \sigma_2^2)$ соответственно, то $\sigma^2(\bar{x}_1-\bar{x}_2) \leqslant \varepsilon$.

ГЛАВА 16

СТАТИСТИЧЕСКИЕ РЕШАЮЩИЕ ФУНКЦИИ

16.1. Общие замечания

В построенной Нейманом и Пирсоном теории проверки статистических гипотез основную роль играет понятие потери (убытка) от неверного принятия или неверного отклонения гипотезы. В двух основополагающих статьях и последовавшей за ними книге Вальд (1947b, 1949a, 1950) распространил подход, основанный на понятии риска, на более широкий класс статистических задач и разработал теорию *статистических решающих функций*. Создание этой теории было частично стимулировано теорией игр, изложенной фон Нейманом и Моргенштерном (1944), частично же — желанием построить более общую теорию риска для различных статистических процедур. После оригинальной работы Вальда появилось много статей и несколько книг, в том числе весьма полная книга Блекуэла и Гиршика (1954), посвященных теории статистических решений для дискретных выборочных пространств. Эта глава служит кратким введением в некоторые из основных идей и результатов теории статистических решений для ситуации простейшего типа. Мы не касаемся теории в наиболее общей форме. Читатели, интересующиеся общей теорией и дальнейшими подробностями, могут обратиться к книгам Вальда и Блекуэла и Гиршика. Более элементарный характер носят книги Чернова и Мозеса (1959), Льюса и Райфы (1957), Райфы и Шлайфера (1961), Сэвиджа (1954), Уэйсса (1961).

16.2. Определения и терминология

Пусть x — дискретная случайная величина, выборочным пространством которой является R, θ — точка параметрического пространства Ω, содержащего конечное число точек $\theta_1, \ldots, \theta_h$. Обозначим через $p(x \mid \theta)$ функцию вероятности x при данном θ. Всего имеется h распределений $p(x \mid \theta_1), \ldots, p(x \mid \theta_h)$, одно из которых предполагается *истинным* всякий раз, когда производится наблюдение над x.

Пусть a — произвольный элемент множества A возможных решений, которые можно вынести относительно θ по результатам наблюдений над x. Например, если $\theta_1, \ldots, \theta_h$ — числа, а c — некоторое критическое значение, то множество A может состоять из двух элементов, a_1 и a_2, где решение a_1 означает, что $\theta \leqslant c$, а решение a_2 означает $\theta > c$. Множество A может содержать любое конечное или бесконечное число элементов. A называется *пространством решений*. Если решение $a \in A$ выносится на основании результата наблюдения над x, то это означает, что имеется *решающая функция* $d(x)$, отображающая R в A. Все рассматриваемые решающие функции являются элементами *класса решающих функций* D.

При каждом $\theta \in \Omega$ можно вынести любое из решений $a \in A$. Предположим, что принятие решения a, когда истинное значение параметра равно θ, приводит к *потере* $L(\theta, a)$. *Функция потерь* $L(\theta, a)$ предполагается вещественной ограниченной функцией, определенной в пространстве $\Omega \times A$. Для всякой решающей функции $d \in D$ и каждой точки $\theta \in \Omega$ определяется *функция риска* $r(\theta, d)$ как среднее значение функции потерь $L(\theta, d(x))$, т. е.

$$r(\theta, d) = EL(\theta, d(x)) = \sum_{x \in R} L(\theta, d(x))\, p(x \mid \theta). \tag{16.2.1}$$

Таким образом, $r(\theta, d)$ задана в каждой точке пространства $\Omega \times D$ и ограничена вследствие ограниченности $L(\theta, d(x))$.

Замечание. Некоторые из введенных понятий допускают интерпретацию в терминах элементарной теории игр. Представим себе природу в виде игрока I, а статистика — в виде игрока II. Природа выбирает стратегию среди $\theta_1, \ldots, \theta_h$, статистик—среди $d \in D$. Величина $L(\theta, d(x))$ измеряет потери статистика, выбравшего по результату наблюдения x стратегию $d(x)$, когда природа выбрала θ. Определенная в (16.2.1) величина $r(\theta, d)$ характеризует средние потери статистика, использующего стратегию d.

Функция риска дает разумный критерий, позволяющий предпочесть одну решающую функцию из D другой. Действительно, пусть d и d^* — две решающие функции из D. Функция d^* *предпочтительнее* d, если

$$r(\theta, d^*) \leqslant r(\theta, d) \tag{16.2.2}$$

при всех $\theta \in \Omega$ и

$$r(\theta, d^*) < r(\theta, d) \tag{16.2.3}$$

хотя бы при одном $\theta \in \Omega$. Если имеют место (16.2.2) и (16.2.3), то d^* называется *равномерно лучшей решающей функцией*, чем d. Если не существует функции в D, равномерно лучшей, чем d^*, то d^* называется *допустимой решающей функцией*. Класс D' решающих функций называется *полным классом*, если для любой $d \overline{\in} D'$ найдется элемент $d^* \in D'$, равномерно лучший, чем d.

Если полный класс решающих функций не содержит никакого (собственного) полного подкласса, то он называется *минимальным полным классом*. Понятие полного класса решающих функций — важное в общей теории решений. Отношение между полным и минимальным полным классами решающих функций описывается теоремами **16.2.1.** и **16.2.2.**

16.2.1. *Если минимальный полный класс существует, то он совпадает с классом всех допустимых решающих функций.*

Доказательство. Обозначим через D_0 класс допустимых решающих функций. Тогда, если D' — минимальный полный класс, то $D_0 \subset D'$. Пусть теперь $d' \in D'$ и $d' \overline{\in} D_0$. Тогда найдется элемент d'', равномерно лучший, чем d'. Но d'' не может принадлежать D', так как D' — минимальный полный класс. Поэтому должен существовать элемент $d''' \in D'$, равномерно лучший, чем d'', и тем самым равномерно лучший, чем d'. Но это невозможно, так как D' — минимальный полный класс. Следовательно, D_0 не может быть собственным подмножеством D' и $D_0 = D'$.

Если класс D_0 всех допустимых решающих функций полный, то он образует минимальный полный класс. Отсюда и из теоремы **16.2.1** следует

16.2.2. *Необходимое и достаточное условие существования минимального полного класса решающих функций состоит в том, что класс допустимых решающих функций полный.*

При довольно общих условиях Вальд (1950) доказал, что класс допустимых решающих функций полный и, в частности, класс всех байесовских решений (рассматриваемых ниже) полный.

16.3. Минимаксный подход к проблеме решения

Вернемся к функции риска $r(\theta, d)$. Для любой решающей функции d определим *вектор риска* $(r(\theta_1, d), \ldots, r(\theta_h, d))$. Рассмотрим максимальную компоненту этого вектора; она, естественно, зависит от d. Пусть $d^* \in D$ — решающая функция, минимизирующая максимальную компоненту; другими словами, d^* такова, что

$$\sup_{\theta \in \Omega} r(\theta, d^*) \leqslant \sup_{\theta \in \Omega} r(\theta, d) \qquad (16.3.1)$$

для всех $d \in D$. Тогда d^* или связанный с ней вектор $(r(\theta_1, d^*), \ldots, r(\theta_h, d^*))$ называют *минимаксным решением* нашей задачи.

Если D содержит только конечное число элементов, то, очевидно, всегда существует по меньшей мере одно минимаксное решение d^*; для него

$$\min_d \{\max_{\theta \in \Omega} r(\theta, d)\} = \max_{\theta \in \Omega} r(\theta, d^*). \qquad (16.3.2)$$

Такое (или такие) d^*, доставляющее *минимаксный риск*, можно найти непосредственным перебором значений $r(\theta, d)$ на конечном множестве точек $\Omega \times D$ в порядке, указанном в (16.3.2).

Если же D содержит бесконечное число элементов, то ситуация становится более сложной. В этом случае полезно геометрическое представление. Для каждой решающей функции $d \in D$ набор h чисел $(r(\theta_1, d), \ldots, r(\theta_h, d))$ можно представлять себе точкой r в эвклидовом пространстве R_h. Множество таких точек, отвечающих всем $d \in D$, обозначим E; оно ограничено, так как функция $r(\theta, d)$ ограничена на $\Omega \times D$. Задача нахождения минимаксного решения эквивалентна выбору максимальной координаты каждой точки r и последующей минимизации этой максимальной координаты. Если минимаксное решение d^* существует, то его можно получить следующим образом. Пусть E_i — подмножество E, на котором $r(\theta_i;\ d) \geqslant \max_{j \neq i} r(\theta_j,\ d)$, $i = 1, \ldots, h$. Тогда $E = E_1 \cup E_2 \cup \ldots \cup E_h$, хотя множества $E_1, \ldots, E_h$ не обязательно дизъюнктны, поскольку точка, у которой имеются две одинаковые максимальные координаты, принадлежит нескольким из множеств $E_1, \ldots, E_h$. Обозначим через M_i точную верхнюю грань i-х координат точек из E_i, $i = 1, \ldots, h$. Если существует $d^* \in D$ с условием

$$\max_i r(\theta_i,\ d^*) = \min(M_1, \ldots, M_h), \tag{16.3.3}$$

то d^* будет минимаксным решением. Если E замкнуто, т. е. содержит пределы всех сходящихся последовательностей точек из E, то по меньшей мере один такой элемент d^* существует.

Покажем, что при слабых условиях множество E замкнуто. Заметим, что для любого элемента из $D \times R$ точка $L = (L(\theta_1,\ d(x)), \ldots, L(\theta_h,\ d(x)))$ принадлежит эвклидову пространству R_h. Обозначим множество всех таких точек через F, тогда F ограничено, так как ограничена $L(\theta,\ d(x))$. Докажем теорему

16.3.1. *Если F замкнуто, то E также замкнуто.*

Занумеруем точки R в последовательность x_t, $t = 1, 2, \ldots$ Для каждой $d \in D$

$$r(\theta_i,\ d) = \sum_{t=1}^{\infty} L(\theta_i,\ d(x_t))\ p(x_t \mid \theta_i). \tag{16.3.4}$$

Всякая решающая функция $d(x)$ порождает последовательность точек

$$(L(\theta_1,\ d(x_t)), \ldots,\ L(\theta_h,\ d(x_t))), \qquad t = 1, 2, \ldots, \tag{16.3.5}$$

в R_h. Объединение таких последовательностей, отвечающих всевозможным $d \in D$, образует ограниченное множество F, которое мы предположили замкнутым.

Рассмотрим произвольную сходящуюся последовательность точек r_α, $\alpha = 1, 2, \ldots$, из E, где

$$r_\alpha = (r(\theta_1,\ d_\alpha), \ldots,\ r(\theta_h,\ d_\alpha)). \tag{16.3.6}$$

Обозначим предел этой последовательности через

$$r^* = (r(\theta_1,\ d^*), \ldots,\ r(\theta_h,\ d^*)). \tag{16.3.7}$$

Покажем, что r^* также принадлежит E. В множестве $F\,d_\alpha$ соответствует последовательность точек

$$(L(\theta_1,\ d_\alpha(x_t)), \ldots,\ L(\theta_h,\ d_\alpha(x_t))), \qquad t = 1, 2, \ldots \tag{16.3.8}$$

Если $\alpha = 1, 2, \ldots$, то (16.3.8) образует двойную последовательность точек $L_{\alpha t}$ в F. Применяя к этой двойной последовательности канторовский диагональный процесс, получим подпоследовательность $d_{\alpha\beta}$,

$\beta=1, 2, \ldots$, последовательности d_α, $\alpha=1, 2, \ldots$, такую, что при каждом t последовательность точек

$$(L(\theta_1, d_{\alpha\beta}(x_t)), \ldots, L(\theta_h, d_{\alpha\beta}(x_t))) \tag{16.3.9}$$

из множества F сходится к

$$(L(\theta_1, d^*(x_t)), \ldots, L(\theta_h, d^*(x_t))) \tag{16.3.10}$$

при $\beta\to\infty$. Так как F замкнуто, то все точки (16.3.10) при $t=1, 2, \ldots$, принадлежат ему. Рассмотрим теперь функцию риска

$$r(\theta, d_{\alpha\beta}) = \sum_{t=1}^{\infty} L(\theta, d_{\alpha\beta}(x_t))\, p(x_t \mid \theta). \tag{16.3.11}$$

Переходя здесь к пределу при $\beta\to\infty$ и пользуясь сходимостью последовательности (16.3.9) к точке (16.3.10), для каждого t будем иметь

$$\lim_{\beta\to\infty} r(\theta, d_{\alpha\beta}) = \lim_{\beta\to\infty} \sum_{t=1}^{\infty} L(\theta, d_{\alpha\beta}(x_t))\, p(x_t \mid \theta) = \sum_{t=1}^{\infty} \lim_{\beta\to\infty} L(\theta, d_{\alpha\beta}(x_t)) \times$$

$$\times\, p(x_t \mid \theta) = \sum_{t=1}^{\infty} L(\theta, d^*(x_t))\, p(x_t \mid \theta). \tag{16.3.12}$$

Перестановка операций суммирования и предельного перехода законна, так как функция риска ограничена, а $p(x_t \mid \theta)$ — распределение вероятностей. Но поскольку точки (16.3.10) принадлежат F, то $r(\theta, d^*) = \sum_{t=1}^{\infty} L(\theta, d^*(x_t))\, p(x_t \mid \theta)$ лежит в E. Следовательно, E замкнуто.

Таким образом, если множество $F \subset R_h$, состоящее из точек $(L(\theta_1, d(x)), \ldots, L(\theta_h, d(x)))$, $d \in D$, $x \in R$, замкнуто, то множество $E \subset R_h$, образованное точками $(r(\theta_1, d), \ldots, r(\theta_h, d))$, $d \in D$, также замкнуто и существует минимаксное решение $d^* \in D$. Множество всех таких решений d^* не пусто и замкнуто.

Отыскание минимаксных решений в конкретных задачах обычно связано с громоздкими вычислениями. Мы выбрали для иллюстрации следующий очень простой пример.

Пример. Пусть x — дискретная случайная величина, $R=\{0, 1\}$,

$$p(x \mid \theta_i) = \theta_i^x (1-\theta_i)^{1-x}, \quad \theta_1 = \frac{1}{4}, \quad \theta_2 = \frac{1}{2}.$$

Пространство решений A предполагаем состоящим из двух элементов, a_1 и a_2. Функция потерь $L(\theta_i, a_j)$ определяется таблицей

	a_1	a_2
θ_1	1	4
θ_2	3	2

Имеется всего четыре решающие функции d_α, $\alpha=1, 2, 3, 4$:

$$\begin{aligned} &d_1(0)=a_1, \quad && d_1(1)=a_1,\\ &d_2(0)=a_1, && d_2(1)=a_2,\\ &d_3(0)=a_2, && d_3(1)=a_1,\\ &d_4(0)=a_2, && d_4(1)=a_2. \end{aligned}$$

По формуле (16.2.1) находятся четыре вектора риска $(r(\theta_1, d_\alpha), r(\theta_2, d_\alpha))$:

$$r(\theta_i, d_\alpha)=\sum_{x=0}^{1} L(\theta_i, d_\alpha(x))\,p(x\mid\theta_i), \quad i=1, 2;\ \alpha=1, 2, 3, 4.$$

В нашем примере они равны (1, 3), (7/4, 5/2), (13/4, 5/2), (4/2). Наибольшие координаты у каждого вектора равны соответственно 3, 5/2, 13/4, 4 и минимальная из них равна 5/2.

Минимаксной решающей функцией будет $d_2(x)$, а соответствующий ей вектор риска есть (7/4, 5/2).

16.4. Байесовский подход к статистической проблеме решения

(а) Решение при заданном априорном распределении. Предположим теперь, что значения параметра θ появляются в соответствии с некоторой априорной функцией вероятности $q(\theta)$, $\theta=\theta_1, \ldots, \theta_h$. Тогда для заданной решающей функции $d\in D$ усредненная по априорному распределению $q(\theta)$ функция $r(\theta, d)$ даст *средний риск* $\bar r(q, d)$:

$$\bar r(q, d)=\sum_{i=1}^{h} r(\theta_i, d)\,q(\theta_i)=$$
$$=\sum_{t=1}^{\infty}\sum_{i=1}^{h} L(\theta_i, d(x_t))\,p(x_t\mid\theta_i)\,q(\theta_i). \qquad (16.4.1)$$

Если бы мы знали $q(\theta)$ и могли найти такую функцию $d^*\in D$, что

$$\bar r(q, d^*)\leqslant \bar r(q, d) \qquad (16.4.2)$$

для всех $d\in D$, то d^* мы могли бы считать оптимальным решением задачи. Решающая функция, минимизирующая $\bar r(q, d)$, называется *байесовским решением*, отвечающим априорному распределению $q(\theta)$.

Отметим, что поскольку $q(\theta_i)\geqslant 0$, $i=1, \ldots, h$, и $q(\theta_1)+\ldots+q(\theta_h)=1$, то набор $(q(\theta_1), \ldots, q(\theta_h))$ можно представлять точкой $(h-1)$-мерного симплекса G в эвклидовом пространстве R_h, порожденного точками $(1, 0, \ldots, 0)$, $(0, 1, \ldots, 0)$, ..., $(0, 0, \ldots, 1)$.

При каких же условиях существует байесовское решение? Представим $\bar r(q, d)$ в виде

$$\bar r(q, d)=\sum_{t=1}^{\infty}\sum_{i=1}^{h} L(\theta_i, d(x_t))\,Q(\theta_i\mid x_t)\,P_q(x_t), \qquad (16.4.3)$$

где

$$Q(\theta_i \mid x_t) = \frac{p(x_t \mid \theta_i)\, q(\theta_i)}{P_q(x_t)} \tag{16.4.4}$$

и

$$P_q(x_t) = \sum_{i=1}^{h} p(x_t \mid \theta_i)\, q(\theta_i), \tag{16.4.5}$$

$Q(\theta_i|x_t)$ — апостериорная вероятность того, что $\theta = \theta_i$ при условии $x = x_t$ и заданном априорном распределении $q(\theta)$.

Пусть для данных $q(\theta)$ и t величина

$$\sum_{i=1}^{h} L(\theta_i,\ d^*(x_t))\, Q(\theta_i \mid x_t) \tag{16.4.6}$$

представляет собой точную нижнюю грань чисел

$$\sum_{i=1}^{h} L(\theta_i,\ d(x_t))\, Q(\theta_i \mid x_t) \tag{16.4.7}$$

по всем $(L(\theta_1,\ d(x_t)),\ \ldots,\ L(\theta_h,\ d(x_t))) \in F$. При фиксированном t точка $(Q(\theta_1 \mid x_t),\ \ldots,\ Q(\theta_h \mid x_t))$ принадлежит G. Таким образом, последовательность точек $x_1, x_2, \ldots$ выборочного пространства R определяет последовательность векторов в F и соответствующую последовательность в G. Пополним *) множество F векторами

$$(L(\theta_1,\ d(x_t))\, Q(\theta_1 \mid x_t),\ \ldots,\ L(\theta_h,\ d(x_t))\, Q(\theta_h \mid x_t)), \tag{16.4.8}$$

$t = 1, 2, \ldots, d \in D$, и минимизирующим вектором

$$(L(\theta,\ d^*(x_t))\, Q(\theta_1 \mid x_t),\ \ldots,\ L(\theta_h,\ d^*(x_t))\, Q(\theta_h,\ x_t)), \tag{16.4.9}$$

$t = 1, 2, \ldots$ Тогда E и, следовательно, D будут замкнуты и (16.4.6) будет минимизировать (16.4.7) при фиксированном t по всем точкам из F. Другими словами, D будет содержать тогда по крайней мере один элемент d^*, для которого

$$\sum_{t=1}^{\infty} \sum_{i=1}^{h} L(\theta_i,\ d^*(x_t))\, Q(\theta_i \mid x_t)\, P_q(x_t) \leqslant r(q,\ d) \tag{16.4.10}$$

при всех $d \in D$. Каждый такой элемент $d^* \in D$ дает байесовское решение задачи, отвечающее априорному распределению $q(\theta)$. Множество таких d^* непусто и замкнуто.

Таким образом, доказана теорема.

16.4.1. *Если при заданном априорном распределении* $q(\theta)$ *множество* F *пополнено векторами* (16.4.8) *для всех* $t = 1, 2, \ldots$ *и всех* $d \in D$ *и векторами* (16.4.9) *для всех* $t = 1, 2, \ldots$, *то* E

*) Это пополнение означает расширение пространства решений D. (*Прим. перев.*)

замкнуто и решающая функция d^*, *отвечающая последовательности векторов* $(L(\theta_1, d^*(x_t)), \ldots, L(\theta_h, d^*(x_t)))$, *определенной в* (16.4.6), *минимизирует средний риск* $\bar r(q, d)$ *из* (16.4.1).

(b) Геометрическая интерпретация. Обозначим левую часть (16.4.10) через $\bar r(q, d^*)$; тогда

$$\bar r(q, d^*) = \min_{d \in D} \bar r(q, d) = \min_{r \in E} \sum_{i=1}^{h} r(\theta_i, d)\, q(\theta_i). \tag{16.4.11}$$

Следовательно, если взять взвешенные средние координат каждой точки из E с весами $q(\theta_1), \ldots, q(\theta_h)$, то $\bar r(q, d^*)$ будет наименьшим среди них. Геометрически: если для заданного $q(\theta)$ построить семейство гиперплоскостей (C — параметр семейства)

$$\sum_{i=1}^{h} r_i q(\theta_i) = C \tag{16.4.12}$$

в R_h, то точка $r = (r_1, \ldots, r_h)$ множества E, отвечающая наименьшему (среди точек E) значению C, дает байесовское решение относительно $q(\theta)$, а соответствующее значение C равно минимуму среднего риска $\bar r(q, d^*)$. Все координаты r конечны, так как E ограничено.

(c) Множество решений, отвечающих всевозможным априорным распределениям. Опишем множество решений, отвечающих всем априорным распределениям. Набор

$$(q(\theta_1), \ldots, q(\theta_h)) \tag{16.4.13}$$

представляется точкой симплекса $G \subset R_h$, порожденного векторами $(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$. Координаты точки (16.4.13) являются направляющими косинусами нормали к гиперплоскости (16.4.12). Следовательно, множеству точек $r \in E$, доставляющих минимальные значения C при всевозможных априорных распределениях $q(\theta)$, отвечает множество гиперплоскостей, образующих нижнюю границу *оболочки* множества E. Очевидно, что множество E^* точек из E, участвующих в построении границы оболочки множества E, само лежит на границе. Оно представляет собой множество всех байесовских решений задачи, отвечающих всевозможным априорным распределениям $q(\theta)$.

Пример. В конце § 16.3 мы привели простой числовой пример, иллюстрирующий процедуру нахождения минимаксного решения. Вернемся к этому примеру и предположим, что θ имеет априорное распределение $q(\theta)$, где $q(\theta_1) = \frac{1}{3}$, $q(\theta_2) = \frac{2}{3}$. Вспоминая, что решающим функциям d_1, d_2, d_3, d_4 отвечают векторы риска (1, 3), (7/4, 5/2), (13/4, 5/2), (4, 2), мы можем вычислить $\sum_i r(\theta_i, d)\, q(\theta_i)$ как скалярное произведение каждого из векторов риска и вектора $(q(\theta_1), q(\theta_2)) = (1/3, 2/3)$. Эти скалярные произведения равны 28/12, 27/12, 33/12, 32/12. Они представляют собой средние риски решений d_1, d_2, d_3, d_4 при данном $q(\theta)$. Решающей функцией, минимизирующей средний риск при выбранном $q(\theta)$, оказывается d_2. Если построить описанную выше оболочку, то мы увидим, что она содержит только первые два вектора риска. Это означает, что только d_1 или d_2 оказываются байесовскими решениями при любых априорных распределениях.

16.5. Замечания о возможных обобщениях

Мы рассмотрели статистическую теорию решений только для простейшего случая, когда случайная величина x дискретна, параметрическое пространство Ω содержит конечное число точек и над x производится одно наблюдение. (Заметим, что x могла бы быть n-мерной случайной величиной. Например, под x можно было бы понимать выборку объема n из пуассоновской совокупности с параметром θ, принимающим конечное число значений.) Были построены минимаксное и байесовское решения в этом сравнительно простом случае.

Полученные результаты можно обобщить в различных направлениях. Например, можно перенести эти результаты на случай, когда распределение x абсолютно непрерывно, а Ω по-прежнему состоит из конечного числа точек. Обобщение в другом направлении связано с отказом от конечности пространства Ω, которое можно предполагать счетным или даже считать его подмножеством эвклидова пространства. Еще одно направление обобщений — введение последовательной выборки.

Читателям, интересующимся различными обобщениями, следует обратиться к книгам Вальда (1950) и Блекуэла и Гиршика (1954) и к статьям Дворецкого, Кифера и Вольфовитца (1953a, 1953b), Карлина и Рубина (1956), Лемана (1957).

ЗАДАЧИ

16.1. Пусть x — случайная величина, имеющая биномиальное распределение

$$p(x\mid\theta)=\binom{3}{x}\theta^x(1-\theta)^{3-x},\quad x=0,\ 1,\ 2,\ 3,$$

а параметрическое пространство состоит из двух точек, $\theta_1=1/100$, $\theta_2=1/10$. Предположим, что пространство решений A содержит два элемента, a_1 и a_2, а функция потерь задана таблицей

	a_1	a_2
θ_1	0	2
θ_2	1	0

Пусть пространство D решающих функций состоит из трех точек, $d_1(x)$, $d_2(x)$, $d_3(x)$, где

$$d_\alpha(x)=\begin{cases}a_1, & x=0,\ 1,\ \ldots,\ \alpha-1,\\ a_2 & \text{в остальных случаях,}\end{cases}$$

$\alpha = 1, 2, 3$. Определить решающую функцию, доставляющую минимаксное решение задачи.

16.2. (*Продолжение*). Найти решающую функцию, которая дает байесовское решение задачи для заданного априорного распределения.

16.3. Пусть x — случайная величина с функцией вероятности

$$p(x \mid \theta) = (1-\theta)\,\theta^{x-1}, \qquad x = 1, 2, \ldots,$$

параметрическое множество Ω состоит из двух точек, $\theta_1 = 1/10$, $\theta_2 = 2/10$. Предположим, что пространство решений A содержит два элемента, a_1 и a_2, и функция потерь определена следующим образом:

	a_1	a_2
θ_1	0	1
θ_2	2	0

Пространство D содержит решающие функции

$$d_\alpha(x) = \begin{cases} a_1, & x = 1, \ldots, \alpha, \\ a_2, & x = \alpha+1, \alpha+2, \ldots, \end{cases}$$

$\alpha = 1, 2, 3, 4, 5$. Какая решающая функция дает минимаксное решение?

16.4. (*Продолжение*). Определить решающую функцию, дающую байесовское решение задачи для заданного априорного распределения.

16.5. Пусть x — случайная величина с одной из двух функций вероятности $p(x \mid \theta_1)$ или $p(x \mid \theta_2)$, пространство решений $A = \{a_1, a_2\}$, функция потерь определена таблицей

	a_1	a_2
θ_1	0	c
θ_2	1	0

Доказать, что все решающие функции, являющиеся байесовскими решениями для априорного распределения $q(\theta)$, имеют такой вид:

$d(x) = a_1$ для всех x, удовлетворяющих неравенству $\dfrac{p(x \mid \theta_2)}{p(x \mid \theta_1)} \leqslant c\,\dfrac{q(\theta_1)}{q(\theta_2)}$,

$d(x) = a_2$ для всех остальных x.

ГЛАВА 17

ВРЕМЕННЫЕ РЯДЫ

17.1. Вводные замечания

Многие задачи науки и техники связаны с процессами, которые можно представить как семейство случайных величин (случайный процесс) на некотором интервале (времени) T. Значение процесса x_t в каждый момент t является случайной величиной. Такие процессы называются *временными рядами*. Примерами временных рядов служат: напряжение в цепи в течение некоторого промежутка времени, уровень фабричного шума, высота волны в море.

Временные ряды возникают также при наложении случайных флуктуаций на систематический тренд. Примеры таких временных рядов — температура в городе в течение суток, цены на сырье в различные месяцы, национальное производство в различные годы.

Разработаны различные статистические и математические методы анализа временных рядов. Многие из этих методов созданы для оценки тренда, возникающего во временных рядах при «усреднении», в более или менее эмпирическом смысле, случайных флуктуаций.

В последние годы значительное внимание уделялось изучению временных рядов как случайных процессов. Такой подход часто позволяет глубже (по сравнению с эмпирическим) проникнуть внутрь механизма, порождающего процесс.

Имеется обширная литература, включающая несколько книг, в которой временные ряды рассматриваются с точки зрения случайных процессов. Здесь мы даем только краткое введение в предмет и приводим небольшое число сравнительно простых основных результатов. Дальнейшие результаты можно найти в различных книгах, особенно у Бартлетта (1955), Дуба (1953), Гренандера и Розенблатта (1957), Волда (1938).

17.2. Стационарные временные ряды

Вспомним § 4.1, где случайный процесс определяется как такое семейство случайных величин $\{x_t;\ t \in T\}$, что для каждого конечного набора $t_1, \ldots, t_n$ значений $t \in T$ имеется совместная функция распределения случайных величин $x_{t_1}, \ldots, x_{t_n}$.

Будем представлять себе переменную t временем, а T считать подмножеством вещественной оси. Так, в некоторых случаях T может состоять из точек арифметической прогрессии. В этих случаях можно выбрать единицу времени таким образом, чтобы $T = \ldots, -1,\ 0,\ +1, \ldots$ *Выборки* образованы значениями x_t в моменты $1, \ldots, n$ или $n-h,\ n-h+1, \ldots,\ n-1$.

Произвольный случайный процесс $\{x_t;\ t \in T\}$ оказывается слишком общим объектом, чтобы его можно было плодотворно изучать. Мы ограничимся так называемыми *стационарными процессами* или *стационарными временными рядами*. Имеется два важных типа таких процессов.

Стационарный в узком смысле процесс определяется тем свойством, что случайная величина

$$(x_{t_1}, \ldots, x_{t_n}) \tag{17.2.1}$$

при любых $t_1, \ldots, t_n$ и произвольном h распределена одинаково с величиной

$$(x_{t_1+h}, \ldots, x_{t_n+h}). \tag{17.2.2}$$

(Заметим, что можно считать x_t комплексной или k-мерной случайной величиной, хотя мы будем почти всегда рассматривать одномерную вещественную x_t.) Следовательно, распределение величины (17.2.1) зависит только от разностей

$$t_2 - t_1, \quad t_3 - t_2, \quad \ldots, \quad t_n - t_{n-1}. \tag{17.2.3}$$

Читатель легко докажет утверждение

17.2.1. *Если* $\{x_t;\ t \in T\}$ — *стационарный в узком смысле процесс и* $\mathscr{E}(|x_t|) < \infty$, *то* $\mathscr{E}(x_t) = \mathrm{const}$ *при всех* t.

Приведем пример стационарного в узком смысле процесса.

Пример. Рассмотрим вещественный процесс

$$x_t = \sum_{p=1}^{r} a_p \cos(t\omega_p + u_p),\ t = \ldots,\ -1,\ 0,\ +1, \ldots, \tag{17.2.4}$$

где $a_1, \ldots, a_r$ — вещественные постоянные, $\omega_1, \ldots, \omega_r$ — числа из интервала $[-\pi,\ +\pi]$, которые можно считать упорядоченными: $\omega_1 < \ldots < \omega_r$, $u_1, \ldots, u_r$ — независимые случайные величины, каждая из которых равномерно распределена на $[-\pi,\ +\pi]$. Пусть $t_1, \ldots, t_n$ — произвольный набор значений t. Тогда

$$x_{t_\xi + h} = \sum_{p=1}^{r} a_p \cos(t_\xi\omega_p + v_p), \qquad \xi = 1, \ldots, n, \tag{17.2.5}$$

где v_p — независимые величины, равномерно распределенные на интервалах $[h\omega_p - \pi,\ h\omega_p + \pi]$, $p = 1, \ldots, r$. Очевидно, что совместное распределение величин (x_{t_1}, x_{t_n}) такое же, как и $(x_{t_1+h}, \ldots, x_{t_n+h})$.

Следовательно, процесс (17.2.4.) стационарен в узком смысле.

Отметим, что комплексным вариантом процесса (17.2.4) является также стационарный в узком смысле процесс

$$x_t = \sum_{p=1}^{r} a_p e^{\mathrm{i}(t\omega_p + u_p)},\ t = \ldots,\ -1,\ 0,\ +1, \ldots \tag{17.2.6}$$

Некоторые из наиболее полезных результатов о временны́х рядах верны без предположения о стационарности процесса в узком смысле при следующих более слабых условиях: (i) $\mathscr{E} x_t$ постоянно при всех t, в таком случае можно считать $\mathscr{E} x_t = 0$; (ii) матрица ковариаций величин (17.2.2) не зависит от h. Процессы $\{x_t;\ t \in T\}$, удовлетворяющие этим двум условиям, называются *стационарными в широком смысле.* Это означает, что матрица ковариаций зависит только от разностей (17.2.3) и, следовательно, ковариация x_{t+h} и x_t (считаем $\mathscr{E} x_t = 0$) оказывается функцией только h:

$$\mathscr{E}(x_{t+h}\, x_t) = \gamma_h. \tag{17.2.7}$$

Рассматриваемая как функция h, γ_h называется *функцией ковариаций* процесса $\{x_t;\ t \in T\}$; иногда ее называют *ковариацией с запаздыванием аргумента* или *автоковариацией* с запаздыванием. Коэффициент корреляции между x_{t+h} и x_t равен $\rho_h = \gamma_h/\gamma_0$. Рассматриваемый как функция h, ρ_h называется *сериальным коэффициентом корреляции процесса.*

Для комплексных x_t функция ковариаций γ_h определяется как

$$\gamma_h = \mathscr{E}(x_{t+h}\, \bar{x}_t), \tag{17.2.8}$$

где $\bar{x}_t$ — величина, комплексно сопряженная с x_t.

Легко устанавливается следующий результат.

17.2.2. *Стационарный в узком смысле процесс с конечной матрицей ковариаций стационарен также в широком смысле.*

Моделью некоторых временных рядов $\{y_t;\ t \in T\}$ служат процессы вида

$$y_t = m_t + x_t, \tag{17.2.9}$$

где m_t постоянна при всех t, а процесс $\{x_t;\ t \in T\}$ стационарный с $\mathscr{E} x_t = 0$. Таким образом,

$$\mathscr{E} y_t = m_t \tag{17.2.10}$$

и функция ковариаций процесса $\{y_t;\ t \in T\}$ одинакова с функцией ковариаций процесса $\{x_t;\ t \in T\}$, т. е. (в вещественном случае)

$$\gamma_h = \mathscr{E}[(y_{t+h} - m_{t+h})(y_t - m_t)] = \mathscr{E}(x_{t+h} x_t). \tag{17.2.11}$$

Одной из задач теории временны́х рядов является оценивание параметров m_t и γ_h или проверка гипотез относительно m_t и γ_h по результатам наблюдений отрезка процесса длины n, включая и случай $n \to \infty$.

17.3. Спектральная функция стационарного процесса

(а) Частный случай. Если рассматривать только два первых момента стационарного процесса, то для описания процесса достаточно функции ковариаций γ_h. Покажем, что сама функция ковариаций может быть выражена через *спектральную функцию* процесса.

Рассмотрим сначала связь между функцией ковариаций и спектральной функцией для стационарного процесса (17.2.4). Легко проверяется, что при всех t

$$\mathscr{E}x_t = \sum_{p=1}^{r} a_p \mathscr{E}[\cos(t\omega_p + u_p)]. \qquad (17.3.1)$$

Для функции ковариаций $\gamma_h = E(x_{t+h}x_t)$ имеем

$$\gamma_{t-s} = \sum_{p,\, q=1}^{r} a_p a_q \; \mathscr{E}[\cos(s\omega_p + u_p)\, \cos(t\omega_q + u_q)]. \qquad (17.3.2)$$

Но $\mathscr{E}[\cos(s\omega_p + u_p)\cos(t\omega_q + u_q)] = 0$, $q \neq p$. Пользуясь тем, что при $q = p$

$$\cos(s\omega_p + u_p)\cos(t\omega_p + u_p) = \frac{1}{2}\cos((s+t)\,\omega_p + 2u_p) + \frac{1}{2}\cos(s-t)\;\omega_p,$$

находим

$$\mathscr{E}[\cos(s\omega_p + u_p)\cos(t\omega_p + u_p)] = \frac{1}{2}\cos(s-t)\omega_p.$$

Следовательно, полагая $t - s = h$, будем иметь

$$\gamma_h = \frac{1}{2}\sum_{p=1}^{r} a_p^2 \cos(h\omega_p) = \int_{-\pi}^{+\pi} \cos h\omega\, dF(\omega), \qquad (17.3.3)$$

где $F(\omega)$ — неубывающая ступенчатая функция на $[-\pi, +\pi]$, определяемая как

$$F(\omega) = \frac{1}{2}\sum_{\omega_p \leqslant \omega} a_p^2. \qquad (17.3.4)$$

В конечных точках интервала $[-\pi, +\pi]$ $F(-\pi) = 0$ и $F(+\pi) = \gamma_0$. $F(\omega)$ называется *спектральной функцией*, отвечающей функции ковариаций γ_h из (17.3.3). Крайнее справа выражение в (17.3.3) называется также *спектральным представлением* γ_h.

Для комплексных процессов вида (17.2.6) так же просто может быть получено спектральное представление γ_h:

$$\gamma_h = \frac{1}{2}\sum_{p=1}^{r} a_p^2 e^{ih\omega_p} = \int_{-\pi}^{+\pi} e^{ih\omega} dF(\omega), \qquad (17.3.5)$$

где $F(\omega)$ — та же спектральная функция, что и в (17.3.3).

(b) Общий случай. Пусть теперь

$$x_t,\ t = \ldots, -1, 0, +1, \ldots \qquad (17.3.6)$$

— произвольный вещественный стационарный процесс с $\mathscr{E}(x_t) = 0$ и функцией ковариаций γ_h. Для такого процесса $\gamma_{-h} = \gamma_h$. Для любого целого M определим функцию $F_M(\omega)$ на $[-\pi, +\pi]$ следующим образом:

$$F_M(\omega)=\frac{1}{\pi M}\int_{-\pi}^{\omega} \mathscr{E}(x_1\cos\omega+\ldots+x_M\cos M\omega)^2\,d\omega=$$

$$=\frac{1}{\pi M}\int_{-\pi}^{\omega}\sum_{p,q=1}^{M}\gamma_{p-q}\cos p\omega\cos q\omega\,d\omega=$$

$$=\frac{1}{\pi M}\left\{\sum_{p\neq q=1}^{M}\gamma_{p-q}\left[\frac{\sin(p+q)\omega}{2(p+q)}+\frac{\sin(p-q)\omega}{2(p-q)}\right]+\right.$$

$$\left.+\gamma_0\sum_{p=1}^{M}\frac{\sin p\omega\cos p\omega}{2p}+\frac{\gamma_0 M}{2}(\omega+\pi)\right\}. \quad (17.3.7)$$

Функция $F_M(\omega)$ не убывает на $[-\pi, +\pi]$, $F_M(-\pi)=0$, $F_M(+\pi)=$ $=\gamma_0$. При $M\to\infty$ $F_M(\omega)$ сходится к неубывающей функции $F(\omega)$ с $F(-\pi)=0$, $F(+\pi)=\gamma_0$.

Так как $\gamma_{-h}=\gamma_h$, достаточно рассматривать только неотрицательные значения h. Покажем, что при $h=0, 1, 2,\ldots$

$$\gamma_h=\int_{-\pi}^{+\pi}\cos h\omega\,dF(\omega). \quad (17.3.8)$$

Для $h=0,1,2,\ldots, M$ рассмотрим

$$\int_{-\pi}^{+\pi}\cos h\omega\,dF_M(\omega)=\frac{1}{\pi M}\int_{-\pi}^{+\pi}\sum_{p,\,q=1}^{M}\gamma_{p-q}\cos p\omega\cos q\omega\cos h\omega\,d\omega=$$

$$=\frac{1}{\pi M}(A_h+B_h),$$

где

$$A_h=\int_{-\pi}^{+\pi}\sum_{p\neq q=1}^{M}\gamma_{p-q}\cos p\omega\cos q\omega\cos h\omega\,d\omega,$$

$$B_h=\gamma_0\int_{-\pi}^{+\pi}\sum_{p=1}^{M}\cos^2 p\omega\cos h\omega\,d\omega.$$

Пользуясь равенствами

$$\cos p\omega\cos q\omega\cos h\omega=\frac{1}{4}[\cos(p+q+h)\omega+\cos(p+q-h)\omega+$$
$$+\cos(p-q+h)\omega+\cos(p-q-h)\omega],$$

$$\cos^2 p\omega\cos h\omega=\frac{1}{4}[\cos(2p+h)\omega+\cos(2p-h)\omega+2\cos h\omega],$$

$$\gamma_{p-q}=\gamma_{q-p},$$

получим

$$
\begin{aligned}
&A_h = \pi\gamma_h(M-h), && h \neq 0,\\
&A_h = 0, && h = 0,\\
&B_h = \pi\gamma_0\delta_h, && h \neq 0,\\
&B_h = \pi\gamma_0 M, && h = 0,
\end{aligned}
$$

где $\delta_h = \frac{1}{2}[1+(-1)^h]$.

Следовательно, при $h = 0, 1, \ldots, M$

$$
\int_{-\pi}^{+\pi} \cos h\omega \, dF_M(\omega) = \begin{cases} \gamma_h\left(1-\frac{h}{M}\right)+\frac{\gamma_0\delta_h}{M}, & h\neq 0,\\ \gamma_0, & h = 0. \end{cases} \tag{17.3.9}
$$

Если теперь положить $M \to \infty$, то мы получим (17.3.8) для неотрицательных h. Используя симметричность приращений $F(\omega)$ и функции $\cos h\omega$ относительно $\omega = 0$, (17.3.8) можно записать в виде

$$
\gamma_h = 2\int_0^{\pi} \cos h\omega \, dF(\omega). \tag{17.3.10}
$$

$F(\omega)$ называется *спектральной функцией* стационарного процесса x_t. Если нормировать $F(\omega)$, поделив ее на γ_0, то отношение $F(\omega')/\gamma_0$ будет представлять собой долю *спектральной массы*, приходящуюся на частоты $\omega \leqslant \omega'$. Если $F(\omega)$ абсолютно непрерывна и $f(\omega)$ — ее производная, то $f(\omega)$ называют *спектральной плотностью* процесса. Сформулируем полученный результат:

17.3.1. *Если x_t, $t = \ldots, -1, 0, +1, \ldots$ — вещественный стационарный процесс с $\mathcal{E}(x_t) = 0$, то функция ковариаций γ_h, $h = 0, 1, 2, \ldots$, допускает представление* (17.3.10), *где спектральная функция $F(\omega) = \lim\limits_{M\to\infty} F_M(\omega)$, а $F_M(\omega)$ определена в* (17.3.7).

Для комплексных стационарных процессов x_t, $t = \ldots, -1, 0, +1, \ldots$, с функцией ковариаций $\gamma_h = \mathcal{E}(x_{t+h}\bar{x}_t)$, $h = -1, 0, +1, \ldots$, теорема, соответствующая **17.3.1**, принадлежит Герглотцу (1911) и утверждает, что

$$
\gamma_h = \int_{-\pi}^{+\pi} e^{ih\omega} \, dF(\omega), \tag{17.3.11}
$$

где

$$
F(\omega) = \lim_{M\to\infty} F_M(\omega) \tag{17.3.12}
$$

и

$$
F_M(\omega) = \frac{1}{2\pi M}\int_{-\pi}^{\omega} \mathcal{E}\left(\sum_{p=1}^{M} x_p e^{-ip\omega} \sum_{q=1}^{M} \bar{x}_q e^{iq\omega}\right) d\omega. \tag{17.3.13}
$$

Можно показать, что подынтегральное выражение вещественно и неотрицательно.

Доказательство теоремы для комплексного случая аналогично доказательству **17.3.1** и предоставляется читателю.

Для описания большинства стационарных процессов, встречающихся в физических задачах, достаточно формул (17.3.10) и (17.3.11), так как в такого рода задачах можно так выбрать единицу времени, чтобы считать $t=\ldots,-1,0,+1,\ldots$ Обобщение (17.3.11) на случай, когда множеством значений h служит вещественная ось (времени), принадлежит Бохнеру (1932). Он доказал, что для комплексных процессов $\{x_t; -\infty<t<+\infty\}$ с непрерывной при $h=0$ функцией γ_h справедлива формула

$$\gamma_h=\int_{-\infty}^{+\infty} e^{ih\omega}\, dF(\omega). \tag{17.3.14}$$

Для вещественных стационарных процессов обобщением (17.3.10) является представление

$$\gamma_h=2\int_{0}^{\infty} \cos h\omega\, dF(\omega). \tag{17.3.15}$$

(c) Белый шум. Если вещественный стационарный процесс x_t, $t=\ldots,-1,0,+1,\ldots$, имеет постоянную спектральную плотность $f(\omega)=\gamma_0/(2\pi)$, $\omega\in[-\pi,+\pi]$, то из (17.3.8) следует, что $\gamma_h=0$ при $h\neq 0$. Обратно, пусть $\gamma_h=0$ при $h\neq 0$; тогда

$$\int_{-\pi}^{+\pi} \cos h\omega\, f(\omega)\, d\omega=\begin{cases}0, & h\neq 0,\\ \gamma_0, & h=0.\end{cases} \tag{17.3.16}$$

Если функция $f(\omega)$ симметрична относительно $\omega=0$ и может быть представлена рядом Фурье $b_0+b_1\cos\omega+b_2\cos 2\omega+\ldots$, то

$$\int_{-\pi}^{+\pi} \cos h\omega\, f(\omega)\, d\omega=\int_{-\pi}^{+\pi} [b_0+b_1\cos\omega+b_2\cos 2\omega+\ldots]\cos h\omega\, d\omega. \tag{17.3.17}$$

Но (17.3.17) приводится к виду

$$\int_{-\pi}^{+\pi} \cos h\omega\, f(\omega)\, d\omega=b_h\int_{-\pi}^{+\pi} \cos^2 h\omega\, d\omega. \tag{17.3.18}$$

Левая часть обращается в нуль при $h\neq 0$ в соответствии с (17.3.16); следовательно, $b_h=0$ для $h\neq 0$. Отсюда $f(\omega)=b_0$ и из того, что

$$\int_{-\pi}^{+\pi} f(\omega)\, d\omega=\gamma_0,$$

получаем $b_0=\gamma_0/(2\pi)$. Таким образом, на интервале $[-\pi,+\pi]$

$$f(\omega)=\frac{\gamma_0}{2\pi}. \tag{17.3.19}$$

Итак,

17.3.3. *Если вещественный стационарный процесс* $x_t, t = \ldots, -1, 0, +1, \ldots$, *имеет симметричную спектральную плотность* $f(\omega)$, *представимую рядом Фурье* $b_0 + b_1 \cos\omega + b_2 \cos 2\omega + \ldots$, *то необходимым и достаточным условием обращения в нуль функции ковариаций* γ_h *при* $h \neq 0$ *является постоянство* $f(\omega)$. *В этом случае* $f(\omega) = \gamma_0/(2\pi)$, $\omega \in [-\pi, +\pi]$.

