Д. ХУДСОН

СТАТИСТИКА ДЛЯ ФИЗИКОВ

Лекции по теории вероятностей и элементарной статистике

Второе дополненное издание

Перевод с английского

ИЗДАТЕЛЬСТВО «МИР»

МОСКВА 1970

STATISTICS

Lectures on Elementary Statistics and Probability

by

DEREK J. HUDSON

Geneva 1964

Предисловие редактора перевода к первому изданию

Среди разделов математики, завоевавших прочное место в арсенале современной физики, важную роль играют теория вероятностей и математическая статистика. С формированием молекулярно-кинетических представлений о строении вещества и созданием теории микромира статистика превратилась в неотъемлемую часть аппарата теоретической физики. Одновременно статистика сделалась важным инструментом и экспериментальных исследований. В многообразных применениях теории вероятностей и математической статистики можно разграничить три типа взаимоотношений этих разделов математики с физикой:

1) Создание математического аппарата таких наук, как статистическая физика и квантовая механика.

2) Описание случайных процессов.

3) Обработка результатов наблюдений.

В основе этих взаимоотношений лежат совершенно объективные основания. В первом случае — это статистический характер ряда фундаментальных законов природы, во втором — случайный характер событий, образующих сложный физический процесс, и, наконец, в третьем — экспериментальный характер физики. Последнее означает, что в физике имеют дело прежде всего с результатами измерений, которые по своей природе представляют собой случайные величины.

Для каждого из этих аспектов характерно использование в значительной степени специфического круга вопросов теории вероятностей и математической статистики. Что касается первой из отмеченных выше точек соприкосновения статистики с физикой, то в большинстве курсов теоре-

тической физики основное внимание уделяется, естественно, изложению физических проблем, а не обоснованию используемого математического аппарата. Теория случайных процессов как самостоятельная дисциплина возникла сравнительно недавно, хотя отдельные применения этой теории давно известны и, разумеется, выходят далеко за рамки физики [59].

Существенные изменения за последние 10—15 лет произошли и в методах обработки результатов экспериментальных исследований. Статистика оказалась мощным средством извлечения ценной информации из экспериментальных данных. В современных исследованиях не часто удается непосредственно измерять физические величины, представляющие интерес. И лишь статистический анализ позволяет делать надежные выводы о многих явлениях. Здесь уместно напомнить, что если у физиков есть основания судить о том, что́ происходит за промежутки времени масштаба 10^{-22} *сек* или на расстояниях порядка 10^{-13} *см*, то только потому, что они в совершенстве владеют статистическими методами обработки опытных данных. Если еще сравнительно недавно основное употребление статистики при обработке результатов заключалось в отыскании средних значений и их погрешностей, то усложнение экспериментов и их косвенный характер заставили физиков искать поддержки в более сложных разделах, посвященных методам оценки параметров и проверки гипотез. Стал широко применяться анализ регрессий. Все эти вопросы достаточно детально разработаны в математической статистике и изложены в ряде прекрасных руководств [21, 36, 57] и особенно в фундаментальных монографиях Т. Андерсона [40], Э. Лемана [45], Г. Шеффе [58]. Однако большинство этих книг в силу их большого объема и последовательного характера изложения требует от читателя, заинтересованного прежде всего в использовании теории для решения конкретных проблем, значительных затрат времени и сил. Для современного специалиста, захваченного высоким темпом собственных исследований и до предела загруженного работой по своей узкой специальности, академический путь совершенствования своих знаний путем всестороннего ознакомления с литературой оказывается малоподходящим. На самом

деле трудно рассчитывать, что люди, получающие информацию даже в своей области не из журналов или книг, а, как правило, в результате переписки, посещения семинаров и т. п., смогут уделить достаточно времени изучению статистики. Но поскольку потребность в расширении знаний по статистике весьма велика, то время от времени в различных лабораториях и научных центрах предпринимаются попытки создать курс статистики для специалиста-физика, содержащий краткое и доступное изложение основ и результатов, проиллюстрированных примерами применений.

Такие попытки предпринимались в США — в Калифорнийском университете [31, 50], Брукхейвенской национальной лаборатории [53], Университете штата Мэриленд [16], в СССР — в Объединенном институте ядерных исследований [22] и в Европейском центре ядерных исследований, ЦЕРНе (Женева) [19]. Аналогичные тенденции явились причиной появления и фундаментальной монографии [61], а также работ [49, 52]. Большинство упомянутых работ — это запись лекций, посвященных избранным вопросам математической статистики. Этим они, на наш взгляд, выгодно отличаются от объемистых руководств, которые помимо основных сведений содержат множество конкретных рецептов и скорее носят характер учебников, а не справочников.

Настоящая книга — также запись лекций, прочитанных для сотрудников ЦЕРНа. Мы думаем, что уже сам по себе факт ознакомления с опытом одного из крупнейших мировых научных центров окажет большую пользу нашим специалистам. Кроме того, лекции Худсона, на наш взгляд, удовлетворяют существующим запросам. Эти лекции в основном посвящены трем очень важным для физиков вопросам: проверке гипотез, методу максимального правдоподобия, анализу регрессий. В изложении отсутствует традиционная вводная часть, посвященная основам линейной алгебры, теории вероятностей и математической статистики, которые, как правило, хорошо известны из учебных курсов.

В процессе изложения основ теории вероятностей, которым посвящены первые две главы, автор сразу же вводит и ряд статистических понятий, таких, как точечные

и интервальные оценки, различные статистики, используемые для оценки параметров и построения критериев проверки. Гл. 3 содержит описание ряда специальных распределений, широко используемых при проверке статистических гипотез. Большой интерес представляет использование графического построения функции правдоподобия для получения оценок, о котором подробно рассказано в гл. 4. Прекрасно изложены и многие важные стороны метода наименьших квадратов, одним из важнейших приложений которого является описание данных с помощью полинома. Процедура анализа регрессий и выбора формы полинома, наилучшим способом описывающего данные, детально излагается в гл. 5.

Мы включили в качестве дополнения к книге Худсона сводку статистических данных (автор Малви), которая может служить в качестве справочника по обработке экспериментальных результатов. Учитывая, что лекции нуждались в значительной редакционной обработке, мы позволили себе в ряде случаев внести некоторые уточнения или исправления непосредственно в текст. Это же касается рубрикации разделов текста. Мы надеемся, что лаконизм и насыщенность окупят отсутствие строгости изложения. Для тех, кто хотел бы познакомиться с более обстоятельным обоснованием затронутых вопросов, мы рекомендуем прекрасную монографию Линника [55]. Помимо цитированных в тексте руководств по теории вероятностей и статистике, можно порекомендовать известные книги Бернштейна [48], Гнеденко [51], Дунина-Барковского и Смирнова [57]. Книга Худсона, безусловно, представит интерес не только для физиков, но и для специалистов других областей науки и техники, а также инженеров и всех, кто использует статистику при обработке результатов своих экспериментальных исследований.

Е. Лейкин

Предисловие редактора перевода ко второму изданию

Интерес, проявленный читателями к лекциям Д. Худсона три года назад, не только послужил оправданием их выпуска в переводе на русский язык, но и стимулировал нынешнее переиздание. За истекшее время приложение статистики к физике продолжало расширяться. Не последнюю роль здесь играло увеличение удельного веса систем автоматической обработки экспериментальных данных, например создание обширного программного обеспечения систем обработки информации с различных фильмовых камер. Поскольку ряд новых приложений статистики не вошел в лекции Д. Худсона, мы воспользовались новым изданием, чтобы пополнить книгу изложением разделов, отвечающих современным требованиям. В книге введено дополнение II, в качестве которого использован ряд разделов из прекрасной монографии по статистике Гамильтона [53]. Незначительно увеличив объем книги, мы тем не менее получили возможность познакомить читателя с вопросами, важность которых отмечалась еще в лекциях Худсона (см. стр. 172). Таким образом, новое дополнение является логическим продолжением основного текста. Подготовленный же читатель может воспользоваться этим материалом независимо от текста лекций.

Дополнение содержит обобщение метода наименьших квадратов на случай коррелированных измерений и наличия связи между параметрами модели. Оно ликвидирует имевшийся в лекциях пробел в изложении практически важных проблем нелинейной зависимости от параметров. Наконец, в дополнении обсуждается подход к проблеме планирования эксперимента и другие вопросы. Кроме того, новое издание книги содержит отсутствовавшие ранее

в ней таблицы распределения $\mathcal{F}$, крайне важного в задачах проверки гипотез; несколько расширилась и библиография. Для удобства пользования книгой обозначения в дополнении II приведены в соответствие с обозначениями в основном тексте книги.

Е. Лейкин

Из предисловия автора

Настоящие лекции были прочитаны в 1963—1964 гг. в ЦЕРНе, Женева. В первой части лекций собраны некоторые важнейшие результаты теории случайных величин. Большинство доказательств основано на аксиомах теории вероятностей в их классическом виде. Здесь же приводятся дискретные и непрерывные распределения, в том числе распределения χ^2 и t. Для вычисления новых распределений используется метод производящих функций моментов; этот же метод использован для доказательства центральной предельной теоремы.

В процессе изложения вводится ряд статистических понятий, на которых основан более глубокий подход к формулировке статистических выводов. Преследуемая при этом цель заключается в том, чтобы развить эти подходы во второй половине лекций, где основной упор сделан на достаточно строгое рассмотрение ряда более серьезных проблем.

Задача проверки гипотез решается с применением распределения t; для проверки нормальности распределения предлагается способ, основанный на использовании вероятностной бумаги; рассмотрено распределение $\mathcal{F}$ и получено выражение для его плотности.

Принцип максимального правдоподобия вводится с интуитивной позиции при попытке найти наиболее правдоподобное объяснение полученным экспериментальным данным.

Приводятся графические изображения некоторых функций правдоподобия, причем чаще всего в полулогарифмическом масштабе. Теорема Бейеса показывает, как следует модифицировать функцию правдоподобия при

наличии априорной информации об исследуемых параметрах.

Глава, посвященная методу наименьших квадратов, имеет дело с задачей аппроксимации полиномом экспериментальных результатов, которые представлены в виде набора пар (x_i, y_i), где $\{x_i\}$ фиксированы заранее, а $\{y_i\}$ являются случайными величинами. Для получения фундаментальных результатов (в частности, для доказательства теоремы Гаусса — Маркова) широко используется матричная алгебра. Вводятся ортогональные полиномы, позволяющие упростить анализ и контролировать ошибки округления при численных расчетах. В том случае, когда степень аппроксимирующего полинома неизвестна, выбрать ее помогает критерий $\mathscr{F}$.

ГЛАВА ПЕРВАЯ

Введение в теорию вероятностей

Теория вероятностей — это один из разделов чистой математики. Строится эта теория дедуктивно, исходя из некоторых аксиом и определений. Наиболее строгий подход связан с использованием теории множеств, теории меры и интеграла Лебега. Обычно начинают с построения «элементарной теории вероятностей», в которой рассматриваются случайные события с конечным числом исходов. Например, если мы бросаем игральную кость, то имеется всего шесть возможных исходов, а именно: 1, 2, 3, 4, 5 или 6. Затем теория распространяется на случай, когда число возможных исходов бесконечно. Например, если величина x обозначает человеческий рост, то она может принимать бесконечное число различных значений. С подходом, основанным на теории меры, читатель может ознакомиться по книге Колмогорова [23].

Применение теорем к решению различных задач теории вероятностей связано с использованием сочетаний, перестановок, операций суммирования и интегрирования. Вычисления обычно не очень сложны. Некоторые применяющиеся в теории вероятностей методы, такие, например, как преобразование Лапласа, используются и в других разделах математики.

В противоположность теории вероятностей статистика — это раздел прикладной математики. Для нее характерно главным образом индуктивное построение, поскольку в этом случае мы идем в обратном направлении — от наблюдения события к гипотезе. При этом аргументация основывается на выводах теории вероятностей, всестороннее знание которой, таким образом, оказывается совершенно необходимым.

Пример 1. *Типичная задача теории вероятностей.* Когда подбрасывается монета, то имеется известная вероятность p, что выпадет «орел», и вероятность $(1 - p)$, что выпадет «решка». Какова вероятность того, что в результате N бросаний «орел» выпадет n раз?

Используя биномиальное распределение, о котором будет идти речь позднее, мы получим следующий результат:

$$\mathcal{P}(n) = C_N^n p^n (1-p)^{N-n}.$$

Пример 2. *Типичная задача статистики.* Монета подбрасывается N раз, при этом «орел» выпадает n раз. Что можно сказать о неизвестном параметре p?

Очевидно, нельзя надеяться получить на этот вопрос столь же определенный ответ, что и в предыдущем случае. С самого начала мы знаем, что $0 \leqslant p \leqslant 1$. Кроме того, $p \neq 0$, если $n \neq 0$, и $p \neq 1$, если $n \neq N$.

Рассмотрим понятие *наиболее правдоподобного значения* параметра. В данном случае мы могли бы сказать, что *наиболее правдоподобное значение* p равно n/N. Затем следовало бы рассмотреть и другие почти столь же правдоподобные значения p. В результате получим малый интервал

$$p_1 < \frac{n}{N} < p_2, \tag{1}$$

который, как мы надеемся, будет содержать истинное значение p.

Пусть $\delta p = p_2 - p_1$. Чем больше δp, тем с большей достоверностью p попадает в интервал (1). С другой стороны, более широкий интервал дает нам меньшую информацию относительно самой величины p. Таким образом, в статистическом анализе всегда присутствует принципиальная неопределенность. Мы можем во всяком случае рассчитывать, что оценим эту неопределенность.

§ 1. Определение вероятности

К сожалению, между специалистами по теории вероятностей и статистиками нет согласия относительно того, как наилучшим образом определить понятие вероятности. Попытки ввести определение, основанное на строгой логи-

ке, просто показали, что имеется несколько способов формулировки этого фундаментального понятия. В таком случае выбор наиболее приемлемого из них оказывается в значительной мере делом вкуса.

Комбинаторное определение

Если событие может приводить к N различным равновозможным исходам и если в n случаях проявляется признак A, то вероятность A есть n/N.

Такое определение можно подвергнуть критике на том основании, что выражение «равновозможно» на самом деле означает «равновероятно», и в рассуждении, таким образом, содержится порочный круг.

Пример. Пусть совершенно случайным образом отобраны 23 человека. Какова вероятность того, что по крайней мере у двоих окажется один и тот же день рождения?

Полное число возможных комбинаций из 23 дней рождения (в году 365 дней) равно

$$N = 365^{23}.$$

Число комбинаций, в которых имеется 23 *различных* дня рождения, равно

$$N - n = 365 \times 364 \times \ldots \times 343.$$

Вероятность того, что все дни рождения различны, равна

$$\frac{N-n}{N} = \frac{365 \times 364 \times \ldots \times 343}{365^{23}}.$$

Вероятность того, что *не все* дни рождения различны, равна

$$\frac{n}{N} = 1 - \frac{N-n}{N} = 1 - 1\left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\ldots$$

$$\ldots\left(1 - \frac{22}{365}\right) \approx 0{,}5.$$

Таким образом, для группы из 23 или более человек вероятность того, что по крайней мере у двоих окажется общий день рождения, превышает 0,5 (см. [10], § 11.3). Для группы из 57 человек эта вероятность равна 0,99.

Частотное определение

Мы наблюдаем некое событие E, повторяющееся большое число (N) раз. При этом в n случаях событие обладает признаком A. Если исходы событий в этой последовательности взаимно независимы, то вероятность признака A определяется как

$$\mathscr{P}(A) = \lim_{N\to\infty} \frac{n}{N}.$$

В примере 1 (стр. 14) мы могли бы предположить, что с монетой был предварительно произведен гипотетический эксперимент. Бросая монету N раз и наблюдая n раз выпадение «орла», мы полагаем [17]

$$\mathscr{P}(\text{«орел»}) = p = \lim_{N\to\infty} \frac{n}{N}.$$

Современное определение, основанное на теории меры

Введем пространство Ω, элементы которого e_1, e_2, ... назовем элементарными событиями. Определим в нем неотрицательную меру, называемую вероятностью, со свойствами:

1) $\mathscr{P}(\Omega) = 1$, т. е. вероятность того, что «наблюдаемое событие принадлежит к числу тех, которые могут наблюдаться», равна единице.

2) Если множества элементарных событий A и B не имеют общих элементов, то $\mathscr{P}(A \cup B) = \mathscr{P}(A) + \mathscr{P}(B)$, т. е. вероятность того, что событие принадлежит либо множеству A, либо множеству B, равна сумме вероятностей обнаружить событие отдельно в множестве A и множестве B.

Рассмотрим теперь *следствие*, которое служит примером использования этих аксиом. Пусть Φ — пустое множество событий, иначе говоря, Φ означает отсутствие событий. Тогда $(\Omega \cup \Phi) = \Omega$, и Ω не имеет общих элементов с Φ.

Следовательно,

$$\mathscr{P}(\Omega \cup \Phi) = \mathscr{P}(\Omega) + \mathscr{P}(\Phi) = \mathscr{P}(\Omega)$$

и

$$\mathscr{P}(\Phi) = 0.$$

Иными словами, если известно, что одно из событий должно обязательно происходить, то вероятность отсутствия событий равна нулю. (Подробнее об этом см. у Колмогорова [23].)

«Субъективное» определение

Вероятность того, что событие произошло или произойдет, служит иногда мерой нашей уверенности в происходящем.

Пример. У человека одно из трех заболеваний: D_1, D_2 или D_3, но неизвестно какое. Субъективный человек может говорить о «вероятности» заболевания D_2. Для объективного человека такое утверждение бессмысленно, поскольку он должен был бы сказать:

$$\mathscr{P}(D_2) = 0 \text{ или } 1.$$

§ 2. Основные законы теории вероятностей

Пусть $C = (A \cup B) \equiv A + B$ есть множество элементарных событий, которые встречаются либо в A, либо в B, либо и в том и в другом. Пусть далее $D = (A \cap B) \equiv AB$ есть множество событий, которые встречаются одновременно как в A, так и в B.

Сложение вероятностей

$$\mathscr{P}(A \cup B) = \mathscr{P}(A) + \mathscr{P}(B) - \mathscr{P}(A \cap B),$$

или

$$\mathscr{P}(A + B) = \mathscr{P}(A) + \mathscr{P}(B) - \mathscr{P}(AB).$$

Итак, чтобы найти вероятность того, что событие принадлежит множеству A или множеству B, мы прежде всего должны сложить вероятности A и B ([27], § 2.8). Однако некоторые события, принадлежащие одновременно как A, так и B, были бы тогда сосчитаны дважды; поэтому необходимо из суммы вероятностей вычесть $\mathscr{P}(AB)$.

Пример. Вытащим из колоды, содержащей 52 карты, одну. Пусть A означает «король», B означает «трефовая масть». Тогда вероятность того, что данная карта окажется либо королем, либо трефовой мастью, равна

$$\mathscr{P}(A+B) = \mathscr{P}(\text{«король»}) + \mathscr{P}(\text{«трефовая масть»}) - \mathscr{P}(\text{«король трефовой масти»}) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52}.$$

Следствие 1. Для m множеств $E_1, E_2, \ldots, E_m$ имеем

$$\mathscr{P}(E_1 + E_2 + \ldots + E_m) = \sum_i \mathscr{P}(E_i) - \sum_{i<j} \mathscr{P}(E_iE_j) + \sum_{i<j<k} \mathscr{P}(E_iE_jE_k) - \ldots + (-1)^{m+1}\mathscr{P}(E_1E_2 \ldots E_m). \quad (2)$$

Доказательство может быть проведено методом индукции по m.

Следствие 2. Если события взаимно исключают друг друга, т. е. если $(E_i \cap E_j) = E_iE_j = \Phi$, то

$$\mathscr{P}\left(\sum_i E_i\right) = \sum_i \mathscr{P}(E_i). \quad (3)$$

В частности, $\mathscr{P}(E_1 + E_2) = \mathscr{P}(E_1) + \mathscr{P}(E_2)$, если $E_1E_2 = \Phi$.

Пример. Игра в «снэп». Два игрока имеют одинаковые колоды, содержащие по m карт. Они тасуют карты и затем одновременно открывают сверху по одной. Если карты совпадают, игроки выкрикивают «снэп». Какова вероятность, что совпадений не будет?

Пусть E_i означает совпадение карт i-й пары. Тогда вероятность отсутствия совпадений в игре имеет следующий вид:

$$\mathscr{P}(\text{«нет»}) = 1 - \mathscr{P}(E_1 + E_2 + \ldots + E_m).$$

В одной из колод карты оказываются расположенными в некоторой последовательности от первой до m-й. Чтобы произошло событие E_i, i-я карта первой колоды должна находиться и во второй колоде на i-м месте. Число способов расположения остальных $(m-1)$ карт равно $(m-1)!$

Всего же существует $m!$ способов расположения карт в колоде, так что

$$\mathscr{P}(E_i) = \frac{(m-1)!}{m!} = \frac{1}{m}.$$

Подобным образом для $(E_i \cap E_j)$ можно найти $n = (m-2)!$, а $N = m!$, так что

$$\mathscr{P}(E_iE_j) = \frac{(m-2)!}{m!} = \frac{1}{m(m-1)}.$$

Далее,

$$\mathscr{P}(E_iE_jE_k) = \frac{1}{m(m-1)(m-2)} \text{ и, наконец,}$$

$$\mathscr{P}(E_1E_2 \ldots E_m) = \frac{1}{m!}.$$

Следовательно,

$$\mathscr{P}\left(\sum_{i=1}^{m} E_i\right) = \sum \mathscr{P}(E_i) - \sum \mathscr{P}(E_iE_j) +$$

$$+ \sum \mathscr{P}(E_iE_jE_k) - \ldots$$

$$\ldots + (-1)^{m+1}\mathscr{P}(E_1E_2 \ldots E_m) =$$

$$= m\mathscr{P}(E_i) - C_m^2\mathscr{P}(E_iE_j) + C_m^3\mathscr{P}(E_iE_jE_k) - \ldots$$

$$\ldots + (-1)^{m+1}\mathscr{P}(E_1E_2 \ldots E_m) =$$

$$= m\frac{1}{m} - C_m^2\frac{(m-2)!}{m!} +$$

$$+ C_m^3\frac{(m-3)!}{m!} - \ldots + (-1)^{m+1}\frac{1}{m!} =$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \ldots + (-1)^{m+1}\frac{1}{m!}.$$

Значит,

$$\mathscr{P}(\text{«нет»}) = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \ldots + (-1)^{m+1}\frac{1}{m!} \to$$

$$\to \frac{1}{e} = 0{,}3679 \ldots .$$

Ряд сходится к $1/e$ очень быстро. Так, уже при $m = 7$ $\mathscr{P}$(«нет») $= 0{,}3679$. Другими словами, значение $\mathscr{P}$(«нет») практически не зависит от объема колоды карт.

Условная вероятность

До сих пор можно было полагать, что теория вероятностей представляет собой не более чем специальный случай абстрактной теории меры. Но мы введем сейчас еще одно определение, которое приведет к отличию вероятности от всех прочих мер.

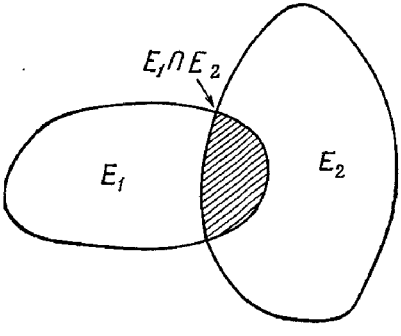
Предположим, что событие E_1 может осуществляться n_1 способами, событие E_2 — n_2 способами, а событие $(E_1 \cap E_2)$ — n_{12} способами (причем полное число возможных исходов равно N).

Тогда

$$\mathscr{P}(E_1) = \frac{n_1}{N}, \qquad \mathscr{P}(E_1E_2) = \frac{n_{12}}{N},$$
$$\mathscr{P}(E_1E_2) = \frac{n_1}{N}\frac{n_{12}}{n_1} = \mathscr{P}(E_1)\frac{n_{12}}{n_1}. \tag{4}$$

Из n_1 случаев, в которых происходило событие E_1, в относительной доле случаев, равной n_{12}/n_1, происходит также и событие E_2.

Таким образом, n_{12}/n_1 есть вероятность E_2 при условии, что произошло событие E_1. Эта вероятность записывается в виде $\mathscr{P}(E_2/E_1)$ и называется *условной вероятностью* события E_2 при условии, что произошло событие E_1.



Фиг. 1.

Умножение вероятностей

Из формулы (4) получаем

$$\mathscr{P}(E_1E_2) = \mathscr{P}(E_1)\cdot\mathscr{P}(E_2/E_1). \tag{5}$$

На языке теории множеств имеем (фиг. 1)

$$E_1E_2 \in E_1;$$

следовательно,

$$\mathscr{P}(E_1E_2) \leqslant \mathscr{P}(E_1).$$

Величина, на которую нужно умножить меру множества E_1, чтобы получить меру заштрихованного множества $E_1 \cap E_2$, есть новая мера, называемая $\mathscr{P}(E_2/E_1)$.

Соотношение (5) можно переписать в виде

$$\mathcal{P}(E_2/E_1) = \frac{\mathcal{P}(E_1E_2)}{\mathcal{P}(E_1)}, \qquad (6)$$

полезном для практического вычисления условных вероятностей.

Независимость событий

Если тот факт, что событие E_1 произошло, не влияет на вероятность события E_2, то говорят, что событие E_2 *не зависит* от события E_1. Тогда $\mathcal{P}(E_2/E_1) = \mathcal{P}(E_2)$ и, в частности,

$$\mathcal{P}(E_1E_2) = \mathcal{P}(E_1)\,\mathcal{P}(E_2). \qquad (7)$$

Следствие. В случае трех событий закон умножения вероятностей имеет вид

$$\mathcal{P}(E_1E_2E_3) = \mathcal{P}(E_1)\,\mathcal{P}(E_2/E_1)\,\mathcal{P}(E_3/E_1E_2).$$

Три события взаимно независимы в том и только в том случае, если

$$\mathcal{P}(E_1E_2E_3) = \mathcal{P}(E_1)\,\mathcal{P}(E_2)\,\mathcal{P}(E_3).$$

Пример 1. *Независимые события.* Из колоды в 52 карты выбирается одна. Пусть E_1 означает «туз», а E_2 — «пиковая масть». Тогда

$$\mathcal{P}(E_1) = \frac{4}{52}, \qquad \mathcal{P}(E_2) = \frac{13}{52}.$$

Так как $\mathcal{P}(E_1E_2) = \mathcal{P}(\text{«туз пик»}) = 1/52$, то

$$\mathcal{P}(E_1E_2) = \mathcal{P}(E_1)\,\mathcal{P}(E_2),$$

и, следовательно, E_1 и E_2 — независимые события.

Пример 2. *Зависимые события.* В колоду карт добавлен джокер (карта, не имеющая масти). Тогда

$$\mathcal{P}(E_1) = \frac{4}{53}, \quad \mathcal{P}(E_2) = \frac{13}{53}, \quad \text{а} \quad \mathcal{P}(E_1E_2) = \frac{1}{53}.$$

В то же время $\mathcal{P}(E_1)\,\mathcal{P}(E_2) = 1/53 - 1/53^2$, и, следовательно, события E_1 и E_2 не независимы. Этот результат

объясняется следующим рассуждением. Если мы знаем, что данная карта — туз, то она не может быть джокером, следовательно, мы уже получили некоторую информацию для определения масти.

§ 3. Дискретные распределения

Событие E с несколькими различными исходами называется *случайным*. Если E — переменная *числовая* величина, то такая величина называется *случайной*.

Если существует лишь конечное число ν исходов (e_i) события E, то каждому e_i приписывается вероятность P_i, т. е.

$$\mathcal{P}(E = e_i) = P_i.$$

Число ν может быть также бесконечным счетным, т. е. $\nu = \aleph_0$. Множество значений вероятностей P_i мы называем *дискретным распределением* вероятности [1]).

Биномиальное распределение ($\nu = 2$)

Пример. Пусть событие E есть результат бросания монеты. Имеется два возможных исхода E: e_1 = «орел», e_2 = «решка». Если мы обозначим оба эти исхода соответствующими числами, например «орел» — 1, «решка» — 0, то тем самым каждому бросанию монеты будет сопоставлена случайная величина. При описании таких событий обычно принята следующая терминология: «благоприятный исход» и «неблагоприятный исход». Назовем выпадение «орла» благоприятным исходом, а выпадение «решки» — неблагоприятным исходом. Найдем вероятность r благоприятных исходов при N бросаниях; в этом случае r также будет случайной величиной. Вероятность того, что при первых r бросаниях выпадет «орел», а при последующих $N - r$ бросаниях — «решка», равна

$$p^r (1-p)^{N-r}.$$

Такой порядок выпадения «орла» является лишь одним из C_N^r возможных способов реализации r благоприятных

[1]) См. также [27], гл. 3.

исходов при N испытаниях. Следовательно, полная вероятность равна

$$\mathcal{P}(r) = C_N^r p^r (1-p)^{N-r}. \tag{8}$$

Множество вероятностей $\{\mathcal{P}(r)\}$ называется *биномиальным распределением*. Заметим, что

$$\sum_{r=0}^{N} \mathcal{P}(r) = [p + (1-p)]^N = 1,$$

как и для любого дискретного распределения. Итак, случайная величина r подчиняется распределению, которое относится к *типу* (8), и полностью характеризуется двумя *параметрами* N и p.

Предельные формы биномиального распределения

1. Если p фиксировано, то распределение (8) стремится к нормальному распределению при стремлении N к бесконечности. Это утверждает центральная предельная теорема, которая будет доказана позже.

2. Если Np фиксировано, то распределение (8) стремится к распределению Пуассона, когда N стремится к бесконечности. Это утверждение будет доказано в конце данного параграфа.

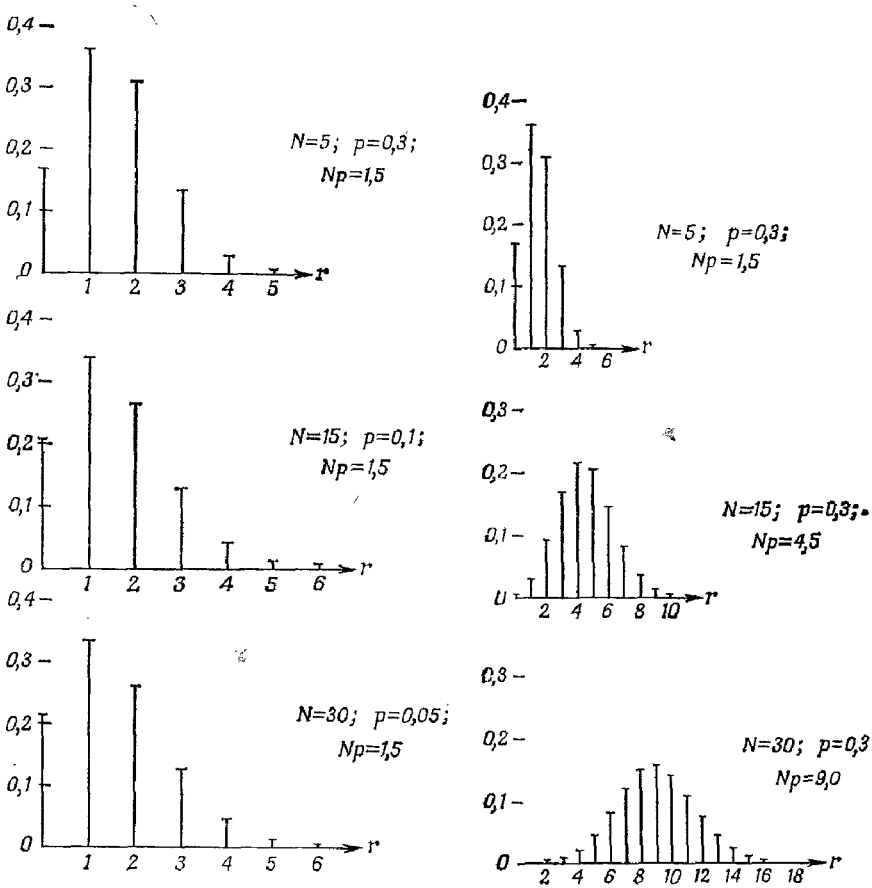
Фиг. 2 иллюстрирует различные степени приближения биномиального распределения к его предельным формам. Распределение с параметрами $N = 30$, $p = 0{,}05$ весьма близко к распределению Пуассона со средним значением 1,5. Распределение с параметрами $N = 30$, $p = 0{,}3$ напоминает по своему характеру нормальное распределение со средним значением 9,0.

Распределение Пуассона ($\nu = \nu_0$)

Допустим, нас интересует вероятность того, что за данный промежуток времени произойдет r событий. При этом выполняются следующие предположения:

1. Произойдет или не произойдет событие в момент t, не зависит от событий, предшествующих моменту t.

2. Вероятность отдельного события за малый интервал времени δt возрастает пропорционально длительности



Фиг. 2.

этого интервала. Иными словами, вероятность отдельного события за промежуток времени $(t,\ t+\delta t)$ равна $\mu\cdot\delta t + o(\delta t)$, где $o(\delta t)$ — величина более высокого порядка малости по сравнению с δt.

3. Вероятность двух или большего числа событий за тот же промежуток времени $(t,\ t+\delta t)$ равна нулю: $0 + o(\delta t)$.

Вывод распределения Пуассона. Вычислим в этих предположениях вероятность того, что в интервале $(0,\ t)$

произойдет r событий. Для этого сравним величины $P(t)$ и $P(t+\delta t)$ и затем составим дифференциальное уравнение для $P(t)$.

За промежуток времени $(0, t+\delta t)$ не произойдет ни одного события, если не будет событий в интервалах $(0, t)$ и $(t, t+\delta t)$, т. е.

$$P_0(t+\delta t) = P_0(t)(1-\mu\delta t + o(\delta t)).$$

Следовательно,

$$\frac{P_0(t+\delta t)-P_0(t)}{\delta t} = -\mu P_0(t) + \frac{o(\delta t)}{\delta t}.$$

Устремив δt к нулю, получим дифференциальное уравнение

$$\frac{dP_0(t)}{dt} = -\mu P_0(t).$$

Решение этого уравнения имеет вид

$$\ln P_0(t) = -\mu t + C,$$

или

$$P_0(t) = Ae^{-\mu t}.$$

При $t=0$ имеем $P_0(0)=1$; таким образом, $A=1$ и

$$P_0(t) = e^{-\mu t}.$$

Если $r=1$, то имеются две возможности: либо событие произошло в интервале $(0, t)$, либо в интервале $(t, t+\delta t)$. Отсюда

$$P_1(t+\delta t) = P_1(t)(1-\mu\delta t + o(\delta t)) + P_0(t)(\mu\delta t + o(\delta t)),$$

$$\frac{P_1(t+\delta t)-P_1(t)}{\delta t} = -\mu P_1(t) + \mu P_0(t) + \frac{o(\delta t)}{\delta t}.$$

Полагая $\delta t \to 0$, получаем

$$\frac{dP_1(t)}{dt} = -\mu P_1(t) + \mu e^{-\mu t}.$$

Решение этого уравнения имеет вид

$$P_1(t) = \mu t e^{-\mu t}. \tag{9}$$

В общем случае (для $r \geqslant 1$)

$$\frac{dP_r(t)}{dt} = -\mu P_r(t) + \mu P_{r-1}(t).$$

Решение для общего случая имеет вид

$$P_r(t) = \frac{(\mu t)^r}{r!} e^{-\mu t}. \qquad (9')$$

Заметим, что $\sum_{r=0}^{\infty} P_r(t) = e^{-\mu t} \sum_{r=0}^{\infty} \frac{(\mu t)^r}{r!} = 1$, как и следовало ожидать. Распределение (9′) содержит единственный параметр μ.

Вывод распределения Пуассона как предельной формы биномиального распределения. Пусть радиоактивный источник испускает в среднем μt частиц в интервале $(0, t)$. Этот интервал можно разбить на большое число N подинтервалов, в каждом из которых имеется вероятность p наблюдать одну частицу. Вероятность наблюдать r частиц в N подинтервалах (благоприятный исход) равна

$$P_r(t) = C_N^r p^r (1-p)^{N-r}$$

при среднем числе $Np = \mu t$. Так как $p = \mu t/N$, то

$$P_r(t) = \frac{N(N-1)\ldots(N-r+1)}{r!}\left(\frac{\mu t}{N}\right)^r\left(1-\frac{\mu t}{N}\right)^{N-r} =$$

$$= \frac{N(N-1)\ldots(N-r+1)}{N^r}\frac{(\mu t)^r}{r!}\left(1-\frac{\mu t}{N}\right)^{N-r} =$$

$$= 1\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\ldots\left(1-\frac{r-1}{N}\right)\frac{(\mu t)^r}{r!}\times$$

$$\times\frac{\left(1-\frac{\mu t}{N}\right)^N}{\left(1-\frac{\mu t}{N}\right)^r}.$$

При увеличении числа подинтервалов N получим

$$\lim_{N\to\infty} P_r(t) = \frac{(\mu t)^r}{r!} e^{-\mu t}.$$

В качестве иного примера рассмотрим район некоего озера, в котором имеется какое-то количество рыбы. Нас интересует распределение вероятности обнаружить в этом районе r рыб (при среднем значении μ). Будет ли этим распределением $P_r = (\mu^r/r!)\, e^{-\mu}$?

Предположение 1 (см. стр. 23), которое в этом случае относится не ко времени, а к пространству, равноценно допущению, что наличие или отсутствие одной рыбы в данном районе озера не зависит от того, где находится другая рыба. Но такое допущение на самом деле вряд ли справедливо, и поэтому распределение Пуассона здесь как раз неприменимо.

Позже мы проанализируем с помощью распределения χ^2 судьбу офицеров прусской кавалерии, гибель которых подчиняется распределению Пуассона.

§ 4. Непрерывные распределения

Пусть X — случайная величина, используемая как мера человеческого роста. Тогда X может принимать любое значение, лежащее в интервале, скажем, от 0 до 300 *см*; таким образом, число возможных значений величины X оказывается бесконечным и несчетным. Поэтому мы называем X *непрерывной случайной величиной.*

Парадокс нулевой вероятности

При попытке вычислить вероятность $\mathscr{P}(X=x)$ мы сталкиваемся с некоторой логической проблемой.

Если

$$\mathscr{P}(X=x)\neq 0,$$

то мы получим

$$\sum_x \mathscr{P}(X=x)=\infty>1$$

для x в интервале (0,300), как бы ни было мало $\mathscr{P}(X=x)$. С другой стороны, если $\mathscr{P}(X=x)=0$, то, казалось бы,

$$\sum_x \mathscr{P}(X=x)=0$$

для x в интервале (0,300). Однако последнее утверждение неверно, поскольку мы производим суммирование по несчетному множеству точек x.

Обойдем эту трудность, рассмотрев вероятность того, что X заключено в некотором малом интервале. Пусть

$$f(x)\,dx=\mathscr{P}(x<X<x+dx)\geqslant 0, \qquad (10)$$

т. е.

$$f(x)\,dx = \mathcal{P}(0 \leqslant X \leqslant x + dx) - \mathcal{P}(0 \leqslant X \leqslant x) = \\ = F(x + dx) - F(x).$$

Следовательно,

$$f(x) = \frac{dF(x)}{dx} \geqslant 0. \tag{11}$$

Таким образом, $f(x)$ — это скорость возрастания вероятности $F(x)$; $f(x)$ называют *плотностью вероятности* или *функцией непрерывного распределения*. Значит,

$$F(x_0) = \int_0^{x_0} f(x)\,dx,$$

откуда следует

$$\int_0^{300} f(x)\,dx = \mathcal{P}(0 \leqslant X \leqslant 300) = 1.$$

В наиболее общем случае область изменения x может распространяться от $-\infty$ до $+\infty$, и тогда

$$F(x) = \mathcal{P}(-\infty \leqslant X \leqslant x) = \int_{-\infty}^{x} f(x)\,dx, \tag{12}$$

причем

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1. \tag{13}$$

Мы называем $F(x)$ *распределением накопленной вероятности*. Это неубывающая функция, предельные значения которой:

$$F(-\infty) = 0, \qquad F(+\infty) = 1.$$

Равномерное (прямоугольное) распределение

Если все значения случайной величины, лежащие в интервале от a до b, равновероятны, то

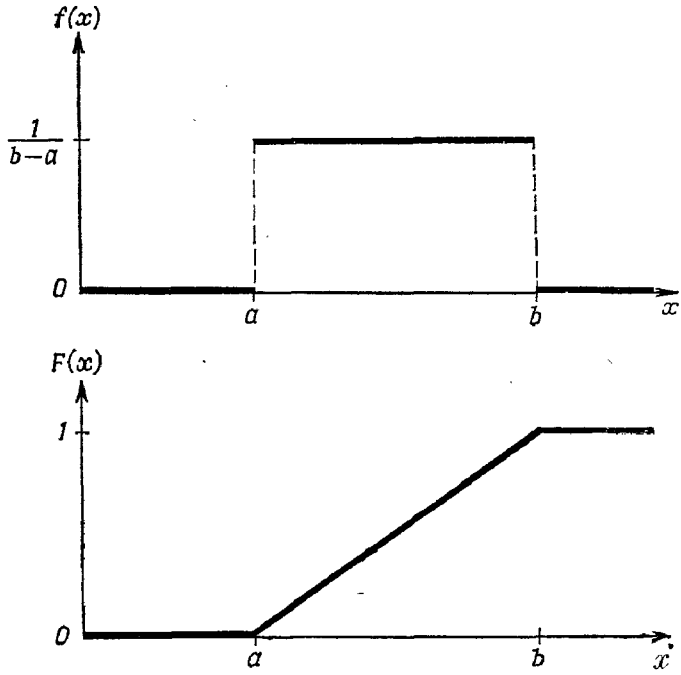
$$f(x) = \begin{cases} 0 & \text{при} \quad -\infty < x < a, \\ \dfrac{1}{b-a} & \text{при} \quad a \leqslant x \leqslant b, \\ 0 & \text{при} \quad b < x < +\infty. \end{cases} \tag{14}$$

Распределение накопленной вероятности:

$$F(x)=\begin{cases}0 & \text{при} \quad -\infty<x<a,\\ \dfrac{x-a}{b-a} & \text{при} \quad a\leqslant x\leqslant b,\\ 1 & \text{при} \quad b<x<+\infty.\end{cases}$$

Вид этих функций изображен на фиг. 3.

Пример. Длина деревянного бруска измерена с точностью до 1 *см*. Пусть x — *ошибка*, т. е. разность между



Фиг. 3.

истинной и измеренной длиной. Все значения x между $-0{,}5$ и $+0{,}5$ равновероятны, так что

$$f(x)=1 \quad \text{при} \quad -0{,}5\leqslant x\leqslant +0{,}5.$$

Теперь допустим, что *единицей* измерения длины является метр. Мы произвели измерение с точностью до 0,01 *м*,

и ошибка x' лежит между $-0{,}005$ и $+0{,}005$ *м*; следовательно,

$$f(x') = 100 \quad \text{при} \quad -0{,}005 \leqslant x' \leqslant +0{,}005.$$

Таков первый пример распределения функции случайной величины x'. Здесь x' есть функция x, а именно $x'(x) = 0{,}01x$.

Нормальное распределение

Одним из наиболее важных распределений, встречающихся в статистике, является нормальное, или гауссово, распределение [1]). Оно имеет вид симметричной колоколообразной кривой, которая распространяется до бесконечности как в положительном, так и в отрицательном направлениях.

Плотность этого распределения имеет вид

$$f(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-(x-\mu)^2/2\sigma^2}. \tag{15}$$

Распределение (15) полностью определяется заданием двух параметров: μ и σ. И наоборот, во многих статистических задачах, предполагая, что распределение является нормальным, пытаются оценить его параметры μ и σ. В более сложных случаях сама величина μ есть функция неизвестных параметров $\mu(\beta_0, \beta_1, \ldots, \beta_r)$, которые подлежат определению; например, $\mu = \beta_0 + \beta_1 t$, где t — независимая переменная.

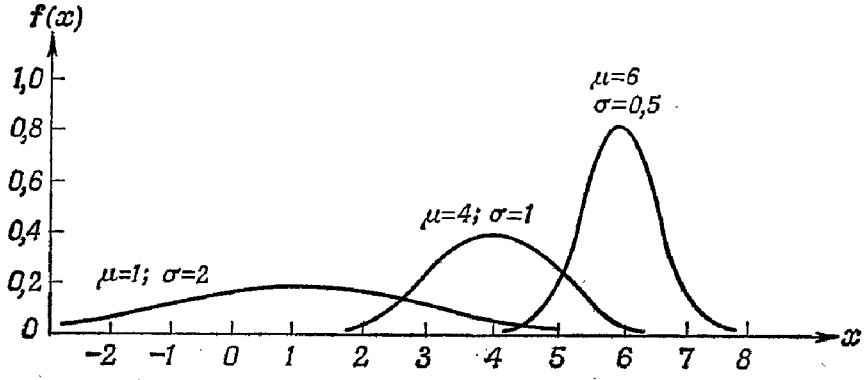
На фиг. 4 изображены функции $f(x)$ с различными значениями μ и σ.

Распределение накопленной вероятности представляет собой S-образную кривую

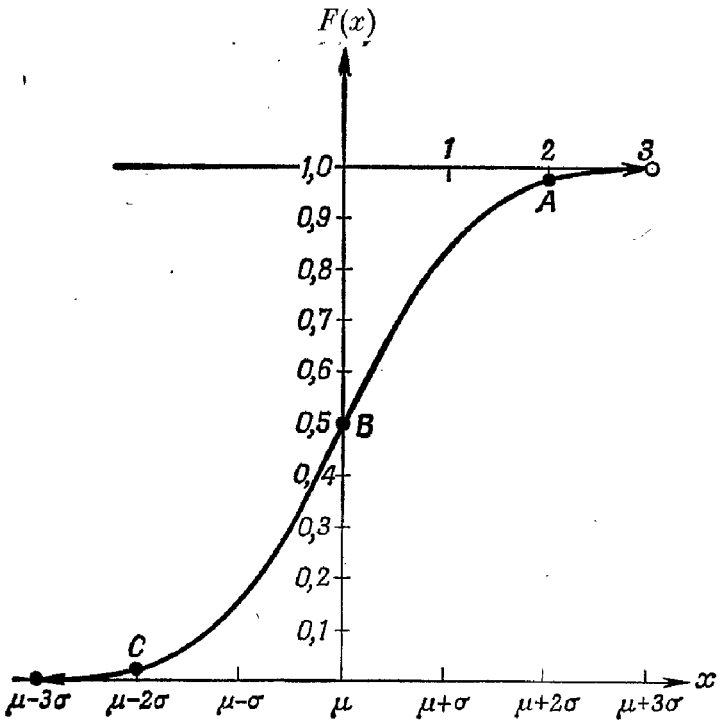
$$F(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \int_{-\infty}^{x} e^{-(t-\mu)^2/2\sigma^2}\, dt \tag{16}$$

(фиг. 5). Изменением шкалы по вертикальной оси эту кривую можно превратить в прямую линию. Графическая бумага такого типа с соответствующей шкалой ординат называется *вероятностной бумагой*.

[1]) См. также [27], гл. 6; [9], гл. 5.


f(x)
1,0
0,8
0,6
0,4
0,2
0
μ=6
σ=0,5
μ=4; σ=1
μ=1; σ=2
x
−2 −1 0 1 2 3 4 5 6 7 8

Фиг. 4.


F(x)
1,0
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
1
2
3
A
B
C
x
μ−3σ
μ−2σ
μ−σ
μ
μ+σ
μ+2σ
μ+3σ

Фиг. 5.

§ 5. Математическое ожидание и дисперсия случайной величины

Дискретное распределение вероятности случайной величины x полностью задается множеством значений вероятности или функцией распределения $\{\mathcal{P}(x = r)\}$. Непрерывное распределение полностью задается его плотностью $f(x)$. Во многих случаях бывает необходимо выделить наиболее важные свойства распределения. Для этого используются такие характеристики, как среднее значение, дисперсия, асимметрия и др.

Определение 1. *Математическое ожидание* случайной величины x есть среднее значение x с учетом вероятности (или плотности вероятности) осуществления каждого значения x. Для дискретного распределения

$$\mathcal{E}(x) = \sum_r r\mathcal{P}(x = r), \tag{17}$$

для непрерывного распределения

$$\mathcal{E}(x) = \int_{-\infty}^{+\infty} xf(x)\,dx. \tag{18}$$

Принято обозначать $\mathcal{E}(x)$ через μ. Мы будем пользоваться величиной μ как характеристикой положения распределения x.

Определение 2. Математическое ожидание функции $h(x)$ случайной величины x есть средняя величина $h(x)$ по всем возможным значениям переменной x, т. е.

$$\mathcal{E}[h(x)] = \sum_r h(r)\mathcal{P}(x = r), \tag{19}$$

или

$$\mathcal{E}[h(x)] = \int_{-\infty}^{+\infty} h(x)f(x)\,dx. \tag{20}$$

Дисперсия. Пусть $h(x) = [x - \mathcal{E}(x)]^2 = (x - \mu)^2$. Тогда *дисперсия* x равна $\mathcal{E}[(x - \mu)^2]$; ее принято обозначать через $\mathcal{D}(x)$ или $\sigma^2(x)$. Корень квадратный из этой величины σ называют среднеквадратичным или стандартным отклонением x и используют как меру разброса x относительно среднего значения μ.

Асимметрия. *Асимметрия* характеризуется параметром

$$\gamma_1 = \frac{\mathscr{E}[(x-\mu)^3]}{\sigma^3}.$$

Асимметрия отрицательна, если $f(x)$ сильно вытянуто влево от μ, и положительна, если $f(x)$ вытянуто вправо от μ. Если распределение симметрично, то параметр γ_1 равен нулю.

Закон сложения. Пусть $h(x) = h_1(x) + h_2(x)$. Тогда

$$\mathscr{E}[h(x)] = \sum [h_1(r) + h_2(r)]\, \mathscr{P}(x = r),$$

или

$$\mathscr{E}[h(x)] = \int [h_1(x) + h_2(x)]\, f(x)\, dx.$$

Следовательно, в обоих случаях

$$\mathscr{E}[h_1(x) + h_2(x)] = \mathscr{E}[h_1(x)] + \mathscr{E}[h_2(x)]. \tag{21}$$

Аналогично, полагая $h_1(x) = kh(x)$, где k — произвольная постоянная, найдем

$$\mathscr{E}(kh) = k\mathscr{E}(h). \tag{22}$$

Пример:

$$\sigma^2 = \mathscr{E}[(x-\mu)^2] = \mathscr{E}(x^2 - 2\mu x + \mu^2) = \mathscr{E}(x^2) - 2\mu\mathscr{E}(x) + \\ + \mu^2 = \mathscr{E}(x^2) - 2\mu^2 + \mu^2 = \mathscr{E}(x^2) - \mu^2.$$

Аналогичным образом для любой постоянной a

$$\mathscr{E}[(x-a)^2] = \mathscr{E}(x^2) - 2a\mu + a^2.$$

Кроме того,

$$\mathscr{E}[(kx - ka)^2] = k^2\mathscr{E}[(x-a)^2].$$

В частности, для дисперсии $\mathscr{D}$:

$$\mathscr{D}(kx) = \mathscr{E}[(kx - k\mu)^2] = k^2\mathscr{D}(x). \tag{23}$$

Биномиальное распределение

Рассмотрим биномиальное распределение с параметрами N и p:

$$\mathscr{P}(x = r) = C_N^r p^r (1-p)^{N-r}.$$

В этом случае, полагая $q=1-p$, имеем

$$\mu=\sum_{r=0}^{N}\frac{rN!}{(N-r)!\,r!}\,p^r q^{N-r}=Np\sum_{r=1}^{N}\frac{(N-1)!}{(N-r)!\,(r-1)!}\,p^{r-1}q^{N-r}=$$

$$=Np\sum_{s=0}^{N-1}\frac{(N-1)!}{(N-s-1)!\,s!}\,p^s q^{N-s-1},$$

где $s=r-1$.

Пусть $n=N-1$, тогда $\mu=Np\sum\limits_{0}^{n}C_n^s p^s q^{n-s}=Np$. Аналогично получим

$$\mathscr{E}(x^2)=\sum_{0}^{N}r^2P_r=\sum_{2}^{N}r(r-1)P_r+\sum_{1}^{N}rP_r=$$

$$=N(N-1)p^2+Np;$$

следовательно,

$$\sigma^2=\mathscr{E}(x^2)-\mu^2=Np(1-p)=Npq.$$

Заметим, что σ^2 растет с N.

Можно показать, что $\mathscr{E}[(x-\mu)^3]=Npq(q-p)$. Таким образом, асимметрия

$$\gamma_1=\frac{Npq(q-p)}{(Npq)^{3/2}}=\frac{q-p}{\sqrt{Npq}}=\frac{1-2p}{\sqrt{Np(1-p)}},$$

причем

$$\gamma_1<0,\quad \text{если}\quad p>\frac{1}{2},$$

$$\gamma_1=0,\quad \text{если}\quad p=\frac{1}{2},$$

$$\gamma_1>0,\quad \text{если}\quad p<\frac{1}{2}.$$

Если p фиксировано, то $\gamma_1\to 0$ при $N\to\infty$ для любого p.

Пример. Допустим, что мы ввели новую случайную величину $h(x)=x/N$. Тогда

$$\mathscr{E}\left(\frac{x}{N}\right)=\frac{Np}{N}=p, \tag{24}$$

$$\mathscr{D}\left(\frac{x}{N}\right)=\frac{1}{N^2}\mathscr{D}(x)=\frac{Npq}{N^2}=\frac{pq}{N}, \tag{25}$$

а среднеквадратичное отклонение σ:

$$\sigma\left(\frac{x}{N}\right)=\sqrt{\frac{pq}{N}}. \tag{26}$$

При решении статистической задачи об оценке неизвестного параметра p можно использовать *статистику* [1]) x/N.

Пусть

$$\hat{p}=\frac{x}{N}, \tag{27}$$

причем наличие у параметра «крышки» означает *оценку* p. Поскольку x — случайная величина, то $\hat{p}$ — тоже случайная величина. Мы говорим, что $\hat{p}$ — *несмещенная оценка* параметра p, так как, согласно (24), $\mathscr{E}(\hat{p})=p$. Эта оценка *состоятельна*, поскольку, согласно (25), $\mathscr{D}(\hat{p})\to 0$ при возрастании N. Увеличение точности оценки $\hat{p}$ характеризуется уменьшением среднеквадратичного отклонения и пропорционально $\sqrt{N}$, как следует из формулы (26).

Распределение Пуассона

Пусть $\mathscr{P}(x=r)=(\mu^r/r!)\,e^{-\mu}$. Как видно из обозначения, здесь предполагается, что $\mathscr{E}(x)=\mu$. Докажем это:

$$\mathscr{E}(x)=\sum_{r=0}^{\infty} r\,\frac{\mu^r}{r!}\,e^{-\mu}=\mu\sum_{r=1}^{\infty}\mu^{r-1}\,\frac{e^{-\mu}}{(r-1)!}=\mu\sum_{s=0}^{\infty}\mu^s\,\frac{e^{-\mu}}{s!},$$

где $s=r-1$.

Таким образом,

$$\mathscr{E}(x)=\mu.$$

Аналогично

$$\mathscr{E}(x^2)=\sum_{0}^{\infty}[r(r-1)+r]\,\mu^r\,\frac{e^{-\mu}}{r!}=$$

$$=\mu^2\sum_{2}^{\infty}\mu^{r-2}e^{-\mu}\,\frac{1}{(r-2)!}+\mu=\mu^2+\mu.$$

[1]) Под статистикой понимается случайная величина, являющаяся функцией результатов наблюдений.— *Прим. ред.*

Отсюда следует, что

$$\mathscr{D}(x) = \mathscr{E}(x^2) - [\mathscr{E}(x)]^2 = \mu^2 + \mu - \mu^2 = \mu$$

и

$$\sigma(x) = \sqrt{\mu},$$

т. е. разброс x относительно среднего значения пропорционален корню квадратному из среднего значения. Наконец, $\mathscr{E}[(x-\mu)^3] = \mu$, и поэтому асимметрия

$$\gamma_1 = \frac{\mu}{\mu^{3/2}} = \mu^{-1/2}.$$

Асимметрия всегда положительна и стремится к нулю по мере возрастания μ. Другими словами, с ростом μ распределение становится все более симметричным.

Равномерное распределение

В этом случае

$$\mu = \int_a^b \frac{x}{b-a}\,dx = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2},$$

как и следовало ожидать из соображений симметрии;

$$\mathscr{E}(x^2) = \int_a^b \frac{x^2}{b-a}\,dx = \frac{b^2+ab+a^2}{3},$$

$$\mathscr{D}(x) = \mathscr{E}(x^2) - [\mathscr{E}(x)]^2 = \frac{(b-a)^2}{12}.$$

В частном случае $b-a=1$ имеем

$$\mathscr{D}(x) = \frac{1}{12} \quad \text{и} \quad \sigma(x) = \frac{\sqrt{3}}{6} = 0{,}29.$$

Допустим, что μ неизвестно, но известно, что $b-a=1$. Пусть наблюдалось единственное значение x. Что при этом можно сказать о μ? Мы знаем, что

$$\mathscr{P}\left[\mu - \frac{\alpha}{2} \leqslant x \leqslant \mu + \frac{\alpha}{2}\right] = \alpha, \tag{28}$$

если $\alpha \leqslant 1$.

Если мы сделаем утверждение, содержащееся в формуле (28), в которую подставлено наблюдавшееся значение $x = x_0$, то мы будем правы в 100α случаях из 100. Выражение (28) можно записать в другом виде:

$$\mathscr{P}\left[-x-\frac{\alpha}{2}\leqslant-\mu\leqslant-x+\frac{\alpha}{2}\right]=\alpha,$$

и, следовательно,

$$\mathscr{P}\left[x+\frac{\alpha}{2}\geqslant\mu\geqslant x-\frac{\alpha}{2}\right]=\alpha. \tag{29}$$

Полученная формула дает нам *доверительные пределы* с *доверительной вероятностью* α, или $100\alpha\%$. Поскольку μ — постоянная, а не случайная величина, то было бы недопустимо трактовать выражение (29) как «вероятность того, что μ больше $x - \alpha/2$ и меньше $x + \alpha/2$». Поэтому формулу (29) лучше переписать в виде

$$\mathscr{P}\left[\left(x-\frac{\alpha}{2},\ x+\frac{\alpha}{2}\right)\ni\mu\right]=\alpha \tag{30}$$

и сказать, что с вероятностью α истинное значение μ лежит в интервале от $x-\alpha/2$ до $x+\alpha/2$. Так,

$$\mathscr{P}\left[(x-0{,}475,\ x+0{,}475)\ni\mu\right]=0{,}95.$$

В действительности такое утверждение не имеет особого смысла, так как мы знаем, что μ должно быть заключено между $x - 0{,}5$ и $x + 0{,}5$. Причина столь скудной информации относительно величины μ заключается в том, что ее оценка получена на основании единичной выборки.

Нормальное распределение

В этом случае

$$\mathscr{E}(x)=\int_{-\infty}^{\infty}\frac{x}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}\,dx.$$

Положим

$$w=\frac{x-\mu}{\sigma},\qquad dw=\frac{dx}{\sigma},$$

тогда, учитывая, что интеграл от нечетной функции равен 0, получаем

$$\mathscr{E}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma w + \mu)\, e^{-w^2/2}\, dw =$$

$$= 0 + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\, dw =$$

$$= \mu \times \text{Интеграл от плотности вероятности.}$$

Итак, как и следовало ожидать из обозначений,

$$\mathscr{E}(x) = \mu.$$

Дисперсия

$$\mathscr{D}(x) = \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\, dx = \int_{-\infty}^{\infty} \frac{\sigma^2 w^2}{\sqrt{2\pi}} e^{-w^2/2}\, dw.$$

Интегрируя по частям, получаем

$$\mathscr{D}(x) = \frac{1}{2\pi}\left[-\sigma^2 w e^{-w^2/2}\right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 e^{-w^2/2}\, dw = \sigma^2,$$

что также следует из принятых обозначений. И наконец, $\mathscr{E}[(x-\mu)^3] = 0$, поскольку распределение плотности вероятности в данном случае симметрично относительно μ, и асимметрия равна нулю.