Стационарный процесс x_t, $t = \ldots, -1, 0, +1, \ldots$, для которого x_t и $x_{t'}$ некоррелированы при $t \neq t'$, называется *белым шумом.* Теорема **17.3.3** дает необходимое и достаточное условие того, что процесс является белым шумом (для определенного класса спектральных плотностей).

17.4. Оценка среднего и функции ковариаций стационарного процесса

(a) Оценка среднего. Пусть x_t, $t = \ldots, -1, 0, +1, \ldots$, — вещественный стационарный процесс с $\mathscr{E}(x_t) = \mu$ и мы хотим оценить μ по выборке $(x_1, \ldots, x_n)$. Среднее $\bar{x} = \frac{x_1 + \ldots + x_n}{n}$ будет несмещенной оценкой μ, так как

$$\mathscr{E}(\bar{x}) = \mu. \tag{17.4.1}$$

Дисперсия $\bar{x}$ равна

$$\sigma^2(\bar{x}) = \frac{1}{n}\left[\gamma_0 + 2\sum_{\xi=1}^{n-1}\left(1 - \frac{\xi}{n}\right)\gamma_\xi\right]. \tag{17.4.2}$$

Если $F(\omega)$ имеет производную $f(\omega)$, непрерывную в нуле и изображаемую рядом Фурье

$$\frac{1}{2\pi}(\gamma_0 + 2\gamma_1 \cos\omega + 2\gamma_2 \cos 2\omega + \ldots), \tag{17.4.3}$$

то нетрудно проверить, что величина в квадратных скобках в (17.4.2) стремится при $n \to \infty$ к $2\pi f(0)$. Следовательно, в этом случае при больших n

$$\sigma^2(\bar{x}) \approx \frac{2\pi f(0)}{n}. \tag{17.4.4}$$

(b) Оценка ковариации. Для оценки

$$c_h = \frac{1}{n-h}\sum_{\xi=1}^{n-h}(x_\xi - \bar{x})(x_{\xi+h} - \bar{x}) \tag{17.4.5}$$

ковариации γ_h после алгебраических преобразований получим

$$\mathscr{E}(c_h) = \gamma_h - \sigma^2(x) + \frac{2}{n}R, \tag{17.4.6}$$

где

$$R=\sum_{\xi=1}^{n-1}\left(1-\frac{\xi}{n}\right)\gamma_\xi-\sum_{\xi=1}^{n-h-1}\left(1-\frac{\xi}{n-h}\right)(\gamma_\xi+\gamma_{\xi+h})-\sum_{\xi=1}^{h}\gamma_\xi. \quad (17.4.7)$$

Можно показать, что $R\to 0$ при $n\to\infty$, так что при больших n смещение c_h как оценки γ_h приблизительно равно $-\sigma^2(x)$. Если имеется непрерывная в нуле спектральная плотность $f(\omega)$, то при больших n

$$\mathscr{E}(c_h)\approx\gamma_h-\frac{2\pi f(0)}{n}. \quad (17.4.8)$$

Заметим, что если μ известно и в (17.4.5) $\bar{x}$ заменено на μ, то среднее значение полученной оценки в точности равно γ_h.

Вычисление дисперсии c_h требует предположения о конечности моментов четвертого порядка величин x_t, $t=\ldots,-1,0,+1,\ldots$, и связано с некоторыми алгебраическими преобразованиями. Однако для стационарных в узком смысле процессов x_t можно получить приближенное выражение для дисперсии c_h. Рассмотрим процесс $z_t=(x_t-\mu)(x_{t+h}-\mu)$, $t=\ldots,-1,0,+1,\ldots$, который стационарен в узком смысле, если таковым был x_t; $\mathscr{E}(z_t)=\gamma_h$. Предположим, что существует функция ковариаций $\gamma^*_{h,g}$ процесса z_t

$$\gamma^*_{h,g}=\mathscr{E}(z_t z_{t+g}). \quad (17.4.9)$$

Можно показать, что при больших n дисперсия c_h приблизительно равна дисперсии c^*_h, где

$$c^*_h=\frac{1}{n-h}\sum_{\xi=1}^{n-h}x_\xi x_{\xi+h}. \quad (17.4.10)$$

Дисперсия c^*_h имеет ту же структуру, что и $\sigma^2(\bar{x})$, и равна

$$\sigma^2(c^*_h)=\frac{1}{n-h}\left[\gamma^*_{h,0}+2\sum_{\xi=1}^{n-h-1}\left(1-\frac{\xi}{n-h}\right)\gamma^*_{h,\xi}\right]. \quad (17.4.11)$$

Если z_t имеет спектральную плотность, то при больших n

$$\sigma^2(c^*_h)\approx\frac{\gamma^*_{h,0}\left(3-\frac{1}{\pi}\right)}{n}, \quad (17.4.12)$$

где

$$\gamma^*_{h,0}=\mathscr{E}\,[(x_t-\mu)(x_{t+h}-\mu)]^2.$$

17.5. Оценка спектральной функции

Пусть стационарный процесс x_t, $t=\ldots,-1,0,+1,\ldots$, имеет среднее μ и спектральную плотность $f(\omega)$. Рассмотрим задачу оценивания $f(\omega)$ по выборке $(x_1,\ldots,x_n)$, считая μ известным.

Если вспомнить определение $F(\omega)$ как предела последовательности функций $F_M(\omega)$, определенных в (17.3.7), при $M \to \infty$, то в качестве оценки $f(\omega)$ по выборке $(x_1, \ldots, x_n)$ напрашивается выражение

$$f_n(\omega) = \frac{1}{n\pi}(y_1 \cos \omega + y_2 \cos 2\omega + \ldots + y_n \cos n\omega)^2, \quad (17.5.1)$$

где $y_t = x_t - \mu$. Для нее легко получим

$$\mathscr{E} f_n(\omega) = \frac{1}{2\pi}\left\{\gamma_0\left[1 + \frac{\sin n\omega \cos(n+1)\omega}{n \sin \omega}\right] + \right.$$
$$\left. + 2\sum_{\xi=1}^{n-1}\gamma_\xi\left[\left(1 - \frac{\xi}{n}\right)\cos h\omega + \frac{\sin(n-\xi)\omega \cos(n+1)\omega}{n \sin \omega}\right]\right\}. \quad (17.5.2)$$

Устремляя n к бесконечности и используя ряд Фурье (17.4.3) функции $f(\omega)$, найдем, что для всех $\omega \in [-\pi, +\pi]$, кроме $\omega = 0$,

$$\lim_{n \to \infty} \mathscr{E} f_n(\omega) = f(\omega), \quad \text{а} \lim_{n \to \infty} \mathscr{E} f_n(0) = 2f(0).$$

Таким образом, $f_n(\omega)$ оказывается асимптотически несмещенной оценкой $f(\omega)$ для $\omega \in [-\pi, +\pi]$, $\omega \neq 0$. С другой стороны, при слабых условиях можно доказать, что дисперсия $f_n(\omega)$ не стремится к нулю при $n \to \infty$. В частности, если $(x_1, \ldots, x_n)$ — n-мерная нормальная случайная величина, то $f_n(\omega)$ при всех n только постоянным множителем отличается от величины, имеющей χ^2-распределение с одной степенью свободы.

Мы обнаружили, что естественная оценка $f_n(\omega)$ спектральной плотности $f(\omega)$ оказывается несостоятельной и следует искать другие оценки. Различные состоятельные оценки $f(\omega)$ предлагались и изучались Бартлеттом (1950), Гренандером и Розенблаттом (1953), Дженкинсом и Пристли (1957), Парзеном (1957), Тьюки (1949а) и другими. Бо́льшая часть этих процедур предназначена для оценивания $f(\omega)$ в фиксированной точке ω интервала $[-\pi, +\pi]$. Но метод Тьюки позволяет оценивать спектральную массу на подынтервалах $[-\pi, +\pi]$. Мы рассмотрим метод, аналогичный методу Тьюки, и не будем обсуждать другие оценки. Читатель, интересующийся ими, может обратиться к книгам Бартлетта (1950) и Гренандера и Розенблатта (1957).

Предположим, что спектральная плотность $f(\omega)$ стационарного процесса, наблюдения над которым дали выборку $(x_1, \ldots, x_n)$, представляется рядом Фурье (17.4.3). Для заданного m будем оценивать по выборке $(x_1, \ldots, x_n)$ спектральные массы

$$\Delta_p = F\left(\frac{p\pi}{m} + \frac{\pi}{2m}\right) - F\left(\frac{p\pi}{m} - \frac{\pi}{2m}\right), \quad (17.5.3)$$

содержащиеся в интервалах

$$I_p = \left(\frac{p\pi}{m} - \frac{\pi}{2m}, \frac{p\pi}{m} + \frac{\pi}{2m}\right], \qquad p = 0, \pm 1, \ldots, \pm m, \quad (17.5.4)$$

где $F(\omega)$ — спектральная функция.

Число m выбирается так, чтобы, с одной стороны, не пришлось оценивать слишком много спектральных масс, а с другой — чтобы по спектральным массам, содержащимся в интервалах I_p, $p=0$, $\pm 1, \ldots, \pm m$, можно было составить представление о спектральной функции. Грубо говоря, это означает, что m выбирается равным числу ковариаций, дающих «приемлемое» описание функции ковариаций процесса.

Спектральная масса Δ_p может быть записана в виде

$$\Delta_p = \int\limits_{\frac{p\pi}{m}-\frac{\pi}{2m}}^{\frac{p\pi}{m}+\frac{\pi}{2m}} f(\omega)\, d\omega. \tag{17.5.5}$$

Используя ряд Фурье (17.4.3) функции $f(\omega)$ и произведя интегрирование, получим

$$\Delta_p = \frac{\gamma_0}{2m} + \frac{1}{m} \sum_{h=1}^{\infty} \gamma_h \frac{\sin(h\pi/2m)}{h\pi/2m} \cos\frac{hp\pi}{m}. \tag{17.5.6}$$

Если считать, что ковариации $\gamma_0, \gamma_1, \ldots, \gamma_m$ содержат «достаточное» количество информации о стационарном процессе, то для спектральных масс Δ_p, $p=0, \pm 1, \ldots, \pm m$, можно предложить следующее приближенное выражение:

$$\tilde{\Delta}_p = \frac{\gamma_0}{2m} + \frac{1}{m} \sum_{h=1}^{m} \gamma_h \frac{\sin(h\pi/2m)}{h\pi/2m} \cos\frac{hp\pi}{m}, \tag{17.5.7}$$

$$p = 0, \pm 1, \ldots, \pm m.$$

Если μ известно, то естественной оценкой $\tilde{\Delta}_p$ по выборке $(x_1, \ldots, x_n)$ будет

$$\tilde{\Delta}_{pn} = \frac{c_0}{2m} + \frac{1}{m} \sum_{h=1}^{m} c_h \frac{\sin(h\pi/2m)}{h\pi/2m} \cos\frac{hp\pi}{m}, \tag{17.5.8}$$

где $c_h = \frac{1}{n-h} \sum_{\xi=1}^{n-h} y_\xi y_{\xi+h}$, $h=0, 1, \ldots, n-1$, $m \leqslant n-1$, $y_\xi = x_\xi - \mu$. Ясно, что, поскольку $\mathscr{E} c_h = \gamma_h$, $h=0, 1, \ldots$,

$$\mathscr{E}(\tilde{\Delta}_{pn}) = \tilde{\Delta}_p, \tag{17.5.9}$$

так что $\tilde{\Delta}_{pn}$ — несмещенная оценка $\tilde{\Delta}_p$.

Отметим, что оценка $\tilde{\Delta}_{pn}$ спектральной массы несколько отличается от оценки, предложенной Тьюки (1949 а). Если заменить $\frac{\sin(h\pi/2m)}{h\pi/2m}$ приближением $0{,}46\cos(h\pi/m) + 0{,}54$ и ввести поправку в крайней точке приписыванием коэффициента $1/2$ последнему члену суммы в (17.5.8), то получится оценка Тьюки для Δ_p.

Так как $\tilde{\Delta}_{pn}$ служит оценкой спектральной массы на интервале $(p\pi/m - \pi/2m,\ p\pi/m + \pi/2m]$, то несмещенную оценку среднего значения $\overline{f}_p(\omega)$ спектральной плотности на интервале I_p можно получить делением $\tilde{\Delta}_{pn}$ на длину интервала π/m, т. е.

$$\mathscr{E}^{-1}(\overline{f}_p(\omega)) = \frac{m}{\pi}\tilde{\Delta}_{pn}. \tag{17.5.10}$$

Можно показать, что дисперсия оценки (17.5.10) при больших n и больших m $(m = O(n))$ равна

$$\sigma^2\left(\frac{m}{\pi}\tilde{\Delta}_{pn}\right) \approx \frac{2}{n}\overline{f}_p^2(\omega)\sum_{h=-m}^{m}\left(\frac{\sin(h\pi/2m)}{h\pi/2m}\right)^2\cos\frac{hp\pi}{m}. \tag{17.5.11}$$

Если среднее μ для процесса неизвестно, то в определении c_h следует заменить μ на $\bar{x}$ и пользоваться в (17.5.8.) модифицированными значениями c_h, $h = 0, 1, \ldots, m$, для получения оценки спектральной массы $\tilde{\Delta}_{pn}$. Можно показать, что при таком изменении c_h (17.5.9) и (17.5.11) имеют место при больших n.

17.6. Статистические критерии для процессов с конечным числом параметров

Полное изложение различных задач статистических выводов для стационарных процессов лежит вне сферы данной книги. Здесь мы рассмотрим задачи оценивания и проверки гипотез только для некоторых классических процессов с конечным числом параметров. Дальнейшее изучение этих задач можно найти в книгах Бартлетта (1955) и Гренандера и Розенблатта (1957).

(а) Метод переменных разностей. Пусть процесс x_t, $t = \ldots, -1, 0, +1, \ldots$, имеет вид

$$x_t = \sum_{p=0}^{k}\beta_p t^p + \xi_t = \mu_t + \varepsilon_t, \tag{17.6.1}$$

где постоянные $\beta_0, \beta_1, \ldots, \beta_k$ неизвестны, а процесс ε_t представляет собой белый шум с дисперсией σ^2.

Если k известно, то оценки $\beta_0, \beta_1, \ldots, \beta_k$ с минимальной дисперсией по выборке $(x_1, \ldots, x_n)$, $n \geqslant k+1$, можно получить методом наименьших квадратов, а дисперсии этих оценок вычислить методами §10.3 (с). Эта же процедура применима, если в (17.6.1) все t^p заменены полиномами $g_p(t)$ степени p, $p = 0, 1, \ldots, k$, где $g_0(t), g_1(t), \ldots, g_k(t)$ функционально независимы.

Если же в (17.6.1) k неизвестно, то для оценки k можно воспользоваться полуэмпирической процедурой, известной под названием *метода переменных разностей*. Этот метод применяется следующим образом. Пусть $y_{h,t}$ — процесс, образованный h-ми правыми разностями процесса x_t, т. е.

$$y_{h,t}=\Delta^h x_t=\Delta^h \mu_t+\Delta^h \varepsilon_t, \tag{17.6.2}$$

где

$$\Delta^h \varepsilon_t=\varepsilon_{t+h}-\binom{h}{1}\varepsilon_{t+h-1}+\binom{h}{2}\varepsilon_{t+h-2}-\ldots+(-1)^h \varepsilon_t \tag{17.6.3}$$

и $\Delta^h \mu_t$ выражается аналогично. Рассмотрим последовательность выборок $(x_1, \ldots, x_{n+h})$, $h=1, 2, \ldots$, и образуем отношения

$$Q_h=\sum_{\xi=1}^{n} y^2_{h,\xi}\Big/\left[n\binom{2h}{h}\right], \qquad h=1, 2, \ldots \tag{17.6.4}$$

Среднее значение Q_h равно

$$\mathscr{E}Q_h=\mathscr{E}\sum_{\xi=1}^{n}(\Delta^h \mu_\xi+\Delta^h \varepsilon_\xi)^2\Big/\left[n\binom{2h}{h}\right]. \tag{17.6.5}$$

Так как ε_t — белый шум, а μ_t — неслучайная функция t, то

$$\mathscr{E}(\Delta^h \mu_\xi+\Delta^h \varepsilon_\xi)^2=(\Delta^h \mu_\xi)^2+\mathscr{E}(\Delta^h \varepsilon_\xi)^2. \tag{17.6.6}$$

Но

$$\mathscr{E}(\Delta^h \varepsilon_\xi)^2=\sigma^2\sum_{p=0}^{h}\binom{h}{p}^2=\binom{2h}{h}\sigma^2. \tag{17.6.7}$$

Объединяя теперь (17.6.7), (17.6.6) и (17.6.5), найдем, что

$$\mathscr{E}Q_h=R_h+\sigma^2, \qquad h=1, 2, \ldots, \tag{17.6.8}$$

где $R_h=\frac{1}{n}\sum_{\xi=1}^{n}(\Delta^h \mu_\xi)^2\Big/\binom{2h}{h}$.

Мы видим, что $R_h \geqslant 0$ и $R_h=0$ для $h \geqslant k+1$. Следовательно, при $h \geqslant k+1$ $\mathscr{E}Q_h=\sigma^2$.

На практике поступают так. Вычисляют значения Q_h, $h=1, 2, \ldots$, и в качестве оценки k выбирают значение h, начиная с которого Q_h становится почти постоянной (с точностью до «малых» случайных флуктуаций); самое это значение Q_h служит оценкой σ^2. Читателю, интересующемуся подробным изложением метода переменных разностей и различных его модификаций, следует обратиться к работам Тинтнера (1940) и Куэнуила (1948).

(b) Тригонометрические схемы с известными коэффициентами и периодом. Предположим, что процесс x_t, $t=\ldots, -1, 0, +1, \ldots$, имеет следующую параметрическую форму:

$$x_\xi=\sum_{p=1}^{k}[a_p \cos \omega_p \xi+b_p \sin \omega_p \xi]+\varepsilon_\xi, \tag{17.6.9}$$

где ε_ξ — белый шум с известной дисперсией σ^2, (вещественные) постоянные a_p, b_p, ω_p известны, $0 \leqslant \omega_p \leqslant \pi$.

Умножим обе части (17.6.9) на $\frac{1}{n} \cos \omega\xi$ и просуммируем по ξ от 1 до n. Получим

$$\alpha(\omega) = \sum_{p=1}^{k} [a_p u_p(\omega) + b_p v_p(\omega)] + \delta(\omega), \tag{17.6.10}$$

где

$$\begin{aligned} \alpha(\omega) &= \frac{1}{n} \sum_{\xi=1}^{n} x_\xi \cos \omega\xi, \\ u_p(\omega) &= \frac{1}{n} \sum_{\xi=1}^{n} \cos \omega_p\xi \cos \omega\xi, \\ v_p(\omega) &= \frac{1}{n} \sum_{\xi=1}^{n} \sin \omega_p\xi \cos \omega\xi, \\ \delta(\omega) &= \frac{1}{n} \sum_{\xi=1}^{n} \varepsilon_\xi \cos \omega\xi. \end{aligned} \tag{17.6.11}$$

Аналогично

$$\alpha'(\omega) = \sum_{p=1}^{k} [a_p u'_p(\omega) + b_p v'_p(\omega)] + \delta'(\omega), \tag{17.6.12}$$

где $\alpha'(\omega)$, $u'_p(\omega)$, $v'_p(\omega)$, $\delta'(\omega)$ получаются соответственно из $\alpha(\omega)$, $u_p(\omega)$, $v_p(\omega)$, $\delta(\omega)$ заменой $\cos \omega\xi$ на $\sin \omega\xi$.

Заметим, что $u_p(\omega)$, $v_p(\omega)$, $u'_p(\omega)$, $v'_p(\omega)$ можно записать в виде

$$\begin{aligned} u_p(\omega) &= \frac{1}{2n} \sum_{\xi=1}^{n} [\cos(\omega_p + \omega)\xi + \cos(\omega_p - \omega)\xi], \\ v_p(\omega) &= \frac{1}{2n} \sum_{\xi=1}^{n} [\sin(\omega_p + \omega)\xi + \sin(\omega_p - \omega)\xi], \\ u'_p(\omega) &= \frac{1}{2n} \sum_{\xi=1}^{n} [\sin(\omega_p + \omega)\xi - \sin(\omega_p - \omega)\xi], \\ v'_p(\omega) &= \frac{1}{2n} \sum_{\xi=1}^{n} [-\cos(\omega_p + \omega)\xi - \cos(\omega_p - \omega)\xi], \end{aligned} \tag{17.6.13}$$

откуда легко получить, что при $n \to \infty$

$$\mathscr{E}[\alpha(\omega)] \to \begin{cases} 0, & \omega \neq \omega_p, \quad p = 1, \ldots, k, \\ \dfrac{a_p}{2}, & \omega = \omega_p. \end{cases} \tag{17.6.14}$$

Точно так же при $n \to \infty$

$$\mathscr{E}\,[\alpha'(\omega)] \to \begin{cases} 0, & \omega \neq \omega_p, \quad p = 1, \ldots, k, \\ -\dfrac{b_p}{2}, & \omega = \omega_p. \end{cases} \tag{17.6.15}$$

Если для данного процесса периоды $2\pi/\omega_p$, $p = 1, \ldots, k$, известны, но коэффициенты a_p, b_p, $p = 1, \ldots, k$, неизвестны, то в соответствии с § 10.3 (c) линейные оценки этих коэффициентов с минимальной дисперсией можно найти методом наименьших квадратов.

(c) Тригонометрические схемы с неизвестными параметрами — анализ периодограмм*). Пусть процесс имеет вид (17.6.9), где a_p, b_p, ω_p и даже k неизвестны. Для этого случая Шастер (1898) предложил метод отыскания возможных периодов, а Фишер (1929) разработал метод проверки значимости предполагаемых периодов. Мы рассмотрим последний метод, называемый *анализом периодограмм.*

Описанное выше поведение средних значений функций $\alpha(\omega)$ и $\alpha'(\omega)$ показывает, что эти функции полезны для «просеивания» истинных периодов, если таковые существуют.

Предположим, что на самом деле периодов нет, т. е. все a_p и b_p равны нулю, тогда x_t — белый шум. Для нечетного n, $n = 2k + 1$, положим

$$\omega_p = \frac{2\pi p}{2k+1}, \qquad p = 1, \ldots, k. \tag{17.6.16}$$

Рассмотрим теперь $\alpha(\omega_p)$, $\alpha'(\omega_p)$, $p = 1, \ldots, k$, которые являются линейными функциями x_t. Ковариация $\alpha(\omega_p)$ и $\alpha(\omega_q)$, $p \neq q$, равна

$$\operatorname{cov}(\alpha(\omega_p), \alpha(\omega_q)) = \frac{\sigma^2}{(2k+1)} \sum_{\xi=1}^{2k+1} \cos\omega_p\xi \cos\omega_q\xi = \frac{\sigma^2}{2\,(2k+1)^2} \sum_{\xi=1}^{2k+1} \times$$
$$\times [\cos(\omega_p + \omega_q)\,\xi + \cos(\omega_p - \omega_q)\,\xi]. \tag{17.6.17}$$

Но

$$\sum_{\xi=1}^{2k+1} \cos(\omega_p \pm \omega_q)\,\xi = \begin{matrix}\text{вещественной}\\ \text{части}\end{matrix} \left\{ \frac{e^{i(\omega_p \pm \omega_q)}\,[1 - e^{i(\omega_p \pm \omega_q)(2k+1)}]}{1 - e^{i(\omega_p \pm \omega_q)}} \right\}.$$

Так как $(\omega_p \pm \omega_q)(2k+1)$ кратно 2π, то величина в квадратных скобках предыдущего равенства обращается в нуль.

Следовательно,

$$\operatorname{cov}(\alpha(\omega_p), \alpha(\omega_q)) = 0, \; p \neq q.$$

---

*) В § 17.6 (c) используется сходимость по вероятности $\alpha(\omega)$ $(\alpha'(\omega))$ к нулю при $\omega \neq \omega_p$ и к $\frac{a_p}{2}\left(-\frac{b_p}{2}\right)$ при $\omega = \omega_p$, которая обеспечивается применимостью закона больших чисел к последовательности $\varepsilon_\xi \cos\omega\xi$, $\xi = 1, 2, \ldots$ (*Прим. перев.*)

Аналогично получим

$$\operatorname{cov}(\alpha'(\omega_p),\ \alpha'(\omega_q)) = 0,\ p \neq q,$$

и

$$\operatorname{cov}(\alpha'(\omega_p),\ \alpha(\omega_q)) = 0,\ p,\ q = 1,\ \ldots,\ k.$$

Для $p = q$ из (17.6.17) будем иметь

$$\sigma^2(\alpha(\omega_p)) = \frac{\sigma^2}{2(2k+1)}$$

и точно так же

$$\sigma^2(\alpha'(\omega_p)) = \frac{\sigma^2}{2(2k+1)}.$$

Подытожим все сказанное.

17.6.1. *Если $x_1, \ldots, x_{2k+1}$ — независимые случайные величины со средним 0 и дисперсией σ^2, то $\alpha(\omega_p)$, $\alpha'(\omega_p)$, $p = 1, \ldots, k$, где числа ω_p определены в* (17.6.16), *будут случайными величинами с нулевыми средними, нулевыми ковариациями и дисперсиями* $\sigma^2/(4k+2)$.

Теперь мы в состоянии проверить выборку $(x_1, \ldots, x_{2k+1})$ на предмет обнаружения периодов длины $2\pi/\omega_p$; $p = 1, \ldots, k$. Если $2\pi/\omega_p$ — истинный период, то мы знаем *), что при $n \to \infty$ $\alpha(\omega_p)$ и $\alpha'(\omega_p)$ сходятся по вероятности к значениям, отличным от нуля. Это означает, что $\alpha^2(\omega_p) + \alpha'^2(\omega_p)$ сходится по вероятности к положительному значению, если $2\pi/\omega_p$ — истинный период. Поэтому необходимо иметь критерий, который позволил бы проверить, значимо ли отклонение наибольшей величины (или m-й по порядку, начиная с наибольшей) среди $\alpha^2(\omega_p) + \alpha'^2(\omega_p)$, $p = 1, \ldots, k$, от нуля, если x_t — белый шум. Дополнительная трудность связана с тем, что дисперсия σ^2 обычно неизвестна.

Построим *периодограмму* выборки $(x_1, \ldots, x_{2k+1})$:

$$I_{2k+1}(\omega) = \frac{2k+1}{2\pi}\,[\alpha^2(\omega) + \alpha'^2(\omega)]. \tag{17.6.18}$$

Сделаем еще одно предположение, именно, что x_t — нормальный белый шум, тогда $x_1, \ldots, x_{2k+1}$ — независимые случайные величины, распределенные по одному и тому же закону $N(0, \sigma^2)$.
Величины

$$\frac{\sqrt{2(2k+1)}}{\sigma}\,\alpha(\omega_p), \qquad \frac{\sqrt{2(2k+1)}}{\sigma}\,\alpha'(\omega_p), \qquad p = 1, \ldots, k,$$

независимы в совокупности и имеют распределения $N(0, 1)$. Положим

$$u_p = \frac{2k+1}{\sigma^2}[\alpha^2(\omega_p) + \alpha'^2(\omega_p)] = \frac{2\pi}{\sigma^2} I_{2k+1}(\omega_p), \qquad p = 1, \ldots, k. \tag{17.6.19}$$

*) См. примечание на стр. 529. (*Прим. перев.*)

Тогда $u_1, \ldots, u_k$ будут независимыми случайными величинами, совместное распределение которых задается элементом вероятности

$$e^{-(u_1+\ldots+u_k)}du_1\ldots du_k \tag{17.6.20}$$

в области $u_p \geqslant 0$, $p=1,\ldots, k$ и 0 в остальных точках.

Проверка значимости отклонения от нуля наибольшей из величин $\alpha^2(\omega_p)+\alpha'^2(\omega_p)$ (или $I_{2k+1}(\omega_p)$) сводится теперь к проверке значимости отклонения от нуля наибольшей среди $u_1,\ldots u_k$. Так как σ^2 неизвестна, то естественно проверять, значимо ли отклонение от нуля наибольшего из отношений

$$\frac{u_p}{u_1+\ldots+u_k}, \qquad p=1,\ldots, k, \tag{17.6.21}$$

которые, очевидно, не зависят от σ^2.

Пусть v_r обозначает r-ю по величине (начиная с наименьшей) среди $u_1,\ldots, u_k$, $r=1, \ldots, k$. Элемент вероятности порядковых статистик $v_1,\ldots, v_k$ равен

$$k!\ e^{-(v_1+\ldots+v_k)}\,dv_1\ldots dv_k \tag{17.6.22}$$

в области $0<v_1<v_2<\ldots<v_k<+\infty$. Положим

$$g=\frac{v_k}{v_1+\ldots+v_k}. \tag{17.6.23}$$

Величина g совпадает с наибольшим из отношений (17.6.21). Получим распределение g, определив сначала моменты с помощью простого, но эффективного приема.

Заметим, что h-й момент g равен

$$\mathscr{E}g^h=\int\limits_{-\infty}^{0}\ldots\int\limits_{-\infty}^{0}\mathscr{E}\left(e^{(v_1+\ldots+v_k)\varphi}\,v_k^h\right)dz_1\ldots dz_h, \tag{17.6.24}$$

где $\varphi=z_1+\ldots+z_h$. Интегрируя по $v_1,\ldots, v_{k-1}$, находим

$$\mathscr{E}\left(e^{(v_1+\ldots+v_k)\varphi}\,v_k^h\right)=k(1-\varphi)^{-(k-1)}\int\limits_0^\infty\left(1-e^{-v_k(1-\varphi)}\right)^{k-1}e^{-v_k(1-\varphi)}\,v_k^h\,dv_k. \tag{17.6.25}$$

Подставляя это выражение в (17.6.24), разлагая $\left(1-e^{-v_k(1-\varphi)}\right)^{k-1}$, производя интегрирование по v_k и затем по $z_1,\ldots, z_h$, будем иметь

$$\mathscr{E}(g^h)=k(k-1)\sum_{p=0}^{k-1}(-1)^p\binom{k-1}{p}\frac{\Gamma(h+1)\,\Gamma(k-1)}{(p+1)^{h+1}\,\Gamma(h+k)}. \tag{17.6.26}$$

Но

$$\frac{\Gamma(h+1)\,\Gamma(k-1)}{\Gamma(h+k)}=\int\limits_0^1 w^h(1-w)^{k-2}\,dw. \tag{17.6.27}$$

Следовательно,

$$\mathcal{E}(g^h)=k(k-1)\sum_{p=0}^{k-1}(-1)^p\binom{k-1}{p}\int_0^1\left(\frac{w}{p+1}\right)^h(1-w)^{k-2}d\left(\frac{w}{p+1}\right). \tag{17.6.28}$$

Положим $w/(p+1)=y$, тогда (17.6.28) перейдет в

$$\mathcal{E}(g)^h=k(k-1)\sum_{p=0}^{k-1}(-1)^p\binom{k-1}{p}\int_0^{\frac{1}{p+1}}y^h[1-(p+1)y]^{k-2}\,dy. \tag{17.6.29}$$

Если ввести теперь функцию

$$\theta_p(y)=\begin{cases}[1-(p+1)y]^{k-2}, & 0<y<\dfrac{1}{p+1},\\ 0 & \text{в остальных точках,}\end{cases}$$

то (17.6.29) запишется в виде

$$\mathcal{E}(g^h)=k(k-1)\int_0^1\sum_{p=0}^{k-1}(-1)^p\binom{k-1}{p}y^h\,\theta_p(y)\,dy, \tag{17.6.30}$$

где $h=1,2\ldots$.

Из **5.5.1а** теперь заключаем, что g и y распределены одинаково. Следовательно, плотность вероятности g есть

$$f(g)=k(k-1)\sum_{p=0}^{k-1}(-1)^p\binom{k-1}{p}\theta_p(g). \tag{17.6.31}$$

Отсюда

$$P(g>g')=\sum_{p=0}^{r}(-1)^p\binom{k}{p+1}[1-(p+1)g']^{k-1}, \tag{17.6.32}$$

где r — наибольшее целое, не превосходящее $k-1$, для которого $1-(r+1)g'\geqslant 0$. Этот результат был впервые получен Фишером (1929) другим методом. Дэвис (1941) составил таблицы значений $P(g>g_a)$ для $g_ak=0{,}10\,(0{,}10)\,10{,}0$; $k=10\,(10)\,70$ и для $g_ak=5{,}1\,(0{,}1)\,10{,}1$; $k=80\,(10)\,160\,(20)\,300$.

Сформулируем полученный результат.

17.6.2 *Пусть* $x_1,\ldots,x_{2k+1}$ — *независимые случайные величины, распределенные по закону* $N(0,\sigma^2)$; u_p, $p=1,\ldots,k$, *определены* в (17.6.19), *а* $\alpha(\omega_p)$ *и* $\alpha'(\omega_p)$ *определены в* (17.6.11) *и* (17.6.12). *Тогда, если* $g=\max\{u_p\}/(u_1+\ldots+u_k)$, *то плотность вероятности* g *задается формулой* (17.6.31), *а* $P(g>g')$ *для любого* $g'\in(0,1)$ — *формулой* (17.6.32).

Методами, аналогичными использованным, можно показать, что если g — m-я по величине (начиная с наибольшей) среди $u_1, \ldots, u_k$, деленная на $u_1 + \ldots + u_k$, то

$$P(g > g') = \frac{k!}{(m-1)!} \sum_{p=m}^{r} \frac{(-1)^{p-m}(1-pg')^{k-1}}{p(k-p)!\,(p-m)!}, \qquad (17.6.33)$$

где r — наибольшее целое, для которого $1 - rg' \geqslant 0$.

Отметим, наконец, что полученные результаты относятся к случаю $\mathscr{E} x_t = 0$. Если среднее значение неизвестно, то приближенные формулы, по крайней мере при больших n, можно получить, заменив всюду x_ξ на $x_\xi - \bar{x}$, $\xi = 1, \ldots, 2k+1$.

17.7. Проверка отклонения нормального процесса от белого шума

Пусть по данным выборки $(x_1, \ldots, x_n)$ мы хотим проверить гипотезу о том, что нормальный процесс $\{x_t\}$ представляет собой белый шум, т. е. что $x_1, \ldots, x_n$ — независимые случайные величины, распределенные по закону $N(\mu, \sigma^2)$. Для этой цели Р. Андерсон (1942) предложил критерий, основанный на отношении

$$R = \frac{c_1'}{c_0'}, \qquad (17.7.1)$$

где

$$\sigma^2 c_h' = \sum_{\xi=1}^{n} (x_\xi - \bar{x})(x_{\xi+h} - \bar{x}), \quad h = 0,\ 1,$$

и $x_{n+1} \equiv x_1$. Хотя условие $x_{n+1} \equiv x_1$ выглядит произвольным, употребление *циклического* коэффициента корреляции c_1' значительно облегчает вычисление распределения R. Если $\{x_t\}$ — белый шум, то при больших n значения R лежат около нуля; в противном случае этого не будет. Далее, c_1' и c_0' суть квадратичные формы переменных от $x_1, \ldots, x_n$. Разность $c_1' - R'c_0'$, где R' — число, также квадратичная форма $x_1, \ldots, x_n$. Так как

$$P(c_1' - R'c_0' > 0) = P(R > R'), \qquad (17.7.2)$$

то распределение R определяется распределением квадратичной формы $c_1' - R'c_0'$. Если $x_1, \ldots, x_n$ нормальны и независимы, то характеристическая функция $c_1' - R'c_0'$ равна

$$\varphi(t) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int\limits_{R_n} \exp\left[i(c_1' - R'c_0')t - \frac{1}{2\sigma^2}\sum_{\xi=1}^{n}(x_\xi - \mu)^2\right] dx_1 \ldots dx_n. \qquad (17.7.3)$$

Используя методы и результаты § 8.4 и замечая, что $c_1' - R'c_0'$ является квадратичной формой также относительно $x_1 - \mu, \ldots, x_n - \mu$, получим

$$\varphi(t) = \prod_{\xi=1}^{n-1} [1 - 2it(d_\xi - R')]^{-\frac{1}{2}}, \qquad (17.7.4)$$

где $d_\xi = \cos(2\pi\xi/n)$, $\xi = 1, \ldots, n-1$. Так как $d_\xi = d_{n-\xi}$, то для нечетных n

$$\varphi(t) = \prod_{\xi=1}^{\frac{1}{2}(n-1)} [1 - 2it(d_\xi - R')]^{-1}. \qquad (17.7.5)$$

Плотность вероятности величины $y = c_1' - R'c_0'$ согласно **5.1.2** равна

$$g(y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iyt} \varphi(t)\, dt. \qquad (17.7.6)$$

Функция $\varphi(z)$ имеет простые полюсы в точках $z_\xi = -\frac{1}{2} i/(d_\xi - R')$, $\xi = 1, \ldots, (n-1)/2$ комплексной плоскости. Вычисляя интеграл (17.7.6) методом вычетов, получим

$$g(y) = \frac{1}{2} \sum_{d_\xi > R'} A_\xi \exp\left\{-\frac{y}{2(d_\xi - R')}\right\}, \qquad (17.7.7)$$

где

$$A_\xi = \frac{(d_\xi - R')^{\frac{1}{2}(n-5)}}{\prod\limits_{j=1}^{(n-1)/2} (d_\xi - d_j)}. \qquad (17.7.8)$$

Так как $P(R > R') = P(y > 0)$, то

$$P(R > R') = \int_0^\infty g(y)\, dy = \sum_{d_\xi > R'} A_\xi (d_\xi - R'), \qquad (17.7.9)$$

откуда можно вычислить критические значения R', отвечающие заданным уровням значимости. Р. Андерсон распространил этот метод на случай четных n. Он составил таблицу значений R_α, где $P(R > R_\alpha) = \alpha$, для $\alpha = 0{,}99$; $0{,}95$; $0{,}05$; $0{,}01$ и $n = 5\,(1)\,15\,(5)\,75$. Дальнейшее изучение задач сериальной корреляции проводилось Андерсоном (1948), Диксоном (1944), Дёрбином и Ватсоном (1951), Хэннаном (1955), Купмэнсом (1942), Мораном (1948), Огаварой (1951), Куэнуилом (1948), Уиттлом (1951) и другими.

17.8. Линейное прогнозирование стационарных процессов

Предположим, что нам известны значения процесса $\{x_t\}$ в моменты $t = \ldots, -2, -1, 0$, и мы хотим предсказать x_1 с помощью линейной комбинации значений $\ldots, x_{-2}, x_{-1}, x_0$, воспользовавшись методом наименьших квадратов. Возникает вопрос: при каких условиях на процесс среднее значение квадрата ошибки предиктора положительно и при каких условиях оно равно нулю? Ответ на этот вопрос был дан Колмогоровым (1941) и Винером (1949). Мы рассмотрим эту задачу вкратце, пользуясь элементарными средствами анализа.

Обозначим через $\hat{x}_{1,n}$ предиктор для x_1, построенный по $x_{-n}, \ldots$ $\ldots, x_{-1}, x_0$. Тогда

$$\hat{x}_{1,n} = \sum_{\xi=-n}^{0} \hat{a}_{\xi n} x_{\xi}, \tag{17.8.1}$$

где коэффициенты $\hat{a}_{\xi n}$, $\xi = -n, \ldots, -1, 0$, минимизируют относительно $a_{\xi n}$ выражение

$$\mathscr{E}\left(x_1 - \sum_{\xi=-n}^{0} a_{\xi n} x_{\xi}\right)^2. \tag{17.8.2}$$

Согласно § 3.8 находим, что

$$\hat{a}_{\xi n} = \sum_{\xi'=-n}^{0} \sigma^{\xi\xi'} \gamma_{\xi'-1}, \tag{17.8.3}$$

где $\|\sigma^{\xi\xi'}\| = \|\sigma_{\xi\xi'}\|^{-1}$ и $\|\sigma_{\xi\xi'}\|$ — матрица размера $(n+1) \times (n+1)$

$$\begin{Vmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_n \\ \gamma_1 & \gamma_0 & \cdots & \gamma_n \\ \cdot & \cdot & \cdot & \cdot \\ \gamma_n & \gamma_{n-1} & \cdots & \gamma_0 \end{Vmatrix}, \tag{17.8.4}$$

образованная числами

$$\gamma_h = \gamma_{-h} = \int_{-\pi}^{\pi} \cos h\omega \, f(\omega) \, d\omega. \tag{17.8.5}$$

Если обозначить через Δ_{n+1} определитель матрицы $\|\sigma_{\xi\xi'}\|$, то средний квадрат ошибки предиктора $\hat{x}_{1,n}$ запишется в виде

$$\mathscr{E}(x_1 - \hat{x}_{1,n})^2 = \frac{\Delta_{n+2}}{\Delta_{n+1}}. \tag{17.8.6}$$

Для $\xi = 1, \ldots, n$ имеем

$$\frac{\Delta_{n+2+\xi}}{\Delta_{n+1+\xi}} \leqslant \frac{\Delta_{n+2}}{\Delta_{n+1}} \leqslant \frac{\Delta_{n+2-\xi}}{\Delta_{n+1-\xi}} \tag{17.8.7}$$

и, следовательно,

$$\prod_{\xi=1}^{n} \frac{\Delta_{n+2+\xi}}{\Delta_{n+1+\xi}} \leqslant \left(\frac{\Delta_{n+2}}{\Delta_{n+1}}\right)^n \leqslant \prod_{\xi=1}^{n} \frac{\Delta_{n+2-\xi}}{\Delta_{n+1-\xi}}, \tag{17.8.8}$$

откуда

$$\left(\frac{\Delta_{2n+2}}{\Delta_{n+2}}\right)^{1/n} \leqslant \frac{\Delta_{n+2}}{\Delta_{n+1}} \leqslant \left(\frac{\Delta_{n+1}}{\Delta_1}\right)^{1/n}. \tag{17.8.9}$$

Переходя к логарифмам и полагая $\Delta_{n+2}/\Delta_{n+1}=\sigma_n^2$, получим

$$A_n \leqslant \log \sigma_n^2 \leqslant B_n, \tag{17.8.10}$$

где

$$\begin{gathered} A_n=\frac{1}{n}\sum_{\xi=1}^{2n+2}\log\theta_{\xi,\,2n+2}-\frac{1}{n}\sum_{\xi=1}^{n+2}\log\theta_{\xi,\,n+2}, \\ B_n=\frac{1}{n}\sum_{\xi=1}^{n+1}\log\theta_{\xi,\,n+1}-\frac{1}{n}\log\gamma_0, \end{gathered} \tag{17.8.11}$$

а $\theta_{1,\,k},\ldots,\ \theta_{k,\,k}$ — собственные значения матрицы, получающейся из (17.8.4) заменой n на $k-1$. Поскольку при всех n матрицы (17.8.4) положительно определены и диагональные элементы их равны γ_0, все собственные значения таких матриц лежат на интервале $(0,\ \gamma_0)$ и, следовательно, логарифмы собственных значений лежат на $(-\infty,\ \log\gamma_0)$.

Если

$$\lim_{n\to\infty}\frac{1}{n}\sum_{\xi=1}^{n+1}\log\theta_{\xi,n+1}=K,$$

то, очевидно,

$$\lim_{n\to\infty}A_n=\lim_{n\to\infty}B_n=K.$$

Полагая

$$\sigma^2=\lim_{n\to\infty}\mathscr{E}(x_1-\hat{x}_{1,n})^2, \tag{17.8.12}$$

будем иметь

$$\sigma^2=e^K. \tag{17.8.13}$$

Если $K=-\infty$, то $\sigma^2=0$, т. е. средний квадрат ошибки $\hat{x}_{1,n}$ стремится к нулю при $n\to\infty$.

Используя более мощные средства анализа, можно показать, что если $K>-\infty$, то

$$K=\log 2\pi+\frac{1}{2\pi}\int_{-\pi}^{+\pi}\log f(\omega)\,d\omega. \tag{17.8.14}$$

Доказательство можно найти у Колмогорова (1941) и Винера (1949), а также в книгах Дуба (1953) и Гренандера и Розенблатта (1957).

Другая задача, связанная с линейным прогнозированием, состоит в анализе процессов *авторегрессии*. Говорят, что $\{x_t\}$ — процесс авторегрессии r-го порядка, если при некоторых $a_1,\ldots,\ a_r$

$$y_t=x_t-a_1x_{t-1}-\ldots-a_rx_{t-r} \tag{17.8.15}$$

является белым шумом. Оценка наименьших квадратов $a_1,\ldots,a_r$ по выборке $(x_1,\ldots,x_n)$ может быть найдена обычным образом. Диксон (1944) рассмотрел для нормального белого шума $\{y_t\}$ задачу проверки гипотезы о равенстве всех $a_1, \ldots, a_r$ нулю против альтернативы, согласно которой $a_m = a_{m+1} = \ldots = a_r = 0$, используя метод отношения правдоподобия. Авторегрессионные схемы изучались Манном и Вальдом (1943). Вальд (1938) рассмотрел задачу проверки гипотезы о том, что процесс $\{x_t\}$ имеет вид $y_t + a_1 y_{t-1} + \ldots + a_r y_{t-r}$, где $\{y_t\}$ — белый шум.

ЗАДАЧИ

17.1. Пусть x_t, $t = \ldots, -1, 0, +1, \ldots,$ — стационарный процесс, спектральная плотность которого равна

$$f(\omega) = \frac{1}{\pi^2}(\pi - |\omega|), \quad -\pi \leqslant \omega \leqslant +\pi.$$

Показать, что функция ковариаций имеет вид

$$\gamma_h = \begin{cases} 1, & h = 0, \\ \left(\dfrac{2}{\pi h}\right)^2, & \text{если } h \text{ нечетно}, \\ 0, & \text{если } h \text{ четно}. \end{cases}$$

17.2 (*Теорема Слуцкого* (1937)). Пусть x_t, $t = \ldots, -1, 0, +1, \ldots,$ — вещественный стационарный процесс со спектральной плотностью $f(\omega)$. Доказать, что сглаженный процесс

$$y_t = \sum_{p=0}^{k} a_p x_{t-k+p}, \qquad t = \ldots, -1, 0, +1, \ldots,$$

где $a_0, a_1, \ldots, a_k$ — вещественные постоянные, также стационарный и его спектральная плотность равна

$$\left[\sum_{q=0}^{k} \sum_{p=0}^{k} a_p a_q \cos(p-q)\,\omega\right] f(\omega).$$

В частности, покажите, что если $a_p = 1/(k+1)$, $p = 0, 1, \ldots, k$, т. е. сглаженный процесс представляет собой среднее из $(k+1)$ элементов, то $\{y_t\}$ имеет спектральную плотность

$$\frac{1 - \cos(k+1)\,\omega}{(k+1)^2\,(1 - \cos\omega)}\, f(\omega).$$

17.3. (*Продолжение*). Применим к стационарному процессу x_t, $t = \ldots, -1, 0, +1, \ldots,$ операцию сглаживания, определенную в предыдущей задаче, r раз. Доказать, что при этом получится стационарный процесс со спектральной плотностью

$$\left[\sum_{q=0}^{k} \sum_{p=0}^{k} a_p a_q \cos(p-q)\,\omega\right]^r f(\omega).$$

17.4. Пусть $\{x_t\}$ — белый шум с дисперсией 1. Доказать, что функция ковариаций процесса, образованного n-ми правыми разностями x_t, равна

$$\gamma_h = \begin{cases} (-1)^h \dfrac{(2n)!}{(n-h)!\,(n+h)!}, & -n \leqslant h \leqslant n, \\ 0 & \text{в противном случае.} \end{cases}$$

17.5. Предположим, что x_t, $t = \ldots, -1, 0, +1, \ldots,$ — линейный процесс, т. е.

$$x_t = \sum_{p=-\infty}^{+\infty} a_{t-p}\varepsilon_p,$$

где $\ldots, a_{-1}, a_0, a_1, \ldots$ — вещественные числа, $\sum\limits_{p=-\infty}^{+\infty} a_p^2 < \infty$, а $\ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots$ — независимые случайные величины с нулевыми средними и единичными дисперсиями.

Показать, что спектральная плотность процесса $\{x_t\}$ есть

$$f(\omega) = \frac{1}{2\pi} \sum_{q=-\infty}^{+\infty} \sum_{p=-\infty}^{+\infty} a_p a_{p+q} \cos q\omega.$$

17.6. Доказать, что для стационарного процесса $\{x_t\}$ с $\mathscr{E}x_t = 0$ спектральная функция $F(\omega) = \lim\limits_{M\to\infty} F_M^*(\omega)$, где

$$F_M^*(\omega) = \frac{1}{\pi M} \int_{-\pi}^{\omega} \mathscr{E}(x_1 \sin\omega + \ldots + x_M \sin M\omega)^2 \, d\omega,$$

а также $F(\omega) = \lim\limits_{M\to\infty} F_M^{**}(\omega)$, где $I_M(\omega)$ — периодограмма, $I_M(\omega) = \frac{1}{2\pi M} \times$ $\times \left| \sum\limits_{p=1}^{M} x_p e^{-ip\omega} \right|^2$,

$$F_M^{**}(\omega) = \int_{-\pi}^{\omega} \mathscr{E} I_M(\omega) \, d\omega.$$

17.7. Доказать, что функция ковариаций

$$\gamma_h = \mathscr{E}(x_{t+h} \bar{x}_t), \qquad h = \ldots, -1, 0, +1, \ldots,$$

комплексного стационарного процесса дается формулой (17.3.11), где $F(\omega) = \lim\limits_{M\to\infty} F_M(\omega)$ и $F_M(\omega)$ определена в (17.3.12).

17.8. Пусть x_t, $t = \ldots, -1, 0, +1, \ldots,$ — стационарный процесс с функцией ковариаций

$$\gamma_h = \gamma_0 \rho^h, \quad 0 < \rho < +1, \quad h = 0, 1, \ldots$$

Доказать, что средний квадрат ошибки линейного прогнозирования x_1 по $x_{-n}, x_{-n+1}, \ldots, x_{-1}, x_0$ методом наименьших квадратов равен $\sigma_n^2 = \gamma_0(1 - \rho^2)$.

17.9. (*Продолжение*). Покажите, что спектральная плотность процесса из задачи 17.8 равна

$$f(\omega) = \frac{\gamma_0 (1 - \rho\cos\omega)}{\pi(1 + \rho^2 - 2\rho\cos\omega)}.$$

17.10. Доказать формулу (17.6.33) методами, аналогичными использованным при выводе (17.6.32).

ГЛАВА 18

МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ

В последнюю четверть века линейный и квадратичный анализ выборок из многомерных совокупностей вполне оформились в отдельный раздел математической статистики, получивший название *многомерного статистического анализа.* Бо́льшая часть теории в этом круге задач относится к выборкам из нормальных совокупностей. Имеется множество важных и интересных результатов в многомерной статистической теории, рассеянных в статистической литературе. В книге такого типа, как настоящая, невозможно осветить все их. Читателю, интересующемуся дальнейшими математическими деталями, следует обратиться к книгам Андерсона (1958) и Роя (1957). Подробности, связанные с использованием этих методов для вычислений, можно найти в книгах Рао (1952) и Кендалла (1957). В книге Андерсона имеется исчерпывающая библиография по многомерному статистическому анализу.

18.1. Многомерное статистическое рассеивание

В качестве введения в многомерный статистический анализ удобно определить и рассмотреть *рассеивание* выборки относительно точки.

(а) Одномерный случай. Для начала пусть $(x_1, \ldots, x_n)$ — выборка из одномерной совокупности $F(x)$ и x_0 — произвольное вещественное число. Числа $x_1, \ldots, x_n$ и x_0 можно представлять себе точками вещественной прямой. *Рассеивание* ${}_1S_{x_0, n}$ выборки относительно *центра* x_0 определяется следующим образом:

$$
{}_1S_{x_0, n} = \sum_{\xi=1}^{n} (x_\xi - x_0)^2 = {}_1S_{\bar{x}, n} + n(\bar{x} - x_0)^2. \tag{18.1.1}
$$

Заметим, что ${}_1S_{x_0, n}$ есть просто сумма квадратов длин n отрезков, соединяющих точку x_0 со всеми точками x_ξ, принадлежащими выборке. Отметим также, что рассеивание выборки относительно точки x_0 равно ее рассеиванию относительно среднего значения выборки $\bar{x}$ плюс взвешенное рассеивание точки $\bar{x}$ относительно x_0 с весом, равным объему выборки n. Следовательно, *рассеивание выборки минимально, если оно вычислено относительно выборочного среднего.*

Мы знаем также из § 8.2, что если $F(x)$ имеет среднее значение μ и дисперсию σ^2, то

$$\mathscr{E}({}_1S_{x_0,\,n}) = n[\sigma^2 + (\mu - x_0)^2], \tag{18.1.2}$$

откуда очевидно, что *минимальное значение* $\mathscr{E}({}_1S_{x_0,\,n})$ *достигается при* $x_0 = \mu$.

В механических терминах ${}_1S_{x_0,\,n}$ есть момент инерции выборочных точек $x_1, \ldots, x_n$ относительно центра x_0.

Предположим теперь, что имеются две выборки $(x_1^{(1)}, \ldots, x_{n_1}^{(1)})$ и $(x_1^{(2)}, \ldots, x_{n_2}^{(2)})$ из одномерных совокупностей $F_1(x)$ и $F_2(x)$ соответственно. Пусть $\bar{x}^{(1)}$ и $\bar{x}^{(2)}$ — выборочные средние, а $\bar{x}$ — среднее значение *объединенной выборки* объема $n_1 + n_2$, построенной по выборкам $(x_1^{(1)}, \ldots, x_{n_1}^{(1)})$ и $(x_1^{(2)}, \ldots, x_{n_2}^{(2)})$. Пусть, далее, ${}_1S_{\bar{x}^{(1)},\,n_1}$, ${}_1S_{\bar{x}^{(2)},\,n_2}$ — рассеивания выборок относительно соответствующих выборочных средних, взятых в качестве центров, и ${}_1S_{\bar{x},\,n_1+n_2}$ — рассеивание объединенной выборки относительно центра $\bar{x}$. Читатель легко проверит, что

$${}_1S_{\bar{x},\,n_1+n_2} = {}_1S_{\bar{x}^{(1)},\,n_1} + {}_1S_{\bar{x}^{(2)},\,n_2} + n_1(\bar{x}^{(1)} - \bar{x})^2 + n_2(\bar{x}^{(2)} - \bar{x})^2. \tag{18.1.3}$$

Таким образом, при жестких сдвигах каждой выборки минимальное значение рассеивания объединенной выборки относительно $\bar{x}$ равно сумме рассеиваний отдельных выборок относительно соответствующих выборочных средних; этот минимум достигается только, когда оба выборочных средних совпадают. Результат, выражаемый равенством (18.1.3), непосредственно обобщается на любое число выборок.

(b) Двумерный случай. Рассмотрим теперь выборку $(x_{1\xi}, x_{2\xi}; \xi = 1, \ldots, n)$ объема n из двумерной совокупности $F(x_1, x_2)$. Пусть (x_{10}, x_{20}) — произвольная пара чисел. Элементы выборки вместе с парой чисел (x_{10}, x_{20}) можно рассматривать как множество (самое большее $n+1$) точек плоскости. Условимся называть это множество *выборочной гроздью*. Возьмем два любых элемента выборки $(x_{1\xi}, x_{2\xi})$ и $(x_{1\eta}, x_{2\eta})$ и точку (x_{10}, x_{20}). Эти три пары можно в общем случае рассматривать как три точки плоскости (x_1, x_2). Если эти три точки не коллинеарны, к ним можно тремя различными способами добавить четвертую точку так, чтобы полученная четверка лежала в вершинах некоторого параллелограмма. Все три возможных параллелограмма имеют одинаковые (с точностью до знака) площади $A_{\xi\eta,\,x_0}$.

Условимся называть абсолютное значение этой площади *двумерным объемом*, определяемым точками $(x_{1\xi}, x_{2\xi})$, $(x_{1\eta}, x_{2\eta})$, (x_{10}, x_{20}). Квадрат этого объема равен

$$A^2_{\xi\eta,\,x_0} = \begin{vmatrix} x_{1\xi} - x_{10} & x_{2\xi} - x_{20} \\ x_{1\eta} - x_{10} & x_{2\eta} - x_{20} \end{vmatrix}^2. \tag{18.1.4}$$

Возведя в квадрат определитель в правой части, (18.1.4) можно записать в виде

$$A^2_{\xi\eta,\,x_0}=\Delta_{\xi\xi}+\Delta_{\xi\eta}+\Delta_{\eta\xi}+\Delta_{\eta\eta}, \tag{18.1.5}$$

где

$$\Delta_{\xi\eta}=\begin{vmatrix}(x_{1\xi}-x_{10})^2 & (x_{1\eta}-x_{10})(x_{2\eta}-x_{20})\\ (x_{1\xi}-x_{10})(x_{2\xi}-x_{20}) & (x_{2\eta}-x_{20})^2\end{vmatrix}, \tag{18.1.6}$$

и аналогичные соотношения имеют место для $\Delta_{\xi\xi}$, $\Delta_{\eta\xi}$, $\Delta_{\eta\eta}$. Отметим, что $\Delta_{\xi\xi}$ и $\Delta_{\eta\eta}$ равны 0.

Пара (ξ, η) может быть любой парой чисел $1, \ldots, n$. Всего можно образовать таких пар $\binom{n}{2}$ и так как $A^2_{\xi\xi x_0}=0$, сумма значений $A^2_{\xi\eta,\,x_0}$ по всем парам (ξ, η) равна $\sum\limits_{\eta=1}^{n}\sum\limits_{\xi=1}^{n}A^2_{\xi\eta,\,x_0}$. Но эта сумма равна $\sum\limits_{\eta=1}^{n}\sum\limits_{\xi=1}^{\eta}\Delta_{\xi\eta}=\sum\limits_{\eta=1}^{n}\sum\limits_{\xi=1}^{n}\Delta_{\eta\xi}$. Обозначим последнюю сумму через ${}_2S_{x_0,n}$ и примем во внимание, что для любого множества определителей $\begin{vmatrix}a_\xi & b_\eta\\ c_\xi & d_\eta\end{vmatrix}$, $\xi, \eta=1, \ldots, n$, верно соотношение

$$\sum_\eta\sum_\xi\begin{vmatrix}a_\xi & b_\eta\\ c_\xi & d_\eta\end{vmatrix}=\begin{vmatrix}\sum\limits_\xi a_\xi & \sum\limits_\eta b_\eta\\ \sum\limits_\xi c_\xi & \sum\limits_\eta d_\eta\end{vmatrix}. \tag{18.1.7}$$

Тогда будем иметь следующий результат:

$${}_2S_{x_0,\,n}=\begin{vmatrix}\sum\limits_\xi (x_{1_\xi}-x_{10})^2 & \sum\limits_\eta (x_{1_\eta}-x_{10})(x_{2_\eta}-x_{20})\\ \sum\limits_\xi (x_{2_\xi}-x_{20})(x_{1_\xi}-x_{10}) & \sum\limits_\eta (x_{2_\eta}-x_{20})^2\end{vmatrix}, \tag{18.1.8}$$

где суммирование ведется по ξ и η от 1 до n.

Матрица, определитель которой есть ${}_2S_{x_0,\,n}$, является *матрицей Грама*, т. е. матрицей вида $A'A$, где

$$A=\begin{Vmatrix}x_{11}-x_{10} & x_{21}-x_{20}\\ \cdot\;\cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot\;\cdot\\ x_{1n}-x_{10} & x_{2n}-x_{20}\end{Vmatrix} \tag{18.1.9}$$

и A' — матрица, транспонированная по отношению к A.

Величина ${}_2S_{x_0,\,n}$ характеризует *рассеивание* выборки ($x_{1\xi}$, $x_{2\xi}$; $\xi=1, \ldots, n$) относительно центра (x_{10}, x_{20}); она равна сумме квадратов двумерных объемов, определяемых точкой (x_{10}, x_{20}) и всевозможными парами выборочных точек ($x_{1\xi}$, $x_{2\xi}$; $\xi=1, \ldots, n$). Матрица с определителем ${}_2S_{x_0,\,n}$ называется *матрицей рассеивания* выборки относительно (x_{10}, x_{20}).

Если положить $\bar{x}_i = \frac{1}{n}\sum_{\xi=1}^{n} x_{i\xi}$, $i = 1, 2$, и

$$u_{ij} = u_{ji} = \sum_{\xi=1}^{n} (x_{i\xi} - \bar{x}_i)(x_{j\xi} - \bar{x}_j), \qquad i, j = 1, 2, \quad (18.1.10)$$

то будем иметь

$$\sum_{\xi=1}^{n} (x_{i\xi} - x_{i0})(x_{j\xi} - x_{j0}) = u_{ij} + n(\bar{x}_i - x_{i0})(\bar{x}_j - x_{j0}). \quad (18.1.11)$$

Следовательно,

$${}_2S_{x_0,\,n} = \begin{vmatrix} u_{11} + n(\bar{x}_1 - x_{10})^2 & u_{12} + n(\bar{x}_1 - x_{10})(\bar{x}_2 - x_{20}) \\ u_{21} + n(\bar{x}_2 - x_{20})(\bar{x}_1 - x_{10}) & u_{22} + n(\bar{x}_2 - x_{20})^2 \end{vmatrix}, \quad (18.1.12)$$

что является двумерным аналогом (18.1.1). Если записать определитель в правой части (18.1.12) обычным образом в виде суммы четырех определителей, то выражение для ${}_2S_{x_0,\,n}$ примет вид

$${}_2S_{x_0,n} = \begin{vmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{vmatrix} \left[1 + n\sum_{i,j=1}^{2} u^{ij}(\bar{x}_i - x_{i0})(\bar{x}_j - x_{j0})\right], \quad (18.1.13)$$

где $\| u^{ij} \| = \| u_{ij} \|^{-1}$, и эта обратная матрица существует, конечно, только в том случае, когда матрица рассеивания $\| u_{ij} \|$ выборки относительно $(\bar{x}_1, \bar{x}_2)$ неособая с вероятностью 1. Матрицу $\| u_{ij} \|$ будем называть *матрицей внутреннего рассеивания*, а ее определитель $| u_{ij} |$ — *внутренним рассеиванием* выборки. Будет показано, что $\| u_{ij} \|$ — неособая матрица в том и только в том случае, когда n выборочных точек $(x_{1\xi}, x_{2\xi};\ \xi = 1, \ldots, n)$ неколлинеарны. Если матрица $\| u_{ij} \|$ неособая, то квадратичная форма, стоящая внутри квадратных скобок в правой части (18.1.13), положительно определена и мы имеем следующий результат.

18.1.1. *Рассеивание выборки относительно центра (x_{10}, x_{20}) минимально в том случае, если в качестве центра взято выборочное среднее $(\bar{x}_1, \bar{x}_2)$. Это минимальное рассеивание — внутреннее рассеивание выборки — равно*

$${}_2S_{\bar{x},\,n} = \begin{vmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{vmatrix}. \quad (18.1.14)$$

Покажем теперь, что

18.1.2. *Среднее значение двумерного рассеивания относительно (x_{10}, x_{20}) равно*

$$\mathfrak{E}({}_2S_{x_0,\,n}) = 2\binom{n}{2} \begin{vmatrix} \sigma_{11} + (\mu_1 - x_{10})^2 & \sigma_{12} + (\mu_1 - x_{10})(\mu_2 - x_{20}) \\ \sigma_{21} + (\mu_2 - x_{20})(\mu_1 - x_{10}) & \sigma_{22} + (\mu_2 - x_{20})^2 \end{vmatrix}, \quad (18.1.15)$$

где $\|\sigma_{ij}\|$ — *матрица ковариаций и* (μ_1, μ_2) — *вектор средних распределения* $F(x_1, x_2)$. *Минимум* $\mathfrak{E}({}_2S_{x_0,\,n})$ *достигается тогда и только тогда, когда* $(x_{10}, x_{20}) = (\mu_1, \mu_2)$; *в этом случае* $\min \mathfrak{E}({}_2S_{x_0,\,n}) = \mathfrak{E}({}_2S_{\mu,\,n}) = 2\binom{n}{2}|\sigma_{ij}|$.

Формула (18.1.15) — двумерный аналог (18.1.2). Чтобы установить ее, напишем выражение для $\mathfrak{E}\Delta_{\xi\eta}$:

$$\mathfrak{E}(\Delta_{\xi\eta}) = \int_{R_2}\int_{R_2} \Delta_{\xi\eta} dF(x_{1_\xi}, x_{2_\xi})\, dF(x_{1_\eta}, x_{2_\eta}). \qquad (18.1.16)$$

Заметив, что равенство (18.1.7) справедливо для интегралов так же, как для конечных сумм, можно правую часть (18.1.6) записать в виде определителя $|D_{ij}|$, где

$$D_{21} = \int_{R_2} (x_{1_\xi} - x_{10})(x_{2_\xi} - x_{20})\, dF(x_{1_\xi}, x_{2_\xi})$$

и D_{11}, D_{12}, D_{22} имеют аналогичный вид. Отсюда

$$D_{ij} = \mathfrak{E}[(x_i - x_{i0})(x_j - x_{j0})] = \sigma_{ij} + (\mu_i - x_{i0})(\mu_j - x_{j0}).$$

Следовательно, для $\xi \neq \eta = 1, \ldots, n$

$$\mathfrak{E}(\Delta_{\xi\eta}) = \begin{vmatrix} \sigma_{11} + (\mu_1 - x_{10})^2 & \sigma_{12} + (\mu_1 - x_{10})(\mu_2 - x_{20}) \\ \sigma_{21} + (\mu_2 - x_{20})(\mu_1 - x_{10}) & \sigma_{22} + (\mu_2 - x_{20})^2 \end{vmatrix}.$$

Но

$$\mathfrak{E}(A^2_{\xi\eta,\,x_0}) = \mathfrak{E}(\Delta_{\xi\xi}) + \mathfrak{E}(\Delta_{\xi\eta}) + \mathfrak{E}(\Delta_{\eta\xi}) + \mathfrak{E}(\Delta_{\eta\eta}) = 2\mathfrak{E}(\Delta_{\xi\eta}),$$

а ${}_2S_{x_0,\,n}$ является суммой величин $A^2_{\xi\eta,\,x_0}$ в количестве $\binom{n}{2}$. Поэтому

$$\mathfrak{E}({}_2S_{x_0,\,n}) = 2\binom{n}{2}\mathfrak{E}\Delta_{\xi\eta},$$

что эквивалентно (18.1.15). Доказательство утверждения **18.1.2** закончено.

Отметим, что если матрица $\|\sigma_{ij}\|$ положительно определена, то обратная матрица $\|\sigma_{ij}\|^{-1} = \|\sigma^{ij}\|$ существует и также положительно определена; в этом случае

$$\mathfrak{E}({}_2S_{x_0,\,n}) = 2\binom{n}{2}|\sigma_{ij}|\left[1 + \sum_{i,\,j=1}^{2} \sigma^{ij}(\mu_i - x_{i0})(\mu_j - x_{j0})\right]. \quad (18.1.17)$$

Из (18.1.17) очевидным образом следует

18.1.3. *Если* $\|\sigma_{ij}\|$ *положительно определена, то* $\mathfrak{E}({}_2S_{x_0,\,n})$ *достигает минимума при выборе в качестве центра вектора средних* (μ_1, μ_2) *распределения* $F(x_1, x_2)$. *Это минимальное значение равно*

$$\mathfrak{E}({}_2S_{\mu,\,n}) = E|v_{ij}| = 2\binom{n}{2}|\sigma_{ij}|, \qquad (18.1.18)$$

где

$$v_{ij}=v_{ji}=\sum_{\xi=1}^{n}(x_{i_\xi}-\mu_i)(x_{j_\xi}-\mu_j), \qquad i,\ j=1,\ 2. \quad (18.1.19)$$

Предположим теперь, что $(x_{1\xi_1}^{(1)},\ x_{2\xi_1}^{(1)};\ \xi_1=1,\ \ldots,\ n_1)$ и $(x_{1\xi_2}^{(2)},\ x_{2\xi_2}^{(2)};\ \xi_2=1,\ \ldots,\ n_2)$ — выборки из совокупностей $F_1(x_1,\ x_2)$ и $F_2(x_1,\ x_2)$ соответственно. Пусть $(\bar{x}_1^{(1)},\ \bar{x}_2^{(1)})$ и $(\bar{x}_1^{(2)},\ \bar{x}_2^{(2)})$ — средние, а $\|u_{ij}^{(1)}\|$ и $\|u_{ij}^{(2)}\|$, $i,\ j=1,\ 2,$ — матрицы внутреннего рассеивания этих выборок.

Если $(\bar{x}_1,\ \bar{x}_2)$ — вектор средних объединенной выборки, то можно непосредственно убедиться в том, что рассеивание объединенной выборки относительно $(\bar{x}_1,\ \bar{x}_2)$ равно

$$ {}_2S_{\bar{x},\ n_1+n_2}=|\,u_{ij}^{(1)}+u_{ij}^{(2)}+n_1(\bar{x}_i^{(1)}-\bar{x}_i)(\bar{x}_j^{(1)}-\bar{x}_j)+ \\ +n_2(\bar{x}_i^{(2)}-\bar{x}_i)(\bar{x}_j^{(2)}-\bar{x}_j)\,|. \quad (18.1.20)$$

Определитель второго порядка в правой части (18.1.20) является двумерным аналогом правой части (18.1.3).

(c) k-мерный случай. Никаких трудностей нового порядка не встречается при определении рассеивания ${}_kS_{x_0,\ n}$ относительно центра $(x_{10},\ \ldots,\ x_{k0})$ выборки $(x_{1_\xi},\ \ldots,\ x_{k_\xi};\ \xi=1,\ \ldots,\ n)$ из k-мерной совокупности $F(x_1,\ \ldots,\ x_k)$.

Выборку вместе с точкой $(x_{10},\ \ldots,\ x_{k0})$ можно рассматривать как совокупность $n+1$ точек в эвклидовом пространстве R_k, которую мы назовем *выборочной гроздью.* Любой набор k различных выборочных точек $(x_{1\xi_i},\ \ldots,\ x_{k\xi_i})$, $i=1,\ \ldots,\ k,$ вместе с точкой $(x_{10},\ \ldots,\ x_{k0})$ предоставляют $k+1$ способов выбора еще одной точки так, чтобы эта точка вместе с уже отмеченными $k+1$ точками образовала k-мерный параллелепипед. Все эти различные параллелепипеды имеют одинаковые (с точностью до знака) k-мерные объемы, абсолютные значения которых назовем *k-мерным объемом,* определяемым точками $(x_{1\xi_i},\ \ldots,\ x_{k\xi_i})$, $i=1,\ \ldots,\ k$ и $(x_{10},\ \ldots,\ x_{k0})$. *Рассеивание* ${}_kS_{x_0,\ n}$ определяется теперь как сумма квадратов k-мерных объемов, определяемых точкой $(x_{10},\ \ldots,\ x_{k0})$ и каждым из $\binom{n}{k}$ различных наборов точек $(x_{1\xi_i},\ \ldots,\ x_{k\xi_i})$, $i=1,\ \ldots,\ k.$ Непосредственно устанавливается, что ${}_kS_{x_0,\ n}$ является обобщением величины ${}_2S_{x_0,\ n}$, задаваемой формулой (18.1.11), на случай k измерений, т. е. что

$${}_kS_{x_0,\ n}=|\,u_{ij}+n(\bar{x}_i-x_{i0})(\bar{x}_j-x_{j0})\,|, \qquad (18.1.21)$$

где определение u_{ij}, $i,\ j=1,\ \ldots,\ k,$ понятно из (18.1.10).

Если матрица $\|u_{ij}\|$ внутреннего рассеивания выборки положительно определена, — а это будет тогда и только тогда, когда n выборочных точек $(x_{1_\xi},\ \ldots,\ x_{k_\xi};\ \xi=1,\ \ldots,\ n)$ не лежат в гиперплоскости размерности, меньшей k, — то ${}_kS_{x_0,\,n}$ может быть записано в иной форме:

$$ {}_kS_{x_0,\,n} = |u_{ij}| \left[1 + n \sum_{i,\,j=1}^{k} u^{ij}(\bar{x}_i - x_{i0})(\bar{x}_j - x_{j0})\right], \qquad (18.1.22) $$

где, как и прежде, $\|u^{ij}\| = \|u_{ij}\|^{-1}$ и эта матрица также положительно определена. Отметим, что (18.1.22) является k-мерным обобщением (18.1.13). Так как квадратичная форма, стоящая внутри квадратных скобок, положительно определена, непосредственно получаем k-мерный вариант утверждения **18.1.1**: ${}_kS_{x_0,\,n}$ достигает минимального значения, когда в качестве $(x_{10}, \ldots, x_{k0})$ выбран вектор выборочных средних $(\bar{x}_1, \ldots, \bar{x}_k)$, и это минимальное значение равно

$$ {}_kS_{\bar{x},\,n} = |u_{ij}|, \qquad i,\ j = 1, \ldots, k, \qquad (18.1.23) $$

что представляет собой внутреннее рассеивание выборки.

Заметим, что ${}_kS_{\bar{x},\,n}$ — определитель матрицы Грама $A'A$, где A — матрица $(n \times k)$

$$ \left\| \begin{matrix} x_{11} - x_{10}, & \ldots, & x_{k1} - x_{k0} \\ \cdot & \cdot \; \cdot & \cdot \\ x_{1n} - x_{10}, & \ldots, & x_{kn} - x_{k0} \end{matrix} \right\| $$

и A' — матрица, транспонированная относительно A. Матрица $\|u_{ij}\|$ также является *матрицей Грама* $B'B$, где B получается из A заменой x_{i0} на $\bar{x}_i$, $i = 1, \ldots, k$.

Если $(x_{1\xi}, \ldots, x_{k\xi};\ \xi = 1, \ldots, n)$ — выборка из k-мерной совокупности с вектором средних $(\mu_1, \ldots, \mu_k)$ и (неособой) матрицей ковариаций $\|\sigma_{ij}\|$, то с помощью рассуждений, аналогичных использованным при выводе (18.1.15), можно получить, что

$$ \mathfrak{E}({}_kS_{x_0,\,n}) = k!\binom{n}{k} |\sigma_{ij} + (\mu_i - x_{i0})(\mu_j - x_{j0})|. \qquad (18.1.24) $$

Это может быть также записано в виде

$$ \mathfrak{E}({}_kS_{x_0,\,n}) = k!\binom{n}{k} |\sigma_{ij}| \left[1 + \sum_{i,\,j=1}^{k} \sigma^{ij}(\mu_i - x_{i0})(\mu_j - x_{j0})\right]. \qquad (18.1.25) $$

Последняя формула представляет собой k-мерный аналог (18.1.7); из (18.1.25) сразу следует k-мерный вариант утверждения **18.1.2**: $\mathfrak{E}({}_kS_{x_0,\,n})$ достигает минимального значения в том и только том случае, если в качестве $(x_{10}, \ldots, x_{k0})$ выбран вектор средних $(\mu_1, \ldots, \mu_k)$. Минимальное значение равно

$$ \mathfrak{E}({}_kS_{\mu,\,n}) = \mathfrak{E}\,|v_{ij}| = k!\binom{n}{k} |\sigma_{ij}|, \qquad (18.1.26) $$

где

$$ v_{ij} = v_{ji} = \sum_{\xi=1}^{n} (x_{i\xi} - \mu_i)(x_{j\xi} - \mu_j). \qquad (18.1.27) $$

Определитель $|\sigma_{ij}|$ иногда называют *обобщенной дисперсией* распределения $F(x_1, \ldots, x_k)$. Его можно также рассматривать как *внутреннее рассеивание* этого распределения.

Читателю, разобравшемуся в выводе формулы (18.1.20), будет очень легко установить ее k-мерный вариант, именно: рассеивание ${}_kS_{\bar{x}, n_1+n_2}$ объединенной выборки относительно среднего этой выборки дается формулой (18.1.20), где $i, j = 1, \ldots, k$. Подробное исследование многомерного статистического рассеивания можно найти у Уилкса (1960а).

18.2. Распределение Уишарта

(а) Вывод распределения Уишарта. В предыдущем параграфе мы видели, что вычисление среднего значения рассеивания выборки из k-мерной совокупности является сравнительно легким делом для произвольной совокупности с конечным вектором средних и матрицей ковариаций. Однако вычисление старших моментов рассеивания выборки из таких совокупностей, по-видимому, весьма трудно, за исключением случая k-мерной нормальной совокупности. На самом деле можно не только вычислить моменты рассеивания, но и получить ф. р. элементов матрицы рассеивания выборки из k-мерной нормальной совокупности, если в качестве центра взять вектор средних этой нормальной совокупности. Вывод распределения элементов матрицы рассеивания в этом случае, известного как *распределение Уишарта* (1928), является замечательным результатом, играющим фундаментальную роль в далеко продвинутой теории многомерных статистических выводов. Мы приведем здесь вывод распределения Уишарта, основанный на использовании характеристических функций. Он действует для любой точки пространства $\{v_{ij}\}$, в которой матрица $\|v_{ij}\|$ положительно определена, в предположении, что $\|\sigma_{ij}\|$ положительно определена и $k \leqslant n$.

Пусть $(x_{1\xi}, \ldots, x_{k\xi};\ \xi = 1, \ldots, n)$ — выборка из k-мерного нормального распределения $N(\{\mu_i\}, \|\sigma_{ij}\|)$, плотность которого дается формулой (7.4.1). Выборка объема n из k-мерного нормального распределения определена в § 8.1. Рассмотрим рассеивание ${}_kS_{\mu, n}$ этой выборки относительно среднего $(\mu_1, \ldots, \mu_k)$. Имеем

$$ {}_kS_{\mu, n} = |v_{ij}|, \tag{18.2.1}$$

где

$$v_{ij} = v_{ji} = \sum_{\xi=1}^{n} (x_{i\xi} - \mu_i)(x_{j\xi} - \mu_j), \qquad i, j = 1, \ldots, k. \tag{18.2.2}$$

Если мы обозначим через $\varphi(\{t_{ij}\})$, $t_{ij} \equiv t_{ji}$, характеристическую функцию случайного вектора $\{v_{ii},\ i = 1, \ldots, k;\ 2v_{ij},\ i > j = 1, \ldots, k\}$, то

$$\varphi(\{t_{ij}\}) = \left[\frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{k/2}}\right]^n \int\limits_{R_{nk}} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k} \sigma^{ij}v_{ij} + \mathrm{i}\sum_{i,j=1}^{k} v_{ij}t_{ij}\right] \prod_{i,\xi} dx_{i,\xi}. \tag{18.2.3}$$

Так как $\varphi(0)=1$, то мы видим, что интеграл в (18.2.3) равен $[(2\pi)^{k/2}/\sqrt{|\sigma^{ij}-2it_{ij}|}]^n$. Следовательно,

$$\varphi(\{t_{ij}\})=|\tau_{ij}|^{\frac{1}{2}n}\,|\tau_{ij}-it_{ij}|^{-\frac{1}{2}n}, \tag{18.2.4}$$

где

$$\tau_{ij}=\frac{1}{2}\sigma^{ij}, \qquad i,\ j=1,\ \ldots,\ k. \tag{18.2.5}$$

Используя многомерную форму теоремы Леви, именно теорему **5.2.1**, п. в. вектора $\{v_{ii},\ i=1,\ \ldots,\ k,\ 2v_{ij},\ i>j=1,\ \ldots,\ k\}$ можно представить в виде

$$f(\{v_{ii},\ 2v_{ij}\})=\left(\frac{1}{2\pi}\right)^{\frac{1}{2}k(k+1)}\times$$
$$\times\int\limits_{R_{\frac{1}{2}k(k+1)}}\exp\Big[-i\sum_{i,j=1}^{k}v_{ij}t_{ij}\Big]\varphi(t_{ij})\prod_{i\geq j=1}^{k}dt_{ij}. \tag{18.2.6}$$

Положим $\tau_{ij}-it_{ij}=\theta_{ij}$. Тогда $\theta_{ij}=\theta_{ji}$, $-it_{ij}=\theta_{ij}-\tau_{ij}$, $dt_{ij}=$ $=id\theta_{ij}$ и

$$f(\{v_{ii},\ 2v_{ij}\})=A\int\limits_{\substack{\tau_{ij}-i\infty\\ i\geq j=1,\ldots,k}}^{\tau_{ij}+i\infty}\exp\Big[\sum_{i,j=1}^{k}v_{ij}\theta_{ij}\Big]|\theta_{ij}|^{-\frac{1}{2}n}\prod_{i\geq j=1}^{k}d\theta_{ij}, \tag{18.2.7}$$

где

$$A=\frac{|\tau_{ij}|^{\frac{1}{2}n}\exp\Big[-\sum\limits_{i,j=1}^{k}\tau_{ij}v_{ij}\Big]}{(2\pi i)^{\frac{1}{2}k(k+1)}}. \tag{18.2.8}$$

Будем проводить интегрирование последовательно в таком порядке. Сначала интегрируем по θ_{kk}, затем по $\theta_{1k},\ \ldots,\ \theta_{k-1,k}$, далее интегрируем по $\theta_{k-1,k-1}$ и затем по $\theta_{1,k-1},\ \ldots,\ \theta_{k-2,k-1}$ и т. д. Можно записать (18.2.7) в виде

$$f(\{v_{ii},\ 2v_{ij}\})=AH_1\int\limits_{\substack{\tau_{ik}-i\infty\\ i=1,\ldots,k-1}}^{\tau_{ik}+i\infty}\exp\Big[2\sum_{i=1}^{k-1}v_{ik}\theta_{ik}\Big]H_2\prod_{i=1}^{k-1}d\theta_{ik}, \tag{18.2.9}$$

где

$$H_1=\int\limits_{\tau_{ij}-i\infty}^{\tau_{ij}+i\infty}\exp\Big[\sum_{i,j=1}^{k-1}v_{ij}\theta_{ij}\Big]|\theta_{ij}|_{(k-1)}^{-\frac{1}{2}n}\prod_{i\geq j=1}^{k-1}d\theta_{ij},$$
$$H_2=\int\limits_{\tau_{kk}-i\infty}^{\tau_{kk}+i\infty}\exp(v_{kk}\theta_{kk})\Big[\theta_{kk}-\sum_{i,j=1}^{k-1}\theta_{(k-1)}^{ij}\theta_{ik}\theta_{jk}\Big]^{-\frac{1}{2}n}d\theta_{kk} \tag{18.2.10}$$

и

$$\| \theta^{ij}_{(k-1)} \| = \| \theta_{ij} \|^{-1}_{(k-1)}, \quad \| \theta_{ij} \|_{(k-1)} = \left\| \begin{matrix} \theta_{11} & \dots & \theta_{1,k-1} \\ \theta_{k-1,1} & \dots & \theta_{k-1,k-1} \end{matrix} \right\|.$$

Положим теперь в выражении для H_2

$$\theta_{kk} - \sum_{i,\ j=1}^{k-1} \theta^{ij}_{(k-1)} \theta_{ik} \theta_{jk} = w. \tag{18.2.11}$$

Тогда

$$H_2 = \exp\left[v_{kk} \sum_{i,\ j=1}^{k-1} \theta^{ij}_{(k-1)} \theta_{ik} \theta_{jk}\right] \int\limits_{c-i\infty}^{c+i\infty} e^{v_{kk} w} w^{-\frac{1}{2} n}\, dw, \tag{18.2.12}$$

где c — вещественное положительное число (точное значение которого нам не требуется). Интеграл в (18.2.12) есть, по существу, интеграл Ганкеля (см. пример в § 5.1), и его значение I_1 равно

$$I_1 = \frac{(v_{kk})^{\frac{1}{2} n - 1}\, 2\pi i}{\Gamma\left(\frac{1}{2} n\right)}. \tag{18.2.13}$$

Подставляя значение H_2 в (18.2.9), получим

$$f(\{v_{ii},\ 2v_{ij}\}) = A \cdot H_1 \cdot I_1 \cdot G, \tag{18.2.14}$$

где

$$G = \int\limits_{\substack{\tau_{ik} - i\infty \\ i=1, \dots, k-1}}^{\tau_{ik} + i\infty} \exp\left[v_{kk} \sum_{i,\ j=1}^{k-1} \theta^{ij}_{(k-1)} \theta_{ik} \theta_{jk} + \sum_{i=1}^{k-1} v_{ik} \theta_{ik}\right] \prod_{i=1}^{k-1} d\theta_{ik}. \tag{18.2.15}$$

После подстановки $\theta_{ik} = i\psi_{ik}$, $i = 1, \dots, k-1$, G запишется в виде

$$G = i^{k-1} \int\limits_{\substack{-\infty + i\tau_{ik} \\ i=1, \dots, k-1}}^{+\infty + i\tau_{ik}} \exp\left[-v_{kk} \sum_{i,\ j=1}^{k-1} \theta^{ij}_{(k-1)} \psi_{ik} \psi_{jk} + 2i \sum_{i=1}^{k-1} v_{ik} \psi_{ik}\right] \prod_{i=1}^{k-1} d\psi_{ik}. \tag{18.2.16}$$

Мы видим, что интеграл в (18.2.16) берется от функции комплексных переменных по $(k-1)$-мерному пространству вещественных значений этих комплексных переменных при фиксированных значениях их мнимых частей. Можно показать, что значение интеграла не зависит от того, каковы фиксированные значения (конечные) этих мнимых частей. Следовательно, мнимые части можно считать равными нулю. Поэтому интеграл имеет то же значение, как если бы $\psi_{1k}, \dots, \psi_{k-1,k}$ были вещественными и интегрирование велось по $(k-1)$-мерному пространству переменных $\psi_{1k}, \dots, \psi_{k-1,k}$. Обращаясь к (7.4.13) и (7.4.17), мы можем получить значение этого интеграла:

$$G = J_1 \left| \theta^{ij}_{(k-1)} \right|^{-\frac{1}{2}} \exp\left[-\sum_{i,\ j=1}^{k-1} \theta_{ij} \frac{v_{ik} v_{jk}}{v_{kk}}\right], \tag{18.2.17}$$

где

$$J_1 = \frac{\pi^{\frac{1}{2}(k-1)} \mathrm{i}^{k-1}}{(v_{kk})^{\frac{1}{2}(k-1)}}. \tag{18.2.18}$$

Подставляя значение G в (18.2.14), будем иметь

$$f(\{v_{ii}, 2v_{ij}\}) = \\ = A(I_1 \cdot J_1) \int\limits_{\substack{\tau_{ij}-\mathrm{i}\infty \\ i \geq j = 1, \ldots, k-1}}^{\tau_{ij}+\mathrm{i}\infty} \exp\left[\sum_{i,j=1}^{k-1} v_{ij}^{(1)}\theta_{ij}\right] \left|\theta_{ij}\right|_{(k-1)}^{\frac{1}{2}(n-1)} \prod_{i \geq j = 1}^{k-1} d\theta_{ij}, \tag{18.2.19}$$

где

$$v_{ij}^{(1)} = v_{ij} - \frac{v_{ik}v_{jk}}{v_{kk}}, \qquad i, j = 1, \ldots, k-1. \tag{18.2.20}$$

Теперь мы видим, что интеграл, фигурирующий в (18.2.19), имеет ту же структуру, что и интеграл из (18.2.7), только k, n, v_{ij} заменены соответственно на $k-1$, $n-1$, $v_{ij}^{(1)}$.

Последовательное осуществление циклов интегрирования, аналогичных уже проделанному, дает

$$f(\{v_{ii}, 2v_{ij}\}) = A(I_1 \cdot J_1)(I_2 \cdot J_2) \ldots (I_k \cdot J_k), \tag{18.2.21}$$

где I_1 и J_1 указаны в (18.2.13) и (18.2.18), а

$$I_2 = \frac{\left(v_{k-1,\,k-1}^{(1)}\right)^{\frac{1}{2}(n-1)-1} \cdot 2\pi i}{\Gamma\left(\frac{n-1}{2}\right)}, \quad \ldots, \quad I_k = \frac{\left(v_{11}^{(k-1)}\right)^{\frac{1}{2}(n-k+1)-1} 2\pi \mathrm{i}}{\Gamma\left(\frac{n-k+1}{2}\right)}; \tag{18.2.22}$$

$$J_2 = \frac{\pi^{\frac{1}{2}(k-2)} \mathrm{i}^{k-2}}{\left(v_{k-1,\,k-1}^{(1)}\right)^{\frac{1}{2}(k-2)}}, \quad \ldots, \quad J_k = \frac{\pi^{\frac{1}{2}(k-k)} \mathrm{i}^{k-k}}{\left(v_{11}^{(k-1)}\right)^{\frac{1}{2}(k-k)}} = 1;$$

$v_{ij}^{(1)}$ определена в (18.2.20) и

$$v_{ij}^{(2)} = v_{ij}^{(1)} - \frac{v_{i,\,k-1}^{(1)} v_{j,\,k-1}^{(1)}}{v_{k-1,\,k-1}^{(1)}}, \qquad i, j = 1, \ldots, k-2,$$

$$v_{11}^{(k)} = v_{11}^{(k-1)} - \frac{v_{12}^{(k-1)} v_{12}^{(k-1)}}{v_{22}^{(k-1)}}. \tag{18.2.23}$$

Подставляя значения $I_1, \ldots, I_k, J_1, \ldots, J_k$ в (18.2.21), получим

$$f(\{v_{ii}, 2v_{ij}\}) = \\ = \frac{|\tau_{ij}|^{\frac{1}{2}n} \left[v_{kk} v_{k-1,\,k-1}^{(1)} \ldots v_{11}^{(k-1)}\right]^{\frac{1}{2}(n-k-1)} \exp\left[-\sum_{i,j=1}^{k} \tau_{ij} v_{ij}\right]}{2^{\frac{1}{2}k(k-1)} \pi^{\frac{1}{4}k(k-1)} \Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{n-1}{2}\right) \ldots \Gamma\left(\frac{n-k+1}{2}\right)} \tag{18.2.24}$$

в каждой точке выборочного пространства $\{v_{ii}, 2v_{ij}\}$, в которой матрица $\| v_{ij} \|$ положительно определена, и $f(\{v_{ii}, 2v_{ij}\}) = 0$ в остальных точках. Но структура членов последовательности $v_{kk}, v_{k-1, k-1}^{(1)}, \ldots$ $\ldots, v_{11}^{(k-1)}$ вполне аналогична структуре членов последовательности $\sigma^{11}, \sigma_{(1)}^{22}, \ldots, \sigma_{(k-1)}^{kk}$, определенной в § 7.4 (а). Поэтому из (7.4.6) следует, что

$$v_{kk} v_{k-1, k-1}^{(1)} \ldots v_{11}^{(k-1)} = | v_{ij} |. \tag{18.2.25}$$

Выражение в правой части (18.2.24) есть п. в. вектора $\{v_{ii}, i = 1, \ldots$ $\ldots, k; 2v_{ij}, i > j = 1, \ldots, k\}$. Следовательно, если $g(\{v_{ij}\})$ — п. в. вектора $\{v_{ij}, i \geq j = 1, \ldots, k\}$, то

$$g(\{v_{ij}\}) = 2^{\frac{1}{2} k(k-1)} f(\{v_{ii}, 2v_{ij}\}). \tag{18.2.26}$$

Вспоминая, что $\tau_{ij} = \frac{1}{2} \sigma^{ij}$, мы приходим, наконец, к следующему фундаментальному результату Уишарта (1928):

18.2.1. Если $(x_{1\xi}, \ldots, x_{k\xi}; \xi = 1, \ldots, n)$, $k \leq n$ *— выборка из* k*-мерной нормальной совокупности* $N(\{\mu_i\}, \| \sigma_{ij} \|)$ *и* $\| v_{ij} \|$ *— матрица рассеивания выборки относительно среднего* $(\mu_1, \ldots, \mu_k)$, *определенная в* (18.2.2), *то распределение элементов матрицы* $\| v_{ij} \|$ *задается п. в.*

$$g(\{v_{ij}\}) = \frac{| \sigma^{ij} |^{\frac{1}{2} n} \cdot | v_{ij} |^{\frac{1}{2}(n-k-1)} \exp\left[-\frac{1}{2} \sum_{i, j=1}^{k} \sigma^{ij} v_{ij}\right]}{2^{\frac{1}{2} kn} \pi^{\frac{1}{4} k(k-1)} \Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{n-1}{2}\right) \ldots \Gamma\left(\frac{n-k+1}{2}\right)} \tag{18.2.27}$$

в $\frac{1}{2} k(k+1)$*-мерной области, в которой матрица* $\| v_{ij} \|$ *положительно определена, и* $g(\{v_{ij}\}) = 0$ *в остальных точках.*

Будем говорить, что матрица $\| v_{ij} \|$ случайных величин, элементы которой распределены с п. в. (18.2.27), имеет *распределение Уишарта* $W(k, n, \| \sigma_{ij} \|)$, которое характеризуется параметрами k, n, $\| \sigma_{ij} \|$.