. Пример. Допустим, что величина σ известна, и мы хотим на основании единственного значения x оценить величину μ. Как и прежде, начнем со следующего утверждения:

$$\mathscr{P}[\mu - k_\alpha\sigma \leqslant x \leqslant \mu + k_\alpha\sigma] = \alpha$$

и перепишем затем его в виде

$$\mathscr{P}[x - k_\alpha\sigma \leqslant \mu \leqslant x + k_\alpha\sigma] = \alpha.$$

Значения k_α можно найти из соответствующих таблиц в справочниках по математической статистике [32].

В приведенной ниже табл. 1 дано несколько значений величины k_α.

Таблица 1

α	k_α
0,000	0
0,500	0,674
0,683	1,000
0,900	1,645
0,950	1,960 *
0,990	2,576
0,9975	3,000
0,999	3,291

* Этот коэффициент, соответствующий доверительной вероятности 0,95, принято считать равным 2,00.

У физиков общепринято приводить экспериментальные данные в виде величины $\hat{\mu}$ плюс или минус одно среднеквадратичное отклонение, т. е.

$$\mu = \hat{\mu} \pm \sigma. \tag{31}$$

Если $\hat{\mu}$ имеет нормальное распределение, то из табл. 1 видно, что $\alpha = 0{,}683$ при $k_\alpha = 1$. Это значит, что утверждение, содержащееся в (31), может оказаться несправедливым примерно в одной трети всех случаев его применения. С другой стороны, утверждение

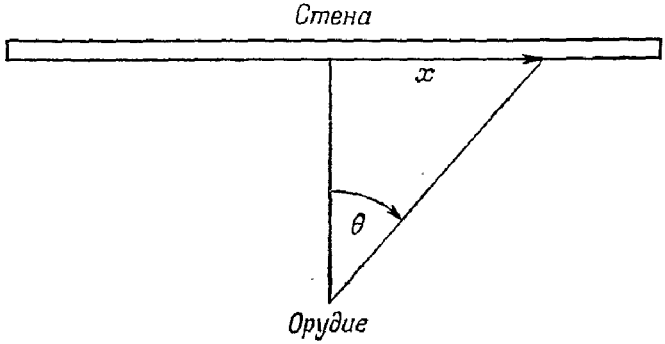
$$\mu = \hat{\mu} \pm 2\sigma \tag{32}$$

будет несправедливым уже только в одной двадцатой всех случаев. Запись (31) удобна, поскольку она содержит обе величины: $\hat{\mu}$ и σ, но необходимо иметь в виду, что ей соответствует невысокая доверительная вероятность (0,683). Если величина σ сама оценена из результатов эксперимента, то доверительная вероятность становится еще ниже.

Распределение Коши

Это экзотическое распределение часто упоминают с иллюстративной целью вследствие его необычного свойства: дисперсия распределения Коши оказывается бесконечной.

Предположим, что артиллерийское орудие расположено на единичном расстоянии от сколь угодно длинной стены. Лафет орудия вращается с постоянной скоростью.



Фиг. 6.

На каждом обороте в случайный момент времени из орудия производится выстрел. Как при этом распределены попадания по стене?

Обозначим через θ (фиг. 6) угол, на который повернулся ствол орудия относительно начального положения (по нормали к стене), а через x — расстояние вдоль стены от ее середины до точки попадания. Снаряд попадет в стену только при условии, что θ заключено в пределах от $-\pi/2$ до $+\pi/2$. В силу того что все значения углов равновероятны, мы используем для величины θ равномерное распределение

$$f(\theta)\,d\theta = \frac{d\theta}{\pi}, \quad -\frac{\pi}{2} \leqslant \theta \leqslant \frac{\pi}{2}.$$

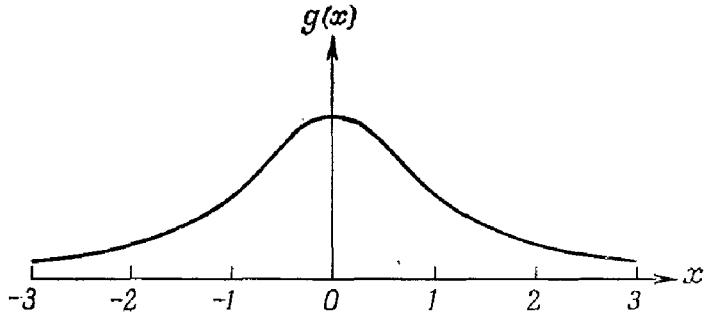
Поскольку $\theta = \operatorname{arctg} x$, то

$$d\theta = \frac{dx}{1+x^2}$$

Таким образом, плотность распределения величины x равна

$$g(x)\,dx = \frac{1}{\pi}\,d\theta = \frac{1}{\pi}\,\frac{dx}{1+x^2}, \qquad -\infty \leqslant x \leqslant \infty. \tag{33}$$

Эта функция изображена на фиг. 7; она симметрична относительно нуля, и ее среднее значение поэтому равно



Фиг. 7.

нулю, что можно доказать с помощью обычного выражения

$$\mu = \lim_{R\to\infty} \int_{-R}^{R} x g(x)\,dx.$$

Из этого выражения следует

$$\mu = \lim_{R\to\infty} \frac{1}{2\pi}\,[\ln(1+x^2)]_{-R}^{R} = \frac{1}{2\pi} \lim_{R\to\infty} \left[\ln \frac{1+R^2}{1+(-R)^2}\right] = 0.$$

Дисперсия такого распределения

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 g(x)\,dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x^2}{1+x^2}\,dx =$$

$$= \frac{1}{\pi} \int_{-\infty}^{\infty} \left[1 - \frac{1}{1+x^2}\right] dx = \infty - 1.$$

Следовательно, $\sigma^2 = \infty$.

Пусть некая общая теорема доказана нами в предположении, что рассматриваемая случайная величина обладает

конечной дисперсией. Если эта теорема окажется несправедливой для распределения Коши, то это будет означать, что сделанное предположение *необходимо*.

§ 6. Моменты случайной величины. Производящие функции моментов

Нам уже известно, как находить математическое ожидание функции случайной величины x. Математические ожидания степенных функций от x или от разностей между x и ее средним называются *моментами*.

Определение 1. *Начальный момент* n-го порядка случайной величины x равен

$$\mu'_n = \mathscr{E}(x^n) = \begin{cases} \int\limits_{-\infty}^{\infty} x^n f(x)\,dx, \\ \sum\limits_{r=0}^{\infty} r^n P_r. \end{cases} \tag{34}$$

Заметим, что $\mu'_0 = 1$, $\mu'_1 = \mu$.

Определение 2. *Центральный момент* n-го порядка случайной величины x равен

$$\mu_n = \mathscr{E}[(x-\mu)^n]. \tag{35}$$

Заметим, что

$$\mu_0 = 1, \quad \mu_1 = 0, \quad \mu_2 = \sigma^2.$$

Помимо важности отдельных моментов, существенное теоретическое значение имеет полный набор моментов $\{\mu'_n\}$ или $\{\mu_n\}$. При весьма общих условиях набор моментов полностью определяет распределение вероятности. Есть немало случаев, когда проще найти полный набор моментов неизвестного распределения, нежели непосредственно само распределение. В таких случаях распределение удается построить исходя из данного набора моментов.

Теорема. Если две плотности вероятности $f_1(x)$ $f_2(x)$ имеют одинаковые моменты (μ'_n) и если можно

$f_1(x) - f_2(x)$ разложить в ряд по степеням x, то $f_1(x) \equiv f_2(x)$.

Доказательство. Пусть $f_1(x) - f_2(x) = c_0 + c_1 x + c_2 x^2 + \ldots$. Тогда

$$\int_{-\infty}^{\infty} [f_1(x) - f_2(x)]^2 \, dx =$$

$$= \int_{-\infty}^{\infty} [f_1(x) - f_2(x)] [c_0 + c_1 x + c_2 x^2 + \ldots] \, dx =$$

$$= c_0 \int (f_1 - f_2) \, dx + c_1 \int (x f_1 - x f_2) \, dx +$$

$$+ c_2 \int (x^2 f_1 - x^2 f_2) \, dx + \ldots = c_0 (1 - 1) +$$

$$+ c_1 (\mu_1' - \mu_1') + c_2 (\mu_2' - \mu_2') + \ldots = 0.$$

Следовательно,

$$f_1(x) \equiv f_2(x).$$

Производящие функции моментов (ПФМ)

Набор моментов $\{\mu_n'\}$ важен, поскольку он характеризует распределение случайной величины. Целесообразно ввести функцию, зависящую от всех моментов μ_n', которая свела бы весь набор $\{\mu_n'\}$ к единственному выражению.

Введем вспомогательную переменную t и сопоставим моменту μ_n' соответствующую степень t^n так, чтобы по этой степени можно было всегда найти μ_n'.

Определение 1. *Производящая функция начальных моментов* случайной величины x:

$$M_x'(t) = \mathscr{E}(e^{xt}) = \begin{cases} \int_{-\infty}^{\infty} e^{xt} f(x) \, dx, \\ \sum_{r=0}^{\infty} e^{rt} P_r. \end{cases} \qquad (36)$$

Экспоненту можно разложить в ряд

$$M'_x(t) = \mathcal{E}\left[1 + xt + \frac{(xt)^2}{2!} + \frac{(xt)^3}{3!} + \ldots\right] =$$

$$= 1 + \mu'_1 t + \mu'_2 \frac{t^2}{2!} + \mu'_3 \frac{t^3}{3!} + \ldots,$$

так что μ'_n будет просто коэффициентом при $t^n/n!$. Иными словами,

$$\mu'_n \equiv n! \times \text{Коэффициент при } t^n. \tag{37}$$

Другой способ нахождения μ'_n из ПФМ $M'_x(t)$ состоит в том, что эту функцию дифференцируют n раз.

В результате имеем

$$\frac{\partial^n M'_x(t)}{\partial t^n} = \int_{-\infty}^{\infty} x^n e^{xt} f(x)\, dx.$$

Отсюда следует

$$\frac{\partial^n M'_x(0)}{\partial t^n} = \int_{-\infty}^{\infty} x^n f(x)\, dx = \mu'_n. \tag{38}$$

Определение 2. *Производящая функция центральных моментов* случайной величины x:

$$M_x(t) = \mathcal{E}\left[e^{(x-\mu)t}\right], \quad \mu = \mu'_1. \tag{39}$$

Разлагая в ряд, получаем

$$M_x(t) = \mathcal{E}\left[1 + (x-\mu)t + (x-\mu)^2 \frac{t^2}{2!} + \ldots\right] =$$

$$= 1 + 0 + \mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \ldots,$$

так что μ_n — коэффициент при $t^n/n!$. Заметим, что $M(t) = e^{-\mu t} M'(t)$ и $M'(t) = e^{\mu t} M(t)$. Наконец, если две ПФМ $M'_1(t)$ и $M'_2(t)$ одинаковы, то одинаковы и исходные распределения вероятности.

Мы будем использовать ПФМ двояким образом. По заданному распределению вероятности очень легко найти ПФМ. Тогда моменты могут быть найдены по формулам (37) или (38). Этот способ может оказаться проще вычисления моментов из самого распределения.

Кроме того, может возникнуть необходимость найти функцию распределения нескольких случайных величин. Возможно, что сначала легче отыскать ПФМ неизвестного распределения, а затем уже попытаться найти само распределение.

Эти методы имеют много общего с использованием преобразования Лапласа (t — комплексное число) [1]):

$$L(t) = \int_0^\infty e^{-xt} f(x)\, dx.$$

---

[1]) В различных приложениях теории вероятностей, например в теории случайных процессов [49, 59], широко используются и другие вспомогательные величины.

Аналогично ПФМ можно ввести *производящие функции вероятности* (ПФВ). Если t — вспомогательная переменная, то ПФВ $G_x(t)$ имеет вид

$$G_x(t) = \mathscr{E}(t^x) = \begin{cases} \int_{-\infty}^{\infty} t^x f(x)\, dx, \\ \sum_{r=0}^{\infty} t^r P_r, \end{cases}$$

причем $G_x(1) = 1$. Нетрудно связать ПФВ с моментами. Обозначая $\partial G/\partial t$ через G', получаем

$$\mu_1' = G'(1); \quad \mu_2' = G''(1) + G'(1); \quad \mu_3' = G'''(1) + 3G''(1) + G'(1) \text{ и т. д.}$$

Часто используются также величины, представляющие собой коэффициенты разложения в ряд Тэйлора логарифма ПФМ, а именно

$$\log M_x'(t) = \varkappa_1 t + \varkappa_2 \frac{t^2}{2!} + \varkappa_3 \frac{t^3}{3!} + \dots,$$

т. е.

$$\varkappa_n = \left[\frac{\partial^n}{\partial t^n} \log M_x'(t)\right]_{t=0}.$$

Величины $\varkappa_n$ называются *семиинвариантами* или *кумулянтами*. Они просто связаны с центральными моментами, в частности $\varkappa_1 = \mu_1$, $\varkappa_2 = \mu_2$, $\varkappa_3 = \mu_3$. — *Прим. ред.*

Биномиальное распределение

$$M'(t)=\sum_0^\infty e^{rt}P_r=\sum_0^\infty C_N^r(pe^t)^r(1-p)^{N-r}=(pe^t+1-p)^N,$$

$$\frac{\partial M'(t)}{\partial t}=Npe^t(pe^t+1-p)^{N-1}.$$

Следовательно,

$$\mu_1'=\frac{\partial M'(0)}{\partial t}=Np.$$

Распределение Пуассона

$$M'(t)=e^{-\mu}\sum_0^\infty\frac{(\mu e^t)^r}{r!}=e^{-\mu}e^{\mu e^t}=e^{\mu(e^t-1)},$$

$$\frac{\partial M'(t)}{\partial t}=\mu e^t e^{\mu(e^t-1)},$$

т. е.

$$\mu_1'=\mu.$$

Производящая функция центральных моментов

$$M(t)=e^{-\mu t}M'(t)=e^{\mu(e^t-t-1)}=e^{\mu\left(1+t+\frac{t^2}{2!}+\frac{t^3}{3!}+\ldots-t-1\right)}=$$

$$=e^{\mu\left(\frac{t^2}{2!}+\frac{t^3}{3!}+\ldots\right)}=1+\mu\left(\frac{t^2}{2!}+\frac{t^3}{3!}+\ldots\right)+$$

$$+\mu^2\frac{1}{2}\left(\frac{t^2}{2!}+\frac{t^3}{3!}+\ldots\right)^2+\ldots.$$

Коэффициент при $t^2/2!$ равен μ. Значит, $\mu_2=\sigma^2=\mu$.

Нормальное распределение

$$M'(t)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{xt}e^{-(x-\mu)^2/2\sigma^2}\,dx=$$

$$=e^{\mu t+\sigma^2\frac{t^2}{2}}=e^{\mu t}\cdot e^{\sigma^2\frac{t^2}{2}}=$$

$$=\left[1+\mu t+\mu^2\frac{t^2}{2!}+\ldots\right]\left[1+\sigma^2\frac{t^2}{2!}+\ldots\right].$$

Коэффициент при t

$$\mu_1' = \mu.$$

Коэффициент при $t^2/2!$

$$\mu_2' = \mu^2 + \sigma^2.$$

Следовательно,

$$\mu_2 = \mu_2' - (\mu_1')^2 = \sigma^2.$$

Согласно формуле (39),

$$M(t) = e^{\sigma^2 \frac{t^2}{2!}} = \sum_{k=0}^{\infty} \left(\sigma^2 \frac{t^2}{2!}\right)^k \cdot \frac{1}{k!} = \sum_{n=0}^{\infty} \mu_n \frac{t^n}{n!}.$$

Таким образом,

$$\mu_{2n} = \frac{(2n)!}{2^n \cdot n!} \sigma^{2n}$$

и

$$\mu_{2n+1} = 0.$$

Распределение $y = ax + b$, где a и b — постоянные величины

$$M_y'(t) = \mathscr{E}(e^{yt}) = \mathscr{E}(e^{bt} \cdot e^{axt}) = e^{bt} \mathscr{E}(e^{xat}).$$

Значит,

$$M_y'(t) = e^{bt} M_x'(at). \tag{40}$$

Допустим, что x имеет нормальное распределение со средним значением μ и дисперсией σ^2. Тогда

$$M_x'(t) = e^{\mu t + \sigma^2 \frac{t^2}{2}}$$

и

$$M_y'(t) = e^{\mu a t + \sigma^2 a^2 \frac{t^2}{2} + bt} = e^{(a\mu + b)t + a^2\sigma^2 \frac{t^2}{2}}.$$

Следовательно, y распределено по нормальному закону со средним $(a\mu + b)$ и дисперсией $a^2\sigma^2$.

Гамма-распределение

Гамма-распределение имеет следующую плотность:

$$f(x) = \frac{1}{\alpha!\, \beta^{\alpha+1}} x^{\alpha} e^{-x/\beta}, \quad 0 \leqslant x \leqslant \infty, \tag{41}$$

т. е.

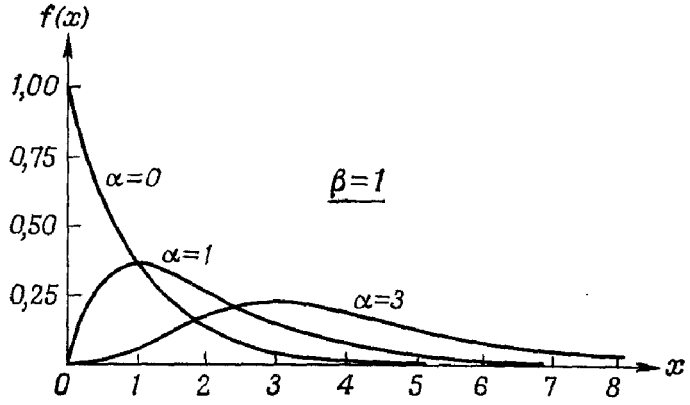
$$f(x) = f(x, \alpha, \beta).$$

Параметры α и β имеют допустимые области: $\alpha > -1$; $\beta > 0$. Форма $f(x)$ определяется параметром α (фиг. 8),

а параметр β носит характер масштабного множителя. Обратим внимание на то, что при $\alpha = 0$

$$f(x) = \frac{1}{\beta} e^{-x/\beta},$$

т. е. превращается в экспоненциальное распределение. В дальнейшем мы увидим, что распределение χ^2 есть



Фиг. 8.

частный случай гамма-распределения. ПФМ в случае гамма-распределения

$$M'(t) = \frac{1}{\alpha!\,\beta^{\alpha+1}} \int_{-\infty}^{\infty} x^{\alpha} e^{-x/\beta} e^{xt}\, dx.$$

Преобразуем это выражение к константе, умноженной на интеграл от плотности. Тогда $M'(t)$ будет просто равна этой константе.

Подставим $-x/\beta + xt = -x\,[\beta/(1-\beta t)]^{-1}$ и, кроме того, $\beta^{\alpha+1} = [\beta/(1-\beta t)]^{\alpha+1}(1-\beta t)^{\alpha+1}$. В таком случае

$$M'(t) = \frac{1}{(1-\beta t)^{\alpha+1}} \int_{-\infty}^{\infty} \frac{1}{\alpha!} \left(\frac{1-\beta t}{\beta}\right)^{\alpha+1} x^{\alpha} e^{-x(1-\beta t)/\beta}\, dx =$$

$$= \frac{1}{(1-\beta t)^{\alpha+1}} \int_{-\infty}^{\infty} f\left(x, \alpha, \frac{\beta}{1-\beta t}\right) dx.$$

Следовательно,

$$M'(t) = (1-\beta t)^{-\alpha-1}. \qquad (42)$$

ГЛАВА ВТОРАЯ

Совместные распределения вероятностей

§ 7. Функции распределения вероятностей двух или нескольких случайных величин

Пусть X_1 и X_2 — дискретные случайные величины. Введем

$$\mathcal{P}[X_1 = r_1 \text{ и } X_2 = r_2] = P_{r_1, r_2}, \tag{43}$$

причем

$$\sum_{r_1} \sum_{r_2} P_{r_1, r_2} = 1.$$

Для непрерывных случайных величин введем также

$$f(x_1, x_2)\, dx_1\, dx_2 = \mathcal{P}[x_1 \leqslant X_1 \leqslant x_1 + dx_1 \text{ и } x_2 \leqslant X_2 \leqslant x_2 + dx_2]$$

при условии

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1\, dx_2 = 1.$$

Пример 1. Пусть X_1 означает вертикальный номер квадрата на шахматной доске, а X_2 — горизонтальный номер. Величины X_1 и X_2 могут принимать значения: 1, 2, . . . , 8. Выберем произвольный квадрат. Какова при этом вероятность того, что $X_1 = r_1$ и $X_2 = r_2$? Поскольку шахматная доска имеет 64 равноценных квадрата, то

$$\mathcal{P}(X_1 = r_1 \text{ и } X_2 = r_2) = \frac{1}{64}.$$

Пример 2. Предположим, что шахматная доска имеет длину b (единиц) и ширину c (единиц). Выберем на доске произвольную точку. Какова плотность вероятности получить при этом координаты (x_1, x_2)?

Очевидно, что

$$f(x_1, x_2)\, dx_1\, dx_2 = \frac{dx_1\, dx_2}{bc} \tag{44}$$

при условии $0 \leqslant x_1 \leqslant b$; $0 \leqslant x_2 \leqslant c$. Это — равномерное распределение для случая двух измерений.

Пример 3. Если эксперимент повторяется n раз, то его результаты удобно рассматривать как n случайных величин. Если n результатов представляют собой *независимые* выборки с одинаковым распределением, то в частном случае $n = 2$

$$f(x_1, x_2)\, dx_1\, dx_2 = g(x_1)\, dx_1 g(x_2)\, dx_2, \tag{45}$$

где $g(x)$ — плотность вероятности отдельного результата. Этот результат следует из закона умножения вероятностей (7).

В более общем случае плотность вероятности для n независимых случайных величин с одинаковыми распределениями равна

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} g(x_i). \tag{46}$$

Сумма случайных величин

Часто нам приходится складывать случайные величины (например, для получения среднего). Чтобы найти распределение суммы случайных величин, можно воспользоваться производящей функцией моментов суммы. Пусть

$$M'_{\Sigma x_i}(t) = \mathscr{E}\,[e^{t\Sigma x_i}].$$

Итак, нам необходимо усреднить по всем $\{x_i\}$, учитывая при этом относительные частоты появления каждой величины x_i. Поэтому

$$M'_{\Sigma x_i}(t) = \int\int \ldots \int e^{t\Sigma x_i} \prod_{i=1}^{n} g(x_i)\, dx_i =$$

$$= \int\int \ldots \int \prod_{i=1}^{n} e^{tx_i} g(x_i)\, dx_i = \prod_{i=1}^{n} \int e^{tx_i} g(x_i)\, dx_i =$$

$$= \prod_{i=1}^{n} M'_{x_i}(t).$$

Таким образом,

$$\text{ПФМ}\left(\sum x_i\right) = \prod_{i=1}^{n} \text{ПФМ}\,(x_i). \qquad (47)$$

Другими словами, производящая функция моментов суммы независимых случайных величин есть просто произведение ПФМ отдельных слагаемых.

Если все величины x_i имеют одинаковые распределения, то их ПФМ также будут одинаковы; значит, в этом случае

$$\text{ПФМ}\left(\sum x_i\right) = [\text{ПФМ}\,(x_i)]^n. \qquad (48)$$

Пусть $\{m'_k\}$ — начальные моменты суммы $\sum x_i$, а $\{\mu'_k\}$ — начальные моменты слагаемых x_i. Тогда из формул (36) и (48) следует

$$\text{ПФМ}\left(\sum x_i\right) = 1 + m'_1 t + m'_2 \frac{t^2}{2!} + \ldots \equiv$$

$$\equiv \left[1 + \mu'_1 t + \mu'_2 \frac{t^2}{2!} + \ldots\right]^n = 1 + n\left(\mu'_1 t + \frac{\mu'_2 t^2}{2!}\right) +$$

$$+ \frac{n(n-1)}{2}(\mu'_1 t)^2 + \ldots$$

с точностью до t^3. Приравнивая коэффициенты при t, получаем $m'_1 = n\mu'_1$, или

$$\mathscr{E}\left(\sum x_i\right) = n\mathscr{E}\,(x_i), \qquad (49)$$

т. е. математическое ожидание суммы равно сумме математических ожиданий слагаемых.

Далее, приравнивая коэффициенты при $t^2/2!$, получаем

$$m'_2 = n\mu'_2 + (n^2 - n)(\mu'_1)^2.$$

Дисперсия суммы $\sum x_i$ равна

$$\mathscr{D}\left(\sum x_i\right) = m'_2 - (m'_1)^2 = n\mu'_2 + (n^2 - n)(\mu'_1)^2 - n^2(\mu'_1)^2 =$$

$$= n[\mu'_2 - (\mu'_1)^2],$$

откуда

$$\mathscr{D}\left(\sum x_i\right) = n\mathscr{D}\,(x_i), \qquad (50)$$

т. е. дисперсия суммы равна сумме дисперсий слагаемых.

Это свойство аддитивности дисперсий широко используется в дисперсионном анализе, а также в задаче о переносе ошибок.

Выборочное среднее

Если рассмотреть *выборочное среднее*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

то легко заметить, что

$$\mathscr{E}(\bar{x}) = \frac{1}{n} \mathscr{E}\left(\sum x_i\right) = \mathscr{E}(x_i),$$

т. е.

$$\mathscr{E}(\bar{x}) = \mu. \tag{51}$$

Таким образом, выборочное среднее есть несмещенная оценка среднего значения μ генеральной совокупности.

Аналогичным образом решается и вопрос о дисперсии:

$$\mathscr{D}(\bar{x}) = \mathscr{D}\left(\frac{1}{n} \sum x_i\right) = \frac{1}{n^2} \mathscr{D}\left(\sum x_i\right).$$

Следовательно,

$$\mathscr{D}(\bar{x}) = \frac{1}{n} \mathscr{D}(x_i). \tag{52}$$

Мы видим, что выборочное среднее $\bar{x}$ является значительно более точной оценкой μ, чем отдельное значение x_i, так как $\bar{x}$ обладает меньшим разбросом относительно истинного среднего. Точность оценки определяется выражением

$$\sigma(\bar{x}) = \frac{\sigma(x_i)}{\sqrt{n}},$$

которое показывает, что точность возрастает пропорционально $\sqrt{n}$; с этим результатом мы уже встречались на примере биномиального распределения (см. стр. 34—35).

§ 8. Закон больших чисел

Мы убедились в том, что выборочное среднее является хорошей оценкой среднего μ генеральной совокупности. Теперь докажем более сильное утверждение: с ростом

объема выборки выборочное среднее с очень большой вероятностью стремится к μ как к своему пределу. Сначала докажем для произвольной случайной величины x *неравенство Чебышева*:

$$\mathcal{P}[|x-\mu| \geqslant k\sigma] \leqslant \frac{1}{k^2}, \tag{53}$$

где $\mathcal{E}(x)=\mu$ и $\mathcal{D}(x)=\sigma^2$.

Отметим важное обстоятельство: это неравенство справедливо лишь для тех распределений, у которых среднее и дисперсия конечны.

Начнем доказательство с того, что запишем

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

и разобьем область интегрирования на три части:

$$(-\infty,\ \mu-k\sigma),\ (\mu-k\sigma,\ \mu+k\sigma) \text{ и } (\mu+k\sigma,\ \infty).$$

Тогда

$$\sigma^2 = \left[\int_{-\infty}^{\mu-k\sigma} + \int_{\mu-k\sigma}^{\mu+k\sigma} + \int_{\mu+k\sigma}^{\infty}\right](x-\mu)^2 f(x)\, dx.$$

Опуская средний из трех интегралов, приходим к неравенству

$$\sigma^2 \geqslant \left[\int_{-\infty}^{\mu-k\sigma} + \int_{\mu+k\sigma}^{\infty}\right](x-\mu)^2 f(x)\, dx.$$

В пределах каждой из этих областей интегрирования справедливо соотношение $(x-\mu)^2 \geqslant k^2\sigma^2$. Значит,

$$\sigma^2 \geqslant k^2\sigma^2\left[\int_{-\infty}^{\mu-k\sigma} + \int_{\mu+k\sigma}^{\infty}\right] f(x)\, dx =$$

$$= k^2\sigma^2 \mathcal{P}[-\infty \leqslant x \leqslant \mu-k\sigma \text{ или } \mu+k\sigma \leqslant x \leqslant \infty] =$$

$$= k^2\sigma^2 \mathcal{P}[|x-\mu| \geqslant k\sigma],$$

откуда и следует

$$\mathcal{P}[|x-\mu| \geqslant k\sigma] \leqslant \frac{1}{k^2}.$$

Это неравенство, вообще говоря, весьма слабое. В случае нормального распределения мы можем сравнить величину $1/k^2$ с точным значением доверительной вероятности, соответствующей выбранному интервалу.

В табл. 2 проводится такое сравнение для трех значений k, причем доверительные вероятности взяты из таблиц нормального распределения [30].

Таблица 2

k	Табличные значения вероятностей	Значения вероятностей согласно неравенству Чебышева
1	0,32	$\leqslant 1{,}000$
2	0,045	$\leqslant 0{,}250$
3	0,003	$\leqslant 0{,}111$

Выборочное среднее

Применим теперь неравенство Чебышева к величине $\bar{x}$ со среднеквадратичным отклонением $\sigma/\sqrt{n}$:

$$\mathscr{P}\left[|\bar{x}-\mu| \geqslant \frac{k\sigma}{\sqrt{n}}\right] \leqslant \frac{1}{k^2},$$

где n — объем выборки. Перепишем то же самое неравенство, введя обозначения $\varepsilon = k\sigma/\sqrt{n}$ и $k = \varepsilon\sqrt{n}/\sigma$. Тогда

$$\mathscr{P}[|\bar{x}-\mu| \geqslant \varepsilon] \leqslant \frac{\sigma^2}{n\varepsilon^2}. \qquad (54)$$

Если в квадратных скобках поменять смысл неравенства на противоположный, то можно получить соотношение, логически эквивалентное (54):

$$\mathscr{P}[|\bar{x}-\mu| < \varepsilon] \geqslant 1 - \frac{\sigma^2}{n\varepsilon^2}$$

или

$$\mathscr{P}[-\varepsilon < \bar{x}-\mu < \varepsilon] \geqslant 1 - \frac{\sigma^2}{n\varepsilon^2}. \qquad (55)$$

В таком случае утверждение слабого закона больших чисел сводится к следующему:

«Выберем сколь угодно малое ε, чтобы при этом $\bar{x}$ почти совпадало с μ. Из формулы (55) следует, что, выбирая n достаточно большим, утверждение $|\bar{x} - \mu| < \varepsilon$ можно сделать сколь угодно достоверным». Иными словами, вероятность того, что $\bar{x} \to \mu$ при $n \to \infty$, стремится к единице [1]).

Различие между слабым и усиленным законами больших чисел, грубо говоря, подобно различию между понятиями сходимости по точкам и равномерной сходимости, которые встречаются в математическом анализе [2]).

Центральная предельная теорема

Выше было показано, что точность оценки μ на основе выборочного среднего возрастает пропорционально $\sqrt{n}$. Далее с помощью закона больших чисел было показано, что можно доказать более сильное утверждение относительно сходимости $\bar{x}$ к μ по вероятности. Центральная предельная теорема приводит к еще более сильному утверждению относительно распределения $\bar{x}$ в целом, а не только его параметров $\mathscr{E}(\bar{x})$ и $\mathscr{D}(\bar{x})$.

Это наиболее важная теорема статистики и вообще одна из самых замечательных математических теорем. В ее разработке принимали участие многие крупнейшие математики, среди которых следует отметить Муавра, Лапласа, Гаусса, Чебышева и Ляпунова.

Теорема. Пусть случайная величина x имеет среднее значение μ и дисперсию σ^2. Если σ^2 конечно, то при стремлении объема выборки n к бесконечности распределение выборочного среднего $\bar{x}$ будет стремиться к нормальному со средним μ и дисперсией σ^2/n.

Доказательство. Используем производящие функции центральных моментов. По определению

$$M_x(t) = \mathscr{E}[e^{(x-\mu)t}] = 1 + \mu_1 t + \mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \ldots =$$

$$= 1 + 0 + \sigma^2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \ldots .$$

[1]) См. [17], гл. 11.

[2]) См., например, [21], § 7.23.

Пусть

$$w = (x - \mu)\frac{1}{\sigma\sqrt{n}}.$$

Для величины w ПФМ равна

$$M_w(t) = \mathscr{E}\left[e^{\frac{(x-\mu)t}{\sigma\sqrt{n}}}\right] = M_x\left(\frac{t}{\sigma\sqrt{n}}\right).$$

С другой стороны,

$$M_w(t) = 1 + \sigma^2\frac{t^2}{2!\,\sigma^2 n} + \mu_3\frac{t^3}{3!\,\sigma^3 n\sqrt{n}} + \ldots =$$

$$= 1 + \frac{t^2}{2n} + \mu_3\frac{t^3}{6\sigma^3 n\sqrt{n}} + \ldots.$$

Введем новую переменную z, являющуюся линейной функцией от $\bar{x}$:

$$z = \frac{\bar{x} - \mu}{\sigma(\bar{x})} = \frac{(\bar{x} - \mu)\sqrt{n}}{\sigma} = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - \mu)\sqrt{n}}{\sigma} = \sum_{i=1}^{n} w_i. \quad (56)$$

Согласно формуле (48), ПФМ суммы одинаково распределенных случайных величин есть n-я степень их ПФМ, т. е.

$$M_z(t) = [M_w(t)]^n =$$

$$= \left(1 + \frac{t^2}{2n} + \frac{\mu_3 t^3}{6\sigma^3 n\sqrt{n}} + \ldots\right)^n \to e^{t^2/2} \text{ при } n \to \infty.$$

На стр. 46 было показано, что ПФМ нормального распределения с нулевым средним значением и дисперсией σ^2 равна $e^{\sigma^2 t^2/2}$. Следовательно, величина z в пределе обладает нормальным распределением с нулевым средним и единичной дисперсией. Согласно формуле (56),

$$\bar{x} = \frac{\sigma z}{\sqrt{n}} + \mu.$$

На стр. 47 было показано, что линейная функция (в данном случае $\bar{x}$) случайной величины (z), распределенной по нормальному закону, также распределена по нормальному закону. Следовательно, величина $\bar{x}$ в пределе распре-

делена по нормальному закону со средним μ и дисперсией σ^2/n.

Отметим интересное обстоятельство, заключающееся в том, что при доказательстве теоремы ни слова не было сказано относительно характера распределения величины x. Более того, оно даже не предполагалось непрерывным, хотя таково предельное распределение для $\bar{x}$. Налагаемое на дисперсию требование конечности не вносит сильного ограничения применимости теоремы, поскольку почти во всех практических случаях область изменения рассматриваемой величины ограничена, что автоматически делает ее дисперсию конечной. Тем не менее условие конечности дисперсии существенно необходимо.

Использование таблиц нормального распределения для вычисления биномиального распределения [1])

Предположим, что игральную кость бросали 300 раз. Появлению чисел 1 или 2 поставим в соответствие величину $x=1$, а остальным вариантам $x=0$. Это означает, что $p={}^1/_3$, $q={}^2/_3$. Полное число благоприятных исходов равно $r=\sum_{i=1}^{300} x_i$; среднее значение x равно $\bar{x}=r/300$, причем $\mathscr{E}(\bar{x})={}^1/_3$.

Распределение (дискретное) величины r имеет вид

$$P_r=C_{300}^r\left(\frac{1}{3}\right)^r\left(\frac{2}{3}\right)^{300-r}.$$

Пусть нас интересует вероятность того, что число благоприятных исходов будет отличаться от 100 не более чем на 15. Чтобы найти эту вероятность, необходимо вычислить

$$\mathscr{P}=\sum_{85}^{115} P_r,$$

что весьма утомительно.

Можно переформулировать задачу вычисления вероятности $\mathscr{P}(85\leqslant r\leqslant 115)$, сведя ее к эквивалентной задаче вычисления вероятности $\mathscr{P}(85/300\leqslant r/300\leqslant 115/300)$, т. е.

[1]) См. [27], § 7.7.

к вычислению

$$\mathscr{P}\left(\frac{85}{300} \leqslant \bar{x} \leqslant \frac{115}{300}\right).$$

Поскольку теперь нам предстоит иметь дело со средним значением $\bar{x}$, воспользуемся центральной предельной теоремой. Соответствующее нормальное распределение для $\bar{x}$ обладает средним $\bar{x} = 1/3$ и дисперсией $pq/n = 2/2700$. Запишем это в виде [1])

$$\bar{x} \sim N\left(\frac{1}{3}, \frac{2}{2700}\right)$$

и вычислим

$$\mathscr{P} = \int_{85/300}^{115/300} N\left(\frac{1}{3}, \frac{2}{2700}\right) d\bar{x}. \tag{57}$$

В таблицах обычно приводятся распределения для вероятностей $N(0, 1)$. Введем стандартизованную переменную

$$w = \frac{\bar{x} - \frac{1}{3}}{\sqrt{2/2700}} = \frac{\bar{x} - \mathscr{E}(\bar{x})}{\sigma(\bar{x})},$$

такую, что $w \sim N(0, 1)$. Выражение (57) при этом преобразуется в

$$\mathscr{P} = \int_{-1{,}84}^{1{,}84} N(0, 1)\, dw,$$

и из таблиц находим

$$\mathscr{P} = 0{,}93.$$

Иными словами, вероятность того, что полное число благоприятных исходов не заключено в пределах от 85 до 115, составляет лишь 7%.

Пример, относящийся к распределению Коши

Мы хотим продемонстрировать необходимость конечной дисперсии для справедливости центральной предельной теоремы. В случае распределения Коши не существует

[1]) Символ $\sim$ обозначает «распределено как».— *Прим. ред.*

ПФМ. В связи с этим воспользуемся *характеристической функцией* $\varphi(t)$. По своему определению она сходна с ПФМ:

$$\varphi_x(t) = \mathscr{E}(e^{ixt}). \tag{58}$$

В случае распределения Коши

$$\varphi_x(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{itx}}{1+x^2}\, dx.$$

Этот интеграл можно вычислить, интегрируя вдоль контура, что дает

$$\varphi_x(t) = e^{-|t|}.$$

Можно показать, что характеристическая функция суммы n одинаковых случайных величин получается возведением $\varphi_x(t)$ в n-ю степень. Выражение для характеристической функции, аналогичное формуле (48), имеет вид

$$\varphi_{\sum x_i}(t) = e^{-n|t|} = e^{-|nt|}.$$

Используя формулу (40), можно, наконец, показать, что

$$\varphi_{\bar{x}}(t) = \varphi_{\sum x_i}\left(\frac{t}{n}\right) = e^{-|t|}.$$

В итоге распределение величины $\bar{x}$ вновь оказывается распределением Коши, что и подтверждает несправедливость в данном случае центральной предельной теоремы. Причиной этого является бесконечная дисперсия исходного распределения.

§ 9. Распределение вероятности для функции случайной величины

Формулы (22) и (23) показывают, как находить математическое ожидание функции $h(x)$ случайной величины x. С помощью так называемого правила *переноса ошибок* (propagation of errors) можно также в первом приближении найти и дисперсию функции $h(x)$. В конечном итоге может быть найдено и полное распределение вероятности для функции $h(x)$.

Перенос ошибок

Ограничимся несколькими первыми членами разложения $h(x)$ в ряд Тэйлора относительно точки μ. В этом случае

$$h(x) = h(\mu) + (x-\mu)h'(\mu) + (x-\mu)^2 \frac{h''(\mu)}{2!} + \\ + \text{Члены более высокого порядка.} \qquad (59)$$

Наши приближения сводятся к следующим:

1) $$\mathscr{E}[h(x)] = h(\mu) + 0 + \frac{\sigma^2 h''(\mu)}{2!} + \ldots \approx \\ \approx h(\mu) = h[\mathscr{E}(x)],$$

2) $$\mathscr{D}[h(x)] = \mathscr{E}[h(x) - \mathscr{E}[h(x)]]^2 \approx \\ \approx \mathscr{E}[h(x) - h(\mu)]^2.$$

Из формулы (59) имеем

$$h(x) - h(\mu) = (x-\mu)h'(\mu) + (x-\mu)^2 \frac{h''(\mu)}{2!} + \ldots .$$

Следовательно,

$$\mathscr{D}[h(x)] \approx \mathscr{E}[(x-\mu)h'(\mu)]^2 \approx \\ \approx [h'(\mu)]^2 \mathscr{D}(x), \qquad (60)$$

$$\approx [h'(x)]^2 \mathscr{D}(x). \qquad (61)$$

Аналогично можно получить

$$\sigma[h(x)] \approx |h'(\mu)| \sigma(x), \qquad (62)$$

$$\approx |h'(x)| \sigma(x). \qquad (63)$$

Соотношение (63) часто записывается в виде

$$\delta[h(x)] \approx \frac{dh(x)}{dx} \delta x. \qquad (64)$$

Соотношения (60) — (64) точны лишь в том случае, когда ряд Тэйлора оканчивается членом, содержащим $h'(\mu)$, т. е. когда $h(x)$ представляет собой линейную функцию

$$h(x) = ax + b.$$

Тогда выражения

$$\mathscr{E}[h(x)] = a\mu + b$$

и

$$\mathscr{D}[h(x)] = a^2\sigma^2$$

точны. Во всех прочих случаях мы делаем ряд приближений, считая функцию линейной. Тем не менее получающиеся результаты весьма полезны, ибо они указывают порядок величины дисперсии новой случайной величины $h(x)$. Аналогичным образом, если h есть функция m случайных величин, то, используя m-мерное разложение в ряд Тэйлора, получаем

$$h(x_1, \ldots, x_m) \approx h(\mu_1, \ldots, \mu_m) + \sum_{i=1}^{m} (x_i - \mu_i) \frac{\partial h(\mu)}{\partial x_i},$$

где

$$\mathscr{E}(x_i) = \mu_i \quad \text{и} \quad \mu = (\mu_1, \mu_2, \ldots, \mu_m).$$

Отсюда следует

$$\mathscr{E}[h(\mathbf{x})] \approx h(\mu)$$

и

$$\mathscr{D}[h(\mathbf{x})] \approx \mathscr{E}[h(\mathbf{x}) - \mathscr{E}(h(\mathbf{x}))]^2 \approx$$

$$\approx \mathscr{E}[h(\mathbf{x}) - h(\mu)]^2 \approx \sum_{i=1}^{m} \left[\frac{\partial h(\mu)}{\partial x_i}\right]^2 \mathscr{E}(x_i - \mu_i)^2,$$

так как $\mathscr{E}[(x_i - \mu_i)(x_j - \mu_j)] = 0$ при $i \neq j$.

Таким образом,

$$\mathscr{D}[h(\mathbf{x})] = \sum_{i=1}^{m} \left[\frac{\partial h(\mu)}{\partial x_i}\right]^2 \mathscr{D}(x_i),$$

$$\mathscr{D}[h(\mathbf{x})] \approx \sum_{i=1}^{m} \left[\frac{\partial h(\mathbf{x})}{\partial x_i}\right]^2 \sigma_i^2 \tag{65}$$

и соответственно

$$\sigma[h(\mathbf{x})] = \sqrt{\mathscr{D}[h(\mathbf{x})]} \approx$$

$$\approx \sqrt{\sum_{i=1}^{m} \left[\frac{\partial h(\mu)}{\partial x_i}\right]^2 \sigma_i^2} \approx \sqrt{\sum_{i=1}^{m} \left[\frac{\partial h(\mathbf{x})}{\partial x_i}\right]^2 \sigma_i^2}, \tag{66}$$

если заменить величину μ ее оценкой $\mathbf{x}$.

Аналогичное выражение в дифференциальной форме имеет вид

$$\delta h(\mathbf{x}) \approx \sum_{i=1}^{m} \frac{\partial h(\mathbf{x})}{\partial x_i} \delta x_i. \tag{67}$$

Все приведенные формулы представляют собой приближения, справедливые лишь в случае небольших отклонений x_i относительно их средних μ_i.

Пример. Рассмотрим преобразование $h(x) = \cos x$, причем $\mathscr{D}(x) = \sigma^2$. Тогда

$$\frac{dh(x)}{dx} = -\sin x$$

и

$$\frac{dh(\mu)}{dx} = \frac{dh(x)}{dx}\Big|_{x=\mu} = -\sin \mu.$$

Следовательно,

$$\mathscr{D}(h) \approx \sigma^2 \sin^2 \mu, \tag{68}$$

$$\mathscr{D}(h) \approx \sigma^2 \sin^2 x. \tag{69}$$

Поскольку дисперсия h есть величина постоянная, то случайная величина (69) служит лишь оценкой дисперсии (68).

Распределение вероятности для функции дискретной величины

Предположим, что величина x имеет дискретное распределение. Тогда распределение для $h(x)$ легко получить путем непосредственного вычисления.

а) Если существует лишь одно значение $x = x_0$, при котором $h(x) = h(x_0)$, то

$$\mathscr{P}[h(x) = h_0] = \mathscr{P}[h(x) = h(x_0)] = \mathscr{P}(x = x_0). \tag{70}$$

Пример. Пусть $h(x) = \cos x$, а распределение величины x задается таблицей

x	0	$\pi/4$	$\pi/2$
$\mathscr{P}(x)$	1/6	2/6	3/6

В таком случае распределение для $h(x)$ оказывается следующим:

h	0	0,71	1
$\mathcal{P}(h)$	3/6	2/6	1/6

б) Если существует несколько значений $x = x_0$, $x_1, \ldots, x_m$, при которых $h(x) = h_0$, то

$$\mathcal{P}[h(x) = h_0] = \mathcal{P}[x = x_0, \text{ или } x = x_1, \text{ или } \ldots, \text{ или } x = x_m] =$$

$$= \sum_{i=0}^{m} \mathcal{P}(x = x_i).$$

Следовательно,

$$\mathcal{P}[h(x) = h_0] = \sum_{h(x_i)=h_0} \mathcal{P}(x = x_i). \qquad (71)$$

Пример. Пусть $h(x) = \cos x$, а распределение величины x задается таблицей

x	$-\pi/2$	$-\pi/4$	0	$\pi/4$	$\pi/2$
$\mathcal{P}(x)$	1/35	3/35	6/35	10/35	15/35

Тогда распределение для $h(x)$ имеет вид

h	0	0,71	1
$\mathcal{P}(h)$	16/35	13/35	6/35

Распределение вероятности для функции непрерывной величины [1])

Пусть дана

$$f(x)\,dx = \mathscr{P}[x \leqslant X \leqslant x + dx] \tag{72}$$

и требуется найти

$$g(h)\,dh = \mathscr{P}[h \leqslant H \leqslant h + dh], \tag{73}$$

где $h = h(x)$.

Допустим сначала, что $h(x)$ — однозначная функция x. В случае дискретной величины мы находили значение $h = h_0$, соответствующее заданному значению $x = x_0$. При этом вероятность x_0 просто сопоставлялась h_0. Для непрерывной величины мы можем найти малый *интервал значений* h, соответствующий заданному малому интервалу значений x с известной вероятностью $f(x)\,dx$.

Воспользуемся обратной функцией

$$x = x(h), \tag{74}$$

которая позволяет определить значение x, соответствующее заданному h. Подставляя $x = x(h)$, получаем сначала выражение для $f(x)$ как функции h

$$f(x) = f[x(h)]. \tag{75}$$

Необходимо, кроме того, преобразовать бесконечно малое приращение согласно тем же правилам, которыми пользуются при вычислении интегралов методом подстановки. Единственное отличие при этом состоит в необходимости сохранить преобразованное выражение положительным в соответствии с тем, что плотность вероятности должна оставаться неотрицательной. Тогда получаем

$$dx = \frac{dx(h)}{dh}\,dh, \tag{76}$$

если $dx(h)/dh$ положительно, и

$$dx = -\frac{dx(h)}{dh}\,dh, \tag{77}$$

если $dx(h)/dh$ отрицательно.

[1]) См. [27], гл. 10.

Выражения (76) и (77) эквивалентны:

$$dx = \left| \frac{dx(h)}{dh} \right| dh. \tag{78}$$

Объединяя соотношения (75) и (78), имеем

$$f(x)\,dx = f[x(h)] \cdot \left| \frac{dx(h)}{dh} \right| dh. \tag{79}$$

Если сравнить эту формулу с формулой (73), можно получить выражение для новой плотности вероятности, зависящей от h:

$$g(h) = f[x(h)] \cdot \left| \frac{dx(h)}{dh} \right|. \tag{80}$$

Пример. Пусть $h(x) = \cos x$, а распределение вероятности для величины x дается выражением

$$f(x)\,dx = a + bx, \quad 0 \leqslant x \leqslant \frac{\pi}{2},$$

где $a = 1/\pi$, $b = 4/\pi^2$. С помощью обратного преобразования получаем $x = \arccos h$, причем

$$\frac{dx}{dh} = \frac{-1}{\sqrt{1-h^2}} \quad \text{для} \quad 0 \leqslant x \leqslant \frac{\pi}{2}.$$

Если теперь заменить $f(x)$ на $f[x(h)]$ и dx на $|dx(h)/dh|\,dh$, то мы получим

$$g(h)\,dh = f[x(h)] \left| \frac{dx(h)}{dh} \right| dh = \\ = [a + b \cdot \arccos h] \frac{dh}{\sqrt{1-h^2}}, \qquad 0 \leqslant h \leqslant 1.$$

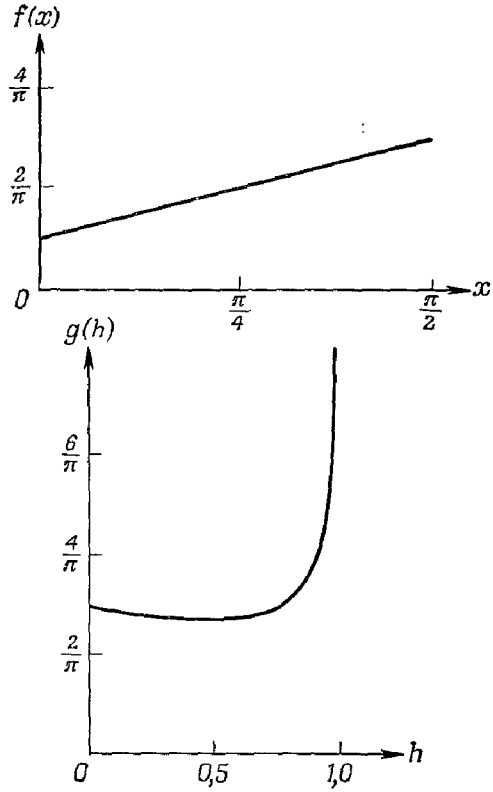
Этот результат проиллюстрирован на фиг. 9.

Аналогичное рассмотрение можно провести и в случае неоднозначной функции $h(x)$, т. е. в том случае, когда одному и тому же значению $h(x)$ соответствует несколько различных значений x. Один из возможных примеров представлен на фиг. 10.

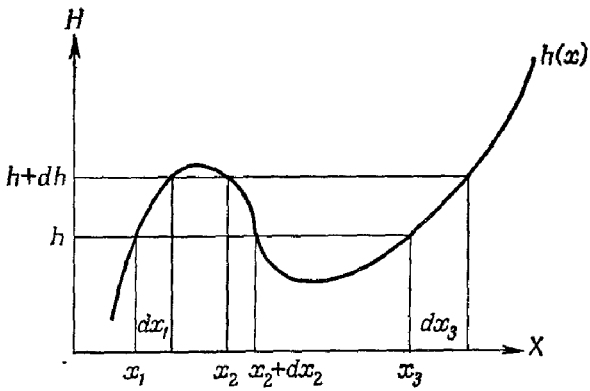
Исходным соотношением является равенство

$$\mathcal{P}[h \leqslant H \leqslant h + dh] = \mathcal{P}[x_1 \leqslant X \leqslant x_1 + dx_1,$$

или $x_2 \leqslant X \leqslant x_2 + dx_2$, или ..., или $x_m \leqslant X \leqslant x_m + dx_m]$.


f(x)
4/π
2/π
0
π/4
π/2
x
g(h)
6/π
4/π
2/π
0
0,5
1,0
h

Фиг. 9.


H
h(x)
h+dh
h
dx1
dx3
X
x1
x2
x2+dx2
x3

Фиг. 10.

Можно показать, что при этом

$$g(h)\,dh = f(x_1)\left|\frac{dx_1}{dh}\right|dh + f(x_2)\left|\frac{dx_2}{dh}\right|dh + \ldots$$
$$\ldots + f(x_m)\left|\frac{dx_m}{dh}\right|dh, \qquad (81)$$

или

$$g(h)\,dh = \sum_{h(x_i)=h} f[x_i(h)]\left|\frac{dx_i(h)}{dh}\right|dh. \qquad (82)$$

Распределение χ^2

Пусть x_i — независимые случайные величины, распределенные по нормальному закону с нулевым средним значением и единичным средним крадратичным отклонением.

Введем

$$u = \sum_{i=1}^{k} x_i^2 \qquad (83)$$

и будем называть u случайной величиной χ^2 с k степенями свободы.

Сначала найдем распределение для функции $h(x) = x^2$, имея в виду, что

$$f(x)\,dx = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx.$$

Рассматриваемая функция является двузначной, поскольку $h(x) = x^2 = (-x)^2$. Это означает, что существуют две обратные ей функции:

$$x_1(h) = \sqrt{h} \quad \text{и} \quad x_2(h) = -\sqrt{h}.$$

Их производные соответственно равны

$$\frac{dx_1(h)}{dh} = \frac{1}{2\sqrt{h}} \quad \text{и} \quad \frac{dx_2(h)}{dh} = -\frac{1}{2\sqrt{h}}.$$

Воспользуемся формулой (81) для получения плотности распределения по h. Тогда

$$g(h)\,dh = f[x_1(h)]\left|\frac{dx_1(h)}{dh}\right|dh + f[x_2(h)] \times$$
$$\times \left|\frac{dx_2(h)}{dh}\right|dh = f(\sqrt{h})\left|\frac{1}{2\sqrt{h}}\right|dh + f(-\sqrt{h}) \times$$
$$\times \left|\frac{-1}{2\sqrt{h}}\right|dh = \frac{1}{\sqrt{2\pi}}\left[e^{-h/2} + e^{-h/2}\right]\frac{dh}{2\sqrt{h}}.$$

Делая дальнейшие преобразования, получаем

$$g(h)\,dh = \frac{1}{\sqrt{2\pi}}\,h^{-1/2}e^{-h/2}\,dh =$$

$$= \frac{1}{\left(-\frac{1}{2}\right)!\,2^{-1/2+1}}\,h^{-1/2}e^{-h/2}\,dh =$$

$$= \frac{1}{\alpha!\,\beta^{\alpha+1}}\,h^{\alpha}e^{-h/\beta}\,dh, \qquad 0 \leqslant h \leqslant \infty, \tag{84}$$

где $\alpha = -1/2$ и $\beta = 2$.

Итак, $g(h)$ представляет собой гамма-распределение, которое рассматривалось на стр. 47. В гл. 1, § 6, было показано, что производящая функция моментов для гамма-распределения $M_h'(t) = (1-\beta t)^{-\alpha-1}$, или в данном случае

$$M_h'(t) = (1-2t)^{-1/2}. \tag{85}$$

Из определения величины χ_k^2 [формула (83)] вытекает, что выражение для u можно переписать в виде

$$u = \sum_{i=1}^{k} h_i,$$

т. е. представить как сумму независимых случайных величин с одинаковым гамма-распределением.

Таким образом,

$$M_u'(t) = [M_h'(t)]^k = (1-2t)^{-k/2}.$$

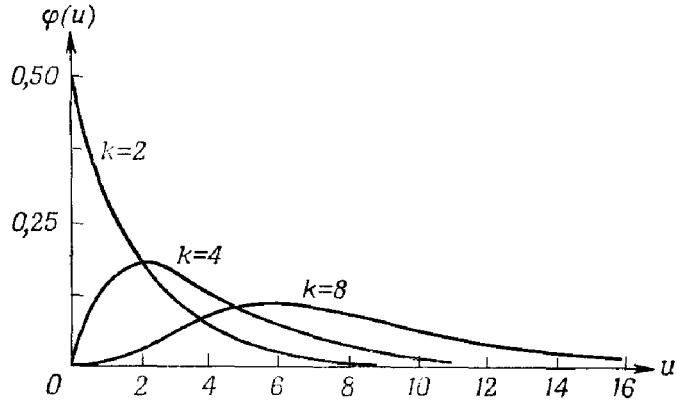
Полученное выражение совпадает с (42), причем $\alpha = (k/2) - 1$, $\beta = 2$. Это означает, что величина u имеет гамма-распределение с приведенными выше параметрами, т. е.

$$\varphi(u) = \frac{1}{\left(\frac{k}{2}-1\right)!\,2^{k/2}} \cdot u^{(k/2)-1} \cdot e^{-u/2}, \qquad 0 \leqslant u \leqslant \infty. \tag{86}$$

Тем самым по существу найдено и распределение χ_k^2.

На фиг. 8 были изображены кривые, соответствующие значениям параметра $\alpha = 0$, 1 и 3. Они представляют собой графики функции $\varphi(u)$ соответственно при $k = 2$, 4, 8 с той лишь разницей, что масштаб по оси абсцисс должен быть удвоен (фиг. 11). Следует заметить, что

распределение достигает максимума при значении $u = k - 2$, а математическое ожидание величины u равно k.



Фиг. 11.

Найдем математическое ожидание величины χ_k^2, используя ПФМ:

$$M'(t) = (1-2t)^{-k/2}.$$

Тогда

$$\frac{\partial M'(t)}{\partial t} = k(1-2t)^{-(k/2)-1}$$

и

$$\mu_1' = \mathscr{E}(\chi_k^2) = \frac{\partial M'(0)}{\partial t} = k.$$

Подобным же образом можно получить

$$\mathscr{D}(\chi_k^2) = 2k.$$

Случай неединичной дисперсии

Пусть величина x распределена по нормальному закону $N(0, \sigma^2)$ вместо $N(0, 1)$, т. е. $x \sim N(0, \sigma^2)$. Тогда величина x_i/σ будет обладать единичной дисперсией, и выражение $u = \sum_{i=1}^{k} (x_i/\sigma)^2$ вновь окажется величиной χ_k^2. В этом случае $\sum_{i=1}^{k} x_i^2 = \sigma^2 u$ представляет собой случайную величину $\sim \sigma^2 \chi_k^2$.

Преобразование нескольких случайных величин

Допустим, что необходимо преобразовать случайные величины

$$x_1, x_2, \ldots, x_n$$

в

$$h_1, h_2, \ldots, h_n,$$

причем $h_i = h_i(x_1, \ldots, x_n)$ при каждом i.

Совместная плотность вероятности $f(x_1, \ldots, x_n)$, зависящая от x_i, преобразуется в новую плотность, зависящую от h_i, посредством операции, которая является обобщением правила

$$g(h)\,dh = f[x(h)]\left|\frac{dx(h)}{dh}\right|dh$$

для случая одной переменной.

Прежде всего найдем обратное преобразование

$$x_i = x_i(h_1, h_2, \ldots, h_n) = x_i(\mathbf{h}).$$

Плотностью вероятности для h_i является

$$g(h_1, \ldots, h_n) = f[x_1(\mathbf{h}), x_2(\mathbf{h}), \ldots, x_n(\mathbf{h})] \times \left|\frac{\partial(x_1, x_2, \ldots, x_n)}{\partial(h_1, h_2, \ldots, h_n)}\right|, \qquad (87)$$

где $|\partial(x_1, \ldots, x_n)/\partial(h_1, \ldots, h_n)|$ — абсолютное значение якобиана преобразования.