Это распределение было первоначально получено Уишартом (1928) с помощью геометрических рассуждений; несколько иной вариант геометрических рассуждений был предложен позже Махаланобисом, Бозе и Роем (1937). Оно было получено также различными другими методами Ингамом (1933), Уишартом и Бартлеттом (1932), Мэдоу (1938), Сюем (1939а), Свердрупом (1947) и Рэсчем (1948). Приведенный здесь вывод принадлежит, по существу, Ингаму, Уишарту и Бартлетту.

В случае $k = 1$, $\sigma^{11} = \frac{1}{\sigma^2}$, где σ^2—дисперсия исходного нормального распределения, и распределение Уишарта $W(1, n, \sigma^2)$ сводится к распределению, при котором величина $\sigma^{11} v_{11}$ имеет χ^2-распределение с n степенями свободы.

Отметим, что число случайных величин в выборке nk, так что распределение Уишарта является $\frac{1}{2}k(k+1)$-мерным распределением, задаваемым nk-мерным распределением элементов выборки. Джеймс (1954) показал, что существует такое преобразование величин $x_{i\xi}$, которое переводит элемент вероятности выборки в произведение элементов вероятности трех распределений, одно из которых — распределение Уишарта, второе задает положение гиперплоскости, порожденной kn-мерными векторами

$$[(x_{i1}-\mu_i), \ldots, (x_{in}-\mu_i)], \qquad i=1, \ldots, k,$$

а третье является распределением ортогональной $(k \times k)$-матрицы, определяющей ориентацию этих k векторов на гиперплоскости.

(b) Моменты и распределение рассеивания выборок из нормальной совокупности. Формула (18.1.26) показывает, что среднее значение рассеивания $|v_{ij}|$ выборки $(x_{1\xi}, \ldots, x_{k\xi};\ \xi=1, \ldots, n)$ относительно вектора средних $(\mu_1, \ldots, \mu_k)$ распределения, из которого извлечена выборка, равно $k!\binom{n}{k}|\sigma_{ij}|$, где $\|\sigma_{ij}\|$ — матрица ковариаций этого распределения. В специальном случае k-мерного нормального распределения r-й момент $|v_{ij}|$ может быть вычислен следующим образом.

Прежде всего, положим

$$g(\{v_{ij}\})=K(k, n, \{\sigma_{ij}\})\cdot h(k, n, \{\sigma_{ij}\}, \{v_{ij}\}), \tag{18.2.28}$$

где

$$K(k, n, \{\sigma_{ij}\})=\frac{|\sigma^{ij}|^{\frac{1}{2}n}}{2^{\frac{1}{2}kn}\pi^{\frac{1}{4}k(k-1)}\prod_{i=1}^{k}\Gamma\left(\frac{n+1-i}{2}\right)} \tag{18.2.29}$$

и

$$h(k, n, \{\sigma_{ij}\}, \{v_{ij}\})=|v_{ij}|^{\frac{1}{2}(n-k-1)}\exp\left[-\frac{1}{2}\sum_{i, j=1}^{k}\sigma^{ij}v_{ij}\right]. \tag{18.2.30}$$

Тогда

$$\mathcal{E}|v_{ij}|^r=K(k, n, \{\sigma_{ij}\})\int\limits_{R_{\frac{1}{2}k(k+1)}}|v_{ij}|^r h(k, n, \{\sigma_{ij}\}, \{v_{ij}\})\prod_{i\geqslant j=1}^{k}dv_{ij}. \tag{18.2.31}$$

Так как интеграл от $g(\{v_{ij}\})$ по всему пространству переменных $\{v_{ij}\}$ равен единице, имеем

$$\int\limits_{R_{\frac{1}{2}k(k+1)}}h(k, n, \{\sigma_{ij}\}, \{v_{ij}\})\prod_{i\geqslant j=1}^{k}dv_{ij}=\frac{1}{K(k, n, \{\sigma_{ij}\})}. \tag{18.2.32}$$

Но подинтегральная функция в (18.2.31) есть $h(k, n+2r, \{\sigma_{ij}\}, \{v_{ij}\})$ и интеграл от нее по пространству переменных $\{v_{ij}\}$ равен $1/K(k, n+2r, \{\sigma_{ij}\})$. Следовательно,

$$\mathscr{E}\,|v_{ij}|^r = \frac{K(k, n, \{\sigma_{ij}\})}{K(k, n+2r, \{\sigma_{ij}\})} = |2\sigma_{ij}|^r \prod_{i=1}^{k} \frac{\Gamma\left(\frac{n+1-i}{2}+r\right)}{\Gamma\left(\frac{n+1-i}{2}\right)}. \qquad (18.2.33)$$

Полагая $r=1$, находим для рассеивания $|v_{ij}|$ выборки объема n из $N(\{\mu_i\}, \|\sigma_{ij}\|)$ формулу

$$\mathscr{E}\,|v_{ij}| = k!\binom{n}{k}|\sigma_{ij}|.$$

Напомним, что этот результат, согласно (18.1.26), верен для более общего случая выборки объема n из произвольной k-мерной совокупности с матрицей ковариаций $\|\sigma_{ij}\|$.

Заметим, что $\mathscr{E}\,|v_{ij}|^r$ можно записать в следующем виде:

$$\mathscr{E}\,|v_{ij}|^r = \int_0^\infty \cdots \int_0^\infty (|2\sigma_{ij}|\,z_1 \ldots z_k)^r \prod_{i=1}^{k} \left[\frac{z_i^{\frac{1}{2}(n+1-i)-1} e^{-z_i}}{\Gamma\left(\frac{n+1-i}{2}\right)} dz_i\right]. \qquad (18.2.34)$$

Отсюда видно, что распределение $|v_{ij}|$ совпадает с распределением величины $|2\sigma_{ij}|(z_1 \ldots z_k)$, где z_i — независимые случайные величины, имеющие гамма-распределения $G\left(\frac{1}{2}(n+1-i)\right)$, $i=1, \ldots, k$, соответственно.

Итак, получен следующий результат.

18.2.2. *Если $|v_{ij}|$ — рассеивание выборки объема n из совокупности $N(\{\mu_i\}, \|\sigma_{ij}\|)$ относительно вектора средних $(\mu_1, \ldots, \mu_k)$, то распределение $|v_{ij}|$ совпадает с распределением $|2\sigma_{ij}|\prod_{i=1}^{k} z_i$, где z_i — независимые случайные величины, имеющие гамма-распределения $G\left(\frac{1}{2}(n+1-i)\right)$, $i=1, \ldots, k$, соответственно.*

Явную форму ф. р. $|v_{ij}|$ в виде интеграла читатель найдет у Уилкса (1932).

(c) Воспроизводимость распределения Уишарта. Закон воспроизводимости распределения Уишарта может быть сформулирован следующим образом.

18.2.3. *Если $\{v_{ij}^{(1)}\}$ и $\{v_{ij}^{(2)}\}$ — независимые случайные векторы, имеющие распределения Уишарта $W(k, n_1, \|\sigma_{ij}\|)$ и $W(k, n_2, \|\sigma_{ij}\|)$, то случайный вектор $\{v_{ij}^{(1)}+v_{ij}^{(2)}\}$ имеет распределение Уишарта $W(k, n_1+n_2, \|\sigma_{ij}\|)$.*

Это может быть сразу доказано с помощью характеристических функций. Действительно, согласно (18.2.4) характеристические функции $\{v_{ij}^{(1)}\}$ и $\{v_{ij}^{(2)}\}$ равны

$$|\tau_{ij}|^{\frac{1}{2}n_1}\,|\tau_{ij}-i\varepsilon_{ij}t_{ij}|^{-\frac{1}{2}n_1} \quad \text{и} \quad |\tau_{ij}|^{\frac{1}{2}n_2}\,|\tau_{ij}-i\varepsilon_{ij}t_{ij}|^{-\frac{1}{2}n_2}$$

соответственно, где $\tau_{ij}=\frac{1}{2}\sigma^{ij}$ и $\varepsilon_{ii}=1$, $\varepsilon_{ij}=\frac{1}{2}$ при $i\neq j$. Так как $\{v_{ij}^{(1)}\}$ и $\{v_{ij}^{(2)}\}$ независимы, то характеристическая функция $\{v_{ij}^{(1)}+v_{ij}^{(2)}\}$ равна произведению характеристических функций $\{v_{ij}^{(1)}\}$ и $\{v_{ij}^{(2)}\}$, т. е. она равна

$$|\tau_{ij}|^{\frac{1}{2}(n_1+n_2)}\cdot|\tau_{ij}-i\varepsilon_{ij}t_{ij}|^{-\frac{1}{2}(n_1+n_2)}.$$

Но это есть характеристическая функция распределения Уишарта

$$W(k, n_1+n_2, \|\sigma_{ij}\|).$$

Основной результат типа закона воспроизводимости для распределений Уишарта и нормального, который пригодится в дальнейшем, таков:

18.2.4. *Пусть* $\{a_{ij}\}$ — *случайный вектор, имеющий распределение Уишарта* $W(k, n, \|\sigma_{ij}\|)$; $(b_{1\beta}, \ldots, b_{k\beta})$, $\beta=1, \ldots, p$, — *независимые случайные векторы, имеющие одинаковые распределения* $N(\{0\}, \|\sigma_{ij}\|)$ *и не зависящие также от* $\{a_{ij}\}$. *Тогда случайный вектор* $\{a_{ij}+\sum_{\beta=1}^{p} b_{i\beta}b_{j\beta}\}$ *имеет распределение Уишарта* $W(k, n+p, \|\sigma_{ij}\|)$.

Доказательство **18.2.4** с помощью характеристических функций аналогично доказательству теоремы **18.2.3** и предоставляется читателю в качестве упражнения. При $p\geqslant k$ **18.2.4** является непосредственным следствием **18.2.3.**

18.3. Независимость вектора средних и матрицы внутреннего рассеивания в выборках из k-мерных нормальных совокупностей

Пусть $(x_{1\xi}, \ldots, x_{k\xi};\ \xi=1, \ldots, n)$ — выборка из k-мерной нормальной совокупности $N(\{\mu_i\}, \|\sigma_{ij}\|)$ и $\|u_{ij}\|$ — матрица внутреннего рассеивания выборки. Мы покажем, что

18.3.1. *Элементы матрицы внутреннего рассеивания* $\|u_{ij}\|$ *и вектор выборочных средних* $(\bar{x}_1, \ldots, \bar{x}_k)$ *образуют независимые множества случайных величин с распределениями* $W(k, n-1, \|\sigma_{ij}\|)$ *и* $N(\{\mu_i\}, \|\sigma_{ij}/n\|)$ *соответственно.*

Чтобы установить этот результат, рассмотрим характеристическую функцию величин $\{v_{ii},\ i=1,\ldots,k;\ 2v_{ij}, i>j=1,\ldots,k\}$ и $(\bar{x}_1-\mu_1,\ldots,\bar{x}_k-\mu_k)$:

$$\varphi(\{t_{ij}\},\ \{t_i\}) = \mathscr{E}\left\{\exp\left[\mathrm{i}\sum_{i,j=1}^{k} v_{ij}t_{ij} + \mathrm{i}\sum_{i=1}^{k}(\bar{x}_i-\mu_i)t_i\right]\right\}, \tag{18.3.1}$$

где $\|v_{ij}\|$ — матрица рассеивания выборки относительно вектора средних $(\mu_1,\ldots,\mu_k)$, определенная в (18.1.27). Заметим, что правая часть (18.3.1) может быть представлена в виде

$$\left(\mathscr{E}\left\{\exp\left[\mathrm{i}\sum_{i,j=1}^{k} y_iy_jt_{ij} + \mathrm{i}\sum_{i=1}^{k}\frac{y_i}{n}t_i\right]\right\}\right)^n, \tag{18.3.2}$$

где $y_i = x_i - \mu_i$. Вычисляя (18.3.2) с помощью результатов § 7.4, находим

$$\varphi(\{t_{ij}\},\ \{t_i\}) = |\sigma^{ij}|^{\frac{1}{2}n}\,|\sigma_*^{ij}|^{-\frac{1}{2}n}\exp\left[-\frac{1}{2}\sum_{i,j=1}^{k}\sigma_{ij}^{*}\frac{t_it_j}{n}\right], \tag{18.3.3}$$

где $\|\sigma_*^{ij}\| = \|\sigma^{ij} - 2\mathrm{i}t_{ij}\|$ и $\|\sigma_{ij}^*\| = \|\sigma_*^{ij}\|^{-1}$.

Применяя теперь многомерную форму теоремы Леви **5.2.1**, п. в. величин $\{v_{ii},\ \bar{x}_i-\mu_i,\ i=1,\ldots,k;\ 2v_{ij},\ i>j=1,\ldots,k\}$ можно представить в виде

$$\left(\frac{1}{2\pi}\right)^{\frac{1}{2}k(k+3)}\int\limits_{R_{\frac{1}{2}k(k+3)}}\exp\left[-\mathrm{i}\sum_{i,j=1}^{k}v_{ij}t_{ij} - \mathrm{i}\sum_{i=1}^{k}(\bar{x}_i-\mu_i)t_i\right]\times$$
$$\times\,\varphi(\{t_{ij}\},\ \{t_i\})\prod_{i=1}^{k}dt_i\prod_{i\geqslant j=1}^{k}dt_{ij}. \tag{18.3.4}$$

Используя результаты § 7.4 и производя интегрирование по t_i, получаем, что выражение (18.3.4) приводится к

$$I\frac{\sqrt{|n\sigma^{ij}|}}{(2\pi)^{\frac{1}{2}k}}\exp\left[-\frac{n}{2}\sum_{i,j=1}^{k}\sigma^{ij}(\bar{x}_i-\mu_i)(\bar{x}_j-\mu_j)\right], \tag{18.3.5}$$

где

$$I = \left(\frac{1}{2\pi}\right)^{\frac{1}{2}k(k+1)}\int\limits_{R_{\frac{1}{2}k(k+1)}}|\sigma^{ij}|^{\frac{1}{2}(n-1)}\,|\sigma^{ij}-2\mathrm{i}t_{ij}|^{-\frac{1}{2}(n-1)}\times$$
$$\times\exp\left[-\mathrm{i}\sum_{i,j=1}^{k}u_{ij}t_{ij}\right]\prod_{i\geqslant j=1}^{k}dt_{ij} \tag{18.3.6}$$

и

$$u_{ij} = v_{ij} - n(\bar{x}_i - \mu_i)(\bar{x}_j - \mu_j) = \sum_{\xi=1}^{n} (x_{i_\xi} - \bar{x}_i)(x_{j_\xi} - \bar{x}_j).$$

Теперь, сравнивая (18.3.6) с (18.2.6), видим, что I является п. в. случайных величин $\{u_{ii},\ i = 1, \ldots, k;\ 2u_{ij},\ i > j = 1, \ldots, k\}$, где $\{u_{ij}\}$ имеют распределение Уишарта $W(k, n-1, \|\sigma_{ij}\|)$.

Таким образом, (18.3.5) представляет собой произведение п. в. распределения Уишарта $W(k, n-1, \|\sigma_{ij}\|)$ и п. в. нормального распределения $N(\{\mu_i\}, \|\sigma_{ij}/n\|)$, что и доказывает **18.3.1.**

18.4. Хотеллинговское обобщенное распределение Стьюдента

(а) Случай одной выборки. Пусть имеется выборка из некоторого k-мерного распределения и мы хотим проверить гипотезу о том, что вектор средних этого распределения есть $(\mu_1^0, \ldots, \mu_k^0)$. Сам собой напрашивается критерий, основанный на анализе рассеивания: сравнить внутреннее рассеивание $|u_{ij}|$ выборки с рассеиванием ее $|v_{ij}|$ относительно проверяемого среднего $(\mu_1^0, \ldots, \mu_k^0)$.

Рассмотрим отношение

$$R_1 = \frac{|u_{ij}|}{|v_{ij}|} = \frac{|u_{ij}|}{|u_{ij} + n(\bar{x}_i - \mu_i^0)(\bar{x}_j - \mu_j^0)|}. \tag{18.4.1}$$

Поскольку

$$|v_{ij}| = |u_{ij}| \left[1 + n \sum_{i, j=1}^{k} u^{ij} (\bar{x}_i - \mu_i^0)(\bar{x}_j - \mu_j^0)\right], \tag{18.4.2}$$

мы можем записать R_1 в другом виде:

$$R_1 = \frac{1}{1 + nD^2/(n-1)}, \tag{18.4.3}$$

где

$$D^2 = (n-1) \sum_{i, j=1}^{k} u^{ij} (\bar{x}_i - \mu_i^0)(\bar{x}_j - \mu_j^0) \tag{18.4.4}$$

и $\|u^{ij}\| = \|u_{ij}\|^{-1}$.

Очевидно, что R_1 заключено между 0 и 1 и равно единице в том и только том случае, если вектор выборочных средних $(\bar{x}_1, \ldots, \bar{x}_k)$ совпадает с вектором средних $(\mu_1^0, \ldots, \mu_k^0)$ исходного распределения. Мы предполагаем, конечно, что матрица $\|u_{ij}\|$ положительно определена, и это имеет место с вероятностью 1, если $n > k$. Величина D^2 известна как (*возведенное в квадрат*) *расстояние Махаланобиса* (1936) *между выборкой и совокупностью*, или просто как расстояние D^2 Махаланобиса. Чем больше D^2, тем, конечно, меньше R_1. Величина T^2, определяемая как

$$T^2 = nD^2, \tag{18.4.5}$$

называется (*возведенным в квадрат*) *хотеллинговским обобщенным отношением Стьюдента*, или просто отношением T^2 Хотеллинга (1931).

Следует отметить, что $R_1 = \lambda^{\frac{2}{n}}$, где λ — отношение правдоподобия Неймана — Пирсона для сложной гипотезы $\mathscr{H}(\omega; \Omega)$ (см. § 13.3). Здесь Ω — допустимое множество точек в пространстве значений параметра, состоящее из всех значений σ_{ij}, $i \geqslant j = 1, \ldots, k$, для которых $\|\sigma_{ij}\|$ положительно определена, и всех вещественных значений μ_i, а ω — подмножество Ω, состоящее из тех точек, для которых $\mu_i = \mu_i^0$, $i = 1, \ldots, k$. Доказательство этого утверждения предоставляется читателю в качестве упражнения.

Найдем теперь распределение отношения R_1, когда гипотеза $\mathscr{H}(\omega; \Omega)$ верна, т. е. выборка извлечена из нормальной совокупности с вектором средних $(\mu_1^0, \ldots, \mu_k^0)$. Это можно сделать сравнительно простым путем, именно: сначала вычислить моменты R_1, а затем восстановить распределение по моментам. Итак, мы хотим вычислить

$$\mathscr{E}(R_1^r) = E(|u_{ij}|^r |v_{ij}|^{-r}), \qquad r = 1, 2, \ldots$$

Так как элементы матрицы $\|v_{ij}\|$ имеют распределение Уишарта $W(k, n, \|\sigma_{ij}\|)$, то значение $\mathscr{E}(|v_{ij}|^{-r})$ дается формулой (18.2.33), в которой r надо заменить на $-r$. Но $v_{ij} = u_{ij} + b_i b_j$, где $b_i = \sqrt{n}(\bar{x}_i - \mu_i^0)$, $\{u_{ij}\}$ и $\{b_i\}$ являются независимыми множествами случайных величин с распределениями $W(k, n-1, \|\sigma_{ij}\|)$ и $N(\{0\}, \|\sigma_{ij}\|)$ соответственно, как мы видели в **18.3.1**.

Следовательно,

$$K(k, n-1, \{\sigma_{ij}\}) \left[\frac{|\sigma_{ij}|}{(2\pi)^k}\right]^{\frac{1}{2}} \int\limits_{R_{\frac{1}{2}k(k+1)}} \int\limits_{R_k} |v_{ij}|^{-r} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k} \sigma^{ij} b_i b_j\right] \times$$

$$\times h(k, n-1, \{\sigma_{ij}\}, \{u_{ij}\}) \prod_{i=1}^{k} db_i \prod_{i \geqslant j=1}^{k} du_{ij} =$$

$$= |2\sigma_{ij}|^{-r} \prod_{i=1}^{k} \frac{\Gamma\left(\frac{n+1-i}{2} - r\right)}{\Gamma\left(\frac{n+1-i}{2}\right)}, \qquad (18.4.6)$$

где $K(k, n-1, \{\sigma_{ij}\})$ и $h(k, n-1, \{\sigma_{ij}\}, \{u_{ij}\})$ были определены в § 18.2 (b).

Заменим в (18.4.6) n на $n+2r$ и обе части равенства умножим на $K(k, n-1, \{\sigma_{ij}\})/K(k; n+2r-1, \{\sigma_{ij}\})$. Мы видим, что левая часть полученного равенства есть интеграл, определяющий $\mathscr{E}R_1^r$, а правая часть — значение этого интеграла. Таким образом,

$$\mathscr{E}R_1^r = \frac{|2\sigma_{ij}|^{-r} K(k, n-1, \{\sigma_{ij}\})}{K(k, n+2r-1, \{\sigma_{ij}\})} \prod_{i=1}^{k} \frac{\Gamma\left(\frac{n+1-i}{2}\right)}{\Gamma\left(\frac{n+1-i}{2} + r\right)}. \qquad (18.4.7)$$

Подставляя значения $K(k, n-1, \{\sigma_{ij}\})$, $K(k, n+2r-1, \{\sigma_{ij}\})$ из (18.2.29) в (18.4.7), после упрощений получим

$$\mathscr{E}R_1^r = \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{n-k}{2}+r\right)}{\Gamma\left(\frac{n}{2}+r\right)\Gamma\left(\frac{n-k}{2}\right)}, \tag{18.4.8}$$

где $r=1, 2, \ldots$ Замечаем, что правая часть (18.4.8) представляет собой r-й момент бета-распределения:

$$\mathscr{E}R_1^r = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n-k}{2}\right)}\int_0^1 R_1^{\frac{1}{2}(n-k)-1+r}(1-R_1)^{\frac{1}{2}k-1}dR_1,\ r=1, 2, \ldots$$

Из **5.5.1a** выводим, что R_1 имеет бета-распределение $\mathrm{Be}\left(\frac{1}{2}(n-k), \frac{1}{2}k\right)$, т. е. элемент вероятности R_1 равен

$$\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n-k}{2}\right)} R_1^{\frac{1}{2}(n-k)-1}(1-R_1)^{\frac{1}{2}k-1}dR_1. \tag{18.4.9}$$

Сформулируем полученный результат.

18.4.1. *Если* $(x_{1\xi}, \ldots, x_{k\xi};\ \xi=1, \ldots, n)$, $n>k$, *— выборка из нормальной совокупности* $N(\{\mu_i^0\}, \|\sigma_{ij}\|)$, *то отношение рассеиваний* $R_1=|u_{ij}|/|v_{ij}|$ *имеет бета-распределение* $\mathrm{Be}\left(\frac{1}{2}(n-k), \frac{1}{2}k\right)$.

Применяя к (18.4.9) преобразование

$$R_1 = \frac{1}{1+T^2/(n-1)},$$

получим, что элемент вероятности распределения T^2 Хотеллинга равен

$$\frac{\Gamma\left(\frac{n}{2}\right)}{(n-1)^{\frac{1}{2}k}\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n-k}{2}\right)}\left(1+\frac{T^2}{(n-1)}\right)^{-\frac{1}{2}n}(T^2)^{\frac{1}{2}k-1}d(T^2) \tag{18.4.10}$$

в интервале $(0, +\infty)$ значений T^2. Отметим, что при $k=1$ (18.4.10) сводится к элементу вероятности обычного t^2 Стьюдента с $(n-1)$ степенями свободы (см. § 7.8 (b)).

(b) Случай двух выборок. Предположим, что $(x_{1\xi_1}^{(1)}, \ldots, x_{k\xi_1}^{(1)};\ \xi_1=1, \ldots, n_1)$ и $(x_{1\xi_2}^{(2)}, \ldots, x_{k\xi_2}^{(2)};\ \xi_2=1, \ldots, n_2)$ — две выборки из нормальных совокупностей с одинаковыми параметрами. Обозначим общее распределение через $N(\{\mu_i\}, \|\sigma_{ij}\|)$. Пусть $\|u_{ij}^{(1)}\|$ и $\|u_{ij}^{2}\|$ —

матрицы внутреннего рассеивания выборок, а $\|u_{ij}\|$ — матрица внутреннего рассеивания объединенной выборки, получающейся из двух данных. Матрицу $\|u_{ij}^{(1)}+u_{ij}^{(2)}\|$ будем называть *матрицей внутривыборочного рассеивания* для двух выборок. Геометрически она представляет собой матрицу рассеивания k-мерной грозди, полученной следующим образом. Одна выборочная гроздь жестко сдвигается (без вращения) относительно другой так, чтобы совпали средние обеих выборок. Затем обе выборочные грозди объединяются в одну.

Рассмотрим отношение

$$R_1'=\frac{|u_{ij}^{(1)}+u_{ij}^{(2)}|}{|u_{ij}|}. \tag{18.4.11}$$

Покажем, что R_1' имеет структуру и распределение, аналогичные R_1, определенному в (18.4.3). Положим

$$u_{ij}^{w}=u_{ij}^{(1)}+u_{ij}^{(2)}. \tag{18.4.12}$$

Тогда общий член u_{ij} определителя в правой части (18.1.20) можно представить в виде

$$u_{ij}=u_{ij}^{w}+b_ib_j, \tag{18.4.13}$$

где

$$b_i=\sqrt{\frac{n_1n_2}{n_1+n_2}}\,(\bar{x}_i^{(1)}-\bar{x}_i^{(2)}). \tag{18.4.14}$$

Имеем

$$R_1'=\frac{|u_{ij}^{w}|}{|u_{ij}^{w}+b_ib_j|}. \tag{18.4.15}$$

Из **18.2.3** следует, что $\{u_{ij}^{w}\}$ имеет распределение Уишарта $W(k,\ n_1+n_2-2,\ \|\sigma_{ij}\|)$. В то же время, так как b_i являются функциями от векторов средних двух выборок, вследствие **18.3.1** $\{b_i\}$ не зависит от $\{u_{ij}^{w}\}$. Легко проверить, что $\{b_i\}$ имеет распределение $N(\{0\};\ \|\sigma_{ij}\|)$. Поэтому ясно, что задача нахождения распределения R_1' аналогична задаче нахождения распределения R_1. Фактически мы обнаружили, что распределение R_1' совпадает с распределением величины R_1, в которой $n-1$ заменено на n_1+n_2-2.

Запишем

$$R_1'=\frac{1}{1+\dfrac{n_1n_2}{(n_1+n_2)(n_1+n_2-2)}D'^2}, \tag{18.4.16}$$

где

$$D'^2=\frac{(n_1+n_2)(n_1+n_2-2)}{n_1n_2}\sum_{i,j=1}^{k}u_w^{ij}b_ib_j=$$

$$=(n_1+n_2-2)\sum_{i,j=1}^{k}u_w^{ij}(\bar{x}_i^{(1)}-\bar{x}_i^{(2)})(\bar{x}_j^{(1)}-\bar{x}_j^{(2)}) \tag{18.4.17}$$

и $\|u_w^{ij}\|=\|u_{ij}^{w}\|^{-1}$.

Величина D'^2 называется (*возведенным в квадрат*) *обобщенным расстоянием Махаланобиса* (1930, 1936) *между двумя выборками*.

Хотеллинговское обобщенное отношение Стьюдента для задачи о двух выборках определяется как

$$T'^2 = \frac{n_1 n_2}{n_1 + n_2} D'^2.$$

Элемент вероятности распределения T'^2 дается выражением (18.4.10) с $n-1$, замененным на $n_1 + n_2 - 2$; конечно, предполагается, что обе выборки извлечены независимо из одинаковых k-мерных нормальных совокупностей.

Отметим, наконец, что $R_1' = \lambda^{2/(n_1+n_2)}$, где λ — отношение правдоподобия Неймана — Пирсона для сложной гипотезы $\mathscr{H}$ (ω; Ω). Здесь Ω — параметрическое множество, в точках которого $\|\sigma_{ij}^{(1)}\| = \|\sigma_{ij}^{(2)}\| = \|\sigma_{ij}\|$ является положительно определенной матрицей, а векторы $(\mu_1^{(1)}, \ldots, \mu_k^{(1)})$ и $(\mu_1^{(2)}, \ldots, \mu_k^{(2)})$ имеют произвольные значения, в точках же $\omega \subset \Omega$ $\mu_i^{(1)} = \mu_i^{(2)}$, $i = 1, \ldots, k$. Сами же выборки извлекаются из нормальных совокупностей $N(\{\mu_i^{(1)}\}; \|\sigma_{ij}\|)$ и $N(\{\mu_i^{(2)}\}; \|\sigma_{ij}\|)$ соответственно. Доказательство этого предоставляется читателю.

18.5. Критерий для многомерной Модели I дисперсионного анализа

Критерий для многомерной модели I дисперсионного анализа представляет собой в наиболее общей форме критерий для проверки линейной гипотезы $\mathscr{H}$ в теории нормальной регрессии, определенной в § 13.3 (b). Отношение правдоподобия для этой гипотезы имеет вид

$$\lambda = \left[\frac{S_\Omega}{S_\Omega + (S_\omega - S_\Omega)}\right]^{\frac{1}{2}n}, \tag{18.5.1}$$

где S_Ω и $S_\omega - S_\Omega$ — независимые случайные величины, распределенные по одномерным законам Уишарта $W(1, m_1, \sigma^2)$ и $W(1, m_2, \sigma^2)$ соответственно, когда $\mathscr{H}$ истинна. Здесь $m_1 = n - k$ и $m_2 = k - k'$ — числа степеней свободы случайных величин S_Ω и $S_\omega - S_\Omega$ соответственно. Как было отмечено в § 13.3 (b), критерий, основанный на отношении правдоподобия λ, эквивалентен критерию, основанному на статистике F Снедекора, где

$$F = \frac{m_1 (S_\omega - S_\Omega)}{m_2 S_\Omega}. \tag{18.5.2}$$

Эти критерии также эквивалентны критерию, базирующемуся на статистике $\lambda^{2/n}$, которая просто равна отношению, стоящему в квадратных скобках (18.5.1).

Числитель и знаменатель этого отношения являются одномерными рассеиваниями, причем числитель меньше знаменателя.

Многомерное обобщение критерия для многомерной модели дисперсионного анализа в наиболее общей форме состоит в обобщении отношения внутри квадратных скобок (18.5.1). Этим обобщением является отношение

$$R_s = \frac{|a_{ij}|}{\left| a_{ij} + \sum\limits_{\beta=1}^{s} b_{i\beta} b_{j\beta} \right|}, \tag{18.5.3}$$

$i,\ j = 1, \ldots, k,\ k \leqslant m$, где $\{a_{ij}\}$ имеет распределение Уишарта $W(k, m, \|\sigma_{ij}\|)$ и $(b_{1\beta}, \ldots, b_{k\beta})$, $\beta = 1, \ldots, s$, — независимые случайные векторы, распределенные по одному и тому же нормальному закону $N(\{0\}; \|\sigma_{ij}\|)$ и не зависящие в совокупности от $\{a_{ij}\}$.

Используя для определения r-го момента R_s тот же метод, что и при отыскании r-го момента R_1, получим

$$\mathscr{E}R_s^r = \prod_{i=1}^{k} \frac{\Gamma\left(\frac{m+1-i}{2}+r\right)\Gamma\left(\frac{m+s+1-i}{2}\right)}{\Gamma\left(\frac{m+1-i}{2}\right)\Gamma\left(\frac{m+s+1-i}{2}+r\right)}. \tag{18.5.4}$$

Заметим, что при $s \geqslant k$ $\mathscr{E}R_s^r$ может быть записано в виде

$$\mathscr{E}R_s^r = \prod_{i=1}^{k}\left[\frac{\Gamma\left(\frac{m+s+1-i}{2}\right)}{\Gamma\left(\frac{m+1-i}{2}\right)\Gamma\left(\frac{s}{2}\right)} \cdot \frac{\Gamma\left(\frac{m+1-i}{2}+r\right)\Gamma\left(\frac{s}{2}\right)}{\Gamma\left(\frac{m+s+1-i}{2}+r\right)}\right], \tag{18.5.4a}$$

откуда видно, что распределение R_s совпадает с распределением $\prod\limits_{i=1}^{k} z_i$, где z_i — независимые случайные величины, имеющие бета-распределения

$$\mathrm{Be}\left(\frac{m+1-i}{2},\ \frac{s}{2}\right), \qquad i = 1, \ldots, k,$$

соответственно.

Аналогично при $1 \leqslant s < k$ $\mathscr{E}R_s^r$ может быть представлено в виде

$$\mathscr{E}R_s^r = \prod_{i=1}^{s}\left[\frac{\Gamma\left(\frac{m+i}{2}\right)}{\Gamma\left(\frac{m-k+i}{2}\right)\Gamma\left(\frac{k}{2}\right)}\ \frac{\Gamma\left(\frac{m-k+i}{2}+r\right)\Gamma\left(\frac{k}{2}\right)}{\Gamma\left(\frac{m+i}{2}+r\right)}\right], \tag{18.5.4b}$$

откуда ясно, что в этом случае распределение R_s одинаково с распределением $\prod\limits_{i=1}^{s} w_i$, где w_i — независимые случайные величины, имеющие соответственно бета-распределения

$$\mathrm{Be}\left(\frac{m-k+i}{2},\ \frac{k}{2}\right), \qquad i = 1, \ldots, s.$$

Суммируем все полученное выше в виде следующего результата.

18.5.1. *Пусть* $\|a_{ij}\|$ — *симметричная матрица, положительно определенная с вероятностью* 1, *случайные элементы которой имеют совместное распределение Уишарта* $W(k,\ m,\ \|\sigma_{ij}\|)$; *далее, пусть*

$$(b_{1\beta},\ \ldots,\ b_{k\beta}),\quad \beta=1,\ \ldots,\ s,$$

— *независимые k-мерные случайные векторы, распределенные по одному и тому же закону* $N(\{0\};\ \|\sigma_{ij}\|)$ *и в совокупности не зависящие от* $\{a_{ij}\}$. *Если*

$$R_s=\frac{|a_{ij}|}{\left|a_{ij}+\sum\limits_{\beta=1}^{s}b_{i\beta}b_{j\beta}\right|},$$

то

1) *при* $s\geqslant k$ *распределение* R_s *совпадает с распределением произведения* k *независимых случайных величин, имеющих бета-распределения*

$$\mathrm{Be}\left(\frac{m+1-i}{2},\ \frac{s}{2}\right),\qquad i=1,\ \ldots,\ k;$$

2) *при* $1\leqslant s<k$ *распределение* R_s *совпадает с распределением произведения* s *независимых случайных величин, имеющих бета-распределения*

$$\mathrm{Be}\left(\frac{m-k+i}{2},\ \frac{k}{2}\right),\qquad i=1,\ \ldots,\ s.$$

Отметим, что определенная в (18.4.1) величина R_1 есть частный случай R_s при $s=1$, $m=n-1$, $a_{ij}=u_{ij}$, $b_i=\sqrt{n}\,(\bar{x}_i-\mu_i)$, $i,\ j=1,\ \ldots,\ k$. Точно так же R_1' из (18.4.11) или (18.4.15) есть частный случай R_s при $s=1$, $m=n_1+n_2-2$, $a_{ij}=u_{ij}^{(1)}+u_{ij}^{(2)}$ и

$$b_i=\sqrt{\frac{n_1n_2}{n_1+n_2}}\,(\bar{x}_i^{(1)}-\bar{x}_i^{(2)}),\qquad i,\ j=1,\ \ldots,\ k.$$

Интересный специальный случай R_s получается при $s=2$; именно, $\sqrt{R_2}$ следует бета-распределению $\mathrm{Be}(m+1-k,\ k)$ и элемент вероятности $\sqrt{R_2}$ равен

$$\frac{\Gamma(m+1-k)}{\Gamma(m+1)\,\Gamma(k)}\left(\sqrt{R_2}\right)^{m-k}\left(1-\sqrt{R_2}\right)^{k-1}d\sqrt{R_2}. \tag{18.5.5}$$

Чтобы убедиться в этом, вычислим с помощью (18.5.4b) r-й момент R_2:

$$\mathcal{E}R_2^r=\frac{\Gamma\left(\frac{m+2}{2}\right)\Gamma\left(\frac{m+1}{2}\right)\Gamma\left(\frac{m+2-k}{2}+r\right)\Gamma\left(\frac{m+1-k}{2}+r\right)}{\Gamma\left(\frac{m+2}{2}+r\right)\Gamma\left(\frac{m+1}{2}+r\right)\Gamma\left(\frac{m+2-k}{2}\right)\Gamma\left(\frac{m+1-k}{2}\right)}. \tag{18.5.6}$$

Используя формулу удвоения Лежандра

$$\Gamma(\nu)\,\Gamma\left(\nu+\frac{1}{2}\right)=\frac{\sqrt{\pi}\,\Gamma(2\nu)}{2^{2\nu-1}}, \tag{18.5.7}$$

можно объединить попарно гамма-функции в (18.5.6) и получить

$$\mathscr{E}R_2^r=\frac{\Gamma(m+1)\,\Gamma(m+1-k+2r)}{\Gamma(m+1+2r)\,\Gamma(m+1-k)}. \tag{18.5.8}$$

Это может быть записано как

$$\mathscr{E}R_2^r=\frac{\Gamma(m+1)}{\Gamma(m+1-k)\,\Gamma(k)}\int\limits_0^1 z^{m-k+2r}(1-z)^{k-1}\,dz. \tag{18.5.9}$$

Из (18.5.9) и **5.5.1а** мы видим, что распределение R_2 совпадает с распределением z^2, где z имеет бета-распределение $\mathrm{Be}(m+1-k,\ k)$. Следовательно, это же распределение имеет $\sqrt{R_2}$. Выражение для ф. р. R_s в форме интеграла дано Уилксом (1932).

Пример. Предположим, что $(x_{1\xi_\gamma}^{(\gamma)},\ \ldots,\ x_{k\xi_\gamma}^{(\gamma)};\ \xi_\gamma=1,\ \ldots,\ n_\gamma)$, $\gamma=1,\ 2,\ 3,$ — три независимые выборки из одинаковых нормальных совокупностей $N(\{\mu_i\};\ \|\sigma_{ij}\|)$. Пусть $\|u_{ij}^{(\gamma)}\|$, $\gamma=1,\ 2,\ 3$, будут матрицами внутреннего рассеивания трех выборок, а $\|u_{ij}\|$ — матрицей внутреннего рассеивания большой выборки, полученной при объединении трех выборок в одну. Наконец, пусть $\|u_{ij}^{(1)}+u_{ij}^{(2)}+u_{ij}^{(3)}\|=\|u_{ij}^{W}\|$ будет матрицей внутривыборочного рассеивания для трех выборок. Тогда отношение

$$R_2'=\frac{|u_{ij}^{W}|}{|u_{ij}|} \tag{18.5.10}$$

служит обобщением отношения R_2' из (18.4.15) на случай трех выборок. R_2'' является частным случаем R_s из (18.5.3) при $s=2$ и $m=n_1+n_2+n_3-3$ и, следовательно, элемент вероятности задается формулой (18.5.5) с $m=n_1+$ $+n_2+n_3-3$.

Отметим, что система случайных величин $\{u_{ij}^{W}\}$ имеет распределение Уишарта $W(k,\ n_1+n_2+n_3-3,\ \|\sigma_{ij}\|)$. Определим, кроме того, *матрицу межвыборочного рассеивания*

$$u_{ij}^{B}=u_{ij}-u_{ij}^{W}=\sum_{\gamma=1}^{3}n_\gamma\,(\bar{x}_i^{(\gamma)}-\bar{x}_i)\,(\bar{x}_j^{(\gamma)}-\bar{x}_j), \tag{18.5.11}$$

$i,\ j=1,\ \ldots,\ k$, где $(\bar{x}_1,\ \ldots,\ \bar{x}_k)$ — вектор средних объединенной выборки. Можно найти два независимых вектора $(b_{1\beta},\ \ldots,\ b_{k\beta})$, $\beta=1,2$, которые вместе не зависят также от $\{u_{ij}^{W}\}$, допускают представление $b_{i\beta}=\sum\limits_{\gamma=1}^{3}c_{i\beta}^{(\gamma)}\bar{x}_i^{(\gamma)}$, распределены по одному и тому же нормальному закону $N(\{0\};\ \|\sigma_{ij}\|)$, и с помощью этих векторов представить правую часть (18.5.11) как $\sum\limits_{\beta=1}^{2}b_{i\beta}b_{j\beta}$. Однако нет необходимости определять эти векторы, чтобы отыскать распределение R_2'. Оно может быть легко найдено по своим моментам с помощью процедуры, аналогичной той, которая уже использовалась при получении распределения R_1.

Можно показать, конечно, что R_2' есть эквивалент обобщения для трех выборок отношения правдоподобия Неймана — Пирсона, описанного в конце § 18.4 (b).

18.6. Главные компоненты

(a) Собственные значения и собственные векторы матрицы рассеивания. В этом параграфе мы будем заниматься следующей задачей. Рассматривается выборка объема n из k-мерного распределения, $k \leqslant n$. Эта выборка может быть геометрически представлена как выборочная гроздь из n точек в k-мерном эвклидовом пространстве R_k. Предположим, что мы хотим ортогонально спроектировать эту выборочную гроздь на s-мерное эвклидово подпространство R_s, $s \leqslant k$, так, чтобы получить максимально возможное значение s-мерного рассеивания проекций. Задача состоит в определении: 1) направления проектирования относительно системы координат R_k и 2) величины рассеивания в R_s.

Решение этой задачи может быть сформулировано в виде следующего результата, принадлежащего Хотеллингу (1933).

18.6.1. *Предположим, что* $(x_{1\xi}, \ldots, x_{k\xi};\ \xi = 1, \ldots, n)$ — *выборка объема* $n > k$ *из* k-*мерной совокупности с положительно определенной матрицей ковариаций. Пусть* $\| u_{ij} \|$ — *матрица внутреннего рассеивания этой выборки — положительно определена с вероятностью 1. Введем* s k-*мерных единичных векторов*

$$(c_{1p}, \ldots, c_{kp}), \qquad p = 1, \ldots, s, \tag{18.6.1}$$

так что $\sum_{i=1}^{k} c_{ip}^2 = 1,\ p = 1, \ldots, s,$ *и положим*

$$z_{p\xi} = \sum_{i=1}^{k} c_{ip}\, x_{i\xi}, \qquad p = 1, \ldots, s. \tag{18.6.2}$$

Пусть $\| \tilde{u}_{pq} \|$ *будет матрицей внутреннего рассеивания выборки* $(z_{1\xi}, \ldots, z_{s\xi};\ \xi = 1, \ldots, n)$.

Векторы (18.6.1), *которые максимизируют рассеивание* $| \tilde{u}_{pq} |$, *определяются как решения* s *систем уравнений:*

$$\sum_{j=1}^{k} (u_{ij} - l_p \delta_{ij})\, c_{jp} = 0, \qquad i = 1, \ldots, k, \tag{18.6.3}$$

$p = 1, \ldots, s,$ *где* $l_1, \ldots, l_s$ — s *наибольших корней характеристического уравнения*

$$| u_{ij} - l \delta_{ij} | = 0,$$

δ_{ij} — *обычный символ Кронекера, и где* $l_1 > \ldots > l_s$ *с вероятностью*

1. *Эти векторы, кроме того, ортогональны и максимальное значение* $|\tilde{u}_{pq}|$ *равно* $l_1\ l_2 \ldots l_s$.

Чтобы доказать **18.6.1**, рассмотрим следующие общие линейные формы компонент каждого элемента выборки:

$$z_\xi = \sum_{i=1}^{k} c_i x_{i\xi}, \qquad \xi = 1, \ldots, k, \tag{18.6.4}$$

и образуем сумму квадратов

$$Q = \sum_{\xi=1}^{n} (z_\xi - \bar{z})^2 = \sum_{i,j=1}^{k} u_{ij} c_i c_j, \tag{18.6.5}$$

где

$$\sum_{i=1}^{k} c_i^2 = 1.$$

Найдем теперь значения $(c_1, \ldots, c_k)$, при которых Q достигает экстремума. Используя метод множителей Лагранжа, мы получим эти векторы из уравнений

$$\frac{\partial \varphi}{\partial c_i} = 0, \qquad i = 1, \ldots, k, \tag{18.6.6}$$

где

$$\varphi = Q + l\left(1 - \sum_{1}^{k} c_i^2\right).$$

Уравнения (18.6.6) имеют вид

$$\sum_{j=1}^{k} (u_{ij} - l\delta_{ij})\, c_j = 0, \qquad i = 1, \ldots, k. \tag{18.6.7}$$

Для существования решений (18.6.7), отличных от тривиального $c_i = 0$, $i = 1, \ldots, k$ (которое, конечно, исключается требованием $\sum_i c_i^2 = 1$), необходимо, чтобы

$$|u_{ij} - l\delta_{ij}| = 0. \tag{18.6.8}$$

Так как мы предполагаем выборку объема $n > k$ извлеченной из k-мерной совокупности, а матрицу $\|u_{ij}\|$ положительно определенной с вероятностью 1, то уравнение (18.6.8) имеет k положительных корней, которые будем предполагать различными, так что с вероятностью 1 $l_1 > \ldots > l_k > 0$. Эти корни называются *собственными значениями* матрицы $\|u_{ij}\|$. Их называют также *характеристическими числами* матрицы $\|u_{ij}\|$.

Если $l_{p'}$ — какой-либо из этих корней, то вектор $(c_{1p'}, \ldots, c_{kp'})$ будет p'-м *собственным вектором*; он удовлетворяет уравнению

$$\sum_{j=1}^{k} (u_{ij} - l_{p'}\delta_{ij})\, c_{jp'} = 0. \tag{18.6.9}$$

Умножая обе части (18.6.9) на $c_{ip'}$ и суммируя по i, получим

$$\sum_{i,\,j=1}^{k}(u_{ij}-l_{p'}\delta_{ij})\,c_{ip'}\,c_{jp'}=0, \tag{18.6.10}$$

откуда следует, что

$$l_{p'}=\sum_{i,\,j=1}^{k}u_{ij}c_{ip'}c_{jp'}. \tag{18.6.11}$$

Аналогично для любых двух собственных значений $l_{p'}$, $l_{q'}$, $p'\neq q'$ имеем

$$\begin{gathered}\sum_{i,\,j=1}^{k}(u_{ij}-l_{p'}\delta_{ij})\,c_{ip'}c_{jq'}=0,\\ \sum_{i,\,j=1}^{k}(u_{ij}-l_{q'}\delta_{ij})\,c_{ip'}c_{jq'}=0.\end{gathered} \tag{18.6.12}$$

Отсюда, вычитая первое уравнение из второго, найдем

$$(l_{p'}-l_{q'})\sum_{i=1}^{k}c_{ip'}c_{iq'}=0. \tag{18.6.13}$$

Следовательно, при $p'\neq q'$

$$\sum_{i=1}^{k}c_{ip'}c_{iq'}=0, \tag{18.6.14}$$

т. е. собственные векторы $(c_{1p},\ldots,c_{kp})$, $p=1,\ldots,k$, попарно ортогональны. Эти векторы называются также *характеристическими векторами* матрицы $\|u_{ij}\|$.

С учетом (18.6.14) из любого уравнения (18.6.12) найдем, что при $p'\neq q'$

$$\sum_{i,\,j=1}^{k}u_{ij}c_{ip'}c_{jq'}=0. \tag{18.6.15}$$

Выберем теперь s собственных векторов, соответствующих $s\,(s\leqslant k)$ наибольшим характеристическим числам $l_1>\ldots>l_s$, и образуем выборку $(z_{1\xi},\ldots,z_{k\xi};\ \xi=1,\ldots,n)$, где

$$z_{p\xi}=\sum_{i=1}^{k}c_{ip}x_{i\xi},\qquad p=1,\ldots,s.$$

Из ортогональности собственных векторов следует, что эта выборка есть проекция исходной выборки на s-мерное подпространство. Внутреннее рассеивание спроектированной выборки равно, следовательно, $|\tilde{u}_{pq}|$, p, $q=1,\ldots,s$, где

$$\tilde{u}_{pq}=\sum_{\xi=1}^{n}(z_{p\xi}-\bar{z}_p)(z_{q\xi}-\bar{z}_q)=\sum_{i,\,j=1}^{k}u_{ij}c_{ip}c_{jq}. \tag{18.6.16}$$

Используя (18.6.11) и (18.6.15), находим

$$|\tilde{u}_{pq}| = l_1 l_2 \ldots l_s, \tag{18.6.17}$$

что завершает доказательство **18.6.1.**

Читатель заметит, что если положить $s = k$, то $|\tilde{u}_{pq}| = |u_{ij}| = l_1 \ldots l_k$. Впрочем, это следует и из того, что произведение корней уравнения (18.6.8) равно $|u_{ij}|$.

Обращаясь снова к (18.6.8), мы видим, что

$$l_1 + \ldots + l_k = \sum_{i=1}^{k} u_{ii}. \tag{18.6.18}$$

Другими словами, сумма собственных значений матрицы $\|u_{ij}\|$ равна сумме рассеиваний отдельных компонент выборки относительно своих средних значений. Собственные значения $l_1, \ldots, l_k$ матрицы $\|u_{ij}\|$ внутреннего рассеивания выборки, поделенные на $n-1$, Хотеллинг (1933) назвал *главными компонентами* полной выборочной дисперсии, так как

$$\frac{l_1}{n-1} + \ldots + \frac{l_k}{n-1} = s_1^2 + \ldots + s_k^2, \tag{18.6.19}$$

где $s_i^2 = u_{ii}/(n-1)$, как это видно из (18.6.18).

(b) Выборочная теория собственных значений матрицы рассеивания. Так как собственные значения $l_1, \ldots, l_k$ являются корнями уравнения (18.6.8), где $\|u_{ij}\|$ — матрица внутреннего рассеивания выборки, то они сами являются случайными величинами. Естественно, встает вопрос, что можно сказать о распределении этих корней. Точная выборочная теория собственных значений матрицы рассеивания построена только для случая, когда выборка $(x_{1\xi}, \ldots, x_{k\xi};\ \xi = 1, \ldots, n)$ извлечена из k-мерного сферически симметричного нормального распределения, т. е. распределения, у которого матрица ковариаций имеет равные характеристические числа. В этом параграфе мы изложим теорию собственных значений матрицы рассеивания. Она была построена приблизительно одновременно Фишером (1939), Гиршиком (1939), Сюем (1939 b), Мудом (1951) и Роем (1939). Результаты Муда опубликованы через двенадцать лет после их получения. Позже задачей о распределении собственных значений занимались Джеймс (1954), Олкин (1951), Олкин и Рой (1954) и другие, используя различные методы.

Основной результат может быть сформулирован в следующем виде.

18.6.2. *Пусть* $(x_{1\xi}, \ldots, x_{k\xi};\ \xi = 1, \ldots, n)$ — *выборка из нормальной совокупности* $N(\{\mu_i\}; \|\sigma_{ij}\|)$ *и* $\|u_{ij}\|$ — *матрица внутреннего рассеивания выборки. Распределение корней* $l_1 \geqslant \ldots \geqslant l_k > 0$ *характеристического уравнения*

$$|u_{ij} - l\sigma_{ij}| = 0 \tag{18.6.20}$$

задается элементом вероятности

$$\frac{\pi^{\frac{1}{2}k}\left(\prod_{i=1}^{k} l_i\right)^{\frac{1}{2}(n-k-2)} \prod_{i>j=1}^{k}(l_j-l_i)\exp\left(-\frac{1}{2}\sum_{i=1}^{k} l_i\right)}{2^{\frac{1}{2}k(n-1)}\prod_{i=1}^{k}\left[\Gamma\left(\frac{k+1-i}{2}\right)\Gamma\left(\frac{n-i}{2}\right)\right]}\,dl_1\ldots dl_k \qquad (18.6.21)$$

в области $0 \leqslant l_k \leqslant l_{k-1} \leqslant \ldots \leqslant l_1 < +\infty$ *и* 0 *в остальных точках.*

Чтобы установить **18.6.2,** поступим следующим образом. Мы знаем, что $\{u_{ij}\}$ имеет распределение Уишарта $W(k, n-1, \|\sigma_{ij}\|)$. Далее, из теории положительно определенных квадратичных форм известно (см., например, Биркгоф и Маклэйн (1953)), что найдется такое невырожденное линейное преобразование

$$x_{i\xi} - \mu_i = \sum_{j=1}^{k} c_{ij} y_{j\xi}, \qquad i = 1, \ldots, k;\ \xi = 1, \ldots, n, \qquad (18.6.22)$$

при котором $(y_{1\xi}, \ldots, y_{k\xi};\ \xi = 1, \ldots, n)$ будет выборкой из $N(\{0\}; \|\delta_{ij}\|)$, где $\|\delta_{ij}\|$ — единичная матрица. Если обозначить матрицы $\|c_{ij}\|$, $\|\sigma_{ij}\|$ и $\|\delta_{ij}\|$ через c, σ и I, обратные к c и σ — через c^{-1} и σ^{-1}, а операцию транспонирования отмечать штрихом, то приведенный выше результат можно записать так:

$$c'\sigma^{-1}c = I, \qquad (c^{-1})'\sigma c^{-1} = I. \qquad (18.6.23)$$

Если $\|u^*_{ij}\|$ — матрица рассеивания выборки $(y_{1\xi}, \ldots, y_{k\xi};\ \xi = 1, \ldots, n)$, то $\{u^*_{ij}\}$ имеет распределение Уишарта $W(k, n-1, \|\delta_{ij}\|)$. Обозначив матрицы $\|u_{ij}\|$ и $\|u^*_{ij}\|$ через u и u^*, можно выразить матрицы u и u^* одну через другую следующим образом:

$$u = c'u^*c, \qquad u^* = (c^{-1})'uc^{-1}. \qquad (18.6.24)$$

Предположим теперь, что $l_1 > \ldots > l_k$ являются корнями уравнения (18.6.20), т. е.

$$|u - l\sigma| = 0. \qquad (18.6.25)$$

Вероятность совпадения двух и более корней равна нулю. Умножая матрицу $\|u - l\sigma\|$ слева на $(c^{-1})'$ и справа на c^{-1}, получим матрицу

$$(c^{-1})'uc^{-1} - l(c^{-1})'\sigma c^{-1}, \qquad (18.6.26)$$

которая в силу вторых уравнений из (18.6.23) и (18.6.24) есть

$$u^* - lI. \qquad (18.6.27)$$

Определители матриц (18.6.26) и (18.6.27), конечно, равны. Но определитель матрицы (18.6.26) есть $|(c^{-1})'|\,|u - l\sigma|\,|c^{-1}|$ и множитель $|u - l\sigma|$ обращается в нуль при $l = l_1, \ldots, l_k$. Следовательно, $|u^* - lI|$ обращается в нуль при тех же значениях l.

Таким образом, распределение корней уравнения $|u - l\sigma| = 0$, где $\{u_{ij}\}$ имеет распределение Уишарта $W(k, n-1, \|\sigma_{ij}\|)$, совпадает с распределением корней уравнения $|u^* - lI| = 0$, где $\{u_{ij}^*\}$ имеет распределение Уишарта $W(k, n-1, \|\delta_{ij}\|)$. Элемент вероятности $\{u_{ij}^*\}$ равен

$$\frac{|u_{ij}^*|^{\frac{1}{2}(n-k-2)} \exp\left(-\frac{1}{2}\sum_{i=1}^{k} u_{ii}^*\right) \prod_{i \geqslant j = 1}^{k} du_{ij}^*}{2^{\frac{1}{2}k(n-1)} \pi^{\frac{1}{4}k(k-1)} \Gamma\left(\frac{n-1}{2}\right) \ldots \Gamma\left(\frac{n-k}{2}\right)}. \tag{18.6.28}$$

Согласно теории положительно определенных квадратичных форм найдутся ортогональная матрица $\|e_{ij}\|$ случайных величин и набор положительных случайных величин $f_1 > \ldots > f_k$ такие, что

$$u^* = e'fe, \tag{18.6.29}$$

где $e = \|e_{ij}\|$, а f — диагональная матрица с элементами $f_1, \ldots, f_k$ на диагонали. Так как e — ортогональная матрица, то $e'e = I$. Заменив теперь в матрице $\|u^* - lI\|$ u^* на $e'fe$ и I на $e'e$, получим

$$u^* - lI = e'fe - le'e = e'(f - l)e. \tag{18.6.30}$$

Если рассмотреть определители матриц из (18.6.30), мы увидим, что корни $l_1 > \ldots > l_k$ равны соответственно величинам $f_1 > \ldots > f_k$, так как $|e| \neq 0$. Поэтому распределение корней $l_1, \ldots, l_k$ совпадает с распределением случайных величин $f_1, \ldots, f_k$. Чтобы найти распределение $f_1, \ldots, f_k$, применим преобразование (18.6.29) к элементу вероятности (18.6.28). Так как e — ортогональная матрица, $\{e_{ij}\}$ являются функциями от $\frac{1}{2}k(k-1)$ параметров, которые мы обозначим через g_t, $t = 1, 2, \ldots, \frac{1}{2}k(k-1)$. Якобиан преобразования равен

$$J = \left|\frac{\partial(u_{ij}^*)}{\partial(g_t, f_{i'})}\right| = \overbrace{\left|\begin{array}{c} \dfrac{\partial(u_{ij}^*)}{\partial(g_t)} \\ \cdots \\ \dfrac{\partial(u_{ij}^*)}{\partial(f_{i'})} \end{array}\right|}^{\frac{1}{2}k(k+1)\text{ столбцов}} \begin{array}{l} \left.\vphantom{\dfrac{a}{b}}\right\} \frac{1}{2}k(k-1) \text{ строк,} \\ \\ \left.\vphantom{\dfrac{a}{b}}\right\} k \text{ строк,} \end{array} \tag{18.6.31}$$

где $i \geqslant j = 1, \ldots, k$; $i' = 1, \ldots, k$; $t = 1, \ldots, \frac{1}{2}k(k-1)$. Каждый элемент J, принадлежащий одной из верхних $\frac{1}{2}k(k-1)$ строк, имеет вид $\sum_{j=1}^{k} q_j f_j$, где q_j зависит только от $\{e_{ij}\}$ и их первых производных по g_t. Каждый элемент, стоящий в одной из нижних k строк, зависит только от $\{e_{ij}\}$. Следовательно, J представляет собой однородный полином от $f_1, \ldots, f_k$ степени $\frac{1}{2}k(k-1)$ с коэффициентами, зависящими от $\{e_{ij}\}$ и их производных по $\{g_t\}$.

Преобразование (18.6.29) неоднозначно, если $f_i = f_j$ при некоторых $i \neq j$. Поэтому для каждых $i > j$ член $(f_j - f_i)^{a_i}$ должен входить сомножителем в J, где $a_{ij} \geqslant 1$ — целое положительное число. Якобиан J должен иметь, таким образом, вид $\prod\limits_{i>j=1}^{k} (f_j - f_i)^{a_{ij}} G$. Примем теперь во внимание, что, во-первых, все a_{ij} — целые числа, не меньшие 1, во-вторых, имеются $\frac{1}{2}k(k-1)$ множителей вида $(f_j - f_i)^{a_{ij}}$ и, в-третьих, J есть однородный полином от $f_1, \ldots, f_k$ степени $\frac{1}{2}k(k-1)$. Отсюда заключаем, что каждое $a_{ij} = 1$ и G не зависит от f_i, а зависит от e_{ij} и их первых производных по g_t. Следовательно, G зависит только от g_t и может быть записана как $G(\{g_t\})$. Таким образом,

$$J = \prod_{i>j=1}^{k} (f_j - f_i)\, G(\{g_t\}). \tag{18.6.32}$$

Так как $|e_{ij}| = 1$, из (18.6.29) получаем

$$|u_{ij}^*| = |e'| \cdot |f| \cdot |e| = f_1 \ldots f_k. \tag{18.6.33}$$

Кроме того,

$$\sum_{i=1}^{k} u_{ii}^* = \sum_{i=1}^{k} f_i. \tag{18.6.34}$$

Используя теперь результаты (18.6.32), (18.6.33), (18.6.34), получим, что выражение (18.6.28) после применения преобразования (18.6.29) переходит в

$$KG(g_t) \prod_{t=1}^{\frac{1}{2}k(k-1)} dg_t \left(\prod_{i=1}^{k} f_i\right)^{\frac{1}{2}(n-k-2)} \prod_{i>j=1}^{k} (f_j - f_i) \times \\ \times \exp\left(-\frac{1}{2}\sum_{i=1}^{k} f_i\right) \prod_{i=1}^{k} df_i, \tag{18.6.35}$$

где K — постоянная. Так как $\{g_t\}$ и $\{f_i\}$ — независимые множества случайных величин, элемент вероятности $\{f_i\}$ равен

$$C\left\{\left(\prod_{i=1}^{k} f_i\right)^{\frac{1}{2}(n-k-2)} \prod_{i>j=1}^{k} (f_j - f_i) \exp\left(-\frac{1}{2}\sum_{i=1}^{k} f_i\right)\right\} \prod_{i=1}^{k} df_i, \tag{18.6.36}$$

где C — постоянная, которую надо определить. Обозначив выражение, стоящее в фигурных скобках (18.6.36), через $R(k, n, \{f_i\})$, мы можем написать

$$\frac{1}{C}=\int\limits_{E_k} R(k, n, \{f_i\})\, df_1 \ldots df_k, \tag{18.6.37}$$

где E_k — подмножество R_k, в точках которого $0<f_k<f_{k-1}<\ldots<<f_1<+\infty$. Если обозначить интеграл в правой части (18.6.37) через $\varphi_k\left[\frac{1}{2}(n-k-2)\right]$ и вспомнить, что $|u_{ij}^*|^r=\left(\prod\limits_{i=1}^{k} f_j\right)^r$, то r-й момент величины $|u_{ij}^*|$ можно представить в виде

$$\mathscr{E}\,|u_{ij}^*|^r=\frac{\varphi_k\left(\frac{n-k-2}{2}+r\right)}{\varphi_k\left(\frac{n-k-2}{2}\right)}. \tag{18.6.38}$$

Но так как $\{u_{ij}^*\}$ имеет распределение Уишарта $W(k, n-1, \|\delta_{ij}\|)$, элемент вероятности которого равен (18.6.28), то r-й момент $|u_{ij}^*|$ определяется формулой (18.2.33) с $\|\sigma_{ij}\|=\|\delta_{ij}\|$ и n, замененным на $n-1$. Следовательно,

$$\frac{\varphi_k\left(\frac{n-k-2}{2}+r\right)}{\varphi_k\left(\frac{n-k-2}{2}\right)}=2^{rk}\prod_{i=1}^{k}\frac{\Gamma\left(\frac{n-i}{2}+r\right)}{\Gamma\left(\frac{n-i}{2}\right)}. \tag{18.6.39}$$

Если положить $r=-\frac{1}{2}(n-k-2)$, то будем иметь

$$C=\frac{1}{\varphi_k\left(\frac{n-k-2}{2}\right)}=\frac{2^{-\frac{1}{2}k(n-k-2)}}{\varphi_k(0)}\prod_{i=1}^{k}\frac{\Gamma\left(\frac{k+2-i}{2}\right)}{\Gamma\left(\frac{n-i}{2}\right)}. \tag{18.6.40}$$

Но

$$\varphi_k(0)=\int\limits_{E_k}\prod_{i>j=1}^{k}(f_j-f_i)\exp\left(-\frac{1}{2}\sum_{i=1}^{k}f_i\right)df_1\ldots df_k. \tag{18.6.41}$$

Произведя преобразование $f_k=f_k'$, $f_{k-1}=f_{k-1}'+f_k'$, ..., $f_1=f_1'+f_k'$ мы найдем, что

$$\varphi_k(0)=\frac{2}{k}\varphi_{k-1}(1). \tag{18.6.42}$$

Но из (18.6.40) при $n=k+4$ выводим, что

$$\frac{\varphi_k(0)}{\varphi_k(1)}=2^{-k}\prod_{i=1}^{k}\frac{\Gamma\left(\frac{k+2-i}{2}\right)}{\Gamma\left(\frac{k+4-i}{2}\right)}=\frac{1}{(k+1)!}. \tag{18.6.43}$$

Из (18.6.42) и (18.6.43) следует

$$\varphi_k(1) = (k+1)!\,\frac{2}{k}\,\varphi_{k-1}(1).$$

Заменяя k последовательно на $k-1$, $k-2, \ldots, 2$ и замечая, что $\varphi_1(1)=4$, получим

$$\varphi_k(1) = 2^k (k+1)! \prod_{i=1}^{k} \Gamma(k+1-i). \tag{18.6.44}$$

Подставляя значение $\varphi_k(1)$ в (18.6.43), находим, что

$$\varphi_k(0) = 2^k \prod_{i=1}^{k} \Gamma(k+1-i). \tag{18.6.45}$$

Используя теперь формулу удвоения Лежандра для гамма-функции

$$\Gamma(k+1-i) = \frac{2^{k-i}}{\sqrt{\pi}}\,\Gamma\left(\frac{k+1-i}{2}\right)\Gamma\left(\frac{k+2-i}{2}\right), \tag{18.6.46}$$

будем иметь при последовательном применении формул (18.6.46), (18.6.45) и (18.6.40)

$$C = \frac{\pi^{\frac{1}{2}k}}{2^{\frac{1}{2}k(n-1)} \prod\limits_{i=1}^{k} \Gamma\left(\frac{k+1-i}{2}\right)\Gamma\left(\frac{n-i}{2}\right)}. \tag{18.6.47}$$

Подставляя это значение C в (18.6.36), мы получим, наконец, распределение $f_1, \ldots, f_k$. Вспомним теперь, что $f_1, \ldots, f_k$ распределены одинаково с $l_1, \ldots, l_k$. Поэтому распределение случайных величин $l_1, \ldots, l_k$ дается формулой (18.6.36) с заменой $f_1, \ldots, f_k$ на $l_1, \ldots, l_k$. Это завершает доказательство **18.6.2.**

Распределение (18.6.21) величин $l_1, \ldots, l_k$ иногда называется распределением в *нулевом случае.* Общий случай — это тот, когда матрица ковариаций нормального распределения, из которого извлекается выборка, отлична от диагональной матрицы $\|\sigma_{ij}\|$ в (18.6.20). Распределение собственных значений в этом более общем случае было получено Джеймсом (1960). Оно значительно сложнее, чем распределение (18.6.21).

18.7. Дискриминантный анализ

(а) Случай двух выборок. Задача, которая будет рассматриваться в этом параграфе, состоит в следующем. Имеются две выборки из k-мерных совокупностей. Их можно геометрически представить как две выборочные грозди в k-мерном эвклидовом пространстве. Мы хотим спроектировать эти грозди на такую прямую, чтобы расхождение между проекциями выборок было максимальным (в сравнении с расхождением между проекциями на другие прямые). Задача состоит в том, чтобы найти направление, в котором нужно спроектировать грозди для достижения максимума расхождения между проекциями выборок. Другими словами, мы хотим спроектировать выборочные грозди на одномерное подпространство так, чтобы после проектирования они были удалены настолько, насколько это возможно в соответствии с колебаниями внутри выборок. В практических ситуациях, если мы сумеем найти такое направление, что проекции в этом направлении на одномерное подпространство двух k-мерных выборочных гроздей оказываются достаточно хорошо разделенными, мы получаем тем самым способ для различения выборок из двух совокупностей с помощью подходящих линейных комбинаций компонент наблюдаемых векторов. Эта задача была первоначально рассмотрена Фишером (1938), а метод статистического анализа, который развился из решения этой задачи, получил название *дискриминантного анализа*. Рассмотрим точную формулировку задачи.

Имеются две выборки $(x_{1\xi_\gamma}^{(\gamma)}, \ldots, x_{k\xi_\gamma}^{(\gamma)}$, $\xi_\gamma = 1, \ldots, n_\gamma)$, $\gamma = 1, 2$, где $n_1 > k$, $n_2 > k$. Пусть $(\bar{x}_1^{(\gamma)}, \ldots, \bar{x}_k^{(\gamma)})$, $\gamma = 1, 2$, будут векторы средних этих выборок, а $\|u_{ij}^{(\gamma)}\|$, $\gamma = 1, 2$, — их матрицы внутреннего рассеивания, которые мы предположим неособыми с вероятностью 1. Пусть $(\bar{x}_1, \ldots, \bar{x}_k)$ — вектор средних, $\|u_{ij}\|$ — матрица внутреннего рассеивания выборки, получающейся при объединении двух исходных, и $\|u_{ij}^W\| = \|u_{ij}^{(1)} + u_{ij}^{(2)}\|$ — матрица внутривыборочного рассеивания для двух выборок. Матрица $\|u_{ij}^B\|$ есть матрица межвыборочного рассеивания. Для произвольного вектора $(c_1, \ldots, c_k)$ положим

$$z_{\xi_\gamma}^{(\gamma)} = \sum_{i=1}^{k} c_i x_{i\xi_\gamma}^{(\gamma)}, \quad \xi_\gamma = 1, \ldots, n_\gamma, \qquad \gamma = 1, 2. \tag{18.7.1}$$

Тогда $(z_1^{(1)}, \ldots, z_{n_1}^{(1)})$ и $(z_1^{(2)}, \ldots, z_{n_2}^{(2)})$ будут, с точностью до масштаба, одномерными выборками, полученными соответственно проектированием исходных k-мерных выборок на прямую, направляющие косинусы которой пропорциональны $(c_1, \ldots, c_k)$. Пусть $\bar{z}^{(1)}$ и $\bar{z}^{(2)}$ — средние значения выборок $(z_1^{(1)}, \ldots, z_{n_1}^{(1)})$ и $(z_1^{(2)}, \ldots, z_{n_2}^{(2)})$, а $\bar{z}$ — среднее объединенной выборки $(z_1^{(1)}, \ldots, z_{n_2}^{(2)})$. Положим

$$S_W = \sum_{\gamma=1}^{2} \sum_{\xi_\gamma=1}^{n_\gamma} (z_{\xi_\gamma}^{(\gamma)} - \bar{z}^{(\gamma)})^2, \qquad S_B = \sum_{\gamma=1}^{2} n_\gamma (\bar{z}^{(\gamma)} - \bar{z})^2. \tag{18.7.2}$$

Отметим, что, если S — рассеивание объединенной выборки $(z_1^{(1)}, \ldots$ $\ldots, z_{n_2}^{(2)})$, то $S = S_W + S_B$. S_W представляет собой внутривыборочную компоненту, а S_B — межвыборочную компоненту S. Основная задача теперь состоит в определении $(c_1, \ldots, c_k)$ так, чтобы максимизировать S_B (т. е. минимизировать $S_W/(S_W + S_B)$) при фиксированном значении S_W.

Основной результат здесь может быть сформулирован следующим образом.

18.7.1. *Пусть* $\|u_{ij}^W\|$ *и* $\|u_{ij}^B\|$ — *матрицы внутривыборочного и межвыборочного рассеивания двух выборок из* k-*мерной совокупности. Предположим, что объем каждой выборки больше* k *и матрица* $\|u_{ij}^W\|$ *положительно определена с вероятностью* 1. *Пусть* S_B *и* S_W *определены согласно* (18.7.2). *Значение* $(c_1, \ldots, c_k)$ — *будем обозначать его* $(c_1', \ldots, c_k')$, — *которое минимизирует отношение*

$$Q = \frac{S_W}{S_W + S_B} \tag{18.7.3}$$

при условии, что $S_W = C \neq 0$, *является решением системы уравнений*

$$\sum_{j=1}^{k} (u_{ij}^B - l_1 u_{ij}^W) c_j = 0, \qquad i = 1, \ldots, k, \tag{18.7.4}$$

где l_1 — *ненулевой корень характеристического уравнения*

$$|u_{ij}^B - l u_{ij}^W| = 0, \tag{18.7.5}$$

определяемый формулой

$$l_1 = \frac{n_1 n_2}{n_1 + n_2} \sum_{i,\, j=1}^{k} u_W^{ij} (\bar{x}_i^{(1)} - \bar{x}_i^{(2)}) (\bar{x}_j^{(1)} - \bar{x}_j^{(2)}) = m^2. \tag{18.7.6}$$

Здесь

$$m = \frac{1}{\sqrt{C}} \sum_{i=1}^{k} c_i' b_i, \qquad \|u_W^{ij}\| = \|u_{ij}^W\|^{-1} \tag{18.7.7}$$

и b_i *определены в* (18.4.14). *Минимальное значение* Q *для* $S_W = C \neq 0$ *равно* $1/(1 + l_1)$.

Для доказательства **18.7.1** заметим, что

$$\begin{aligned} S_B &= \sum_{\gamma=1}^{2} n_\gamma \Big[\sum_{i=1}^{k} c_i (\bar{x}_i^{(\gamma)} - \bar{x}_i)\Big]^2 = \sum_{i,\, j=1}^{k} u_{ij}^B c_i c_j, \\ S_W &= \sum_{\gamma=1}^{2} \sum_{\xi_\gamma=1}^{n_\gamma} \Big[\sum_{i=1}^{k} c_i (x_{i\xi_\gamma}^{(\gamma)} - \bar{x}_i^{(\gamma)})\Big]^2 = \sum_{i,\, j=1}^{k} u_{ij}^W c_i c_j. \end{aligned} \tag{18.7.8}$$

Минимизация $S_W/(S_W + S_B)$ при условии $S_W = C \neq 0$ сводится (введением множителя Лагранжа l) к нахождению максимума выражения $\varphi = S_B + (C - S_W) l$ по $(c_1, \ldots, c_k)$ и l. Чтобы найти максимум φ, надо решить уравнения

$$\begin{aligned} &\frac{\partial \varphi}{\partial c_i} = 0, \qquad i = 1, \ldots, k, \\ &\frac{\partial \varphi}{\partial l} = 0, \end{aligned} \tag{18.7.9}$$

которые могут быть записаны как

$$\sum_{j=1}^{k} (u_{ij}^{B} - l u_{ij}^{W}) c_j = 0, \qquad i = 1, \ldots, k. \tag{18.7.10}$$

Чтобы существовало решение $(c_1', \ldots, c_k')$ системы (18.7.10), отличное от $(0, \ldots, 0)$, необходимо выполнение условия:

$$| u_{ij}^{B} - l u_{ij}^{W} | = 0. \tag{18.7.11}$$

Вспоминая из § 18.4 (b), что

$$u_{ij}^{B} = b_i b_j,$$

где, как было показано в (18.4.14),

$$b_i = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} (\bar{x}_i^{(1)} - \bar{x}_i^{(2)}), \tag{18.7.12}$$

мы видим, что (18.7.11) эквивалентно уравнению

$$l^k \left| u_{ij}^{W} - \frac{b_i b_j}{l} \right| = 0.$$

Оно может быть записано как

$$l^{k-1} | u_{ij}^{W} | \cdot \left[l - \sum_{i, j=1}^{k} u_W^{ij} b_i b_j \right] = 0, \tag{18.7.13}$$

где $\| u_W^{ij} \| = \| u_{ij}^{W} \|^{-1}$. Ненулевой корень l_1 уравнения (18.7.11), следовательно, равен

$$l_1 = \sum_{i, j=1}^{k} u_W^{ij} b_i b_j. \tag{18.7.14}$$

Поэтому решение $(c_1', \ldots, c_k')$ (18.7.10) определяется системой

$$\sum_{j=1}^{k} (u_{ij}^{B} - l_1 u_{ij}^{W}) c_j' = 0, \qquad i = 1, \ldots, k. \tag{18.7.15}$$

Умножая i-е уравнение (18.7.15) на c_i и суммируя затем по i, получим

$$\sum_{i, j=1}^{k} u_{ij}^{B} c_i' c_j' - l_1 C = 0, \tag{18.7.16}$$

при этом надо было учесть, что

$$\hat{S}_W = \sum_{i, j=1}^{k} u_{ij}^{W} c_i c_j' = C.$$

Из (18.7.16) следует, что

$$l_1 = \frac{1}{C}\left(\sum_{i=1}^{k} c_i' b_i\right)^2,$$

где b_i дается формулой (18.7.12). Это доказывает (18.7.6).

Подставляя c_i' вместо c_i в (18.7.8) и используя (18.7.6), найдем, что минимальное значение Q равно $1/(1+l_1)$. Доказательство теоремы **18.7.1** закончено.

Задача дискриминантного анализа в случае двух малых выборок из нормальных распределений в пространстве большого числа измерений, т. е. когда n_1 и n_2 меньше k, рассматривалась Демпстером (1958).

(b) Случай нескольких выборок. В задаче о двух выборках мы показали, как надо спроектировать две k-мерные выборки на прямую, чтобы средние значения двух получающихся при этом одномерных выборок были удалены друг от друга на максимально возможное (в соответствии с внутривыборочным рассеиванием одномерных выборок) расстояние. В случае трех выборок мы хотим спроектировать эти три выборки на двумерное подпространство (плоскость) так, чтобы рассеивание объединенной двумерной выборки было максимально возможным в соответствии с внутривыборочным рассеиванием проекций трех выборок. И вообще, если имеются $(s+1)$ k-мерных выборок, $s+1 \leqslant k$, мы хотим спроектировать эти выборки на s-мерное эвклидово подпространство так, чтобы доставить рассеиванию объединенной s-мерной выборки, получившейся при этом, максимально возможное, в соответствии с внутривыборочным рассеиванием $(s+1)$ s-мерных выборок, значение.

Более точно, пусть $(x_{1\xi_\gamma}^{(\gamma)}, \ldots, x_{k\xi_\gamma}^{(\gamma)}; \xi_\gamma = 1, \ldots, n_\gamma)$, $\gamma = 1, \ldots, s+1$, будут $(s+1)$ k-мерных выборок, $s+1 \leqslant k$. Далее, пусть $\|u_{ij}^{(\gamma)}\|$ — матрицы рассеивания этих выборок относительно соответствующих средних $(\bar{x}_1^{(\gamma)}, \ldots, \bar{x}_k^{(\gamma)})$, $\gamma = 1, \ldots, s+1$; $\|u_{ij}\|$ — матрица внутреннего рассеивания объединенной выборки; $\|u_{ij}^W\| = \|u_{ij}^{(1)} + \ldots + u_{ij}^{(s+1)}\|$ — матрица внутривыборочного рассеивания, которую мы предположим неособой с вероятностью 1; $\|u_{ij}^B\| = \|u_{ij} - u_{ij}^W\|$ — матрица межвыборочного рассеивания выборок. Заметим, что если $\|u_{ij}^W\|$ — неособая матрица с вероятностью 1, то этим же свойством обладает и $\|u_{ij}\|$. Определим $z_{\xi_\gamma}^{(\gamma)}$ как в (18.7.1) с $\gamma = 1, \ldots, s+1$, где $(c_{1p}, \ldots, c_{kp})$, $p = 1, \ldots, s+1$, — линейно независимые векторы. Для выборки с номером γ, $\gamma = 1, \ldots, s+1$, пусть координаты проекций выборочных точек будут

$$z_{p\xi}^{(\gamma)} = \sum_{i=1}^{k} c_{ip} x_{i\xi}^{(\gamma)}, \qquad p = 1, \ldots, s, \; \xi = 1, \ldots, n_\gamma. \quad (18.7.17)$$

Обозначим через $\|\tilde{u}_{pq}^{(\gamma)}\|$ матрицу внутреннего рассеивания выборки $(z_{1\xi}^{(\gamma)}, \ldots, z_{s\xi}^{(\gamma)}, \xi = 1, \ldots, n_\gamma)$, $\gamma = 1, \ldots, s+1$, и через $\|\tilde{u}_{pq}\|$ — матрицу внутреннего рассеивания объединения этих выборок. Пусть, наконец, $\|\tilde{u}_{pq}^W\| = \|\tilde{u}_{pq}^{(1)} + \ldots + \tilde{u}_{pq}^{(s+1)}\|$ и $\|\tilde{u}_{pq}^B\| = \|\tilde{u}_{pq} - \tilde{u}_{pq}^W\|$. Наша задача теперь состоит в том, чтобы найти s линейно независимых векторов $(c_{1p}, \ldots, c_{kp})$, $p = 1, \ldots, s$, для которых рассеивание $|\tilde{u}_{pq}|$ достигало бы максимума при фиксированном $|\tilde{u}_{pq}^W|$. Это эквивалентно отысканию направления, в котором надо спроектировать $(s+1)$ выборок на s-мерное подпространство R_s таким образом, чтобы отношение внутривыборочного рассеивания к рассеиванию объединенной выборки, получившейся в R_s, было бы тем же самым, что и отношение внутривыборочного рассеивания к рассеиванию объединенной выборки в R_k. Другими словами, мы хотим найти такие векторы $(c_{1p}, \ldots, c_{kp})$, $p = 1, \ldots, s$, чтобы

$$\frac{|\tilde{u}_{pq}^W|}{|\tilde{u}_{pq}|} = \frac{|u_{ij}^W|}{|u_{ij}|}. \tag{18.7.18}$$

Мы поступим так же, как в случае двух выборок, и для произвольного вектора $(c_1, \ldots, c_k)$ определим S_B и S_W по формулам (18.7.8). Найдем затем значение этого вектора, для которого S_B достигает экстремума при условии, что S_W имеет фиксированное значение $C \neq 0$. Для этого надо решить систему уравнений (18.7.9), где $\varphi = S_B + (C - S_W)\, l$. Она записывается в виде (18.7.10), где, конечно, $\|u_{ij}^B\|$ и $\|u_{ij}^W\|$ — матрицы внутривыборочного и межвыборочного рассеивания для $(s+1)$ выборок. Условие, при котором (18.7.10) имеет ненулевое решение в случае $(s+1)$ выборок, вполне аналогично (18.7.11):

$$|u_{ij}^B - l u_{ij}^W| = 0. \tag{18.7.19}$$

Далее, если $n_1 + \ldots + n_{s+1} - (s+1) \geqslant k$ и $s + 1 \leqslant k$, то можно показать, что ранг $\|u_{ij}^B\|$ равен s и ранг $\|u_{ij}\|$ равен k (с вероятностью 1). Следовательно, $u^W = \|u_{ij}^W\|$ положительно определена, а $u^B = \|u_{ij}^B\|$ положительно полуопределена. Из теории таких матриц известно (см., например, Биркгоф и Маклэйн (1953)), что существуют вещественная неособая матрица $\|e_{ij}\| = e$ и s вещественных чисел, различных с вероятностью 1, $+\infty > l_1 > \ldots > l_s > 0$ с условием

$$e' u^W e = I, \qquad e' u^B e = L, \tag{18.7.20}$$

где L — диагональная матрица $k \times k$ с диагональными элементами $l_1, \ldots, l_s, 0, \ldots, 0$. Умножив матрицу $u^B - l u^W$ слева на e' и справа на e, получим матрицу

$$e' u^B e - l e' u^W e. \tag{18.7.21}$$

Вычислим определитель этой матрицы:

$$|e' u^B e - l e' u^W e| = |L - lI| = l^{k-s}(l_1 - l) \ldots (l_s - l). \tag{18.7.22}$$

Но так как

$$|e'u^Be - le'u^We| = |e'| \cdot |u^B - lu^W| \, |e| \tag{18.7.23}$$

и $|e| = |e'| \neq 0$, то значения l, при которых определитель $|u^B - lu^W|$ обращается в нуль, обращают в нуль и определитель $|e'u^Be - le'u^We|$. Из (18.7.22) мы видим, что ненулевыми корнями последнего (собственными значениями) являются $l_1, \ldots, l_s$. Пусть теперь $(c_{1p}, \ldots, c_{kp})$, $p = 1, \ldots, s$, — решения (собственные векторы) системы (18.7.10) в случае $(s+1)$ выборок; они удовлетворяют условиям:

$$\sum_{j=1}^{k} (u_{ij}^B - l_p u_{ij}^W) c_{jp} = 0, \qquad i = 1, \ldots, k, \tag{18.7.24}$$

$$\sum_{i,j=1}^{k} u_{ij}^W c_{ip} c_{jp} = C. \tag{18.7.24a}$$

Отметим, что (18.7.24a) есть просто уравнение $C - S_W = 0$, где S_W вычислено для $(c_1, \ldots, c_k) = (c_{1p}, \ldots, c_{kp})$.

Покажем, что $(c_{1p}, \ldots, c_{kp})$, $p = 1, \ldots, s$, суть требуемые векторы. Если умножить i-е уравнение из (18.7.24) на c_{ip} и просуммировать по i, приняв во внимание (18.7.24a), то получим

$$\sum_{i,j=1}^{k} u_{ij}^B c_{ip} c_{jp} = l_p C. \tag{18.7.25}$$

Если теперь умножить i-е уравнение из (18.7.24) на c_{iq}, $q \neq p$, и просуммировать по i, то будем иметь

$$\sum_{i,j=1}^{k} u_{ij}^B c_{jp} c_{iq} - l_p \sum_{i,j=1}^{k} u_{ij}^W c_{jp} c_{iq} = 0. \tag{18.7.26}$$

Аналогично

$$\sum_{i,j=1}^{k} u_{ij}^B c_{jq} c_{ip} - l_q \sum_{i,j=1}^{k} u_{ij}^W c_{jq} c_{ip} = 0. \tag{18.7.27}$$

Так как $l_p \neq l_q$, то вычитание (18.7.27) из (18.7.26) дает

$$\sum_{i,j=1}^{k} u_{ij}^W c_{jp} c_{iq} = 0. \tag{18.7.28}$$

Тогда из (18.7.26) или (18.7.27) выводим

$$\sum_{i,j=1}^{k} u_{ij}^B c_{jp} c_{iq} = 0. \tag{18.7.29}$$

Теперь легко получить, что

$$\tilde{u}_{pq}^W = \sum_{i,j=1}^{k} u_{ij}^W c_{ip} c_{jq} = \delta_{pq} C \tag{18.7.30}$$

и

$$\tilde{u}_{pq} = \tilde{u}_{pq}^W + \tilde{u}_{pq}^B = \sum_{i,\,j=1}^{k} (u_{ij}^W + u_{ij}^B)\, c_{ip} c_{jq} = \delta_{pq} C (1 + l_p), \qquad (18.7.31)$$

где $\delta_{pq} = 1$, если $p = q$, и $\delta_{pq} = 0$, если $p \neq q$. Следовательно,

$$\frac{|\tilde{u}_{pq}^W|}{|\tilde{u}_{pq}|} = \frac{1}{(1 + l_1) \ldots (1 + l_s)}. \qquad (18.7.32)$$

Рассмотрим теперь отношение $|u_{ij}^W| / |u_{ij}|$ рассеиваний в исходном k-мерном пространстве. Так как согласно (18.6.20)

$$|e' u^W e| = |e'| \cdot |u^W| \cdot |e| = 1,$$

то $|u_{ij}^W| = 1/|e_{ij}|^2$. Поскольку $|u_{ij}| = |u_{ij}^W + u_{ij}^B|$, то аналогичным образом из (18.6.20) выводим, что $|u_{ij}| = (1 + l_1) \ldots (1 + l_s)/|e_{ij}|^2$. Итак,

$$\frac{|u_{ij}^W|}{|u_{ij}|} = \frac{1}{(1 + l_1) \ldots (1 + l_s)}, \qquad (18.7.33)$$

т. е. отношения рассеиваний, стоящие в левых частях (18.7.32) и (18.7.33), равны между собой.

Суммируем полученные результаты.

18.7.2. *Пусть* $(x_{1\xi_\gamma}^{(\gamma)}, \ldots, x_{k\xi_\gamma}^{(\gamma)};\ \xi_\gamma = 1, \ldots, n_\gamma)$, $\gamma = 1, \ldots, s + 1$, $n_1 + \ldots + n_{s+1} - (s + 1) \geqslant k$, $s + 1 \leqslant k$, — $(s + 1)$ *независимых k-мерных выборок. Обозначим отношение внутривыборочного рассеивания к внутреннему рассеиванию объединенной выборки через* $|u_{ij}^W|/|u_{ij}|$, *где* $|u_{ij}^W| \neq 0$ *с вероятностью 1. Рассмотрим проекцию* (18.7.17) *выборочных точек на s-мерное подпространство и обозначим через* $|\tilde{u}_{pq}^W| / |\tilde{u}_{pq}|$ *отношение внутривыборочного рассеивания проекций выборок к внутреннему рассеиванию объединения проекций выборок. Собственные векторы* $(c_{1p}, \ldots, c_{kp})$, $p = 1, \ldots$ $\ldots, s < k$, *подчиненные условию* (18.7.24а), *для которых отношения рассеиваний* $|u_{ij}^W|/|u_{ij}|$ *и* $|\tilde{u}_{pq}^W| / |\tilde{u}_{pq}|$ *равны, находятся как решения уравнений* (18.7.24), *в которых собственные значения* $+\infty > l_1 > \ldots > l_s > 0$ *являются s ненулевыми корнями уравнения* (18.7.9). *Общее значение отношений* $|u_{ij}^W|/|u_{ij}|$ *и* $|\tilde{u}_{pq}^W| / |\tilde{u}_{pq}|$ *есть* $1/[(1 + l_1) \ldots (1 + l_s)]$. *Собственные значения* l_p, $p = 1, \ldots, s$ *даны в* (18.7.25); *все они попарно различны с вероятностью* 1. *Любые два собственных вектора* $(c_{1p}, \ldots, c_{kp})$ и $(c_{1q}, \ldots, c_{kq})$, $p \neq q$, *удовлетворяют условиям* (18.7.28) *и* (18.7.29).

Главный смысл этой теоремы состоит в том, что если использовать s собственных векторов $(c_{1p}, \ldots, c_{kp})$, $p = 1, \ldots, s$, для линейного отображения (18.7.17) k-мерных выборочных точек в s-мерное подпространство, то отношение рассеиваний $|\tilde{u}_{pq}^W| / |\tilde{u}_{pq}|$ образов точек в R_s будет в точности равно отношению рассеиваний $|u_{ij}^W|/|u_{ij}|$

выборочных точек в исходном k-мерном пространстве R_k. В эквивалентной форме это утверждение можно сформулировать так: если спроектировать k-мерные выборки на s-мерное подпространство, порожденное собственными векторами $(c_{1p}, \ldots, c_{kp})$, $p = 1, \ldots, s$, то рассеивание проекции объединенной выборки $| \tilde{u}_{pq} |$ будет максимально возможным при заданном внутривыборочном рассеивании $| u_{pq}^W |$. Если бы мы захотели спроектировать $(s+1)$ выборок на одномерное подпространство так, чтобы получить наибольшее рассеивание точек на прямой, возможное при заданном внутривыборочном рассеивании проекций выборочных точек, то в (18.7.17) надо было бы использовать собственный вектор $(c_{11}, \ldots, c_{k1})$, отвечающий максимальному собственному значению l_1. Отношение внутривыборочного рассеивания к рассеиванию объединенной выборки, составленной из проекций выборочных точек на прямую, равно $1/(1+l_1)$. Аналогично, если мы хотим спроектировать выборочные точки на t-мерное подпространство, $t < s$, то будем использовать в (18.7.17) собственные векторы $(c_{1p}, \ldots$ $\ldots, c_{kp})$, $p = 1, \ldots, t$, отвечающие собственным числам $l_1, \ldots, l_t$. Отношение внутривыборочного рассеивания к рассеиванию объединенной выборки из проекций выборочных точек на это t-мерное подпространство, равно $1/[(1+l_1)\ldots(1+l_t)]$.

Следует отметить, что, если $(s+1)$ выборок предполагаются извлеченными из k-мерных нормальных совокупностей с одной и той же матрицей ковариаций $\| \sigma_{ij} \|$, то отношение рассеиваний $| u_{ij}^W | / | u_{ij} |$ равно отношению правдоподобия Неймана — Пирсона для проверки гипотезы о равенстве векторов средних нормальных распределений. Кроме того, если эта гипотеза верна, то отношение $| u_{ij}^W | / | u_{ij} |$ имеет распределение, одинаковое с R_s из (18.5.3) при $m = n_1 + \ldots + n_{s+1} -$ $-(s+1)$.

18.8. Распределение собственных значений в дискриминантном анализе

В задаче дискриминантного анализа из предыдущего параграфа собственные числа $l_1 > \ldots > l_s$ и соответствующие собственные векторы $(c_{1p}, \ldots, c_{kp})$, $p = 1, \ldots, s$, играли главную роль. Степень разделения k-мерных выборочных гроздей зависит от величины собственных значений: чем больше собственные значения, тем сильнее разделение. Если выборки извлечены из одинаковых k-мерных нормальных совокупностей, то распределение собственных значений может быть найдено без большого труда. В частном случае двух выборок из одинаковых k-мерных нормальных совокупностей отношение внутривыборочного рассеивания к суммарному равно $1/(1+l_1)$, где l_1 — единственное собственное значение. Но $1/(1+l_1)$ в точности то же, что и величина R'_1, определенная формулой (18.4.15) в §18.4(b). Поэтому $1/(1+l_1)$ имеет одинаковое с R'_1 распределение. Ниже будет показано, что $(n_1+n_2-2)l_1$ есть отношение T^2 Хотеллинга для задачи о двух выборках.