Определение. Линейным ортогональным преобразованием называется линейное преобразование

$$h_i = \sum_{j=1}^{n} c_{ij}x_j \qquad (i = 1, 2, \ldots, n)$$

при условии, что

$$\sum_{i=1}^{n} h_i^2 = \sum_{j=1}^{n} x_j^2.$$

Якобиан такого преобразования равен единице. Кроме того, если

$$h_1 = c_j \sum x_j,$$

то для $h_2, h_3, \ldots, h_n$ имеем

$$\sum_{j=1}^{n} c_{ij} = 0 \quad \text{при} \quad i \geqslant 2.$$

В матричных обозначениях рассматриваемое преобразование имеет вид

$$\mathbf{h} = C\mathbf{x}.$$

Поскольку должно выполняться соотношение

$$\mathbf{h}'\mathbf{h} = \mathbf{x}'C'C\mathbf{x} = \mathbf{x}'\mathbf{x},$$

то отсюда следует

$$C'C = I.$$

Такое преобразование эквивалентно вращению координатных осей в n-мерном евклидовом пространстве.

Сумма квадратов отклонений от среднего

Рассмотрим случайную выборку, содержащую n независимых величин x_i из совокупности, которая подчиняется нормальному закону $N(\mu, \sigma^2)$. Пусть выборочное среднее равно $\bar{x}$.

Введем величину

$$w = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2. \tag{88}$$

Чтобы найти распределение случайной величины w, воспользуемся ортогональным линейным преобразованием. Введем новую переменную

$$h_1 = \frac{x_1}{\sqrt{n}} + \frac{x_2}{\sqrt{n}} + \ldots + \frac{x_n}{\sqrt{n}} = \bar{x}\sqrt{n}.$$

Остальные h_i обладают следующим свойством:

$$\mathscr{E}(h_i) = \mathscr{E}\left(\sum c_{ij}x_j\right) = \sum c_{ij}\mathscr{E}(x_j) = \mu \sum c_{ij},$$

т. е.

$$\mathscr{E}(h_i) = 0 \quad \text{при} \quad i \geqslant 2.$$

Кроме того,

$$\sum_{i=1}^{n} h_i^2 = \sum_{j=1}^{n} x_j^2.$$

Из формулы (88) следует

$$w = \sum_{i=1}^{n} h_i^2 - h_1^2 = \sum_{i=2}^{n} h_i^2,$$

где $h_i \sim N(0; \sigma^2)$ при $i \geqslant 2$.

Таким образом,

$$w \sim \sigma^2 \chi^2_{n-1}.$$

Это означает, что сумма квадратов отклонений от среднего пропорциональна случайной величине χ^2 с $(n-1)$ степенями свободы. Грубо говоря, мы показали, что сумма квадратов n независимых нормально распределенных случайных величин $(x_i - \bar{x})$, т. е.

$$w = \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

эквивалентна сумме квадратов $(n-1)$ независимых нормально распределенных случайных величин h_i.

Появление зависимости случайных величин объясняется тем, что n величин $z_i = x_i - \bar{x}$ связаны соотношением

$$\sum_{i=1}^{n} z_i = 0.$$

В общем случае число степеней свободы ν распределения χ^2 для суммы квадратов типа (88) равно

$$\nu = n - \text{Число линейных соотношений.}$$

Выборочная дисперсия

Выборочная дисперсия s^2 определяется следующим образом:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=2}^{n} h_i^2$$

независимо от вида распределения величин h_i или x_i.

Так как $\mathscr{E}(h_i) = 0$, то $\mathscr{E}(h_i^2) = \sigma^2$ и, следовательно,

$$\mathscr{E}(s^2) = \frac{1}{n-1} \sum_{i=2}^{n} \sigma^2 = \sigma^2$$

при любом распределении.

В случае нормального распределения

$$s^2 \sim \frac{\sigma^2 \chi^2_{n-1}}{n-1}, \tag{89}$$

причем

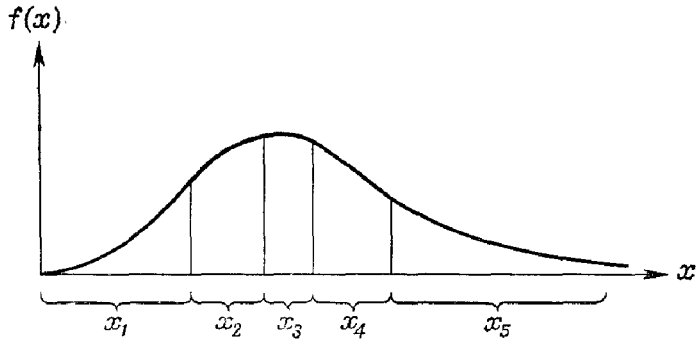
$$\sigma(s^2) = \sigma^2 \sqrt{\frac{2}{n-1}}.$$

ГЛАВА ТРЕТЬЯ

Проверка гипотез

§ 10. Критерий согласия χ^2 [1])

Рассмотрим сначала случайную величину x, которая может принимать k возможных значений $x_1, \ldots, x_k$. Это означает, что x обладает дискретным распределением. В случае непрерывного распределения k исходов можно определить подобно тому, как это сделано на фиг. 12, где $k = 5$.



Фиг. 12.

Пусть $\theta_i = 1$, когда, $x = x_i$ и $\theta_i = 0$ в остальных случаях (где $i = 1, 2, \ldots, k$). Иными словами, θ_i представляют собой случайные числа, подчиняющиеся условию

$$\sum_{i=1}^{k} \theta_i = 1.$$

В более общем случае, когда опыт повторяется n раз, введем величину θ_{ij}, причем $\theta_{ij} = 1$, если $x = x_i$ в j-м опыте, и $\theta_{ij} = 0$ в остальных случаях.

[1]) См. [37], § 79, или [21], пример 15.3.

Полную частоту появления r_i мы получим, просуммировав θ_{ij} по полному числу опытов $j = 1, 2, \ldots, n$, т. е.

$$r_i = \sum_{j=1}^{n} \theta_{ij}, \qquad i = 1, 2, \ldots, k. \tag{90}$$

Полученное выражение является прямым обобщением биномиального распределения с $k = 2$; распределение величины r_i называется *полиномиальным распределением*.

Запишем его в виде

$$\mathcal{P}(r_1, r_2, \ldots, r_k) = \frac{n!}{r_1!\, r_2! \ldots r_k!}\, p_1^{r_1} p_2^{r_2} \ldots p_k^{r_k}, \tag{91}$$

где $\mathcal{P}(x = x_i) = p_i$ и $\sum_{i=1}^{k} r_i = n$.

Выражение (90) для частоты r_i подчеркивает тот факт, что каждое r_i является суммой большого числа n случайных величин θ_{ij} с одинаковым распределением. Таким образом, для каждого r_i справедлива центральная предельная теорема.

Важно отметить, что в том случае, когда величина r_i описывается полиномиальным распределением, можно весьма точно определить скорость, с которой оно сходится к нормальному распределению.

Положим

$$\mathcal{E}(r_i) = np_i = m_i.$$

Каждое из r_i в отдельности распределено по биномиальному закону. Мы уже видели (стр. 23 и 26), что биномиальное распределение быстро сходится к распределению Пуассона со средним значением и дисперсией, равными $\mu = np_i$. В свою очередь распределение Пуассона сходится к нормальному распределению по мере возрастания μ, причем сходимость можно легко оценить. Для удобства мы можем считать, что при $\mu \geqslant 5$ распределение Пуассона уже достаточно хорошо аппроксимирует нормальное распределение. На фиг. 13 изображено распределение Пуассона для случая $\mu = 5$.

Таким образом, при $np_i \geqslant 5$ случайная величина r_i распределена примерно по нормальному закону со средним значением np_i и дисперсией np_i. Значит, распределе-

ние величины

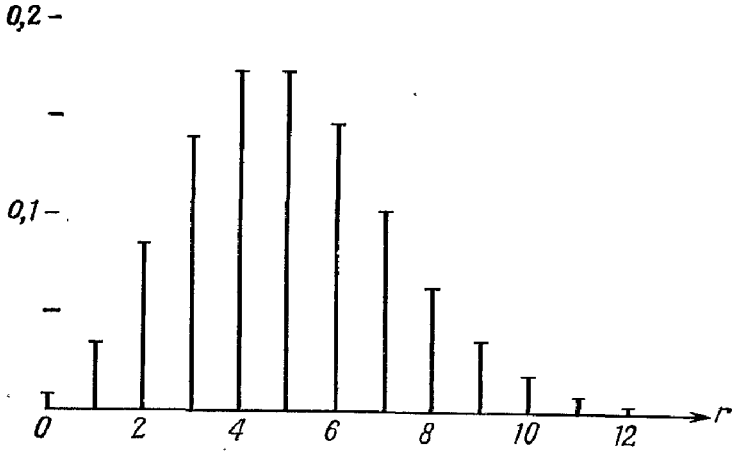
$$z_i = \frac{r_i - np_i}{\sqrt{np_i}} = \frac{r_i - m_i}{\sqrt{m_i}} \qquad (92)$$

примерно совпадает с $N(0, 1)$. Таким образом, величина

$$u = \sum_{i=1}^{k} z_i^2 = \sum_{i=1}^{k} \frac{(r_i - m_i)^2}{m_i} \qquad (93)$$

распределена примерно по закону χ^2.

Допустим, что мы хотим проверить гипотезу, состоящую в том, что $\mathscr{E}(r_i) = m_i$ $(i = 1, 2, \ldots, k)$. Для этого вычислим величину u и посмотрим, *настолько ли хорошо*



Фиг. 13.

согласие между r_i *и* m_i, *что и можно рассматривать как приемлемую выборку из распределения* χ^2. Когда r_i имеет распределение, у которого $\mathscr{E}(r_i)$ не совпадает с величиной m_i, проверяемой данной гипотезой, формула (93) будет предсказывать очень большое значение u.

Число степеней свободы

Выше было получено соотношение

$$\sum_{i=1}^{k} r_i = n,$$

связывающее k величин r_i. Часто бывает известно, что распределение $\{p_i\}$ принадлежит к определенному классу,

однако полностью найти его не удается до тех пор, пока из экспериментальных данных не произведена оценка одного или нескольких параметров этого распределения. Если данные используются для оценки t параметров, то это подразумевает наличие t дополнительных соотношений, и, таким образом, число степеней свободы оказывается равным

$$\nu = k - 1 - t. \tag{94}$$

Пример. Двадцать лет собирались сведения о количестве кавалеристов прусской армии, погибших в результате гибели под ними коня [4]. Эти данные основывались на ежегодных докладах десяти армейских корпусов, представленных за двадцать лет, что в целом составляет 200 донесений.

Обозначим через x число людей, погибших в одном корпусе за год. Соответствующие частоты r_x даны в табл. 3.

Таблица 3

Число погибших офицеров, x	0	1	2	3	4	5	6	Всего
Число донесений r_x с указанием x погибших	109	65	22	3	1	0	0	200
Ожидаемые (по Пуассону) частоты m_x	108,7	66,3	20,2	4,1	0,6	0,07	0,01	200

Из таблицы следует, что полное число погибших равно $0 \times 109 + 1 \times 65 + 2 \times 22 + 3 \times 3 + 4 \times 1 = 122$. Таким образом, среднее число погибших в одном корпусе за год составляло $122/200 = 0{,}61$. Проверим гипотезу о том, что эти данные описываются распределением Пуассона, т. е. что

$$\mathcal{P}(x = x_i) = p_i = \frac{\mu^i e^{-\mu}}{i!}.$$

Мы видим, что для полного определения p_i нужно оценить параметр μ. Пусть этой оценкой будет выборочное среднее $\hat{\mu} = 0{,}61$.

Тогда для ожидаемых частот получаем

$$m_i = np_i = 200 \cdot 0{,}61^i \cdot e^{-0{,}61} \cdot (i!)^{-1}.$$

Эти значения приведены в третьей строке табл. 3. Поскольку ожидаемые частоты для $x > 2$ малы, мы объединим последние четыре значения x и составим новую табл. 4.

Таблица 4

x	0	1	2	$\geqslant 3$
r_x	109	65	22	4
m_x	108,7	66,3	20,2	4,8

Для проверки гипотезы вычислим

$$u = \sum (r_x - m_x)^2 \cdot \frac{1}{m_x} = \frac{(0,3)^2}{108,7} + \frac{(1,3)^2}{66,3} + \frac{(1,8)^2}{20,2} + \frac{(0,8)^2}{4,8} = 0{,}32.$$

Наблюдаемые частоты связаны двумя линейными соотношениями

$$\sum r_x = 200 \quad \text{и} \quad \sum x r_x = 122.$$

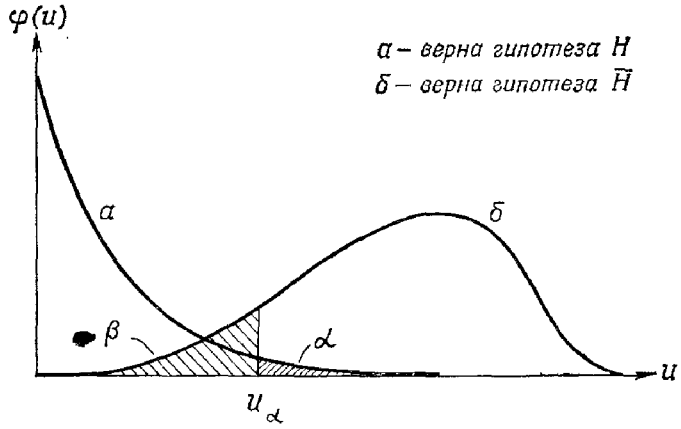
Второе соотношение обусловлено тем, что мы оценивали параметр μ по данной выборке. Таким образом, число степеней свободы равно $\nu = 4 - 1 - 1 = 2$. Следовательно, величина u должна быть распределена примерно как χ^2_2, т. е. $u \sim \chi^2_2$.

Выясним теперь, не окажется ли величина u столь большой, что исходная гипотеза (о распределении Пуассона) неправдоподобна. Из таблиц распределения χ^2_2 (см., например, [7, 32]) [1]) находим, что с вероятностью 0,95

[1]) См. также табл. IV, стр. 258.— *Прим. ред.*

величина u должна лежать в интервале от 0 до 6,0. Поскольку значение u попадает в этот *незначимый* интервал, проверяемая гипотеза принимается.

Но, с другой стороны, если бы мы получили для u значение 6,1 или больше, то должны были бы считать его *значимым*. Дело в том, что если проверяемая гипотеза верна, то не более чем в 5 случаях из 100 величина u



Фиг. 14.

будет принимать столь большие значения. Мы говорим, что $u = 6{,}0$ есть *5%-ная граница* распределения χ^2_2.

Наряду с этим укажем и другие значения χ^2_2, соответствующие уровням значимости 1% и 0,1%:

$$[\chi^2_2]_{1\%} = 9{,}2, \qquad [\chi^2_2]_{0{,}1\%} = 13{,}8.$$

Рассмотренный пример иллюстрирует общую теорию проверки гипотез. Сначала по экспериментальным данным составляется функция u и на основе «нулевой гипотезы» H вычисляется ее распределение. Затем проверяют, как изменится u, если гипотеза H неверна. Обычно u в этом случае принимает намного большие значения.

На фиг. 14 изображены два распределения величины u, соответствующие двум конкурирующим гипотезам H и $\bar{H}$. Если из опыта получается большое значение u, то можно сказать, что либо справедлива гипотеза H, но произошло чрезвычайно маловероятное событие, либо более правдо-

подобна справедливость гипотезы $\overline{H}$. Если оставаться в пределах одной лишь гипотезы, то результат следует признать *значимым* в том смысле, что он выходит за рамки обычного. Теория проверки гипотез позволяет оценить и неопределенность, присущую сделанному выводу.

Отвергая *нулевую гипотезу* H, когда она справедлива, мы совершаем *ошибку первого рода*. Принимая *нулевую гипотезу* H, в то время как на самом деле справедлива конкурирующая с ней гипотеза $\overline{H}$, мы совершаем *ошибку второго рода* (подробнее см. [27], гл. 12) [1]).

§ 11. Распределение t Стьюдента и его применения

Начнем с примера. Пусть мы хотим проверить, действительно ли средний рост большой группы людей равен 167 *см*. В выборке из десяти человек рост распределился следующим образом:

160, 160, 167, 170, 173, 176, 178, 178, 181, 181.

---

[1]) Таким образом, при проверке *нулевой гипотезы* H возможны следующие ситуации:

1) справедлива гипотеза H и критерий допускает H;
2) справедлива гипотеза H, но критерий отвергает H;
3) справедлива гипотеза $\overline{H}$ и критерий отвергает H;
4) справедлива гипотеза $\overline{H}$, но критерий допускает H.

Только в случаях 1 и 3 проверка гипотезы приводит к правильному результату. Вероятность *ошибки первого рода* (случай 2) численно равна уровню значимости α, задаваемому при проверке гипотезы. Если вероятность *ошибки второго рода* (случай 4) равна β, то $1 - \beta$ называют *мощностью критерия*. Часто мощность критерия удается увеличить лишь за счет увеличения α. Иными словами, возможен компромисс между уровнем значимости и мощностью критерия, причем иногда большая мощность оказывается существеннее малого α. В примере, приведенном на фиг. 14, ситуация оказывается именно такой. Действительно, пусть u — выборочная статистика. Тогда, если $u > u_\alpha$, мы отвергаем H и принимаем $\overline{H}$, и наоборот, если $u < u_\alpha$. Площадь под распределением H вправо от u_α равна уровню значимости α, т. е. вероятности ошибки первого рода. Площадь под распределением $\overline{H}$ влево от u_α равна вероятности ошибки второго рода β, а вправо от u_α — мощности критерия. Таким образом, чем больше α, тем больше $1 - \beta$. Подробнее см., например, [45, 53, 57, 58], а также дополнение II. — *Прим. ред.*

Нулевая гипотеза состоит в предположении, что величина роста x распределена по нормальному закону со средним значением 167, и дисперсией σ^2. Конкурирующая гипотеза $\overline{H}$ заключается в допущении, что $x \sim N(\mu, \sigma^2)$, где $\mu > 167$.

Выборочное среднее

$$\bar{x} = 172{,}4.$$

Можно ли считать эту величину достаточно большой, чтобы оправдать принятие гипотезы $\overline{H}$, или тот факт, что $\bar{x} > 167$, обусловлен, просто игрой случая?

Введем величину

$$z = \frac{(\bar{x} - \mu)\sqrt{n}}{\sigma}, \tag{95}$$

распределенную по закону $N(0, 1)$.

Если бы дисперсия σ^2 была известна, можно было бы воспользоваться таблицами нормального распределения вероятности и проверить, является ли z значимо большим нуля при условии, что $\mu = 167$. Но так как мы не знаем величины σ^2, то ее надо оценить с помощью выборочной дисперсии s^2. Найдем

$$s^2 = \sum_{1}^{10} \frac{(x_i - 172{,}4)^2}{9} = 62{,}9,$$

т. е.

$$s = 7{,}93.$$

Оценка среднеквадратичного отклонения для величины $\bar{x}$:

$$\frac{s}{\sqrt{n}} = \sqrt{\frac{62{,}9}{10}} = 2{,}51.$$

По аналогии с z [формула (95)] введем величину

$$t = \frac{(\bar{x} - \mu)\sqrt{n}}{s}. \tag{96}$$

Величина t служит критерием проверки, и нам необходимо вычислить ее распределение в предположении нулевой гипотезы.

Если переписать выражение (96) в виде

$$t=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\Big/\sqrt{\frac{s^2}{\sigma^2}}=\frac{N}{\sqrt{U}}\,, \tag{97}$$

то становится очевидным, что числитель N распределен по нормальному закону $N(0, 1)$, а знаменатель U имеет распределение $\chi^2_{n-1}/(n-1)$. Таким образом, t есть функция двух случайных величин N и U, чьи распределения известны, так что распределение величины t можно вычислить по известным правилам.

Запишем двумерную плотность распределения по N и U:

$$\varphi(N, U)=\frac{e^{-N^2/2}}{\sqrt{2\pi}}\cdot\frac{U^{\frac{n-1}{2}-1}\cdot e^{-U/2}}{\left(\frac{n-1}{2}-1\right)!\,2^{(n-1)/2}}$$

при $-\infty\leqslant N\leqslant\infty$ и $0\leqslant U\leqslant\infty$.

Сначала преобразуем ее, положив

$$h_1(N, U)=t=\frac{N}{\sqrt{U}}\,,$$

$$h_2(N, U)=U,$$

а затем вычислим плотность распределения $g(t, U)$. Интегрируя по U, получаем

$$f(t)=\int_0^\infty g(t, U)\,dU=\frac{C}{\left[1+\frac{t^2}{n-1}\right]^{n/2}}\,, \tag{98}$$

где t имеет $(n-1)$ степеней свободы, поскольку исходная величина χ^2_{n-1} имела $(n-1)$ степеней свободы.

В выражении (98) постоянная C равна

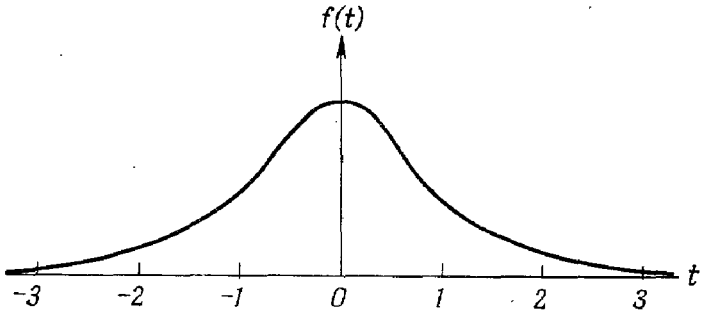
$$C=\frac{\left(\frac{n-2}{2}\right)!}{\sqrt{\pi(n-1)}\cdot\left(\frac{n-3}{2}\right)!}\,. \tag{99}$$

При $n=2$ распределение (98) переходит в распределение Коши (33). В более общем случае у t_ν-распределения с ν степенями свободы существуют только такие начальные моменты μ_r, для которых $r<\nu$,

В нашем примере

$$t_9 = \frac{172{,}4 - 167{,}0}{2{,}51} = 2{,}15.$$

Если принять нулевую гипотезу, то оказывается, что с вероятностью 0,95 величина $t_9 \leqslant 1{,}83$. Таким образом, наблюдаемое значение 2,15 является *значимым при 5%-ном уровне значимости*. На фиг. 15 изображена соответствующая плотность распределения $f(t)$.



Фиг. 15.

Итак, можно сделать вывод, что значение 167 вряд ли может служить надежной оценкой среднего роста группы людей.

Различие между двумя выборочными средними

Покажем теперь, как можно свести к предыдущему случаю несколько более сложную задачу. Пусть дана выборка из n_1 значений нормально распределенной величины x и n_2 значений нормально распределенной величины y. Предположим, что

$$x \sim N(\mu_1, \sigma^2) \quad \text{и} \quad y \sim N(\mu_2, \sigma^2),$$

где величина σ^2 неизвестна, и мы хотим проверить гипотезу

$$\mu_1 = \mu_2.$$

Предположение относительно нормальности распределения указанных величин не входит в проверяемую гипотезу. Выборочные средние $\bar{x}$ и $\bar{y}$ имеют соответствующие

дисперсии: σ^2/n_1 и σ^2/n_2. Если бы величина σ^2 была известна, то мы могли бы вычислить

$$w = \frac{\bar{x} - \bar{y}}{\sigma(\bar{x} - \bar{y})} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \tag{100}$$

и проверить, будет ли нормально распределенная величина w значимо отлична от нуля.

Когда σ^2 неизвестна, мы воспользуемся объединенной оценкой σ^2, полученной из обеих выборок:

$$s^2 = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x})^2 + \sum_{i=1}^{n_2} (y_i - \bar{y})^2}{n_1 + n_2 - 2}. \tag{101}$$

Вместо w мы используем *статистику* t с $\nu = n_1 + n_2 - 2$ степенями свободы:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} = \frac{\bar{x} - \bar{y}}{\hat{\sigma}(\bar{x} - \bar{y})}. \tag{102}$$

Использование распределения t в (102) необходимо обосновать. Числитель этого выражения представляет собой разность двух нормально распределенных величин и, следовательно, тоже описывается нормальным распределением. Утверждение о том, что $\mu_1 = \mu_2$, эквивалентно $\mu_1 - \mu_2 = 0$, и тогда, согласно нулевой гипотезе, среднее значение числителя равно нулю. Необходимо также показать, что объединенная оценка s^2 следует распределению χ^2, которое не зависит от $(\bar{x} - \bar{y})$. Действительно, $\bar{x}$ не зависит от $\sum (x_i - \bar{x})^2$, а $\bar{y}$ не зависит от $\sum (y_i - \bar{y})^2$; значит, s^2 не зависит от $(\bar{x} - \bar{y})$. Кроме того,

$$\sum_1^{n_1} (x_i - \bar{x})^2 \sim \sigma^2 \chi^2_{n_1 - 1}$$

и

$$\sum_1^{n_2} (y_i - \bar{y})^2 \sim \sigma^2 \chi^2_{n_2 - 1},$$

т. е. по существу надо доказать:

$$\chi^2_{n_1-1} + \chi^2_{n_2-1} \sim \chi^2_{n_1+n_2-2}.$$

Производящие функции моментов для $\chi^2_{n_1-1}$ и $\chi^2_{n_2-1}$ соответственно равны $(1 - 2s)^{-(n_1-1)/2}$ и $(1 - 2s)^{-(n_2-1)/2}$ (см. стр. 68). ПФМ суммы двух независимых случайных величин получается в результате перемножения соответствующих ПФМ слагаемых [формула (47)], следовательно,

$$\text{ПФМ}\,(\chi^2_{n_1-1} + \chi^2_{n_2-1}) = (1-2s)^{-(n_1+n_2-2)/2}.$$

Стоящее справа выражение есть ПФМ величины χ^2 с $(n_1 + n_2 - 2)$ степенями свободы.

Пример. Требуется подвергнуть испытанию 13 образцов пряжи, чтобы выяснить, не изменяется ли при намокании ее способность к вытягиванию. Шесть произвольно отобранных образцов были проверены на растяжение, для чего подвешивался груз заданной величины. Их относительные удлинения (y) оказались равны (в %)

$$y = 12{,}3,\ 13{,}7,\ 10{,}4,\ 11{,}4,\ 14{,}9,\ 12{,}6.$$

Остальные семь образцов были тщательно намочены, после чего их относительные удлинения (x) при такой же проверке с тем же грузом дали следующие результаты:

$$x = 15{,}7,\ 10{,}3,\ 12{,}6,\ 14{,}5,\ 12{,}6,\ 13{,}8,\ 11{,}9.$$

При этом $\bar{x} = 13{,}06$, $\bar{y} = 12{,}55$, так что $\bar{x} - \bar{y} = 0{,}51$.

Случайно ли полученное различие? Из приведенных данных следует

$$\sum_1^7 (x_i - 13{,}06)^2 = 18{,}98;\quad \sum_1^6 (y_i - 12{,}55)^2 = 12{,}86.$$

Отсюда

$$s^2 = \frac{12{,}86 + 18{,}98}{11} = 2{,}89,$$

и оценка $\mathscr{D}(\bar{x} - \bar{y})$ равна

$$2{,}89\left(\frac{1}{7} + \frac{1}{6}\right) = 0{,}895,$$

что дает

$$\hat{\sigma}(\bar{x}-\bar{y})=0{,}95.$$

Следовательно,

$$t_{11}=\frac{0{,}51}{0{,}95}=0{,}54.$$

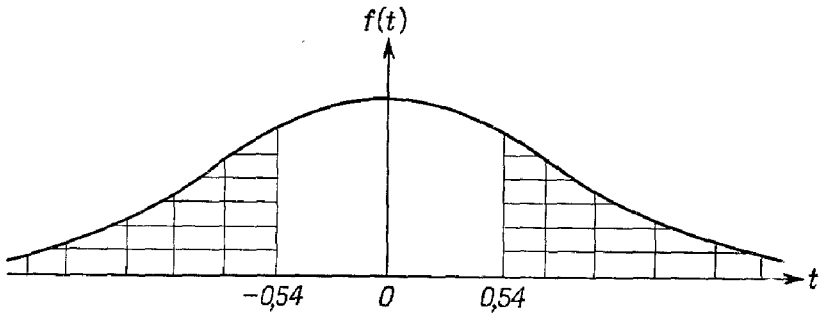
Используя таблицы распределения t[1]), находим, что при $t=0{,}54$ и $\nu=11$

$$P(t)=\int_{t}^{\infty} f(t)\,dt=0{,}30.$$

Поскольку необходимо выяснить, не приводит ли намокание как к увеличению, так и к уменьшению длины пряжи, следует применить *двухсторонний критерий*:

$$P(t)=\int_{-\infty}^{-t} f(t)\,dt+\int_{t}^{\infty} f(t)\,dt=0{,}60. \qquad (103)$$

В необходимости двухстороннего критерия можно убедиться, заметив, что с самого начала лишь из соображений



Фиг. 16.

удобства одна выборка была обозначена через x, а другая через y. Если поменять местами эти обозначения, то получилось бы $\bar{x}-\bar{y}=-0{,}51$. Поэтому в данной задаче большие отрицательные значения следовало бы считать так же значимыми, как и большие положительные (фиг. 16).

[1]) См., например, [32] (см. также табл. V, стр. 260.— *Прим. ред.*).

Соотношение (103) показывает, что даже если между двумя сравниваемыми партиями пряжи нет различия, мы будем в среднем получать величину t, равную или большую 0,54, примерно в шести из каждых десяти испытаний. Следовательно, нет оснований считать необычным полученный в данном случае результат. Таким образом, обнаруженная разница *незначима*.

Используемое в статистике выражение «сильно значимый» надо понимать как «оказавшийся за пределами разумных сомнений». Но то, что «оказалось», не следует автоматически рассматривать как нечто действительно важное или полезное.

Представим себе аналогичный, но больший по объему эксперимент. Пусть объем выборки возрос в 25 раз, и при этом обнаружена такая же разница $(\bar{x} - \bar{y})$. Из выражения (102) видно, что новое значение t будет равно

$$t = 0{,}54\sqrt{25} = 2{,}7.$$

Двухсторонний критерий указывает, что эта величина значима при 2%-ном уровне значимости. Имеется лишь один шанс из 50 наблюдать подобное или большее значение t, если на самом деле обе партии пряжи одинаковы. В этом случае мы предпочитаем говорить, что различие, вероятно, существует, и рассматривать $(\bar{x} - \bar{y}) = 0{,}51$ в качестве наилучшей оценки величины $(\mu_1 - \mu_2)$.

Однако изготовитель пряжи еще мог бы рекламировать ее как не имеющую «заметной усадки» длины после намокания. Даже если во втором эксперименте обнаружится значимо отличная от нуля разница 0,51, он не сочтет, что это имеет существенное значение. В конце концов всегда можно считать себя введенным в заблуждение игрой случая, приведшей к такому значению t, которое встречается в одном случае из пятидесяти в подобной выборке большого объема при $\mu_1 - \mu_2 = 0$.

Можно также представить, какими были бы наши выводы, если бы в первом опыте различие оказалось равным 5,1, а не 0,51. В этом случае $t = 5{,}4$ и оказывается *сильно значимым*. Такой результат уже не только значим, но и важен. Изготовитель теперь должен был бы обрабатывать всю пряжу прежде, чем пускать ее в продажу.

Доверительный интервал

Вернемся к исходной задаче и покажем, как можно представить полученный результат, используя понятие доверительного интервала.

Из таблицы распределения t с числом степеней свободы $\nu = 11$ получим

$$\int_{2,2}^{\infty} f(t)\,dt = 0{,}025,$$

так что

$$\int_{-\infty}^{-2,2} f(t)\,dt + \int_{2,2}^{\infty} f(t)\,dt = 0{,}05,$$

т. е. случайная величина t_{11} с вероятностью 0,95 заключена в интервале от $-2{,}2$ до $+2{,}2$. Допустим, что действительное различие между μ_1 и μ_2 есть некоторая постоянная

$$\delta = \mu_1 - \mu_2. \tag{104}$$

Тогда

$$\frac{(\bar{x}-\bar{y})-\delta}{\sqrt{\frac{s^2}{n_1}+\frac{s^2}{n_2}}} = t_{11} \tag{105}$$

и

$$-2{,}2 \leqslant \frac{(\bar{x}-\bar{y})-\delta}{\sqrt{\frac{s^2}{n_1}+\frac{s^2}{n_2}}} \leqslant 2{,}2 \tag{106}$$

с вероятностью 0,95. В выражении (106) случайными величинами являются, конечно, не δ, а $\bar{x}$, $\bar{y}$ и s^2.

Перепишем формулу (106) в виде

$$(\bar{x}-\bar{y}) - 2{,}2s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \leqslant \delta \leqslant (\bar{x}-\bar{y}) + 2{,}2s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \tag{107}$$

с вероятностью 0,95. Этим вновь подчеркивается, что именно крайние точки интервала искомых значений представляют собой случайные величины. Подставив найден-

ные значения, получим

$$0{,}51 - 2{,}08 = -1{,}57 \leqslant \delta \leqslant 2{,}59 = 0{,}51 + 2{,}08. \quad (108)$$

Строго говоря, смысл неравенств (108) заключается в том, что интервал

$$(-1{,}57;\ 2{,}59) \quad (109)$$

с вероятностью 0,95 содержит истинное значение δ. Поскольку этот интервал включает нуль, мы говорим, что найденное значение $(\bar{x} - \bar{y})$ с вероятностью 0,95 незначимо отлично от нуля. Интервал (109) называется *95%-ным доверительным интервалом* для величины δ (или интервалом, соответствующим доверительной вероятности 0,95).

Как уже говорилось, физики обычно предпочитают приводить результат $\pm$ одно среднеквадратичное отклонение; в соответствии с этим

$$\delta = (\bar{x} - \bar{y}) \pm \hat{\sigma}(\bar{x} - \bar{y}) = 0{,}51 \pm 0{,}95.$$

По существу такая запись содержит указание доверительного интервала

$$-0{,}44 \leqslant \delta \leqslant 1{,}46. \quad (109а)$$

Из таблиц распределения t следует, что доверительный интервал полушириной в одно среднеквадратичное отклонение содержит истинный результат с вероятностью всего 0,66.

Утверждение (109а) может оказаться несправедливым с вероятностью 0,34, в то время как утверждение (108) несправедливо лишь с вероятностью 0,05.

В рассмотренном выше гипотетическом втором эксперименте 95%-ный доверительный интервал равен

$$0{,}51 - \frac{2{,}08}{5} = 0{,}09 \leqslant \delta \leqslant 0{,}93 = 0{,}51 + \frac{2{,}08}{5}. \quad (110)$$

Вероятность того, что истинное значение δ заключено в максимально возможном не содержащем нуль интервале $0 < \delta < 1{,}02$, равна 0,98.

§ 12. Анализ сделанных предположений

1. Специальный характер распределений. Наиболее очевидное предположение, использованное выше, заключается в том, что случайные величины считались распределенными по нормальному закону. Позже будет показано, как можно провести грубую проверку нормальности распределения, используя специальную вероятностную бумагу.

2. Равенство дисперсий. Менее очевидно сделанное на стр. 83 предположение о том, что два нормальных распределения имеют одну и ту же дисперсию σ^2. Ниже будет показано, как построить критерий проверки для отношения двух дисперсий

$$\mathcal{F} = \frac{\sigma_1^2}{\sigma_2^2}.$$

Мы могли бы углубиться в анализ, обнаружив при этом проблемы и методологического характера.

3. Рандомизация. Предположим, что в шляпу опущено 13 номеров, которые соответствуют 13 различным образцам пряжи. Шляпа хорошенько встряхивается, и из нее вынимают шесть номеров, чтобы решить, какие из образцов должны первыми подвергнуться проверке. Могло случиться, что среди оставшихся семи образцов оказалось несколько таких, которые по своей природе склонны к заметному растяжению независимо от того, намачивали их или нет. Иными словами, сам факт проведения рандомизации мог бы способствовать решению вопроса благодаря пристрастному выбору для намачивания семи образцов. Можно ли с научной точки зрения допустить, что процедура перемешивания номеров в шляпе влияет на исход эксперимента? Ответ заключается в том, что: а) отсутствие рандомизации не гарантирует от возможной ошибки из-за наличия семи «растягивающихся» образцов пряжи; б) лучше всего, чтобы ошибка могла проявиться равновероятным образом в обеих выборках.

4. Априорное знание. Выше мы действовали так, как если бы величина $\delta = \mu_1 - \mu_2$ была абсолютно неизвест-

на. Однако экспериментатор мог бы сказать, что маловероятно, чтобы, например, $|\delta| > 100$; таким образом, уже до начала эксперимента мы обладаем некоторой априорной информацией относительно величины δ. Хотя мы и не использовали это априорное знание, кто-то мог бы предложить метод анализа, в котором такое знание используется; это позволило бы на основании тех же экспериментальных данных сделать более точное заключение.

В идеальном случае необходима процедура вычисления, которая строго была бы математически эквивалентна следующему утверждению:

(Априорное знание) + (Результаты эксперимента) $\longrightarrow$ $\longrightarrow$ (Апостериорное знание).

Этой цели служит теорема Бейеса, изложенная на стр. 138.

5. **Смысл** $\mathscr{P}\{|t| > |t_0|\}$. Пусть получено значение $t = 0{,}54$. Пытаясь выяснить значимость такого результата, мы вычисляем вероятность получить $|t| = |t_0| = 0{,}54$ плюс вероятность получить $|t| > 0{,}54$. Почему мы учитываем вероятность второго события, которое не происходило?

6. **Проблема закона научной индукции.** Существует и эта фундаментальная проблема. Любой вывод, основанный на статистическом анализе, зависит от эксперимента, проведенного в прошлом. Насколько можно быть уверенным в том, что любой из открытых «законов» природы останется незыблемым в будущем?

Очевидно, что подобная проблема выходит за рамки содержания книги.

Первые два из перечисленных вопросов будут кратко рассмотрены ниже [1]). Вопрос 4 разбирается в гл. 4, посвященной принципу максимального правдоподобия.

---

[1]) Вопросу критериев проверки нормальности распределения посвящен § 5.5 книги Диксона и Месси [9]. Общее обсуждение проблемы дано в первой части гл. 10 книги Арлея и Буха [1]. В гл. 4 книги Кроу и др. [7] рассмотрен ряд наиболее современных критериев проверки.

Использование вероятностной бумаги для проверки распределения внутри выборки

На фиг. 5 (стр. 31) была изображена S-образная кривая накопленной нормальной вероятности. Функцию $F(x)$ можно интерпретировать следующим образом.

Выберем значение x; *доля* значений случайной величины $X \leqslant x$ в очень большой (практически бесконечной) выборке равна $F(x)$. В случае выборки конечного объема n положение различных точек на кривой $F(x)$ можно оценить, вычисляя доли выборки, которые содержат величины, не превышающие x. Получается *упорядоченная выборка*:

$$-\infty < x_{(1)} < x_{(2)} < \ldots < x_{(n-1)} < x_{(n)} < \infty. \quad (111)$$

Функция распределения накопленной вероятности имеет вид

$$F(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \int_{-\infty}^{x} e^{-(t-\mu)^2/2\sigma^2}\, dt.$$

Если имеется $(n+1)$ интервалов

$$(-\infty,\ x_{(1)}),\ (x_{(1)},\ x_{(2)}),\ \ldots,\ (x_{(n-1)},\ x_{(n)}),\ (x_{(n)},\ \infty), \quad (112)$$

то обычно говорят, что на каждый интервал приходится $1/(n+1)$ часть выборки. Таким образом, на долю значений $\leqslant x_{(1)}$ приходится $1/(n+1)$ часть выборки, т. е. $1/(n+1)$ есть *оценка* $F(x_{(1)})$; доля выборки, в которой $X \leqslant x_{(2)}$, равна $2/(n+1)$; . . .; доля выборки, в которой $X \leqslant x_{(n)}$, равна $n/(n+1)$. Таким образом, $(n+1)/(n+1)$ есть точное значение $F(\infty)$. Поэтому условимся оценивать кривую $F(x)$ точками

$$[x;\ F(x)] = \left[x_{(i)};\ \frac{i}{n+1}\right],\ i = 1,\ 2,\ \ldots,\ n. \quad (113)$$

Строго говоря, надо было бы написать

$$[x(F);\ F] = \left[x_{(i)};\ \frac{i}{n+1}\right],$$

показывая тем самым, что $F = i/(n+1)$ заранее фиксировано и $x_{(i)}$ служит оценкой для $x(F) = x[i/(n+1)]$. Доказательство того, что $x_{(i)}$ является хорошей оценкой,

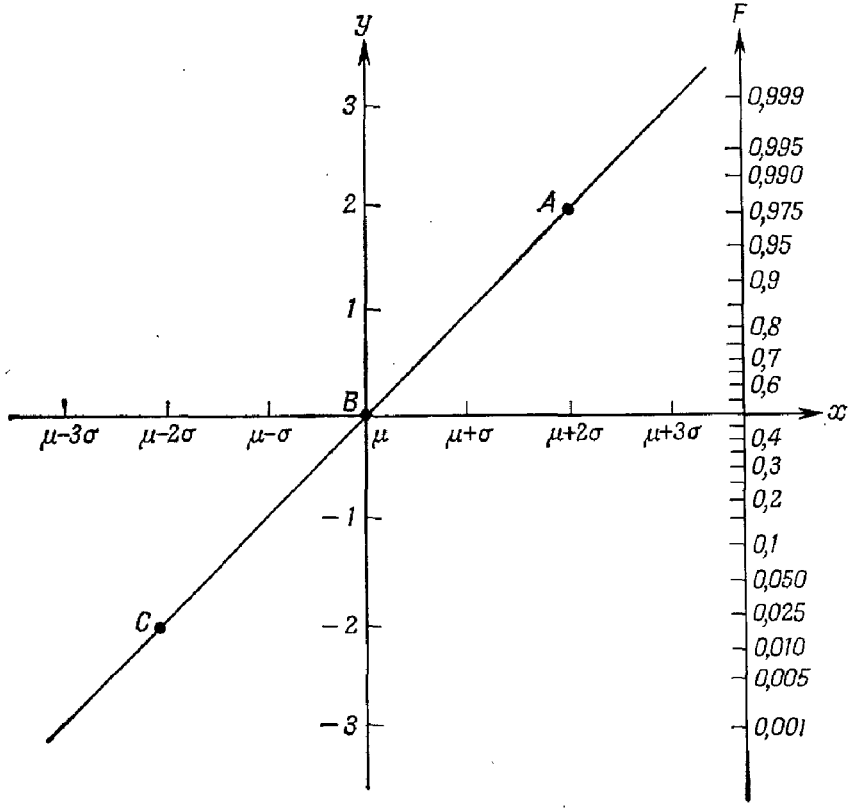
довольно сложно и основано на том, что

$$\mathscr{E}[F(x_{(i)})] = \frac{i}{n+1} \qquad (114)$$

для всех случайных величин, имеющих непрерывное распределение. Соотношение (114) *не зависит от характера распределения*, т. е. является *непараметрическим*.

Линеаризация кривой

Сравнение небольшого числа точек на графике с теоретической S-образной кривой весьма затруднительно.



Фиг. 17.

Поэтому кривую $F(x)$ линеаризуют, преобразуя масштаб по оси ординат (F). После этого значительно проще сравнивать отдельные точки с прямой линией.

Преобразуем шкалу F в интервале от 0 до 1 в новую шкалу по y в интервале от $-\infty$ до $+\infty$; для этого запишем

$$F = \int_{-\infty}^{y(F)} \frac{1}{\sqrt{2\pi}} e^{-s^2/2}\, ds. \qquad (115)$$

Это уравнение следует решить относительно y (F) при заданном F. Допустим, что x подчиняется нормальному распределению со средним значением μ и среднеквадратичным отклонением σ; тогда уравнение кривой $F(x)$ в новой шкале будет иметь вид

$$y = \frac{x-\mu}{\sigma} \qquad (116)$$

(фиг. 17). На фиг. 17 видно, как расположились три точки A, B и C после преобразования кривой (см. фиг. 5). Обычно на такой графической бумаге указывается только преобразованная шкала F, так что соответствующие значения y откладываются автоматически вместо того, чтобы решать уравнение (115) для каждого F.

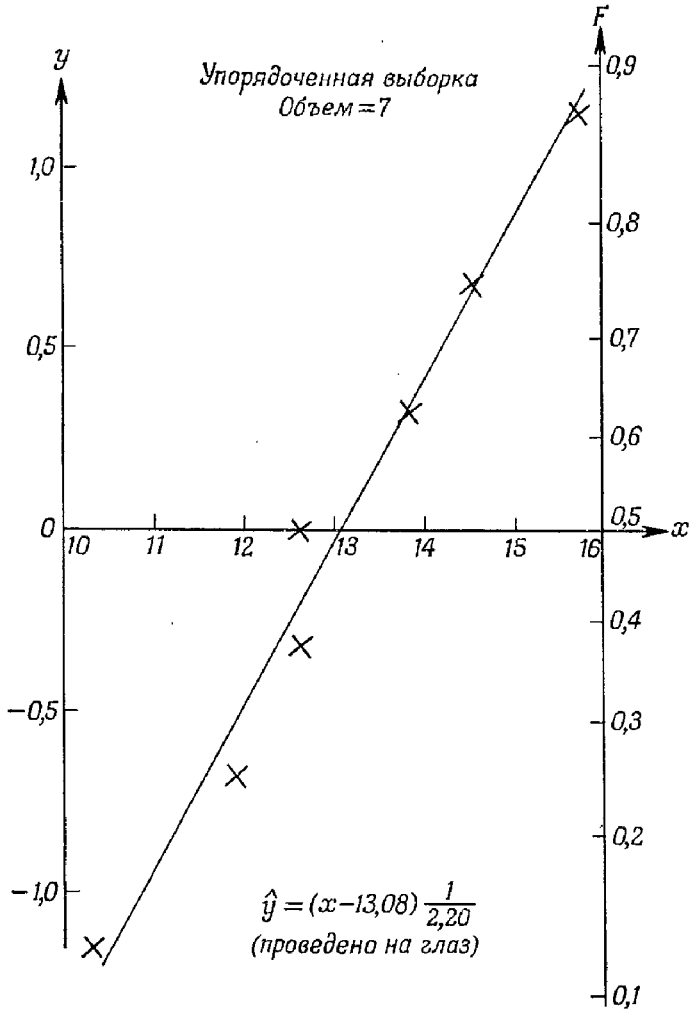
Пример. Мы можем расположить результаты случайной выборки (см. стр. 85) в возрастающем порядке и иметь дело с *порядковой статистикой*. Соответствующие значения F и y приведены в табл. 5.

Таблица 5

i	$x_{(i)}$	F	y
1	10,3	0,125	—1,150
2	11,9	0,250	—0,674
3	12,6	0,375	—0,319
4	12,6	0,500	0
5	13,8	0,625	0,319
6	14,5	0,750	0,674
7	15,7	0,875	1,150

Значения y получены с помощью стандартных статистических таблиц [1]). На фиг. 18 порядковая статистика

[1]) См., например, [32] (см. также табл. II и III стр. 256 и 257.— *Прим. ред.*).



Фиг. 18.

изображена на вероятностной бумаге для нормального распределения. Линейный характер расположения точек на этом графике свидетельствует о том, что выборка распределена приблизительно по нормальному закону. Проведя на глаз прямую линию через точки, получаем следующие оценки:

$$\mu_2 = 13{,}08, \qquad \sigma_2 = 2{,}20.$$

Эти значения можно сравнить с оценками, полученными на стр. 85:

$$\hat{\mu} = 13{,}06, \quad \hat{\sigma} = \sqrt{\frac{18{,}98}{6}} = 1{,}78.$$

§ 13. Распределение $\mathcal{F}$ Фишера

Для проверки, не являются ли значимо различными средние значения двух нормальных выборок, можно использовать критерий t. Когда необходимо проверить, не являются ли значимо различными *дисперсии* двух нормальных выборок, применяется критерий $\mathcal{F}$. Допустим, что s_1^2 представляет собой оценку дисперсии σ_1^2; более точно

$$s_1^2 \sim \sigma_1^2 \frac{\chi_{\nu_1}^2}{\nu_1}. \tag{117}$$

Сравним эту величину с *независимой* оценкой s_2^2 дисперсии σ_2^2, причем

$$s_2^2 \sim \sigma_2^2 \frac{\chi_{\nu_2}^2}{\nu_2}. \tag{118}$$

Проверяемая гипотеза состоит в том, что

$$\sigma_1^2 = \sigma_2^2 = \sigma^2.$$

Согласно этой гипотезе, и s_1^2 и s_2^2 представляют собой оценки дисперсии σ^2.

При использовании распределения t для проверки различия между двумя *выборочными средними* целесообразно полагать, что t пропорционально разности этих величин

$$t = k(\bar{x}_1 - \bar{x}_2).$$

Таким образом, статистика, на основе которой строится критерий проверки, не зависит от неизвестного параметра μ, характеризующего положение выборки.

При сравнении двух *выборочных дисперсий* удобно подвергнуть проверке их отношение.

Введем величину

$$\mathcal{F}_{\nu_1, \nu_2} = \frac{s_1^2}{s_2^2}. \tag{119}$$

Чтобы найти распределение $\mathcal{F}$, можно воспользоваться тем фактом, что

$$\frac{s_1^2}{s_2^2} \sim \frac{\sigma^2\chi^2_{\nu_1}/\nu_1}{\sigma^2\chi^2_{\nu_2}/\nu_2} = \frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2} \tag{120}$$

не зависит от неизвестного параметра σ^2. Поскольку обе величины χ^2 в выражении (120) делятся на соответствующие числа степеней свободы, то и числитель и знаменатель оказываются порядка единицы. Следовательно, плотность распределения $\mathcal{F}$ будет группироваться в окрестности единицы.

Нам нужно найти плотность вероятности функции двух независимых случайных величин $\chi^2_{\nu_1}$ и $\chi^2_{\nu_2}$. Дело в сущности сводится к решению простой задачи на применение правил вычисления плотности вероятности функции случайных величин (см. стр. 69). Результат дается формулой (127), промежуточные выкладки могут быть при первом чтении опущены.

Введем для удобства обозначения

$$V = \chi^2_{\nu_1} \text{ и } W = \chi^2_{\nu_2}$$

и, кроме того, обозначим $m = \nu_1$, $n = \nu_2$. Совместная плотность вероятности двух переменных V и W равна

$$p(V, W) = \frac{1}{\left(\frac{m}{2}-1\right)!}\frac{1}{2^{m/2}}V^{\frac{m}{2}-1}e^{-V/2} \times$$

$$\times \frac{1}{\left(\frac{n}{2}-1\right)!}\frac{1}{2^{n/2}}W^{\frac{n}{2}-1}e^{-W/2}. \tag{121}$$

Введем новую переменную

$$\mathcal{F} = \frac{n}{m}\frac{V}{W}. \tag{122}$$

Для проведения вычислений указанным методом необходимо, чтобы число новых переменных равнялось числу старых. Поэтому, кроме $\mathcal{F}$, введем еще

$$V \equiv V.$$

Теперь необходимо выразить в новых переменных старые. Поскольку $V = V$ и

$$W = \frac{n}{m}\frac{V}{\mathcal{F}}, \tag{123}$$

то якобиан перехода будет равен

$$\frac{\partial(V, W)}{\partial(\mathcal{F}, V)} = \begin{vmatrix} \frac{\partial V}{\partial \mathcal{F}} & \frac{\partial V}{\partial V} \\ \frac{\partial W}{\partial \mathcal{F}} & \frac{\partial W}{\partial V} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ -\frac{n}{m}\frac{V}{\mathcal{F}^2} & \frac{n}{m}\frac{1}{\mathcal{F}} \end{vmatrix} = \frac{n}{m}\frac{V}{\mathcal{F}^2}. \tag{124}$$

Подставим в соотношение (121) вместо V и W их выражения через $\mathcal{F}$ и V (123) и полученный результат умножим на абсолютное значение якобиана (124). Мы получим плотность вероятности в переменных $\mathcal{F}$ и V

$$q(\mathcal{F}, V) = p[V(\mathcal{F}, V),\ W(\mathcal{F}, V)]\left|\frac{\partial(V, W)}{\partial(\mathcal{F}, V)}\right| =$$

$$= \frac{\left(\frac{n}{m}\right)^{n/2}\cdot 2^{-\frac{n+m}{2}}}{\left(\frac{m}{2}-1\right)!\left(\frac{n}{2}-1\right)!}\frac{V^{\frac{m+n}{2}-1}}{\mathcal{F}^{\frac{n}{2}+1}}e^{-\frac{V}{2}-\frac{nV}{2m\mathcal{F}}}. \tag{125}$$

Проинтегрировав это выражение по V, получим *маргинальное* (безусловное) распределение величины $\mathcal{F}$. Итак,

$$r(\mathcal{F}) = \int_0^\infty q(\mathcal{F}, V)\,dV. \tag{126}$$

Основная часть интегрирования сводится к вычислению [см. (41)]

$$\int_0^\infty V^\alpha e^{-V/\beta}\,dV = \alpha!\,\beta^{\alpha+1},$$

где

$$\alpha = \frac{m+n}{2} - 1, \quad \beta = \frac{2m\mathcal{F}}{m\mathcal{F}+n}.$$

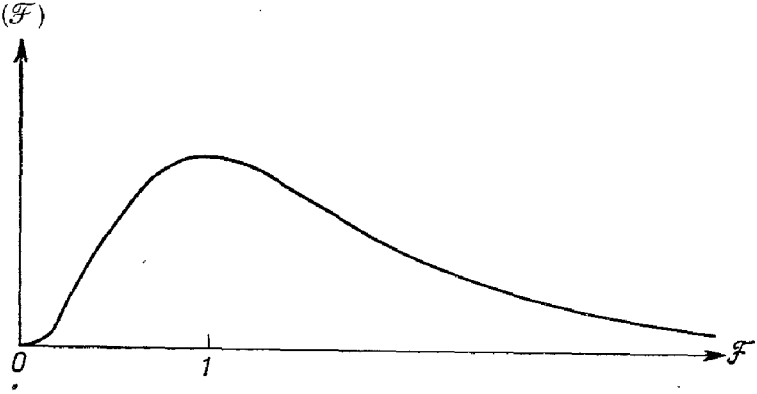
В итоге

$$r(\mathcal{F}) = \frac{C\mathcal{F}^{\frac{m}{2}-1}}{(m\mathcal{F}+n)^{\frac{m+n}{2}}}, \tag{127}$$

причем

$$C = \frac{\Gamma\left(\frac{m}{2}+\frac{n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} n^{n/2} m^{m/2}. \qquad (128)$$

На фиг. 19 изображена функция $r(\mathcal{F})$. По своему виду она аналогична распределению χ^2_m при малых значениях



Фиг. 19.

$V = \chi^2_m$. Действительно, как следует из формулы (86), плотность вероятности

$$s(V) = \frac{C' V^{\frac{m}{2}-1}}{e^{V/2}}.$$

Однако у распределения $\mathcal{F}$ заметен более длинный «хвост»[1]). Это означает, что величины, соответствующие достаточно высокому уровню вероятности, будут располагаться значительно правее моды распределения. Так, при $m = 12$ и $n = 6$ мода соответствует значению $\mathcal{F} = 0{,}625$, в то время как 5%-ная граница соответствует значению $\mathcal{F} = 4{,}0$. Иными словами, для значимого различия при 5%-ном уровне значимости s_1^2 в 4 раза должно превышать s_2^2. Такое различие может показаться чрезмерно большим. Однако столь большое значение $\mathcal{F}$ вполне разумно, если вспомнить, что $s_1^2/s_2^2 > 4$ означает всего лишь $s_1/s_2 > 2$.

[1]) Распределение $\mathcal{F}$ приведено в табл. VII, стр. 284.— *Прим. ред.*

Пример. В задаче, приведенной на стр. 85, рассматривались две партии пряжи. Были получены следующие значения выборочных дисперсий для величин растяжения пряжи под нагрузкой:

$$s_x^2 = \frac{18{,}98}{6} = 3{,}16$$

$$s_y^2 = \frac{12{,}86}{5} = 2{,}57.$$

Значимо ли различие этих величин?

Находим отношение $\mathcal{F}$:

$$\mathcal{F}_{6,5} = \frac{s_x^2}{s_y^2} = \frac{3{,}16}{2{,}57} = 1{,}23.$$

Вероятности 0,95 соответствует значение $\mathcal{F}_{6,5}$, равное 4,95; таким образом, значение 1,23 незначимо отлично от единицы.

Применение критерия $\mathcal{F}$ к решению задачи о проведении кривой по точкам

Важнейшее применение критерия $\mathcal{F}$ связано с рассмотрением отношения двух независимых оценок дисперсии σ^2 из *одной и той же* выборки.

Пусть необходимо произвести оценку величины α, причем можно построить сумму квадратов

$$s_1^2 = k\alpha^2 + \sigma^2 \frac{V}{m},$$

где V следует распределению χ_m^2. Если величина W дает независимую оценку σ^2, так что $s_2^2 = \sigma^2 W/n$ (причем W следует распределению χ_n^2), то можно подвергнуть проверке отношение

$$\frac{s_1^2}{s_2^2} = \frac{\sigma^2 \dfrac{V}{m} + k\alpha^2}{\sigma^2 \dfrac{W}{n}} \sim \frac{\chi_m^2/m + k\dfrac{\alpha^2}{\sigma^2}}{\chi_n^2/n}. \qquad (129)$$

Нетрудно видеть, что отношение (129) имеет распределение $\mathcal{F}$ в том и только в том случае, когда $\alpha = 0$. С дру-

гой стороны, если $\alpha \neq 0$, это отношение будет принимать значения, много большие единицы, при условии, что k достаточно велико. Детальный анализ показывает, что k растет по мере увеличения объема эксперимента, так что более обширный эксперимент обеспечивает более мощный критерий проверки нулевой гипотезы $\alpha = 0$.

Выражение (129) указывает на общую структуру критериев, объединенных общим наименованием *дисперсионный анализ* [1]). Это наименование может несколько вводить в заблуждение, поскольку чаще всего интерес представляет не дисперсия σ^2, а параметр α. Пусть, например, необходимо аппроксимировать экспериментальные точки некоторой кривой, и мы используем для аппроксимации следующую форму:

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_r x^r + \varepsilon. \qquad (130)$$

Здесь ε — *ошибка* с дисперсией σ^2, а значение r нам неизвестно. Мощным методом решений этой задачи является критерий $\mathcal{F}$, с помощью которого проверяется значимость коэффициента α_r. Именно в такой постановке критерий $\mathcal{F}$ будет нами использован в дальнейшем при изложении метода наименьших квадратов [2]).

---

[1]) В тех случаях, когда $\alpha \neq 0$, критерий проверки может быть сформулирован на основе нецентральных распределений χ^2 и $\mathcal{F}$ (см. [53, 58], а также дополнение II). Использование подобных распределений может оказаться целесообразным, например, при сопоставлении результатов эксперимента с предсказаниями теории.— *Прим. ред.*

[2]) См. об этом примечание на стр. 168.— *Прим. ред.*

ГЛАВА ЧЕТВЕРТАЯ

Принцип максимального правдоподобия[1])

§ 14. Функция правдоподобия

Оценка параметров составляет основную цель всех научных исследований, и часто подходящим средством для получения таких оценок оказывается принцип (или метод) *максимального правдоподобия*. Дальнейшее применение этого метода позволяет найти приближенные ошибки полученных оценок. Можно показать, что в большинстве случаев эти ошибки оказываются меньше, нежели при любых других методах оценки параметров.

Подход к оценкам, основанным на методе максимального правдоподобия, мы изложим в три этапа. На первом, проиллюстрированном двумя примерами, рассматриваются дискретная случайная величина r и неизвестный параметр p, принимающий конечное число возможных значений. На втором этапе величина r остается по-прежнему дискретной, а параметр p будет непрерывным и в рассмотренном примере заключен в интервале $0 \leqslant p \leqslant 1$. И наконец, в третьем случае непрерывными будут как случайная величина x, так и неизвестный параметр μ, причем $-\infty < \mu < \infty$.

Биномиальное распределение

Пусть имеются две урны, содержащие каждая черные и белые шары. В урне A на каждый белый шар приходится три черных, в урне B на каждый черный шар приходится три белых. Сначала произвольно выбирается одна из урн, а затем из нее вынимаются (с последующим возвращением) три шара. Вероятность того, что среди этих

---

[1]) См. также [27], гл. 8.

шаров окажется r черных, равна [см. формулу (8)]

$$\mathscr{P}(r) = C_3^r p^r (1-p)^{3-r}, \qquad (131)$$

где $p = {}^3/_4$ для урны A и $p = {}^1/_4$ для урны B.