В этом параграфе мы рассмотрим распределение собственных значений $l_1, \ldots, l_s$ в случае, когда $(s+1)$ выборок извлекаются из одинаковых k-мерных нормальных совокупностей. Надо изучить два важных случая: 1) случай $s+1>k$, т. е. число собственных значений равно k; 2) случай $s+1 \leqslant k$, когда имеется s собственных значений.

(а) Случай k собственных значений. При построении теории распределения собственных значений удобно установить сначала следующий общий результат.

18.8.1. *Пусть $\{v_{ij}\}$ и $\{v'_{ij}\}$ — две независимые системы случайных величин, имеющих распределения Уишарта $W(k, n, \|\sigma_{ij}\|)$ и $W(k, n', \|\sigma_{ij}\|)$ соответственно, где $n>k$, $n'>k$. Тогда распределение корней $1 \geqslant g_1 \geqslant \ldots \geqslant g_k \geqslant 0$ уравнения*

$$|v'_{ij}-(v_{ij}+v'_{ij})g|=0 \tag{18.8.1}$$

задается элементом вероятности

$$K\left[\prod_{i=1}^{k}(1-g_i)\right]^{\frac{1}{2}(n-k-1)}\left[\prod_{i=1}^{k}g_i\right]^{\frac{1}{2}(n'-k-1)}\prod_{i>j=1}^{k}(g_j-g_i)\,dg_1\ldots dg_k \tag{18.8.2}$$

в области $0 \leqslant g_k \leqslant g_{k-1} \leqslant \ldots \leqslant g_1 \leqslant 1$ и 0 в остальных точках. Постоянная K равна

$$K=\pi^{\frac{1}{2}k}\prod_{i=1}^{k}\frac{\Gamma\left(\frac{n+n'+1-i}{2}\right)}{\Gamma\left(\frac{n+1-i}{2}\right)\Gamma\left(\frac{n'+1-i}{2}\right)\Gamma\left(\frac{k+1-i}{2}\right)}.$$

Рассуждениями, аналогичными уже использованным при доказательстве **18.6.1**, можно показать, что **18.8.1** эквивалентно утверждению о том, что корни уравнения

$$|v'^*_{ij}-(v^*_{ij}+v'^*_{ij})g|=0 \tag{18.8.1a}$$

имеют элемент вероятности (18.8.2), если только $\{v^*_{ij}\}$ и $\{v'^*_{ij}\}$ представляют собой независимые множества случайных величин с распределениями Уишарта $W(k, n, \|\delta_{ij}\|)$ и $W(k, n', \|\delta_{ij}\|)$. К доказательству этого утверждения мы и приступим.

Так как матрицы $\|v^*_{ij}+v'^*_{ij}\|=v^*+v'^*$ и $\|v'^*_{ij}\|=v'^*$ положительно определены (с вероятностью 1), то существует вещественная неособая матрица $\|e_{ij}\|=e$ и диагональная матрица m с попарно различными (с вероятностью 1) диагональными элементами $1>m_1>\ldots>m_k>0$, такие, что

$$v^*+v'^*=e'e, \qquad v'^*=e'me. \tag{18.8.3}$$

Но уравнения (18.8.3) эквивалентны следующим:

$$v'^*=e'me, \qquad v^*=e'(I-m)e. \tag{18.8.3a}$$

Из (18.8.3) следует, что (18.8.1а) может быть записано как

$$|e'me - ge'e| = 0, \tag{18.8.4}$$

или как

$$|e'|\ |m - gI| \cdot |e| = 0. \tag{18.8.4a}$$

Следовательно, корни $g_1, \ldots, g_k$ равны соответственно $m_1, \ldots, m_k$. Элемент вероятности $\{v_{ij}^*\}$ и $\{v_{ij}'^*\}$ равен

$$A\,|v_{ij}^*|^{\frac{1}{2}(n-k-1)}\,|v_{ij}'^*|^{\frac{1}{2}(n'-k-1)}\,e^{-\frac{1}{2}\sum\limits_{i=1}^{k}(v_{ii}^* + v_{ii}'^*)}\prod_{i \geqslant j=1}^{k} dv_{ij}^* dv_{ij}'^*, \tag{18.8.5}$$

где A — нормирующая постоянная.

Если применить преобразование (18.8.3а) к (18.8.5), то якобиан его будет иметь следующий вид:

$$J = \begin{vmatrix} \overbrace{\dfrac{\partial(v_{ij}^*)}{\partial(e_{i'j'})}}^{k^2\text{ столбцов}} & \overbrace{\dfrac{\partial(v_{ij}^*)}{\partial(m_{i'})}}^{k\text{ столбцов}} \\ \cdots & \\ \dfrac{\partial(v_{ij}'^*)}{\partial(e_{i'j'})} & \dfrac{\partial(v_{ij}'^*)}{\partial(m_{i'})} \end{vmatrix} \begin{array}{l} \Big\}\,\frac{1}{2}k(k+1)\text{ строк,} \\ \\ \Big\}\,\frac{1}{2}k(k+1)\text{ строк.} \end{array} \tag{18.8.6}$$

Прибавим к первой строке верхней половины матрицы первую строку нижней, ко второй строке верхней половины — вторую строку нижней и т. д. Тогда элементы правого верхнего блока J будут иметь вид $\partial(v_{ij}^* + v_{ij}'^*)/\partial m_{i'}$ и в силу (18.8.3) окажутся равными нулю. В этой новой форме якобиана J только элементы левого нижнего блока зависят от $m_1, \ldots, m_k$ и представляют собой однородные линейные формы этих переменных с коэффициентами, зависящими только от e_{ij}. Если теперь разложить J методом Лапласа (см. Бохер (1929)) по верхним $\frac{1}{2}k(k+1)$ строкам, то станет ясно, что любой дополнительный минор, образованный из нижних $\frac{1}{2}k(k+1)$ строк, будет иметь $\frac{1}{2}k(k-1)$ столбцов, выбранных из левого нижнего блока. Поэтому каждый член в разложении Лапласа будет однородным полиномом от $m_1, \ldots, m_k$ степени $\frac{1}{2}k(k-1)$. Если $m_i = m_j$, $i > j$, то преобразование (18.8.3а) неопределенно и $J = 0$; следовательно, $(m_j - m_i)^{a_{ij}}$ входит сомножителем в J, где $a_{ij} \geqslant 1$ — целое положительное число. Таким образом, J имеет вид $G \cdot \prod\limits_{i>j=1}^{k}(m_j - m_i)^{a_{ij}}$, где G

зависит только от e_{ij}. Но из того, что J — полином от $m_1, \ldots, m_k$ степени $\frac{1}{2}k(k-1)$, следует $a_{ij}=1$ для всех i, j. Итак,

$$J = G \prod_{i>j=1}^{k} (m_j - m_i). \tag{18.8.7}$$

Из (18.8.3) и (18.8.3а) мы видим, что

$$\begin{aligned} |v_{ij}'^{*}| &= |e_{ij}|^2 \cdot \prod_{i=1}^{k} m_i, \\ |v_{ij}^{*}| &= |e_{ij}|^2 \prod_{i=1}^{k} (1-m_i), \\ \sum_{i=1}^{k} (v_{ii}^{*} &+ v_{ii}'^{*}) = k. \end{aligned} \tag{18.8.8}$$

Следовательно, применяя формулы (18.8.7) и (18.8.8) к (18.8.5), находим, что $\{e_{ij}\}$ и $\{m_i\}$ являются независимыми множествами случайных величин и элемент вероятности $\{m_i\}$ равен

$$K\left\{\left[\prod_{i=1}^{k}(1-m_i)\right]^{\frac{1}{2}(n-k-1)} \left[\prod_{i=1}^{k} m_i\right]^{\frac{1}{2}(n'-k-1)} \prod_{i>j=1}^{k} (m_j - m_i)\right\} \times \\ \times dm_1 \ldots dm_k \tag{18.8.9}$$

в области $E_k: 1 > m_1 > \ldots > m_k > 0$. Здесь K — постоянная, подлежащая определению.

Обозначим интеграл по области E_k от функции, стоящей в фигурных скобках в (18.8.9), через $\varphi_k\left[\frac{1}{2}(n-k-1), \frac{1}{2}(n'-k-1)\right]$. Тогда

$$\frac{1}{K} = \varphi_k\left(\frac{n-k-1}{2}, \frac{n'-k-1}{2}\right). \tag{18.8.10}$$

Так как $\{v_{ij}^{*}\}$ и $\{v_{ij}'^{*}\}$ распределены независимо по законам Уишарта $W(k, n, \|\delta_{ij}\|)$ и $W(k, n', \|\delta_{ij}\|)$, непосредственно из (18.2.33) получаем

$$\mathscr{E}(|v_{ij}^{*}|^r \cdot |v_{ij}'^{*}|^{-r}) = \prod_{i=1}^{k} \frac{\Gamma\left(\frac{n+1-i}{2}+r\right)\Gamma\left(\frac{n'+1-i}{2}-r\right)}{\Gamma\left(\frac{n+1-i}{2}\right)\Gamma\left(\frac{n'+1-i}{2}\right)}. \tag{18.8.11}$$

Но из (18.8.8) следует, что

$$\mathscr{E}(|v_{ij}^{*}|^r \cdot |v_{ij}'^{*}|^{-r}) = \mathscr{E}\left(\left[\prod_{i=1}^{k}(1-m_i)\right]^r \left[\prod_{i=1}^{k} m_i\right]^{-r}\right) = \\ = \frac{\varphi_k\left(\frac{n-k+1}{2}+r, \frac{n'-k-1}{2}-r\right)}{\varphi_k\left(\frac{n-k-1}{2}, \frac{n'-k-1}{2}\right)}. \tag{18.8.12}$$

Полагая в (18.8.11) и (18.8.12) $r = -\frac{1}{2}(n-k-1)$, будем иметь

$$\frac{\varphi_k\left(0, \frac{n+n'-2k-2}{2}\right)}{\varphi_k\left(\frac{n-k-1}{2}, \frac{n'-k-1}{2}\right)} = \prod_{i=1}^{k} \frac{\Gamma\left(\frac{k+2-i}{2}\right)\Gamma\left(\frac{n+n'-k-i}{2}\right)}{\Gamma\left(\frac{n+1-i}{2}\right)\Gamma\left(\frac{n'+1-i}{2}\right)}. \tag{18.8.13}$$

Задача определения K сводится теперь к вычислению $\varphi_k(0, M)$, где $M = \frac{1}{2}(n+n'-2k-2)$. По определению функции φ_k

$$\varphi_k(0, M) = \int\limits_{E_k} \left(\prod_{i=1}^{k} m_i\right)^M \prod_{i>j=1}^{k} (m_j - m_i)\, dm_1 \ldots dm_k. \tag{18.8.14}$$

Положим $m_1 = t$, $m_2 = tm_2'$, ..., $m_k = tm_k'$. Мы найдем тогда, что

$$\varphi_k(0, M) = \frac{1}{k\left(M + \frac{k+1}{2}\right)} \varphi_{k-1}(1, M). \tag{18.8.15}$$

Из (18.8.13) имеем

$$\frac{\varphi_k(0, M)}{\varphi_k(1, M-1)} = \prod_{i=1}^{k} \frac{M + \frac{k-i}{2}}{\frac{k+2-i}{2}}. \tag{18.8.16}$$

Решая систему разностных уравнений (18.8.15) и (18.8.16), получим

$$\varphi_k\left(0, \frac{n+n'-2k-2}{2}\right) = \\ = \pi^{-\frac{1}{2}k} \prod_{i=1}^{k} \frac{\Gamma\left(\frac{k+2-i}{2}\right)\left(\frac{n+n'-k-i}{2}\right)\Gamma\left(\frac{k+1-i}{2}\right)}{\Gamma\left(\frac{n+n'+1-i}{2}\right)}. \tag{18.8.17}$$

Подставляя это значение $\varphi_k\left(0, \frac{n+n'-2k-2}{2}\right)$ в (18.8.13) и вспоминая, что $K = 1/\varphi_k\left[\frac{1}{2}(n-k-1), \frac{1}{2}(n'-k-1)\right]$, найдем для K значение, приведенное в **18.8.1**.

В задаче дискриминантного анализа преимущественный интерес представляют скорее корни уравнения

$$|v_{ij}' - lv_{ij}| = 0, \tag{18.8.18}$$

чем уравнения (18.8.1).

Но корни уравнений (18.8.18) и (18.8.1) связаны соотношением

$$g_1 = \frac{l_1}{1+l_1}, \ldots, g_k = \frac{l_k}{1+l_k}. \tag{18.8.19}$$

Поэтому, если применить преобразование (18.8.19) к (18.8.2), получим следующий результат.

18.8.2. *Если $\{v_{ij}\}$ и $\{v'_{ij}\}$ — независимые множества случайных величин, имеющих распределения Уишарта $W(k, n, \|\sigma_{ij}\|)$ и $W(k, n', \|\sigma_{ij}\|)$, то корни $l_1 \geqslant \ldots \geqslant l_k \geqslant 0$ уравнения $|v'_{ij} - l v_{ij}| = 0$ имеют элемент вероятности*

$$K\left[\prod_{i=1}^{k}(1+l_i)\right]^{-\frac{1}{2}(n+n')}\left[\prod_{i=1}^{k} l_i\right]^{\frac{1}{2}(n'-k-1)}\prod_{i>j=1}^{k}(l_j - l_i)\,dl_1\ldots dl_k \qquad (18.8.20)$$

в области $0 \leqslant l_k \leqslant l_{k-1} \leqslant \ldots \leqslant l_1 < +\infty$ и 0 в остальных точках. Постоянная K определена в **18.8.1.**

Возвращаясь теперь к задаче дискриминантного анализа для $(s+1)$ выборок, где $s+1 > k$, в качестве следствия **18.8.2** будем иметь

18.8.2a. *Если $\|u^B_{ij}\|$ и $\|u^W_{ij}\|$ — соответственно матрица внутривыборочного и межвыборочного рассеивания $(s+1)$ выборок объемов $n_1, \ldots, n_{s+1}$, где $s+1 > k$ и $n_1 + \ldots + n_{s+1} - (s+1) \geqslant k$, то распределение корней уравнения*

$$|u^B_{ij} - l u^W_{ij}| = 0 \qquad (18.8.21)$$

задается элементом вероятности

$$K_0\left[\prod_{i=1}^{k}(1+l_i)\right]^{-\frac{1}{2}(n+s)}\left[\prod_{i=1}^{k} l_i\right]^{\frac{1}{2}(s-k-1)}\prod_{i>j=1}^{k}(l_j - l_i)\,dl_1\ldots dl_k \qquad (18.8.22)$$

в области $0 \leqslant l_k \leqslant l_{k-1} \leqslant \ldots \leqslant l_1 < +\infty$ и 0 в остальных точках. Постоянная равна

$$K_0 = \pi^{\frac{1}{2}k}\prod_{i=1}^{k}\frac{\Gamma\left(\frac{n+s+1-i}{2}\right)}{\Gamma\left(\frac{n+1-i}{2}\right)\Gamma\left(\frac{s+1-i}{2}\right)\Gamma\left(\frac{k+1-i}{2}\right)}, \qquad (18.8.23)$$

где $n = n_1 + \ldots + n_{s+1} - (s+1)$.

Доказательство этого утверждения сразу получается из **18.8.2**, если заметить, что при сделанных предположениях $\{u^B_{ij}\}$ и $\{u^W_{ij}\}$ являются независимыми множествами случайных величин с распределениями Уишарта $W(k, s, \|\sigma_{ij}\|)$ и $W(k, n, \|\sigma_{ij}\|)$, где $n = n_1 + \ldots + n_{s+1} - (s+1)$.

(b) Случай s собственных значений ($s < k$). Пусть теперь в **18.8.1** $n' < k$; тогда величины $\{v'_{ij}\}$ будут иметь вырожденное распределение Уишарта. В самом деле, $\{v'_{ij}\}$ распределены одинаково с $\left\{\sum_{\xi=1}^{n'} z_{i\xi} z_{j\xi}\right\}$, где $(z_{1\xi}, \ldots, z_{k\xi};\ \xi = 1, \ldots, n')$ — выборка объема n' из k-мерной нормальной совокупности $N(\{0\}, \|\sigma_{ij}\|)$. Уравнение (18.8.1) имеет теперь n' корней $g_1 > \ldots > g_{n'} > 0$, и ясно, что элемент вероятности величин $g_1, \ldots, g_{n'}$ имеет форму (18.8.2) с k, n, n', замененными на n', $n + n' - k$, k соответственно.

Возвращаясь к задаче дискриминантного анализа для $(s+1)$ выборок, где $(s+1) \leqslant k$, мы видим, что в этом случае уравнение (18.8.21) имеет корни $l_1 > \ldots > l_s > 0$, $s < k$. Следовательно, элемент вероятности величин $l_1, \ldots, l_s$ дается выражением (18.8.22) с k, n, s, замененными соответственно на s, $n+s-k$, k, где, конечно, $n = n_1 + \ldots + n_{s+1} - (s+1)$.

Результаты **18.8.1** и **18.8.2** близки к **18.6.2**; они были получены разными авторами, упомянутыми в начале §18.6(b).

18.9. Каноническая корреляция

(а) Определение канонических коэффициентов корреляции. В этом параграфе будем заниматься следующей задачей. Имеется выборка из k-мерной совокупности. Мы хотим найти линейную функцию от первых s компонент и линейную функцию от остальных t компонент $(t+s=k)$ выборочных точек так, чтобы коэффициент корреляции между этими линейными функциями принял наибольшее возможное значение. В практических задачах две линейные функции можно рассматривать как показатели, построенные соответственно по первым s и последним t переменным, и один из них обычно используется для предсказания или оценки другого. Естественно выяснить, как следует строить эти две линейные функции, чтобы обычный коэффициент корреляции между ними принял максимально возможное значение.

Пусть $(x_{i\xi};\ i=1, \ldots, k;\ \xi=1, \ldots, n)$, $n > k$, — выборка объема n из k-мерной совокупности, $\|u_{ij}\|$ — матрица внутреннего рассеивания этой выборки. Пусть $(x_{p\xi};\ p=1, \ldots, s;\ \xi=1, \ldots, n)$ будут первые s компонент выборки, а $|u_{pq}|$, $p, q=1, \ldots, s$, — матрица внутреннего рассеивания этих компонент; аналогично пусть $(x_{v\xi};\ v=s+1, \ldots, k;\ \xi=1, \ldots, n)$ — последние t компонент $(s+1=k$ и $t \geqslant s > 0)$ и $\|u_{vw}\|$, $v, w=s+1, \ldots, k$, — матрица внутреннего рассеивания этих компонент. Отметим, что

$$\|u_{ij}\| = \left\|\begin{array}{c:c} u_{pq} & u_{pw} \\ \hdashline u_{vq} & u_{vw} \end{array}\right\|. \tag{18.9.1}$$

Предположим, что матрица $\|u_{ij}\|$ неособая с вероятностью 1. Возьмем вещественные векторы $(c_{1p};\ p=1, \ldots, s)$ и $(c_{2v};\ v=s+1, \ldots, k)$ и положим

$$z_{1\xi} = \sum_{p=1}^{s} c_{1p} x_{p\xi}, \qquad z_{2\xi} = \sum_{v=s+1}^{k} c_{2v} x_{v\xi}, \tag{18.9.2}$$

$\xi = 1, \ldots, n$. Пусть

$$\left\|\begin{matrix} \tilde{u}_{11} & \tilde{u}_{12} \\ \tilde{u}_{21} & \tilde{u}_{22} \end{matrix}\right\|$$

— матрица рассеивания выборки ($z_{1\xi}$, $z_{2\xi}$; $\xi=1, \ldots, n$) относительно ее среднего. Наша задача состоит в том, чтобы определить векторы (c_{1p}; $p=1, \ldots, s$) и (c_{2v}; $v=s+1, \ldots, k$) при некоторых условиях нормировки, которые без потери общности можно взять в виде $\tilde{u}_{11}=\tilde{u}_{22}=1$, так чтобы коэффициент корреляции

$$R=\frac{\tilde{u}_{12}}{\sqrt{\tilde{u}_{11}\tilde{u}_{22}}} \tag{18.9.3}$$

был максимальным.

Решение этой задачи принадлежит Хотеллингу (1935) и может быть сформулировано в виде следующего утверждения.

18.9.1. *Собственные векторы* ($c_{1p}^{(1)}$; $p=1, \ldots, s$) *и* ($c_{2v}^{(1)}$; $v=s+1, \ldots, k$), *которые максимизируют* R *при условиях* $\tilde{u}_{11}=\tilde{u}_{22}=1$, *являются решениями уравнений*

$$\begin{aligned} -\sqrt{l_1}\sum_{q=1}^{s} u_{pq}c_{1q}+\sum_{w=s+1}^{k} u_{pw}c_{2w}&=0, \qquad p=1, \ldots, s,\\ \sum_{q=1}^{s} u_{vq}c_{1q}-\sqrt{l_1}\sum_{w=s+1}^{k} u_{vw}c_{2w}&=0, \qquad v=s+1, \ldots, k, \end{aligned} \tag{18.9.4}$$

где l_1 — *наибольший корень уравнения*

$$\begin{vmatrix} lu_{pq} & -u_{pw} \\ -u_{vq} & u_{vw} \end{vmatrix}=0. \tag{18.9.5}$$

Максимальное значение R *равно* $\sqrt{l_1}$.

Вообще, уравнение (18.9.5) *имеет* s *корней* $l_1, \ldots, l_s$ *в интервале* (0,1). *Предположим, что выборка* ($x_{i\xi}$; $i=1, \ldots, k$; $\xi=1, \ldots, n$) *извлечена из такой совокупности, что эти корни попарно различны с вероятностью* 1; *в этом случае их можно расположить в виде* $1>l_1>\ldots>l_s>0$. *Собственные векторы* ($c_{1p}^{(g)}$; $p=1, \ldots, s$) *и* ($c_{2v}^{(g)}$; $v=s+1, \ldots, k$) *определим как решения уравнений* (18.9.4) *с* l_1, *замененным на* l_g, $g=1, \ldots, s$. *Значение* R, *отвечающее этим собственным векторам, равно* $\sqrt{l_g}$. *Значение* R, *полученное при использовании собственных векторов* ($c_{1p}^{(g)}$; $p=1, \ldots, s$) и ($c_{2v}^{(g')}$; $v=s+1, \ldots, k$), $g\neq g'$, *есть* 0.

Коэффициенты корреляции $R^{(1)}, \ldots, R^{(s)}$, имеющие соответственно значения $\sqrt{l_1}, \ldots, \sqrt{l_s}$, называются *каноническими коэффициентами корреляции* между первыми s компонентами ($x_{1\xi}, \ldots, x_{s\xi}$; $\xi=1, \ldots, n$) и последними t компонентами ($x_{s+1,\xi}, \ldots, x_{k\xi}$; $\xi=1, \ldots, n$) выборки ($x_{1\xi}, \ldots, x_{k\xi}$; $\xi=1, \ldots, n$). Наибольший практический интерес представляет, конечно, максимальный канонический коэффициент корреляции $R^{(1)}$.

Чтобы доказать **18.9.1**, поступим следующим образом. Можно проверить, что

$$\tilde{u}_{11}=\sum_{p,q} u_{pq}c_{1p}c_{1q}, \quad \tilde{u}_{22}=\sum_{v,w} u_{vw}c_{2v}c_{2w}, \quad \tilde{u}_{12}=\sum_{p,w} u_{pw}c_{1p}c_{2w}, \tag{18.9.6}$$

где p, q пробегают числа $1, \ldots, s$, а v, w — числа $s+1, \ldots, k$. Рассмотрим теперь векторы $(c_{1p};\ p=1, \ldots, s)$ и $(c_{2v};\ v=s+1, \ldots, k)$, которые доставляют экстремум R при условиях $\tilde{u}_{11}=\tilde{u}_{22}=1$. Если ввести множители Лагранжа λ и μ, то эти же векторы будут доставлять экстремум функции

$$\varphi=\tilde{u}_{12}+(1-\tilde{u}_{11})\lambda+(1-\tilde{u}_{22})\mu \tag{18.9.7}$$

и могут быть найдены как решения уравнений

$$\begin{aligned} &\frac{\partial\varphi}{\partial c_{1p}}=0, \qquad p=1, \ldots, s, \\ &\frac{\partial\varphi}{\partial c_{2v}}=0, \qquad v=s+1, \ldots, k. \end{aligned} \tag{18.9.8}$$

Но (18.9.8) записываются в виде

$$\begin{aligned} &-\lambda\sum_{q} u_{pq}c_{1q}+\sum_{w} u_{pw}c_{2w}=0, \ p=1, \ldots, s, \\ &\sum_{q} u_{vq}c_{1q}-\mu\sum_{w} u_{vw}c_{2w}=0, \qquad v=s+1, \ldots, k. \end{aligned} \tag{18.9.9}$$

Умножая p-е из верхних уравнений (18.9.9) на c_{1p} и суммируя по p от 1 до s, а v-е из нижних уравнений — на c_{2v} и суммируя по v от $s+1$ до k, мы видим, что $\lambda=\mu$. Поэтому в (18.9.9) можно заменить μ на λ. Для существования нетривиальных решений уравнения (18.9.9) необходимо, чтобы

$$\begin{vmatrix} -\lambda u_{pq} & u_{pw} \\ u_{vq} & -\lambda u_{vw} \end{vmatrix}=0. \tag{18.9.10}$$

Но это уравнение легко приводится к виду

$$(-1)^k\lambda^{k-2s}\begin{vmatrix} l u_{pq} & -u_{pw} \\ -u_{vq} & u_{vw} \end{vmatrix}=0, \tag{18.9.11}$$

где $l=\lambda^2$. Определитель в уравнении (18.9.11) совпадает с определителем из (18.9.5) и представляет собой полином относительно l степени s. Можно показать, что все его корни вещественны и положительны; обозначим их $l_1, \ldots, l_s$. Собственные векторы $(c_{1p}^{(g)};\ p=1,\ldots$

$\ldots, s)$ и $(c_{2v}^{(g)};\ v=s+1, \ldots, k)$, отвечающие l_g, являются решениями уравнений

$$\begin{gathered}-\sqrt{l_g}\sum_q u_{pq}c_{1q}+\sum_w u_{pw}c_{2w}=0,\quad p=1, \ldots, s,\\ \sum_q u_{vq}c_{1q}-\sqrt{l_g}\sum_w u_{vw}c_{2w}=0,\quad v=s+1, \ldots, k.\end{gathered} \tag{18.9.12}$$

Если подставить эти векторы в (18.9.12), умножить p-е из верхних уравнений на $c_{1q}^{(g)}$ и просуммировать по p от 1 до s, то, учитывая, что $\sum_{p,q} u_{pq}c_{1p}^{(g)}c_{1q}^{(g)}=1$, получим

$$\sum_{p,w} u_{pw}c_{2w}^{(g)}c_{1p}^{(g)}=\sqrt{l_g}. \tag{18.9.13}$$

Но в силу условий $\tilde{u}_{11}=\tilde{u}_{22}=1$ левая часть (18.9.13) есть значение коэффициента корреляции R между z_1 и z_2 из (18.9.2), если в качестве $(c_{1p};\ p=1, \ldots, s)$ и $(c_{2v};\ v=s+1, \ldots, k)$ взяты собственные векторы, определяемые уравнениями (18.9.12). Следовательно, $l_g<1$, $g=1, \ldots, s$, и корни $l_1, \ldots, l_s$, которые по нашему предположению попарно различны с вероятностью 1 и могут быть расположены в виде $l_1>\ldots>l_s$, лежат на интервале (0, 1). Так как l_1 — наибольший корень, то собственные векторы, отвечающие ему, являются решениями уравнений (18.9.4). Можно показать, что при объеме выборки $n>k$ $l_1=1$ с вероятностью 1 тогда и только тогда, когда первые s и последние t $(s+t=k)$ случайных величин k-мерной совокупности, из которой извлечена выборка, линейно зависимы.

Чтобы установить равенство нулю коэффициента корреляции между z_1 и z_2 для любых двух различных собственных векторов $(c_{1p}^{(g)};\ p=1,\ldots$ $\ldots, s)$ и $(c_{2v}^{(g')};\ v=s+1, \ldots, k)$, $g\neq g'$, рассмотрим уравнения (18.9.12), в которые уже подставлены векторы $(c_{1p}^{(g)};\ p=1, \ldots, s)$ и $(c_{2v}^{(g')};\ v=$ $=s+1, \ldots, k)$. Умножим p-е уравнение первой группы на $c_{1p}^{(g)}$, v-е уравнение второй — на $c_{2v}^{(g')}$ и просуммируем по p и v. Затем повторим эту операцию, поменяв местами g и g'. Так как $l_g\neq l_{g'}$, то из подходящей комбинации уравнений можно вывести, что $\sum_{p,w} u_{pw}c_{1p}^{(g)}c_{2w}^{(g')}=0$.

Так что коэффициент корреляции R, получаемый при использовании векторов $(c_{1p}^{(g)};\ p=1, \ldots, s)$ и $(c_{2v}^{(g')};\ v=s+1, \ldots, k)$, обращается в нуль при $g\neq g'$. Аналогичным образом можно получить, что при $g\neq g'$

$$\sum_{p,q} u_{pq}c_{1p}^{(g)}c_{1q}^{(g')}=\sum_{v,w} u_{vw}c_{2v}^{(g)}c_{2w}^{(g')}=0.$$

(b) Выборочная теория канонических коэффициентов корреляции. Получить распределение квадратов канонических коэффициентов корреляции выборки $l_1, \ldots, l_s$ в общих условиях очень сложно. Но при некоторых ограничениях распределение величин $l_1, \ldots, l_s$ оказывается специальным случаем распределения (18.8.2). Точнее, имеет место следующий результат.

18.9.2. *Пусть* $(x_{p\xi};\ p=1, \ldots, s;\ \xi=1, \ldots, n)$ *и* $(x_{v\xi};\ v=s+1, \ldots, k;\ \xi=1, \ldots, n)$ *— независимые выборки, первая — из s-мерной нормальной совокупности $N(\{\mu_p\}, \|\sigma_{pq}\|)$, вторая — из произвольной t-мерной совокупности. Если $s \leqslant t$ и матрица внутреннего рассеивания второй выборки неособая с вероятностью* 1, *то корни* $1 \geqslant l_1 \geqslant \ldots \geqslant l_s \geqslant 0$ *уравнения* (18.9.5) *имеют распределение, элемент вероятности которого равен*

$$K^* \left[\prod_{p=1}^{s}(1-l_p)\right]^{\frac{1}{2}(n-t-s-2)} \left[\prod_{p=1}^{s} l_p\right]^{\frac{1}{2}(t-s-1)} \times$$
$$\times \prod_{p>q=1}^{s}(l_q-l_p)\, dl_1 \ldots dl_s \qquad (18.9.14)$$

в области $0 \leqslant l_s \leqslant l_{s-1} \leqslant \ldots \leqslant l_1 \leqslant 1$ *и* 0 *в остальных точках; при этом*

$$K^* = \pi^{\frac{1}{2}s} \prod_{p=1}^{s} \frac{\Gamma\left(\frac{n-p}{2}\right)}{\Gamma\left(\frac{n-t-p}{2}\right)\Gamma\left(\frac{t+1-p}{2}\right)\Gamma\left(\frac{s+1-p}{2}\right)}. \qquad (18.9.15)$$

При доказательстве **18.9.2** достаточно рассмотреть случай, когда все $(x_{v\xi};\ v=s+1, \ldots, k;\ \xi=1, \ldots, n)$ фиксированы (и матрица $\|u_{vw}\|$ для этого набора неособая). Мы покажем, что в этом случае $l_1, \ldots, l_s$ имеют распределение (18.9.14). Отсюда, конечно, будет следовать, что этот результат верен и тогда, когда $(x_{v\xi};\ v=s+1, \ldots, k;\ \xi=1, \ldots, n)$ — случайная выборка из произвольной совокупности, для которой матрица $\|u_{vw}\|$ неособая с вероятностью 1.

Начнем доказательство **18.9.2** с того, что умножим обе части уравнения (18.9.5) (слева) на определитель

$$\begin{vmatrix} \delta_{pq} & \sum_v u_{pv} u^{vw} \\ \cdots & \cdots \\ 0 & u^{vw} \end{vmatrix},$$

где $\|u^{vw}\| = \|u_{vw}\|^{-1}$ и δ_{pq} — символ Кронекера. Тогда после умножения будем иметь

$$\begin{vmatrix} l u_{pq} - \sum_{v,w} u^{vw} u_{pv} u_{wq} & -u_{pw} + \sum u^{vw'} u_{pv} u_{w'w} \\ \cdots & \cdots \\ -\sum_v u_{vq} u^{vw} & \sum_{w'} u^{vw'} u_{w'w} \end{vmatrix} = 0, \qquad (18.9.16)$$

где p, q пробегают числа $1, \ldots, s$, а v, w, w' — числа $s+1, \ldots, k$. Так как $\sum\limits_{w'} u^{vw'}u_{w'w} = \delta_{vw}$, то ясно, что каждый элемент в правом верхнем блоке определителя из (18.9.16) есть 0. Поэтому корни уравнения (18.9.5) совпадают с корнями уравнения

$$\Big|\sum_{v,\,w} u^{vw}u_{pv}u_{wq} - lu_{pq}\Big| = 0. \qquad (18.9.17)$$

Обозначим теперь $x_{p\xi} - \bar{x}_p$ через $y_{p\xi}$ и $x_{v\xi} - \bar{x}_v$ через $y_{v\xi}$. Тогда

$$u_{pq} = \sum_{\xi=1}^{n} y_{p\xi}y_{q\xi} = u_{pq}^{(1)} + u_{pq}^{(2)}, \qquad (18.9.18)$$

где

$$u_{pq}^{(1)} = \sum_{\xi=1}^{n}\Big(y_{p\xi} - \sum_{w} b_{pw}y_{w\xi}\Big)\Big(y_{q\xi} - \sum_{v} b_{qv}y_{v\xi}\Big),$$

$$u_{pq}^{(2)} = \sum_{v,\,w=s+1}^{k} u^{vw}b_{pw}b_{qv},$$

$$b_{pw} = \sum_{v=s+1}^{k} u^{vw}u_{pw}.$$

Подставляя выражение для b_{pw} в выражение для $u_{pq}^{(2)}$, найдем, что $u_{pq}^{(2)} = \sum\limits_{v,\,w} u^{vw}u_{pv}u_{wq}$. Следовательно, (18.9.17) можно записать в виде

$$|u_{pq}^{(2)} - (u_{pq}^{(1)} + u_{pq}^{(2)})\,l| = 0. \qquad (18.9.19)$$

Можно проверить, что в условиях теоремы **18.9.2** $\{u_{pq}^{(1)}\}$ и $\{u_{pq}^{(2)}\}$ будут независимыми множествами случайных величин, имеющими распределения Уишарта $W(s, n-t-1, \|\sigma_{pq}\|)$ и $W(s, t, \|\sigma_{pq}\|)$. Из **18.8.1** тогда следует, что корни уравнения (18.9.8) имеют распределение (18.8.2) с заменой k на s, n' на t и n на $n-t-1$. При этом получится распределение (18.9.14). Доказательство **18.9.2** закончено.

ЗАДАЧИ

18.1. Пусть $(x_{1\xi}, \ldots, x_{k\xi};\ \xi = 1, \ldots, n)$ — выборка объема $n\ (n > k)$ из нормальной совокупности $N(\{\mu_i\}, \|\sigma_{ij}\|)$. Показать, что оценка максимального правдоподобия для $(\mu_1, \ldots, \mu_k)$ есть $(\bar{x}_1, \ldots, \bar{x}_k)$, а для $\|\sigma_{ij}\|$ — $\|u_{ij}/n\|$, где $(\bar{x}_1, \ldots, \bar{x}_k)$ — вектор выборочных средних и $\|u_{ij}\|$ — матрица внутреннего рассеивания выборки.

18.2. Показать с помощью характеристических функций, что если симметрическая матрица $\|a_{ij}\|$, $i, j = 1, \ldots, k < m$, составлена из случайных величин, имеющих совместное распределение Уишарта $W(k, m, \|\sigma_{ij}\|)$, то элементы матрицы $\|a_{pq}\|$, $p, q = 1, \ldots, s < k$, имеют распределение Уишарта $W(s, m, \|\sigma_{pq}\|)$.

18.3. *Распределение коэффициентов корреляции в выборке из k-мерного нормального распределения с независимыми компонентами.* Покажите, что если выборка $(x_{1\xi}, \ldots, x_{k\xi}, \xi = 1, \ldots, n)$, $n > k$, извлечена из нормальной совокупности $N(\{\mu_i\}, \|\sigma_{ij}\delta_{ij}\|)$ и $\|u_{ij}\|$ — матрица внутреннего рассеивания выборки, то распределение элементов корреляционной матрицы $\|r_{ij}\|$, $r_{ij} = \dfrac{u_{ij}}{\sqrt{u_{ii}u_{jj}}}$, задается п. в.

$$\frac{\Gamma^k\left(\frac{n-1}{2}\right)|r_{ij}|^{\frac{1}{2}(n-k-2)}}{\pi^{k(k-1)/4}\prod_{i=1}^{k}\Gamma\left(\frac{n-i}{2}\right)}$$

в тех точках пространства $\{r_{ij}\}$, в которых $\|r_{ij}\|$ положительно определена, и 0 в остальных точках.

18.4. (*Продолжение*). Показать, что

$$\mathscr{E}(|r_{ij}|^g) = \prod_{i=2}^{k}\left[\frac{\Gamma\left(\frac{n-i}{2}+g\right)\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-i}{2}\right)\Gamma\left(\frac{n-1}{2}+g\right)}\right], \qquad g = 0, 1, 2, \ldots$$

и, следовательно, распределение $|r_{ij}|$ одинаково с распределением $\prod_{i=2}^{k} z_i$, где z_i — независимые случайные величины с бета-распределениями Ве $\left(\frac{1}{2}(n-i), \frac{1}{2}(i-1)\right)$, $i = 2, \ldots, k$ соответственно.

18.5. *Распределение выборочного коэффициента корреляции.* Пусть $\|u_{ij}\|$ — матрица внутреннего рассеивания выборки объема n из двумерной совокупности $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j = 1, 2$. Используя то обстоятельство, что элементы матрицы $\|u_{ij}\|$ имеют совместное распределение Уишарта $W(2, n-1, \|\sigma_{ij}\|)$, покажите, что п. в. выборочного коэффициента корреляции $r = u_{12}/\sqrt{u_{11}u_{22}}$ равна

$$\frac{(1-\rho)^{\frac{1}{2}(n-1)}(1-r^2)^{\frac{1}{2}(n-4)}}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}\sum_{i=0}^{\infty}\frac{(2\rho r)^i}{i!}\Gamma^2\left(\frac{n-1+i}{2}\right)$$

для $-1 \leqslant r \leqslant +1$ и 0 в противном случае, где $\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{12}}$ — коэффициент корреляции в исходной совокупности [Фишер (1915)].

18.6. (*Продолжение*). Преобразуя элемент вероятности распределения Уишарта $W(2, n-1, \|\sigma_{ij}\|)$ по формулам

$$r = u_{12}/\sqrt{u_{11}u_{22}}, \quad s = \sqrt{u_{11}u_{22}}, \quad t = \frac{1}{2}\log\frac{u_{22}}{u_{11}},$$

покажите, что п. в. r может быть также записана в форме (принадлежащей Хотеллингу (1953))

$$\frac{(n-2)\,\Gamma(n-1)\,(1-\rho^2)^{\frac{1}{2}(n-1)}(1-r^2)^{\frac{1}{2}(n-4)}(1-\rho r)^{-\frac{1}{2}(2n-3)}}{\sqrt{2}\,\pi^{\frac{3}{2}}}\times$$

$$\times\sum_{i=0}^{\infty}\frac{\left(\frac{1}{2}+\frac{1}{2}\rho r\right)^i\Gamma^2\left(\frac{1}{2}+i\right)}{\left(i!\,\Gamma n-\frac{1}{2}+i\right)}.$$

18.7. (*Продолжение*). Показать, что принадлежащая Хотеллингу форма элемента вероятности r может быть преобразована к виду

$$\frac{(n-2)\,\Gamma(n-1)\,(1-\rho^2)^{\frac{1}{2}(n-1)}(1-r^2)^{\frac{1}{2}(n-4)}}{\sqrt{2}\,\pi\Gamma\left(n-\frac{1}{2}\right)(1-\rho r)^{\frac{1}{2}(2n-3)}}\times$$

$$\times\left\{\int_0^1\int_0^1 f(x)\,g(y)\left[1-\frac{1}{2}(1+\rho r)\,xy\right]^{-1}dy\,dx\right\}dr,$$

где $f(x)$ и $g(y)$ — п. в. бета-распределений $\mathrm{Be}\left(\frac{1}{2},\,n-1\right)$ и $\mathrm{Be}\left(\frac{1}{2},\,\frac{1}{2}\right)$ соответственно. Отсюда, используя преобразование

$$r=\rho+\frac{1-\rho^2}{\sqrt{n}}\,u,\quad x=\frac{v}{n},\qquad y=w,$$

получить, что предельным распределением для

$$u=\frac{(r-\rho)\sqrt{n}}{1-\rho^2}$$

при $n\to\infty$ будет $N(0,\,1)$.

18.8. (*Продолжение*). Покажите с помощью **9.3.1**, что если

$$z=\frac{1}{2}\log\frac{1+r}{1-r}$$

и

$$\zeta=\frac{1}{2}\log\frac{1+\rho}{1-\rho},$$

то предельное распределение для $(z-\zeta)\sqrt{n}$ при $n\to\infty$ есть $N(0,\,1)$.

18.9. *Доверительные эллипсоиды для вектора средних нормального распределения.* Если $(\bar{x}_1,\,\ldots,\,\bar{x}_k)$ — вектор средних и $\|u_{ij}\|$ — матрица внутреннего рассеивания выборки объема n, $n>k$, из k-мерной нормальной совокупности $N(\{\mu_i\},\,\|\sigma_{ij}\|)$, то покажите, что эллипсоид

$$n\sum_{i,\,j=1}^{k}u^{ij}(\mu_i-\bar{x}_i)(\mu_j-\bar{x}_j)-(1<z_\gamma)/z_\gamma$$

будет 100γ-процентным доверительным эллипсоидом для $(\mu_1, \ldots, \mu_k)$. Здесь z_γ — 100γ-процентная квантиль бета-распределения $\mathrm{Be}\left(\frac{1}{2}(n-k), \frac{1}{2}k\right)$ [Хотеллинг (1931)].

18.10. *Эквивалентность отношения T^2 Хотеллинга отношению правдоподобия для гипотезы о равенстве вектора средних нормального распределения $N(\{\mu_i\}, \|\sigma_{ij}\|)$ заданному вектору.* Пусть $\mathcal{H}(\omega; \Omega)$ — гипотеза, в которой Ω — $\frac{1}{2}k(k+3)$-мерное параметрическое пространство точек $(\{\mu_i\}, \{\sigma_{ij}\})$ таких, что $\|\sigma_{ij}\|$ положительно определена; ω — подмножество Ω, в точках которого $\mu_1=\mu_1^0, \ldots, \mu_k=\mu_k^0$. Покажите, что отношение правдоподобия λ для проверки гипотезы $\mathcal{H}$ по выборке объема n из совокупности $N(\{\mu_i\}, \|\sigma_{ij}\|)$ есть

$$\lambda = R_1^{\frac{1}{2}n},$$

где R_1 определено в (18.4.1). Отсюда R_1 (так же, как и T^2 Хотеллинга) эквивалентно λ и, когда $\mathcal{H}$ верна, R_1 имеет бета-распределение

$$\mathrm{Be}\left(\frac{1}{2}(n-k), \frac{1}{2}k\right).$$

18.11. (*Продолжение*). *Эквивалентность отношения T^2 Хотеллинга и D^2 Махаланобиса отношению правдоподобия для гипотезы о равенстве векторов средних двух нормальных распределений с одинаковыми матрицами ковариаций.* Пусть $\mathcal{H}(\omega; \Omega)$ — гипотеза, в которой Ω — $\frac{1}{2}k(k+5)$-мерное пространство параметрических точек $(\{\mu_i^{(1)}\}, \{\mu_i^{(2)}\}, \{\sigma_{ij}\})$, где $\|\sigma_{ij}\|$ положительно определена: ω — $\frac{1}{2}k(k+3)$-мерное подпространство Ω, в точках которого $\mu_i^{(1)}=\mu_i^{(2)}$, $i=1, \ldots, k$. Показать, что отношение правдоподобия λ для проверки гипотезы $\mathcal{H}$ по независимым выборкам объемов n_1 и n_2 из k-мерных нормальных совокупностей

$$N(\{\mu_i^{(i)}\}, \|\sigma_{ij}\|) \text{ и } N(\{\mu_i^{(2)}\}, \|\sigma_{ij}\|)$$

соответственно, $n_1>k$, $n_2>k$, равно

$$\lambda = (R_1')^{\frac{1}{2}(n_1+n_2)}.$$

Следовательно, R_1' (так же как T^2 Хотеллинга и D^2 Махаланобиса для двух выборок) эквивалентно λ и распределение R_1', когда $\mathcal{H}$ верна, есть

$$\mathrm{Be}\left(\frac{1}{2}(n_1+n_2-k-1), \frac{1}{2}k\right).$$

18.12. Обобщите задачу **18.11** на случай $(s+1)$ выборок объемов $n_1, \ldots, n_{s+1}$ из нормальных совокупностей с общей матрицей ковариаций и докажите, что отношение правдоподобия λ для проверки гипотезы $\mathcal{H}$ о равенстве векторов средних всех исходных совокупностей эквивалентно R_s из (18.5.11), где $\|a_{ij}\|$ — сумма матриц внутреннего рассеивания выборок, $\left\|a_{ij}+\sum_{\beta=1}^{s} b_{i\beta}b_{j\beta}\right\|$ — матрица внутреннего рассеивания объединенной выборки и $m=n_1+\ldots+n_{s+1}-s-1$. Выведите отсюда утверждение **18.5.1**.