Задача состоит в оценке параметра p на основании произведенной выборки.

В табл. 6 приведены значения $\mathscr{P}(r, p)$ для всех возможных r и двух значений p, вычисленные по формуле (131).

Таблица 6

r	0	1	2	3
$\mathscr{P}(r, {}^3/_4)$	1/64	9/64	27/64	27/64
$\mathscr{P}(r, {}^1/_4)$	27/64	27/64	9/64	1/64

Предположим, что среди трех шаров черного не оказалось, т. е. $r = 0$. В этом случае мы делаем заключение, что полученный результат должен был наблюдаться в 27 раз с большей вероятностью при $p = {}^1/_4$, нежели при $p = {}^3/_4$. В общем можно утверждать, что наиболее

Таблица 7

r	0	1	2	3
$\mathscr{P}(r, {}^1/_{10})$	0,729	0,243	0,027	0,001
$\mathscr{P}(r, {}^2/_{10})$	0,512	0,384	0,096	0,008
$\mathscr{P}(r, {}^3/_{10})$	0,343	0,441	0,189	0,027
$\mathscr{P}(r, {}^4/_{10})$	0,216	0,432	0,288	0,064
$\mathscr{P}(r, {}^5/_{10})$	0,125	0,375	0,375	0,125
$\mathscr{P}(r, {}^6/_{10})$	0,064	0,288	0,432	0,216
$\mathscr{P}(r, {}^7/_{10})$	0,027	0,189	0,441	0,343
$\mathscr{P}(r, {}^8/_{10})$	0,008	0,096	0,384	0,512
$\mathscr{P}(r, {}^9/_{10})$	0,001	0,027	0,243	0,729

правдоподобной оценкой параметра $\hat{p}$ будет $^1/_4$, если $r = 0$ или 1, и $^3/_4$, если $r = 2$ или 3.

Мы можем обобщить эту задачу на случай, когда возможные значения p равны 0,1; 0,2; . . .; 0,9, предположив, что произвольно выбирается одна из девяти урн, в каждой из которых находится необходимая смесь белых и черных шаров.

В табл. 7 содержатся значения биномиальных вероятностей, относящиеся к указанному примеру [29].

Как и в предыдущем случае, для каждого возможного значения r существует набор значений $^1/_{10}$, $^2/_{10}$, . . ., $^9/_{10}$ рассматриваемого параметра p. Мы выбираем значение $\hat{p}$, которое с наибольшей вероятностью должно приводить к появлению данного результата.

В табл. 8 приведены оценки $\hat{p}$, основанные на принципе максимального правдоподобия.

Таблица 8

r	0	1	2	3
$\hat{p}$	$^1/_{10}$	$^3/_{10}$	$^7/_{10}$	$^9/_{10}$

Непрерывный параметр

Вообще говоря, можно допустить, что в урне имеется совершенно неизвестная смесь черных и белых шаров, так что p может принимать любое значение от 0 до 1.

Функция

$$l(p) = l(p \mid r) = C_3^r p^r (1-p)^{3-r} \qquad (132)$$

при известном r и неизвестном p называется *функцией правдоподобия* для параметра p. Например, при $r = 1$

$$l(p \mid 1) = 3p(1-p)^2. \qquad (133)$$

Чтобы найти величину p, которая с наибольшей вероятностью приводит к наблюдаемому значению $r = 1$, необходимо определить, при каком p выражение (133) достигает максимума. Когда параметр непрерывен, максимум

функции правдоподобия можно найти, как обычно, посредством дифференцирования. При этом

$$\frac{dl(p)}{dp} = 3(1-p)^2 - 6p(1-p) = 3(3p^2 - 4p + 1) = 0,$$

откуда следует

$$\hat{p} = \frac{1}{3}.$$

Заметим, что множитель 3 в выражении (133) оказывается общим для всех p, поэтому в функции правдоподобия его можно опустить и, как правило, вместо $C_n^r p^r (1-p)^{n-r}$ пишут

$$l(p \mid r) = p^r (1-p)^{n-r}. \tag{134}$$

В общем случае биномиального параметра выражение (134) достигает максимума при

$$\hat{p} = \frac{r}{n} \tag{135}$$

[см. также (27)].

Широко используется также *логарифмическая функция правдоподобия*, представляющая собой иную запись (134):

$$L(p \mid r) = \ln l(p \mid r) = r \ln p + (n - r) \ln (1 - p). \tag{136}$$

Максимум функции $L(p \mid r)$ достигается при том же значении $\hat{p}$, что и максимум функции $l(p \mid r)$. Следует отметить, что отбрасывание постоянного множителя у функции $l(p \mid r)$ эквивалентно добавлению к $L(p \mid r)$ (или вычитанию) постоянной.

Параметры непрерывного распределения

Допустим, что случайная величина x имеет непрерывное распределение с плотностью $f(x \mid \theta)$, причем θ — единственный параметр. Тогда распределение вероятности для выборки объема n будет иметь вид

$$d\mathcal{F}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i \mid \theta)\, dx_i. \tag{137}$$

Если величины x_i известны, а θ неизвестна, то функция правдоподобия для θ имеет вид

$$l(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i \mid \theta) \tag{138}$$

(множитель $\prod dx_i$ опущен, так как он не зависит от θ); логарифмическая функция правдоподобия для θ:

$$L(\theta \mid \mathbf{x}) = \sum_{i=1}^{n} \ln f(x_i \mid \theta).$$

Пример. Пусть случайная величина распределена по нормальному закону с единичной дисперсией и неизвестным средним значением, т. е.

$$f(x \mid \mu) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2}.$$

Тогда

$$l(\mu \mid \mathbf{x}) = e^{-\Sigma(x_i-\mu)^2/2} \tag{139}$$

[множитель $(2\pi)^{-n/2}$ опущен], а

$$L(\mu \mid \mathbf{x}) = -\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2}.$$

Функция $L(\mu)$ представляет собой квадратичную параболу по переменной μ для любого n. Эта функция достигает максимума при таком значении μ, для которого минимальна сумма

$$\sum_{i=1}^{n} (x_i-\mu)^2.$$

Мы получаем результат, хорошо известный из «метода наименьших квадратов»:

$$\hat{\mu} = \sum_{i=1}^{n} \frac{x_i}{n} = \bar{x}.$$

Случай неизвестной дисперсии. Если в рассмотренном выше примере дисперсия распределения неизвестна, то воз-

никает двухпараметрическая задача. В этом случае

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Следовательно,

$$l(\mu, \sigma^2 \mid \mathbf{x}) = (\sigma^2)^{-n/2} e^{-\frac{\Sigma(x_i-\mu)^2}{2\sigma^2}}, \qquad (140)$$

а

$$L(\mu, \sigma^2 \mid \mathbf{x}) = -\frac{n}{2} \ln \sigma^2 - \sum \frac{(x_i-\mu)^2}{2\sigma^2}. \qquad (140')$$

Необходимо найти пару значений $(\hat{\mu}, \hat{\sigma}^2)$, которые с наибольшей вероятностью приводили бы к наблюдаемым результатам. Для этого нужно решить совместно два уравнения:

$$\frac{\partial L}{\partial \mu} = 0 \text{ и } \frac{\partial L}{\partial \sigma^2} = 0.$$

Согласно (140'),

$$\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2} \sum (x_i - \mu) = 0;$$

следовательно, $\hat{\mu} = \bar{x}$.

Далее из формулы (140')

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum (x_i - \mu)^2 = 0;$$

следовательно,

$$\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2. \qquad (141)$$

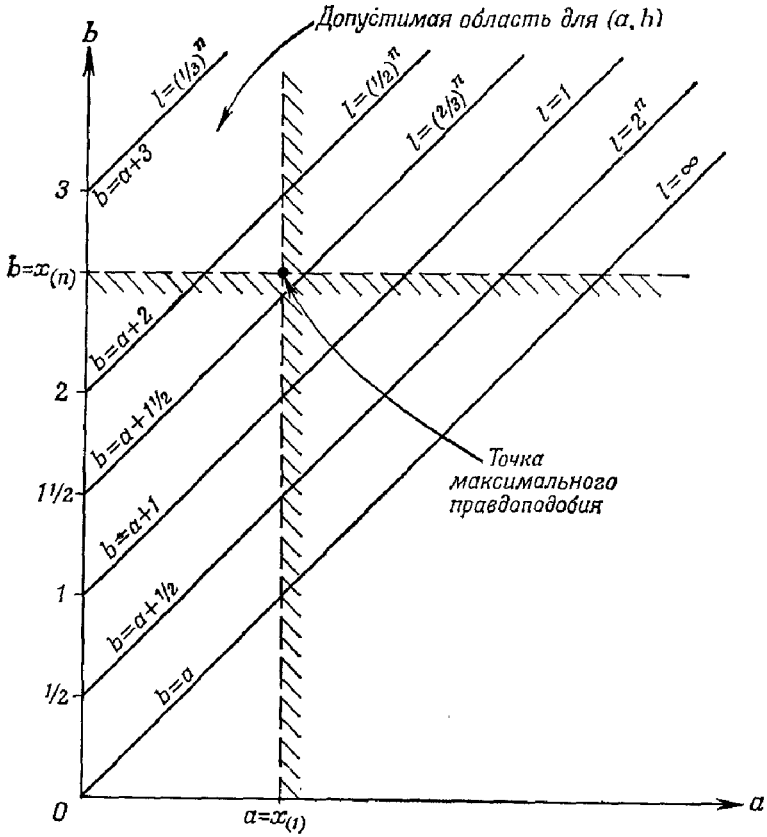
Как видно, $\hat{\sigma}^2$ оказывается несколько смещенной оценкой дисперсии σ^2, поскольку

$$\mathscr{E}(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2 \qquad (142)$$

[ср. формулу (89)].

Равномерное распределение. В этом случае

$$f(x \mid a, b) = \frac{1}{b-a}, \qquad a \leqslant x \leqslant b.$$



Ф и г. 20.

Упорядочим выборку, расположив ее элементы в порядке возрастания:

$$x_{(1)} < x_{(2)} < \ldots < x_{(n-1)} < x_{(n)},$$

так что $a \leqslant x_{(1)}$, $b \geqslant x_{(n)}$. Функция правдоподобия двух параметров a и b изображена на фиг. 20.

Очевидно, что

$$l(a, b \mid \mathbf{x}) = \frac{1}{(b-a)^n}, \qquad a \leqslant x_{(1)}, \quad b \geqslant x_{(n)}.$$

Этот пример выбран для того, чтобы показать, что максимум функции правдоподобия не всегда удается найти

дифференцированием, даже если параметры непрерывны в некоторой области. На фиг. 20 указаны контуры функции правдоподобия. Эта функция достигает максимума при наименьшем из возможных значений ее знаменателя, т. е. когда

$$b = x_{(n)} = \hat{b},$$

$$a = x_{(1)} = \hat{a}.$$

Оставшаяся часть упорядоченной выборки $\{x_i\}$

$$x_2 < x_3 < \ldots < x_{(n-1)}$$

уже не добавляет информации к оценкам a и b.

Поскольку вся информация, которую несет выборка относительно оценок параметров a и b, заключена в элементах $x_{(1)}$ и $x_{(n)}$, последние называют *достаточными* для a и b. Если получение выборки преследовало лишь единственную цель оценить a и b, то все промежуточные элементы не представляют интереса и могут быть отброшены.

Достаточные статистики [1])

В примере с нормальным распределением мы видели что $\hat{\mu}$ и $\hat{\sigma}^2$ являются функциями $\sum x_i$ и $\sum x_i^2$. Нельзя ли извлечь какую-либо полезную информацию и из других функций от исходных величин, таких, например, как $\sum |x_i|$, или $\sum x_i^3$, или $(x_{(n)} - x_{(1)})$? Мы увидим, что относительно μ и σ вообще нельзя получить никаких дополнительных сведений, и фактически, коль скоро известны $\sum x_i$ и $\sum x_i^2$, то исходная выборка $\{x_i\}$ может быть «уничтожена». Мы называем $\sum x_i$ и $\sum x_i^2$ *достаточными статистиками*, а любые оценки параметров, зависящих от $\sum x_i$ и $\sum x_i^2$, представляют собой *достаточные оценки.* (Тем не менее для решения вопроса о том, *какими* функциями от $\sum x_i$ и $\sum x_i^2$ нужно пользоваться, могут оказаться необходимы иные критерии.)

[1]) При первом чтении этот раздел можно опустить.

Пусть имеется выборка $x_1, \ldots, x_n$ из генеральной совокупности с плотностью вероятности $f(x \mid \theta)$. В качестве оценки параметра используется функция этих данных

$$\hat{\theta} = \hat{\theta}(\mathbf{x}).$$

Совместная плотность распределения выборки:

$$\prod_{i=1}^{n} f(x_i \mid \theta).$$

Предположим, что эту функцию можно разложить на множители

$$\prod_{i=1}^{n} f(x_i \mid \theta) = g(\mathbf{x} \mid \hat{\theta})\, h(\hat{\theta} \mid \theta), \tag{143}$$

где $g(\mathbf{x} \mid \hat{\theta})$ — функция только от $\mathbf{x}$ и *не зависит* от θ. Из формулы (143) следует, что $g(\mathbf{x} \mid \hat{\theta})$ есть условное распределение $\mathbf{x}$ при заданном значении $\hat{\theta}(\mathbf{x})$. Поскольку параметр θ уже не входит в это распределение, функцию $g(\mathbf{x} \mid \hat{\theta})$ нельзя использовать для получения информации относительно θ. Аналогично множитель $h(\hat{\theta} \mid \theta)$ зависит от $\mathbf{x}$ только через посредство $\hat{\theta}(\mathbf{x})$, так что в $\hat{\theta}$ должна содержаться вся информация, касающаяся оценки θ.

Пример. Пусть

$$f(x \mid \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

есть плотность распределения с единственным неизвестным параметром μ. Плотность распределения для выборки имеет вид

$$\prod_{i=1}^{n} f(x_i \mid \mu) = (2\pi)^{-n/2} e^{-\frac{\sum (x_i-\mu)^2}{2}} =$$

$$= (2\pi)^{-n/2} e^{-\sum x_i^2/2}\, e^{-\frac{n\mu^2 - 2\mu \sum x_i}{2}}. \tag{144}$$

Из сопоставления формул (144) и (143) следует, что $\sum_i x$ является *достаточной статистикой* для μ, и поэтому

$$\bar{x} = \frac{\sum x_i}{n}$$

представляет собой *достаточную оценку*. Величина $\sum x_i^2$ в формуле (144) не имеет отношения к оценке μ.

Не в каждой задаче оценки параметров имеется достаточная статистика, но когда у параметра существует достаточная оценка, метод максимального правдоподобия дает именно эту оценку.

§ 15. Графический анализ функции правдоподобия

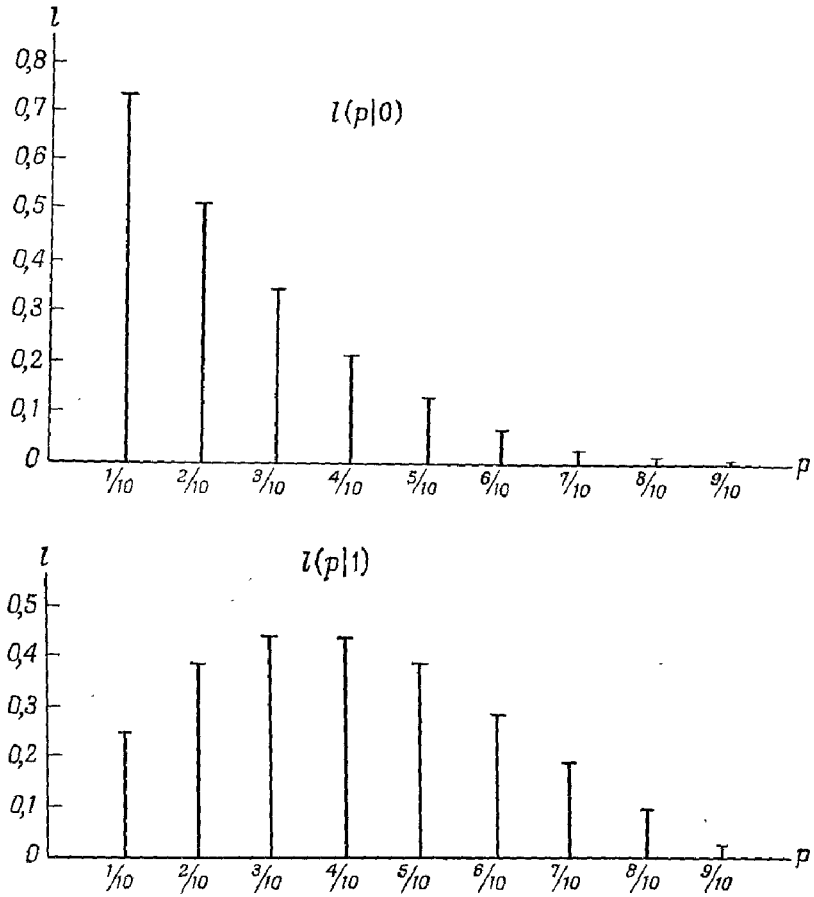
Одно из достоинств использования функции правдоподобия при анализе экспериментальных данных состоит в том, что часто удается найти графический вид этой функции и тем самым наглядно представить полученный результат. Этот метод неудобен, если имеется три или большее число параметров, оценки которых не независимы.

Допустим, что имеется один неизвестный параметр θ. Изобразив графически функцию $l(\theta)$ или функцию $L(\theta) = \ln l(\theta)$, можно сразу же найти значение $\theta = \hat{\theta}$, которому соответствует «наивысшая» точка кривой. Это избавляет нас от трудоемких численных расчетов в тех случаях, когда соответствующее максимуму значение $\hat{\theta}$ не может быть найдено аналитически и приходится прибегать к итерациям. Кроме того, используя график, можно тотчас указать доверительный интервал или ошибку полученной оценки, проведя на графике соответствующую линию. Все это будет видно из дальнейшего.

В том случае, когда имеется два параметра θ и φ, функция правдоподобия представляет собой поверхность, контуры которой можно изобразить на плоскости (θ, φ), как это было сделано на фиг. 20. При этом вновь без труда можно найти точку максимального правдоподобия. Обычно, когда максимум расположен внутри допустимой области, контурами равного правдоподобия являются замкнутые кривые (см. фиг. 27, стр. 135). В этом случае двумерным доверительным интервалом служит область, охватываемая контуром.

Случай дискретного параметра

В примере, рассмотренном на стр. 103, был найден набор функций правдоподобия биномиального параметра p для всевозможных результатов эксперимента. Функ-



Фиг. 21.

ции правдоподобия для $r = 0$ и $r = 1$ изображены графически на фиг. 21. Из фиг. 21 (вверху) видно, что $\hat{p} = {}^1/_{10}$. Кроме того, мы могли бы определить, насколько значение ${}^1/_{10}$ более вероятно, нежели ближайшее к нему значение ${}^2/_{10}$. Поскольку подобный вопрос возникает часто,

целесообразно нанести функцию $L(p\,|\,0) = \ln l(p\,|\,0)$, дающую ответ непосредственно. Такая функция особенно полезна в случае непрерывного параметра. В данном примере значение $\hat{p} = {}^1/_{10}$ всего в 1,5 раза более вероятно, чем $\hat{p} = {}^2/_{10}$, так что эксперимент малого объема не позволяет получить достаточно надежного разделения. В логарифмическом масштабе мы получили бы

$$\ln l\left(\frac{1}{10}\right) - \ln l\left(\frac{2}{10}\right) \approx \ln 1{,}5.$$

В случае, изображенном на фиг. 21 (внизу), $\hat{p} = {}^3/_{10}$, но почти столь же вероятно и $\hat{p} = {}^4/_{10}$. Действительно, набор значений ${}^2/_{10}$, ${}^3/_{10}$, ${}^4/_{10}$ и ${}^5/_{10}$ можно грубо рассматривать как доверительный интервал, который с большой вероятностью содержит истинное значение p. В следующем разделе будет дано более точное выражение этого утверждения.

Случай непрерывного параметра

Пусть имеется нормальное распределение с единичной дисперсией и неизвестным параметром μ. Функция правдоподобия μ, полученная из выборки объема n, имеет вид

$$l(\mu\,|\,\mathbf{x}) = e^{-\frac{\sum(x_i-\mu)^2}{2}}.$$

Она представляет собой такую же колоколообразную кривую, что и плотность вероятности с центром в точке $\mu = \bar{x}$. В данном случае удобнее иметь дело с логарифмической функцией правдоподобия

$$L(\mu\,|\,\mathbf{x}) = -\frac{1}{2}\sum(x_i-\mu)^2. \qquad (145)$$

К выражению (145) можно добавить любую произвольную постоянную, не содержащую μ. Если в качестве такой постоянной взять

$$\frac{1}{2}\left[\sum x_i^2 - n(\bar{x})^2\right],$$

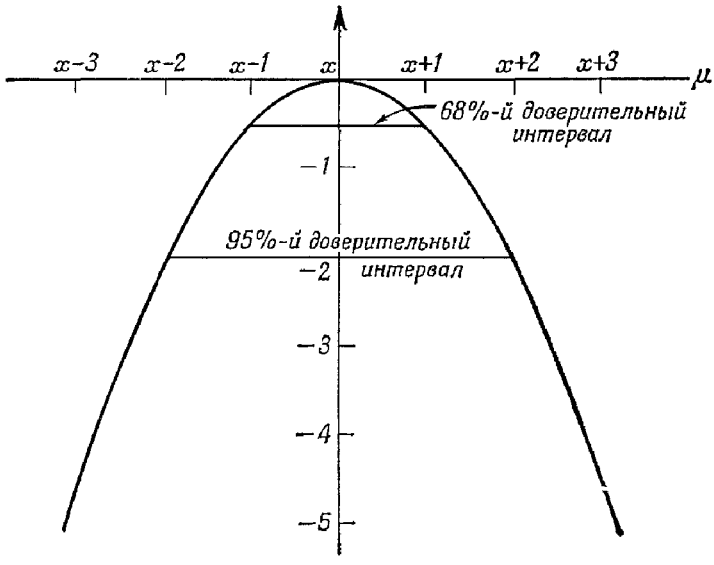
то мы получим функцию

$$L(\mu\,|\,\mathbf{x}) = -\frac{n}{2}(\mu-\bar{x})^2, \qquad (146)$$

представляющую собой параболу, ширина которой растет пропорционально $\sqrt{n}$, а максимум, равный нулю, расположен при $\mu = \bar{x}$. Если $n = 1$, то

$$L(\mu \mid x) = -\frac{1}{2}(\mu - \bar{x})^2. \tag{147}$$

Эта функция изображена на фиг. 22.



Фиг. 22.

Подобный график лежит в основе анализа большого класса функций правдоподобия, которые зависят от одного параметра и обладают следующими свойствами:

$$\left.\begin{array}{ll} 1) & L(\mu) \text{ — непрерывны,} \\ 2) & L(\mu) \text{ — имеют единственный локальный максимум в интересующей нас области.} \end{array}\right\} \tag{148}$$

Приведенная выше логарифмическая функция правдоподобия имеет максимум при $\mu = \bar{x}$. Отрезок прямой, соединяющий ветви параболы на уровне $L = -1/2$,

характеризует доверительный интервал

$$(x - \sigma) \leqslant \mu \leqslant (x + \sigma), \tag{149}$$

который часто записывают в виде

$$\mu = x \pm \sigma,$$

где σ — среднеквадратичное отклонение.

Согласно свойствам нормального распределения, выражению (149) соответствует доверительная вероятность, равная 68,3%. Аналогично отрезок прямой, проведенной на уровне $L = -2$, соответствует 95%-ному доверительному интервалу (на самом деле 95,5%-ному, так как в случае нормального распределения 95%-ному доверительному интервалу соответствует $L = -1{,}96$, а не $-2{,}0$).

Приведем теперь нестрогое доказательство теоремы, имеющей важное практическое значение. Но сначала докажем лемму.

ЛЕММА. Пусть $g(\mu)$ — любая однозначная функция μ. Тогда функция правдоподобия для $g_0 = g(\mu_0)$ совпадает с функцией правдоподобия для μ_0.

Переформулируем лемму с тем, чтобы провести доказательство менее формально. Функция правдоподобия для μ_0 характеризует возможность «объяснить» μ_0 с помощью имеющихся данных. Если преобразовать неизвестный параметр μ с помощью взаимно однозначного соответствия в параметр g новой шкалы, то величина g_0 будет точно соответствовать величине μ_0 по старой шкале. В этом случае g_0 в той же степени, что и μ_0, позволяет «объяснить» имеющиеся данные и, следовательно, имеет ту же функцию правдоподобия, что и μ_0.

Доказательство. По определению функция правдоподобия для μ_0

$$l_1(\mu_0 \mid x) = \mathcal{P}(x \mid \mu_0).$$

Но $\mathcal{P}(x \mid \mu_0) = \mathcal{P}(x \mid g_0)$, поскольку $\mathcal{P}(x \mid \mu_0)$ полностью определено, если известно μ_0; g_0 можно вычислить, зная μ_0, а $\mathcal{P}(x \mid g_0)$ полностью определено, если известно g_0. По определению

$$l_2(g_0 \mid x) = \mathcal{P}(x \mid g_0);$$

следовательно,

$$l_2(g_0 \mid x) = l_1(\mu_0 \mid x),$$

что и требовалось доказать.

ТЕОРЕМА. Свойство «инвариантности». Наиболее вероятное (в смысле максимального правдоподобия) значение μ преобразуется в наиболее вероятное значение g, т. е. $\hat{g} = g(\hat{\mu})$.

Доказательство. Пусть $g' = g(\hat{\mu})$. Из леммы следует, что $l_2(g') = l_1(\hat{\mu}) \geqslant l_1(\mu)$ для всех μ (по определению $\hat{\mu}$). Но для любого μ существует такое g, что $l_1(\mu) = l_2(g)$. Следовательно, $l_2(g') \geqslant l_2(g)$ для всех μ, а значит, и для всех g.

Таким образом, $g' = \hat{g}$, что и требовалось доказать.

Пример. Пусть $x \sim N(\mu, 1)$, и пусть $g = \mu^3$. Тогда $f_1(x \mid \mu_0) = Ce^{-\frac{(x-\mu_0)^2}{2}}$ и $l_1(\mu_0 \mid x) = Ce^{-\frac{(\mu_0-x)^2}{2}}$. С другой стороны,

$$f_2(x \mid g_0) = Ce^{-\frac{(x-g_0^{1/3})^2}{2}} = f_1(x \mid \mu_0)$$

и

$$l_2(g_0 \mid x) = Ce^{-\frac{(g_0^{1/3}-x)^2}{2}} = l_1(\mu_0 \mid x).$$

Причина того, что лемма справедлива в столь простой форме, заключается в различии между понятиями вероятности и функции правдоподобия. Вероятность $\mathcal{P}(x)$ остается вероятностью x независимо от того, является ли параметром просто μ_0 или $g(\mu_0)$.

Итак, мы указали простой способ нахождения *функции правдоподобия* для функции *параметра* в противоположность более сложной операции нахождения *вероятности* (или плотности вероятности) функции *случайной величины*.

Принцип правдоподобия

Определение 1. Предположим, что в двух экспериментах E_1 и E_2, каждый из которых содержит неизвестный параметр θ_1 и θ_2, получены соответствующие

результаты $\mathbf{x}_1$ и $\mathbf{x}_2$. Если функции правдоподобия в этих случаях оказались одинаковыми, т. е. $l_1(\theta_1 \mid \mathbf{x}_1) = l_2(\theta_2 \mid \mathbf{x}_2)$, то мы логически вынуждены сделать одинаковые выводы относительно θ_1 и θ_2.

Определение 2. Сказанное выше справедливо и в том случае, когда $l_1(\theta_1 \mid \mathbf{x}_1) = kl_2(\theta_2 \mid \mathbf{x}_2)$, если только k не зависит от θ_1 или θ_2.

При всей кажущейся наивности эта аксиома содержит весьма мудрые результаты. В конечном смысле она приводит к выводу, что экспериментатор может заготовить математическую часть анализа эксперимента один раз на все случаи жизни. Результаты всех последующих опытов можно преобразовать так, что они станут эквивалентными первому, для которого ответ уже получен. Тогда все, что останется сделать,— это преобразовать ответ к исходной шкале последнего эксперимента. Как это делать, будет видно из теоремы о доверительных интервалах, получаемых непосредственно из функции правдоподобия.

ТЕОРЕМА. Пусть $L_\theta(\theta \mid \mathbf{x})$ — непрерывная функция θ с единственным максимумом в области допустимых значений θ. Доверительный интервал для θ можно найти с помощью той же процедуры, что и доверительный интервал для среднего значения μ в случае единичной выборки x из распределения $N(\mu, 1)$.

1. Добавим к функции $L_\theta(\theta \mid \mathbf{x})$ константу так, чтобы

$$\max_\theta L_\theta(\theta \mid \mathbf{x}) = L_\theta(\hat{\theta} \mid \mathbf{x}) = 0.$$

2. На графике логарифмической функции правдоподобия L_θ проведем прямую линию на уровне $L = -1/2$, что даст 68%-ный доверительный интервал, либо на уровне $L = -2$, что даст 95%-ный доверительный интервал, и т. д.

Доказательство. Существует преобразование $g(\theta)$ параметра θ, которое превращает кривую $L_\theta(\theta \mid \mathbf{x})$ в параболу

$$L_g(g \mid \mathbf{x}) = -\frac{1}{2}(g - x)^2.$$

Напомним, что в случае, когда среднее значение μ оценивалось по нормальной единичной выборке x, логарифми-

ческая функция правдоподобия имела вид

$$L_{\mu}(\mu \mid x) = -\frac{1}{2}(\mu - x)^2.$$

В соответствии с принципом максимального правдоподобия мы должны относительно величины g сделать тот же вывод, что и для μ. Это означает, что в качестве *точечной оценки* можно выбрать

$$\hat{g} = x,$$

и, кроме того, проведя на графике логарифмической функции правдоподобия прямую линию на уровне $L_g = -2$, определим 95%-ный доверительный интервал для g

$$\hat{g} - 2 \leqslant g \leqslant \hat{g} + 2.$$

(Строго говоря, этот интервал следовало бы называть «правдоподобным интервалом», подчеркивая тем самым, что он получен на основании принципа максимального правдоподобия.)

Получив доверительный интервал для g, необходимо преобразовать его в доверительный интервал для исходного параметра θ. Однако, как следует из леммы, в подобной сложной процедуре обратного преобразования нет необходимости. Пусть прямая линия, проведенная на уровне $L = -2$, пересекает кривую L_g в точках, соответствующих g_0 и g_1, т. е.

$$L_g(g_0) = L_g(g_1) = -2.$$

Нам нужно найти значения θ_0 и θ_1, такие, что

$$g_0 = g(\theta_0) \text{ и } g_1 = g(\theta_1).$$

Согласно лемме,

$$L_{\theta}(\theta_0) = L_g(g_0) = -2$$

и

$$L_{\theta}(\theta_1) = L_g(g_1) = -2.$$

Это означает, что точки θ_0 и θ_1 можно найти по известным значениям ординат кривой L_{θ}; проведя же прямую $L = -2$, мы найдем ее отрезок, заключенный между ветвями кривой L_{θ}, который и представляет собой иско-

мый доверительный интервал. Таким образом, с вероятностью 0,95

$$\theta_0 \leqslant \theta \leqslant \theta_1,$$

что и требовалось доказать. Полученный результат проиллюстрирован на фиг. 23.

Свойства оценок максимального правдоподобия в случаях выборки малого и большого объема [1])

Верхняя граница точности, которая может быть достигнута при решении задачи об оценке параметров, удовлетворяющей некоторым условиям регулярности, устанавливается неравенством Крамера — Рао. Это неравенство не зависит от вида распределения случайных величин x в данном эксперименте. Оно записывается через величину среднего квадрата ошибки оценки $t = t(\mathbf{x})$ параметра θ.

Средний квадрат ошибки определяется следующим образом:

$$\mathscr{E}[(t-\theta)^2] = \mathscr{E}[t-\mathscr{E}(t)]^2 + [\mathscr{E}(t)-\theta]^2.$$

Первое слагаемое представляет собой дисперсию оценки, которая характеризует средний квадрат отклонения случайной величины t от ее среднего значения $\mathscr{E}(t)$. Второе — *смещение* (или *систематическую ошибку*) оценки, характеризующее разность между $\mathscr{E}(t)$ и величиной, оценкой которой является t, а именно θ. Следовательно,

$$\mathscr{E}[(t-\theta)^2] = \sigma_t^2(\theta) + \beta^2(\theta). \tag{150}$$

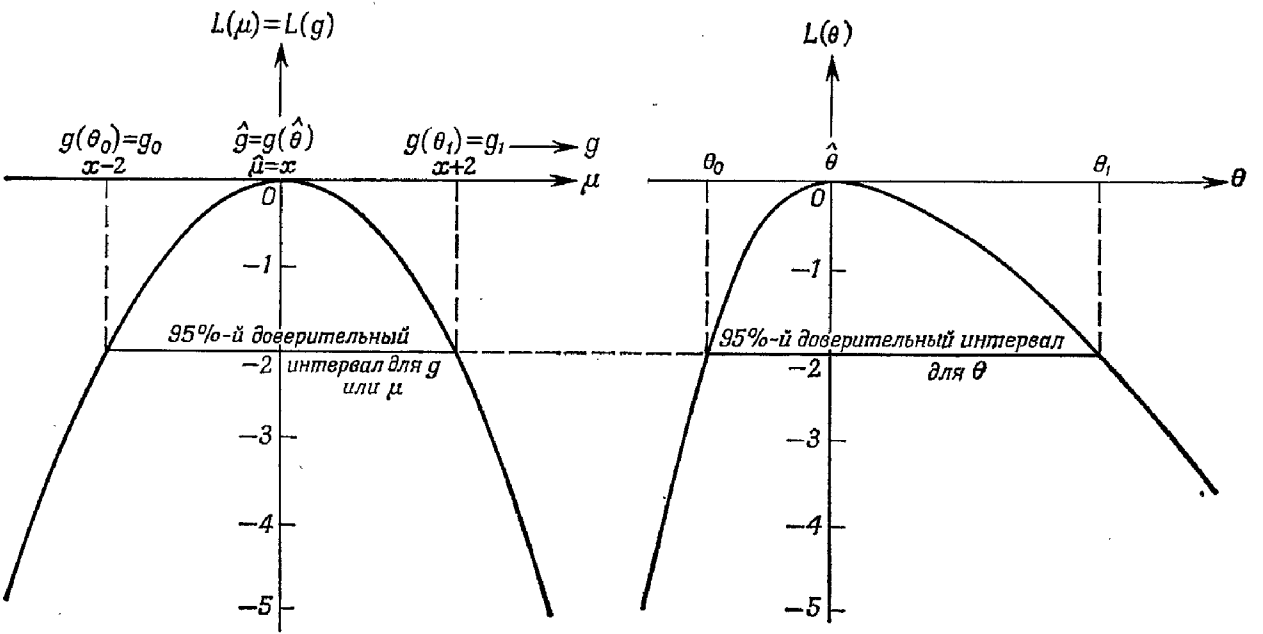
Допустим, что объем выборки равен n. Тогда неравенств Крамера — Рао утверждает, что

$$\mathscr{E}[(t-\theta)^2] \geqslant \frac{\gamma^2(\theta)}{n}, \tag{151}$$

т. е. что точность оценки не превышает

$$\frac{\sqrt{n}}{\gamma(\theta)}, \tag{152}$$

[1]) Приводятся без доказательства.


L(μ)=L(g)
g(θ0)=g0
x−2
ĝ=g(θ̂)
μ̂=x
g(θ1)=g1
x+2
g
μ
0
−1
−2
−3
−4
−5
95%-й доверительный
интервал для g
или μ
L(θ)
θ0
θ̂
θ1
θ
95%-й доверительный интервал
для θ

Фиг. 23.

где

$$\gamma^2(\theta) = \frac{\left[1 + \frac{d\beta(\theta)}{d\theta}\right]^2}{\int \left[\frac{\partial \ln f(x \mid 0)}{\partial 0}\right]^2 f(x \mid \theta)\, dx}. \qquad (153)$$

Оценку t называют *эффективной*, если ее точность в действительности достигает верхней границы (152).

Приведем теперь свойства оценок, основанных на методе максимального правдоподобия.

Первые два свойства справедливы для любого n, однако их называют свойствами «малой выборки», чтобы отличить от остальных свойств, справедливых лишь при большом n.

1. Если существует *достаточная оценка*, то метод максимального правдоподобия дает именно эту оценку и тем самым исчерпывает всю имеющуюся информацию.

2. Если существует *эффективная оценка*, то метод максимального правдоподобия дает именно эту оценку и более точной оценки найти нельзя.

Следующие два свойства справедливы асимптотически при большом n. Иными словами, они строго справедливы в пределе при $n \to \infty$ и оказываются приближенными в случае большого, но конечного n.

3. Оценка максимального правдоподобия является *асимптотически эффективной*.

4. Оценка максимального правдоподобия имеет *асимптотически нормальное распределение* со средним значением θ и дисперсией

$$\sigma_l^2 = \frac{1}{-\mathscr{E}\left[\frac{\partial^2 L}{\partial \theta^2}\right]}, \qquad (154)$$

где L — логарифмическая функция правдоподобия θ. Выражение для σ_l^2 можно в данном случае приближенно записать в виде

$$\sigma_l^2 \approx \frac{1}{-\left[\frac{\partial^2 L}{\partial \theta^2}\right]_{\theta = \hat{\theta}}}. \qquad (155)$$

«Интуитивное» обоснование формулы (155) следует из параболического графика на фиг. 22. Несколько обобщая,

представим себе, что x распределено по закону $N(\mu, \sigma_x^2)$, где σ_x^2 известно, но необязательно равно единице. Тогда

$$L(\mu \mid x) = -\frac{(\mu - x)^2}{2\sigma_x^2},$$

и легко видеть, что

$$\frac{1}{-\left[\frac{\partial^2 L}{\partial \mu^2}\right]_{\mu=x}} = \sigma_x^2 = \mathscr{D}(\hat{\mu}).$$

Иначе говоря, мы утверждаем только то, что форма параболы и получающиеся доверительные интервалы полностью определяются значением второй производной в максимуме. Параболический вид логарифмической функции правдоподобия, конечно, обеспечивается асимптотической нормальностью распределения. Следующее свойство представляет общетеоретический интерес.

5. Оценка максимального правдоподобия *инвариантна* относительно преобразований параметра, как было доказано на стр. 116. Иными словами, если $g = g(\theta)$, то $\hat{g} = g(\hat{\theta})$. Например, если $s^2 = (1/n) \sum (x_i - \bar{x})^2$ — оценка максимального правдоподобия дисперсии σ^2, то оценкой максимального правдоподобия σ будет просто s. Не так обстоит дело с критерием несмещенности, согласно которому несмещенной оценкой σ^2 служит $s_W^2 = [1/(n-1)] \times \times \sum (x_i - \bar{x})^2$, а несмещенной оценкой σ служит $s_V = \{[(n-3)/2]! / \sqrt{2}\,[(n-2)/2]!\} \sqrt{\sum (x_i - \bar{x})^2}$. Поскольку $s_W^2 \neq (s_V)^2$, то этот критерий определенно нарушается.

6. Главная трудность, возникающая при использовании метода максимального правдоподобия в качестве критерия оценки, заключается в том, что при этом приходится предполагать известным вид распределения вероятности. Так, мы считали, что $x \sim N(\mu; 1)$ (μ неизвестно). Это допущение может наложить сильное ограничение. Для определенного класса задач метод наименьших квадратов дает точную оценку при любых распределениях случайной величины, т. е. *не зависящую от распределения.*

7. В добавление к сказанному выше следует отметить, что относительно использования принципа максимального правдоподобия существуют противоречивые мнения.

В последних статьях собраны различные мнения специалистов из разных стран, а также приведены результаты всех обсуждений [2, 35].

Пример использования функции правдоподобия

Пусть $x \sim N(0, \sigma^2)$, где значение σ^2 неизвестно. Вероятность получить в результате выборки величину $\mathbf{x}$ равна

$$f(\mathbf{x} \mid \sigma^2)\, d\mathbf{x} = (2\pi\sigma^2)^{-n/2} e^{\sum x_i^2/2\sigma^2} d\mathbf{x}, \tag{156}$$

так что функция правдоподобия для σ^2 имеет вид

$$l_1(\sigma^2 \mid \mathbf{x}) = (\sigma^2)^{-n/2} e^{-\frac{1}{2}\sum x_i^2/\sigma^2}. \tag{157}$$

Если преобразовать параметр σ^2 в формуле (157) в новый параметр $\sigma = +\sqrt{\sigma^2}$, то, согласно лемме, приведенной на стр. 115, мы получим то же самое численное значение правой части:

$$l_2(\sigma \mid \mathbf{x}) = (\sigma)^{-n} e^{-\frac{1}{2}\sum x_i^2/\sigma^2}, \tag{158}$$

т. е., очевидно, $l_1(\sigma^2 \mid \mathbf{x}) = l_2(\sigma \mid \mathbf{x})$. Из выражения (158) следует, что $\sum x_i^2$ представляет собой достаточную статистику. Предположим, что нам известна только величина

$$w = \sum x_i^2, \tag{159}$$

а исходные данные $x_1, x_2, \ldots, x_n$ утеряны. Найдем плотность вероятности случайных величин $\mathbf{x}$ по заданному значению w. Известно, что $w \sim \sigma^2 \chi_n^2$. На стр. 68 мы уже встречались с плотностью распределения $u \sim \chi_n^2$, а именно

$$\varphi(u) = \frac{1}{\left(\frac{n}{2} - 1\right)!\, 2^{n/2}}\, u^{\frac{n}{2}-1} e^{-u/2}.$$

На основании этого получим плотность распределения преобразованной величины $w = \sigma^2 u$

$$\psi(w \mid \sigma^2) = \frac{1}{\left(\frac{n}{2} - 1\right)!\, 2^{n/2}} \left(\frac{w}{\sigma^2}\right)^{\frac{n}{2}-1} e^{-w/2\sigma^2} \frac{1}{\sigma^2}. \tag{160}$$

Плотность распределения $\mathbf{x}$, которая дается выражением (156)

$$f(\mathbf{x} \mid \sigma^2) = (2\pi\sigma^2)^{-n/2} e^{-w/2\sigma^2},$$

можно разложить на множители:

$$f(\mathbf{x} \mid \sigma^2) = \left[\left(\frac{n}{2} - 1\right)! \, \pi^{-n/2} w^{-\frac{n}{2}+1}\right] [\psi(w \mid \sigma^2)] =$$
$$= g(\mathbf{x} \mid w) \cdot \psi(w \mid \sigma^2), \qquad (161)$$

где $g(\mathbf{x} \mid w)$ не зависит от σ^2 и постоянно при заданном w. Это, кроме того, означает, что если известно w, но неизвестны отдельные $\{x_i\}$, то эти величины будут равномерно распределены по сфере

$$\sum x_i^2 = w.$$

Прежде чем записать функцию правдоподобия для σ^2, используя выражение (161), обратим внимание на то, что первый множитель есть константа, и поэтому можно записать

$$l_3(\sigma^2 \mid w) = \psi(w \mid \sigma^2). \qquad (162)$$

Все выводы из формулы (162) используют данные об $\mathbf{x}$ только через посредство величины w, и, таким образом, оценка максимального правдоподобия оказывается функцией достаточной статистики. Можно еще больше упростить соотношение (162) [или вернуться к первоначальному определению $f(\mathbf{x} \mid \sigma^2)$] и получить

$$l_4(\sigma^2 \mid w) = (\sigma^2)^{-n/2} e^{-w/2\sigma^2}. \qquad (163)$$

Логарифмическую функцию правдоподобия можно записать в виде

$$L(\sigma^2 \mid w) = -\frac{n}{2} \ln \sigma^2 - \frac{w}{2\sigma^2} + \frac{n}{2} \ln \frac{we}{n}, \qquad (164)$$

причем последнее слагаемое добавлено для того, чтобы L в максимуме было равно нулю.

Случайная выборка малого объема

С помощью таблицы случайных чисел произведена выборка объемом 6 из генеральной совокупности с распределением $N(0, \sigma^2)$, где $\sigma^2 = 1$. Получены следующие

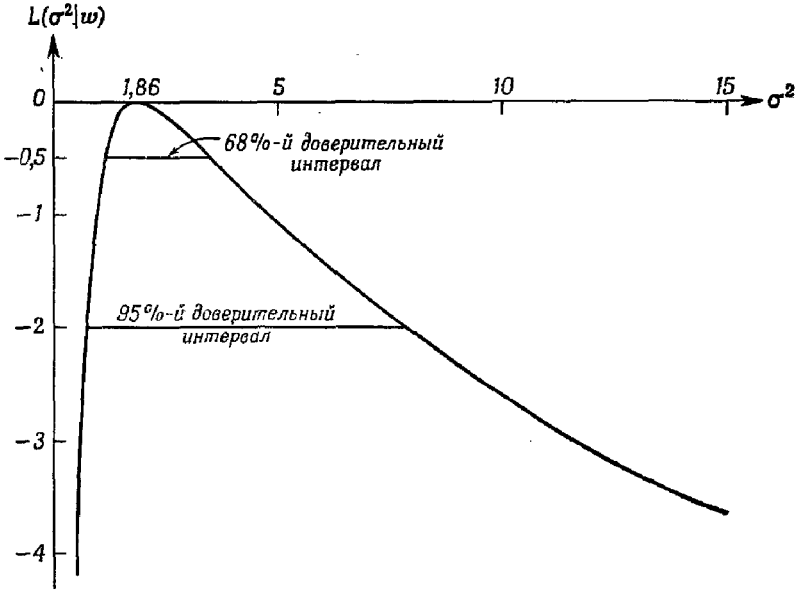
значения x:

$$1{,}468,\ -2{,}097,\ 0{,}146,\ -0{,}295,\ -2{,}120,\ 0{,}017. \qquad (165)$$

По формуле (164) вычислим логарифмическую функцию правдоподобия при условии, что $w = 11{,}55$, $n = 6$:

$$L(\sigma^2 \mid w) = 4{,}860 - 3 \ln \sigma^2 - \frac{5{,}578}{\sigma^2}.$$

В табл. 9 приведены результаты численного расчета $L(\sigma^2 \mid w)$. Оценку максимального правдоподобия σ^2 найдем,



Фиг. 24.

дифференцируя L. В максимуме

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{w}{2\sigma^4} = 0.$$

Отсюда

$$\hat{\sigma}^2 = \frac{w}{n} = \frac{\sum x_i^2}{n} = 1{,}86.$$

На фиг. 24 изображена эта функция. Найдя точечную оценку, можно также определить доверительные интер-

валы, проведя на графике соответствующие линии, параллельные оси абсцисс. В данном случае мы поступаем аналогично определению доверительного интервала $\hat{\mu} = x \pm \sigma$

Таблица 9

σ^2	0	0,5	0,75	1	1,4
$-3 \ln \sigma^2$		2,080	0,863	0	−1,009
$-5{,}578/\sigma^2$	$-\infty$	−11,155	−7,437	−5,578	−3,986
$+4{,}860$		4,860	4,860	4,860	4,860
$L(\sigma^2/w)$	$-\infty$	−4,215	−1,714	−0,718	−0,135

σ^2	1,86	3	5	10	15	∞
$-3 \ln \sigma^2$		−3,296	−4,828	−6,908	−8,124	$-\infty$
$-5{,}578/\sigma^2$		−1,859	−1,116	−0,558	−0,372	
$+4{,}860$		4,860	4,860	4,860	4,860	
$L(\sigma^2/w)$	0	−0,295	−1,084	−2,606	−3,636	$-\infty$

в задаче для $N(\mu \mid 1)$, выбрав точки пересечения кривой $L(\sigma^2 \mid w)$ с прямой линией $L = -1/2$. Это дает

$$1{,}10 \leqslant \sigma^2 \leqslant 3{,}49 \quad (\text{ср. } \hat{\sigma}^2 = 1{,}86).$$

Мы видим, что подобный «узкий» доверительный интервал не содержит известного истинного значения $\sigma^2 = 1$. Можно ожидать, что при использовании узкого доверительного интервала подобное обстоятельство будет встречаться в 32 случаях из 100.

Интервалу $\hat{\mu} = x \pm 2\sigma$ в задаче для $N(\mu \mid 1)$ будут эквивалентны точки пересечения кривой $L(\sigma^2 \mid w)$ с прямой линией $L = -2$; в этом случае

$$0{,}71 \leqslant \sigma^2 \leqslant 7{,}8 \quad (\text{ср. } \hat{\sigma}^2 = 1{,}86).$$

Эффективность оценки $\hat{\sigma}^2$

Воспользовавшись формулой (153), покажем, что $\hat{\sigma}^2$ обладает 100%-ной эффективностью. Действительно,

$$\hat{\sigma}^2 \sim \sigma^2 \frac{\chi_n^2}{n}.$$

Поскольку

$$\mathscr{E}(\chi_n^2) = n \quad \text{и} \quad \mathscr{D}(\chi_n^2) = 2n,$$

то

$$\mathscr{E}(\hat{\sigma}^2) = \sigma^2 \quad \text{и} \quad \mathscr{D}(\hat{\sigma}^2) = \frac{2\sigma^4}{n}. \tag{166}$$

По определению смещение $\beta(\hat{\sigma}^2) = \mathscr{E}(\hat{\sigma}^2) - \hat{\sigma}^2 = 0$. Далее вычислим

$$\frac{\partial \ln f(x \mid \sigma^2)}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{x^2}{2\sigma^4}.$$

Следовательно,

$$\int_{-\infty}^{\infty} \left[\frac{\partial \ln f}{\partial \sigma^2}\right]^2 f(x \mid \sigma^2)\, dx =$$

$$= \frac{1}{4\sigma^4 \sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \left(1 - \frac{x^2}{\sigma^2}\right)^2 e^{-x^2/2\sigma^2}\, dx = \frac{1}{2\sigma^4}.$$

Из неравенства Крамера — Рао видно, что наименьшая из возможных дисперсий регулярной оценки в данном случае равна $\mathscr{D}_{\text{мин}} = 2\sigma^4/n$.

Таким образом, оценка $\hat{\sigma}^2$ имеет 100%-ную эффективность, ибо ее дисперсия равна нижнему пределу; это означает, что данная оценка позволяет извлечь максимум информации из имеющихся результатов.

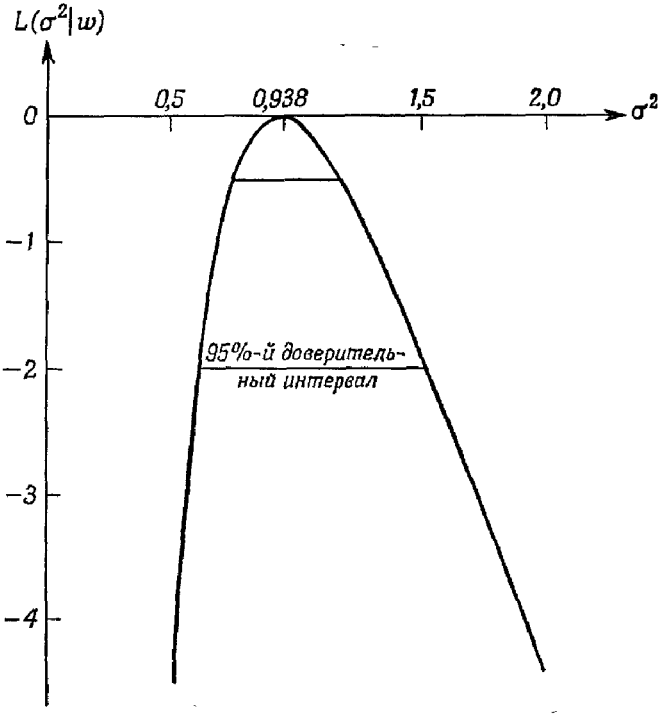
Случайная выборка большого объема

Представим себе случайную выборку объемом 40 из генеральной совокупности с распределением $N(0, \sigma^2)$, причем, как и в предыдущем случае, $\sigma^2 = 1$. Величина

$$w = \sum_1^{40} x_i^2 \tag{167}$$

должна подчиняться распределению χ_{40}^2. Для наших целей достаточно знать величину w; поэтому для получения

единственного значения w можно также воспользоваться таблицами случайных чисел. Пусть выбрано случайное



Ф и г. 25.

число $r = 0{,}4176$. Соответствующая случайная величина с распределением χ^2_{40} равна

$$w = 37{,}52.$$

Отсюда следует

$$\hat{\sigma}^2 = 0{,}938. \tag{168}$$

Согласно (164), логарифмическая функция правдоподобия

$$L(\sigma^2 \mid w) = 18{,}72 - 20 \ln \sigma^2 - \frac{18{,}76}{\sigma^2}. \tag{169}$$

Эта функция изображена графически на фиг. 25. Прямая линия $L = -1/2$ дает нам доверительный интервал

$$0{,}765 \leqslant \sigma^2 \leqslant 1{,}18,$$

который расположен примерно симметрично относительно значения $\hat{\sigma}^2$, т. е.

$$\sigma^2 = \hat{\sigma}^2{}^{+0,24}_{-0,17}. \tag{170}$$

Близость распределения полученной оценки к нормальному следует из почти параболической формы кривой. Кроме того, можно проверить приближенное выражение для дисперсии $\mathscr{D}(\hat{\sigma}^2)$ [формула (155)], часто используемое на практике:

$$\mathscr{D}(\hat{\sigma}^2) \approx -\frac{1}{\left[\frac{\partial^2 L(\sigma^2 \mid \mathbf{x})}{(\partial\sigma^2)^2}\right]_{\sigma^2=\hat{\sigma}^2}}.$$

[Такая процедура эквивалентна несколько необычному описанию $L(\sigma^2 \mid w)$ параболой с теми же значениями: $L(\hat{\sigma}^2)$, $L_{\sigma^2}(\hat{\sigma}^2)$ и $L_{\sigma^2\sigma^2}(\hat{\sigma}^2)$.] Из формулы (164) следует

$$\frac{\partial^2 L(\sigma^2 \mid x)}{(\partial\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{w}{(\sigma^2)^3} = -\frac{n}{2(\hat{\sigma}^2)^2} \quad \text{при} \quad \sigma^2 = \hat{\sigma}^2.$$

Отсюда

$$\mathscr{D}(\hat{\sigma}_2) \approx \frac{(\hat{\sigma}^2)^2}{20} \quad \text{для } n = 40$$

и среднеквадратичное отклонение, т. е.

$$\sigma(\hat{\sigma}^2) \approx \frac{\hat{\sigma}^2}{\sqrt{20}} = 0{,}21.$$

Этот результат близок к полученному более точным методом (170).

Для сопоставления приведем результаты, полученные на основании выборки объемом 6:

$$\hat{\sigma}^2 = 1{,}86,$$

$$\sigma^2 = \hat{\sigma}^2{}^{+1,63}_{-0,76} \quad \text{при} \quad L = -\frac{1}{2};$$

среднеквадратичное отклонение $\sigma(\hat{\sigma}^2)$ равно

$$\frac{\hat{\sigma}^2}{\sqrt{3}} = 1{,}06.$$

Оценка

$$\sigma^2 = \hat{\sigma}^2 \pm 1{,}06$$

совершенно упускает из виду то обстоятельство, что истинное значение σ^2 вполне может более чем вдвое отклоняться в обе стороны от $\hat{\sigma}^2$. Следует задуматься, компенсирует ли эта упрощенная запись только двух величин ($\hat{\sigma}^2$ и ее среднеквадратичного отклонения) весьма существенную потерю информации.

Предположение относительно нормальности распределения

Допустим, что все 40 значений x_i оказались отрицательными. Независимый наблюдатель должен был тогда отвергнуть гипотезу $\mu = 0$, так как ее вероятность составляет 2^{-40}. Однако метод максимального правдоподобия приведет к той же оценке σ^2, поскольку он имеет дело лишь с $\{x_i^2\}$. Поэтому надо тщательно проверять, справедливы ли допущения, на которых основано применение метода.

§ 16. Двумерная функция правдоподобия

Рассмотренный выше подход можно без труда обобщить на случай двух параметров. Если функцию правдоподобия можно представить в виде произведения двух сомножителей

$$l(\theta, \varphi \mid \mathbf{x}) = l_1(\theta \mid \mathbf{x})\, l_2(\varphi \mid \mathbf{x}), \tag{171}$$

то для решения задачи каждую из функций l_1 и l_2 можно рассматривать отдельно, поскольку оценки $\hat{\theta}$ и $\hat{\varphi}$ будут независимы. Таким образом, теперь мы имеем дело с более общим случаем. Нам предстоит определить $\hat{\theta}$ и $\hat{\varphi}$ одновременно. Разобранный ниже пример выбран специально для того, чтобы полученные результаты можно было использовать впоследствии при изложении метода наименьших квадратов.

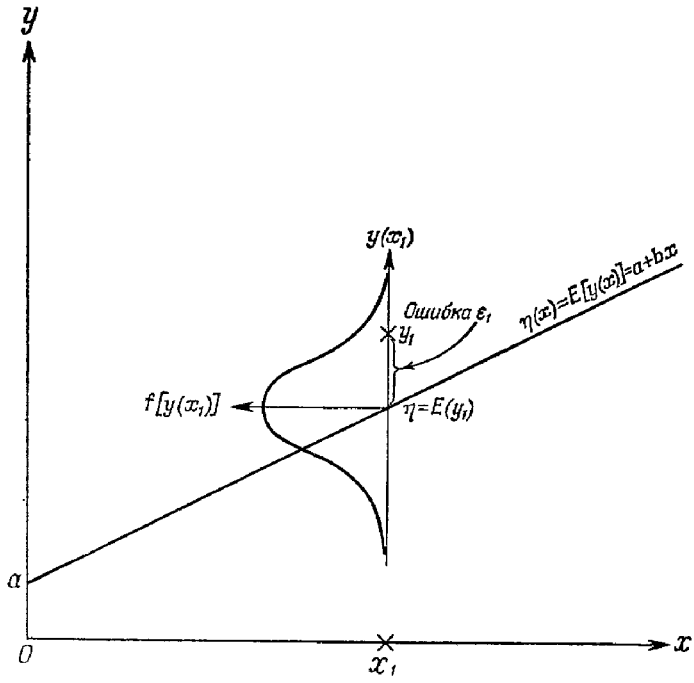
Аппроксимация результатов прямой линией. Пусть необходимо произвести оценку параметров a и b, причем

$$y \sim N(a + bx;\ 1), \tag{172}$$

где x — независимая переменная. Задачу иллюстрирует фиг. 26. Прямая линия

$$\eta(x) = \mathscr{E}[y(x)] = a + bx \tag{173}$$

представляет собой геометрическое место точек, соответствующих средним значениям случайной величины y (x), распределенной по нормальному закону. Такое геометри-



Фиг. 26.

ческое место точек называется *кривой регрессии*. Допустим, что экспериментальные данные включают точку (x_1, y_1). Разность

$$\varepsilon_1 = y_1 - \mathcal{E}(y_1) \tag{174}$$

называется *ошибкой* или *случайным отклонением*, и в данном случае

$$\varepsilon \sim N(0, 1). \tag{175}$$

Плотность распределения для n пар (x_i, y_i)

$$f(\mathbf{y}, a, b) = (2\pi)^{-n/2} e^{-1/2 \sum_{i=1}^{n} (y_i - a - bx_i)^2}, \tag{176}$$

откуда функция правдоподобия параметров a и b:

$$l(a,\ b\,|\,\mathrm{y}) = e^{-1/2\sum (y_i - a - bx_i)^2}. \qquad (177)$$

Логарифмическая функция правдоподобия, в которую добавлена константа, обеспечивающая $L_{\text{макс}} = 0$, имеет вид

$$L(a,\ b\,|\,\mathrm{y}) = -\frac{\sum (y_i - a - bx_i)^2}{2} + \\ + \frac{\sum y_i^2 - \left(\sum y_i\right)^2 \frac{1}{n} - \frac{\sum (x_i y_i)^2}{\sum x_i^2}}{2}. \qquad (178)$$

Это выражение необходимо исследовать на максимум по отношению к a и b. Тот же результат получится, если оценку этих параметров провести из условия минимума по a и b выражения

$$S = \sum (y_i - a - bx_i)^2. \qquad (179)$$

Такой подход лежит в основе хорошо известного *критерия наименьших квадратов*, который будет рассмотрен несколько позже.

Итак, вновь обратимся к логарифмической функции правдоподобия. Из выражения (178) следует

$$\frac{\partial L}{\partial a} = \sum (y_i - a - bx_i) = 0$$

и

$$\frac{\partial L}{\partial b} = \sum x_i (y_i - a - bx_i) = 0.$$

Мы получили систему уравнений

$$na + b \sum x_i = \sum y_i \qquad (180)$$

и

$$a \sum x_i + b \sum x_i^2 = \sum x_i y_i.$$

Решением этой системы являются величины

$$\tilde{a} = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2} \qquad (181)$$

и

$$\hat{b} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}.$$

Пример. Снова воспользуемся нормальной выборкой (165) объемом 6. Эту выборку можно рассматривать как результаты эксперимента. Предположим, что кривая регрессии имеет вид

$$\eta(x) = \mathscr{E}[y(x)] = 1 + 2x$$

и что данные имеют вид

$$[x,\ \eta(x) + \varepsilon],$$

где $\varepsilon \sim N(0, 1)$; $x = -5, -3, -1, 1, 3, 5$, а набор значений $\{\varepsilon_i\}$ взят из (165). Итак, при $n = 6$ имеем

x	-5	-3	-1	1	3	5
y	$-7{,}532$	$-7{,}097$	$-0{,}854$	$2{,}705$	$4{,}880$	$11{,}017$

Отсюда следует, что

$$\hat{a} = 0{,}520, \quad \hat{b} = 1{,}89. \tag{182}$$

Существует большое число задач, в которых может возникнуть необходимость сделать выводы относительно a и b в отдельности. Здесь же мы по-прежнему будем иметь дело с двумерным параметром (a, b). Мы получили *точечную оценку* $(\hat{a}, \hat{b})$ [см. (182)], и теперь нам надо найти двумерную доверительную область, которая дает *интервальную оценку* $(\hat{a}, \hat{b})$.

Вычислим

$$L = -\frac{1}{2}[(-7{,}532 - a + 5b)^2 + (-7{,}097 - a + 3b)^2 + \ldots$$
$$\ldots + (11{,}017 - a - 5b)^2] + \text{const} \tag{183}$$

для различных значений a и b, что позволит изобразить контуры поверхности L в плоскости (a, b). Ранее мы нахо-

дили пределы ошибок параметра θ в одномерной задаче, «смещаясь» по ветвям логарифмической функции правдоподобия относительно максимума до уровня $L = -1/2$ (или -2). Таким способом мы определяли наименьший из возможных доверительных интервалов, соответствующий заданной доверительной вероятности 0,68 (или 0,95).

Аналогично найдем наименьшую из возможных *доверительных областей* для (a, b), «смещаясь» по логарифмической поверхности правдоподобия до заданного уровня L относительно максимума. По определению границей такой области является просто контур сечения поверхности. Соответствующая доверительная вероятность определяется, как обычно в случае нормальных параметров со средними значениями, например μ и ν. Пусть x и y независимы, так что

$$f(x, y) = \frac{1}{2\pi} e^{\frac{-(x-\mu)^2}{2} - \frac{(y-\nu)^2}{2}}.$$

Тогда

$$L(\mu, \nu) = -\frac{1}{2}(x-\mu)^2 - \frac{1}{2}(y-\nu)^2 = 0 \quad \text{при} \quad x = \hat{\mu}, \; y = \hat{\nu}.$$

С другой стороны,

$$(x-\mu)^2 + (y-\nu)^2 \sim \chi_2^2$$

и, таким образом, величина $-2L$ обладает распределением χ_2^2. Доверительная вероятность определяется следующим образом:

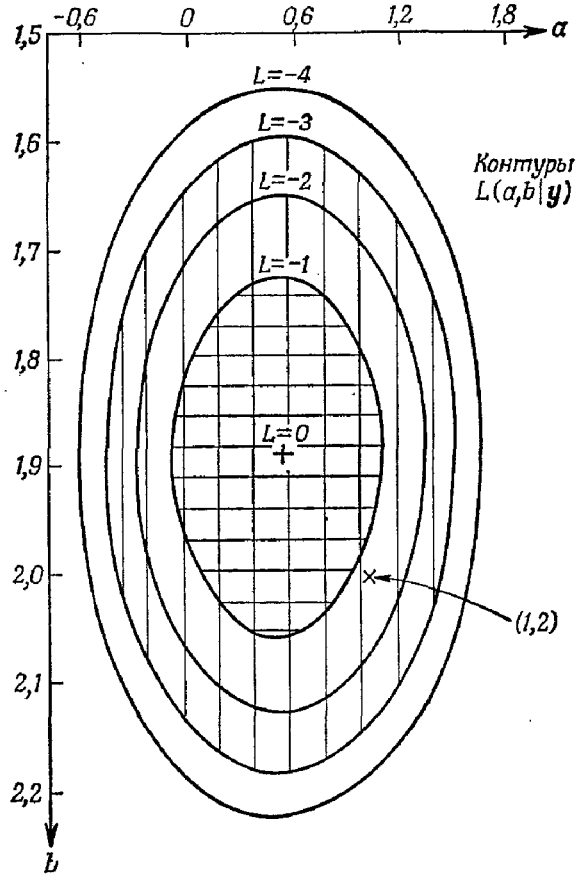
$$\mathcal{P}(\chi_2^2 < -2L), \tag{184}$$

причем $L_{\text{макс}} = 0$. Ниже приведено несколько характерных значений:

L	0	$-1/2$	-1	$-1\,1/2$	-2	$-2\,1/2$	-3	$-\infty$
$\mathcal{P}$	0	39,4%	63,2%	77,7%	86,5%	91,8%	95%	100%

Грубо говоря, область, ограниченная контуром $L(a, b) = -1{,}15$, соответствует такой же доверительной вероятности, что и интервал, соответствующий $L(\mu) =$

$= -1/2$, а контур $L(a, b) = -3{,}1$ можно считать эквивалентным интервалу, отвечающему $L(\mu) = -2$. Поэтому разумно считать «малой» доверительной областью для



Фиг. 27.

(a, b) область, лежащую внутри контура

$$L(a, b) = -1,$$

и «большой» — область внутри контура

$$L(a, b) = -3.$$

На фиг. 27 изображен ряд контуров $L(a, b \mid y)$ для некоторых значений L. Как и в одномерном случае, малая

доверительная область не содержит известного истинного значения $(a, b) = (1, 2)$, что имеет место в 37 случаях из 100. Большая доверительная область содержит точку (1, 2).

Прямая максимального правдоподобия имеет вид

$$\hat{y} = \hat{a} + \hat{b}x. \tag{185}$$

Таким образом, это выражение представляет собой уравнение «случайной» прямой на плоскости, причем среднее значение задается «истинной» прямой $y = a + bx$.

Насколько эта случайная прямая может отклоняться от средней, очевидно, зависит от отклонений точки $(\hat{a}, \hat{b})$ от своего среднего (a, b). Поскольку (a, b), вообще говоря, неизвестно, необходимо учитывать разброс точек $\{(a_j, b_j)\}$ относительно $(\hat{a}, \hat{b})$. Обозначим через R_L область в плоскости (a, b), которая ограничена контуром, расположенным на уровне $-L$. Каждой точке $(a_j, b_j) \in R_L$ соответствует прямая, аналогичная (185). Точнее, доверительная область R_L для (a, b) эквивалентна пучку прямых в плоскости (x, y)

$$y_j = a_j + b_j x, \tag{186}$$

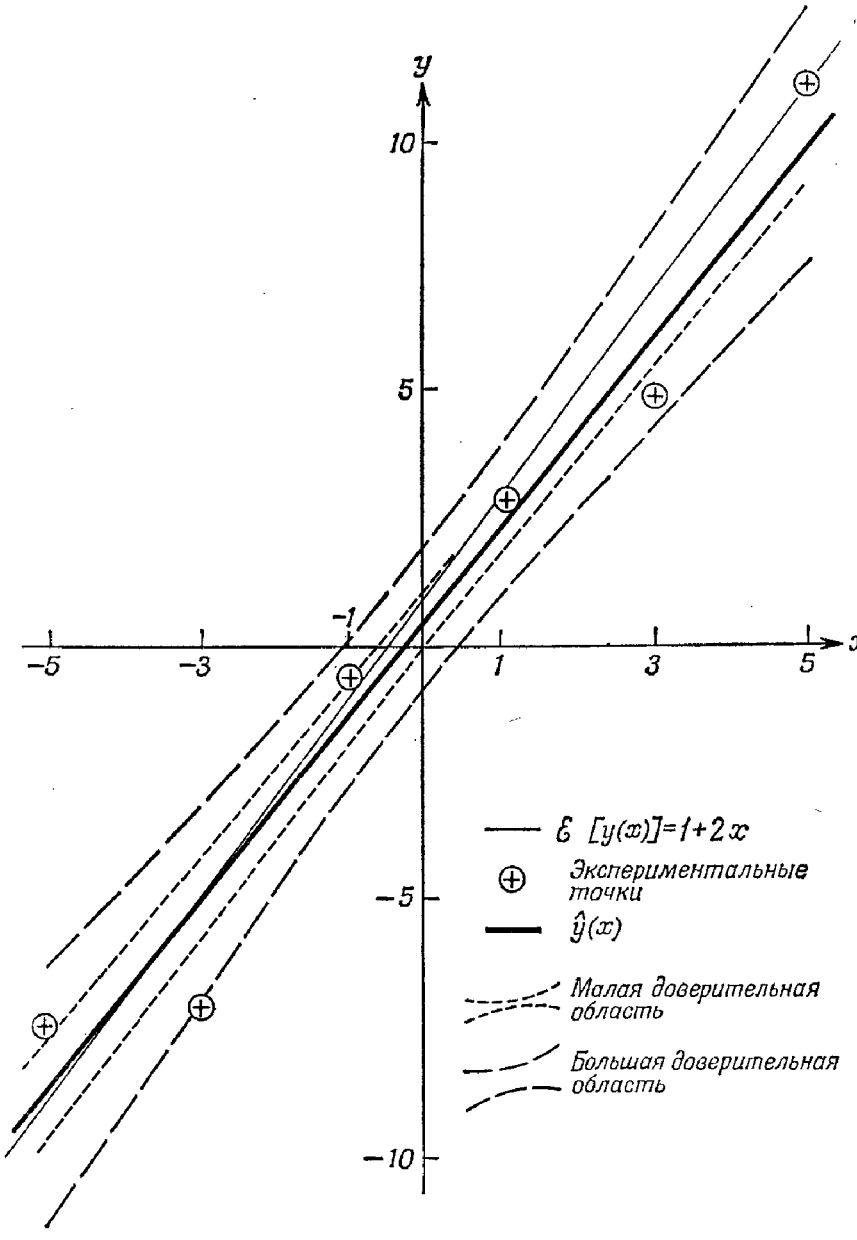
где $(a_j, b_j) \in R_L$. Беря огибающую этого пучка прямых, мы получаем доверительную область для истинной прямой.

Пусть $y_0(x)$ — верхняя граница огибающей. По определению огибающей

$$y_j = a_j + b_j x \leqslant y_0(x)$$

для всех x и для всех $(a_j, b_j) \in R_L$. Таким образом, $y_0(x)$ может быть найдено путем определения максимума выражения $y_j = a_j + b_j x$ при любом x и при всех $(a_j, b_j) \in R_L$. Нетрудно видеть, что при заданном x значение (a_j, b_j), обеспечивающее максимум $y_0(x)$, должно лежать на границе области R_L. С помощью детального анализа, основанного на использовании множителей Лагранжа, можно показать, что

$$y_0(x) = \hat{y}(x) \pm \left[(-2L)\left(\frac{1}{n} + \frac{x^2}{\sum x_i^2}\right)\right]^{1/2}, \tag{187}$$


y
10
5
−5
−10
−1
−5
−3
1
3
5
x
E [y(x)]=1+2x
Экспериментальные точки
ŷ(x)
Малая доверительная область
Большая доверительная область

Фиг. 28.

т. е. представляет собой гиперболу. Знак минус после величины $\hat{y}(x)$ соответствует нижней границе огибающей, и, таким образом, доверительная область расположена симметрично относительно $\hat{y}(x)$ [1]). Малая доверительная область в плоскости (a, b) соответствует малой доверительной области для $y(x)$, определяемой из уравнения (187) с $L = -1$, т. е.

$$y(x) = 0{,}520 + 1{,}89x \pm \left[2\left(\frac{1}{6} + \frac{x^2}{70}\right)\right]^{1/2}. \qquad (188)$$

На фиг. 28 изображены: набор данных

$$(x_i, y_i),$$

истинная кривая регрессии

$$\mathscr{E}[y(x)] = 1 + 2x, \qquad (189)$$

ее оценка максимального правдоподобия

$$\hat{y}(x) = 0{,}520 + 1{,}89x, \qquad (190)$$

малая доверительная область для $y(x)$ [см. (188)] и большая доверительная область для $y(x)$, которая соответствует выражению (187) с $L = -3$,

$$y(x) = \hat{y}(x) \pm \left[6\left(\frac{1}{6} + \frac{x^2}{70}\right)\right]^{1/2}. \qquad (191)$$

При $x = 3$ верхняя граница выражения (189) для малой доверительной области равна 6,95, в то время как $\mathscr{E}[y(3)] = 7$. Таким образом, некоторый участок истинной линии регрессии $\mathscr{E}[y(x)]$ лежит вне малой доверительной области. Этого следовало ожидать, поскольку мы видели (стр. 135), что точка $(a, b) = (1, 2)$ также не попадает в пределы малой доверительной области для $(\hat{a}, \hat{b})$.

§ 17. Теорема Бейеса

На стр. 91 мы поставили вопрос: каким образом учесть априорное знание при анализе экспериментальных результатов? Ответ на него был дан Томасом Бейесом в его известной работе «An essay towards solving a problem

[1]) См. также [21], § 28.25.

in the doctrine of chances» («Попытка решения задачи в проблеме случая»), опубликованной посмертно в 1763 г. Теорема Бейеса требует использования понятия распределения вероятности для характеристики предполагаемой возможности, что неизвестный параметр принимает определенное значение. Таким образом, мы распространяем обычное представление о вероятности, сводящееся к утверждению «$\mathcal{P}(x \leqslant 1) = {}^1/_3$» для *случайной величины* x на тот случай, когда утверждение «$\mathcal{P}(\theta \leqslant 2) = {}^1/_2$» относится к неизвестному *параметру* θ.

Проиллюстрируем различие между этими двумя представлениями о вероятности. Многократно повторяя определение случайной величины x, мы ожидаем, что доля случаев, в которых ее значение $\leqslant 1$, примерно равна одной трети. Но мы не можем повторять эксперимент для получения различных значений θ, ибо θ постоянно. Понятие «вероятности» по отношению к θ скорее отражает предполагаемую возможность. Конечно, выгоднее держать пари, что $\theta \leqslant 2$ с неравными ставками, нежели с равными, и если в дальнейшем выяснится истинное значение θ, то можно ожидать выигрыша пари в 50 случаях из 100.

Определение 1. Априорная вероятность $p(\theta)$ параметра θ характеризует предполагаемую возможность осуществления различных значений θ до того, как проведен эксперимент.

Объективный подход требует, чтобы эта предполагаемая возможность основывалась на известных фактах, таких, как результаты предыдущих экспериментов [1].

Определение 2. Апостериорная вероятность $q(\theta \mid \mathbf{x})$ характеризует возможность осуществления различных значений θ после того, как к априорному знанию добавлено знание, извлеченное из экспериментальных данных $\mathbf{x}$.

ТЕОРЕМА БЕЙЕСА. Утверждается, что апостериорная вероятность параметра θ получается умножением априор-

[1]) Этот вопрос подробно разобран в книге Фитера [11], гл. 2 и 3.

ной вероятности на функцию правдоподобия. Более точно:

$$q(\theta \mid \mathbf{x}) = \frac{p(\theta)\, l(\theta \mid \mathbf{x})}{C}, \tag{192}$$

где C — константа, выбранная так, чтобы

$$\int_{-\infty}^{\infty} q(\theta \mid \mathbf{x})\, d\theta = 1 \quad \text{(в случае непрерывного } \theta)$$

или

$$\sum_i q(\theta_i \mid \mathbf{x}) = 1 \quad \text{(в случае дискретного } \theta).$$

Это значит, что

$$C = \int_{-\infty}^{\infty} p(\theta)\, l(\theta \mid \mathbf{x})\, d\theta \tag{193}$$

или

$$C = \sum_i p(\theta_i)\, l(\theta_i \mid \mathbf{x}). \tag{194}$$

Из формул (192) — (194) следует, что любой постоянный множитель у функции $l(\theta \mid \mathbf{x})$ может быть опущен и это не изменит результата.

Доказательство теоремы состоит из простых операций над вероятностями с учетом определений совместной, условной и маргинальной (безусловной) вероятности.

Рассмотрим совместную вероятность для θ и $\mathbf{x}$. Согласно правилу умножения,

$$\mathscr{P}(\theta \cap \mathbf{x}) = \mathscr{P}(\mathbf{x}) \cdot \mathscr{P}(\theta \mid \mathbf{x}),$$

т. е.

$$\mathscr{P}(\theta \mid \mathbf{x}) = \frac{\mathscr{P}(\theta \cap \mathbf{x})}{\mathscr{P}(\mathbf{x})}.$$

Подобным образом

$$\mathscr{P}(\theta \cap \mathbf{x}) = \mathscr{P}(\theta) \cdot \mathscr{P}(\mathbf{x} \mid \theta),$$

откуда следует

$$\mathscr{P}(\theta \mid \mathbf{x}) = \frac{\mathscr{P}(\theta) \cdot \mathscr{P}(\mathbf{x} \mid \theta)}{\mathscr{P}(\mathbf{x})}.$$

Если θ неизвестно, а x получено на опыте, то

$$\mathscr{P}(\theta \mid x) = \frac{\mathscr{P}(\theta) \cdot l(\theta \mid x)}{\mathscr{P}(x)}$$

в силу определения $l(\theta \mid x)$. Тогда остается лишь вычислить $\mathscr{P}(x)$. Распределение маргинальной (безусловной) вероятности $\mathscr{P}(x)$ получается усреднением $\mathscr{P}(x \cap \theta)$ по всем возможным значениям θ.

В случае непрерывного параметра θ

$$\mathscr{P}(x) = \int_{-\infty}^{\infty} \mathscr{P}(x \cap \theta)\, d\theta = \int_{-\infty}^{\infty} \mathscr{P}(\theta)\, l(\theta \mid x)\, d\theta$$

на основании предыдущего.

В случае дискретного параметра θ

$$\mathscr{P}(x) = \sum \mathscr{P}(\theta_i)\, l(\theta_i \mid x).$$

Этим и исчерпывается доказательство.

Следует подчеркнуть, что теорема оказывается логически неуязвимой только при условии, что мы приняли существование распределения вероятности для параметров.

Применение априорной вероятности

Прежде всего возникает задача вычисления априорной вероятности параметра θ. Вернемся к задаче с двумя урнами (стр. 102). Мы имеем здесь дело с дискретным параметром, принимающим лишь два возможных значения: $\theta = {}^3/_4$ или ${}^1/_4$. Мы должны считать, что каждая урна имеет равные шансы быть выбранной, так что априорные вероятности

$$p\left(\theta = \frac{3}{4}\right) = \frac{1}{2} = p\left(\theta = \frac{1}{4}\right).$$

Функция правдоподобия параметра θ для случая, когда из n выбранных шаров r оказались черными, равна

$$l(\theta \mid r) = p^r (1-p)^{n-r}.$$

Следовательно, апостериорная вероятность $\theta = 1/4$ есть

$$q\left(\frac{1}{4}\middle|r\right) = \frac{p\left(\frac{1}{4}\right) l\left(\frac{1}{4}\middle|r\right)}{p\left(\frac{1}{4}\right) l\left(\frac{1}{4}\middle|r\right) + p\left(\frac{3}{4}\right) l\left(\frac{3}{4}\middle|r\right)} =$$

$$= \frac{\left(\frac{1}{4}\right)^r \left(\frac{3}{4}\right)^{n-r}}{\left(\frac{1}{4}\right)^r \left(\frac{3}{4}\right)^{n-r} + \left(\frac{3}{4}\right)^r \left(\frac{1}{4}\right)^{n-r}}, \qquad (195)$$

поскольку можно сократить на $p(1/4) = p(3/4) = 1/2$.

Аналогичным образом

$$q\left(\frac{3}{4}\middle|r\right) = \frac{\left(\frac{3}{4}\right)^r \cdot \left(\frac{1}{4}\right)^{n-r}}{\left(\frac{1}{4}\right)^r \left(\frac{3}{4}\right)^{n-r} + \left(\frac{3}{4}\right)^r \left(\frac{1}{4}\right)^{n-r}}. \qquad (196)$$

В любом случае в качестве точечной оценки параметра θ мы выбираем такое значение $\hat{\theta}$, которое делает апостериорную вероятность максимальной. Так, сравнивая выражения (195) и (196), следовало бы выбрать $\hat{\theta} = 1/4$ или $\hat{\theta} = 3/4$ в соответствии с наибольшим значением апостериорной вероятности. В частности, при равномерном распределении априорной вероятности для θ точечная оценка параметра θ совпадает с оценкой максимального правдоподобия, поскольку в этом случае выражение (192) превращается в

$$q(\theta \mid \mathbf{x}) = \frac{l(\theta \mid \mathbf{x})}{C}. \qquad (197)$$

Случай, когда априорная вероятность неизвестна

Допустим теперь, что в рассмотренной задаче ни слова не сказано о способе выбора одной из урн. Каким образом тогда правильно вычислить априорную вероятность, которая отражала бы отсутствие сведений о величине θ? Сколько-нибудь удовлетворительного ответа на этот вопрос, по-видимому, не существует. Попыткой выйти из затруднительного положения явился постулат, который был опубликован Бейесом вместе с теоремой.

ПОСТУЛАТ БЕЙЕСА. Если распределение априорной вероятности для θ неизвестно, то, предполагая, что все значения θ априори равновероятны, мы сможем охарактеризовать нашу полную неосведомленность относительно θ. Иными словами, следует положить

$$p(\theta) = \text{const}.$$

Автор сам был настолько неуверен в справедливости данного постулата, что так и не опубликовал его при жизни. Хотя этот постулат логически и несостоятелен, тем не менее он в свое время был поддержан Лапласом, рассматривавшим его как аксиому. Впервые с серьезной критикой постулата Бейеса выступил в 1854 г. Джордж Буль. В его работе «Законы мышления» говорится: «Однако я опасаюсь, что этот метод слишком произволен... Если недостаток (априорных) данных восполняется гипотезой, то решение будет, вообще говоря, зависеть от характера этой гипотезы».

В качестве примера, иллюстрирующего трудности, которые возникают в связи с использованием постулата Бейеса, рассмотрим оценку среднеквадратичного отклонения σ нормально распределенной случайной величины с нулевым средним значением. Пусть для определенности известно, что $1 \leqslant \sigma \leqslant 2$, и это единственное, что нам известно о σ. Согласно постулату, плотность априорной вероятности σ

$$\begin{aligned} f(\sigma)\, d\sigma &= 1 \cdot d\sigma, \quad \text{если } 1 \leqslant \sigma \leqslant 2; \\ &= 0 \qquad\quad \text{для остальных значений } \sigma. \end{aligned}$$

Функция правдоподобия для σ при выборке единичного объема равна

$$l(\sigma \mid x) = \frac{1}{\sigma} e^{-x^2/2\sigma^2}.$$

Распределение апостериорной вероятности σ

$$q(\sigma \mid x) = \frac{1 \cdot e^{-x^2/2\sigma^2} \cdot \dfrac{1}{\sigma}}{\displaystyle\int_1^2 e^{-x^2/2\sigma^2} \cdot \frac{d\sigma}{\sigma}} = \frac{2C}{\sigma} \cdot e^{-x^2/2\sigma^2} \qquad \text{при } 1 \leqslant \sigma \leqslant 2.$$

Интересно также найти распределение апостериорной вероятности σ^2. Пусть $u = \sigma^2$; тогда $du = 2\sigma\, d\sigma = 2\sqrt{u}\, d\sigma$, т. е. $d\sigma = du/2\sqrt{u}$. Отсюда следует

$$r(u \mid x) = \frac{C'}{u} e^{-x^2/2u} \text{ при } 1 \leqslant u \leqslant 4.$$

Допустим теперь, что другой экспериментатор пожелал использовать те же данные с несколько иной целью. Его больше интересует не среднее квадратичное отклонение, а дисперсия, которую он считает важной величиной и намерен оценить. Относительно $u = \sigma^2$ ему известно лишь, что значение u заключено между 1 и 4.

Для получения априорной вероятности экспериментатор вновь обратился бы к постулату Бейеса:

$$g(u)\, du = \frac{1}{3} du, \text{ если } 1 \leqslant u \leqslant 4;$$

$$= 0 \qquad \text{для остальных значений } u.$$

Функцией правдоподобия для такой же выборки будет

$$l_2(u \mid x) = \frac{1}{\sqrt{u}} e^{-x^2/2u},$$

и распределение апостериорной вероятности u

$$s(u \mid x) = \frac{e^{-x^2/2u} \dfrac{1}{3\sqrt{u}}}{\displaystyle\int_1^4 e^{-x^2/2u} \frac{du}{3\sqrt{u}}} \text{ при } 1 \leqslant u \leqslant 4.$$

Полученное выражение отличается от $r(u \mid x)$; таким образом, создается неприглядное положение, когда два различных экспериментатора, произведя математический анализ одних и тех же данных, приходят к различным выводам. Очевидно, что это различие обусловлено произвольным выбором распределения априорной вероятности.

Современный подход к общим проблемам экспериментов с неизвестной априорной вероятностью заключается в том, что для нахождения *точечных оценок* используются *критерии максимального правдоподобия* или *наименьших квадратов*, а для *проверки гипотезы — критерий значимости*.

Библиография

Принцип максимального правдоподобия был использован еще Гауссом [15]. В качестве общего метода оценки параметров метод максимального правдоподобия был введен в 1912 г. Р. Фишером в работе «Об абсолютном критерии для приближения частотных кривых» [12]. Впоследствии Фишер внес большой вклад в развитие этого метода (см., например, [11]).

Легко доступна для чтения по тому же вопросу книга Кендалла и Стьюарта [21], том 1, гл. 8; следует также порекомендовать книгу Ван дер Вардена [36]. Более строгое рассмотрение проблемы дано Уилксом [38], § 6.2, а также Кендаллом и Стьюартом [21], том 2, гл. 18. Фундаментальная «Теория вероятностей» Г. Джеффриса [20] содержит исследование теоремы Бейеса с его постулатом, который рассматривается как аксиома. По этому поводу в книге сказано: «...мы обеспечены... формальным правилом, соответствующим здравому смыслу, которое учит нас практике выбора гипотез».

В книге Муда и Грейбилла [28], § 8.9 и далее, проведено сравнение метода максимального правдоподобия, метода моментов и оценок Бейеса. Недавно опубликованная работа Л. Сейвиджа и др. [34] содержит общетеоретическую дискуссию ведущих специалистов в этой области.

ГЛАВА ПЯТАЯ

Метод наименьших квадратов

§ 18. История развития метода

Первая работа, в которой использовался метод наименьших квадратов, принадлежит Лежандру. В 1805 г. в статье «Новые методы определения орбит комет» он писал: «После того как полностью использованы все условия задачи, необходимо определить коэффициенты так, чтобы величины их ошибок были наименьшими из возможных. Для этого нами указан, вообще говоря, простой метод, который состоит в отыскании минимума суммы квадратов ошибок».

Первым, кто попытался поставить этот метод на прочную математическую основу, был, по-видимому, Р. Адрейн. В его работе «Исследования, касающиеся вероятностей ошибок, которые появляются при измерениях» (1808 г.) можно найти грубое «доказательство» того, что в некоторых условиях эти ошибки ε подчиняются распределению вероятности

$$\varphi(\varepsilon) = Ce^{-h^2\varepsilon^2},$$

т. е. нормальному распределению. Отсюда автоматически вытекал метод наименьших квадратов, дающий «наиболее вероятное» или наиболее правдоподобное решение для неизвестных параметров некоторой линейной формы.

В 1809 г. Гаусс в известной работе о вычислении орбит дал второе обоснование нормального закона распределения ошибок. Кроме того, Гаусс отстаивал приоритет использования метода наименьших квадратов с 1795 г.

Интересно упомянуть и другие методы, использовавшиеся в те времена. Так, в 1792 г. Лаплас применял критерий, согласно которому наименьшее значение должна принимать сумма абсолютных ошибок $\sum_i |\varepsilon_i|$, а не сумма их квадратов $\sum_i \varepsilon_i^2$. Начиная с 1810 г. метод наи-

меньших квадратов использовали Лаплас, а также Бессель и позже Энке.

В 1831 г. Коши предложил метод, в котором требовалось, чтобы «наибольшая сумма всех ошибок, взятых без учета их знаков, достигала минимума», т. е. он предложил минимизировать $\max \sum_i |\varepsilon_i|$. Этот принцип минимакса иногда еще применяется и по сей день.

Дальнейшее развитие метода наименьших квадратов связано с именем Лапласа, который в 1812 г. в работе «Аналитическая теория вероятностей» показал, что этот метод позволяет найти несмещенные оценки безотносительно к типу исходного распределения.

Гаусс опубликовал свои соображения в 1821 г. Не используя такие понятия, как дисперсия, и не обращаясь к матричной алгебре, он доказал, что среди класса оценок, которые являются: а) линейными комбинациями исходных данных и б) несмещенными оценками параметров, оценки, получаемые методом наименьших квадратов, обладают наименьшими погрешностями. Самое важное свойство оценок, основанных на методе наименьших квадратов, заключается в независимости от типа распределения.

В более общем виде эта теорема была доказана в 1912 г. А. Марковым и в настоящее время известна как теорема Гаусса — Маркова.

И наконец, большой вклад в развитие метода был внесен в 1934 г. Эйткеном, который обобщил теорему на случай коррелированных результатов наблюдений с различными дисперсиями.

Читателю можно порекомендовать работу М. Мерримана «Перечень работ, относящихся к методу наименьших квадратов, с историческими и критическими замечаниями» [25].

§ 19. Анализ регрессий [1])

Введение

Пусть имеются две случайные величины x и y, которые связаны между собой следующим образом. Среднее значение величины y представляет собой не константу, а функ-

[1]) См. также [21], гл. 26—30, том 2.

цию от x

$$\mathscr{E}[y(x)] = \eta(x, \theta),$$

где θ — обозначение набора неизвестных параметров, полностью определяющих функцию $\eta(x, \theta)$. Принятый в статистике классический подход состоит в использовании случайной выборки двумерной величины (x, y) и метода максимального правдоподобия оценки θ.

В качестве примера допустим, что x и y означают величины соответственно наименьшего и наибольшего диаметров картофеля. Произвольно отберем партию картофеля и сделаем измерения (x_i, y_i) каждого экземпляра. Здесь не предполагается, что $\{x_i\}$ заранее определены (см. [21], гл. 28).

С точки зрения экспериментатора часто более естественно считать, что $\{x_i\}$ фиксировано заранее. В этом случае x будет «независимой» величиной, а η —«зависимой». (Слова «независимый» и «зависимый» мы применяем здесь в алгебраическом смысле, подчеркивая тем самым, что η рассматривается как функция x, а не наоборот.) Чтобы получить оценку $\eta(x_i)$, мы используем выборку (одномерной) случайной величины $y(x_i)$. Очень часто нет необходимости точно оценивать $\eta(x)$; достаточно в качестве приближения для $\eta(x)$ использовать разложение в ряд Тэйлора или в ряд Фурье. Но тогда возникает вопрос: *сколько членов ряда следует удерживать в разложении*? При разложении в ряд Тэйлора, например, необходимо знать, какую надо выбрать степень полинома, чтобы он давал наилучшее описание.

Если случайный разброс $y(x)$ относительно $\eta(x)$ подчиняется нормальному распределению, то решение, основанное на использовании метода наименьших квадратов, *мнк*-решение, совпадает с решением, полученным методом максимального правдоподобия. Это обстоятельство служит мощным обоснованием применения *мнк*-критерия в качестве метода оценки. Однако предположение о нормальном характере распределения часто оказывается неоправданным. Мы увидим, что *мнк*-решение $\hat{\eta}(x)$ обладает определенными оптимальными свойствами, которые не зависят от характера распределения $y(x)$. Прежде всего предпримем необходимые меры предосторож-

ности, оперируя с результатами y_i, которые имеют различные дисперсии

$$\mathscr{D}(y_i) = \frac{\sigma^2}{w_i},$$

где w_i называется *весом* y_i. Затем покажем, что среди класса несмещенных оценок $\eta(x)$, являющихся линейными функциями самих данных $\{y_i\}$, *мнк*-оценки обладают наибольшей точностью.

Вся оставшаяся часть этой главы будет посвящена технике вычислений, с помощью которой методом наименьших квадратов в простейших случаях находят решения. Мы увидим, как использование ортогональных функций упрощает и расчет, и интерпретацию результатов. При выяснении вопроса о числе членов, которым следует ограничиться в разложении в ряд Тэйлора, будет использован критерий $\mathscr{F}$.

Аппроксимация полиномом [1])

Стандартный метод оценки линии регрессии основан на использовании следующей *линейной* относительно параметров модели:

$$\eta(x) = \theta_0 f_0(x) + \theta_1 f_1(x) + \ldots + \theta_r f_r(x), \qquad (198)$$

где $f_0(x), f_1(x), \ldots, f_r(x)$ — заданные функции. Несмотря на то что нас будут интересовать главным образом разложения по полиномам, следует отметить применимость излагаемой теории к случаю разложения в ряд Фурье, т. е. когда

$$f_0(x) = \frac{1}{2}, \quad f_1(x) = \sin x, \quad f_2(x) = \cos x, \ldots$$

$$\ldots, f_{2r-1}(x) = \sin rx, \quad f_{2r}(x) = \cos rx.$$

При разложении в ряд Тэйлора

$$f_0(x) = 1, \quad f_1(x) = x, \quad f_2(x) = x^2, \ldots, f_r(x) = x^r,$$

так что

$$\eta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \ldots + \theta_r x^r.$$

[1]) См. также книгу Плэкетта [33].

В общем виде случайную величину y_i можно записать как

$$y_i = \eta(x_i) + \varepsilon_i,$$

где ε_i — *ошибки*, представляющие собой случайные величины с одинаковой дисперсией, причем распределения ε_i вовсе не обязательно должны быть нормальными. Таким образом, ошибки характеризуют разброс $\{y_i\}$ относительно их математического ожидания $\{\mathscr{E}(y_i)\}$:

$$\left.\begin{aligned} \varepsilon_1 &= y_1 - \theta_0 - \theta_1 x_1 - \ldots - \theta_r x_1^r, \\ \varepsilon_2 &= y_2 - \theta_0 - \theta_1 x_2 - \ldots - \theta_r x_2^r, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ \varepsilon_n &= y_n - \theta_0 - \theta_1 x_n - \ldots - \theta_r x_n^r. \end{aligned}\right\} \qquad (199)$$

В примере, рассмотренном на стр. 133, мы полагали $r = 1$.

Произведем оценку неизвестных параметров $\{\theta_j\}$ в выражении (198). Для этого потребуем, чтобы сумма квадратов ошибок достигала минимума. Иными словами, найдем минимум величины

$$S = \sum_{i=1}^{n} \varepsilon_i^2 \qquad (200)$$

по переменным $\theta_0, \theta_1, \ldots, \theta_r$.

Очевидно, что

$$S = \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i - \ldots - \theta_r x_i^r)^2, \qquad (201)$$

и экстремальное значение S получится в результате решения системы линейных уравнений

$$\left.\begin{aligned} \frac{\partial S}{\partial \theta_0} &= -2 \sum (y_i - \theta_0 - \theta_1 x_i - \ldots - \theta_r x_i^r) = 0, \\ \frac{\partial S}{\partial \theta_1} &= -2 \sum x_i (y_i - \theta_0 - \theta_1 x_i - \ldots - \theta_r x_i^r) = 0, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \frac{\partial S}{\partial \theta_r} &= -2 \sum x_i^r (y_i - \theta_0 - \theta_1 x_i - \ldots - \theta_r x_i^r) = 0. \end{aligned}\right\} \qquad (202)$$

Эти уравнения можно переписать в виде так называемых *нормальных уравнений*:

$$\left.\begin{aligned} n\theta_0 + \theta_1 \sum x_i + \ldots + \theta_r \sum x_i^r &= \sum y_i, \\ \theta_0 \sum x_i + \theta_1 \sum x_i^2 + \ldots + \theta_r \sum x_i^{r+1} &= \sum x_i y_i, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ \theta_0 \sum x_i^r + \theta_1 \sum x_i^{r+1} + \ldots + \theta_r \sum x_i^{2r} &= \sum x_i^r y_i. \end{aligned}\right\} \qquad (203)$$

Решение этих уравнений сильно упрощается, если использовать матричную алгебру. Кроме того, можно пока не обращать внимания на то, что в данном случае мы имели дело с $f_j(x) = x^j$. Тогда все рассмотрение приобретает более общий характер.

Введем конструкционную (*структурную*) матрицу

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^r \\ 1 & x_2 & x_2^2 & \ldots & x_2^r \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & x_n & x_n^2 & \ldots & x_n^r \end{bmatrix}, \qquad (204)$$

имеющую n строк и p столбцов, а также матрицу

$$\theta' = (\theta_0, \theta_1, \theta_2, \ldots, \theta_r), \qquad (205)$$

имеющую всего одну строку, содержащую p элементов, причем $p = r + 1$, а штрих означает транспонированную матрицу.

Матричная алгебра применима и в том случае, когда линейная модель носит более общий характер, т. е. когда, например, первая строка матрицы A имеет вид

$$[f_0(x_1), f_1(x_1), \ldots, f_r(x_1)].$$

Перепишем уравнения (199) — (201) в матричной форме

$$\varepsilon = y - A\theta, \qquad (199')$$

$$S = \varepsilon'\varepsilon, \qquad (200')$$

$$S = (y - A\theta)'(y - A\theta), \qquad (201')$$

$$S = y'y - 2\theta'A'y + \theta'A'A\theta. \qquad (206)$$

Если продифференцировать выражение (206) по каждому θ_i, то получится система уравнений

$$-2A'\mathbf{y}+2A'A\theta=0, \qquad (202')$$

которую мы перепишем в виде

$$(A'A)\,\theta=A'\mathbf{y}. \qquad (203')$$

Решение уравнений (203) и (203′) имеет вид

$$\hat{\theta}=(A'A)^{-1}A'\mathbf{y} \qquad (207)$$

при условии, что

$$C=A'A \qquad (208)$$

является матрицей полного ранга. Матрица C будет сингулярной только в том случае, когда производится оценка избыточного параметра, как, например, Φ_2 в линейной форме

$$\eta(x)=\Phi_0+\Phi_1x+\Phi_2x,$$

где должно быть только

$$\eta(x)=\Phi_0+\Phi_1x.$$

Вообще говоря, избыточность имеет место, если строки матрицы A' линейно зависимы.

Остаточная сумма квадратов

Разность между аппроксимирующей величиной

$$\hat{y}_i=\hat{\theta}_0+\hat{\theta}_1x_i+\ldots+\hat{\theta}_rx_i^r \qquad (209)$$

и экспериментальным значением y_i называется *остатком* Так, i-й остаток равен

$$y_i-\hat{y}_i=y_i-\hat{\theta}_0-\hat{\theta}_1x_i-\ldots-\hat{\theta}_rx_i^r,$$

а вектор остатков

$$\mathbf{y}-\hat{\mathbf{y}}=\mathbf{y}-A\hat{\theta}.$$

Сумма квадратов остатков называется *остаточной суммой квадратов* R. Подставив $\hat{\theta}$ в уравнение (201′), найдем

минимальное значение S, которое оказывается равным

$$R = (\mathbf{y} - A\hat{\theta})'(\mathbf{y} - A\hat{\theta}), \tag{201''}$$

$$R = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'A\hat{\theta} + \hat{\theta}'A'A\hat{\theta}, \tag{206'}$$

$$R = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'A(A'A)^{-1}A'\mathbf{y} + \mathbf{y}'A(A'A)^{-1}A'A(A'A)^{-1}A'\mathbf{y} =$$

$$= \mathbf{y}'\mathbf{y} - \mathbf{y}'A(A'A)^{-1}A'\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'A\hat{\theta}. \tag{210}$$

Говоря языком статистики, из исходной суммы квадратов $\mathbf{y}'\mathbf{y}$ вычтена сумма квадратов $\mathbf{y}'A\hat{\theta}$, полученная в результате аппроксимации θ. Из уравнения (210) следует, что остаточную сумму квадратов можно найти, не вычисляя самих остатков. К сожалению, в решение уравнения (210) округлением легко вносятся ошибки, так что часто приходится обращаться к исходному уравнению (201''). Это объясняется тем, что часто многие из p элементов $(\mathbf{y}'A)\,\hat{\theta}$ имеют один и тот же порядок величины, но разные знаки. При этом может оказаться, что на вычислительной машине, оперирующей с восьмизначными цифрами в десятичной системе, придется вычесть одно из другого два восьмизначных числа, у которых первые пять значащих цифр одинаковы. Полученная при этом разность будет содержать три значащие цифры.

Не зависящие от вида распределения свойства оценок наименьших квадратов

Полученная методом наименьших квадратов оценка $\hat{\theta}$ обладает тремя важными свойствами, которые не зависят от вида распределения ε.

1) $\hat{\theta}$ является несмещенной оценкой θ. Докажем это. Ранее было найдено, что

$$\hat{\theta} = (A'A)^{-1}A'\mathbf{y} \quad \text{и} \quad \mathscr{E}(\mathbf{y}) = A\theta.$$

Следовательно,

$$\mathscr{E}(\hat{\theta}) = (A'A)^{-1}A' \cdot \mathscr{E}(\mathbf{y}) = (A'A)^{-1}A'A \cdot \theta,$$

т. е.

$$\mathscr{E}(\hat{\theta}) = \theta. \tag{211}$$

2) Теорема Гаусса — Маркова. Среди класса оценок θ^* величины θ, которые а) являются несмещенными оценками и б) представляют собой линейные комбинации исходных данных $\mathbf{y}$, с помощью критерия наименьших квадратов можно найти такую оценку $\hat{\theta}$, что

$$\mathscr{D}(\hat{\theta}_j) \leqslant \mathscr{D}(\theta_j^*) \text{ для любого } j.$$

Иными словами, $\hat{\theta}$ есть наиболее точная оценка θ из всех возможных, принадлежащих данному классу.

Прежде всего необходимо найти выражение для $\mathscr{D}(\theta_j)$. Для вычисления дисперсий и *ковариаций* всех $\{\hat{\theta}_j\}$ удобно воспользоваться матричной алгеброй. Введем определение *ковариационной матрицы* (*матрицы ошибок*) векторной случайной величины $\mathbf{y}$ [1]):

$$\mathfrak{D}(\mathbf{y}) = \begin{vmatrix} \mathscr{D}(y_1) & \operatorname{Cov}(y_1, y_2) & \dots & \operatorname{Cov}(y_1, y_n) \\ \operatorname{Cov}(y_1, y_2) & \mathscr{D}(y_2) & \dots & \operatorname{Cov}(y_2, y_n) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \operatorname{Cov}(y_1, y_n) & \operatorname{Cov}(y_2, y_n) & \dots & \mathscr{D}(y_n) \end{vmatrix}_{(n \times n)}. \tag{212}$$

Предположим, что величины $\{y_i\}$ независимы и обладают одинаковыми дисперсиями, так что $\mathscr{D}(y_i) = \sigma^2$ (σ^2 неизвестно) и $\operatorname{Cov}(y_i, y_j) = 0$ при $i \neq j$. В этом случае

$$\mathfrak{D}(\mathbf{y}) = \sigma^2 I_{(n \times n)},$$

где $I_{(n \times n)}$ означает единичную матрицу с n строками и n столбцами.

ЛЕММА

$$\mathfrak{D}(\hat{\theta}) = \sigma^2 (A'A)^{-1}.$$

Доказательство. Мы имеем

$$\hat{\theta} = (A'A)^{-1} A' (\mathbf{y}). \tag{213}$$

Поскольку это выражение есть линейная комбинация $\{y_i\}$, воспользуемся правилом переноса ошибок: $\mathfrak{D}(W\mathbf{y}) = W\mathfrak{D}(\mathbf{y})W'$ с тем, чтобы точно вычислить $\mathfrak{D}(\theta)$. При

[1]) См. также стр. 197, 209. — *Прим. ред.*

этом получим

$$\mathfrak{D}(\hat{\theta}) = (A'A)^{-1} A' \mathfrak{D}(\mathbf{y}) A (A'A)^{-1} = \\ = \sigma^2 (A'A)^{-1} A'IA (A'A)^{-1},$$

откуда следует

$$\mathfrak{D}(\hat{\theta}) = \sigma^2 (A'A)^{-1}. \tag{214}$$

При этом нет необходимости проводить дополнительные вычисления, поскольку $(A'A)^{-1}$ мы уже нашли при вычислении $\hat{\theta}$.

Теперь перейдем к доказательству самой теоремы в современной трактовке, принадлежащей Плэкетту.

Пусть θ^* представляет собой любую иную оценку θ, являющуюся линейной комбинацией исходных данных, т. е.

$$\theta^*_{(p\times 1)} = U_{(p\times n)} \cdot \mathbf{y}_{(n\times 1)}, \tag{215}$$

в то время как, согласно формуле (213),

$$\hat{\theta} = (A'A)^{-1} A' (\mathbf{y}).$$

Потребуем, чтобы оценка θ^* была несмещенной, т. е.

$$\mathscr{E}(\theta^*) = U\mathscr{E}(\mathbf{y}) = UA\theta = \theta.$$

Таким образом, необходимо, чтобы

$$UA = I. \tag{216}$$

(Заметим, что $U \neq A^{-1}$, так как эти матрицы не являются квадратными.)

Найдем ковариационную матрицу для θ^*

$$\mathfrak{D}(\theta^*) = U\,\mathfrak{D}(\mathbf{y})\,U' = \sigma^2 UIU';$$

$$\mathfrak{D}(\theta^*) = \sigma^2 UU'. \tag{217}$$

Сравним UU' и $(A'A)^{-1}$, имея в виду соотношение (216). Введем обозначения

$$C = A'A \quad \text{и} \quad K = C^{-1} = (A'A)^{-1}.$$

Согласно формуле (214),

$$\mathfrak{D}(\hat{\theta}) = \sigma^2 K.$$

Мы можем записать UU' в форме, необходимой для доказательства теоремы:

$$UU' \equiv K + (U - KA')(U - KA')' \tag{218}$$

[используя формулу (216), правую часть данного выражения легко свести к виду $K - KA'U - UAK' + UU' + KA'AK' = K - K - K + UU' + K$]. Умножим обе части равенства (218) на σ^2, тогда

$$\mathfrak{D}(\theta^*) = \sigma^2 K + \sigma^2 (U - KA')(U - KA')' =$$
$$= \mathfrak{D}(\hat{\theta}) + \sigma^2 (U - C^{-1}A')(U - C^{-1}A')'.$$

Для $\hat{\theta}$ имеем $U = C^{-1}A'$, благодаря чему второй член последнего выражения обращается в нуль. Для любой другой оценки θ^* этот член неотрицателен, так что

$$\mathfrak{D}(\theta^*) \geqslant \mathfrak{D}(\hat{\theta}),$$

где знак $\geqslant$ означает, что каждый диагональный элемент матрицы $\mathfrak{D}(\theta^*)$ не меньше соответствующего диагонального элемента матрицы $\mathfrak{D}(\hat{\theta})$, что и требовалось доказать.

3) Оценка параметра σ^2. Мы приходим к третьему, не зависящему от распределения свойству оценки $\hat{\theta}$

$$\mathscr{E}(R) = (n - p)\sigma^2. \tag{219}$$

Важность данного свойства заключается в том, что для оценки σ^2 можно использовать остаточную сумму квадратов R, т. е.

$$\hat{\sigma}^2 = \frac{R}{n - p}. \tag{220}$$

Доказательство формулы (219) приведено в приложении I.

Случай исходных данных с неодинаковыми дисперсиями

В формуле (199) для ошибок

$$\varepsilon_i = y_i - \theta_0 - \theta_1 x_i - \ldots - \theta_r x_i^r \tag{221}$$

предполагалось, что $\mathscr{D}(\varepsilon_i) = \sigma^2$. Если же на самом деле $\mathscr{D}(\varepsilon_i) = \sigma^2/w_i$ (w_i — вес), то можно вновь прийти к урав-

нению типа (221), полагая

$$\varepsilon_i^* = \varepsilon_i \sqrt{w_i} = (y_i - \theta_0 - \theta_1 x_i - \ldots - \theta_r x_i^r) \sqrt{w_i}$$

для любого i от 1 до n. Тогда

$$\mathscr{D}(\varepsilon_i^*) = w_i \mathscr{D}(\varepsilon_i) = \sigma^2,$$

и можно, как и прежде, иметь дело с величиной $S = \sum (\varepsilon_i^*)^2 = \sum w_i \varepsilon_i^2$. В таком случае *нормальные уравнения* будут иметь вид

$$\theta_0 \sum w_i x_i^j + \theta_1 \sum w_i x_i^{j+1} + \ldots + \theta_r \sum w_i x_i^{j+r} = \sum w_i x_i^j y_i, \quad (222)$$

где $j = 0, 1, \ldots, r$. Решением этих уравнений опять будет решение максимального правдоподобия, если

$$\varepsilon_i \sim N\left(0; \frac{\sigma^2}{w_i}\right).$$

§ 20. Ортогональные полиномы

В § 16 мы успешно использовали метод наименьших квадратов для оценки параметров θ_0 и θ_1 в линейной модели

$$\eta(x, \theta) = \theta_0 + \theta_1 x. \quad (223)$$

Однако можно было бы произвести оценку Φ_0 и Φ_1 в модели

$$\xi(x, \Phi) = \Phi_0 + \Phi_1 (x - \bar{x}), \quad (224)$$

где $\bar{x} = \sum_{i=1}^{n} x_i/n$, а затем вычислить $\theta_0 = \Phi_0 - \bar{x}\Phi_1$ и $\theta_1 = \Phi_1$.

Оказывается, что в принципе значительно удобнее решать эту задачу с использованием модели (224). Найдем решение Φ методом наименьших квадратов. Для уравнения

$$\mathbf{y} = B\Phi + \boldsymbol{\varepsilon}$$

конструкционной является матрица

$$B = \begin{bmatrix} 1 & x_1 - \bar{x} \\ 1 & x_2 - \bar{x} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n - \bar{x} \end{bmatrix}.$$

Мы вновь имеем дело с *линейной моделью*, поскольку можно положить

$$f_0(x) = 1 \quad \text{и} \quad f_1(x) = x - \bar{x}.$$

Это означает, что для получения решения Φ следует воспользоваться уравнением (207). Метод наименьших квадратов приводит к решению

$$\hat{\Phi} = (B'B)^{-1} B'\mathbf{y},$$

где

$$B'B = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 - \bar{x} & x_2 - \bar{x} & \dots & x_n - \bar{x} \end{bmatrix} \cdot \begin{bmatrix} 1 & x_1 - \bar{x} \\ 1 & x_2 - \bar{x} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n - \bar{x} \end{bmatrix} =$$

$$= \begin{bmatrix} n & \sum x_i - n\bar{x} \\ \sum x_i - n\bar{x} & \sum (x_i - \bar{x})^2 \end{bmatrix} = \begin{bmatrix} n & 0 \\ 0 & \sum (x_i - \bar{x}^2) \end{bmatrix}.$$

Тогда

$$(B'B)^{-1} = \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\sum (x_i - \bar{x})^2} \end{bmatrix}.$$

Здесь следует отметить то обстоятельство, что матрица $(B'B)$ является диагональной и, следовательно, может быть легко и точно обращена без ошибок, вызванных округлением. Эта отличительная черта оказывается особенно важной в тех случаях, когда проводится аппроксимация полиномом более высокой степени. *Обычный* полином в общем случае имеет вид

$$\eta(x, \theta) = \theta_0 + \theta_1 x + \dots + \theta_r x^r. \tag{225}$$

Такая модель приводит к матрице $(A'A)$:

$$A'A = \begin{bmatrix} n & \sum x_i & \sum x_i^2 & \dots & \sum x_i^r \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \dots & \sum x_i^{r+1} \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_i^r & \sum x_i^{r+1} & \sum x_i^{r+2} & \dots & \sum x_i^{2r} \end{bmatrix}, \tag{226}$$

плохо обусловленной при больших r. Более того, плохая обусловленность может еще *ухудшиться* с возрастанием объема выборки; в итоге при вычислении $(A'A)$ могут возникнуть значительные ошибки, вызванные округлением. Почему это так, можно видеть из следующего грубого рассмотрения.

Не нарушая общности, можно считать, что все $\{x_i\}$ заключены в интервале от 0 до 1. Допустим, что множество n $\{x_i\}$ равномерно расширяется при $n \to \infty$. Тогда

$$\sum x_i^r \to n \int_0^1 x^r\, dx = \frac{n}{r+1}$$

и, следовательно,

$$A'A \approx n \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{r+1} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{r+2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{r+1} & \frac{1}{r+2} & \frac{1}{r+3} & \cdots & \frac{1}{2r+1} \end{bmatrix} = nH_{r+1},$$

где H_p — гильбертова матрица ранга p, которая, как известно, плохо обусловлена при больших p.

Линейную модель с ортогональными полиномами, эквивалентную модели (225) с обычными полиномами, можно получить, если произвести специальную перегруппировку в правой части (225). Пусть

$$\xi(x, \Phi) = \Phi_0 + \Phi_1 G_1(x) + \Phi_2 G_2(x) + \ldots + \Phi_r G_r(x) \quad (227)$$
$$(\equiv \theta_0 + \theta_1 x + \ldots + \theta_r x^r),$$

где полиномы

$$G_j(x) = k_{j0} + k_{j1} x + \ldots + k_{j,\,j-1} x^{j-1} + x^j. \quad (228)$$

Коэффициенты k_{j0}, k_{j1}, ..., $k_{j,\,j-1}$ (их число равно j) определяются j соотношениями

$$\left.\begin{aligned} \sum_i G_0(x_i)\, G_j(x_i) &= 0, \\ \sum_i G_1(x_i)\, G_j(x_i) &= 0, \\ \cdots\cdots\cdots\cdots & \\ \sum_i G_{j-1}(x_i)\, G_j(x_i) &= 0, \end{aligned}\right\} \quad (229)$$

причем $G_0(x) = 1$. Ортогональные полиномы выбраны с таким расчетом, чтобы недиагональные элементы матрицы $A'A$, которую следует обратить, оказались равны нулю. [Недиагональные элементы следовало бы заново вычислять для каждого различного набора $\{x_i\}$ в отличие от случая ортогональных полиномов, для которых $\int_0^1 G_j(x) G_m(x)\, dx = 0$, если $j \neq m$.]

Оказывается, что вычисление ортогональных полиномов с последующим обращением диагональной матрицы

$$B'B = \begin{bmatrix} \sum G_0^2(x_i) & \sum G_0(x_i) G_1(x_i) & \dots & \sum G_0(x_i) G_r(x_i) \\ \sum G_1(x_i) G_0(x_i) & \sum G_1^2(x_i) & \dots & \sum G_1(x_i) G_r(x_i) \\ \dots & \dots & \dots & \dots \\ \sum G_r(x_i) G_0(x_i) & \sum G_r(x_i) G_1(x_i) & \dots & \sum G_r^2(x_i) \end{bmatrix} =$$

$$= \begin{bmatrix} \sum G_0^2(x_i) & 0 & \dots & 0 \\ 0 & \sum G_1^2(x_i) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sum G_r^2(x_i) \end{bmatrix}_{(r+1)\times(r+1)} \qquad (230)$$

приводит к меньшим ошибкам округления, чем при вычислении обычных полиномов, хотя обе процедуры алгебраически эквивалентны. Кроме того, из формулы (230) видно, что не представляет труда сравнить аппроксимации полиномами степеней r и $r+1$; новая система *мнк*-уравнений, выраженных через ортогональные полиномы, будет содержать матрицу $(B_+'B_+)$, которую следует обратить для получения решения $\Phi_+' = (\Phi', \Phi_{r+1})_{1\times(r+2)}$. При этом

$$B_+'B_+ = \begin{bmatrix} B'B & \vdots & 0 \\ \dots & \dots & \dots \\ 0 & \vdots & \sum G_{r+1}^2(x_i) \end{bmatrix}, \qquad (231)$$

$$(B_+'B_+)^{-1} = \begin{bmatrix} (B'B)^{-1} & \vdots & 0 \\ \dots & \dots & \dots \\ 0 & \vdots & [\sum G_{r+1}^2(x_i)]^{-1} \end{bmatrix}_{(r+2)\times(r+2)}. \qquad (232)$$

Если для сравнения привести соответствующую матрицу для случая обычных полиномов [см. (226)], то она имеет вид

$$A'_+A_+ = \begin{bmatrix} & & & \vdots & \sum x_i^{r+1} \\ & & & \vdots & \sum x_i^{r+2} \\ & A'A & & \cdot & \\ & & & \cdot & \\ & & & & \sum x_i^{2r+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum x_i^{r+1} & \sum x_i^{r+2} & \ldots \sum x_i^{2r+1} & \vdots & \sum x_i^{2r+2} \end{bmatrix}_{(r+2)\times(r+2)}.$$

Вычисление $(A'_+A_+)^{-1}$, очевидно, оказывается значительно более трудоемким.

Наконец, ниже будет показано, что в том случае, когда из экспериментальных данных необходимо определить степень r аппроксимирующего полинома, использование ортогональных полиномов приводит к чрезвычайно ясному статистическому анализу.

Рекуррентное соотношение Форсайта

Преобразовать форму

$$\eta(x, \theta) = \theta_0 + \theta_1 x + \ldots + \theta_r x^r$$

в линейную комбинацию ортогональных полиномов (227)

$$\xi(x, \Phi) = \Phi_0 + \Phi_1 G_1(x) + \ldots + \Phi_r G_r(x)$$

можно с помощью решения системы совместных уравнений (229). В этом заключается метод ортогонализации Грама — Шмидта. Предложенный Форсайтом метод аналогичен предыдущему, однако значительно проще.

Для упрощения дальнейших выкладок будем оперировать ортогональными нормированными полиномами $Q_j(x)$, такими, что $\sum_i Q_j^2(x_i) = 1$. [Аналогичные вычисления можно было бы провести и с полиномами $G_j(x)$.] Итак, мы имеем

$$Q_0(x) = \frac{1}{\sqrt{n}},$$

и в общем виде

$$Q_j(x) = \frac{G_j(x)}{\left[\sum G_j^2(x)\right]^{1/2}}, \tag{233}$$

или

$$Q_j(x) = q_{j0} + q_{j1}x + \ldots + q_{jj}x^j.$$

Пусть теперь линейная форма, выраженная через ортонормированные полиномы, имеет вид

$$\psi_r(x, \omega) = \omega_0 Q_0(x) + \omega_1 Q_1(x) + \ldots + \omega_r Q_r(x). \tag{234}$$

Чтобы вычислить $G_j(x)$, необходимо было совместно решить j уравнений. Соотношение Форсайта устанавливает следующую связь между ортонормированными полиномами:

$$\lambda Q_j(x) = xQ_{j-1}(x) - \alpha Q_{j-1}(x) - \beta Q_{j-2}(x), \tag{235}$$

где константы α и β можно определить из уравнений

$$\alpha = \sum_{i=1}^{n} x_i Q_{j-1}^2(x_i)$$

и

$$\beta = \sum_{i=1}^{n} x_i Q_{j-1}(x_i) Q_{j-2}(x_i),$$

и тогда константа λ определится из условия

$$\sum_{i=1}^{n} Q_j^2(x_i) = 1.$$

Работа [14], в которой содержится доказательство приведенного выше рекуррентного соотношения, посвящена весьма доступному описанию метода аппроксимации результатов.