18.13. *Проверка гипотезы о равенстве матриц ковариаций двух нормальных распределений.* В задаче **18.11** рассмотрите гипотезу $\mathscr{H}(\omega;\ \Omega)$, где множество Ω прежнее, а ω — подмножество Ω, в точках которого $\|\sigma_{ij}^{(1)}\| = \|\sigma_{ij}^{(2)}\|$. Покажите, что отношение правдоподобия λ для гипотезы $\mathscr{H}$ равно

$$\lambda = \left|\frac{u_{ij}^{(1)}}{n_1}\right|^{\frac{1}{2}n_1} \left|\frac{u_{ij}^{(2)}}{n_2}\right|^{\frac{1}{2}n_2} \Bigg/ \left|\frac{u_{ij}^{(1)} + u_{ij}^{(2)}}{n}\right|^{\frac{1}{2}n},$$

где $n = n_1 + n_2$, и когда $\mathscr{H}$ верна,

$$\mathscr{E}(\lambda^r) = \left[\left(\frac{n}{n_1}\right)^{n_1}\left(\frac{n}{n_2}\right)^{n_2}\right]^{\frac{1}{2}kr} \times$$

$$\times \prod_{i=1}^{k}\left[\frac{\Gamma\left(\frac{n-1-i}{2}\right)\Gamma\left(\frac{n_1-i+n_1 r}{2}\right)\Gamma\left(\frac{n_2-i+n_2 r}{2}\right)}{\Gamma\left(\frac{n-1+i+nr}{2}\right)\Gamma\left(\frac{n_1-i}{2}\right)\Gamma\left(\frac{n_2-i}{2}\right)}\right],$$

$r = 0,\ 1,\ 2,\ \ldots$

18.14. *Модель* I *дисперсионного анализа для векторов в двухфакторном плане эксперимента.* Предположим, что в схеме эксперимента, представленной в виде r строк $R_1, \ldots, R_r$ и s столбцов $C_1, \ldots, C_s$ с ячейкой, образованной ξ-й строкой и η-м столбцом, $\xi = 1, \ldots, r$; $\eta = 1, \ldots, s$, связывается k-мерный случайный вектор $(x_{1\xi\eta}, \ldots, x_{k\xi\eta})$. Пусть эти rs случайных векторов независимы и $(x_{1\xi\eta}, \ldots, x_{k\xi\eta})$ имеет k-мерное нормальное распределение

$$N(\{\mu_i + \mu_{i\xi\cdot} + \mu_{i\cdot\eta}\},\ \|\sigma_{ij}\|),$$

где $\sum\limits_{\xi}\mu_{i\xi\cdot} = \sum\limits_{\eta}\mu_{i\cdot\eta} = 0$, $i = 1, \ldots, k$. Обозначим через $\mathscr{H}(\omega;\ \Omega)$ статистическую гипотезу, в которой Ω — пространство $k(r+s-1) + \frac{1}{2}k(k+1)$ независимых параметров, входящих в указанные выше rs распределений, а ω — $\left(ks + \frac{1}{2}k(k+1)\right)$-мерное подпространство Ω, на котором отсутствуют «эффекты по строкам», т. е.

$$\mu_{i1\cdot} = \mu_{i2\cdot} = \ldots = \mu_{ir\cdot} = 0, \qquad i = 1, \ldots, k.$$

Обозначим вычисленные по $x_{i\xi\eta}$, $\xi = 1, \ldots, r$, $\eta = 1, \ldots, s$, величины $\bar{x}_{\cdot\cdot}, \ldots, \bar{x}_{\xi\cdot}, \bar{x}_{\cdot\eta}, m, m_{\xi\cdot}, m_{\cdot\eta}$, определенные в (10.6.3), через $\bar{x}_{i\cdot\cdot}, \bar{x}_{i\xi\cdot}, \bar{x}_{i\cdot\eta}$, $m_i, m_{i\xi\cdot}, m_{i\cdot\eta}$ соответственно и положим

$$S_{ij\cdot\cdot} = \sum_{\xi,\ \eta}(x_{i\xi\eta} - m_i - m_{i\xi\cdot} - m_{i\cdot\eta})(x_{j\xi\eta} - m_j - m_{j\xi\cdot} - m_{j\cdot\eta}),$$

$$S_{ij\cdot 0} = \sum_{\xi,\ \eta} m_{i\xi\cdot} m_{j\xi\cdot}.$$

Показать, что отношение правдоподобия λ для проверки гипотезы $\mathscr{H}$ равно

$$\lambda = L^{\frac{1}{2}rs},$$

где

$$L = \frac{|S_{ij\cdot\cdot}|}{|S_{ij\cdot\cdot} + S_{ij\cdot 0}|},$$

и при нулевой гипотезе (т. е. при $\mu_{i1}. = \ldots = \mu_{ir}. = 0$, $i = 1, \ldots, k$) распределение L совпадает с распределением R_s из **18.5.1** с m и s, замененными на $(r-1)(s-1)$ и $(r-1)$ соответственно.

18.15. *Распределение суммы квадратов разностей в методе наименьших квадратов.* Пусть $(x_{1\xi}, \ldots, x_{k\xi};\ \xi = 1, \ldots, n)$, $n > k$, выборка из k-мерной совокупности $N(\{\mu_i,\}, \|\sigma_{ij}\|)$, $\|u_{ij}\|$ — матрица внутреннего рассеивания выборки. Показать, что для фиксированной выборки минимум $\sum_{\xi=1}^{n} [x_{1\xi} - \beta_1 - \beta_2 x_{2\xi} - \ldots - \beta_k x_{k\xi}]^2$ по $\beta_1, \ldots, \beta_k$ равен $|u_{ij}|/|u_{vw}|$, $i, j = 1, \ldots, k$; $v, w = 2, \ldots, k$. С помощью методов, использованных в § 18.4 для вычисления моментов, доказать, что

$$\mathscr{E}\left(\frac{|u_{ij}|}{|u_{vw}|}\right)^r = (2/\sigma^{11})^r \frac{\Gamma\left(\frac{n-k}{2} + r\right)}{\Gamma\left(\frac{n-k}{2}\right)}, \quad r = 0, 1, 2 \ldots,$$

где σ^{11} — элемент первой строки и первого столбца матрицы $\|\sigma_{ij}\|^{-1}$. Из этой последовательности моментов получить, что величина

$$\sigma^{11}\, |u_{ij}|/|u_{vw}|$$

имеет χ^2-распределение $C(n-k)$.

18.16. (*Продолжение*). *Выборочное распределение множественного коэффициента корреляции.* Показать, что если минимизирующие значения $\beta_1, \ldots, \beta_k$ суть $\hat{\beta}_1, \ldots, \hat{\beta}_k$, то выборочный коэффициент корреляции R между $x_{1\xi}$ и $\hat{\beta}_1 + \hat{\beta}_2 x_{2\xi} + \ldots + \hat{\beta}_k x_{k\xi}$, $\xi = 1, \ldots, n$, удовлетворяет уравнению

$$1 - R^2 = \frac{|u_{ij}|}{u_{11} \cdot |u_{vw}|}.$$

Положим

$$\psi(r, \theta) = \mathscr{E}\left[\left(\frac{|u_{ij}|}{|u_{vw}|}\right)^r e^{-\theta u_{11}}\right].$$

Доказать, что

$$\psi(r, \theta) = (1 + 2\sigma_{11}\theta)^{-\frac{1}{2}(n-1)} \left(\frac{1}{2}\sigma^{11} + \theta\right)^{-r} \frac{\Gamma\left(\frac{n-k}{2} + r\right)}{\Gamma\left(\frac{n-k}{2}\right)} =$$

$$= \frac{\Gamma\left(\frac{n-k}{2} + r\right)}{(2\sigma_{11})^{\frac{1}{2}(n-1)}\, \Gamma\left(\frac{n-1}{2}\right) \Gamma\left(\frac{n-k}{2}\right)} \times$$

$$\times \sum_{i=0}^{\infty} \frac{\left[\frac{\sigma_{11}\sigma^{11} - 1}{2\sigma_{11}}\right]^i \Gamma\left(\frac{n-1}{2} + i\right) \left(\frac{1}{2}\sigma^{11} + \theta\right)^{-\frac{1}{2}(n-1) - r - i}}{i!}.$$

Замечая, что

$$\mathscr{E}(1 - R^2)^r = \int_0^{\infty} \ldots \int_0^{\infty} \psi(r, \xi_1 + \ldots + \xi_r)\, d\xi_1 \ldots d\xi_r, \quad r = 0, 1, 2, \ldots,$$

вывести формулу

$$\mathscr{E}(1-R^2)^r = \frac{(1-\rho^2)^{\frac{1}{2}(n-1)}}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-k}{2}\right)} \sum_{i=0}^{\infty} \frac{\rho^{2i}\,\Gamma\left(\frac{n-1}{2}+i\right)\Gamma\left(\frac{n-k}{2}+r\right)}{i!\,\Gamma\left(\frac{n-1}{2}+i+r\right)},$$

где ρ^2 — квадрат множественного коэффициента корреляции между x_1 и $x_2, \ldots, x_k$ в исходной совокупности,

$$1-\rho^2 = \frac{1}{\sigma_{11}\sigma^{11}}.$$

Пользуясь выражением для моментов $1-R^2$, показать, что п. в. R^2 равна

$$\frac{(1-\rho^2)^{\frac{1}{2}(n-1)}(1-R^2)^{\frac{1}{2}(n-k-2)}(R^2)^{\frac{1}{2}(k-3)}}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-k}{2}\right)} \sum_{i=0}^{\infty} \frac{(\rho^2R^2)^i\,\Gamma^2\left(\frac{n-1}{2}+i\right)}{i!\,\Gamma\left(\frac{k-1}{2}+i\right)}$$

для $0 \leqslant R^2 \leqslant 1$ и 0 в противном случае [Фишер (1928b)].

18.17. (*Продолжение*). *Распределение рассеивания разностей.* Рассмотрим разности

$$y_{p\xi} = x_{p\xi} - \beta_p - \sum_{v=s+1}^{k} \beta_{pv} x_{v\xi}, \qquad p=1, \ldots, s; \quad \xi=1, \ldots, n; \; s<k,$$

и их рассеивание

$$\left|\sum_{\xi=1}^{n} y_{p\xi}\, y_{q\xi}\right|.$$

Методами, аналогичными использованным в § 18.4, показать, что наименьшее значение рассеивания по $\beta_1, \ldots, \beta_{sk}$ равно

$$|u_{ij}|\,/\,|u_{vw}|, \qquad i, j=1, \ldots, k; \quad v, w=s+1, \ldots, k.$$

Доказать, что если выборка $(x_{1\xi}, \ldots, x_{k\xi};\ \xi=1, \ldots, n)$ извлечена из нормальной совокупности $N(\{\mu_i\}, \|\sigma_{ij}\|)$, то

$$\mathscr{E}\left(\frac{|u_{ij}|}{|u_{vw}|}\right)^r = \left(\frac{2^s\,|\sigma_{ij}|}{|\sigma_{vw}|}\right)^r \prod_{i=k-s+1}^{k} \left[\frac{\Gamma\left(\frac{n-i}{2}+r\right)}{\Gamma\left(\frac{n-i}{2}\right)}\right].$$

Вывести отсюда, что распределение $|u_{ij}|\,/\,|u_{vw}|$ совпадает с распределением

$$\frac{2^s\,|\sigma_{ij}|}{|\sigma_{vw}|} \prod_{i=1}^{s} z_i,$$

где $z_1, \ldots, z_s$ — независимые случайные величины, имеющие соответственно гамма-распределения $G\left(\frac{1}{2}(n-k)\right)$, $G\left(\frac{1}{2}(n-k+1)\right), \ldots, G\left(\frac{1}{2}(n-k+s-1)\right)$.

18.18. *Главные компоненты (собственные значения) k-мерного распределения вероятностей.* Главные компоненты k-мерного распределения вероятностей с матрицей ковариаций $\|\sigma_{ij}\|$ равны по определению собственным значениям $\lambda_1, \ldots, \lambda_k$ этой матрицы (т. е. корням уравнения $|\sigma_{ij} - \lambda\delta_{ij}| = 0$), причем $+\infty > \lambda_1 \geqslant \ldots \geqslant \lambda_k \geqslant 0$. Показать, что $\lambda_1 + \ldots + \lambda_k = \sigma_{11} + \ldots + \sigma_{kk}$, $\lambda_1\lambda_2 \ldots \lambda_k = |\sigma_{ij}|$, единичный собственный вектор $(c_{1p}, \ldots, c_{kp})$, отвечающий собственному значению λ_p, отличному от всех остальных собственных значений, удовлетворяет условию

$$\sum_{j=1}^{k} (\sigma_{ij} - \lambda_p \delta_{ij})\, c_{jp} = 0, \qquad i = 1, \ldots, k.$$

Доказать, что если все собственные значения $\lambda_1, \ldots, \lambda_k$ различны, то собственные векторы попарно ортогональны. Следовательно, при выборе в качестве новых осей направлений собственных векторов $(c_{11}, \ldots, c_{k1}), \ldots$ $\ldots, (c_{1k}, \ldots, c_{kk})$ распределение вероятностей в новой системе координат будет иметь матрицу ковариаций $\|\delta_{ij}\lambda_i\|$.

18.19. (*Продолжение*). Пусть матрица ковариаций $\|\sigma_{ij}\|$ k-мерного распределения вероятностей такова, что $\sigma_{ii} = \sigma^2$, $i = 1, \ldots, k$; $\sigma_{ij} = \rho\sigma^2$, $i \neq j = 1, \ldots, k$. Показать, что $\lambda_1 = \sigma^2 [1 + (k-1)\rho]$, $\lambda_2 = \ldots = \lambda_k = \sigma^2 (1-\rho)$, — единичный вектор $(c_{11}, \ldots, c_{k1})$, отвечающий λ_1, задается компонентами

$$c_{11} = \ldots = c_{k1} = \frac{1}{\sqrt{k}}.$$

18.20. *Критерий независимости двух множеств случайных величин.* Пусть $\|u_{ij}\|$ — матрица внутреннего рассеивания выборки объема $n > k$ из k-мерной нормальной совокупности $N(\{\mu_i\}, \|\sigma_{ij}\|)$; $\|u_{pq}\|$, $p, q = 1, \ldots, s$, — матрица внутреннего рассеивания выборки, образованной первыми s координатами, $\|u_{vw}\|$, $v, w = s+1, \ldots, k$, — матрица внутреннего рассеивания выборки, образованной последними t $(t = k - s,\ s \leqslant t)$ координатами. Обозначим $\mathscr{H}(\omega; \Omega)$ гипотезу, в которой Ω есть $\frac{1}{2}k(k+3)$-мерное пространство параметров ($\mu_1, \ldots, \mu_k$ — вещественные, $\|\sigma_{ij}\|$ — симметрическая матрица), ω — $\left(\frac{1}{2}k(k+3) - st\right)$-мерное подпространство Ω, на котором

$$\sigma_{pw} = 0, \quad p = 1, \ldots, s; \quad w = s+1, \ldots, k.$$

Показать, что отношение правдоподобия для гипотезы $\mathscr{H}$ равно

$$\lambda = L^{\frac{1}{2}n},$$

где

$$L = \frac{|u_{ij}|}{|u_{pq}| \cdot |u_{vw}|}.$$

Методами, аналогичными применявшимся в § 18.4 (а), доказать, что

$$\mathscr{E}(L^r) = \prod_{p=1}^{s} \left[\frac{\Gamma\left(\frac{n-p}{2}\right) \Gamma\left(\frac{n-t-p}{2} + r\right)}{\Gamma\left(\frac{n-p}{2} + r\right) \Gamma\left(\frac{n-t-p}{2}\right)} \right],$$

$r = 0, 1, 2, \ldots$, когда $\mathscr{H}$ верна, Вывести отсюда, что при $s=1$, $t=k-1$ L имеет бета-распределение $\mathrm{Be}\left(\frac{1}{2}(n-k),\ \frac{1}{2}(k-1)\right)$, а при $s=2$, $t=k-2$ L имеет распределение $\mathrm{Be}(n-k,\ k-2)$ [Уилкс (1935)].

18.21. *Критерий сферичности нормального распределения.*

Пусть $\|u_{ij}\|$ — матрица внутреннего рассеивания выборки объема n из k-мерной нормальной совокупности $N(\{\mu_i\},\ \|\sigma_{ij}\|)$. Рассмотрим гипотезу $\mathscr{H}(\omega;\ \Omega)$, в которой Ω — $\frac{1}{2}k(k+3)$-мерное пространство параметров (μ_i вещественны, $\|\sigma_{ij}\|$ — симметрическая), а ω — $(k+1)$-мерное подпространство Ω, на котором $\|\sigma_{ij}\|=\|\delta_{ij}\sigma^2\|$, где δ_{ij} — символ Кронекера, $\sigma^2>0$. Показать, что отношение правдоподобия λ для проверки гипотезы $\mathscr{H}$ равно

$$\lambda=L^{\frac{1}{2}n},$$

где

$$L=\frac{|u_{ij}|}{\bar{u}^k},\qquad \bar{u}=\frac{1}{k}(u_{11}+\ldots+u_{kk}).$$

Показать, далее, что при нулевой гипотезе

$$\mathscr{E}\left(\frac{|u_{ij}|}{\bar{u}^k}\right)^r=\int_0^\infty\ldots\int_0^\infty\psi(r,\ \xi_1+\ldots+\xi_{rk})\,d\xi_1\ldots d\xi_{rk}=$$

$$=\frac{k^{kr}\,\Gamma\left(\frac{(n-1)k}{2}\right)}{\Gamma\left(\frac{(n-1)k}{2}+rk\right)}\prod_{i=1}^{k}\left[\frac{\Gamma\left(\frac{n-i}{2}+r\right)}{\Gamma\left(\frac{n-i}{2}\right)}\right],\qquad r=0,\ 1,\ 2,\ \ldots,$$

где

$$\psi(r,\ \theta)=\mathscr{E}\left(|u_{ij}|^r e^{-\theta\bar{u}}\right)=$$

$$=\left(\frac{2}{\sigma^2}\right)^{kr}\left(1+\frac{2\theta}{k\sigma^2}\right)^{-\frac{1}{2}[(n-1)+2rk]}\prod_{i=1}^{k}\left[\frac{\Gamma\left(\frac{n-i}{2}+r\right)}{\Gamma\left(\frac{n-i}{2}\right)}\right].$$

Проверить, что при $k=2$ $\sqrt{L}$ имеет бета-распределение $\mathrm{Be}(n-2,1)$ [Мочли (1940)].

18.22. (*Продолжение*). *Критерий гипотезы о симметричности k-мерного нормального распределения по k переменным.* Пусть $\mathscr{H}^*(\omega,\ \Omega)$ — гипотеза, в которой Ω то же, что и в предыдущей задаче, а ω — $(k+3)$-мерное подпространство Ω, на котором $\sigma_{ii}=\sigma^2$, $i=1,\ \ldots,\ k$; $\sigma_{ij}=\sigma_{ji}=\rho\sigma^2$, $i\neq j$; $-\frac{1}{k-1}\leqslant\rho<1$.

Показать, что отношение правдоподобия λ для гипотезы $\mathscr{H}^*$ равно

$$\lambda=(L^*)^{\frac{1}{2}n},$$

где

$$L^*=\frac{|u_{ij}|}{(\bar{u}-\bar{u}^*)^{k-1}(\bar{u}+(k-1)\bar{u}^*)}$$

и $\bar{u}=\frac{1}{k}(u_{11}+\ldots+u_{kk})$, $\bar{u}^*=\frac{1}{k(k-1)}\sum_{i\neq j} u_{ij}$, так что статистика L^* эквивалентна λ при проверке $\mathcal{H}^*$.

Доказать, что, когда $\mathcal{H}^*$ верна,

$$\mathscr{E}(L^*)^r=\int_0^\infty\ldots\int_0^\infty \psi(r,\ \xi_1+\ldots+\xi_{r(k-1)},\ \eta_1+\ldots+\eta_k)\,d\xi_1\ldots d\xi_{r(k-1)}\,d\eta_1\ldots d\eta_k=$$

$$=(k-1)^{r(k-1)}\frac{\Gamma\left(\frac{(n-1)(k-1)}{2}\right)}{\Gamma\left(\frac{(n-1)(k-1)}{2}+r(k-1)\right)}\prod_{i=1}^{k}\left[\frac{\Gamma\left(\frac{n-i}{2}+r\right)}{\Gamma\left(\frac{n-i}{2}\right)}\right],$$

$$r=0,\ 1,\ 2,\ \ldots,$$

где

$$\psi(r,\ p,\ q)=\mathscr{E}\left[|u_{ij}|^r e^{-p(\bar{u}-\bar{u}^*)-q(\bar{u}+(k-1)\bar{u}^*)}\right]=$$

$$=2^{kr}\prod_{i=1}^{k}\left[\frac{\Gamma\left(\frac{n-i}{2}+r\right)}{\Gamma\left(\frac{n-i}{2}\right)}\right]\times$$

$$\times(A^{k-1}-B)^{\frac{1}{2}(n-1)}\left[A+\frac{2p}{k-1}\right]^{-\frac{1}{2}(k-1)[(n-1)+2r]}[B+2q]^{-\frac{1}{2}(n-1+2r)}$$

и

$$A=\frac{1}{\sigma^2(1-\rho)},\qquad B=\frac{1}{\sigma^2[1+(k-1)\rho]}.$$

Проверить, что при $k=2$ L^* имеет бета-распределение $\mathrm{Be}\left(\frac{1}{2}(n-2),\ \frac{1}{2}\right)$, а при $k=3$, $\sqrt{L^*}$ имеет распределение $\mathrm{Be}(n-2,1)$ [Уилкс (1946)]. [Относительно обобщения критерия L^* для проверки гипотезы о симметричности распределения $N(\{\mu_i\},\ \|\sigma_{ij}\|)$ по переменным внутри блоков см. Вотау (1948).]

18.23. *Нецентральное распределение* T^2. Показать, что если в обстановке § 18.4 (а) выборка извлечена из совокупности $N(\{\mu_i\},\ \|\sigma_{ij}\|)$, где $(\mu_1,\ \ldots,\ \mu_k)\neq(\mu_1^\circ,\ \ldots,\ \mu_k^\circ)$, то определенная в (18.4.5) величина T^2 имеет п. в.

$$\frac{e^{-\frac{1}{2}\delta^2}(T^2)^{\frac{1}{2}k-1}[1+T^2/(n-1)]^{-\frac{1}{2}n}}{(n-1)^{\frac{1}{2}k}\,\Gamma\left(\frac{n-k}{2}\right)}\sum_{i=0}^{\infty}\frac{\left[\frac{\delta^2T^2}{2(n-1)}\right]^i\Gamma\left(\frac{n}{2}+i\right)}{i!\,\Gamma\left(\frac{k}{2}+i\right)[1+T^2/(n-1)]^i},$$

где

$$\delta^2=n\sum_{i,\ j=1}^{k}\sigma_{ij}(\mu_i-\mu_i^0)(\mu_j-\mu_j^0)$$

[Сюй (1938)].

18.24. *Вычисление моментов рассеивания $|v_{ij}|$ выборки из совокупности $N(\{\mu_i\}, \|\sigma_{ij}\|)$ без использования распределения Уишарта* [Уилкс (1934)]. Пусть выборка $(x_{1\xi}, \ldots, x_{k\xi};\ \xi = 1, \ldots, n)$, $n \geqslant k$, извлечена из k-мерной нормальной совокупности $N(\{\mu_i\}, \|\sigma_{ij}\|)$; $\|v_{ij}\|$ — матрица рассеивания относительно среднего $(\mu_1, \ldots, \mu_k)$.

Показать (без использования распределения Уишарта), что при $2r < n + 1 - k$

$$\mathscr{E}(|v_{ij}|^{-r}) = \pi^{kr} \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \mathscr{E}\left[\exp\left(-\sum_{p=1}^{2r} \sum_{i, j=1}^{k} v_{ij} z_{ip} z_{jq}\right)\right] \prod_{i, p} dz_{ip} =$$

$$= |2\sigma_{ij}|^{-r} \prod_{i=1}^{k} \left[\frac{\Gamma\left(\frac{n+1-i}{2} - r\right)}{\Gamma\left(\frac{n+1-i}{2}\right)}\right].$$

ЛИТЕРАТУРА

[В квадратных скобках указаны страницы, на которых встречаются ссылки на соответствующую работу]

Андерсон (R. L. Anderson, 1942), Distribution of the serial correlation coefficient, Ann. Math. Stat., Vol. 13, pp. 1—13. [533]

Андерсон (T. W. Anderson, 1948), On the theory of testing serial correlation, Skand. Aktuar., Vol. 31, pp. 88—116. [534]

Андерсон и Дарлинг (T. W. Anderson and D. A. Darling, 1952), Asymptotic theory of certain «goodness — of — fit» criteria based on stochastic processes, Ann. Math. Stat., Vol. 23, pp. 193—212. [445]

Андерсон (T. W. Anderson, 1958), An introduction to multivariate statistical analysis, John Wiley, New York (русский перевод: Введение в многомерный статистический анализ, М., Физматгиз, 1963). [539]

Андерсон (T. W. Anderson, 1960), A modification of the sequential probability ratio test to reduce the sample size, Ann. Math. Stat., Vol. 31, pp. 165—197. [495]

Арнольд (H. E. Arnold, 1958), Permutation support for multivariate techniques, Ph. D. Dissertation, Princeton University. [471]

Балмер (M. G. Bulmer, 1957), Approximate confidence limits for components of variance, Biometrika, pp. 159—167. [318]

Баранкин и Гёрланд (E. W. Barankin and J. Gurland, 1951), On asymptotically normal efficient estimators: I, Univ. California Publications in Statistics, Vol. I, № 6, pp. 89—129. [369, 372, 389]

Барнард (G. A. Barnard, 1952), The frequency justification of certain sequential tests, Biometrika, Vol. 39, pp. 144—150. [477]

Бартки (W. Bartky, 1943), Multiple sampling with constant probability, Ann. Math. Stat., Vol. 14, pp. 363—377. [478]

Бартлетт (M. S. Bartlett, 1950), Periodogram analysis and continuous spectra, Biometrika, Vol. 37, pp. 1—16. [524]

Бартлетт (M. S. Bartlett, 1953), Approximate confidence intervals, II. More than one unknown parameter, Biometrika, Vol. 40, pp. 306—317. [396]

Бартлетт (M. S. Bartlett, 1955), An introduction to stochastic processes, Cambridge University Press (русский перевод: Введение в теорию случайных процессов, ИЛ, М., 1958). [515, 524, 526]

Бахадур (R. R. Bahadur, 1954), Sufficiency and statistical decision functions, Ann. Math. Stat., Vol. 25, pp. 423—462. [365]

Бахадур (R. R. Bahadur, 1958), Examples of inconsistency of maximum likelihood extimates, Sankhyā, Vol. 20, pp. 207—210. [372]

Бернулли (J. Bernoulli, 1713), Arc Conjectandé, Paris. [150]

Бертран (J. Bertrand, 1889), Calcul des probabilités, Gauthier—Villars, Paris. [475]

Бил (E. M. L. Beale, 1960), Confidence regions in non-linear estimation, Journ. Roy. Stat. Soc., Ser. B., Vol. 22, pp. 41—88. [396]

Биркгоф и Маклэйн (G. Birkhoff and S. MacLane, 1953), A survey of modern algebra (переработанное издание), Macmillan, New York. [226, 567, 576]

Бирнбаум и Тинджи (Z. W. Birnbaum and F. H. Tingey, 1951), One-sided confidence contours for probability distribution functions, Ann. Math. Stat., Vol. 22, pp. 592—596. [345, 347]

Бирнбаум (Z. W. Birnbaum, 1953a), On the power of a one-sided test of fit for continuous probability functions, Ann. Math. Stat., Vol. 24, pp. 484—489. [445]

Бирнбаум (Z. W. Birnbaum, 1953b), Distribution—free tests of fit for continuous probability functions, Ann. Math. Stat., Vol. 24, pp. 1—8. [445]

Блекуэл (D. Blackwell, 1946), On an equation of Wald, Ann. Math. Stat., Vol. 17, pp. 84—87. [481]

Блекуэл (D. Blackwell, 1947), Conditional expectation and unbiased sequential estimation, Ann. Math. Stat., Vol. 18, pp. 105—110. [366]

Блекуэл и Гиршик (D. Blackwell and M. A. Girshik, 1954), Theory of games and statistical decisions, John Wiley, New York (русский перевод: Теория игр и статистических решений, ИЛ, М., 1958). [505, 513]

Блюм и Уэйсс (J. R. Blum and L. Weiss, 1957), Consistency of certain two-sample tests, Ann. Math. Stat., Vol. 28, pp. 242—246. [457]

Бокс и Уилсон (G. E. P. Box and K. B. Wilson, 1951), On the experimental attainment of optimum conditions, Journ. Roy. Stat. Soc., Ser. B, Vol. 13, pp. 1—45. [308]

Бокс (G. E. P. Box, 1952), Multifactor designs of first order, Biometrika, Vol. 39, pp. 49—57. [308]

Бокс (G. E. P. Box, 1954), The exploration and exploitation of response surfaces: Some general considerations and examples, Biometrics, Vol. 10, pp. 16—60. [308]

Бокс и Хантер (G. E. P. Box and J. S. Hunter, 1957), Multifactor experimental designs for exploring response surfaces, Ann. Math. Stat., Vol. 28, pp. 195—241. [308]

Бокс и Маллер (G. E. P. Box and M. E. Muller, 1958), A note on the generation of random normal deviates, Ann. Math. Stat., Vol. 29, pp. 610—611. [202]

Больцман (L. Boltzmann, 1898), Vorlesungen über Gastheorie, Part I, J. A. Barth, Leipzig. [417]

Боуз (R. C. Bose, 1938), On the application of the properties of Galois fields to the problem of construction of hyper-Graeco-Latin squares, Sankhyā, Vol. 3, pp. 323—338. [246]

Боуз (R. C. Bose, 1939), On the construction of balanced incomplete block designs, Ann. Eugen., Vol. 9, pp. 353—399. [244]

Бохер (M. Bochér, 1929), Introduction to higher algebra, Macmillan, New York (русский перевод: Введение в высшую алгебру, Гостехтеоретиздат, М.—Л., 1933). [102, 226, 299, 581]

Бохнер (S. Bochner, 1932), Vorlesungen über Fouriersche Integrale, Leipzig (русский перевод: Лекции об интеграле Фурье, Физматгиз, М., 1962). [521]

Бэттин (I. L. Battin, 1942), On the problem of multiple matching, Ann. Math. Stat., Vol. 13, pp. 294—305. [168]

Вальд и Вольфовитц (A. Wald and J. Wolfowitz, 1939), Confidence limits for continuous distribution functions, Ann. Math. Stat., Vol. 10, pp. 105—118. [348]

Вальд (A. Wald, 1939), Contributions to the theory of statistical estimation and testing hypotheses, Ann. Math. Stat., Vol. 10, pp. 299—326. [409]

Вальд и Вольфовитц (A. Wald and J. Wolfowitz, 1940), On a test whether two samples are from the same population, Ann. Math. Stat., Vol. 11, pp. 147—162. [404, 448 ,458]

Вальд (A. Wald, 1941a), Asymptotically most powerful tests of statistical hypotheses, Ann. Math. Stat., Vol. 12, pp. 1—19. [411, 422]

Вальд (A. Wald, 1941b), Some examples of asymptotically most powerful tests, Ann. Math. Stat., Vol. 12, pp. 396—408. [411]

Вальд (A. Wald, 1942), Asymptotically shortest confidence intervals, Ann. Math. Stat., Vol. 13, pp. 127—137. [384]

Вальд (A. Wald, 1943), An extension of Wilks' method for setting tolerance limits, Ann. Math. Stat., Vol. 14, pp. 45—55. [251]

Вальд и Вольфовитц (A. Wald and J. Wolfowitz, 1944), Statistical tests based on permutations of the observations, Ann. Math. Stat., Vol. 15, pp. 358—372. [277, 278, 470]

Вальд (A. Wald, 1945), Sequential tests of statistical hypotheses, Ann. Math. Stat., Vol. 16, pp. 117—186. [478, 481, 487, 490, 493, 495, 497]

Вальд и Вольфовитц (A. Wald and J. Wolfowitz, 1946), Tolerance limits for a normal distribution, Ann. Math. Stat. Vol. 17, pp. 208—215. [352]

Вальд (A. Wald, 1947a), Sequential analysis, John Wiley, New York (русский перевод: Последовательный анализ, Физматгиз, М., 1960). [477, 478, 492, 493, 498, 499]

Вальд (A. Wald, 1947b), Foundations of a general theory of statistical decision functions, Econometrica, Vol. 15, pp. 279—313. [505]

Вальд (A. Wald, 1948), Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process, Ann. Math. Stat., Vol. 19, pp. 40—46. [369]

Вальд и Вольфовитц (A. Wald and J. Wolfowitz, 1948), Optimum character of the sequential probability ratio test, Ann. Math. Stat., Vol. 19, pp. 326—339. [495]

Вальд (A. Wald, 1949a), Statistical decision functions, Ann. Math. Stat., Vol. 20, pp. 165—205. [505]

Вальд (A. Wald, 1949b), Note on the consistency of the maximum likelihood estimate, Ann. Math. Stat., Vol. 20, pp. 595—601. [369]

Вальд (A. Wald, 1950), Statistical decision functions, John Wiley, New York. [505, 507, 513]

Винер (N. Wiener, 1949), Extrapolation, interpolation and smoothing of stationary time series, John Wiley, New York. [535, 536]

Волд (H. Wold, 1938), A study in the analysis of stationary time series, University of Uppsala. [515, 537]

Вольфовитц (J. Wolfowitz, 1949), Non-parametric statistical inference, Proc. (First) Berkeley Symposium on Math. Statistics and Probability, pp. 93—113, University of California Press. [435]

Вотау (D. F. Votaw, Jr., 1948), Testing compound symmetry in a normal multivariate distribution, Ann. Math. Stat., Vol. 19, pp. 447—473. [599]

Гарвуд (F. Garwood, 1936), Fiducial limits for the Poisson distribution, Biometrika, Vol. 28, pp. 437—442. [378]

Гаттмен (I. Guttman, 1960), Personal communication. [260]

Гаусс (K. F. Gauss, 1809a), Theoria motus corporum coelestium in sectimibus conicis solem aurbientium, Perthes and Besser, Hamburg. [296]

Гаусс (K. F. Gauss, 1809b), Werke, Vol. 4, Göttingen, pp. 1—93. [268]

Герглотц (G. Herglotz, 1911), Über Potenzreihen mit positivem reelen Teil im Einheitskreis, Berichte Verh. König. Sächs. Ges. Wiss., Leipzig, Math.— Phys. Klasse, Vol. 63, pp. 501—511. [520]

Гиршик (M. A. Girshik, 1939), On the sampling theory of roots of determinantal equations, Ann. Math. Stat., Vol. 10, pp. 203—224. [566]

Гиршик, Мостеллер и Сэвидж (M. A. Girshik, F. Mosteller and L. J. Savage, 1946), Unbiased estimates for certain binomial sampling problems with applications, Ann. Math. Stat., Vol. 17, pp. 13—23. [158]

Б. В. Гнеденко и В. С. Королюк (1951), О максимальном расхождении между двумя эмпирическими функциями распределения, Доклады АН СССР, том 80, № 4, стр. 525—528. [461]

Б. В. Гнеденко и А. Н. Колмогоров (1949), Предельные распределения для сумм независимых случайных величин, Гостехиздат, М., [11, 203, 268]

Госсет (W. S. Gosset, «Student», 1908), The probable error of a mean, Biometrika, Vol. 6, pp. 1—25. [224]

Гофдинг (W. Hoeffding, 1948a), A class of statistics with asymptotically normal distribution, Ann. Math. Stat., Vol. 19, pp. 293—325. [215, 468]

Гофдинг (W. Hoeffding, 1948b), A non-parametric test of independence, Ann. Math. Stat., Vol. 19, pp. 546—557. [473]

Гофдинг (W. Hoeffding, 1952), The large-sample power of tests based on permutations of observations, Ann. Math. Stat., Vol. 23, pp. 169—192. [471]

Гофдинг (W. Hoeffding, 1960), Lower bounds for the expected sample size and the average risk of a sequential procedure, Ann. Math. Stat., Vol. 31, pp. 352—368. [495]

Гренандер и Розенблатт (U. Grenander and M. Rosenblatt, 1953), Statistical spectral analysis of time series arising from stationary stochastic processes, Ann. Math. Stat., Vol. 24, pp. 537—558. [524]

Гренандер и Розенблатт (U. Grenander and M. Rosenblatt, 1957), Statistical analysis of stationary time series, John Wiley, New York; Almqvist and Wiksell, Stockholm. [515, 526, 536]

Гринвуд и Хартли (J. A. Greenwood and H. O. Hartley, 1961), Guide to tables in mathematical statistics, Princeton University press. [171, 201, 308]

Грэйбилл (F. A. Graybill, 1961), An introduction to linear statistical models, Mc Graw-Hill, New York. [308, 318]

Гудмен (L. A. Goodman, 1953), Sequential sampling tagging for population size problems, Ann. Math. Stat., Vol. 24, pp. 56—69. [503]

Гумбель (E. J. Gumbel, 1935), Les valeurs extrêmes des distributions statistiques, Ann. de l'Institut Henri Poincaré, Vol. 4, pp. 115—158. [283]

Гумбель (E. J. Gumbell, 1958), Statistics of extremes, Columbia University press (русский перевод: Статистика экстремальных значений, «Мир», М., 1965). [247, 283]

Данкен (D. B. Duncan, 1952), On the properties of the multiple comparison test, Virginia Journ. of Sci., Vol. 3, pp. 49—67. [301]

Дарлинг (D. A. Darling, 1957), The Kolmogorov — Smirnov, Cramér — von Mises tests, Ann. Math. Stat., Vol. 28, pp. 823—838. [350, 445]

Двайер (P. S. Dwyer, 1938), Combined expansions of products of symmetric power sums and of sums of symmetric power products with applications to sampling, Ann. Math. Stat., Vol. 9, pp. 1—47, 97—132. [213]

Двасс (M. Dwass, 1959), Multiple confidence procedures, Ann. Inst. Stat. Math., Vol. 10, pp. 277—282. [301, 307]

Дворецкий, Кифер и Вольфовитц (A. Dvoretzky, J. Kiefer and J. Wolfowitz, 1953a), Sequential decision problems for processes with continuous time parameter. Testing hypotheses, Ann. Math. Stat., Vol. 24, pp. 254—264. [479, 513]

Дворецкий, Кифер и Вольфовитц (A. Dvoretzky, J. Kiefer and J. Wolfowitz, 1953b), Sequential decision problems for processes with continuous time parameter. Problems of estimation, Ann. Math. Stat., Vol. 24, pp. 403—415. [479, 513]

Демпстер (A. P. Dempster, 1955), Personal communication. [348]

Демпстер (A. P. Dempster, 1958), A high dimensional two-sample significance test, Ann. Math. Stat., Vol. 29, pp. 995—1010. [575]

Дёрбин и Ватсон (J. Durbin and G. S. Watson, 1951), Exact tests of serial correlation using non-circular statistics, Ann. Math. Stat., Vol. 22, pp. 446—451. [534]

Джеймс (A. T. James, 1954), Normal multivariate analysis and the orthogonal group, Ann. Math. Stat., Vol. 25, pp. 40—75. [551, 566]

Джеймс (A. T. James, 1960), The distribution of the latent roots of the covariance matrix, Ann. Math. Stat., Vol. 31, pp. 151—158. [571]

Дженкинс и Пристли (G. M. Jenkins and M. B. Priestly, 1957), The spectral analysis of time series, Journ. Roy. Stat. Soc., Ser. B, Vol. 19, pp. 1—12. [524]

Джири (R. C. Geary, 1936), Distribution of Student's distribution for non-normal samples, Journ. Roy. Stat. Soc., Ser. B, Vol. 3, pp. 178—184. [224]

Диксон (W. J. Dixon, 1940), A criterion for testing the hypothesis that two samples are from the same population, Ann. Math. Stat., Vol. 11, pp. 199—204. [448]

Диксон (W. J. Dixon, 1944), Further contributions to the problem of serial correlation, Ann. Math. Stat., Vol. 15, pp. 119—144. [537, 536]

Диксон и Муд (W. J. Dixon and A. M. Mood, 1946), The statistical sign test, Journ. Amer. Stat. Ass., Vol. 41, pp. 557—566. [437]

Диксон и Муд (W. J. Dixon and A. M. Mood, 1948), A method of obtaining and analyzing sensitivity data, Journ. Amer. Stat. Ass., Vol. 43, pp. 109—126. [478]

Дирихле (G. L. Dirichlet, 1839), Sur un nouvelle methode pour la determination des integrales multiples, Comp. Rend. Acad. Sci., Vol. 8, pp. 156—160. [193]

Додд (E. L. Dodd, 1923), The greatest and least variate under general laws of error, Trans. Amer. Math. Soc., Vol. 25, pp. 525—539. [283]

Додж и Ромиг (H. F. Dodge and H. G. Romig, 1929), A method of sampling inspection, Bell System Techn. Journ., Vol. 8, pp. 613—631. [405, 406, 477]

Додж и Ромиг (H. F. Dodge and H. G. Romig, 1959), Sampling inspection tables, изд. 2, John Wiley, New York. [405, 478]

Донелли (T. G. Donelly, 1957), A family of sequential tests, Ph. D. thesis, University of North Carolina. [495]

Донскер (M. D. Donsker, 1952), Justification and extension of Doob's heuristic approach to the Kolmogorov — Smirnov theorems, Ann. Math. Stat., Vol. 23, pp. 277—281. [350]

Дорфман (R. Dorfman, 1943), The detection of defective members of large populations, Ann. Math. Stat., Vol. 14, pp. 436—440. [166]

Дуб (J. L. Doob, 1949), Heuristic approach to the Kolmogorov — Smirnov theorems, Ann. Math. Stat., Vol. 20, pp. 393—403. [350]

Дуб (J. L. Doob, 1953), Stochastic processes, John Wiley, New York (русский перевод: Вероятностные процессы, ИЛ, М., 1956). [11, 109, 110, 515, 536]

Дэвид (F. N. David, 1950), Two combinatorial tests of whether the sample has come from a given population, Biometrika, Vol. 37, pp. 97—110. [441]

Дэвид и Краскел (H. A. David and W. H. Kruskal, 1956), The WAGR sequential t-test reaches a decision with probability 1, Ann. Math. Stat., Vol. 27, pp. 797—805. [477]

Дэвис (H. T. Davis, 1941), The analysis of economic time series, Principia Press, Bloomington, Indiana. [532]

Дэйли (J. F. Daly, 1940), On the unbiased character of likelihood-ratio tests for independence in normal systems, Ann. Math. Stat., Vol. 11, pp. 1—32. [415]

Дюге (D. Dugué, 1937), Application des proprietés de la limité au sens du calcul des probabilités a l'étude des diverses questions d'estimation, Écol. Poly., Vol. 3, № 4, pp. 305—372. [361, 363]

Ингам (A. E. Ingham, 1933), An integral that occurs in statistics, Proc. Cambr. Phil. Soc., Vol. 29, pp. 271—276. [550]

Ирвин (J. O. Irwin, 1930), On the frequency distribution of the means of samples from populations of certain of Pearson's types, Metron, Vol. 8, № 4, pp. 51—105. [218]

Иэйтс (F. Yates, 1949), Sampling methods for censuses and surv.eys, Charles Griffin, London. [323]

Кавата и Сакамото (T. Kawata and H. Sakamoto, 1949), On the characterization of the normal population by the independence of the sample mean and sample variance, Journ. Math. Soc. Japan, Vol. 1, pp. 111—115. [224]

Калбэк и Лейблер (S. Kullback and R. A. Leibler, 1951), On information and sufficiency, Ann. Math. Stat., Vol. 22, pp. 79—86. [425]

Калбэк (S. Kullback, 1959), Information theory and statistics, John Wiley New York. [417]

Капланский и Риордан (I. Kaplansky and J. Riordan, 1945), Multiple matching and runs by the symbolic method, Ann. Math. Stat., Vol. 16, pp. 272—277. [168]

Каратеодори (C. Carathéodory, 1927), Vorlesungen über reelle Funktionen, Leipzig — Berlin. [26]

Карлин и Рубин (S. Karlin and H. Rubin, 1956), The theory of decision procedures for distributions with monotone likelihood ratio, Ann. Math. Stat., Vol. 27, pp. 272—299. [513]

Карлтон (G. A. Carlton, 1946), Estimating the parameters of a rectangular distribution, Ann. Math. Stat., Vol. 17, pp. 355—358. [260]

Кемперман (J. H. B. Kemperman, 1956), Generalized tolerance limits, Ann. Math. Stat., Vol. 27, pp. 180—186. [254]

Кемпсорн (O. Kempthorne, 1952), The design and analysis of experiments, John Wiley, New York. [244, 308, 318]

Кендалл (D. G. Kendall, 1951), Some problems in the theory of queues, Journ. Roy. Stat. Soc., Ser. B, Vol. 13, pp. 151—185. [201]

Кендалл (D. G. Kendall, 1953), Stochastic processes occurring in the theory of queues and their analysis by the method of the Markov chain, Ann. Math. Stat., Vol. 24, pp. 338—354. [201]

Кендалл и Рао (D. G. Kendall and K. S. Rao, 1950), On the generalized second limit-theorem in the calculus of probabilities, Biometrika, Vol. 37, pp. 224—230. [140]

Кендалл и Смит (M. G. Kendall and B. B. Smith, 1939), The problem of *m* rankings, Ann. Math. Stat., Vol. 10, pp. 275—287. [474]

Кендалл (M. G. Kendall, 1943 and 1946), The advanced theory of statistics, Vol. I (1943), Vol. II (1946) (том I нового трехтомного издания вышел в 1958 г.), Charles Griffin, London. [213]

Кендалл и Сандрум (M. G. Kendall and R. M. Sundrum, 1953), Distribution-free methods and order properties, Rev. Inter. Stat. Inst., Vol. 23, pp. 124—134. [435]

Кендалл (M. G. Kendall, 1953), Rank correlation methods, Charles Griffin, London. [247, 435]

Кендалл (M. G. Kendall, 1957), A course in multivatiate analysis, Charles Griffin, London. [539]

Кимбел (B. F. Kimball, 1947), Some basic theorems for developing tests of fit for the case of non-parametric probability distribution function, I, Ann. Math. Stat., Vol. 18, pp. 540—548. [445]

Китабатаке (S. Kitabatake, 1959), A remark on a non-parametric test, Math. Japan, Vol. 4, pp. 45—49. [444]

Китагава и Митоме (T. Kitagawa and M. Mitome, 1953), Tables for the design of factorial experiments, Baifukan, Tokyo. [308]

Кифер (J. Kiefer, 1948), Sequential determination of the maximum of a function, Ph. D. thesis, Massachusets Institute of Technology. [479]

Кифер (J. Kiefer, 1952), On minimum variance estimators, Ann. Math. Stat., Vol. 23, pp. 627—629. [361]

Кифер (J. Kiefer, 1953), Sequential minimax search for a maximum, Proc. Amer. Math. Soc., Vol. 4, pp. 502—506. [479]

Кифер (J. Kiefer, 1957), Optimum sequential search and approximation methods under minimum regularity conditions, Journ. Industr. Appl. Math., Vol. 5, pp. 105—136. [479]

Кифер и Уэйсс (J. Kiefer and L. Weiss, 1957), Some properties of generalized sequential probability ratio test, Ann. Math. Stat., Vol. 28, pp. 57—74. [495]