Применяя метод наименьших квадратов к линейной форме (234), можно получить следующие очень простые результаты:

1) $$\hat{\omega}_j = \sum_{i=1}^{n} y_i Q_j(x_i). \tag{236}$$

В справедливости этого соотношения можно убедиться, сравнив (225) и (234). В первом случае было получено

решение

$$\hat{\theta} = (A'A)^{-1} A'y.$$

Для ортонормированных полиномов $A'A = I$ и поэтому

$$\hat{\omega} = A'y,$$

причем

$$A = \begin{bmatrix} Q_0(x_1) & Q_1(x_1) & \dots & Q_r(x_1) \\ Q_0(x_2) & Q_1(x_2) & \dots & Q_r(x_2) \\ \dots & \dots & \dots & \dots \\ Q_0(x_n) & Q_1(x_n) & \dots & Q_r(x_n) \end{bmatrix}. \tag{237}$$

2) $$\mathscr{D}(\hat{\omega}_j) = \sigma^2. \tag{238}$$

Докажем это соотношение. На основании формулы (236) имеем

$$\mathscr{D}(\hat{\omega}_j) = \sum_{i=1}^{n} Q_j^2(x_i)\, \mathscr{D}(y_i) = \sigma^2 \sum_{i=1}^{n} Q_j^2(x_i) = \sigma^2.$$

3) Аппроксимирующий полином степени r имеет вид

$$\hat{y}_r = \sum_{j=0}^{r} \hat{\omega}_j Q_j(x). \tag{239}$$

Его можно вновь привести к виду (225), выписывая множители

$$\hat{\omega}_j Q_j(x) = \hat{\omega}_j (q_{j0} + q_{j1}x + \dots + q_{jj}x^j)$$

и затем группируя коэффициенты при x^m, что дает

$$\hat{\theta}_m = \hat{\omega}_r q_{rm} + \hat{\omega}_{r-1} q_{r-1,\,m} + \dots + \hat{\omega}_m q_{mm}.$$

4) Остаточная сумма квадратов при аппроксимации полиномами степени r равна

$$R_r = \sum_{i=1}^{n} y_i^2 - \sum_{j=0}^{r} \hat{\omega}_j^2. \tag{240}$$

Действительно, из формулы (210) следует

$$R_r = \sum_{i=1}^{n} y_i^2 - (y_1, y_2, \dots, y_n) A \begin{bmatrix} \hat{\omega}_0 \\ \hat{\omega}_1 \\ \vdots \\ \hat{\omega}_r \end{bmatrix}.$$

Подставив вместо A его выражение (237), получим

$$(y_1, y_2, \ldots, y_n)\,A = [\sum y_i Q_0(x_i), \sum y_i Q_1(x_i), \ldots, \sum y_i Q_r(x_i)] = \\ = [\hat{\omega}_0, \hat{\omega}_1, \ldots, \hat{\omega}_r].$$

5) Оценкой σ^2 является

$$\hat{\sigma}^2 = \frac{R_r}{n-r-1}. \qquad (241)$$

6) Можно, кроме того, исследовать отклонение от $\hat{y}_r(x)$ при повторных экспериментах. Для этого вычислим

$$\mathscr{D}[\hat{y}_r(x)] = \sum_{j=0}^{r} Q_j^2(x)\,\mathscr{D}(\hat{\omega}_j) = \sigma^2 \sum_{j=0}^{r} Q_j^2(x). \qquad (242)$$

Поскольку $\hat{y}_r(x)$ является оценкой для $\mathscr{E}[y(x)]$, дисперсия (242) служит мерой точности оценки $\eta(x) = \mathscr{E}[y(x)]$ для любого заданного x. Результаты 3, 5 и 6 удобно записать в виде единой формулы

$$\mathscr{E}[y(x)] = \hat{y}_r(x) \pm \hat{\sigma}\,[\sum_{i=0}^{r} Q_j^2(x_i)]^{1/2} \qquad (243)$$

или

$$\mathscr{E}[y(x)] = \hat{y}_r(x) \pm 2\hat{\sigma}\,[\sum_{i=0}^{r} Q_j^2(x_i)]^{1/2} \qquad (244)$$

в зависимости от желания указать одно или два средних квадратичных отклонения.

Увеличение степени полинома

Может понадобиться проверить, не улучшается ли аппроксимация при увеличении на единицу степени полинома. Очевидно, для этой цели очень удобно иметь дело с ортонормированными полиномами.

Найдем сначала

$$Q_{r+1}(x) = [xQ_r(x) - \alpha Q_r(x) - \beta Q_{r-1}(x)]\,\frac{1}{\lambda},$$

где

$$\alpha = \sum_{i=1}^{n} x_i Q_r^2(x_i),$$

$$\beta = \sum_{i=1}^{n} x_i Q_r(x_i)\,Q_{r-1}(x_i)$$

и константа $\hat{\lambda}$ выбрана так, чтобы выполнялось равенство

$$\sum_{i=1}^{n} Q_{r+1}^2(x_i) = 1.$$

Тогда имеем

$$\left.\begin{aligned} \hat{\omega}_{r+1} &= \sum_{1}^{n} y_i Q_{r+1}(x_i), \\ \hat{y}_{r+1}(x) &= \hat{y}_r(x) + \hat{\omega}_{r+1} Q_{r+1}(x), \\ R_{r+1} &= R_r - \hat{\omega}_{r+1}^2, \\ \hat{\sigma}^2 &= \frac{R_{r+1}}{n-r-2}, \\ \frac{\mathscr{D}(\hat{y}_{r+1})}{\sigma^2} &= \frac{\mathscr{D}(\hat{y}_r)}{\sigma^2} + Q_{r+1}^2(x). \end{aligned}\right\} \qquad (245)$$

Улучшение согласия характеризуется вычитанием $\hat{\omega}_{r+1}^2$ из остаточной суммы квадратов. Оно достигается ценой возрастания дисперсии аппроксимирующего полинома на величину $\sigma^2 Q_{r+1}^2(x)$.

Как правило, существует критическое значение $r = k$, соответствующее первым $(k+1)$ членам разложения в ряд Тэйлора функции $\eta(x)$. Это значение k определяет оптимальную аппроксимацию. Если попытаться провести аппроксимацию полиномом степени $r \geqslant k+1$, то можно, конечно, получить лучшее согласие $\hat{y}_r(x)$ с *экспериментальными данными*, но при этом неизбежно получается худшее согласие $\hat{y}_r(x)$ с *истинной кривой* $\eta(x)$, которую мы пытаемся оценить. Следует подчеркнуть, что экспериментальные данные представляют собой случайные величины и содержат лишь ограниченную информацию, которую можно использовать для оценки $\eta(x)$.

Перефразируем этот важный результат следующим образом. Если бы функция $\eta(x)$ была нам известна, то, повышая степень полинома, можно было бы тем самым, безусловно, увеличить точность разложения в ряд Тэйлора (т. е. полинома). Однако когда мы попытаемся использовать точки $\{y(x)\}$, для которых

$$\mathscr{P}[y(x) = \eta(x)] = 0,$$

то это следует делать до тех пор, пока извлекается из них надежная информация относительно $\eta(x)$. В данном случае мы должны продолжать оценку величины $Q_r x^r$ при возрастании r лишь постольку, поскольку статистические критерии указывают на *значимость* величины Q_r. Говоря точнее, повышать степень полинома следует лишь до тех пор, пока $\hat{\omega}_r$ еще *значимо* отлично от нуля. В этом смысле процедура аппроксимации полиномами основана на своего рода «асимптотическом» разложении, которое с ростом числа членов до некоторого момента постепенно улучшается, а затем ухудшается.

В следующем параграфе изложен метод выбора критического значения k в случае нормально распределенных ошибок, а также рассмотрено некое видоизменение обычной процедуры, гарантирующее определенные меры предосторожности.

§ 21. Нормальный регрессионный анализ

Полученные выше *мнк*-результаты не зависят от вида распределения ошибок ε. Теперь допустим, что

$$\varepsilon \sim N(0, \sigma^2 I), \tag{246}$$

т. е. что ошибки ε_i подчиняются независимым, одинаковым нормальным распределениям с нулевым средним значением и дисперсией σ^2. Мы может вычислить плотность вероятности для любых функций исходных данных, которые может оказаться необходимым подвергнуть статистическим проверкам. Так,

$$\hat{\omega}_j = \sum_{i=1}^{n} Q_j(x_i)\, y_i$$

есть линейная комбинация нормально распределенных величин $\{y_i\}$; следовательно, ω_j распределено по нормальному закону. Мы уже видели, что *мнк*-оценки всегда несмещенные, т. е. $\mathscr{E}(\hat{\omega}_j) = \omega_j$, и, кроме того, $\mathscr{D}(\hat{\omega}_j) = \sigma^2$. Отсюда вытекает

$$\hat{\omega}_j \sim N(\omega_j, \sigma^2). \tag{247}$$

Если $\omega_j = 0$, то

$$\hat{\omega}_j^2 \sim \sigma^2 \chi_1^2, \tag{248}$$

Если бы мы знали σ^2, то для проверки гипотезы $\omega_j = 0$ можно было бы воспользоваться распределением χ_1^2, так как

$$\frac{\hat{\omega}_j^2}{\sigma^2} \sim \chi_1^2 \quad \text{при} \quad \omega_j = 0. \tag{249}$$

Когда σ^2 неизвестна, используется ее оценка с ν степенями свободы

$$\hat{\sigma}_\nu^2 = \frac{R_r}{n-r-1}, \tag{250}$$

где

$$R_r = \sum_1^n y_i^2 - \sum_0^r \hat{\omega}_j^2 \tag{251}$$

и

$$\nu = n - r - 1.$$

Необходимо найти распределение $\hat{\sigma}^2$. Мы сделаем это, прибегнув к простой математической уловке. Хорошо известно, что n точек $\{y_i\}$ можно аппроксимировать полиномом $(n-1)$-й степени. В этом случае для каждого i получим $\hat{y}_i = y_i$. [Как было отмечено ранее, если $\{y_i\}$ — случайные величины, то подобная кривая скорее всего не имела бы ничего общего с истинной кривой $\eta(x)$.] В каждой точке i остаток равен нулю, $y_i - \hat{y} = 0$, а значит, и остаточная сумма квадратов равна нулю:

$$R_{n-1} = \sum_1^n y_i^2 - \sum_0^r \hat{\omega}_j^2 = 0. \tag{252}$$

Сравнивая это выражение с формулой (251), получаем

$$R_r = \sum_1^n y_i^2 - \sum_0^r \hat{\omega}_j^2 = R_{n-1} + \sum_{r+1}^{n-1} \hat{\omega}_j^2 = \sum_{r+1}^{n-1} \hat{\omega}_j^2. \tag{253}$$

Таким образом, мы получили сумму квадратов величин, распределенных по нормальному закону. Это означает, что

$$R_r \sim \sigma^2 \chi_\nu^2 \tag{254}$$

при условии, что

$$\omega_{r+1} = 0 = \omega_{r+2} = \ldots = \omega_{n-1}.$$

Из формулы (250) следует

$$\hat{\sigma}_\nu^2 \sim \sigma^2 \frac{\chi_\nu^2}{\nu}, \tag{255}$$

где $\nu = n - r - 1$. Из формулы (249) следует

$$\hat{\omega}_j^2 \sim \sigma^2 \chi_1^2 \text{ (для } j = 0, 1, \ldots, r),$$

но поскольку $\hat{\sigma}_\nu^2$ есть функция от $\omega_{r+1}^2, \ldots, \omega_{n-1}^2$, то $\hat{\sigma}_\nu^2$ и $\hat{\omega}_j^2$ *независимы* (для $j = 0, 1, \ldots, r$).

Тогда

$$\frac{\hat{\omega}_j^2}{\hat{\sigma}_\nu^2} \sim \mathcal{F}_{1,\nu}, \quad \text{если} \quad \omega_j = 0 \tag{256}$$

[согласно определению распределения $\mathcal{F}$, данному на стр. 97, $\hat{\omega}_j^2/\hat{\sigma}_\nu^2 \sim (\sigma^2\chi_1^2/1)/(\sigma^2\chi_\nu^2/\nu) = \mathcal{F}_{1,\nu}$]. Кроме того, из формулы (247) вытекает $\hat{\omega}_j^2 \sim \omega_j^2 + \sigma^2\chi_1^2$, если $\omega_j \neq 0$. В таком случае для проверки гипотезы $\omega_j = 0$ можно использовать выражение (256). В частности, если мы уже аппроксимировали результаты с помощью

$$\hat{y}_r = \hat{\omega}_0 Q_0(x) + \ldots + \hat{\omega}_r Q_r(x)$$

и хотим узнать, не следует ли *расширить* эту модель, превратив ее в полином степени $(r + 1)$, то следует проверить гипотезу $\omega_{r+1} = 0$ [1]).

Мера предосторожности

Всегда может оказаться, что $\omega_{r+1} = 0$, но $\omega_{r+2} \neq 0$. Согласно изложенной выше в общих чертах процедуре, мы должны были бы остановиться на величине ω_r. Чтобы избежать подобной случайности, следует сначала рассмотреть полином более высокой степени, например $r = 6$,

[1]) Аналогичным образом проверяется гипотеза $\theta_{r+1} = 0$ и в случае линейной модели (225), часто используемой на практике. В этом случае можно воспользоваться статистикой

$$\frac{R_r - R_{r+1}}{R_{r+1}}(n - r - 1) \sim \mathcal{F}_{1,\, n-r-1},$$

где $R_r = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2$, w_i — вес, а $\hat{y}_i$ дается формулой (209). Если отношение незначимо, то коэффициент θ_{r+1} при x^{r+1} следует положить равным нулю и проверить на значимость коэффициент θ_r при x^r. Если отношение значимо, то целесообразно проверить гипотезу $\theta_{r+2} = 0$ и т. д. Подробнее см. дополнение II. — *Прим. ред.*

а затем проверить, нельзя ли свести форму

$$\mathscr{E}[y_6(x)] = \omega_0 Q_0(x) + \ldots + \omega_5 Q_5(x) + \omega_6 Q_6(x)$$

к

$$\mathscr{E}[y_5(x)] = \omega_0 Q_0(x) + \ldots + \omega_5 Q_5(x).$$

При этом в качестве статистики, лежащей в основе критерия, используется величина

$$\mathscr{F}_{1,\,n-7} = \frac{\hat{\omega}_6^2}{R_6/(n-7)},$$

где $R_6 = \sum_1^n [y_i - \hat{\omega}_0 Q_0(x) - \ldots - \hat{\omega}_6 Q_6(x)]^2$.

Если окажется, что $\hat{\omega}_6$ незначимо отлично от нуля, то мы исключаем из модели член, содержащий $\hat{\omega}_6$, и затем подвергаем проверке величину

$$\mathscr{F}_{1,\,n-6} = \frac{\hat{\omega}_5^2}{R_5/(n-6)}$$

и т. д.

Пример. Рассмотрим набор данных (x_i, y_i), полученных следующим путем: к 19 значениям ординаты кубической параболы

$$\mathscr{E}[y(x)] = 1 + x - 0{,}055x^2 + 0{,}001x^3$$

были добавлены случайные отклонения, обладающие нормальным распределением с единичной дисперсией. Эти результаты представлены в табл. 10.

Таблица 10

i	x_i	$\mathscr{E}[y(x_i)]$	y_i	i	x_i	$\mathscr{E}[y(x_i)]$	y_i
1	2	2,79	2,84	11	22	7,03	7,35
2	4	4,18	5,50	12	24	7,14	6,11
3	6	5,24	5,96	13	26	7,40	6,67
4	8	5,99	4,50	14	28	7,83	9,67
5	10	6,50	6,45	15	30	8,50	7,35
6	12	6,81	7,39	16	32	9,45	9,99
7	14	6,96	6,67	17	34	10,72	10,31
8	16	7,02	5,72	18	36	12,38	12,03
9	18	7,01	7,95	19	38	14,45	13,51
10	20	7,00	5,93				

Для решения задачи использовалась программа вычислительной машины «Меркурий». Ортогональные полиномы $\{G_j(x)\}$ вычислялись в ненормированном виде, так что в действительности

$$\hat{\omega}_j Q_j(x) = \hat{c}_j G_j(x),$$

что, однако, не вносило различия в анализ. Основными числами первой цифропечати являлись величины отношения $\mathcal{F}$. Согласно статистическим таблицам [7] распределения $\mathcal{F}$, 95%-ный уровень для $\mathcal{F}_{1,12}$ соответствует значению 4,75[1]). Как видно из табл. 11, коэффициенты при шестой, пятой и четвертой степенях полинома оказались незначимыми, т. е. $\omega_6 = \omega_5 = \omega_4 = 0$. В качестве наилучшей оценки было принято выражение

$$\hat{y}_3(x) = 2{,}200 + 0{,}744x - 0{,}04148x^2 + 0{,}0007862x^3.$$

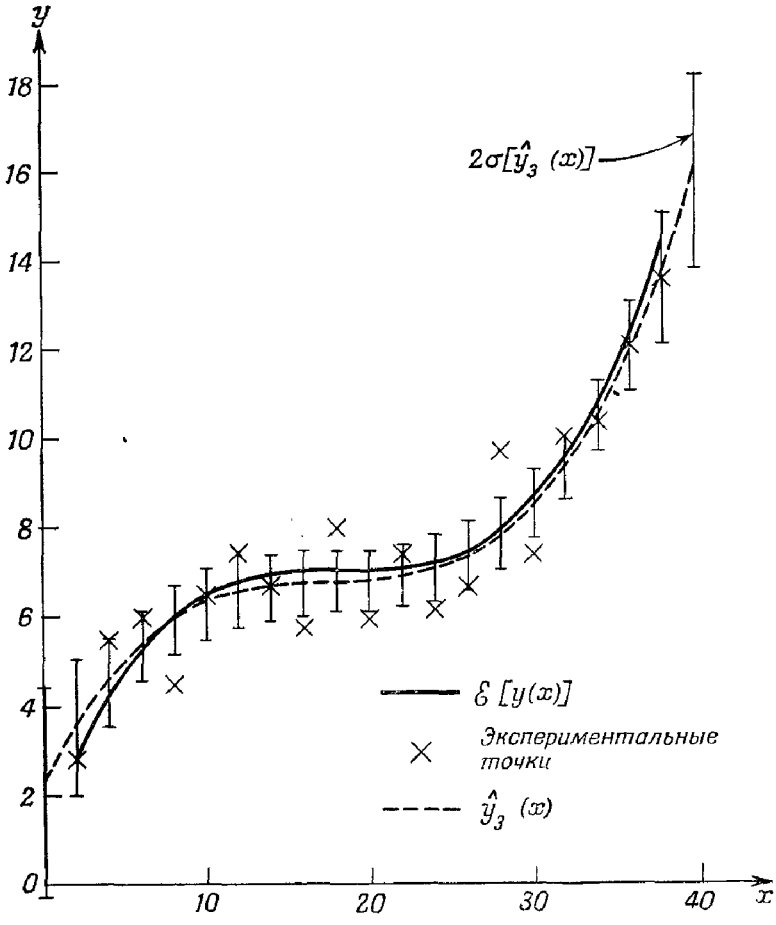
Затем та же программа использовалась вновь для аппроксимации исходных данных полиномом третьей степени и вывода на цифропечать необходимой информации относительно ошибок и т. д. В частности, были получены аппроксимирующие значения $\{\hat{y}_3(x_i)\}$ и в качестве

Таблица 11

Максимальная степень полинома j	$\mathcal{F}_{1,12}$
0	907,1
1	76,3
2	6,0
3	10,4
4	0,3
5	0,0
6	0,1

ошибок удвоенное среднеквадратичное отклонение для $\{\hat{y}_3(x_i)\}$. На фиг. 29 приведены аппроксимирующая кривая $\hat{y}_3(x)$ и истинная кривая $\mathscr{E}[y(x)]$.

[1]) См. табл. VII, стр. 284.—*Прим. ред.*



Ф и г. 29.

Следует отметить, что оценка величины $\sigma = 1$ при числе степеней свободы 15 оказалась равной $\hat{\sigma} = 0{,}982$.

Библиография

Укажем прежде всего на доступное изложение метода наименьших квадратов в книге Уизерборна [37]. Весьма обширное изложение стандартных приемов регрессионного анализа содержится в книге Дэвиса [8], стр. 150.

Одной из типичных монографий на эту же тему является книга Плэкетта [33]. Изложение охватывает случай, когда ошибки имеют разные веса и, кроме того, коррелированы. Подробно рассмотрены примеры применения регрессионного анализа во втором томе книги Кендалла и Стьюарта [21].

В более современных задачах требуется аппроксимировать результаты нелинейной формой. При этом обычно используется линеаризация посредством разложения в ряд Тэйлора. Регрессионные уравнения в этом случае решаются методом итераций, описанным, например, в книге Вильямса [39]. Там же приведены довольно необычные примеры из области лесоводства.

В ряде задач параметры связаны между собой некоторым уравнением, так что число неизвестных параметров можно уменьшить на единицу. Такое уравнение также может быть нелинейным относительно параметров. Подробно этот вопрос рассмотрен в гл. 3 обзора Жоно и Морелля [19].

Если корреляция между ошибками является функцией порядка, в котором были получены экспериментальные данные, то такие данные называют «временным рядом». Современному изложению этого вопроса посвящена книга Гренандера и Розенблатта [18].

ПРИЛОЖЕНИЕ I

Математическое ожидание остаточной суммы квадратов

Докажем, что

$$\mathscr{E}(R) = (n-p)\,\sigma^2. \tag{219}$$

Сначала докажем общую лемму о некоторых квадратичных формах, а затем применим ее к случаю квадратичной формы R.

ЛЕММА. Пусть $Q = \mathbf{y}'U\mathbf{y}$ есть квадратичная форма n случайных величин $\{y_i\}$, а U — квадратная матрица ранга n. Допустим также, что $\boldsymbol{\eta} = \mathscr{E}(\mathbf{y})$ и $\mathfrak{D}(\mathbf{y}) = \sigma^2 I_{n\times n}$. Тогда

$$\mathscr{E}(Q) = \sigma^2 \operatorname{Sp} U + \boldsymbol{\eta}' U \boldsymbol{\eta}$$

(след матрицы U: $\operatorname{Sp} U$ есть сумма ее диагональных элементов).

Доказательство. Обозначим

$$U = \| u_{ij} \|.$$

Тогда

$$\mathbf{y}'U\mathbf{y} = \sum_{i=1}^{n}\sum_{j=1}^{n} u_{ij} y_i y_j =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} u_{ij}\{y_i y_j - \mathscr{E}(y_i)\,\mathscr{E}(y_j)\} + \sum_{i=1}^{n}\sum_{j=1}^{n} u_{ij}\mathscr{E}(y_i)\,\mathscr{E}(y_j).$$

Рассмотрим первую из этих сумм. При $i = j$ имеем

$$\mathscr{E}(y_i^2) - [\mathscr{E}(y_i)]^2 = \mathscr{D}(y_i) = \sigma^2.$$

При $i \neq j$ имеем

$$\mathscr{E}(y_i y_j) - \mathscr{E}(y_i)\,\mathscr{E}(y_j) = \operatorname{Cov}(y_i, y_j).$$

Согласно принятому предположению,

$$\operatorname{Cov}(y_i, y_j) = 0.$$

Вторая сумма представляет собой константу. Следовательно,

$$\mathscr{E}(Q) = \sum_{i=1}^{n} u_{ii}\sigma^2 + [\mathscr{E}(\mathbf{y})]' U [\mathscr{E}(\mathbf{y})] = \sigma^2 \operatorname{Sp} U + \boldsymbol{\eta}' U \boldsymbol{\eta}.$$

Теперь изложим доказательство соотношения (219), принадлежащее Плэкетту. Согласно формуле (210) для остаточной суммы квадратов

$$R = \mathbf{y}'\mathbf{y} - \mathbf{y}' A' \hat{\boldsymbol{\theta}}.$$

Перепишем это выражение в виде

$$R = (\mathbf{y} - A\boldsymbol{\theta})' (I - AC^{-1}A') (\mathbf{y} - A\boldsymbol{\theta}),$$

где члены, содержащие $\boldsymbol{\theta}$, сокращаются, а $C = A'_{p\times n} A_{n\times p}$ — квадратная матрица и ее ранг равен p.

Можно проверить правильность последней записи для R, проведя умножение в правой части:

$$\begin{aligned}
&\mathbf{y}'\mathbf{y} - \mathbf{y}'A\boldsymbol{\theta} - \mathbf{y}'AC^{-1}A'\mathbf{y} + \mathbf{y}'AC^{-1}A'A\boldsymbol{\theta} - \boldsymbol{\theta}'A'\mathbf{y} + \\
&\quad + \boldsymbol{\theta}'A'A\boldsymbol{\theta} + \boldsymbol{\theta}'A'AC^{-1}A'\mathbf{y} - \boldsymbol{\theta}'A'AC^{-1}A'A\boldsymbol{\theta} = \\
&\quad = \mathbf{y}'\mathbf{y} - \mathbf{y}'A\boldsymbol{\theta} - \mathbf{y}'A\hat{\boldsymbol{\theta}} + \mathbf{y}'A\boldsymbol{\theta} - \boldsymbol{\theta}'A'\mathbf{y} + \boldsymbol{\theta}'C\boldsymbol{\theta} + \\
&\quad + \boldsymbol{\theta}'A'\mathbf{y} - \boldsymbol{\theta}'C\boldsymbol{\theta} = \mathbf{y}'\mathbf{y} - \mathbf{y}'A\hat{\boldsymbol{\theta}}.
\end{aligned}$$

Итак, справедливость этого выражения доказана. Далее допустим, что

$$U = I - AC^{-1}A'.$$

Тогда

$$\mathscr{E}(\mathbf{y} - A\boldsymbol{\theta}) = 0$$

и

$$\mathfrak{D}(\mathbf{y} - A\boldsymbol{\theta}) = \mathfrak{D}(\mathbf{y}) = \sigma^2 I_{n\times n}.$$

Чтобы применить доказанную выше лемму, необходимо вычислить

$$\operatorname{Sp} U = \operatorname{Sp}(I_{n\times n} - A_{n\times p} C^{-1}_{p\times p} A'_{p\times n}) = \operatorname{Sp} I_{n\times n} - \operatorname{Sp}(AC^{-1}A').$$

Но

$$\begin{aligned}
\operatorname{Sp}(AC^{-1}A') &= \operatorname{Sp}(C^{-1}A'A) = \\
&= \operatorname{Sp} I_{p\times p} = p.
\end{aligned}$$

Значит,

$$\operatorname{Sp} U = n - p.$$

Следовательно,

$$\mathscr{E}(R) = \sigma^2 (n - p) + 0'U0 = (n - p)\sigma^2,$$

что и требовалось доказать.

ПРИЛОЖЕНИЕ II

Получение выборки случайных величин с заданной плотностью вероятности

Задача заключается в получении случайной величины x, являющейся выборкой из распределения с плотностью $f(x)$; при этом имеется набор случайных чисел $\{r\}$, которые равномерно распределены в интервале от 0 до 1.

ТЕОРЕМА. Пусть величины r представляют собой выборку из генеральной совокупности с плотностью распределения

$$p(r) = 1 \quad \text{при} \quad 0 \leqslant r \leqslant 1.$$

Определим функцию $x = x(r)$, такую, что

$$r = \int_{-\infty}^{x(r)} f(t)\, dt.$$

Тогда $x(r)$ подчиняется требуемому распределению.

Доказательство. Согласно уравнению (80), новая плотность распределения величины x есть

$$g(x)\, dx = p[r(x)] \left| \frac{dr(x)}{dx} \right| dx.$$

Поэтому прежде всего необходимо произвести обратное преобразование, выразив r через x. Очевидно,

$$r(x) = \int_{-\infty}^{x} f(t)\, dt$$

и

$$\frac{dr(x)}{dx} = f(x).$$

Кроме того, поскольку $p(r) = 1$, то и $p[r(x)] = 1$, ибо правая часть не содержит r. Следовательно,

$$g(x)\, dx = 1 \cdot |f(x)|\, dx,$$

т. е.

$$g(x) = f(x).$$

Пример 1. Пусть

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Случайную величину, распределенную по закону $N(0, 1)$, мы получим, решая методом итераций уравнение

$$r = \int_{-\infty}^{x(r)} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

Пример 2. Допустим, что необходимо получить значение случайной величины $w \sim \chi^2_{40}$. Для $w < 0$ имеем $f(w) = 0$, а для $w \geqslant 0$ [см. (86)]

$$f(w) = 2^{-20} w^{19} e^{-w/2} \frac{1}{19!}.$$

Предположим, что из таблицы случайных чисел получено $r = 0{,}4176$. Тогда w определяется из уравнения

$$0{,}4176 = \int_0^w 2^{-20} t^{19} e^{-t/2} \frac{dt}{19!}.$$

И если теперь воспользоваться статистическими таблицами для распределения χ^2 [32], то мы получим

$$w = 37{,}52,$$

что уже было использовано на стр. 128.

Дополнения

I

ДЖ. МАЛВИ

Статистические методы обработки экспериментальных данных[1])

1. ВВЕДЕНИЕ

§ 1. Предварительные замечания

Ниже приводится сводка некоторых данных по математической статистике и основных формул, полезных при оценках неизвестных величин по результатам эксперимента.

Материал изложен по возможности в наиболее общем виде, что позволяет применять его для решения самых разнообразных проблем. Центр тяжести перенесен на результаты математической статистики, без каких-либо попыток доказывать приведенные соотношения или обсуждать предмет теории вероятностей, за исключением тех случаев, где это помогает глубже осмыслить результаты.

Тем, кто заинтересован в более подробном изучении данных вопросов, следует порекомендовать работы, указанные в литературе (стр. 292). Необходимо отметить, что содержание обзора касается самым непосредственным образом лишь таких экспериментов, интерпретация которых базируется на статистическом анализе результатов; в связи с этим для рассмотрения отобраны наиболее эффективные методы анализа.

§ 2. Определения и обозначения

Случайная величина. Величина X, представляющая собой результат случайного опыта, называется *случайной величиной*. Непостоянство результатов такого опыта может быть связано с наличием случайных ошибок измерений

[1]) High Energy and Nuclear Physics Data Handbook, Ed. by W. Galbraith, W.S.C. Williams, Chilton, 1963.

или (кроме того) со статистической природой измеряемой величины (например, времени жизни). Отдельные значения случайной величины будут обозначаться как x_i, где $i = 1, 2, \ldots, n$. Любая функция от x_i также случайная величина.

Выборка и генеральная совокупность. Полезно различать *выборку*, состоящую из конечного числа значений случайной величины, и *генеральную совокупность* — полный набор всех возможных значений, которые может принимать случайная величина.

Распределение вероятности (частотная функция). Генеральную совокупность можно описать с помощью *распределения вероятности*; если распределение является дискретным, то вероятность получить в отдельном испытании результат x_i равна P_i; в случае непрерывного распределения элементарная вероятность $f(x_i)\,dx_i$ представляет собой вероятность результата, заключенного между x_i и $x_i + dx_i$. Распределение вероятности нормировано так, что

$$\sum_{i=1}^{N} P_i = 1, \tag{I.1a}$$

где N — число интервалов дискретного распределения, или

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \tag{I.1б}$$

в случае непрерывного распределения.

Параметры генеральной совокупности. Распределение вероятности, которое описывает генеральную совокупность, содержит постоянные параметры, такие, например, как *среднее значение* и *среднеквадратичное отклонение*; они обычно обозначаются греческими буквами, в частности для указанных параметров это μ и σ.

Оценки. Цель математической статистики — указать методы, с помощью которых по результатам выборки можно было бы получить оценки неизвестных параметров генеральной совокупности и их погрешности, которые сами суть случайные величины. Эти оценки, как и другие характеристики выборки, будут обозначаться курсивными

латинскими буквами; например, $\bar{x}$, s и $s(\bar{x})$ соответственно обозначают *выборочное среднее значение* случайной величины x, ее *выборочное среднеквадратичное отклонение* и *среднеквадратичное отклонение среднего.*

Математическое ожидание. *Математическое ожидание* функции $g(x)$ обозначается через $\mathscr{E}[g(x)]$ и определяется следующим образом:

$$\mathscr{E}[g(x)] = \sum_{i=1}^{N} P_i g(x_i) \qquad \text{(I.2а)}$$

в случае дискретного распределения и

$$\mathscr{E}[g(x)] = \int_{-\infty}^{\infty} f(x)\, g(x)\, dx \qquad \text{(I.2б)}$$

в случае непрерывного распределения.

Если функция $g(x)$ есть некоторая выборочная оценка [например, $g(x) = s$], то ее математическое ожидание представляет собой среднее значение, получаемое при осуществлении повторных выборок объемом n из одной и той же генеральной совокупности. При этом результат может быть функцией как параметров совокупности, так и объема n.

Асимптотика. В дальнейшем будут нередко приводиться результаты, носящие асимптотический характер и соответствующие выборкам предельно большого объема, т. е. при $n \to \infty$.

Несмещенные оценки. Если $\hat{\alpha}$ — оценка параметра α генеральной совокупности, полученная из выборки объемом n, и если

$$\mathscr{E}[\hat{\alpha} - \alpha] = 0,$$

то такая оценка называется *несмещенной.* В тех случаях, когда подобный результат достигается лишь при выборках большого объема (при $n \to \infty$), то говорят, что оценка *асимптотически несмещенная.*

Эффективные оценки. Наиболее эффективную оценку можно определить как оценку, для которой величина

$$\mathscr{E}[\hat{\alpha} - \alpha]^2$$

минимальна.

2. ОПРЕДЕЛЕНИЯ ХАРАКТЕРИСТИК ВЫБОРКИ

§ 1. Меры положения

1) **Среднее $\bar{x}$:**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{I.3}$$

2) **Взвешенное среднее $\bar{x}_w$.** Если каждому значению x_i приписан *вес* w_i, то *взвешенное среднее*:

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}. \tag{I.4}$$

3) **Медиана m.** Если все значения x_i расположены в порядке возрастания их величин, то *медианой* называется

$$m = x_{(n+1)/2} \qquad \text{при } n \text{ нечетном},$$

$$m = \frac{1}{2}(x_{n/2} + x_{(n/2)+1}) \qquad \text{при } n \text{ четном}.$$

Понятие медианы иногда используется в качестве характеристики асимметричного распределения.

4) **Мода.** *Модой* называется значение x_i, соответствующее максимуму в распределении выборки.

§ 2. Меры рассеяния

1) **Среднеквадратичное отклонение s [1]):**

$$s = \left\{ \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right\}^{1/2}. \tag{I.5}$$

Дисперсия есть квадрат среднеквадратичного отклонения s^2.

---

[1]) См. также раздел 3, п. 2.

2) **Среднее отклонение** d:

$$d = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|. \tag{I.6}$$

3) **Размах** r. *Размахом* (или вариационным размахом) называется разность между наибольшим и наименьшим значениями x в выборке:

$$r = x_{\text{макс}} - x_{\text{мин}}. \tag{I.7}$$

§ 3. Коэффициент корреляции

Для выборки, содержащей n пар значений (x_i, y_i) двух случайных величин x и y, вводится понятие *коэффициента корреляции* q:

$$q = \frac{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}, \tag{I.8}$$

или

$$q = \frac{\frac{1}{n} \sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{s_x s_y},$$

где $\bar{x}$ и $\bar{y}$ — средние значения по выборочным распределениям x_i и y_i в отдельности, а s_x и s_y — соответствующие среднеквадратичные отклонения.

Если $q = 0$, то говорят, что x и y не коррелированы; однако эти величины необязательно независимы. Если $q = \pm 1$, то между x и y существует прямая пропорциональность.

3. ОЦЕНКИ ПАРАМЕТРОВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ПО ХАРАКТЕРИСТИКАМ ВЫБОРКИ

Перечисленные выше характеристики сами являются случайными величинами. В этом разделе мы рассмотрим их следующие свойства: 1) приближение среднего значения некоторой характеристики выборки (которое получается на основании *повторных* выборок объема n) к соот-

ветствующему параметру генеральной совокупности и 2) среднеквадратичное отклонение полученной оценки. В дальнейшем мы воспользуемся определением математического ожидания $\mathscr{E}[g(x)]$ [см. (I.2)]. Относительно рассматриваемых выборок предполагается, что они производятся из генеральной совокупности случайной величины X со средним μ и среднеквадратичным отклонением σ.

1) **Среднее.** Каково бы ни было распределение случайной величины X,

$$\mathscr{E}(\bar{x}) = \mu,$$

и если x_i некоррелированы, то дисперсия среднего есть

$$\mathscr{E}[(\bar{x}-\mu)^2] = \frac{\sigma^2}{n}.$$

Согласно центральной предельной теореме [1]), если n велико (практически достаточно, если $n \geqslant 30$) и x_i не коррелированы, то $\bar{x}$ подчиняется нормальному распределению (см. раздел 4, § 1) со средним значением μ и среднеквадратичным отклонением $\sigma/\sqrt{n}$. Этот результат справедлив для любого распределения X в предположении, что оно обладает конечными средним значением и дисперсией. Если X распределено по нормальному закону, то высказанное выше утверждение является строгим при любом n.

На практике μ и σ обычно неизвестны, и в качестве их оценок используются $\bar{x}$ и s. В этом случае среднеквадратичное отклонение $\bar{x}$ приближенно (при больших n) считают равным $s/\sqrt{n}$ и результат серии измерений, как правило, записывают в виде (см. раздел 5, § 2)

$$\bar{x} \pm \frac{s}{\sqrt{n}}.$$

2) **Дисперсия.** Для любого распределения случайной величины X

$$\mathscr{E}(s^2) = \frac{n-1}{n}\sigma^2. \qquad \text{(I.9)}$$

[1]) См. стр. 55 настоящей книги, а также [6].

Таким образом, выборочная дисперсия [формула (I.5)] является смещенной оценкой σ^2. Поэтому для выборок малого объема лучше пользоваться выражением

$$s_c^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \qquad (I.10)$$

Если x_i принадлежит нормальному распределению с дисперсией σ^2, то дисперсия величины s_c^2 равна $2\sigma^4/(n-1)$; в то же время величина $(n-1)s_c^2/\sigma^2$ подчиняется распределению χ^2 с $(n-1)$ степенями свободы (см. раздел 4, § 2).

3) **Среднеквадратичное отклонение.** Если случайная величина X распределена по нормальному закону, то

$$\mathscr{E}(s) \approx \left(1 - \frac{3}{2n}\right)^{1/2}\sigma$$

и, следовательно, s есть асимптотически несмещенная оценка σ. Более того, среднеквадратичное отклонение для s в этом случае равно

$$s(s) \approx \frac{\sigma}{\sqrt{2n}}.$$

Если имеется несколько серий результатов измерений, причем они соответствуют генеральным совокупностям с различными средними, но одинаковыми среднеквадратичными отклонениями σ, то, используя эти данные, можно получить более точную оценку σ. Пусть s_j — среднеквадратичное отклонение [формула (I.5)], полученное в j-й выборке, и пусть N — полное число результатов измерений в m выборках; тогда

$$s = \left[\frac{\sum_{j=1}^{m} s_j^2}{N-m}\right]^{1/2},$$

и среднеквадратичное отклонение s приблизительно равно

$$\frac{s}{[2(N-m)]^{1/2}}.$$

4) **Взвешенное среднее.** Известно, что

$$\mathscr{E}(\bar{x}_w) = \mu,$$

причем его дисперсия равна

$$\mathscr{E}\,[(\bar{x}_w-\mu)^2]=\sigma^2\frac{\sum\limits_{i=1}^{n} w_i^2}{\left(\sum\limits_{i=1}^{n} w_i\right)^2}.$$

Если x_i независимы и получены из генеральных совокупностей с одним и тем же средним μ, но разными дисперсиями σ_i^2, то наиболее эффективной оценкой μ является взвешенное среднее (веса $w_i=1/\sigma_i^2$). Тогда среднеквадратичное отклонение для $\bar{x}_w$ равно

$$s\,(\bar{x}_w)=\left[\sum_{i=1}^{n}\left(\frac{1}{\sigma_i^2}\right)\right]^{-1/2}.$$

5) **Медиана.** Когда распределение X симметрично относительно среднего, то

$$\mathscr{E}\,(m)=\mu.$$

Если распределение нормально, то среднеквадратичное отклонение для медианы (при большом n) равно

$$s\,(m)=\sqrt{\frac{\pi}{2}}\,\frac{\sigma}{\sqrt{n}}\approx 1{,}25\,\frac{\sigma}{\sqrt{n}};$$

при $n\leqslant 10$ это выражение превращается в $\sim 1{,}2\sigma/\sqrt{n}$.

6) **Размах.** В отличие от большинства характеристик выборки размах не является асимптотически нормальной величиной. Тем не менее его нередко используют, чтобы быстро (хотя и грубо) оценить меру рассеяния; при этом в случае нормально распределенных величин x_i можно получить указание о среднеквадратичном отклонении, поделив размах на коэффициент c, значения которого для разных n приведены в табл. I (стр. 254).

7) **Среднее отклонение.** Если x_i подчиняются нормальному распределению, то

$$\mathscr{E}\,(d)=\sigma\left[\frac{2\,(n-1)}{\pi n}\right]^{1/2}$$

и среднеквадратичное отклонение для d приближенно равно

$$s(d) \approx \left(1-\frac{2}{\pi}\right)^{1/2} \frac{\sigma}{\sqrt{n}}.$$

8) **Коэффициент корреляции.** Если две величины x и y распределены по нормальному закону, то

$$s(q) \approx (1-q^2)\frac{1}{\sqrt{n}}.$$

4. СПЕЦИАЛЬНЫЕ РАСПРЕДЕЛЕНИЯ

§ 1. Нормальное (гауссово) распределение

Параметрами *нормального распределения* являются среднее значение μ и среднеквадратичное отклонение σ; плотность вероятности имеет вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \qquad (I.11)$$

Это распределение играет фундаментальную роль в теории ошибок, поскольку

1) оно описывает распределение ошибок, обусловленных множеством *малых независимых вкладов*, носящих случайный характер;

2) многие функции случайных величин (такие, например, как среднее, среднеквадратичное отклонение и т. п.) распределены асимптотически нормально даже в тех случаях, когда исходная величина не следует нормальному закону (см. [6]).

Стандартизованная случайная величина

$$u = \frac{x-\mu}{\sigma}$$

распределена по нормальному закону с нулевым средним и единичным среднеквадратичным отклонением.

Как будет видно из дальнейшего, для вычисления *доверительных пределов* полезно иметь таблицы различных значений u_p и вероятностей p обнаружить значения

$u \geqslant u_p$. Это означает, что p представляет собой интеграл

$$p = \sqrt{\frac{2}{\pi}} \int_{u_p}^{\infty} e^{-u^2/2}\, du.$$

В табл. II (стр. 254) протабулированы значения p и ординаты плотности вероятности нормального распределения $f(u_p)$ как функции от u_p; в табл. III (стр. 255) приводятся значения u_p как функции p.

§ 2. Распределение χ^2

Если x_i — независимые нормально распределенные случайные величины со средним, равным нулю, и среднеквадратичным отклонением, равным единице, то сумма квадратов этих величин

$$\chi^2 = \sum_{i=1}^{n} x_i^2$$

подчиняется распределению χ^2 с n степенями свободы. Плотность вероятности этого распределения

$$f_n(\chi^2) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} e^{-\chi^2/2}(\chi^2)^{(n-2)/2} \quad \text{для} \quad 0 < \chi^2 < \infty.$$

Среднее значение χ^2 равно n, а ее дисперсия равна $2n$.

Сумма m независимых величин χ_j^2, каждая из которых обладает n_j степенями свободы, также подчиняется распределению χ^2 с числом степеней свободы, равным $\sum_{j=1}^{m} n_j$.

Распределение χ^2 имеет важное практическое значение при проверке гипотез (см. раздел 9, § 2), поскольку распределение величины $\sqrt{n}(x - \mu)/\sigma$ стремится к нормальному с нулевым средним и единичным среднеквадратичным отклонением.

В табл. IV (стр. 256) содержатся значения χ_p^2 и вероятностей p обнаружить значения $\chi^2 > \chi_p^2$.

В тех случаях, когда число ν превышает 30, можно пользоваться приближенной формулой

$$\chi_p^2 = \frac{1}{2}\left[(2\nu - 1)^{1/2} + u_p\right]^2,$$

где u_p — значение *стандартизованной* нормально распределенной величины (см. табл. II и III).

§ 3. Распределение t (Стьюдента)

Пусть имеется выборка величины X объемом n из генеральной совокупности, подчиняющейся нормальному распределению со средним μ; тогда оказывается, что величина

$$t = \frac{\sqrt{n-1}}{s}(\bar{x} - \mu)$$

подчиняется распределению t с числом степеней свободы $(n-1)$. Плотность вероятности этого распределения

$$f_{(n-1)}(t) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\left[\pi(n-1)\right]^{-1/2}\left(1 + \frac{t^2}{n-1}\right)^{-n/2}.$$

Распределение t симметрично относительно среднего, равного нулю; его дисперсия равна $(n-1)/(n-3)$.

В табл. V (стр. 258) приведены абсолютные значения t_p и вероятности p обнаружить значения $|t| \geqslant t_p$ в зависимости от числа степеней свободы ν. Вероятность p определяется как

$$p = 2\int_{t_p}^{\infty} f_\nu(t)\,dt.$$

По мере возрастания n распределение t стремится к нормальному распределению; основное различие заключается в том, что для формы распределения t характерны более длинные «хвосты». Это распределение играет важную роль и используется при проверке гипотез (см. [5], а также раздел 9, § 3).

§ 4. Равномерное (прямоугольное) распределение

Если вероятность того, что x заключено в интервале от x_i до $x_i + \delta x_i$, одинакова для любых значений x_i, а область изменения x ограничена интервалом от a до b ($b > a$), то x подчиняется равномерному распределению, причем плотность вероятности

$$f(x) = \frac{1}{b-a}.$$

Среднее значение x равно $(a+b)/2$, а среднее квадратичное отклонение равно $(b-a)/\sqrt{12}$.

§ 5. Биномиальное распределение

Если в результате эксперимента можно получить одно из двух возможных значений A или B (например, речь может идти о событии, относящемся либо к передней, либо к задней полусфере при измерениях углового распределения), то частота k появления результата A (при полном числе испытаний n) подчиняется биномиальному распределению. Пусть вероятность результата A в одном испытании равна p (при этом вероятность B равна $1 - p = q$), тогда вероятность обнаружить A в k из n испытаний равна k-му члену разложения $(p + q)^n$ в ряд, т. е.

$$P(k) = \frac{n!}{k!\,(n-k)!}\, p^k \cdot q^{(n-k)}; \quad k = 0, 1, \ldots, n.$$

Если набор из n испытаний многократно повторяется, то среднее значение величины k будет стремиться к np; при этом среднеквадратичное отклонение для k будет равно $\sqrt{npq}$.

В качестве оценки неизвестной величины p используется $\hat{p} = k/n$. Оценку среднеквадратичного отклонения можно получить, если в выражение $\sqrt{p(1-p)/n}$ подставить $\hat{p}$ вместо p.

В случае большого n (такого, что $npq > 10$) биномиальное распределение хорошо аппроксимируется нормальным распределением; в случае большого n, но мало-

го p для аппроксимации можно использовать распределение Пуассона (см. § 6).

Таблицы биномиального распределения и значения накопленных сумм (т. е. вероятности того, что в n испытаниях число благоприятных исходов $k > k_0$ при заданном p) приводятся в работе [29] для $p = 0{,}01$ (0,01) 0,50 и $n = 1$ (1) 49. Приведенная в книге табл. VI (стр. 260) соответствует значениям $p=0{,}1$ (0,01) 0,5 и $n=2$ (1) 10 [1]).

§ 6. Распределение Пуассона

Если эффективное число испытаний n в некотором эксперименте стремится к бесконечности, тогда как вероятность благоприятного исхода p стремится к нулю, так что произведение np остается конечным и постоянным, то биномиальное распределение переходит в распределение Пуассона. Это распределение очень важно (особенно в экспериментах, где используются различные счетчики), поскольку описывает распределение вероятностей редких событий.

Если вероятность осуществления события в интервале δx (пространства, времени и т. п.) равна $\lambda\delta x$, где λ — постоянная величина, то вероятность того, что в конечном интервале x событие произойдет независимым образом k раз, дается распределением Пуассона

$$p(k) = \frac{K^k}{k!} e^{-K},$$

где $K = \lambda x$ — среднее число событий в интервале x. Среднее число событий равно K, а среднее квадратичное отклонение равно $\sqrt{K}$. Оценкой параметра λ служит величина k/x со средним квадратичным отклонением $\sqrt{k/x}$.

При малом p ($p \leqslant 0{,}10$) распределение Пуассона можно использовать в качестве приближения для биномиального распределения, причем $K = np$. В книге Молина [26] приводятся таблицы значений как отдельных членов распределения Пуассона, так и накопленных вероятностей.

---

[1]) Принятое здесь обозначение указывает пределы изменения величин (крайние цифры) и значение шага (цифра в скобках), с которым изменяются эти величины.— *Прим. перев.*

5. ДОВЕРИТЕЛЬНЫЕ ПРЕДЕЛЫ

Если значение неизвестного параметра α с вероятностью $(1 - p)$ заключено в интервале от α_1 до α_2, то α_1 и α_2 называются $100(1 - p)\%$-ными *доверительными пределами* для α.

§ 1. Выборки из нормального распределения

Понятие доверительных пределов можно использовать для оценок параметров нормального распределения при наличии выборки любого объема.

1) Доверительные пределы для *среднего*:

$$\bar{x} - \frac{t_p s}{\sqrt{n-1}} \quad \text{и} \quad \bar{x} + \frac{t_p s}{\sqrt{n-1}},$$

где t_p — абсолютное значение величины t (критерий Стьюдента), для которого $t \geqslant t_p$ с вероятностью p при числе степеней свободы $n - 1$ (см. табл. V).

2) Доверительные пределы для *среднеквадратичного отклонения*

$$s_1 = \frac{s\sqrt{n}}{\chi_{p''}} \quad \text{и} \quad s_2 = \frac{s\sqrt{n}}{\chi_{p'}},$$

где $p' = 1 - p/2$, $p'' = p/2$, а $\chi^2_{p'}$ и $\chi^2_{p''}$ — значения χ^2, для которых с вероятностью p' и p'' соответственно $\chi^2 \geqslant \chi^2_{p'}$ и $\chi^2 \geqslant \chi^2_{p''}$, при числе степеней свободы $n - 1$ (см. табл. IV).

Следует отметить, что если n велико, то и распределение t, и распределение χ^2 приближаются к нормальному, и в таких случаях обычно применяются более простые методы оценок (см. раздел 9), основанные на нормальном законе.

§ 2. Большие выборки и приближенно нормальные оценки

В случае выборок, не подчиняющихся нормальному распределению, не всегда удается найти доверительный интервал; тем не менее этот метод применяют к выборкам

большого объема, так как при этом считается, что распределение оценки обычно приближается к нормальному [1]).

Таким образом, при большом n можно говорить о приближенных $100\,(1-p)\%$-ных доверительных пределах для параметра α, которые даются в виде $\hat{\alpha} \pm u_p s(\hat{\alpha})$, где $\hat{\alpha}$ — оценка α; $s(\hat{\alpha})$ — ее среднеквадратичное отклонение; u_p — значение стандартизованной нормально распределенной величины, для которой $u \geqslant u_p$ с вероятностью p (см. табл. II и III).

Такой метод дает надежные доверительные пределы для среднего значения, если $n \geqslant 30$ и распределение не очень асимметрично; что касается оценок среднеквадратичного отклонения, коэффициентов корреляции и т. п., то в этих случаях n должно превышать 100.

6. МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

§ 1. Оценка максимального правдоподобия

Метод максимального правдоподобия представляет собой наиболее важный общий метод оценки параметров. Если величины x_i $(i = 1, 2, \ldots, n)$ составляют выборку из генеральной совокупности для x с плотностью вероятности $f(x, \alpha)$, где α — постоянный параметр, то элементарная вероятность получить именно эту выборку равна

$$f(x_1, \alpha)\, f(x_2, \alpha) \ldots f(x_n, \alpha)\, dx_1\, dx_2 \ldots dx_n.$$

Совместная плотность вероятности

$$l(x_1, x_2, \ldots, x_n; \alpha) = \prod_{i=1}^{n} f(x_i, \alpha) \qquad \text{(I.12а)}$$

называется *функцией правдоподобия* выборки.

В случае дискретного распределения функция правдоподобия имеет вид

$$\prod_{i=1}^{n} [P_i(\alpha)]^{r_i}, \qquad \text{(I.12б)}$$

где r_i $(i = 1, 2, \ldots, n)$ — частота появления x_i, а $P_i(\alpha)$ — вероятность получить этот результат.

[1]) О доверительных пределах в случае биномиального распределения см. работу [5].

Метод максимального правдоподобия состоит в том, что в качестве оценки неизвестного параметра α выбирается такое значение, которое максимизирует функцию правдоподобия. В ряде случаев максимум этой функции удается найти аналитически, решая уравнение

$$\frac{\partial L}{\partial \alpha} = \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} \ln [f(x_i; \alpha)] = 0, \tag{I.13}$$

где $L = \ln [l(x_1, \ldots, x_n; \alpha)]$. Пример использования такого метода дается в следующем параграфе.

В случае m неизвестных параметров α_j $(j = 1, \ldots, m)$ их оценки получаются из решений системы m уравнений:

$$\frac{\partial L}{\partial \alpha_j} = 0.$$

Часто на практике не удается получить решение в аналитическом виде, и тогда для нахождения максимума функции правдоподобия прибегают к численным расчетам. В подобных случаях, особенно когда выборка невелика и вид функции правдоподобия далек от нормального, целесообразно изображать графически функцию правдоподобия, откладывая по оси абсцисс значения оцениваемого параметра; в случае двух параметров удобно пользоваться контурными диаграммами (см. гл. 4).

§ 2. Свойства оценок максимального правдоподобия; среднеквадратичное отклонение

Можно показать, что метод максимального правдоподобия позволяет найти наиболее эффективную оценку $\hat{\alpha}$ параметра α, являющуюся при некоторых общих условиях асимптотически нормальной со средним значением α и среднеквадратичным отклонением

$$s(\hat{\alpha}) = \left\{ \mathscr{E} \left[-\frac{\partial^2 L}{\partial \alpha^2} \right] \right\}^{-1/2}. \tag{I.14}$$

Здесь $\mathscr{E}$ — значение математического ожидания, полученное усреднением по распределению совместной вероятности $l(x_1, \ldots, x_n; \alpha)$.

При вычислении величины $s(\hat{\alpha})$ следует использовать тот факт, что

$$\mathscr{E}\left(\frac{\partial L}{\partial \alpha}\right)=0. \qquad (I.15)$$

Для иллюстрации метода вычислим среднеквадратичное отклонение $s(s)$ от оценки s среднеквадратичного отклонения генеральной совокупности σ, предполагая, что x_i распределены по нормальному закону со средним генеральной совокупности μ.

В случае выборки объема n функция правдоподобия

$$l=\prod_{i=1}^{n}\left\{\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}\right\};$$

отсюда

$$L=-n\ln\sigma-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2+\text{const},$$

$$\frac{\partial L}{\partial\sigma}=-\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Полагая $\partial L/\partial\sigma=0$, получаем оценку для σ, т. е.

$$s=\left\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right\}^{1/2}.$$

Далее вычисляем

$$\frac{\partial^2 L}{\partial\sigma^2}=-\frac{2n}{\sigma^2}-\frac{3}{\sigma}\frac{\partial L}{\partial\sigma}.$$

Находя математическое ожидание этого выражения, получаем (с учетом $\mathscr{E}[\partial L/\partial\sigma]=0$)

$$\mathscr{E}\left(\frac{\partial^2 L}{\partial\sigma^2}\right)=-\frac{2n}{\sigma^2}.$$

Следовательно,

$$s(s)=\frac{\sigma}{\sqrt{2n}}.$$

Приближения. Нередко ошибка, которая дается формулой (3.14), представляет собой функцию неизвестных параметров; иногда специальным выбором параметра мож-

но ослабить эту зависимость; обычно для получения приближения, достаточно хорошего в случае малой величины среднеквадратичного отклонения, используются оценочные значения параметров [например, $s(s) \approx s/\sqrt{2n}$].

Во многих случаях математическое ожидание не удается вычислить аналитически, и тогда в качестве приближения величину $\partial^2 L/\partial\hat{\alpha}^2$ определяют из выборки численным путем; Орир [31] предложил использовать следующее выражение:

$$\frac{\int \frac{\partial^2 L}{\partial \hat{\alpha}^2} l(\hat{\alpha})\, d\hat{\alpha}}{\int l(\hat{\alpha})\, d\hat{\alpha}}.$$

§ 3. S-Функция Бартлетта

Использование этой функции лежит в основе метода получения оценок максимального правдоподобия для одного параметра, а также определения доверительных пределов; эти пределы, вообще говоря, более правильны, нежели те, которые получаются методом, указанным в предыдущем параграфе. S-Функция Бартлетта (см. [3]) определяется следующим образом:

$$S(\alpha) = \frac{\partial L}{\partial \alpha} \Big/ \left[-\mathscr{E}\left(\frac{\partial^2 L}{\partial \alpha^2}\right)\right]^{1/2}; \qquad \text{(I.16)}$$

она имеет нулевое среднее и единичную дисперсию.

Вычисляя $S(\alpha)$ как функцию α, можно получить оценку максимального правдоподобия $\hat{\alpha}$, являющуюся решением уравнения $S(\alpha) = 0$. Кроме того, в предположении, что $S(\alpha)$ подчиняется нормальному распределению, можно из уравнения $S(\alpha) = \pm 1$ определить 68,3%-ный доверительный интервал (т. е. плюс или минус одно среднеквадратичное отклонение); подобным образом из уравнения $S(\alpha) = \pm 2$ можно найти доверительные пределы, соответствующие двум среднеквадратичным отклонениям, и т. д.

Примеры использования этого метода в частных задачах оценки времени жизни частиц по их распаду, наблюдаемому в камере Вильсона (или пузырьковой камере), приводятся в работах [3, 16].

7. ПЕРЕНОС ОШИБОК

Часто результат эксперимента представляет собой некоторую функцию от нескольких различных случайных величин X_r. Предположим, что каждое наблюдаемое значение x_r принадлежит генеральной совокупности со средним μ_r и среднеквадратичным отклонением σ_r. Теория переноса ошибок позволяет определить значение среднеквадратичного отклонения, которое следует приписать величине $y(x_1, \ldots, x_m)$.

§ 1. Матрица ошибок (ковариационная матрица)

Для набора m случайных величин X_r $(r = 1, \ldots, m)$ *матрица ошибок* $\mathfrak{D}(\mathbf{x})$ определяется следующим образом:

$$[\mathfrak{D}(\mathbf{x})]_{rs} = \mathscr{E}[(x_r - \mu_r)(x_s - \mu_s)], \qquad \text{(I.17а)}$$

где $r, s = 1, \ldots, m$.

Из этого следует, что $\mathfrak{D}(\mathbf{x})$ — симметричная матрица ранга m, диагональные элементы которой представляют собой дисперсии соответствующих величин:

$$[\mathfrak{D}(\mathbf{x})]_{rr} = \sigma_r^2; \qquad \text{(I.17б)}$$

недиагональные элементы этой матрицы связаны с коэффициентами корреляции ρ_{rs}:

$$[\mathfrak{D}(\mathbf{x})]_{rs} = \sigma_r \sigma_s \rho_{rs}. \qquad \text{(I.17в)}$$

Если величины некоррелированы, то матрица $\mathfrak{D}(\mathbf{x})$ диагональна. На самом деле везде вместо значений σ_r, σ_s и ρ_{rs} фигурируют их оценки: $s(x_r)$, $s(x_s)$ и $q(x_r, x_s)$.