Кифер и Вольфовитц (J. Kiefer and J. Wolfowitz, 1952), Stochastic estimation of the maximum of a regression function, Ann. Math. Stat., Vol. 23, pp. 462—466. [479]

Кишен (K. Kishen, 1945), On the design of experiments for weighing and making other types of measurements, Ann. Math. Stat., Vol. 16, pp. 294—300. [297]

Клоппер и Пирсон (C. J. Clopper and E. S. Pearson, 1934), The use of confidence or fiducial limits illustrated in the case of binomial, Biometrika, Vol. 26, pp. 404—413. [378]

Кокс (D. R. Cox, 1952), Sequential tests for composite hypotheses, Proc. Cambr. Phil. Soc., Vol. 48, pp. 290—299. [477]

Колмогоров (A. Kolmogorov, 1928), Über die Summen durch den Zufall bestimmer unabhängiger Grössen, Math. Ann., Vol. 99, pp. 309—319. [119]

Колмогоров (A. Kolmogorov, 1933a), Grundbegriffe der Wahrscheinlichkeitsrechnung, Ergeb. Math., № 3, Berlin (русский перевод: Основные понятия теории вероятностей, ОНТИ, 1936). [11, 21, 37, 109]

Колмогоров (A. Kolmogorov, 1933b), Sulla determinazióne empirica di una legge di distribuzione, Giorn. dell 'Inst. Ital. Attuari, Vol. 4, pp. 83—91. [348, 350, 445]

А. Н. Колмогоров (1941), Стационарные последовательности в гильбертовом пространстве, Бюллетень Московского унив., том 2, № 6, стр. 1—40. [535, 536]

Коннор (W. S. Connor, Jr., 1952), On the structure of balanced incomplete block designs, Ann. Math. Stat., Vol. 23, pp. 57—71. [244]

Корниш и Фишер (E. A. Cornish and R. A. Fisher, 1937), Moments and cumulants in the specification of distributions, Rev. Inter. Stat. Inst., Vol. 4, pp. 1—14. [213]

Кочрен (W. G. Cochran, 1934), The distribution of quadratic forms in a normal system, with applications to the analysis of covariance, Proc. Cambr. Phil. Soc., Vol. 30, 178—191. [225]

Кочрен (W. G. Cochran, 1953), Sampling techniques, John Wiley, New York. [323]

Кочрен и Кокс (W. G. Cochran and G. M. Cox, 1957), Experimental designs, изд. 2, John Wiley, New York. [244, 308, 318]

Коши (A. L. Cauchy, 1853), Sur les résultats moyens d'observations de mêne nature et sur les résultats les plus probables, Comp. Rend. Acad. Sci., Paris, Vol. 37, pp. 198—206. [142, 266]

Крамер (H. Cramér, 1928), On the composition of elementary errors, Skand. Aktuar., Vol. 11, pp. 13—74, 141—180. [445]

Крамер (H. Cramér, 1936), Über eine Eigenschaft der normalen Verteilungsfunktion, Math. Zeits., Vol. 41, pp. 405—414. [224]

Крамер (H. Cramér, 1937), Random variables and probability distributions, Cambridge tracts in mathematics, №. 36, Cambridge University Press (русский перевод: Случайные величины и распределения вероятностей, ИЛ, М., 1947), [134, 277]

Крамер (H. Cramér, 1946), Mathematical methods of statistics, Princeton University Press (русский перевод: Математические методы статистики, ИЛ, М., 1948). [104, 135, 137, 140, 144, 192, 361, 363]

Краскел (W. H. Kruskal, 1952), A nonparametric test for the several-sample problem, Ann. Math. Stat., Vol. 23, pp. 525—540. [474]

Крафт и Ле Кам (C. Kraft and L. Le Cam, 1956), A remark on the roots of the maximum likelihood equation, Ann. Math. Stat., Vol. 27, pp. 1174—1177. [372]

Крейг (A. T. Craig, 1932), On the distributions of certain statistics, Amer. Journ. Math., Vol. 54, pp. 353—366. [249]

Крейг (C. C. Craig, 1928), An application of Thiele's semiinvariants to the sampling problem, Metron, Vol. 7. № 4, pp. 3—74. [213]

Купмэн (B. O. Koopman, 1936), On distributions admitting sufficient statistics, Trans. Amer. Math. Soc., Vol. 39, pp. 399—409. [401]

Купмэнс (T. Koopmans, 1942), Serial correlation and quadratic forms in normal variables, Ann. Math. Stat., Vol. 13, pp. 14—33. [534]

Куэнуил (M. H. Quenouille, 1948), Some results in the testing of serial correlation coefficients, Biometrika, Vol. 35, pp. 261—267. [527, 534]

Лага (R. G. Laha, 1954), On a characterization of gamma distribution, Ann. Math. Stat., Vol. 25, pp. 784—787. [261]

Лаплас (P. S. Laplace, 1814), Théorie analytique de probabilités, изд. 2, Paris. [158, 217, 268, 374]

Леви (P. Lévy, 1925), Calcul des probabilités, Gauthier — Villars, Paris. [11, 128, 132]

Леви (P. Lévy, 1935), Propriétes asymptotiques des sommes des variables aléatoires indépendantes ou enchoinées, Journ. Math. Pures et Appl., Vol. 14, pp. 347—402. [268]

Леви (P. Lévy, 1937), Théorie de l'addition des variables aléatoires (3-е изд., 1954), Gauthier—Villars, Paris. [11, 134]

Ле Кам (L. Le Cam, 1956), On the asymptotic theory of estimation and testing hypotheses, Proc. Third Berkeley Symposium on Math. Stat. and Prob., Vol. I, pp. 129—156, University of California Press. [369]

Леман и Стейн (E. L. Lehmann and C. Stein, 1948), Most powerful tests of composite hypotheses, I. Normal distributions, Ann. Math. Stat., Vol. 19, pp. 495—516. [411]

Леман и Стейн (E. L. Lehmann and C. Stein, 1949), On the theory of some non-parametric hypotheses, Ann. Math. Stat., Vol. 20, pp. 28—45. [468]

Леман (E. L. Lehmann, 1950), Some principles of the theory of testing hypotheses, Ann. Math. Stat., Vol. 21, pp. 1—26. [411]

Леман и Шеффе (E. L. Lehmann and H. Scheffé, 1950), Completeness, similar regions and unbiased estimates, Sankhyā, Vol. 10, pp. 305—340. [411]

Леман (E. L. Lehmann, 1957), A theory of some multiple decision problems, Ann. Math. Stat., Vol. 28, pp. 1—25. [513]

Леман (E. L. Lehmann, 1959), Testing statistical hypotheses, John Wiley, New York (русский перевод: Проверка статистических гипотез, «Наука», М., 1964). [402, 411, 468]

Либерман и Оуэн (G. J. Lieberman and D. B. Owen, 1961), Tables of the hypergeometric probability distribution, Stanford University Press. [148]

Линдеберг (J. W. Lindeberg, 1922), Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, Math. Zeits., Vol. 15, pp. 211—225. [267—269]

Лорд (F. M. Lord, 1955). Sampling fluctuations resulting from the sampling of test items, Phychometrika, Vol. 20, pp. 1—22. [237]

Лотка (A. J. Lotka, 1939), A contribution to the theory of self-renewing aggregates, with special reference to industrial replacement, Ann. Math. Stat., Vol. 10, pp. 1—25. [204]

Лоэв (M. Lóeve, 1955), Probability theory, D. Van Nostrand, Princeton, N. J. (русский перевод: Теория вероятностей, ИЛ, М., 1962). [11, 32, 110, 135]

Лукач (E. Lukacs, 1942), A characterization of the normal distribution, Ann. Math. Stat., Vol. 13, pp. 91—93. [224, 262]

Льюс и Райфа (R. D. Luce and H. Raiffa, 1957), Games and decisions, John Wiley, New York (русский перевод: Игры и решения, ИЛ, М., 1961). [505]

Льюс (R. D. Luce, 1959), Individual choice behavior, John Wiley, New York. [434]

Ляпунов (A. Lyapunov, 1900), Sur une proposition de la théorie des probabilités, Bull. de l'Acad. Imp. des Sciences de St. Petersbourg, Vol. 13, pp. 359—386. [268]

Ляпунов (A. Lyapunov, 1901), Nouvelle forme de théorème sur la limite de probabilité, Mem. Acad. Science de St. Petersbourg, Vol. 12, pp. 1—24. [268, 277]

Маккарти (P. J. McCarthy, 1947), Approximate solutions for means and variances in a certain class of box problems, Ann. Math. Stat., Vol. 18, pp. 349—383. [158]

Макшейн (E. J. McShane, 1944), Integration, Princeton University Press. [356]

Макшейн и Ботс (E. J. McShane and T. Botts, 1959), Real analysis, D. Van Nostrand, Princeton, N. J. [32, 356]

Малмквист (S. Malmquist, 1954), On certain confidence contours for distribution functions, Ann. Math. Stat., Vol. 25, pp. 523—542. [445]

Манн (H. B. Mann, 1943), On the construction of sets of orthogonal Latin squares, Ann. Math. Stat., Vol. 14, pp. 401—414. [246]

Манн и Вальд (H. B. Mann and A. Wald, 1943), On the statistical treatment of linear stochastic difference equations, Econometrica, Vol. 11, pp. 173—220. [537]

Манн и Уитни (H. B. Mann and D. R. Whitney, 1947), On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat., Vol. 18, pp. 50—60. [465—467]

Манн (H. B. Mann, 1949), Analysis and design of experiments, Dover Publications, New York. [308]

Манроу (M. E. Munroe, 1953), Introduction to measure and integration, Addison—Wesley, Cambridge, Mass. [11]

А. А. Марков (1900, изд. 1), Исчисление вероятностей, изд. 4, ГИЗ, М., 1924. [268, 296]

Масси (F. J. Massey, 1950), A note on the estimation of a distribution function by confidence limits, Ann. Math. Stat., Vol. 21, pp. 116—119. [349, 350]

Масси (F. J. Massey, 1951), The distribution of the maximum deviation between two sample cumulative step functions, Ann. Math. Stat., Vol. 22, pp. 125—128. [461]

Матисен (H. C. Mathisen, 1943), A method of testing the hypothesis that two samples are from the same population, Ann. Math Stat., Vol. 14, pp. 188—194. [448]

Матусита (K. Matusita, 1957), Decision rule based on the distance for the classification problem, Ann. Inst. Stat. Math., Vol. 8, pp. 67—77. [263]

Махаланобис (P. C. Mahalanobis, 1930), On tests and measures of group devergence, Journ. Asiatic Soc. of Bengal, Vol. 26, pp. 541—588. [559]

Махаланобис (P. C. Mahalanobis, 1936), On the generalized distance in statistics, Proc. Nat. Inst. Sci. Calcutta, Vol. 12, pp. 49—55. [555, 559]

Махаланобис, Боуз и Рой (P. C. Mahalanobis, R. C. Bose and S. N. Roy, 1937), Normalization of statistical variates and the use of rectangular coordinates in the theory of sampling distributions, Sankhyā, Vol. 3, pp. 1—40. [550]

Махаланобис (P. C. Mahalanobis, 1940), A sample survey of the acreage under jute in Bengal, with discussion on planning of experiments, Proc. Second Indian Stat. Conference, Statistical Publishing Society, Calcutta. [478]

Мёрфи (R. B. Murphy, 1948), Non-parametric tolerance limits, Ann. Math. Stat., Vol. 19, pp. 581—589. [343]

Мизес (R. von Mises, 1931), Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik, Deuticke, Leipzig. [21, 445]

Молина (E. C. Molina, 1942), Poisson's exponential binomial limit, D. Van Nostrand, Princeton, N. J. [154]

Моран (P. A. P. Moran, 1948), Some theorems on time series, II., The significance of the serial correlation coefficient, Biometrika, Vol. 35, pp. 253—260. [534]

Моран, Уайтфилд и Даниелс (P. A. P. Moran, J. W. Whitfield and H. E. Daniels, 1950), Symposium on ranking methods, Journ. Roy. Stat Soc., Ser. B, Vol. 12, pp. 153—191. [435]

Моригути (S. Moriguti, 1954), Confidence limits for a variance component, Reports of statistical applications in research, Japanese Union of scientists and engineers, Vol. 3, № 2, pp. 29—41. [318]

Морс (P. M. Morse, 1958), Queues, inventories and maintenance, John Wiley, New York. [207]

Мостеллер (F. C. Mosteller, 1946), On some useful «inefficient» statistics, Ann. Math. Stat., Vol. 17, pp. 377—408.]285]

Мочли (J. W. Mauchly, 1940), Significance test for sphericity of a normal n-variate distribution, Ann. Math. Stat., Vol. 11, pp. 204—209. [598]

Муавр (A. De Moivre, 1718), The doctrine of chances, London. [268]

Муд (A. M. Mood, 1940), The distribution theory of runs, Ann. Math. Stat., Vol. 11, pp. 367—392. [158, 164]

Муд (A. M. Mood, 1946), On Hotteling's weighing problem, Ann. Math. Stat., Vol. 17, pp. 432—446. [297]

Муд (A. M. Mood, 1950), Introduction to the theory of statistics, McGraw-Hill, New York. [475]

Муд (A. M. Mood, 1951), On the distribution of the characteristic roots of normal second-moment matrices, Ann Math. Stat., Vol. 22, pp. 266—273. [566]

Муд (A. M. Mood, 1954), On the asympotic efficiency of certain nonparametric two-sample tests, Ann. Math. Stat., Vol. 25, pp. 514—522. [475]

Мэдоу (W. G. Madow, 1938), Contributions to the theory of multivariate statistical analysis, Trans. Amer. Math. Soc., Vol. 44, pp. 454—495. [550]

Наир (K. R. Nair, 1940), Table of confidence intervals for the median in samples from any continuous population, Sankhyā, Vol. 4, pp. 551—558. [340]

Нейман и Пирсон (J. Neyman and E. S. Pearson, 1928), On the use and interpretation of certain test criteria for purposes of statistical inference, Biometrika, Vol. 20A, Part I, pp. 175—240, Part II, pp. 263—294. [402, 410]

Нейман и Пирсон (J. Neyman and E. S. Pearson, 1933), On the problem of the most efficient tests of statistical hypotheses, Trans. Roy. Soc. London, Ser. A, Vol. 231, pp. 289—337. [402, 406, 410]

Нейман (J. Neyman, 1934), On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection, Journ. Roy. Stat. Soc., Vol. 97, pp. 558—625. [326]

Нейман (J. Neyman, 1935), Su un teorema concernente le cosidetti statistiche sufficienti, Giorn. Inst. Ital. Attuari, Vol. 6, pp. 320—334. [364]

Нейман (J. Neyman, 1937), Outline of a theory of statistical estimation based on the classical theory of probability, Phil. Trans. Roy. Soc. London, Ser. A, Vol. 236, pp, 333—380. [375]

Нейман (J. Neyman, 1939), On a new class of «contagious» distributions, applicable in entomology and bacteriology, Ann. Math. Stat., Vol. 10, pp. 35—57. [166]

фон Нейман (J. von Neumann, 1950), Functional operators, Vol. I, Princeton University Press. [60]

фон Нейман и Моргенстерн (J. von Neumann and O. Morgenstern, 1944), Theory of games and economic behavior, изд. 3, 1953, Princeton University Press. [505]

Огавара (M. Ogawara, 1951), A note on the test of serial correlation coefficients, Ann. Math. Stat., Vol. 22, pp. 115—118. [534]

Окамото (M. Okamoto, 1952), On a non-parametric test, Osaka Math. Journ., Vol. 4, pp. 77—85. [444]

Олдс (E. G. Olds, 1938), Distribution of sums of squares of rank differences for small numbers of individuals, Ann. Math. Stat., Vol. 9, pp. 133—148. [473]

Олкин (I. Olkin, 1951), On distribution problems in multivariate analysis, Institute of Statistics mimeograph series, Report № 8, University of North Carolina. [566]

Олкин и Рой (I. Olkin and S. N. Roy, 1954), On multivariate distribution theory, Ann. Math. Stat., Vol. 25, pp. 329—339. [566]

Олкин и Прэтт (I. Olkin and J. W. Pratt, 1958), A multivariate Chebyshev inequality, Ann. Math. Stat., Vol. 29, pp. 226—234. [124]

Парзен (E. Parzen, 1957), On consistent estimates of the spectrum of a stationary time series, Ann. Math. Stat., Vol. 28, pp. 329—348. [524]

Парзен (E. Parzen, 1960), Modern probability theory and its applications, John Wiley, New York. [261]

Пирсон и Хартли (E. S. Pearson and H. O. Hartley, 1954), Biometrika tables for statisticians, Vol. I, Cambridge University Press. [197, 199]

Пирсон (K. Pearson, 1900), On a criterion that a system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen in random sampling, Phil. Mag., Vol. 50, pp. 157—175. [273, 397]

Пирсон (K. Pearson, 1905), On the curves which are most suitable for describing the frequency of random samples of a population, Biometrika, Vol. 5, pp. 172—175. [185, 197]

Пирсон (K. Pearson, 1920), On the probable errors of frequency constants, Part III, Biometrika, Vol. 13, pp. 113—132. [285]

Пирсон (K. Pearson, editor, 1922), Tables of the incomplete gamma function, Cambridge University Press. [185]

Пирсон (K. Pearson, editor, 1934), Tables of the incomplete beta function, Cambridge University Press. [188, 201]

Питмэн (E. J. G. Pitman, 1936), Sufficient statistics and intrinsic accuracy, Proc. Cambr. Phil. Soc., Vol. 32, pp. 567—579. [401]

Питмэн (E. J. G. Pitman, 1937a), The «closest» estimates of statistical parameters, Proc. Cambr. Phil. Soc., Vol. 33, pp. 212—222. [261]

Питмэн (E. J. G. Pitman, 1937b), Significance tests which may be applied to samples from any populations, Journ. Roy. Stat. Soc., Ser. B, Vol. 4, pp. 119—130. [470]

Питмэн (E. J. G. Pitman, 1938), Significance tests which may be applied to samples from any populations, III. The analysis of variance test, Biometrika, Vol. 29, pp. 322—335. [471]

Последовательный анализ статистических данных: применения (Sequential analysis of statistical data: applications, 1945), Statistical research group, Columbia University, Columbia University Press. [498]

Пуассон (S. D. Poisson, 1837), Recherches sur la probabilité des jugements en matière criminelle et en matière civile, precédées des règles générales du calcul des probabilités, Paris. [153]

Райдер (P. R. Rider, 1955), The distribution of the product of maximum values in samples from a rectangular distribution, Journ. Amer. Stat. Assoc., Vol. 50, pp. 1142—1143. [201]

Райфа и Шлайфер (H. Raiffa and R. Schlaifer, 1961), Applied statistical decision theory, Graduate School of Business Administration, Harvard University. [505]

Рао (C. R. Rao, 1945), Information and accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc., Vol. 37, pp. 81—91. [361]

Рао (C. R. Rao, 1949), Sufficient statistics and minimum variance estimates, Proc. Cambr. Phil. Soc., Vol. 45, pp. 213—218. [366, 372, 401]

Рао (C. R. Rao, 1952), Advanced statistical methods in biometric research, John Wiley, New York. [539]

Раштон (S. Rushton, 1950), On a sequential t-test, Biometrika, Vol. 37, pp. 326—333. [477]

Раштон (S. Rushton, 1952), On a two-sided sequential t-test, Biometrika, Vol. 39, pp. 302—308. [477]

Реньи (A. Renyi, 1953), On the theory of order statistics, Act. Math. Acad. Sci. Hungary, Vol. 4, pp. 191—231. [261]

Рикер (W. E. Ricker, 1937), The concept of confidence or fiducial limits applied to the Poisson frequency distribution, Journ. Amer. Stat. Assoc., Vol. 32, pp. 349—356. [378]

Роббинс (H. Robbins, 1944a), On the measure of a random set, Ann. Math. Stat., Vol. 15, pp. 70—74. [261]

Роббинс (H. Robbins, 1944b), On distribution-free tolerance limits in random sampling, Ann. Math. Stat., Vol. 15, pp. 70—74. [343]

Роббинс (H. Robbins, 1948), Convergence of distributions, Ann. Math. Stat., Vol. 19, pp. 72—76. [124]

Роббинс (H. Robbins, 1952), Some aspects of the sequential design of experiments, Bull. Amer. Math. Soc., Vol. 58, pp. 527—535. [479]

Роббинс (H. Robbins, 1954), A remark on the joint distribution of cumulative sums, Ann. Math. Stat., Vol. 25, pp. 614—616. [124]

Роббинс и Монро (H. Robbins and S. Monro, 1951), A stochastic approximation method, Ann. Math. Stat., Vol. 22, pp. 400—407. [479]

Рой (S. N. Roy. 1939), p-statistics or some generalizations in analysis of variance appropriate to multivariate problems, Sankhyā, Vol. 4, pp. 381—396. [566]

Рой и Бозе (S. N. Roy and R. C. Bose, 1953), Simultaneous confidence interval estimation, Ann. Math. Stat., Vol. 24, pp. 513—536. [301]

Рой (S. N. Roy, 1954), Some further results in simultaneous confidence interval estimation, Ann. Math. Stat., Vol. 25, pp. 752—761. [301]

Рой (S. N. Roy, 1957), Some aspects of multivariate analysis, John Wiley, New York; Indian Statistical Institute, Calcutta. [539]

Ромиг (H. G. Romig, 1953), 50—100 Binomial tables, John Wiley, New York. [151]

Рэсч (G. Rasch, 1948), A functional equation for Wishart's distribution, Ann. Math. Stat., Vol. 19, pp. 262—266. [550]

Сайил (T. N. Thiele, 1903), Theory of observations, Layton, London (перепечатано в Ann. Math. Stat., Vol. 2, 1931, pp. 165—307). [127]

Сакс (S. Saks, 1937), Theory of the integral, 2-е изд., Stechert, New York (русский перевод: Теория интеграла, ИЛ, М., 1949). [32, 356]

Свед и Эйзенхарт (F. S. Swed and C. Eisenhart, 1943), Tables for testing randomness of grouping in a sequence of alternatives, Ann. Math. Stat., Vol. 14, pp. 66—87. [163]

Свердруп (E. Sverdrup, 1947), Derivation of the Wishart distribution of the second order sample moments by straightforward integration of a multiple integral, Skand. Aktuar., Vol. 30, pp. 151—166. [550]

Слуцкий (E. Slutzky, 1937), The summation of random causes as the source of cyclic processes, Econometrica, Vol. 5, pp. 105—146. [537]

Смирнов (N. Smirnov, 1935), Über die Verteilung des allgemeinen Gliedes in der Variationsreihe, Metron, Vol. 12, № 2, pp. 59—81. [283, 285]

Н. В. Смирнов (1939a), Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках, Бюлл. Московского ун-та, том 2, № 2, стр. 3—14. [345, 348, 448, 460]

Н. В. Смирнов (1939b), Об уклонениях эмпирической кривой распределения, Матем. сборник, 6(48), стр. 3—26. [445]

Смит (W. L. Smith, 1958), Renewal theory and its ramifications, Journ. Roy. Stat. Soc., Ser. B, Vol. 20, pp. 243—302. [204]

Снедекор (G. W. Snedecor, 1937), Statistical methods, Iowa State College Press, Ames, Iowa. [201]

Сомервилл (P. N. Somerville, 1958), Tables for obtaining non-parametric tolerance limits, Ann. Math. Stat., Vol. 29, pp. 599—601. [344]

Стейн (C. Stein, 1945), A two-sample test for a linear hypothesis whose power is independent of the variance, Ann. Math. Stat., Vol. 16, pp. 243—258. [499]

Стейн (C. Stein, 1946), A note on cumulative sums, Ann. Math. Stat., Vol. 17, pp. 498—499. [488]

Стейн и Вальд (C. Stein and A. Wald, 1947), Sequential confidence intervals for the mean of a normal distribution with known variance, Ann. Math. Stat., Vol. 18, pp. 427—433. [499]

Стефан и Маккарти (F. F. Stephan and P. J. McCarthy, 1958), Sampling opinion, John Wiley, New York. [323]

Стивенс (W. L. Stevens, 1939), Distributions of groups in a sequence of alternatives, Ann. Eugen., Vol. 9, pp. 10—17. [163]

Сукхатмэ (P. V. Sukhatme, 1954), Sampling theory of surveys with applications, Iowa State College Press, Ames, Iowa. [323]

Сэвидж (I. R. Savage, 1953), Bibliography on nonparametric statistics and related topics, Journ. Amer. Stat. Assoc., Vol. 48, pp. 844—906. [247, 435]

Сэвидж (L. J. Savage, 1954), The foundations of statistics, John Wiley, New York. [505]

Сэйвер (S. R. Savur, 1937), The use of the median in tests of significance, Proc. Indian Acad. Sci., Section A, Vol. 5, pp. 564—576. [340]

Сюй (P. L. Hsu, 1938), Notes on Hotteling's generalized T, Ann. Math. Stat., Vol. 9, pp. 231—243. [599]

Сюй (P. L. Hsu, 1939a), A new proof of the joint product moment distribution, Proc. Cambr. Phil. Soc., Vol. 35, pp. 336—338. [550]

Сюй (P. L. Hsu, 1939b), On the distribution of roots of certain determinantal equations, Ann. Eugen., Vol. 9, pp. 250—258. [566]

Сюй (P. L. Hsu, 1945a), The approximate distributions of the mean and variance of a sample of independent variables, Ann. Math. Stat., Vol. 16, pp. 1—29. [277]

Сюй (P. L. Hsu, 1945b), The asymptotic distribution of ratios, Ann. Math. Stat., Vol. 16, pp. 204—210. [277]

Таблицы биномиального распределения (Tables of the cumulative binomial probabilities, 1952), Army Ordnance Corps, ORDP. [151]

Таблицы биномиального распределения (Tables of the cumulative binomial probability distribution, 1955), Harvard computation laboratory, Harvard University. [151]

Таблицы биномиального распределения (Tables of the binomial probability distribution, 1949), National Bureau of Standards, Appl. Math. Series, Vol. 6. [151]

Таблицы функций вероятности (Tables of probability functions, 1942), National Bureau of Standards, Vol. 2, New York. [171]

Техника статистического анализа (Techniques of statistical analysis, 1947), Statistical research group, Columbia University, McGraw-Hill, New York. [498]

Тинтнер (G. Tintner, 1940), The variate difference method, Principia Press, Bloomington, Indiana. [527]

Типпет (L. H. C. Tippet, 1925), On the extreme individuals and the range of samples taken from a normal population, Biometrika, Vol. 17, pp. 364—387. [259]

Томпсон (W. R. Thompson, 1936), On confidence ranges for the median and other expectation distributions for populations of unknown distribution form, Ann. Math. Stat., Vol. 7, pp. 122—128. [340]

Тьюки (J. W. Tukey, 1946), An inequality for deviations from medians, Ann. Math. Stat., Vol. 17, pp. 75—78. [124]

Тьюки (J. W. Tukey, 1947), Nonparametric estimation, II. Statistically equivalent blocks and tolerance regions — the continuous case, Ann. Math. Stat., Vol. 18, pp. 529—539. [251]

Тьюки (J. W. Tukey, 1948), Nonparametric estimation, III. Statistically equivalent blocks and multivariate tolerance regions — the discontinuous case, Ann. Math. Stat., Vol. 19, pp. 30—39. [255]

Тьюки (J. W. Tukey, 1949a), The sampling theory of power spectrum estimates, Proceedings on applications of autocorrelation analysis to physical problems (NAVEXOS—P—735). [524, 525]

Тьюки (J. W. Tukey, 1949b), Moments of random group size distributions, Ann. Math. Stat., Vol. 20, pp. 523—539. [167, 168],

Тьюки (J. W. Tukey, 1950), Some sampling simplified, Journ. Amer. Stat. Assoc., Vol. 45, pp. 501—519. [223, 238]

Тьюки (J. W. Tukey, 1953), The problem of multiple comparisons, не опубликовано, Princeton University. [301, 302, 304, 305]

Тьюки (J. W. Tukey, 1956a), Keeping moment-like sampling computations simple, Ann. Math. Stat., Vol. 27, pp. 37—54. [223]

Тьюки (J. W. Tukey, 1956b), Variances of variance components: I. Balanced designs, Ann. Math. Stat., Vol. 27, pp. 722—736. [318]

Тьюки (J. W. Tukey, 1957a), Variances of variance components: II. The unbalanced single classification, Ann. Math. Stat., Vol. 28, pp. 43—56. [318]

Тьюки (J. W. Tukey, 1957b), Variances of variance components: III. Third moments in a balanced single classification, Ann. Math. Stat., Vol. 28, pp. 378—384. [318]

Тьюки (J. W. Tukey, 1957c), Some examples with fiducial relevance, Ann. Math. Stat., Vol. 28, pp. 687—695. [380]

Тэнг (P. C. Tang, 1938), The power function of the analysis of variance tests with tables and illustrations of their use, Stat. Res. Mem., University College, London, Vol. 2, pp. 128—149. [259]

Уиддер (D. V. Widder, 1947), Advanced calculus, Prentice-Hall, New York. [69]

Уилкоксон (F. Wilcoxon, 1945), Individual comparisons by ranking methods, Biometrics, Vol. 1, pp. 80—83. [465]

Уилкс (S. S. Wilks, 1932), Certain generalizations in the analysis of variance, Biometrika, Vol. 24, pp. 471—494. [552, 562]

Уилкс (S. S. Wilks, 1934), Moment-generating operators for determinants of product moments in samples from a normal system, Ann. Math., Vol. 35, pp. 312—340. [599]

Уилкс (S. S. Wilks, 1935), On the independence of k sets of normally distributed statistical variables, Econometrica, Vol. 3, pp. 309—326. [597]

Уилкс (S. S. Wilks, 1938a), The large-sample distribution of the likelihood ratio for testing composite hypotheses, Ann. Math. Stat., Vol. 9, pp. 60—62. [411, 427]

Уилкс (S. S. Wilks, 1938b), Shortest average confidence intervals from large samples, Ann. Math. Stat., Vol. 9, pp. 166—175. [384]

Уилкс и Дэйли (S. S. Wilks and J. F. Daly, 1939), An optimum property of confidence regions associated with the likelihood function, Ann. Math. Stat., Vol. 10, pp. 225—235. [396]

Уилкс (S. S. Wilks, 1941), On the determination of sample sizes for setting tolerance limits, Ann. Math. Stat., Vol. 12, pp. 91—96. [343, 352]

Уилкс (S. S. Wilks, 1942), Statistical prediction with special reference to the problem of tolerance limits, Ann. Math. Stat., Vol. 13, pp. 400—409. [343]

Уилкс (S. S. Wilks, 1946), Sample criteria for testing equality of means, equality of variances and equality of covariances in a normal multivariate distribution, Ann. Math. Stat., Vol. 17, pp. 257—281. [599]

Уилкс (S. S. Wilks, 1948), Order statistics, Bull. Amer. Math. Soc., Vol. 54, pp. 5—50. [435]

Уилкс (S. S. Wilks, 1959a), Nonparametric statistical inference, Probability and Statistics, The Harald Cramér volume, Almquist and Wiksell, Stockholm. [435]

Уилкс (S. S. Wilks, 1959b), Recurrence of extreme observations, Journ. Austral. Math. Soc., Vol. 1, pp. 106—112. [262]

Уилкс (S. S. Wilks, 1960a), Multidimensional statistical scatter, Contributions to probability and statistics in honour of Harold Hotelling, Stanford University Press. [546]

Уилкс (S. S. Wilks, 1960b), A two-stage scheme for sampling without replacement, Bull. de l'Institut intern. de Statist., Vol. 37, № 2, pp. 241—248. [328]

Уилкс (S. S. Wilks, 1961), A combinatorical test for the problem of two samples from continuous distributions. Proc. Fourth Berkeley Symp. on Math. Stat. and Prob., Vol. 1, University of California Press. [448, 457]

Уилсон (E. B. Wilson, 1927), Probable inference, the law of succession, and statistical inference, Journ. Amer. Stat. Assoc., Vol. 27, pp. 209—212. [375]

Уильямс (J. D. Williams, 1946), An approximation to the probability integral, Ann. Math. Stat., Vol. 17, pp. 363—365. [202]

Уиттекер и Ватсон (E. T. Whittaker and G. N. Watson, 1946), A course in modern analysis, MacMillan, New York (русский перевод: Курс современного анализа, ч. 1, 2, Физматгиз, М., 1963). [131, 192]

Уиттл (P. Whittle, 1951), Hypotheses testing in time series, University of Uppsala. [534]

Уишарт (J. Wishart, 1928), The generalized product moment distribution in samples from a normal multivariate population, Biometrika, Vol. 20A, pp. 32—52. [546, 550]

Уишарт и Бартлетт (J. Wishart and M. S. Bartlett. 1932), The distribution of the second order moment statistics in a normal system, Proc. Cambr. Phil. Soc., Vol. 28, pp. 455—459. [550]

Уоллес (D. L. Wallace, 1958), Asymptotic approximations to distributions, Ann. Math. Stat., Vol. 29, pp. 635—654. [277]

Уоллис (W. A. Wallis, 1939), The correlation ratio for ranked data, Journ. Amer. Stat. Assoc., Vol. 34, pp. 533—538. [474]

Уолш (J. E. Walsh, 1946), Some order statistic distributions for sample of size 4, Ann. Math. Stat., Vol. 17, pp. 246—248. [264]

Уолш (J. E. Walsh, 1949), Some significance tests for the median which are valid under very general conditions, Ann. Math. Stat., Vol. 20, pp. 64—81. [437]

Уоркинг и Хотеллинг (H. Working and H. Hotelling, 1929), Application of the theory of error to the interpretation of trends, Journ. Amer. Stat Assoc., March supplement, pp. 73—85. [307]

Уормлитон (R. Wormleighton, 1959), Some tests of permutation symmetry, Ann. Math. Stat., Vol. 30, pp. 1005—1017. [474]

Уэйсс (L. Weiss, 1961), Statistical decision theory, Mc Graw-Hill, New York. [505]

Уэлч (B. L. Welch, 1937), On the z-test in randomized blocks and Latin squares, Biometrika, Vol. 29, pp. 21—52. [471]

Феллер (W. Feller, 1935), Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung, Math. Zeits., Vol. 40, pp. 521—559. [268, 269]

Феллер (W. Feller, 1948), On the Kolmogorov — Smirnov limit theorems for empirical distributions, Ann. Math. Stat., Vol. 19, pp. 177—189. [350]

Феллер (W. Feller, 1957), An introduction to probability theory and its applications, 2-е изд., John Wiley, New York (русский перевод: Введение в теорию вероятностей и ее приложения, «Мир», М., 1964). [11, 110, 207]

Фергюсон (T. S. Ferguson, 1958), A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities, Ann. Math. Stat., Vol. 29, pp. 1046—1062. [369]

Филлер (E. C. Fieller, 1932), The distribution of the index in a normal bivariate population, Biometrika, Vol. 24, pp. 428—440. [202]

Фишер (R. A. Fisher, 1915), Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, Biometrika, Vol. 10, pp. 507—521. [208, 591]

Фишер (R. A. Fisher, 1922), On the mathematical foundations of theoretical statistics, Phil. Trans. Roy. Soc. London, Ser. A, Vol. 222, pp. 309—368. [359, 361, 362, 364, 369, 371, 389]

Фишер (R. A. Fisher, 1924), On a distribution yielding the error functions of several well-known statistics, Proc. Int. Math. Congress, Vol. II, Toronto, pp. 805—813. [201]

Фишер (R. A. Fisher, 1925a), Statistical methods for research workers, 1-е изд. (12-е изд., 1954), Oliver and Boyd, Edinburgh (русский перевод: Статистические методы для исследователей, Госстатиздат, М., 1958). [191, 199, 201, 287, 311]

Фишер (R. A. Fisher, 1925b), Theory of statistical estimation, Proc. Cambr. Phil. Soc., Vol. 22, pp. 700—725. [359]

Фишер (R. A. Fisher, 1926a), Applications of «Student's» distribution, Metron, Vol. 5, № 4, pp. 90—104. [224]

Фишер (R. A. Fisher, 1926b), On the random sequence, Quart. Journ. Roy. Meteor. Soc., Vol. 52, pp. 250—258. [468, 470]

Фишер (R. A. Fisher, 1928a), Moments and product moments of sampling distributions, Proc. London Math. Soc., Vol. 30, pp. 199—238. [213]

Фишер (R. A. Fisher, 1928b), The general sampling distribution of the multiple correlation coefficient, Proc. Roy. Soc. London, Ser. A, Vol. 121, pp. 654—673. [258, 596]

Фишер и Типпет (R. A. Fisher and L. H. C. Tippett, 1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample, Proc. Cambr. Phil. Soc., Vol. 24, pp. 180—190. [283]

Фишер (R. A. Fisher, 1929), Tests of significance in harmonic analysis, Proc. Roy. Soc. London, Ser. A, Vol. 125, pp. 54—59. [529, 532]

Фишер (R. A. Fisher, 1935a), The design of experiments, Oliver and Boyd, Edinburgh. [308]

Фишер (R. A. Fisher, 1935b), The fiducial argument in statistical inference, Ann. Eugen., Vol. 6, pp. 391—398. [378, 380]

Фишер (R. A. Fisher, 1938), The statistical utilization of multiple measurements, Ann. Eugen., Vol. 8, pp. 376—386. [572]

Фишер (R. A. Fisher, 1939), The sampling distribution of some statistics obtained from non-linear equations, Ann. Eugen., Vol. 9, pp. 238—249. [566]

Фишер и Иэйтс (R. A. Fisher and F. Yates, 1938), Statistical tables for biological, agricultural and medical research (изд. 5, 1957), Oliver and Boyd, Edinburgh. [308]

Фреше (M. Fréchet, 1927), Sur la loi de probabilité de l'écart maximum, Ann. Soc. Polonaise Math., Vol 6, pp. 92—116. [283]

Фридман (M. Friedman, 1937), The use of ranks to avoid the assumption of normality implicit in the analysis of variance, Journ. Amer. Stat. Assoc., Vol. 32, pp. 675—701. [474]

Фрэзер (D. A. S. Fraser, 1951), Sequentially determined statistically equivalent blocks, Ann. Math. Stat., Vol. 22, pp. 372—381. [254]

Фрэзер (D. A. S. Fraser, 1953), Nonparametric tolerance regions, Ann. Math. Stat., Vol. 24, pp. 44—55. [254]

Фрэзер и Гаттмен (D. A. S. Fraser and I. Guttman, 1956), Tolerance regions, Ann. Math. Stat., Vol. 27, pp. 162—179. [254]

Фрэзер (D. A. S. Fraser, 1957), Nonparametric methods in statistics, John Wiley, New York. [247, 435, 468]

Халмош (P. R. Halmos, 1946), The theory of unbiased estimation, Ann. Math. Stat., Vol. 17, pp. 34—43. [291]

Халмош и Сэвидж (P. R. Halmos and L. J. Savage, 1949), Application of the Radon — Nikodym theorem to the theory of sufficient statistics, Ann. Math. Stat., Vol. 20, pp. 225—241. [365]

Халмош (P. R. Halmos, 1950), Measure theory, D. Van Nostrand, Princeton, N. J. (русский перевод: Теория меры, ИЛ, М., 1953). [11, 26, 27, 32, 36]

Хансен, Гурвиц, Мэдоу (M. H. Hansen, W. N. Hurwitz and W. G. Madow, 1953), Sample survey methods and theory, Vol. I and II, John Wiley, New York. [323]

Харрис (T. E. Harris, 1948), Branching processes, Ann. Math. Stat., Vol. 19, pp. 474—494. [145]

Хартли и Дэвид (H. O. Hartley and H. A. David, 1954), Universal bounds for mean range and extreme observation, Ann. Math. Stat., Vol. 25, pp. 85—99. [261]

Хельмерт (F. R. Helmert, 1876a), Ueber die Wahrscheinlichkeit der Potensummen der Beobachtungsfehler und über einige damit im Zusammenhange stehende Fragen, Zeits. für Math. und Phys., Vol. 21, pp. 192—218. [221]

Хельмерт (F. R. Helmert, 1876b), Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers directer Beobachtungen gleicher Genauigkeit, Astron. Nachr., Vol. 88, pp. 112—131. [224]

Хинчин (A. Khintchine, 1929), Sur la loi des grands nombres, Comp. Rend. Acad. Sci., Vol. 188, pp. 477—479. [121, 265]

Холдэйн (J. B. S. Haldane, 1945), On a method of estimating frequencies, Biometrika, Vol. 33, pp. 222—225. [158, 503]

Хотеллинг (H. Hotelling, 1931), The generalization of Student's ratio, Ann. Math. Stat., Vol. 2, pp. 360—378. [301, 556, 593]

Хотеллинг (H. Hotelling, 1933), Analysis of a complex of statistical variables into principal components, Journ. Educ. Psych., Vol. 24, pp. 417—441, 498—520. [563, 566]

Хотеллинг (H. Hotelling, 1935), The most predictable criterion, Journ. Educ. Psych., Vol. 26, pp. 139—142. [586]

Хотеллинг и Пэбст (H. Hotelling and M. R. Pabst, 1936), Rank correlation and tests of significance involving no assumption of normality, Ann. Math. Stat., Vol. 7, pp. 29—43. [473]

Хотеллинг (H. Hotelling, 1941), Experimental determination of the maximum of a function, Ann. Math. Stat., Vol. 12, pp. 20—45. [478]

Хотеллинг (H. Hotelling, 1944), Some improvements in weighing and other experimental techniques, Ann. Math. Stat., Vol. 15, pp. 297—306. [297]

Хотеллинг (H. Hotelling, 1953), New light on the correlation coefficient and its transforms, Journ. Roy. Stat. Soc., Ser. B, Vol. 15, pp. 193—232. [592]

Хузурбазар (V. S. Huzurbazar, 1948), The likelihood equation, consistency and the maxima of the likelihood function, Ann. Eugen., Vol. 14, pp. 185—200. [369]

Хук (R. Hooke, 1956a), Symmetric functions of a two-way array, Ann. Math. Stat., Vol. 27, pp. 55—79. [238]

Хук (R. Hooke, 1956b), Some applications of bipolykays to the estimation of variance components and their moments, Ann. Math. Stat., Vol. 27, pp. 80—98. [318]

Хэннан (E. J. Hannan, 1955), Exact tests for serial correlation, Biometrika, Vol. 42, pp. 133—142. [534]

Чебышев (P. L. Chebyshev, 1867), On mean values, Journ. Math. Pures et Appl., Vol. 12, pp. 177—184. [87]

Чебышев (P. L. Chebyshev, 1890), Sur deux théorèmes relatifs aux probabilités, Acta Math., Vol. 14, pp. 305—315. [268]

Чернов и Мозес (H. Chernoff and L. E. Moses, 1959), Elementary decision theory, John Wiley, New York (русский перевод: Элементарная теория статистических решений, Сов. радио, М., 1962). [505]

Чжун и Де Лёэри (J. H. Chung and D. B. De Lury, 1950), Confidence limits for the hypergeometric distribution, University of Toronto Press. [378]

Чжун (K. L. Chung, 1941), On the probability of the occurrence of at least m events among n arbitrary events, Ann. Math. Stat., Vol. 12, pp. 328—338. [41]

Чжун (K. L. Chung, 1946), The approximate distribution of Student's statistic, Ann. Math. Stat., Vol. 17, pp. 447—465. [277]

Чжун и Феллер (K. L. Chung and W. Feller, 1949), Fluctuations in coin tossing, Proc. Nat. Acad. Sci. USA, Vol. 35, pp. 605—608. [476]

Чэпмен и Роббинс (D. C. Chapman and H. Robbins, 1951), Minimum variance estimation without regularity assumptions, Ann. Math. Stat., Vol. 22, pp. 581—586. [361, 401]

Шастер (A. Schuster, 1898), On the investigation of hidden periodicities with application to a supposed 26-day period of meteorological phenomena, Terrestrial Magnetism, Vol. 3, pp. 13—41. [529]

Шеннон (C. E. Shannon, 1948), A mathematical theory of communication, Bell System Techn. Journ., Vol. 27, pp. 379—423, 623—656 (русский перевод — в сборнике «Работы по теории информации и кибернетике», ИЛ, М., 1963). [417]

Шеппард (W. F. Sheppard, 1898), On the calculation of the most probable values of frequency-constants for data arranged according to equidistant divisions of a scale, Proc. London Math. Soc., Vol. 29, pp. 353—380. [336]

Шерман (B. Sherman, 1950), A random variable related to the spacing of sample values, Ann. Math. Stat., Vol. 21, pp. 339—361. [445]

Шеффе (H. Scheffé, 1943), Statistical inference in the nonparametric case, Ann. Math. Stat., Vol. 14, pp. 305—332. [435]

Шеффе (H. Scheffé, 1947), A useful convergence theorem for probability distributions, Ann. Math. Stat., Vol. 18, pp. 434—438. [124]

Шеффе (H. Scheffé, 1953). A method for judging all contrasts in the analysis of variance, Biometrika, Vol. 40, pp. 87—104. [301, 302]

Шеффе (H. Scheffé, 1959), The analysis of variance, John Wiley, New York (русский перевод: Дисперсионный анализ, Физматгиз, М., 1963). [308, 318]

Шрикхэнд (S. S. Shrikhande, 1952), On the dual of some balanced incomplete block designs, Biometrics, Vol. 8, pp. 66—72. [244]

Шьюарт (W. A. Shewart, 1931), Economic control of quality of manufactured product, D. Van Nostrand, Princeton, N. J. [343]

Эджворт (F. Y. Edgeworth, 1905), The law of error, Trans. Cambr. Phil. Soc., Vol. 20, pp. 36—65. [276]

Эйзенхарт (C. Eisenhart, 1947), The assumptions underlying the analysis of variance, Biometrics, Vol. 3, pp. 1—21. [318]

Элфвинг (G. Elfving, 1947), The asymptotical distribution of range in samples from a normal population, Biometrika, Vol. 34, pp. 111—119. [260]

Эндрюс (F. C. Andrews, 1954), Asymptotic behavior of some rank tests for analysis of variance, Ann. Math. Stat., Vol. 25, pp. 724—736. [474]

Энскомб (F. J. Anscombe, 1953), Sequential estimation, Journ. Roy. Stat. Soc., Ser. B, Vol. 15, pp. 1—29. [499]

Эпстейн (B. Epstein, 1954), Tables for the distribution of the number of exceedances, Ann. Math. Stat., Vol. 25, pp. 762—768. [262]

Эссеен (C. E. Esseen, 1944), Fourier analysis of distribution functions: A mathematical study of the Laplace — Gaussian law, Acta Math., Vol. 77, pp. 1—125. [277]

Юл (G. U. Yule, 1924), A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis, Phil. Trans. Roy. Soc., London, Ser. B, Vol. 213, pp. 21—87. [206]

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