§ 2. Линейные функции

Пусть величины y_r $(r = 1, \ldots, m)$ являются линейными функциями переменных x_s $(s = 1, \ldots, n)$; это означает, что

$$y_r = a_{r1}x_1 + a_{r2}x_2 + \ldots + a_{rn}x_n = \sum_{s=1}^{n} a_{rs}x_s, \qquad \text{(I.18а)}$$

где a_{rs} — постоянные, представляющие собой элементы матрицы A. Уравнение (I.18а) в матричном обозначении

запишется в виде

$$\mathbf{y} = A\mathbf{x}. \tag{I.18б}$$

Если произведено несколько измерений каждой величины x_s, то оценка величины y_r получается при замене x_s на их средние $\bar{x}_s$. Пусть $\mathfrak{D}(\mathbf{x})$ — матрица ошибок величин x_s, тогда матрицей ошибок для функций y_r будет

$$\mathfrak{D}(\mathbf{y}) = A\mathfrak{D}(\mathbf{x})A'. \tag{I.19}$$

§ 3. Нелинейные функции

Результаты предыдущего параграфа можно обобщить и на случай нелинейных функций $y_r = f_r(x_1, \ldots, x_s, \ldots, \ldots, x_n)$, предположив, что функция f_r слабо меняется в области, ограниченной среднеквадратичным отклонением, от ее среднего значения, или, иными словами, что с точностью до членов первого порядка ее разложение в ряд Тэйлора имеет вид

$$y_r = f_r(x_1, \ldots, x_n) = f_r(\bar{x}_1, \ldots, \bar{x}_n) + \\ + \sum_{s=1}^{n} \left\{ (x_s - \bar{x}_s) \frac{\partial f_r}{\partial x_s}\Big|_{\mathbf{x}} \right\}. \tag{I.20}$$

В этом разложении $\bar{x}_s$ есть среднее значение x_s, а $\mathbf{x}$ обозначает набор $(\bar{x}_1, \ldots, \bar{x}_s, \ldots, \bar{x}_n)$. Тогда оценка величины y_r есть

$$\hat{y}_r = f_r(\bar{x}_1, \ldots, \bar{x}_n), \tag{I.21а}$$

а элементами матрицы ошибок являются

$$\mathfrak{D}[(\mathbf{y})]_{rs} = \sum_{i=1}^{n}\sum_{j=1}^{n} \left\{ (x_i - \bar{x}_i)(x_j - \bar{x}_j) \frac{\partial f_r}{\partial x_i}\Big|_{\mathbf{x}} \frac{\partial f_s}{\partial x_j}\Big|_{\mathbf{x}} \right\}. \tag{I.21б}$$

Можно показать, что выражение (I.21б) приводит к следующему результату для $\mathfrak{D}(\mathbf{y})$:

$$\mathfrak{D}(\mathbf{y}) = F\mathfrak{D}(\mathbf{x})F', \tag{I.22}$$

где F — матрица, элементы которой равны

$$F_{rs} = \frac{\partial f_r}{\partial x_s}\Big|_{\mathbf{x}}.$$

§ 4. Среднеквадратичное отклонение отношения двух величин

В качестве простого примера применения общего метода, описанного выше, рассмотрим одномерный случай отношения двух случайных величин x_1 и x_2.

Среднее значение этого отношения

$$\bar{y} \approx \frac{\bar{x}_1}{\bar{x}_2}. \tag{I.23}$$

Матрица ошибок для $\bar{x}_1$ и $\bar{x}_2$

$$\mathfrak{D} = \begin{bmatrix} s_1^2 & s_1 s_2 q_{12} \\ s_1 s_2 q_{12} & s_2^2 \end{bmatrix}, \tag{I.24}$$

и, кроме того,

$$F' = \frac{1}{\bar{x}_2} - \frac{x_1}{(\bar{x}_2)^2}. \tag{I.25}$$

Тогда ошибка величины $\bar{y}$ равна

$$s^2(\bar{y}) = \frac{\bar{x}_1^2}{\bar{x}_2^2}\left\{\frac{s_1^2}{\bar{x}_1^2} + \frac{s_2^2}{\bar{x}_2^2} - 2q_{12}\frac{s_1}{\bar{x}_1}\frac{s_2}{\bar{x}_2}\right\}. \tag{I.26}$$

Примечание. Если x_1 и x_2 распределены по нормальному закону и если $\bar{x}_2/s_2$ больше 5, то $\bar{y}$ будет приблизительно подчиняться нормальному распределению со средним значением (I.23) и дисперсией (I.26).

§ 5. Среднеквадратичное отклонение произведения двух величин

Если исходные величины обладают свойствами, описанными в § 3, то рассмотрение, аналогичное предыдущему, дает

$$\bar{y} \approx \bar{x}_1 \bar{x}_2 \tag{I.27}$$

и

$$s^2(\bar{y}) = \bar{x}_1^2 \bar{x}_2^2\left\{\frac{s_1^2}{\bar{x}_1^2} + \frac{s_2^2}{\bar{x}_2^2} + 2q_{12}\frac{s_1}{\bar{x}_1}\frac{s_2}{\bar{x}_2}\right\}. \tag{I.28}$$

Примечание. Если x_1 и x_2 — независимые нормально распределенные величины, то строгим выражением для дис-

персии их произведения является

$$s^2(\bar{y}) = s_1^2\bar{x}_2^2 + s_2^2\bar{x}_1^2 + s_1^2 s_2^2, \qquad (I.29)$$

и оно может существенно отличаться от приближенного выражения (I.28) (при $q_{12} = 0$), если ошибки $\bar{x}_1$ и $\bar{x}_2$ велики.

§ 6. Общая формула переноса ошибок для независимых переменных

В заключение запишем обычную формулу (приближенную для нелинейных функций) переноса ошибок в случае одного параметра, зависящего от некоторого количества независимых переменных.

Если $y = f(x_i)$, $i = 1, \ldots, n$, то

$$s^2(\hat{y}) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}\right)^2_{\bar{x}_i} s^2(\bar{x}_i).$$

8. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ. АППРОКСИМАЦИЯ ПОЛИНОМОМ

§ 1. Постановка задачи

В этом разделе рассмотрена задача получения наилучшего решения, описывающего соотношение между двумя величинами X и Y, где экспериментальные результаты представляют собой серии измерений Y, y_r ($r = 1, \ldots, n$) со среднеквадратичными отклонениями σ_r, выполненные при соответствующих значениях X, x_r, которые не имеют ошибок. Будем предполагать также, что значения величины Y независимы и распределены по нормальному закону.

Пусть полином, которым предполагается описать экспериментальные данные, представлен в общем виде

$$\mathrm{y} = \mathrm{y}(\alpha_1, \ldots, \alpha_p; \mathrm{x}),$$

где α_j ($j = 1, \ldots, p$) — неизвестные постоянные параметры. В результате процедуры аппроксимации будут получены оценки $\hat{\alpha}_j$ параметров α_j и наилучшие значения $\hat{y}_r(\alpha_1, \ldots, \alpha_p; x_r)$ величины Y в точках x_r. Если y_r под-

чиняются нормальному распределению, то оценками максимального правдоподобия оказываются $\hat{y}$, которые *минимизируют* величину

$$M = \sum_{r=1}^{n} \frac{(y_r - \hat{y}(x_r))^2}{\sigma_r^2}. \qquad (I.30)$$

Эту процедуру обычно называют *методом наименьших квадратов*. Значения $\hat{\alpha}_j$ получаются в результате решения системы p уравнений

$$\frac{\partial M}{\partial \hat{\alpha}_j} = 0. \qquad (I.31)$$

Величина M подчиняется распределению χ^2 с числом степеней свободы, равным $(n - p)$. Следовательно, в данном случае можно применять критерий согласия, основанный на сводке данных раздела 4, § 2 (см. также раздел 9, § 1).

Оценки $\hat{\alpha}_j$ в общем случае будут коррелированными; метод получения матрицы ошибок

$$[\mathfrak{D}(\alpha)]_{ij} = \mathscr{E}[(\hat{\alpha}_i - \alpha_i)(\hat{\alpha}_j - \alpha_j)]$$

состоит в следующем (см. [31]):

$$\mathfrak{D}(\alpha) = C^{-1},$$

причем

$$C_{ij} = \frac{1}{2} \frac{\partial^2 M}{\partial \alpha_i \partial \alpha_j}\Big|_{\hat{\alpha}_1, \ldots, \hat{\alpha}_p}. \qquad (I.32)$$

Если ошибки σ_r неизвестны, но известны их весовые множители w_r, зависящие от x_r, т. е.

$$\sigma_r = \frac{\sigma}{w_r},$$

то уравнения (I.30) — (I.32) следует преобразовать следующим образом:

$$M' = \sum_{r=1}^{n} w_r^2 (y_r - \bar{y}(x_r))^2, \qquad (I.30')$$

$$\frac{\partial M'}{\partial \hat{\alpha}_j} = 0, \qquad (I.31')$$

$$C_{ij} = \frac{1}{2} \frac{\partial^2 M'}{\partial \alpha_i \partial \alpha_j}\Big|_{\hat{\alpha}_1, \ldots, \hat{\alpha}_p}, \qquad (I.32')$$

и

$$\mathfrak{D}(\alpha) = s^2 C^{-1},$$

где s^2 — оценка σ^2. Она может быть найдена из разброса значений y_r относительно аппроксимирующего полинома:

$$s^2 = \frac{M' \text{ (мин.)}}{n-p}. \tag{I.33}$$

В таком случае можно считать, что величина

$$M = \frac{M'}{s^2}$$

подчиняется распределению χ^2 с $(n-p)$ степенями свободы.

§ 2. Линейные функции

Если $\mathbf{y}(\alpha_1, \ldots, \alpha_p; \mathbf{x})$ линейно по α_j, т. е.

$$\mathbf{y}(\alpha_1, \ldots, \alpha_p; \mathbf{x}) = \sum_{j=1} \alpha_j f_j(\mathbf{x}), \tag{I.34}$$

то уравнениями для $\hat{\alpha}_j$ будут

$$\hat{\alpha}_j = \sum_{k=1}^{p} \sum_{r=1}^{n} \frac{y_r f_k(x_r)}{\sigma_r^2} [\mathfrak{D}(\alpha)]_{kj} \tag{I.35}$$

и

$$[\mathfrak{D}^{-1}(\alpha)]_{kj} = \sum_{r=1}^{n} \frac{f_k(x_r) f_j(x_r)}{\sigma_r^2}. \tag{I.36}$$

§ 3. Аппроксимация прямой линией

Уравнение прямой

$$\mathbf{y} = \alpha_1 + \alpha_2 \mathbf{x}. \tag{I.37}$$

Воспользовавшись результатами предыдущих параграфов, найдем

$$f_1(x_r) = 1; \quad f_2(x_r) = x_r, \tag{I.38}$$

$$C = \begin{pmatrix} \sum \frac{1}{\sigma_r^2} & \sum \frac{x_r}{\sigma_r^2} \\ \sum \frac{x_r}{\sigma_r^2} & \sum \frac{x_r^2}{\sigma_r^2} \end{pmatrix}. \tag{I.39}$$

Детерминант матрицы C

$$\Delta = \left(\sum \frac{1}{\sigma_r^2}\right)\left(\sum \frac{x_r^2}{\sigma_r^2}\right) - \left(\sum \frac{x_r}{\sigma_r^2}\right)^2, \qquad (I.40)$$

а матрица $\mathfrak{D}(\alpha)$, обратная C, являющаяся матрицей ошибок для $\hat{\alpha}_1$ и $\hat{\alpha}_2$, оказывается следующей:

$$\mathfrak{D}(\alpha) = \frac{1}{\Delta}\begin{pmatrix} \sum \frac{x_r^2}{\sigma_r^2} & -\sum \frac{x_r}{\sigma_r^2} \\ -\sum \frac{x_r}{\sigma_r^2} & \sum \frac{1}{\sigma_r^2} \end{pmatrix}. \qquad (I.41)$$

Решение задачи

$$\hat{\alpha}_1 = \frac{1}{\Delta}\left\{\left(\sum \frac{y_r}{\sigma_r^2}\right)\left(\sum \frac{x_r^2}{\sigma_r^2}\right) - \left(\sum \frac{y_r x_r}{\sigma_r^2}\right)\left(\sum \frac{x_r}{\sigma_r^2}\right)\right\},$$

$$\hat{\alpha}_2 = \frac{1}{\Delta}\left\{\left(\sum \frac{y_r x_r}{\sigma_r^2}\right)\left(\sum \frac{1}{\sigma_r^2}\right) - \left(\sum \frac{y_r}{\sigma_r^2}\right)\left(\sum \frac{x_r}{\sigma_r^2}\right)\right\}.$$

В данном примере $p = 2$, и, таким образом, M подчиняется распределению χ^2 с $(n - 2)$ степенями свободы.

§ 4. Обобщенный метод наименьших квадратов

Во многих важных приложениях, таких, как кинематическое описание результатов экспериментов, полученных в пузырьковых камерах, метод наименьших квадратов обобщается на случай коррелированных величин и нелинейных функций. Эти вопросы рассмотрены в работах [1, 16] [1]).

9. ПРОВЕРКА ГИПОТЕЗ

§ 1. Общие замечания

Статистическая проверка некоторой гипотезы относительно набора экспериментальных данных сама по себе не дает доказательств, правильна или ложна эта гипотеза;

[1]) См. также [63] и дополнение 2.—*Прим. перев.*

подобная проверка лишь указывает степень согласия гипотезы с результатами эксперимента.

Критерий проверки обычно выбирают таким, чтобы вероятность p *отвергнуть* гипотезу была малой, когда гипотеза *верна*. Значение p называется *уровнем значимости* и, как правило, может быть вычислено. С другой стороны, оценить вероятность того, что гипотеза будет принята, если она *неверна*, вообще говоря, невозможно, поскольку для этого следовало бы рассмотреть все прочие возможные гипотезы, число которых, по-видимому, бесконечно.

Если проверяется несколько различных гипотез, дающих объяснение одним и тем же данным (ситуация, типичная для экспериментов с пузырьковой камерой), то часто возникает неопределенность, обусловленная тем, что заданному критерию удовлетворяет несколько гипотез. Выход из положения может зависеть от эксперимента (к примеру, дополнительная информация может оказаться способной разрешить возникшую неоднозначность), но в общем случае было бы неверно вносить последующие поправки в имеющийся критерий проверки и затем его применять исключительно к сомнительным случаям с целью получить решение. В принципе сомнительное событие следовало бы отнести (при отсутствии какой-либо дополнительной информации) к любому из типов, удовлетворяющих заданному критерию согласия; здесь имеется в виду, что события некоторых типов могут быть плохо различимы и сильно замаскированы событиями более часто встречающихся типов.

Еще одна трудность часто возникает в тех случаях, когда гипотеза формулируется лишь *после* изучения экспериментальных данных; это происходит, например, при проверке на значимость новых «резонансов», обнаруженных в экспериментальных данных, причем положение и характер предполагаемого резонанса неизвестны до того, как обработаны экспериментальные результаты. В подобной ситуации единственным действительно удовлетворительным критерием значимости, по-видимому, может служить повторный эксперимент.

§ 2. Критерий χ^2

Часто бывает необходимо проверить для заданной выборки x_i ($i = 1, \ldots, n$) случайной величины X гипотезу о том, что $f(X)$ является плотностью вероятности для X. Для получения меры отклонения имеющихся данных от ожидаемых согласно гипотетическому распределению используется величина χ^2; критерием проверки гипотезы может служить сопоставление величины χ^2 с табличным значением, которое соответствует заданному $100p\%$-ному уровню значимости.

Прежде всего весь диапазон значений x в данной выборке разбивается на n неперекрывающихся интервалов. Если p_i обозначает гипотетическую вероятность того, что x принимает значение, относящееся к i-му интервалу (причем число значений из выборки, попадающих в i-й интервал, равно ν_i), то

$$\chi^2 = \sum_{i=1}^{n} \frac{(\nu_i - np_i)^2}{np_i}, \tag{I.42}$$

где n — объем выборки, и

$$\sum_{i=1}^{n} (\nu_i - np_i) = 0. \tag{I.43}$$

Если *предполагаемое* число значений в каждом интервале больше десяти (т. е. $np_i \geqslant 10$), то величина

$$\frac{\nu_i - np_i}{\sqrt{np_i}} \tag{I.44}$$

будет распределена асимптотически нормально. В таком случае χ^2 представляет собой сумму квадратов n асимптотически нормальных величин, которые должны подчиняться условию (I.43); из этого следует, что если гипотеза верна, то величина χ^2 (I.42) подчиняется (по мере роста n) распределению χ^2 с числом степеней свободы, равным $(n - 1)$.

Пусть χ_p^2 — значение χ^2, для которого $\chi^2 \geqslant \chi_p^2$ с вероятностью p в случае $n - 1$ степеней свободы.

Тогда проверка гипотезы на $100p\%$-ном уровне значимости состоит в следующем.

Если вычисленная величина χ^2 превысит χ^2_p, то следует считать, что данная выборка обнаруживает значимое отличие от гипотетического распределения, и тогда гипотеза отвергается; в противном случае ($\chi^2 < \chi^2_p$) гипотеза принимается. Согласно этому правилу, вероятность отвергнуть правильную гипотезу равна p.

Чтобы удовлетворялись сделанные предположения, необходимо разбивать диапазон значений x так, чтобы в каждом интервале оказывалось больше десяти ожидаемых значений.

Значения χ^2_p как функции p и числа степеней свободы ν можно получить, пользуясь соответствующей номограммой или табл. IV (стр. 256). В тех случаях, когда ν превышает 30, можно использовать приближенную формулу

$$\chi^2_p = \frac{1}{2}\{(2\nu-1)^{1/2} + u_p\}^2, \qquad (I.45)$$

где u_p соответствует $100p\%$-ному уровню для нормального распределения и может быть найдено из табл. II (стр. 254).

§ 3. Оценка параметров с помощью критерия χ^2

В данном параграфе рассматривается случай, когда гипотетическое распределение не полностью определено заранее. Неизвестные параметры этого распределения можно оценить методом максимального правдоподобия или, что то же самое, методом наименьших квадратов (по минимуму значения χ^2) для приближенно нормально распределенных частей выборки в каждом интервале. Можно показать, что тогда найденная величина χ^2 будет подчиняться распределению χ^2 с числом степеней свободы, равным $(n - s - 1)$, где s — число параметров, оцениваемых из данной выборки. Сам же критерий χ^2 затем применяется аналогично предыдущему случаю.

Если методом наименьших квадратов оценивается единственный параметр α, то результат полезно изображать в виде графика зависимости вероятности p получить большее значение χ^2 от соответствующего значения оценки $\hat{\alpha}$.

Это позволит наглядно убедиться в пригодности полученного решения и одновременно определить доверительный интервал для $\hat{\alpha}$. В случае двух параметров α и β используется графическое представление на плоскости $\hat{a}$, $\hat{b}$, где изображаются контуры постоянных p.

§ 4. Проверка положения среднего значения нормального распределения

В обзоре Камерона [5] рассмотрено несколько критериев проверки положения среднего значения нормально распределенной величины. В данном параграфе будет изложен способ проверки, отлично ли среднее значение от некоторой заданной величины.

Пусть $\bar{x}$ и s — соответственно среднее значение и среднеквадратичное отклонение, найденные на основании выборки объема n из нормального распределения со средним μ и среднеквадратичным отклонением σ.

Необходимо проверить гипотезу $\mu = \mu_0$. Для этого используется статистика

$$t = \frac{|\bar{x} - \mu_0|}{s/\sqrt{n}},$$

и если гипотеза справедлива, то величина t подчиняется распределению t с $(n - 1)$ степенями свободы.

При $100p\%$-ном уровне значимости гипотезу следует отвергнуть, если наблюденное значение t превышает значение t_p, для которого $t \geqslant t_p$ с вероятностью p в случае $n - 1$ степеней свободы. Значения t_p могут быть найдены в табл. V (стр. 258).

Практически при больших n хорошим приближением для распределения t служит нормальное распределение. Какое n можно считать «большим»— в некоторой степени зависит от уровня значимости, на котором проводится проверка; так, при $n = 10$ и $t_p = 3{,}29$ вероятность $t > 3{,}29$ равна $\sim 10^{-3}$ согласно нормальному распределению и $\sim 10^{-2}$ согласно распределению t.

II

У. К. ГАМИЛЬТОН

Метод наименьших квадратов и проверка линейных гипотез[1])

§ 1. Метод наименьших квадратов

Допустим, что результаты эксперимента представляют собой n различных величин

$$y_1, y_2, \ldots, y_n, \tag{II.1}$$

причем известно, что каждая из этих n величин *линейно* зависит от m ($\leqslant n$) параметров

$$\theta_1, \theta_2, \ldots, \theta_m. \tag{II.2}$$

С каждой из наблюдаемых величин y_i связана случайная ошибка ε_i, так что можно написать:

$$\begin{aligned} y_1 &= a_{11}\theta_1 + a_{12}\theta_2 + \ldots + a_{1m}\theta_m + \varepsilon_1, \\ y_2 &= a_{21}\theta_1 + a_{22}\theta_2 + \ldots + a_{2m}\theta_m + \varepsilon_2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n &= a_{n1}\theta_1 + a_{n2}\theta_2 + \ldots + a_{nm}\theta_m + \varepsilon_n. \end{aligned} \tag{II.3}$$

Систему линейных уравнений (II.3) можно записать в матричной форме, а именно

$$\mathbf{y} = A\theta + \boldsymbol{\varepsilon}. \tag{II.3a}$$

Матрица A ранга m предполагается известной. Она часто называется *структурной* или *конструкционной* матрицей. Мы считаем, что параметры θ содержат важную физическую информацию. Поэтому наша задача состоит в получении оценок $\hat{\theta}$ этих параметров.

Если число наблюдаемых величин равно числу параметров и если уравнения (II.3) совместны, то можно найти единственное решение θ, удовлетворяющее системе (II.3); в этом случае ошибки $\boldsymbol{\varepsilon}$ должны обратиться в нуль.

[1]) См. [53], гл. 4, 5.

Поскольку каждое y_i содержит экспериментальную ошибку, мы обычно стараемся, чтобы число наблюдаемых величин превышало количество параметров. Это дает нам возможность получить более точные оценки $\hat{\theta}$ параметров θ и в то же время найти оценки их дисперсий. Условие $n > m$ соответствует «переопределенности» параметров. Из-за наличия ошибок ε данному набору наблюдаемых величин $\mathbf{y}$ отвечает, вообще говоря, не единственное решение для θ. [В случае неизвестных θ и ε система уравнений (II.3) имеет бесконечное число решений.] Конечно, можно было бы выбрать любые n уравнений и решить их относительно m параметров. Однако мы интересуемся наиболее точным решением.

Будем считать, что ошибки $\boldsymbol{\varepsilon}$ подчиняются совместно му распределению с нулевым математическим ожиданием, т. е. что

$$\mathscr{E}(\mathbf{y}) \equiv \mathbf{y}_0 = A\theta, \tag{II.4}$$

и ковариационная матрица $\mathfrak{D}(\mathbf{y})$ ранга n

$$\mathfrak{D}(\mathbf{y}) \equiv \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \dots & \sigma_1\sigma_n\rho_{1n} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \dots & \sigma_2\sigma_n\rho_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_1\sigma_n\rho_{1n} & \sigma_2\sigma_n\rho_{2n} & \dots & \sigma_n^2 \end{pmatrix}, \tag{II.5}$$

где

$$\sigma_i^2 = \mathscr{E}(\varepsilon_i^2) = \mathscr{D}(y_i) \tag{II.6}$$

и

$$\sigma_i\sigma_j\rho_{ij} = \mathscr{E}(\varepsilon_i\varepsilon_j) = \operatorname{Cov}(y_i, y_j). \tag{II.7}$$

Таким образом, одно из основных предположений состоит в том, что ошибки ε_i величин y_i принадлежат распределениям с конечными вторыми моментами. Здесь можно отметить, что это условие является единственным и в методе наименьших квадратов вовсе не требуется предположения о нормальном законе распределения ошибок, хотя обычно это условие ошибочно считают необходимым. Можно показать, что существуют оценки

$$\theta \equiv \{\hat{\theta}_j\}, \tag{II.8}$$

такие, что для любой линейной функции от θ

$$f = U\theta \tag{II.9}$$

оценка

$$\hat{f} = U\hat{\theta} \tag{II.10}$$

будет несмещенной оценкой f с минимальной дисперсией. Оценки $\hat{\theta}$, обладающие таким свойством, не зависят от вида U.

Пример. Допустим, что имеется два параметра θ_1 и θ_2. Линейную функцию f (II.9) можно записать в виде

$$f = u_1\theta_1 + u_2\theta_2$$

с оценкой (II.10)

$$\hat{f} = u_1\hat{\theta}_1 + u_2\hat{\theta}_2.$$

Далее, если оценки параметров являются несмещенными, то

$$\mathscr{E}(\hat{f}) = u_1\mathscr{E}(\hat{\theta}_1) + u_2\mathscr{E}(\hat{\theta}_2) \equiv u_1\theta_1 + u_2\theta_2 = f$$

и

$$\sigma^2(\hat{f}) = u_1^2\sigma^2(\hat{\theta}_1) + u_2^2\sigma^2(\hat{\theta}_2) + 2u_1u_2\,\mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_2).$$

Примечательно, что значения θ_1 и θ_2, приводящие к минимуму $\sigma^2(\hat{f})$, не зависят от u_1 и u_2.

Можно показать, что эти оценки $\hat{\theta}$ идентичны оценкам, полученным простым методом Гаусса, который рассматривал специальный случай

$$\mathfrak{D}(\mathbf{y}) \equiv I, \tag{II.11}$$

т. е. случай некоррелированных ошибок ε_i с одинаковыми дисперсиями. Метод наименьших квадратов Гаусса основан на минимизации остаточной суммы квадратов

$$R = \sum_i (y_i - \hat{y}_i)^2, \tag{II.12}$$

где

$$\hat{\mathbf{y}} = A\hat{\theta}. \tag{II.13}$$

Если ввести вектор остатков

$$\mathbf{V} \equiv \mathbf{y} - \hat{\mathbf{y}} \equiv \mathbf{y} - A\hat{\theta}, \tag{II.14}$$

то минимизация величины R, которая определена равенствами (II.12), эквивалентна минимизации квадратичной формы

$$R = \mathbf{V}'\mathbf{V}. \tag{II.15}$$

Это означает, что в случае неединичной матрицы $\mathfrak{D}(\mathbf{y})$ соответствующая задача состоит в нахождении минимума формы

$$R = \mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{V}. \tag{II.16}$$

Преобразуем это выражение, а именно

$$\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{V} = (\mathbf{y} - A\hat{\theta})'\mathfrak{D}^{-1}(\mathbf{y})\,(\mathbf{y} - A\hat{\theta}) = \mathbf{y}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{y} + \\ + \hat{\theta}'A'\mathfrak{D}^{-1}(\mathbf{y})\,A\hat{\theta} - \mathbf{y}'\mathfrak{D}^{-1}(\mathbf{y})\,A\hat{\theta} - \hat{\theta}'A'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{y}. \tag{II.17}$$

Вводя дифференциальный оператор δ, мы получим условие минимума:

$$\delta[\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{V}] = 2(\delta\hat{\theta})'[A'\mathfrak{D}^{-1}(\mathbf{y})\,A\hat{\theta} - A'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{y}] = 0, \tag{II.18}$$

которое приводит к системе так называемых нормальных уравнений

$$[A'\mathfrak{D}^{-1}(\mathbf{y})\,A]\,\hat{\theta} = A'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{y},$$

или

$$B\hat{\theta} = A'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{y}, \tag{II.19}$$

где $B \equiv A'\mathfrak{D}^{-1}(\mathbf{y})\,A$ — матрица системы нормальных уравнений. Уравнение (II.19) можно решить относительно искомых оценок $\hat{\theta}$:

$$\hat{\theta} = B^{-1}A'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{y}. \tag{II.20}$$

Замечательно, что минимизация обобщенной суммы квадратов остатков приводит к тем же наилучшим оценкам $\hat{\theta}$, что и минимизация дисперсии произвольной линейной комбинации от $\hat{\theta}$. Полученная методом наименьших квадратов оценка (*мнк*-оценка) $\hat{\theta}$ соответствует минимуму дисперсии оценки *любой* линейной комбинации и параметров, и в частности по этой причине *мнк*-оценки имеют столь важное значение. Следовало бы еще раз подчеркнуть, что *мнк*-результаты не требуют специальных ограничений.

на распределение ошибок величин y. Как уже отмечалось, необходима лишь конечность вторых моментов, хотя часто неверно считают, что метод наименьших квадратов основан на предположении о нормальном законе распределения ошибок. Возможно, это заблуждение обусловлено тем, что лишь при нормальном законе распределения ошибок принцип максимального правдоподобия приводит к тому же решению $\hat{\theta}$, что и метод наименьших квадратов. Таким образом, в случае нормально распределенных ошибок *мнк*-оценка идентична оценке, полученной методом максимального правдоподобия.

Теперь допустим, что $\mathfrak{D}(\mathbf{y})$ известна лишь с точностью до некоторого множителя:

$$\mathfrak{D}(\mathbf{y}) = \sigma^2 N \tag{II.21}$$

и

$$\mathfrak{D}^{-1}(\mathbf{y}) = \frac{1}{\sigma^2} N^{-1}, \tag{II.22}$$

где матрица N известна, а величина σ^2 неизвестна. Иными словами, известны относительные, а не абсолютные значения элементов ковариационной матрицы. Тогда нормальные уравнения приобретают вид

$$\frac{1}{\sigma^2} A' N^{-1} A\hat{\theta} = \frac{1}{\sigma^2} A' N^{-1}\mathbf{y}. \tag{II.23}$$

Величина σ^2, стоящая в обеих частях уравнения, сокращается, и решение в итоге не зависит от абсолютной величины элементов ковариационной матрицы. Матрица

$$W \equiv N^{-1} \tag{II.24}$$

называется *весовой матрицей*, а σ^2 — дисперсия наблюдения с единичным весом. Пусть для примера

$$\mathfrak{D}(\mathbf{y}) = \sigma^2 I. \tag{II.25}$$

Тогда весовая матрица

$$I^{-1} = I, \tag{II.26}$$

т. е. веса всех наблюдаемых величин равны единице, а их дисперсии одинаковы и равны σ^2.

Ниже мы убедимся, что σ^2 можно рассматривать как оцениваемый параметр и что эта оценка обладает многими

свойствами критерия согласия. Перепишем *мнк*-результат в виде

$$\hat{\theta} = (A'WA)^{-1}A'W\mathbf{y}. \tag{II.27}$$

В случае нематричных обозначений мы имеем дело с системой линейных уравнений

$$\sum_{j=1}^{m} b_{ij}\hat{\theta}_j = c_i, \qquad i = 1, \ldots, m, \tag{II.28}$$

где

$$b_{ij} = \sum_{k=1}^{n}\sum_{r=1}^{n} a_{ki}a_{rj}w_{kr} \tag{II.29}$$

и

$$c_i = \sum_{k=1}^{n}\sum_{r=1}^{n} a_{ki}y_r w_{kr}. \tag{II.30}$$

Если весовая матрица диагональна, то

$$w_{ij} = 0 \quad \text{при} \quad i \neq j \tag{II.31}$$

и

$$w_i \equiv w_{ii} \tag{II.32}$$

могут быть приписаны отдельным уравнениям. Тогда соотношения (II.29) и (II.30) сведутся к

$$b_{ij} = \sum_{k=1}^{n} a_{ki}a_{kj}w_k, \tag{II.33}$$

$$c_i = \sum_{k=1}^{n} a_{ki}y_k w_k. \tag{II.34}$$

Это эквивалентно умножению каждого из уравнений (II.1) на квадратный корень из соответствующего веса и переходу к новой системе уравнений, не содержащей веса, т. е.

$$a^*_{pq} = w_p^{1/2}a_{pq}, \tag{II.35}$$

$$y^*_p = w_p^{1/2}y_p, \tag{II.36}$$

$$b_{ij} = \sum_{k=1}^{n} a^*_{ki}a^*_{kj}, \tag{II.37}$$

$$c_i = \sum_{k=1}^{n} a^*_{ki}y^*_k. \tag{II.38}$$

Ошибки оценок параметров. После того как в результате решения системы нормальных уравнений получены оценки, остается показать, что ковариационная матрица (или матрица ошибок) этих оценок равна обратной матрице системы нормальных уравнений, т. е. B^{-1}. Искомую матрицу можно записать в виде

$$\mathfrak{D}(\hat{\theta}) = \mathscr{E}[(\hat{\theta}-\theta)(\hat{\theta}-\theta)']. \tag{II.39}$$

Тогда

$$\mathfrak{D}(\theta) = \mathscr{E}[B^{-1}A'\mathfrak{D}^{-1}(\mathbf{y})(\mathbf{y}-\mathbf{y}_0)(\mathbf{y}-\mathbf{y}_0)'\mathfrak{D}^{-1}(\mathbf{y})AB^{-1}] =$$
$$= B^{-1}A'\mathfrak{D}^{-1}(\mathbf{y})\mathscr{E}[(\mathbf{y}-\mathbf{y}_0)(\mathbf{y}-\mathbf{y}_0)']\mathfrak{D}^{-1}(\mathbf{y})AB^{-1}. \tag{II.40}$$

Но поскольку

$$\mathfrak{D}(\mathbf{y}) \equiv \mathscr{E}[(\mathbf{y}-\mathbf{y}_0)(\mathbf{y}-\mathbf{y}_0)'], \tag{II.41}$$

получаем

$$\mathfrak{D}(\theta) = B^{-1}A'\mathfrak{D}^{-1}(\mathbf{y})\mathfrak{D}(\mathbf{y})\mathfrak{D}^{-1}(\mathbf{y})AB^{-1} = B^{-1}, \tag{II.42}$$

что и утверждалось.

Далее, коль скоро

$$B = A'\mathfrak{D}^{-1}(\mathbf{y})A = \frac{1}{\sigma^2}A'WA, \tag{II.43}$$

оказывается, что

$$\mathfrak{D}(\hat{\theta}) \equiv B^{-1} = \sigma^2(A'WA)^{-1}. \tag{II.44}$$

Чтобы получить ковариационную матрицу оценок параметров, нужно либо знать σ^2 — дисперсию наблюдения с единичным весом, и тогда $\mathfrak{D}(\theta)$ полностью определена, либо иметь несмещенную оценку σ^2, и тогда мы будем иметь несмещенные оценки элементов искомой матрицы. Ниже будет показано, что метод наименьших квадратов действительно позволяет получить несмещенную оценку σ^2.

Рассмотрим величину

$$\mathscr{E}(\mathbf{V}'W\mathbf{V}) \equiv \sigma^2\mathscr{E}[\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\mathbf{V}], \tag{II.45}$$

которая представляет собой обобщенную (взвешенную) остаточную сумму квадратов. Далее

$$\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\mathbf{V} = \mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})(\mathbf{y}-A\hat{\theta}) = \mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\mathbf{y}, \tag{II.46}$$

поскольку условие

$$\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})A = 0 \tag{II.47}$$

эквивалентно системе нормальных уравнений. Кроме того, обращаясь к нормальным уравнениям, можно показать, что

$$\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{y} = (\mathbf{y}' - \hat{\theta}'A')\,\mathfrak{D}^{-1}(\mathbf{y})\,(\mathbf{y}) =$$

$$= \mathbf{y}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{y} - \hat{\theta}'A'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{y} = \mathbf{y}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{y} - \hat{\theta}'B\hat{\theta}. \quad \text{(II.48)}$$

Полученное выражение равно

$$(\mathbf{y} - \mathbf{y}_0)'\,\mathfrak{D}^{-1}(\mathbf{y})\,(\mathbf{y} - \mathbf{y}_0) - (\hat{\theta} - \theta)'\,B\,(\hat{\theta} - \theta). \quad \text{(II.49)}$$

Таким образом,

$$\mathscr{E}\,[\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{V}] = \mathscr{E}\,[(\mathbf{y} - \mathbf{y}_0)'\,\mathfrak{D}^{-1}(\mathbf{y})\,(\mathbf{y} - \mathbf{y}_0)] -$$

$$- \mathscr{E}\,[(\hat{\theta} - \theta')\,B\,(\hat{\theta} - \theta)]. \quad \text{(II.50)}$$

Можно показать, что это приводит к результату

$$\mathscr{E}\,[\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{V}] = n - m, \quad \text{(II.51)}$$

где n и m — ранг матриц $\mathfrak{D}^{-1}(\mathbf{y})$ и B соответственно[1]). Итак,

$$\mathscr{E}\,(\mathbf{V}'W\mathbf{V}) = (n - m)\,\sigma^2 \quad \text{(II.52)}$$

и несмещенная оценка σ^2 дается выражением

$$\hat{\sigma}^2 = \frac{\mathbf{V}'W\mathbf{V}}{n - m}. \quad \text{(II.53)}$$

Следовательно, несмещенная оценка $\mathfrak{D}(\theta)$:

$$\hat{\mathfrak{D}}(\hat{\theta}) = \frac{\mathbf{V}'W\mathbf{V}}{n - m}\,(A'WA)^{-1}, \quad \text{(II.54)}$$

---

[1]) Если x_i $(i = 1, \ldots, n)$ — случайные величины с нулевыми средними значениями, конечными дисперсиями и ковариационной матрицей $\mathfrak{D}$ ранга n, то легко показать, что

$$\mathscr{E}\,(\mathbf{X}'\mathfrak{D}^{-1}\mathbf{X}) = n.$$

Доказательство:

$$\mathscr{E}\,(\mathbf{X}'\mathfrak{D}^{-1}\mathbf{X}) \equiv \mathscr{E}\,[\mathrm{Sp}\,(\mathbf{X}'\mathfrak{D}^{-1}\mathbf{X})] =$$

$$= \mathscr{E}\,[\mathrm{Sp}\,(\mathbf{X}'\mathbf{X}\mathfrak{D}^{-1})] - \mathrm{Sp}\,\mathscr{E}\,(\mathbf{X}\mathbf{X}'\mathfrak{D}^{-1}) = \mathrm{Sp}\,\mathfrak{D}\mathfrak{D}^{-1} = \mathrm{Sp}\,I = n.$$

Соотношение (II.51) следует из того факта, что $\mathbf{y} - \mathbf{y}_0$ и $\hat{\theta} - \theta$ суть случайные величины с нулевыми средними и ковариационными матрицами ранга n и m соответственно.

или просто равна обратной матрице нормальных уравнений, умноженной на несмещенную оценку σ^2. Следует особенно подчеркнуть, что ни один из полученных выводов не требует предполагать нормальный закон для распределения ошибок величин y_i. Необходимо лишь, чтобы y_i были несмещенными оценками (с конечной дисперсией) математического ожидания генеральной совокупности. *Мнк*-оценка $\hat{\theta}$ представляет собой несмещенную оценку θ с минимальной дисперсией, которая не зависит от вида распределения; несмещенная оценка матрицы $\mathfrak{D}(\hat{\theta})$ дается формулой (II.54). Однако в задачах проверки гипотез мы будем считать распределение ошибок нормальным. Кроме того, когда распределение предполагается нормальным, оценка σ^2 (II.53), оставаясь несмещенной, обладает некоторыми оптимальными свойствами. Тогда, как уже указывалось, оценка максимального правдоподобия совпадает с *мнк*-оценкой.

Пример. Допустим, что с помощью трех измеренных величин y_1, y_2, y_3 надо определить два параметра θ_1 и θ_2, причем известно

$$y_1 = 3\theta_1 + \theta_2,$$
$$y_2 = 2\theta_1 + \theta_2,$$
$$y_3 = 3\theta_1 + 2\theta_2.$$

Все найденные на опыте величины оказались равными 4,0. Предположим, что дисперсии во всех трех случаях одинаковы и равны, например $2^2 = 4$, и что, кроме того, имеется коэффициент корреляции между первыми двумя величинами, равный 0,5. Тогда ненормированная ковариационная матрица наблюдений будет иметь вид

$$N = \begin{pmatrix} 4{,}0 & 2{,}0 & 0{,}0 \\ 2{,}0 & 4{,}0 & 0{,}0 \\ 0{,}0 & 0{,}0 & 4{,}0 \end{pmatrix}.$$

Весовая матрица

$$W = N^{-1} = \begin{pmatrix} 0{,}3333 & -0{,}1667 & 0{,}0000 \\ -0{,}1667 & 0{,}3333 & 0{,}0000 \\ 0{,}0000 & 0{,}0000 & 0{,}2500 \end{pmatrix}.$$

Кроме того,

$$\mathbf{y} = \begin{pmatrix} 4{,}0 \\ 4{,}0 \\ 4{,}0 \end{pmatrix} \text{ и } A = \begin{pmatrix} 3{,}0 & 1{,}0 \\ 2{,}0 & 1{,}0 \\ 3{,}0 & 2{,}0 \end{pmatrix}.$$

Производя необходимые вычисления, получаем

$$A'WA = \begin{pmatrix} 4{,}583 & 2{,}333 \\ 2{,}333 & 1{,}333 \end{pmatrix}$$

и

$$\hat{\theta} = (A'WA)^{-1} A'W\mathbf{y} = \begin{pmatrix} 1{,}00 \\ 0{,}75 \end{pmatrix}.$$

Вектор остатков

$$\mathbf{V} = \mathbf{y} - A\hat{\theta} = \begin{pmatrix} 0{,}25 \\ 1{,}25 \\ -0{,}50 \end{pmatrix}$$

и

$$\mathbf{V}'W\mathbf{V} = 0{,}5000.$$

Поскольку число наблюдений равно 3 и число параметров равно 2, несмещенная оценка σ^2

$$\hat{\sigma}^2 = \frac{\mathbf{V}'W\mathbf{V}}{n-m} = \frac{0{,}5000}{3-2} = 0{,}5000,$$

а оценка ковариационной матрицы параметров

$$\hat{\mathfrak{D}}(\hat{\theta}) = \frac{\mathbf{V}'W\mathbf{V}}{3-2}(A'WA)^{-1} = \begin{pmatrix} 1{,}000 & -1{,}750 \\ -1{,}750 & 3{,}437 \end{pmatrix}.$$

Оценки средних квадратичных отклонений

$$\hat{\sigma}(\theta_1) = (1{,}000)^{1/2} = 1{,}000,$$

$$\hat{\sigma}(\theta_2) = (3{,}437)^{1/2} = 1{,}854,$$

так что *мнк*-результат можно представить в виде

$$\hat{\theta}_1 = 1{,}00, \qquad \hat{\sigma}_1 = 1{,}00,$$

$$\hat{\theta}_2 = 0{,}75, \qquad \hat{\sigma}_2 = 1{,}85.$$

Однако у корреляционной матрицы

$$\rho_{ij} = \begin{pmatrix} 1{,}00 & -0{,}94 \\ -0{,}94 & 1{,}00 \end{pmatrix}$$

столь большие недиагональные элементы, что бессмысленно приводить оценки среднеквадратичных отклонений, не упоминая о сильной корреляции. Большая положительная ошибка в $\hat{\theta}_1$ потребует большой отрицательной ошибки в $\hat{\theta}_2$.

§ 2. Геометрическая интерпретация матрицы ошибок

Весьма полезной и поучительной может оказаться геометрическая интерпретация матрицы ошибок. Квадратичная форма

$$S_0 = (\hat{\theta} - \theta)' \mathfrak{D}^{-1}(\theta)(\hat{\theta} - \theta) \equiv \Delta' \mathfrak{D}^{-1}(\theta) \Delta \equiv \Delta' B \Delta \quad \text{(II.55)}$$

представляет собой уравнение гиперэллипсоида с центром в точке $\Delta = 0$ (нулевой вектор) или $\hat{\theta} = \theta$ в m-мерном пространстве параметров. В большинстве приложений плотность вероятности ошибок будет функцией лишь S_0, так что приведенное выше уравнение описывает гиперэллипсоид постоянной вероятности.

Рассмотрим гиперэллипсоид, соответствующий специальному случаю $S_0 = 1$; будем называть его гиперэллипсоидом среднеквадратичного отклонения. Если обозначить элементы вектора Δ через δ_i, то проекция этого гиперэллипсоида на плоскость (θ_1, θ_2) есть эллипс, уравнение которого

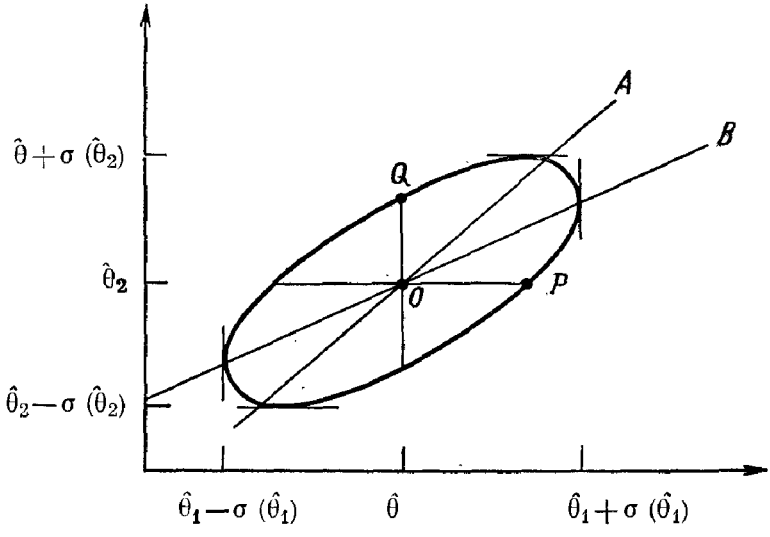
$$\frac{\delta_1^2}{\sigma_1^2} + \frac{\delta_2^2}{\sigma_2^2} - \frac{2\rho_{12}}{\sigma_1\sigma_2}\delta_1\delta_2 = 1 - \rho_{12}^2. \quad \text{(II.56)}$$

Проекции этого эллипса на оси координат характеризуют удвоенные значения *маргинальных* среднеквадратичных отклонений σ_1 и σ_2. Подобным образом, *маргинальное* среднее квадратичное отклонение для каждого из параметров равно половине длины проекции исходного гиперэллипсоида на соответствующую ось. Если δ_1 фиксировано, то

$$\mathscr{E}\{\delta_2 \mid \delta_1\} = \rho_{12}\frac{\sigma_2}{\sigma_1}\delta_1 \equiv -\frac{b_{21}}{b_{22}}\delta_1 \quad \text{(II.57)}$$

— *условное математическое ожидание* величины δ_2 — лежит на линии с наклоном $\rho_{12}\sigma_2/\sigma_1$, обозначенной на фиг. II.1

буквой B. Эта линия называется *линией регрессии* величины δ_2 по δ_1. Аналогичным образом линия с наклоном $(1/\rho_{12})\,\sigma_2/\sigma_1$, обозначенная на той же фигуре буквой A, представляет собой линию регрессии величины δ_1 по δ_2. *Условное среднеквадратичное отклонение* величины δ_2,



Ф и г. II.1. **Эллипс среднеквадратичного отклонения в случае двух коррелированных параметров.**

Проекции эллипса на оси координат являются маргинальными среднеквадратичными отклонениями. Отрезки OQ и OP представляют собой условные среднеквадратичные отклонения $\sigma\,(a \mid 1)$ и $\sigma\,(1 \mid a)$.

$\sigma_{2\,|\,1}$ равно половине длины отрезка прямой, параллельного оси δ_2, и ограничено размерами эллипса. Это условное среднеквадратичное отклонение не зависит от величины δ_1.

В общем случае m-мерного пространства можно записать

$$\mathscr{E}\{\delta_1 \mid \delta_2 \ldots \delta_m\} \equiv \delta_{1\,|\,2\ldots m} = \sum_{j\neq 1} \frac{-b_{1j}}{b_{11}}\theta_j. \qquad \text{(II.58)}$$

Величины $(-b_{ij}/b_{11})$ представляют собой парциальные коэффициенты регрессии θ_1 по θ_j. Подобным образом условная дисперсия дается выражением

$$\mathscr{E}\{\delta_1^2 \mid \delta_2 \ldots \delta_m\} \equiv \sigma^2_{1\,|\,2\ldots m} = \frac{1}{b_{11}}. \qquad \text{(II.59)}$$

Условная дисперсия всегда меньше маргинальной дисперсии, если корреляции между θ_i и θ_j не равны нулю при всех значениях j; в противном случае обе дисперсии равны друг другу.

Для более подробного рассмотрения распределения вероятности необходимо вводить специальные распределения ошибок наблюдаемых величин.

Вводя ортогональное преобразование, которое диагонализирует матрицу $\mathfrak{D}^{-1}(\theta) \equiv B$, можно найти такие линейные комбинации Δ_1 оценок параметров Δ_0, у которых взаимные ковариации равны нулю, что является условием статистической независимости в случае многомерного нормального распределения. Главные оси гиперэллипсоида ошибок будут параллельны новым осям параметров. Если

$$T\mathfrak{D}^{-1}(\theta)\,T' = M, \tag{II.60}$$

где матрица M диагональна, то

$$\Delta_0'\mathfrak{D}^{-1}(\theta)\,\Delta_0 \equiv \Delta_1' M \Delta_1 = \Delta_1' T \mathfrak{D}^{-1}(\theta)\,T'\Delta_1 \tag{II.61}$$

и

$$\Delta_1 = (T')^{-1}\Delta_0 = T\Delta_0, \tag{II.62}$$

поскольку матрица T предполагается ортогональной. Так как диагонализирована матрица B, то для нахождения независимых (в указанном смысле) линейных комбинаций оценок параметров нет необходимости искать *мнк*-решение.

§ 3. Многомерное нормальное распределение и свойства распределений квадратичных форм

Выше было указано, что наилучшие оценки параметров и дисперсий могут быть получены без специальных предположений относительно закона распределения ошибок. Однако ясно, что для определения доверительных интервалов и проверки гипотез такие предположения необходимы. Мы сосредоточим основное внимание на распределениях, имеющих отношение к нормальному закону. Плотность вероятности m-мерного нормального распределения с нулевыми средними может быть записана в виде

$$N(\theta) = C\exp\left(-\tfrac{1}{2}\theta'\mathfrak{D}^{-1}\theta\right), \tag{II.63}$$

где

$$C = \frac{[\det \mathfrak{D}^{-1}]^{1/2}}{(2\pi)^{m/2}} . \tag{II.64}$$

Матрица $\mathfrak{D}^{-1}$ квадратичной формы в экспоненте (II.63) представляет собой обратную ковариационной матрице. Это можно показать, используя интегральное выражение

$$\{\mathrm{Cov}(\theta_1, \theta_2)\} = C \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \theta\theta' \exp\left(-\frac{1}{2}\theta'\mathfrak{D}^{-1}\theta\right) d\theta_1 \dots d\theta_m. \tag{II.65}$$

Пусть

$$\theta = T\mathbf{Y}, \tag{II.66}$$

так что

$$\theta'\mathfrak{D}^{-1}\theta \equiv \mathbf{Y}'\mathbf{Y}. \tag{II.67}$$

Тогда

$$\{\mathrm{Cov}(\theta_1, \theta_2)\} = \frac{1}{(2\pi)^{m/2}} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} T\mathbf{Y}\mathbf{Y}'T' e^{-\mathbf{Y}'\mathbf{Y}/2} dy_1 \dots dy_m =$$

$$= T\left[\frac{1}{(2\pi)^{m/2}} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{-\infty} \mathbf{Y}\mathbf{Y}' e^{-\mathbf{Y}'\mathbf{Y}/2} dy_1 \dots dy_m\right] T'. \tag{II.68}$$

Но матрица в скобках есть единичная матрица, так как величины y_k независимы. Поэтому

$$\mathscr{E}(\theta\theta') = \mathrm{Cov}(\theta_i, \theta_j) = TIT' = TT'. \tag{II.69}$$

И поскольку

$$TT' = \mathfrak{D}, \tag{II.70}$$

в итоге получаем

$$\mathrm{Cov}(\theta_i, \theta_j) = \mathfrak{D}. \tag{II.71}$$

Нас будет интересовать распределение квадратичной формы

$$\theta'\mathfrak{D}^{-1}\theta. \tag{II.72}$$

Произведя линейное преобразование

$$\mathbf{Y} = T\theta, \tag{II.73}$$

получим

$$\theta' \mathfrak{D}^{-1} \theta = \mathbf{Y}'\mathbf{Y} = y_1^2 + y_2^2 + \ldots + y_m^2. \qquad \text{(II.74)}$$

Эта квадратичная форма представляет собой сумму квадратов m нормально распределенных величин с нулевыми средними и единичными среднеквадратичными отклонениями. Следовательно, она распределена как χ^2 с m степенями свободы.

В случае многомерного нормального распределения с отличными от нуля средними

$$N(\theta) = \frac{(\det \mathfrak{D}^{-1})^{1/2}}{(2\pi)^{m/2}} \exp\left[-\frac{1}{2}(\theta - \theta_0)' \mathfrak{D}^{-1} (\theta - \theta_0)\right], \qquad \text{(II.75)}$$

аналогично можно показать, что квадратичная форма

$$\theta' \mathfrak{D}^{-1} \theta \qquad \text{(II.76)}$$

следует нецентральному распределению χ^2 с параметром нецентральности

$$\xi^2 = \theta_0' \mathfrak{D}^{-1} \theta_0. \qquad \text{(II.77)}$$

Если θ_0 равно нулю, параметр нецентральности исчезает, и распределение переходит в обычное распределение χ^2. Полезно отметить, что в общем виде квадратичная форма $\theta' K \theta$ ранга q и порядка m распределена как χ^2 с q степенями свободы в том и только том случае, когда ранг квадратичной формы $\theta' (\mathfrak{D}^{-1} - K) \theta$ равен $m - q$.

§ 4. Распределение квадратичных форм в методе наименьших квадратов

Пусть дана квадратичная форма $\theta' B \theta$ ранга m, кроме того, предположим, что на θ_i наложены p независимых линейных связей

$$Q_{(p \times m)} \theta_{(m \times 1)} = 0_{(p \times 1)}. \qquad \text{(II.78)}$$

При таких условиях форма $\theta' B \theta$ распределена как χ^2 с $m - p$ степенями свободы, поскольку число независимых переменных θ_i теперь равно не m, а $m - p$.

Рассмотрим теперь квадратичную форму

$$\mathbf{V}' \mathfrak{D}^{-1} (\mathbf{y}) \mathbf{V}, \qquad \text{(I.I79)}$$

где $\mathbf{V}$ — вектор остатков, полученных при *мнк*-аппроксимации. В § 1 дополнения II мы показали, что

$$\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\,A = 0, \tag{II.80}$$

так что, введя обозначение

$$Q' = \mathfrak{D}^{-1}(\mathbf{y})\,A, \tag{II.81}$$

получим

$$Q\mathbf{V} = \mathbf{0}. \tag{II.82}$$

Матрица Q имеет ранг m, поэтому форма $\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{V}$ распределена как χ^2_{n-m}. Так как $\mathscr{E}(\chi^2_n) = n$, неудивительно, что для случая нормального распределения мы получили то же самое, что и в общем случае, а именно

$$\mathscr{E}[\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{V}] = n - m. \tag{II.83}$$

Таким образом, $\hat{\sigma} = \mathbf{V}'W\mathbf{V}/(n-m)$ представляет собой несмещенную оценку σ^2 и распределено как

$$\sigma^2 \frac{\chi^2_{n-m}}{n-m}. \tag{II.84}$$

Ранее мы видели, что

$$\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{V} = (\mathbf{y}-\mathbf{y}_0)'\,\mathfrak{D}^{-1}(\mathbf{y})\,(\mathbf{y}-\mathbf{y}_0) - (\hat{\theta}-\theta)\,B\,(\hat{\theta}-\theta).$$

Поскольку $(\mathbf{y}-\mathbf{y}_0)'\,\mathfrak{D}^{-1}(\mathbf{y})\,(\mathbf{y}-\mathbf{y}_0)$ распределено как χ^2_n, а $\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\,\mathbf{V}$ распределено как χ^2_{n-m}, то $(\hat{\theta}-\theta)'\,B\,(\hat{\theta}-\theta)$ должно быть распределено как χ^2_m; это следует также из того, что если существует *мнк*-решение, то ранг матрицы B равен m. Значит, если известен множитель σ^2 ковариационной матрицы, то отклонения от *мнк*-оценок можно было бы проверить с помощью критерия χ^2. Однако, как уже отмечалось, чаще всего σ^2 оценивается с помощью $\hat{\sigma}^2$, критерия согласия, и это вынуждает нас поступать несколько иначе.

Вводя прежнее обозначение оценки ковариационной матрицы искомых параметров

$$\hat{\mathfrak{D}}(\theta) \equiv \hat{\sigma}^2\,(A'WA)^{-1},$$

мы приходим к распределению следующей формы:

$$(\hat{\theta}-\theta)'\,\hat{\mathfrak{D}}^{-1}(\theta)(\hat{\theta}-\theta). \tag{II.85}$$

Так как

$$W = \sigma^2 \mathfrak{D}^{-1}(\mathbf{y}), \tag{II.86}$$

то

$$(\hat{\theta} - \theta)' \hat{\mathfrak{D}}^{-1}(\theta)(\hat{\theta} - \theta) = (\hat{\theta} - \theta)' B (\hat{\theta} - \theta) \frac{\sigma^2}{\hat{\sigma}^2}. \tag{II.87}$$

Поскольку $\hat{\sigma}^2$ распределено как $\sigma^2 \chi^2_{n-m}/(n-m)$ и не зависит от $(\hat{\theta} - \theta)' B (\hat{\theta} - \theta)$, то

$$\frac{(\hat{\theta} - \theta)' \hat{\mathfrak{D}}^{-1}(\theta)(\hat{\theta} - \theta)}{m} \tag{II.88}$$

будет распределено как

$$\frac{\chi^2_m/m}{\chi^2_{n-m}/(n-m)} \equiv \mathcal{F}_{m,\, n-m}. \tag{II.89}$$

Итак, чтобы проверить нулевую гипотезу H_0:

$$\theta = \theta_H, \tag{II.90}$$

необходимо вычислить величину

$$\frac{S_H}{m} = \frac{1}{m} (\hat{\theta} - \theta_H)' \hat{\mathfrak{D}}^{-1}(\theta)(\hat{\theta} - \theta_H)$$

и сравнить ее с критерием $\mathcal{F}_{m,\, n-m}$. Если S_H/m превышает $\mathcal{F}_{m,\, n-m,\, \alpha}$, то на уровне значимости α гипотезу H_0 можно отвергнуть. В дальнейшем (§ 6) этот вопрос будет рассмотрен в общем виде.

§ 5. Метод наименьших квадратов при наличии линейной связи параметров

Вернемся к задаче нахождения *мнк*-оценки параметров, но допустим теперь, что θ_i не независимы, а удовлетворяют системе линейных уравнений

$$Q\theta = \mathbf{Z}, \tag{II.91}$$

причем ранг матрицы Q равен b. Вводя множители Лагранжа $\Lambda_{(1 \times b)}$, построим функционал

$$G = \mathbf{V}' \mathfrak{D}^{-1}(\mathbf{y}) \mathbf{V} - 2\Lambda (Q\theta - \mathbf{Z}). \tag{II.92}$$

Пусть $\mathbf{y}'$ — *мнк*-оценка линейно-связанных параметров. Тогда

$$\begin{aligned} \delta G &= 2\mathbf{V}'\mathfrak{D}^{-1}(\mathbf{y})\,\delta\mathbf{V} - 2\Lambda Q\delta\theta, \\ 0 &= 2\left[-\mathbf{y}'\mathfrak{D}^{-1}(\mathbf{y})\,A + \theta' B - \Lambda Q\right]\delta\theta, \end{aligned} \tag{II.93}$$

т. е.

$$\Lambda Q = \overline{\overline{\theta}}' B - \mathbf{y}'\mathfrak{D}^{-1}(\mathbf{y})\,A. \tag{II.94}$$

Заметим, что

$$\mathbf{y}'\mathfrak{D}^{-1}(\mathbf{y})\,A = \hat{\theta}' B, \tag{II.95}$$

где $\hat{\theta}$ — *мнк*-оценка в отсутствие связи между параметрами. Таким образом,

$$\begin{aligned} \Lambda Q &= (\overline{\overline{\theta}} - \hat{\theta})'\,B, \\ \Lambda Q B^{-1} Q' &= (\overline{\overline{\theta}} - \hat{\theta})'\,Q' = \mathbf{Z}' - \hat{\theta}' Q'. \end{aligned} \tag{II.96}$$

В итоге

$$\Lambda = (\mathbf{Z}' - \hat{\theta}' Q')\,(QB^{-1}Q')^{-1}. \tag{II.97}$$

Исключая Λ из последних двух выражений, находим

$$(\overline{\overline{\theta}} - \hat{\theta})'\,B = (\mathbf{Z}' - \hat{\theta}' Q')\,(QB^{-1}Q')^{-1}\,Q,$$

или

$$\overline{\overline{\theta}}' = \hat{\theta}' + (\mathbf{Z}' - \hat{\theta}' Q')\,(QB^{-1}Q')^{-1}\,QB^{-1}. \tag{II.98}$$

Для решения уравнения (II.98) мы можем опять использовать матрицу $(A'WA)$ вместо матрицы B.

Поскольку B^{-1} — ковариационная матрица $\hat{\theta}$, ковариационной матрицей $\overline{\overline{\theta}}$ будет

$$\mathfrak{D}(\overline{\overline{\theta}}) = B^{-1} - B^{-1}Q'\,(QB^{-1}Q')^{-1}\,QB^{-1}. \tag{II.99}$$

Взвешенная остаточная сумма квадратов при наличии линейной связи между параметрами имеет вид

$$R_Q = \mathbf{V}'W\mathbf{V} + (\overline{\overline{\theta}} - \hat{\theta})'\,(A'WA)\,(\overline{\overline{\theta}} - \hat{\theta}), \tag{II.100}$$

где $\mathbf{V}$ — вектор остатков в случае независимых параметров, т. е. $\mathbf{V} = \mathbf{y} - A\theta$. Математическое ожидание R_a равно $(n - m + b)\,\sigma^2$. Второй член правой части равенства (II.100) можно переписать как

$$R_H = R_Q - R_0, \tag{II.101}$$

где $R_0 = \mathbf{V}'W\mathbf{V}$ — остаточная сумма квадратов при отсутствии связи параметров, а R_H — дополнительная остаточная сумма квадратов, возникающая из-за наличия связи. Простыми преобразованиями можно получить

$$R_H = (\mathbf{Z} - Q\hat{\theta})' (Q (A'WA)^{-1} Q')^{-1} (\mathbf{Z} - Q\hat{\theta}). \quad \text{(II.102)}$$

Из этого соотношения видно, что для вычисления величины R_H при наличии связи (II.91) вовсе нет необходимости решать уравнение (II.98).

§ 6. Многомерные линейные гипотезы

Теперь мы можем использовать некоторые результаты предыдущего параграфа для проверки гипотез и определения доверительных интервалов оценок параметров. Некоторые из этих результатов уже обсуждались. Так, мы показали, что $(1/m)\,(\hat{\theta} - \theta)'\,\mathfrak{D}^{-1}(\hat{\theta})\,(\hat{\theta} - \theta)$ имеет распределение $\mathcal{F}_{m,\,n-m}$. Таким образом, многомерная доверительная область задается гиперэллипсоидом

$$(\hat{\theta} - \theta)'\,\mathfrak{D}^{-1}(\hat{\theta})\,(\hat{\theta} - \theta) = m\mathcal{F}_{m,\,n-m,\,\alpha}. \quad \text{(II.103)}$$

Вероятность того, что эта область не содержит истинного значения θ, равна α. Ниже будет приведена более общая формулировка проблемы проверки гипотез.

Рассмотрим общую гипотезу, которая представляет собой некое утверждение относительно либо всех параметров, либо части из них. Ограничимся классом *линейных гипотез*, которые можно записать в виде

$$Q_{(b\times m)}\theta_{(m\times 1)} = \mathbf{Z}_{(b\times 1)}. \quad \text{(II.104)}$$

Ранг гипотезы равен рангу матрицы Q, который предполагается в свою очередь равным $b \leqslant m$. Нулевая гипотеза состоит в том, что истинное значение θ удовлетворяет приведенному уравнению. Естественной проверкой этой гипотезы могло бы быть сравнение остаточной суммы квадратов в том случае, когда верно H_0, с остаточной суммой квадратов, полученной при отсутствии связи. Использование результатов предыдущих параграфов показывает, что, если гипотеза H_0 справедлива, остаточная сумма квадратов, отвечающая этой гипотезе, $R_H = R_Q - R_0$,

будет распределена как $\sigma^2\chi^2_b$ и не будет зависеть от R_0. Величина же R_0 распределена как $\sigma^2\chi^2_{n-m}$. Следовательно, отношение

$$\frac{R_H}{R_0} = \frac{R_Q - R_0}{R_0} \tag{II.105}$$

должно быть распределено как

$$\frac{b}{n-m}\mathcal{F}_{b,\,n-m}. \tag{II.106}$$

(Важно отметить, что R_Q/R_0 не следует $\mathcal{F}$-распределению, ибо R_Q и R_0 не являются независимыми величинами, распределенными как χ^2.) Как и в случае распределения Стьюдента t, распределение отношения R_H/R_0 не зависит от σ^2, а значит, от нормировки весовой матрицы.

Ранее было показано, что

$$R_H = (\mathbf{Z} - Q\hat{\theta})' \,(Q\,(A'WA)^{-1}\,Q')^{-1}\,(\mathbf{Z} - Q\hat{\theta}).$$

Кроме того,

$$R_0 = \mathbf{V}'W\mathbf{V} = (n-m)\,\hat{\sigma}^2,$$

где $\hat{\sigma}^2$ — несмещенная оценка σ^2 в отсутствие связи. Тогда

$$\frac{R_H}{R_0} = \frac{(\mathbf{Z} - Q\hat{\theta})' \,(Q\,(A'WA)^{-1}Q')^{-1}\,(\mathbf{Z} - Q\hat{\theta})}{(n-m)\,\hat{\sigma}^2}. \tag{II.107}$$

Но поскольку

$$\frac{A'WA}{\hat{\sigma}^2} = \hat{\mathfrak{D}}^{-1}\,(\hat{\theta}),$$

величина

$$\frac{n-m}{b}\,\frac{R_H}{R_0} = \frac{(\mathbf{Z} - Q\hat{\theta})' \,(Q\hat{\mathfrak{D}}^{-1}\,(\theta)\,Q')^{-1}\,(\mathbf{Z} - Q\hat{\theta})}{b} \tag{II.108}$$

распределена как $\mathcal{F}_{b,\,n-m}$. Мы отвергаем гипотезу, если вычисленное значение этой статистики превышает табличное значение $\mathcal{F}_{b,\,n-m,\,\alpha}$.

Одной из наиболее распространенных проверок гипотез является проверка непротиворечивости *мнк*-параметров предсказаниям некоторой теоретической модели. Пусть теоретический набор параметров есть θ^T. Тогда нулевая гипотеза

$$H_0\colon\ \theta = \theta^T, \tag{II.109}$$

т. е.

$$\begin{aligned} \theta_1 &= \theta_1^T, \\ \theta_2 &= \theta_2^T, \\ &\cdots\cdots \\ \theta_m &= \theta_m^T. \end{aligned} \tag{II.110}$$

Введем соответствующие величины Q и $\mathbf{Z}$:

$$Q_{(m\times m)} = I_{(m\times m)}; \qquad \mathbf{Z}_{(m\times 1)} = \theta^T. \tag{II.111}$$

Наша гипотеза является m-мерной, и по критерию $\mathscr{F}_{m,\, n-m}$ следует проверить статистику

$$\frac{(\hat{\theta}-\theta^{T'})\mathfrak{D}^{-1}(\hat{\theta})\,(\hat{\theta}-\theta^T)}{m}. \tag{II.112}$$

Здесь дан более общий подход к вопросу о многомерной доверительной области, нежели в начале параграфа.

В качестве другого крайнего случая можно проверить гипотезу, касающуюся лишь одного параметра безотносительно к значениям всех прочих. Выбирая в качестве такого параметра θ_1, получим

$$\begin{aligned} \mathbf{Q} &= (1, 0, 0, \ldots, 0), \\ \mathbf{Z} &= \theta_1^T. \end{aligned} \tag{II.113}$$

И если верна гипотеза

$$H_0\colon \theta_1 = \theta_1^T,$$

то величина

$$(n-m)\frac{R_H}{R_0} = \frac{(\hat{\theta}_1-\theta_1^T)^2}{\hat{\sigma}_1^2}, \tag{II.114}$$

где $\hat{\sigma}_1^2$ — оценка дисперсии σ_1^2, распределена как $\mathscr{F}_{1,\, n-m}$. Она просто совпадает с квадратом величины t критерия Стьюдента с $(n-m)$ степенями свободы. Значит доверительные интервалы оценок единичного параметра в многомерном случае задаются величиной t с использованием оценки маргинальной дисперсии. При этом, как и следовало ожидать, число степеней свободы равно числу наблюдаемых величин минус число искомых параметров.

Конечно, для большинства гипотез, представляющих интерес, значение b заключено между 1 и m.

Пример. Рассмотрим некоторые гипотезы относительно параметров, полученных во втором примере из § 1. Пусть

$$H_1: \theta_1 = 6{,}0,$$
$$H_2: \theta_2 = 6{,}0,$$
$$H_3: \theta_1 = 6{,}0;\ \theta_2 = 6{,}0,$$
$$H_4: \theta_1 = \theta_2.$$

Для проверки гипотез 1 и 2 вычислим

$$\frac{(\hat{\theta}_i - \theta_i)^2}{\hat{\sigma}_i^2} \qquad (i = 1, 2)$$

и сопоставим с табличным значением $\mathcal{F}_{1,1,\alpha}$. Вычисленные величины равны

$$\frac{5^2}{1} = 25 \text{ и } \frac{(5{,}25)^2}{3{,}437} = 8.$$

Поскольку $\mathcal{F}_{1;1;0{,}05} = 161{,}45$, мы делаем вывод, что ни одну из этих гипотез не следует отвергать на 5%-ном уровне значимости. Гипотеза H_3 имеет ранг 2, так что следует вычислить величину

$$\frac{(\theta^T - \hat{\theta})' \hat{\mathfrak{D}}^{-1}(\hat{\theta})(\theta^T - \hat{\theta})}{2} =$$
$$= \frac{1}{2}(5{,}00;\ 5{,}25)\begin{pmatrix} 9{,}167 & 4{,}667 \\ 4{,}667 & 2{,}667 \end{pmatrix}\begin{pmatrix} 5{,}00 \\ 5{,}25 \end{pmatrix} = 253$$

и сравнить ее с $\mathcal{F}_{2;\,1;\,0{,}05} = 199{,}50$. Таким образом, эта гипотеза должна быть отвергнута на 5%-ном уровне значимости.

Гипотеза H_4 утверждает просто равенство двух параметров безотносительно к их конкретному значению. Поэтому

$$\mathbf{Q} = (1;\ -1),$$
$$\mathbf{Z} \equiv z = 0,$$
$$\mathbf{Q}\hat{\mathfrak{D}}^{-1}(\hat{\theta})Q' = (1;\ -1)\begin{pmatrix} 1{,}000 & -1{,}750 \\ -1{,}750 & 3{,}437 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 7{,}937,$$
$$\frac{R_H}{R_0} = \frac{(z - \hat{\theta}_1 + \hat{\theta}_2)^2}{7{,}937} = \frac{(0{,}25)^2}{7{,}937} = 0{,}008.$$

Эту гипотезу явно нельзя отвергать. Установим 95%-ный доверительный интервал для разности $\theta_1 - \theta_2 = z$, полагая

$$\frac{(z-0{,}25)^2}{7{,}937} \leqslant 161{,}45$$

или

$$|z-0{,}25| < 35{,}8,$$
$$-35{,}6 < \theta_1 - \theta_2 < 36{,}0.$$

Следует заметить, что гипотезы не отвергались из-за большого значения $\mathcal{F}$, характерного для случая одной степени свободы знаменателя. Другими словами, оценка σ^2 является крайне слабой в силу того, что число наблюдаемых величин превышает число параметров лишь на единицу. При бóльшем числе степеней свободы для всех линейных функций, входящих в гипотезы, доверительный интервал будет существенно меньше. Например, увеличив число степеней свободы с единицы до двух, мы уменьшим величину $\mathcal{F}$ примерно в 10 раз.

§ 7. Мощность критерия $\mathcal{F}$ для многомерной линейной гипотезы

Для определения мощности критерия $\mathcal{F}$ необходимо исследовать распределение величины $\frac{R_H/b}{R_0/(n-m)}$ в случае, когда гипотеза не верна. Знаменатель этого отношения распределен как центральное χ^2, деленное на число степеней свободы, независимо от того, верна гипотеза или нет. С другой стороны, числитель распределен как центральное χ^2/b только в том случае, если гипотеза верна; в противном случае он распределен как нецентральное χ^2/b, причем параметр нецентральности

$$\xi^2 = \varphi' (QB^{-1}Q')^{-1} \varphi = \frac{\varphi' (Q (A'WA)^{-1} Q')^{-1} \varphi}{\sigma^2}, \quad \text{(II.115)}$$

где

$$\varphi = Q\theta - Z,$$

характеризует отклонение гипотезы от истины. Таким образом, интересующая нас статистика распределена как нецентральное $\mathcal{F}$. Одна из важнейших особенностей этого

распределения заключается в том, что мощность критерия при заданном числе степеней свободы всегда растет с ростом параметра нецентральности ξ^2.

§ 8. Выбор весов в методе наименьших квадратов

При нахождении *мнк*-решений предполагалось, что весовая матрица W известна с точностью до множителя. Это эквивалентно тому, что известны относительные значения элементов ковариационной матрицы наблюдаемых величин. Множитель σ^2, конечно, оценивается по критерию согласия этих величин с некоторой моделью. На практике весовая матрица часто оказывается диагональной, так что необходимы лишь относительные дисперсии каждой наблюдаемой величины.

На самом деле у нас будут лишь *оценки* дисперсий наблюдаемых величин. При отсутствии дополнительной информации о распределении ошибок в генеральной совокупности, из которой получена выборка данных, можно считать, что эти оценки дают наилучшие веса для проведения *мнк*-анализа. Если такие оценки не соответствуют истинным дисперсиям генеральной совокупности, то оценки дисперсий параметров будут смещенными. С другой стороны, оценки параметров будут несмещенными при условии, что таковы сами наблюдаемые величины. Этот следует из того, что, какой бы ни была ковариационная матрица $\mathfrak{D}$,

$$\mathscr{E}(\hat{\theta}) = \mathscr{E}(B^{-1}A'\mathfrak{D}^{-1}\mathrm{y}) = B^{-1}A'\mathfrak{D}^{-1}\mathscr{E}(\mathrm{y}) =$$
$$= B^{-1}A'\mathfrak{D}^{-1}\mathrm{y}_0 = \theta. \qquad \text{(II.116)}$$

Но если веса известны не точно, то эти несмещенные оценки параметров могут не обладать минимальными дисперсиями.

Одной из распространенных ошибок при использовании весов является пренебрежение корреляциями погрешностей наблюдаемых величин. Если считать весовую матрицу диагональной, в то время как имеются значительные корреляции, то это может сильно повлиять на полученные *мнк*-решения, и особенно оценки дисперсий найденных параметров. Если имеются какие-либо данные отно-

сительно корреляций, то их следует разумным образом ввести в недиагональные элементы весовой матрицы. В ряде случаев оценка таких корреляций оказывается столь же простой, что и оценка дисперсий. Мы проиллюстрируем это приведенным ниже примером.

Пример. Предположим, что число рассеянных на газе электронов измеряется с интервалом углов рассеяния φ, равным 0,01°. Процедура наблюдения состоит в измерении интенсивности $J(\varphi)$ под каждым из углов. Используя эти данные, можно с помощью *мнк*-анализа определить межатомные расстояния в молекулах рассеивающего газа. Предположим далее, что основные ошибки результатов обусловлены шумами с характерной длиной волны, т. е. типичным проявлением ошибки было бы обнаружение резкого возрастания измеряемой величины в одной угловой позиции, менее резкого в следующей и так далее с уменьшением этого эффекта по мере удаления от первоначальной позиции. Таким образом, должна существовать заметная, но монотонно падающая корреляция между ошибками в соседних точках измерений. Возможную форму матрицы корреляций можно было бы представить как

$$\rho_{ij} = \rho_{12}^{|i-j|}.$$

Так, для $\rho_{12} = 0{,}5$ мы имели бы

$$\rho_{ij} = \begin{pmatrix} 1 & 0{,}5 & 0{,}25 & 0{,}125 & \dots \\ 0{,}5 & 1 & 0{,}5 & 0{,}25 & \dots \\ 0{,}25 & 0{,}5 & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}.$$

Если матрица корреляций имеет такой простой вид, то существуют методы надежной оценки величины ρ_{12} (см., например, [33]). На самом же деле корреляции в экспериментах по дифракции электронов, вообще говоря, не столь просты. Поэтому для того, чтобы получить таким путем экспериментальную оценку характера и величины корреляций, присущих данному методу, лучше всего проанализировать результаты нескольких *мнк*-аппроксимаций.

Последнее утверждение относится не только к задаче оценки корреляций, но в равной степени и к задаче оцен-

ки дисперсий. Поэтому часто оказывается более разумно использовать оценки дисперсии генеральной совокупности, полученные на основании большого опыта работы с данной экспериментальной методикой, нежели оценки, полученные просто из имеющихся под рукой данных. Это особенно важно, когда измерения представляют собой выборки крайне малых объемов, поскольку возникает опасность сильно переоценить одни веса по сравнению с другими. Как и в случае дисперсионного анализа, оценки дисперсий при наличии большого числа степеней свободы оказываются более устойчивыми, нежели оценки, полученные из данных с очень малым числом степеней свободы.

Пример. Допустим, что измеренная интенсивность рентгеновских лучей обнаруживает в области отдельного дифракционного пика $J(hkl)$ разброс, оцениваемый в среднем как $0{,}10 \times J(hkl)$. На одной из серий пленок были дважды обмерены несколько дифракционных пиков и найдены величины ожидаемых погрешностей. Два результата измерений $J(101)$ оказались равными 5020 и 5010. Оценка разброса на основе такой выборки равна $\sqrt{50}$. Тем не менее казалось бы неразумным приписать этому измерению вес $^1/_{50}$ для дальнейшего *мнк*-анализа. Надежнее было бы использовать при этом значение веса, равное $(1/501{,}5)^2$, взятое на основании опыта работы с данной экспериментальной методикой. Конечно, если столь малая величина оценки дисперсии встречалась неоднократно, следовало бы пересмотреть оценку разброса измеренных величин, равную $0{,}1J$. В противном случае экспериментатор должен предположить, что в измерениях существует систематическая погрешность. При определении весов важна величина не выборочной дисперсии, а дисперсии генеральной совокупности. Поэтому, если в нашем распоряжении есть оценка дисперсии генеральной совокупности и оценка выборочной дисперсии, то со всех точек зрения следует предпочесть первую из них.

Допустим, что ковариационная матрица наблюдаемых величин, а следовательно, и весовая матрица выбраны плохо. Возникает вопрос, можно ли предугадать, как это скажется на результатах *мнк*-анализа и при последующей проверке гипотез. Исследование такой проблемы проведе-

но в работах [36а, 36б]. Показано, что доверительные интервалы для линейных гипотез могут быть связаны довольно сложным образом с характеристическими числами истинной ковариационной матрицы. Хотя такой метод применим к более сложным случаям, при желании его вполне можно использовать и для выявления неверного взвешивания, проводя *мнк*-анализ и последующую проверку гипотез с варьированием весовой матрицы, возможно, даже за пределами разумных значений. Если результаты проверки гипотез оказываются весьма чувствительны к выбору весов, то целесообразно проверить такие гипотезы при более высоком уровне значимости, чем обычно. С другой стороны, результаты проверок гипотез могут оказаться довольно нечувствительными к значениям используемых весов. Это следует принять в качестве приятной неожиданности и проводить дальнейшую проверку гипотез с бóльшим доверием к результатам.

Пример. Пусть для некоторых измеренных величин не удалось найти выборочные дисперсии. При отсутствии дополнительной информации относительно дисперсий генеральной совокупности *всем* измеренным величинам следует приписать единичный вес. Важно отметить, что в таком случае можно не получить правильных оценок ошибок. Но если в проведенных ранее анализах подобных экспериментов успешно использовалась некая схема взвешивания, то ею, несомненно, следовало бы воспользоваться.

§ 9. Перенос ошибок

Допустим, что в результате *мнк*-анализа получены оценки параметров и их ошибок, и возникает необходимость определить ошибки линейных функций найденных параметров.

Пусть задано многомерное распределение, т. е. мы рассматриваем функцию распределения

$$\Phi(G) = \Phi(g_1, \ldots, g_n), \qquad (II.117)$$

где случайные величины g_i имеют нулевые средние значения. Пусть также задана матрица вторых моментов $\mathfrak{D}(G)$ ранга n. Если величины f_i $(i = 1, \ldots, m)$ представляют

собой линейные функции от g_i:

$$\mathbf{F}_{(m\times 1)} = T_{(m\times n)}\mathrm{G}_{(n\times 1)}, \qquad \text{(II.118)}$$

то

$$\mathfrak{D}(\mathbf{F}) \equiv \mathscr{E}\{\mathbf{F}\mathbf{F}'\} = \mathscr{E}\{T\mathrm{G}\mathrm{G}'T'\} = T\mathscr{E}\{\mathrm{G}\mathrm{G}'\}T' = T\mathfrak{D}(\mathrm{G})T'. \qquad \text{(II.119)}$$

Вновь отметим, что нет необходимости предполагать нормальным характер распределения; требуется лишь существование вторых моментов распределения. Вектор **F** может содержать больше, столько же или меньше компонент, чем вектор G. В любом случае ранг матрицы $\mathfrak{D}(\mathbf{F})$ не превышает ранга $\mathfrak{D}(\mathrm{G})$. Если m больше, чем n, то компоненты вектора **F** не будут линейно независимыми, матрица $\mathfrak{D}(\mathbf{F})$ будет особенной, а соответствующий m-мерный гиперэллипсоид ошибок — вырожденным, т. е. некоторые из его главных осей будут иметь нулевую длину.

Если G распределено по нормальному закону

$$\Phi(\mathrm{G}) = C\exp\left(-\frac{1}{2}\mathrm{G}'\mathfrak{D}^{-1}(\mathrm{G})\mathrm{G}\right), \qquad \text{(II.120)}$$

то, как было показано ранее, плотность распределения **F** будет иметь вид

$$\Phi(\mathbf{F}) = C'\exp\left(-\frac{1}{2}\mathbf{F}'\mathfrak{D}^{-1}(\mathbf{F})\mathbf{F}\right) \qquad \text{(II.121)}$$

при условии, что матрица $\mathfrak{D}(\mathbf{F})$ неособенная. Приведенные здесь соотношения полезны в том случае, когда требуется вычислить ошибки величин, которые представляют интерес с точки зрения физики, но непосредственно не извлекаются из *мнк*-анализа.

§ 10. Метод наименьших квадратов в случае нелинейной зависимости от параметров

Традиционное изложение метода наименьших квадратов и задачи проверки линейных гипотез не отвечает требованиям нелинейных задач. Между тем измеренные на опыте величины y_i нередко оказываются нелинейными функциями подлежащих оценке параметров θ_i. В данном

параграфе изложен подход к решению таких задач и рассмотрены изменения, которым могут подвергнуться полученные ранее результаты.

Линеаризация задачи. В общем случае мы будем иметь дело с набором n наблюдаемых величин y_i, являющихся функциями m параметров θ_j:

$$y_i = y_i(\theta_1, \ldots, \theta_m), \qquad i = 1, \ldots, n.$$

Эти функции можно разложить в ряд Тэйлора в окрестности точки $(\theta_1^0, \ldots, \theta_m^0)$, сохранив лишь члены первого порядка малости, а именно

$$y_i \approx y_i(\theta_1^0, \ldots, \theta_m^0) + \frac{\partial y_i}{\partial \theta_1}(\theta_1 - \theta_1^0) + \ldots + \frac{\partial y_i}{\partial \theta_m}(\theta_m - \theta_m^0) \tag{II.122}$$

или

$$y_i - y_i^0 \approx \sum_{j=1}^{m} \frac{\partial y_i}{\partial \theta_j}(\theta_j - \theta_j^0), \tag{II.123}$$

т. е.

$$\Delta y_i = \sum_{j=1}^{m} \frac{\partial y_i}{\partial \theta_j} \Delta\theta_j \quad (i = 1, \ldots, n). \tag{II.124}$$

Если разложение справедливо, то тем самым задача сведена к линейной, в которой матрица A представляет собой матрицу частных производных первого порядка от исходных функций по соответствующим параметрам, а вектор наблюдаемых величин задается отклонениями измеренных значений от значений функций в точках разложения. Таким образом,

$$A \equiv \{a_{ij}\} \equiv \left\{\frac{\partial y_i}{\partial \theta_j}\right\}, \tag{II.125}$$

$$\mathbf{y} \equiv \{\Delta y_i\}, \tag{II.126}$$

$$\boldsymbol{\theta} \equiv \{\Delta\theta_j\}. \tag{II.127}$$

Дальнейшая процедура такова: сначала выбираются приближенные значения параметров θ_j^0, т. е. по возможности наиболее близкие к искомым значениям θ_j. Значения функций, вычисленные для этих приближенных значений

параметров, равны y_i^0. Разности между наблюдаемыми величинами функций и вычисленными

$$\Delta y_i \equiv y_i - y_i^0 \equiv y_i^{\text{набл}} - y_i^{\text{выч}}$$

играют роль линейных функций в обычном методе наименьших квадратов. Для нахождения $\Delta\theta_j$ решаются обычные *мнк*-уравнения. В итоге получаются уточненные значения параметров:

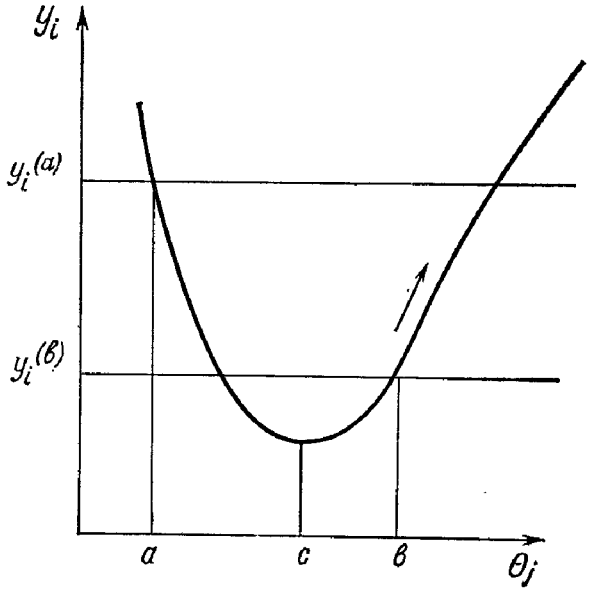
$$\theta_j = \theta_j^0 + \Delta\theta_j. \qquad \text{(II.128)}$$

Эти θ_j используются в качестве новых приближенных значений для вычисления новых величин $y_i^{\text{выч}}$, которые должны лучше согласовываться с наблюдаемыми величинами, чем предыдущие. Это согласие, однако, может оказаться еще не столь хорошим, как требуется, и тогда вся процедура повторяется снова. Если исходная задача, решаемая методом наименьших квадратов, линейная, то уже на первом этапе решение нормальных уравнений дает искомый результат. В случае нелинейной задачи необходимо использовать итерационную процедуру, в процессе которой получаются новые значения параметров, и, следовательно, производные и конструкционная матрица будут изменяться. Итерационная процедура продолжается до тех пор, пока изменения параметров не станут крайне малыми или равными нулю. При этом говорят, что итерационная процедура в методе наименьших квадратов *сходится*. Последовательное приближение к наилучшему решению иногда называют *уточнением* параметров.

В процессе подобного уточнения каждая итерация выполняется так же, как и в линейном случае. Веса как обычно выбираются обратно пропорциональными дисперсиям наблюдаемых величин. После того как уточнение получено, предполагается, что ошибки параметров задаются обратной матрицей системы нормальных уравнений, и проводятся, как обычно, проверки гипотез относительно параметров или функций от параметров. Эту процедуру следует проводить с некоторой осторожностью и тщательно исследовать влияние нелинейных эффектов. В дальнейшем мы вернемся к этому вопросу.

Теперь рассмотрим несколько подробнее ряд этапов метода наименьших квадратов в нелинейном случае.

Справедливость разложения в ряд Тэйлора. Применение метода наименьших квадратов в нелинейных задачах справедливо лишь постольку, поскольку справедливо разложение в ряд Тэйлора в той области значений параметров, где ищется уточнение. Укажем на две из возможных трудностей, присущих данной задаче. Рассмотрим функцию y_i, изображенную на фиг. II.2, в зависимости от



Фиг. II.2. Функция y_i, нелинейно зависящая от параметра θ_j. Если правильное значение θ_j равно a и если в некоторой итерации *мик*-решения нелинейной задачи θ_j оказывается равным b, то при использовании этой функции в последующих итерациях величина параметра θ_j будет изменяться в неправильном направлении, указанном стрелкой.

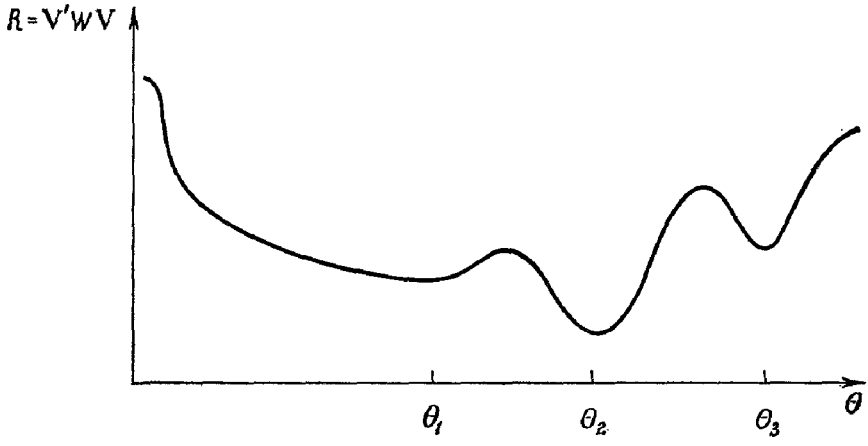
параметра θ_j. Допустим, что наилучшая аппроксимация исходных данных достигается при наборе параметров, в котором θ_j равно a. Соответствующее значение величины y_i есть $y_i^{(a)}$, и для упрощения будем считать, что оно совпадает с наблюдаемым значением y_i. Если теперь окажется, что на некотором этапе уточнения этот параметр принимает значение b, а y_i — соответствующее значение $y_i^{(b)}$, то в соответствии с разложением в ряд Тэйлора согласие будет улучшаться при возрастании θ_j, вместо стремления

к a. И действительно, согласие для этой функции будет улучшено, однако при этом, возможно, ухудшится согласие для других функций. Таким образом, уточнение этой функции будет противопоставлено остальным. Если к тому же она обладает большим весом или если имеется довольно много функций одинакового вида, то возможно (и даже вероятно), что итерационная процедура никогда не сойдется к решению с минимальной дисперсией. Скорее всего эта процедура будет сходиться к неверному решению, которое вполне может дать лишь относительный минимум для остаточной суммы квадратов.

Эту тенденцию процесса уточнения сходиться к ложному минимуму остаточной суммы квадратов можно иногда обнаружить, задавая различные начальные значения параметров. В том случае, когда число параметров невелико, можно эффективно исключить эту опасность, варьируя начальные значения в разумных пределах (если таковые на самом деле известны) и проверяя, всегда ли уточнение сходится к одному и тому же значению. Сходимость к правильному значению в каждом случае зависит от деталей структуры многопараметрической поверхности ошибок в окрестности начальной точки. Тем не менее можно сделать общее утверждение о том, что сходимость очень маловероятна, если значения производных большого числа наблюдаемых величин меняют знак. И наоборот, если изменения знака отсутствуют, то сходимость почти всегда гарантирована.

Данное утверждение проиллюстрировано на фиг. II.3, где отмечена еще одна возникающая при этом опасность. Может так случиться, что два сильно отличающихся друг от друга набора параметров приводят к одинаковым (или почти одинаковым) значениям остаточной суммы квадратов. С математической точки зрения каждый из этих наборов является приемлемым решением, и получение того или другого набора зависит от выбора начальных значений. Если обнаружена подобная ситуация, то основания для отбора одного из двух решений следует искать, не прибегая к статистике. В большинстве случаев мы вынуждены поставить вопрос: какое из решений более разумно с физической точки зрения, и найти достаточно мужества признать, что эксперимент не дал однозначного ответа.

Если начальное значение соответствует экстремуму функции по некоторому параметру, например точка c на фиг. II.2, то возникает еще одна трудность. В этой точке исчезает первая производная, и из разложения в ряд Тэйлора, представленного только первым членом, следует (конечно, неверно), что при изменении параметра согласие не улучшается и не ухудшается. Если подобная ситуация возникает лишь у небольшой части функций, то это не



Фиг. II.3. Схематическое изображение зависимости R от значения параметра θ.

Точка θ_2 соответствует абсолютному минимуму R и является искомой. Точки θ_1 и θ_3 соответствуют относительным минимумам той же величины. Если выбранное для уточнения начальное значение параметра находится вблизи этих точек, то истинное значение параметра скорей всего не будет найдено. Во всяком случае при использовании модифицированного метода Гаусса — Ньютона обеспечена сходимость к «ложному минимуму» θ_1 или θ_3.

приводит к серьезным затруднениям в процедуре уточнения. Но если при некотором значении параметра все функции достигают экстремума и если итерационный процесс происходит вслепую (что, к сожалению, часто бывает в большинстве задач), то получающаяся матрица нормальных уравнений оказывается особенной, и решение не существует.

Один из возможных способов обойти эту трудность состоит в следующем. В данном случае функции y_i можно выразить в виде

$$y_i = y_i^0 + \sum_{k \neq j} \frac{\partial y_i}{\partial \theta_k} \Delta\theta_k + \frac{1}{2} \frac{\partial^2 y_i}{\partial \theta_j^2} (\Delta\theta_j)^2. \qquad \text{(II.129)}$$

Если все функции имеют такой вид, то можно рассматривать $(\Delta\theta_j)^2$ как оцениваемый параметр и действовать обычным образом, используя в конструкционной матрице вторые производные по θ_j. Решение может привести к неожиданному результату — отрицательному значению $(\Delta\theta_j)^2$; это может означать, что в начальной точке или вблизи нее находится истинный минимум остаточной суммы квадратов, т. е. что $\Delta\theta_j$ в действительности равно нулю. В таком случае к этой части задачи на самом деле неприменим метод наименьших квадратов, поскольку этот метод требует, чтобы экстремум лежал внутри области возможных значений параметра. Если значения параметров лежат в некоторой области, заданной соответствующими неравенствами, то для получения решения можно использовать методы *линейного программирования*.

Если подобная ситуация имеет место лишь для небольшой части функций, которые все же желательно учитывать при поиске уточнения, их можно записать как

$$(y_i - y_i^0)^{1/2} = \pm \left(\frac{1}{2}\frac{\partial^2 y_i}{\partial\theta_j^2}\right)^{1/2} \Delta\theta_j + \ldots . \qquad (\text{II}.130)$$

К сожалению, в этом уравнении неизвестен знак, так что в начальной стадии процедуры уточнения лучше всего эти члены опустить. На заключительных этапах, когда обычно дело идет к оценкам ошибок параметров, такие члены могут сыграть большую роль. При некоторых обстоятельствах небольшие отклонения параметров от этих особых значений могут привести к большим расхождениям, и это следовало бы учитывать при оценке ошибок. При этом согласие будет весьма хорошим, так что величина $y_i - y_i^0$ будет малой. Тогда в обычную *мнк*-процедуру можно включить члены, представленные в приведенном выше уравнении, и хотя изменения параметров, которые последуют от включения этих членов, должны быть малыми, достоверность таких оценок, благодаря учету дополнительной информации, должна возрасти.

Выбор начальных значений параметров. Из всего сказанного выше ясно, что трудностей, связанных с разложением в ряд Тейлора, можно избежать, если выбрать

начальные значения параметров по возможности близкими к конечным или истинным значениям. Таким образом, варьируя начальные значения, можно изучить необходимую стратегию их выбора. В некоторых случаях, как, например, в классическом уточнении Дюмондом и Коэном [9a] фундаментальных физических констант, выбор начальных значений очевиден. В качестве исходных брались последние наилучшие значения констант, и метод наименьших квадратов давал истинные уточнения хорошо известных величин. Аналогичным образом можно поступать и во многих других экспериментальных задачах, когда накопленные новые данные позволяют получить более точные значения параметров, уже известных с хорошей степенью точности.

В большинстве других случаев для выбора начальных значений используются графические, аналитические или основанные на интуиции методы. Если известно или предполагается, что начальные значения далеки от истинных, можно применять следующий подход. На первых этапах процедуры уточнения следует использовать лишь те функции, которые довольно плавно зависят от значений параметров, если, конечно, таких функций достаточно для проведения *мнк*-анализа. На этой стадии следует опустить функции, сильно зависящие от параметров, поскольку они скорей всего и приводят к рассмотренным выше трудностям, а именно попаданию на «ложную» ветвь максимума или минимума функции. После того как первое уточнение проведено, такие функции можно учесть, возможно, сначала с малым весом, а впоследствии с полным весом. Из-за высокой чувствительности к изменениям определенных параметров эти функции крайне полезны для получения весьма точных оценок параметров. С другой стороны, приписывая им слишком большой вес на ранней стадии уточнения, мы можем в итоге достичь ложного минимума.

Сходимость *мнк*-процедуры в случае нелинейных задач. В случае нелинейных задач нельзя доказать, что *мнк*-процедура будет сходиться хотя бы к относительному минимуму взвешенной остаточной суммы квадратов $\mathbf{V}'W\mathbf{V}$. Однако Хартли [18a] показал, что при некоторых довольно-

но общих условиях сходимость всегда будет иметь место, если использовать следующую модификацию рассмотренной выше процедуры Гаусса — Ньютона. Для получения поправок $\Delta\theta$ к оценкам параметров как, обычно, решается система нормальных уравнений. Далее вводится скалярный множитель $0 \leqslant q \leqslant 1$, и исправленные значения оценок параметров записываются в виде

$$\theta_{\text{нов}} = \theta_{\text{стар}} + q\Delta\theta. \qquad \text{(II.131)}$$

Теперь квадратичная форма $\mathbf{V}'W\mathbf{V}$ будет функцией от q. Исходным для следующей итерации служит набор параметров, для которого $\mathbf{V}'W\mathbf{V}$ достигает минимума по q. Таким образом, обычное решение дает направление изменения поправок, а численная минимизация величины $\mathbf{V}'W\mathbf{V}$ позволяет получить их значение. Практически наилучшее q удобно находить, вычисляя $\mathbf{V}'W\mathbf{V}$ для $q = 0$, $^1/_2$ и 1 и аппроксимируя результаты параболой.

Найденный минимум может не быть абсолютным. В большинстве случаев судить о том, достигнут истинный минимум или этого не произошло, приходится на основании физического смысла полученных параметров и известного характера наблюдаемых функций, например степени их отклонения от линейности, кривизны и т. п.

На практике плохая сходимость может быть обусловлена ограниченной точностью вычислительного устройства. Если решение лежит вдали от минимума или если у некоторых параметров велики коэффициенты корреляции, т. е. матрица нормальных уравнений почти особенная, то ошибки округления даже при использовании вычислительной машины с большой длиной слова могут оказаться причиной неразрешимости задачи.

Проверка гипотез, нелинейных относительно параметров. В предыдущих параграфах отмечалось, что *мнк*-процедура в нелинейном случае все же часто сходится к искомому набору значений параметров, минимизирующему среднюю квадратичную ошибку. Однако оценить ошибки параметров и проверить линейные гипотезы оказывается совсем не просто.

Если оценки ошибок малы, и функции действительно линейны относительно каждого параметра в пределах нескольких средних квадратичных отклонений, то можно применять обычные методы проверки линейных гипотез. В случае нормального распределения ошибок наблюдаемых величин оценки параметров будут распределены также по нормальному закону, и тогда к соответствующим квадратичным формам можно применять критерий $\mathcal{F}$. Если в тех же пределах заметно отклонение от линейности, такой критерий неприменим, и, не зная точного характера функций, нельзя дать общей рекомендации относительно критериев проверок гипотез.

Часто нас будет интересовать проверка нелинейной гипотезы, имеющей вид

$$Q_k(\theta_1, \theta_2, \ldots, \theta_m) = Z_k, \qquad k = 1, 2, \ldots, b. \quad \text{(II.132)}$$

Мы не в состоянии сформулировать удовлетворительную процедуру проверки столь общей гипотезы и поэтому вновь вынуждены прибегнуть к линеаризации. Разложим Q_k в ряд Тэйлора в окрестности точки $(\theta_1^0, \ldots, \theta_m^0)$:

$$Q_k - Q_k^0 = \sum_{i=1}^{m} \frac{\partial Q_k}{\partial \theta_i} (\theta_i - \theta_i^0) + \ldots . \quad \text{(II.133)}$$

Если ограничиться только первым членом разложения, то мы придем к случаю линейной гипотезы, которую можно проверить стандартным образом, предполагая, что ковариационная матрица для $\Delta\theta_i$ получена методом наименьших квадратов. Следует особенно отметить, что эта линеаризация проводится помимо линеаризации (разложения в ряд Тэйлора) наблюдаемых функций. Поэтому вновь следует быть уверенными в справедливости линеаризации в той области значений параметров, которая используется при проверке гипотез. Эта область, вероятно, должна быть в несколько раз больше среднеквадратичного отклонения каждого из параметров, или лежать внутри гиперэллипсоида ошибок, концентричного с гиперэллипсоидом среднеквадратичных отклонений, но превышающего его в 3—4 раза. При проверке гипотез, нелинейных относительно параметров, которые получены с помощью линейного метода наименьших квадратов, мы,

конечно, должны столкнуться с аналогичными ограничениями.

Если проверяемая гипотеза особенно проста и состоит в том, что всем параметрам должны быть приписаны фиксированные значения, то мы имеем дело со случаем линейной гипотезы относительно параметров, найденных в нелинейной задаче. Тогда справедливость линейной гипотезы зависит только от возможности использовать квадратичную форму $\mathbf{V}'W\mathbf{V}/(n-m)$ в качестве оценки σ^2 и от инвариантности матрицы $B = A'WA$ в области значений параметров, рассматриваемой в этой гипотезе. В нелинейном случае практически, конечно, лучше вычислять $\mathbf{y}$ и, следовательно, $\mathbf{V}$ для проверяемых параметров, нежели использовать уравнение, справедливое в линейном случае:

$$\mathbf{V}'W\mathbf{V} = \mathbf{y}'W\mathbf{y} - \hat{\boldsymbol{\theta}}'B^{-1}\boldsymbol{\theta}. \qquad \text{(II.134)}$$

Согласие величин $\mathbf{V}'W\mathbf{V}$, полученных двумя разными способами, может служить критерием линеаризации в интересующей нас области изменения параметров. Вероятно, лучше всего вычислять обе величины и использовать любую из них на всякий случай, чтобы избежать ошибки при проверке гипотез. Как обычно, в случае грубых проверок может понадобиться провести проверку на большем уровне значимости по сравнению со случаем, когда к задаче строго применимы использованные методы.

Для проверки нелинейных гипотез могут быть использованы и более точные методы, однако они часто оказываются слишком сложными и, конечно, зависят от конкретного характера отклонения от линейности. Во многих физических задачах достаточно использовать критерии проверки, известные в линейных случаях, соблюдая при этом осторожность и тщательно исследуя влияние нелинейности.

Для оценки нелинейности в среднем Биль [3а] предложил использовать специальный параметр. По величине этого параметра можно оценить степень нелинейности и модифицировать обычные доверительные пределы, если нелинейность значительна. Подобная процедура, несмотря на рекомендации, не нашла у физиков широкого применения.

§ 11. Задача планирования эксперимента

Пусть уравнения для наблюдаемых величин записаны в виде

$$\mathbf{y} = A\theta + \varepsilon, \tag{II.135}$$

причем конструкционная матрица A является абстрактной характеристикой реально выполненного эксперимента. Матрица A показывает, какие конкретные функции параметров измерены на опыте. Одной из важных задач планирования эксперимента является предварительное детальное изучение матрицы A.

Какие функции следует измерять для получения нужной информации относительно параметров при минимальной затрате времени?

Как следовало бы выбрать матрицу A, чтобы сделать матрицу системы нормальных уравнений приблизительно диагональной, т. е. по возможности уменьшить корреляции между параметрами?

С какой точностью необходимо измерять функции, чтобы можно было проверить данную гипотезу H_0 относительно альтернативной гипотезы H_1 на уровне значимости α и с мощностью β?

Подобные вопросы составляют содержание задачи планирования эксперимента. Мы надеемся, что читатель уже достаточно хорошо знаком с проблемами и методами как *мнк*-аппроксимаций, так и проверки гипотез и в состоянии самостоятельно сформулировать и решить многие из поставленных вопросов. Здесь мы изложили частное решение одного из аспектов первого из вопросов.

Рассмотрим систему уравнений

$$\mathbf{y} = A\theta + \varepsilon,$$

и пусть ковариационная матрица $\mathbf{y}$ имеет вид

$$\mathfrak{D}(\mathbf{y}) = \sigma^2 W^{-1}. \tag{II.136}$$

При этом мы считаем, что можно изменять веса отдельных наблюдаемых величин путем некоторых изменений в эксперименте, например просто затрачивая больше времени на каждое отдельное измерение. Как уже известно,

общее *мнк*-решение имеет вид

$$(A'WA)\hat{\theta} = A'W\mathbf{y}. \tag{II.137}$$

Если W — диагональная матрица,

$$W = \begin{pmatrix} w_1 & \dots & 0 \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ 0 & \dots & w_n \end{pmatrix} \tag{II.138}$$

(или произвольная матрица, для которой определена величина $W^{1/2}$), то, как было указано ранее, умножая уравнения для наблюдаемых величин (II.137) на матрицу $W^{1/2}$, мы получаем

$$W^{1/2}\mathbf{y} = W^{1/2}A\theta + W^{1/2}\boldsymbol{\varepsilon}.$$

Теперь наблюдаемыми функциями будут

$$\mathbf{g} = W^{1/2}\mathbf{y},$$

а конструкционной матрицей $W^{1/2}A$. Ковариационной матрицей величин $\mathbf{g}$ будет матрица $\sigma^2 I$, характерная для метода наименьших квадратов без учета весов. Мы сделаем ряд замечаний в связи с этим аспектом планирования эксперимента, т. е. в связи с задачей выбора весов в случае линейной регрессии (если имеется возможность такого выбора). При этом мы ограничимся случаем диагональной весовой матрицы.

Прежде чем проводить эксперимент, нужно выяснить, какую цель он преследует. Вообще говоря, надо с самого начала формулировать линейные гипотезы, которые должны подвергаться проверке. Допустим, что цель эксперимента состоит в проверке линейной гипотезы

$$Q\theta = \mathbf{Z}. \tag{II.139}$$

Поставим вопрос: можно ли найти оптимальные веса для проверки этой гипотезы, и если да, то каким образом? С этой целью выберем веса так, чтобы сделать максимальной мощность критерия проверки гипотезы, т. е. свести к минимуму вероятность принятия ложной гипотезы.

Предварительно отметим, что мощность критерия $\mathscr{F}$ монотонно возрастает с ростом параметра нецентральности ξ^2. Следовательно, задача сводится к максимизации ξ^2 при любом дополнительном условии, наложенном по желанию.

Величина ξ^2 дается выражением

$$\xi^2 = \frac{\theta' [Q (A'WA)^{-1} Q']^{-1} \theta}{\sigma^2}. \qquad \text{(II.140)}$$

Если предполагать, что весовая матрица нормирована на величину σ^2, так что $W \equiv \mathfrak{D}^{-1}(y)$, то

$$\xi^2 = \theta' [Q (A'WA)^{-1} Q']^{-1} \theta. \qquad \text{(II.141)}$$

Теперь максимизируем ξ^2 при условии, что некоторая линейная функция весов

$$\sum_i r_i w_i \qquad \text{(II.142)}$$

остается постоянной величиной (причина такого условия вскоре будет очевидна). Для решения этой задачи будем варьировать функцию

$$G = \xi^2 + \lambda \sum_i r_i w_i, \qquad \text{(II.143)}$$

и если искомый экстремум существует, то его можно найти, решая систему нелинейных уравнений

$$\frac{\partial \xi^2}{\partial w_i} = -\lambda r_i, \quad i = 1, \ldots, n, \qquad \text{(II.144)}$$

т. е.

$$\frac{1}{r_i} \frac{\partial \xi^2}{\partial w_i} = \frac{1}{r_j} \frac{\partial \xi^2}{\partial w_j}, \quad i, j = 1, \ldots, n. \qquad \text{(II.145)}$$

Эту систему уравнений можно решить с помощью итерационной процедуры и определить величины w_i. Однако существует затруднение, связанное с тем, что w_i могут принимать только положительные значения. Решение, обеспечивающее максимум ξ^2, часто будет находиться на границе области

$$w_i \geqslant 0, \qquad i = 1, \ldots, n. \qquad \text{(II.146)}$$

Таким образом, некоторые из весов будут равны нулю. Если бы мощность критерия не возрастала с увеличением числа степеней свободы, то лишь m весов имели бы конеч-

ные значения. Поскольку не всегда удается достичь экстремум, данный метод оказывается недостаточным и задачу следует решать иными способами. В частности, для решения экстремальных задач, в которых наилучшие значения могут лежать на границе области, заданной неравенствами, применим метод *нелинейного программирования*. Существующие методы программирования могут оказаться недостаточными для решения имеющейся задачи в общем виде.

Предположим теперь, что экспериментальные данные представляют собой результаты измерений числа отсчетов, причем каждая из наблюдаемых величин подчиняется распределению Пуассона. (В дальнейшем будем полагать, что число отсчетов достаточно велико, чтобы распределение ошибок было приблизительно нормальным и для проверки гипотез был бы применим критерий $\mathcal{F}$.) Пусть необходимо измерить n различных скоростей счета с математическим ожиданием каждой из них, равным y_i^0 отсчетов в 1 *мин*. Для проведения измерений в нашем распоряжении имеется время T. Пусть на измерение величины y_i тратится T_i минут. Если количество отсчетов обозначить через N_i, то

$$y_i = \frac{N_i}{T_i} \quad \text{и} \quad y_i^0 = \frac{N_i^0}{T_i}. \tag{II.147}$$

Так как

$$\sigma^2(N_i^0) = N_i^0, \tag{II.148}$$

то дисперсия измеренной величины y_i равна

$$\sigma^2(y_i) = \frac{N_i^0}{T_i^2} = \frac{y_i^0}{T_i}. \tag{II.149}$$

Соответствующий вес i-го измерения

$$w_i = \frac{T_i}{y_i^0}. \tag{II.150}$$

Мы хотим максимизировать значение ξ^2 при условии, что

$$T = \sum_i T_i = \sum_i y_i^0 w_i \tag{II.151}$$

постоянно. Как было указано ранее, для сложных задач нет готовых методов такой максимизации. Однако ряд простых примеров продемонстрирует возможность подобного подхода.

Пример 1. Допустим, что имеются две измеренные величины и два параметра, связанные с ними простым соотношением

$$y_1^0 = \theta_1, \qquad y_2^0 = \theta_2,$$

причем веса этих величин равны соответственно w_1 и w_2. Тогда

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad (A'WA)^{-1} = \begin{pmatrix} 1/w_1 & 0 \\ 0 & 1/w_2 \end{pmatrix}.$$

Рассмотрим гипотезу

$$a\theta_1 + b\theta_2 = z,$$

т. е.

$$Q = (a, b).$$

В таком случае

$$Q\,(A'WA)^{-1}Q' = \frac{a^2}{w_1} + \frac{b^2}{w_2}$$

и величина ξ^2 равна

$$\xi^2 = \varphi^2\,\frac{w_1 w_2}{a^2 w_2 + b^2 w_1}.$$

Мы хотим максимизировать ξ^2 при условии

$$T_1 + T_2 = y_1^0 w_1 + y_2^0 w_2 = 1.$$

(Нормируем полное время на единицу.) Пусть

$$G = \xi^2 + \lambda\,(y_1^0 w_1 + y_2^0 w_2).$$

Тогда

$$\frac{\partial G}{\partial w_1} = \frac{a^2 w_2^2}{a^2 w_2 + b^2 w_1} + \lambda y_1^0 = 0,$$

$$\frac{\partial G}{\partial w_2} = \frac{b^2 w_1^2}{a^2 w_2 + b^2 w_1} + \lambda y_2^0 = 0.$$

Исключая величину λ, получаем

$$\frac{a^2 w_2^2}{b^2 w_1^2} = \frac{y_1^0}{y_2^0},$$

или

$$\frac{w_2}{w_1} = \left|\frac{b}{a}\right| \left(\frac{y_1^0}{y_2^0}\right)^{1/2}$$

и

$$\frac{T_2}{T_1} = \left|\frac{b}{a}\right| \left(\frac{y_2^0}{y_1^0}\right)^{1/2}.$$

Теперь допустим, что проверяемая гипотеза имеет вид

$$Q\theta = \mathbf{Z},$$

причем

$$Q = (1\ 1),$$

т. е.

$$\theta_1 + \theta_2 = z.$$

В таком случае $b/a = 1$. Предположим далее, что $y_2^0/y_1^0 = 4$. Тогда $T_2 = 2T_1$, т. е. на измерение большей величины затрачивается вдвое больше времени. Поскольку

$$\sigma^2(\hat{\theta}_1 + \hat{\theta}_2) = \sigma^2(\hat{\theta}_1) + \sigma^2(\hat{\theta}_2) = \frac{1}{w_1} + \frac{1}{w_2} = \frac{y_1^0}{T_1} + \frac{y_2^0}{T_2},$$

то для получения оптимального решения

$$T_1 = \frac{1}{3},\quad T_2 = \frac{2}{3} \quad \text{и} \quad y_1 = \frac{1}{4},\quad y_2 = 1.$$

Таким образом,

$$\sigma^2(\hat{\theta}_1 + \hat{\theta}_2) = 9.$$

Пусть теперь на самом деле относительные ошибки измеренных величин одинаковы. Это значит, что

$$y_1^0 T_1 = y_2^0 T_2$$

или в данном примере

$$T_1 = \frac{4}{5},\qquad T_2 = \frac{1}{5}$$

и

$$\sigma^2(\hat{\theta}_1 + \hat{\theta}_2) = 21{,}25.$$

Итак, добиваясь максимума ξ^2, мы получили минимальное значение $\sigma^2(\hat{\theta}_1 + \hat{\theta}_2)$, что всегда характерно для случая одномерной гипотезы. Отметим очевидный факт: если $a = 1$, $b = 0$, то все время следует затратить на измерение

величины y_1. Из приведенного примера видно, что для определения наилучших весов или интервалов времени измерения необходимо заранее знать скорость счета в каждом отдельном измерении. Если предложенную здесь процедуру применять к реальному эксперименту, то сначала следует на опыте грубо оценить значения измеряемых величин. Тогда данный метод позволит найти оптимальное распределение времени для проведения эксперимента с целью проверки конкретных гипотез.

Пример 2. Допустим, что нам приблизительно известна структура некоторого кристалла. Это значит, что приблизительно известны интенсивности дифракционного рассеяния рентгеновских лучей. Планируется новый эксперимент, в котором интенсивность дифракционного рассеяния будет измеряться с помощью счетчиков. Целью этого эксперимента является проверка равенства длин некоторых связей. Развитый в этом параграфе формализм мог бы дать указания на то, под какими углами следует производить измерения. Знание приблизительной структуры здесь необходимо не только для того, чтобы знать относительные величины $\mathbf{y}$, но и для выяснения применимости линейной аппроксимации $\mathbf{y}^0 = A\theta$, где A — известная конструкционная матрица. Хотя мы здесь не рассматриваем примеры многомерных задач, из следующего примера будет видно, как можно получить общие качественные результаты.

Пример 3. Пусть имеется серия наблюдаемых величин, описываемых выражением

$$y_i^0 = q_i \exp(-B\rho_i),$$

где B — искомый параметр. Если мы хотим проверить гипотезу $B = B_0$, то, используя линейную аппроксимацию (справедливую в окрестности B_0)

$$dy_i^0 = -\rho_i q_i \exp(-B\rho_i)\, dB = -\rho_i y_i^0\, dB,$$

можно показать, что

$$\xi^2 = (B - B_0)^2 \sum_i \rho_i^2 (y_i^0)^2 w_i =$$

$$= (B - B_0)^2 \sum_i \rho_i y_i t_i,$$

если справедливы сделанные выше предположения относительно времени измерения числа отсчетов. Параметр нецентральности достигает наибольшего значения, если все время затрачено на измерение той величины, для которой произведение $\rho_i^2 y_i^0$ максимально. Для специалистов по кристаллографии это обстоятельство не удивительно, так как им известно, что «температурный фактор» B наиболее чувствителен к рассеянию с большой интенсивностью при больших значениях $\rho_i = (\sin \alpha_i/\lambda)^2$, где α_i — угол рассеяния в i-м наблюдении. В этом примере использована конкретная зависимость, имеющая место при некоторых упрощающих предположениях, без учета возможных корреляций с остальными параметрами.

Как уже отмечалось, в проведенном выше обсуждении и примерах мы упускали из вида одно важное обстоятельство, полагая, что мощность критерия проверки гипотез увеличивается с ростом ξ^2. Это верно лишь в том случае, когда при оценке σ^2 число степеней свободы остается неизменным, а для большинства задач попытка максимизировать ξ^2 приводит к выводу, что на $n - m$ наблюдений вовсе не следует тратить времени. Однако с точки зрения практики число наблюдаемых величин должно превышать число искомых параметров, даже если σ^2 можно получить, вычисляя дисперсии наблюдаемых величин из распределения Пуассона. Независимая оценка σ^2 необходима, чтобы иметь возможность обнаружить неизвестные источники экспериментальных ошибок или несостоятельность теоретической модели. Поэтому число степеней свободы желательно поддерживать большим. Следовательно, при необходимости использовать нелинейное программирование вместо условия $T_i \geqslant 0$, вероятно, лучше пользоваться условием $T_i \geqslant T_0$, где T_0 — минимальный интервал времени, который надо затрачивать на каждое отдельное измерение.

Исследование в общем виде вопроса о связи ξ^2 с числом степеней свободы при определении мощности критерия чересчур громоздко, чтобы пользоваться на практике.

Более простой путь планирования эксперимента с *мнк*-анализом состоит в выборе весов, при котором $\det (A'WA)^{-1}$ минимален. Это не дает предпочтения какой-либо частной гипотезе, но сводит к минимуму объем гипер-

эллипсоида среднеквадратичных отклонений, что является вполне разумным шагом. Такая процедура непосредственно связана с максимизацией мощности критерия проверки гипотез, когда Q задано в виде единичной матрицы. Для подробного знакомства с приложением к нелинейным задачам мы отсылаем читателя к работе [4а] [1]).

[1]) См. также [63].— *Прим. ред.*

Статистические таблицы

Таблица I

Соотношение между размахом и среднеквадратичным отклонением для нормальных выборок [5, 32]

n	5	10	20	30	100
c	2,3	3,1	3,7	4,1	5,0

В таблице указаны значения c (см. дополнение I, раздел 3) для различных объемов выборки

Таблица II

Нормальное распределение

$$f(u_p) = \frac{1}{\sqrt{2\pi}}\, e^{-u_p^2/2}, \quad p = \sqrt{\frac{2}{\pi}} \int_{u_p}^{\infty} e^{-u^2/2} du$$

u_p	$f(u_p)$	p	u_p	$f(u_p)$	p
0,0	0,39894	1,00000	1,4	0,14973	0,16152
0,1	0,39695	0,92034	1,5	0,12952	0,13362
0,2	0,39104	0,84148	1,6	0,11092	0,10960
0,3	0,38139	0,76418	1,7	0,09405	0,08914
0,4	0,36827	0,68916	1,8	0,07895	0,07186
0,5	0,35207	0,61708	1,9	0,06562	0,05744
0,6	0,33322	0,54850	2,0	0,05399	0,04550
0,7	0,31225	0,48392	2,1	0,04398	0,03572
0,8	0,28969	0,42372	2,2	0,03547	0,02780
0,9	0,26609	0,36812	2,3	0,02833	0,02144
1,0	0,24197	0,31732	2,4	0,02239	0,01640
1,1	0,21785	0,27134	2,5	0,01753	0,01242
1,2	0,19419	0,23014	2,6	0,01358	0,00932
1,3	0,17137	0,19360	2,7	0,01042	0,00694

Продолжение табл. II

u_p	$f(u_p)$	p	u_p	$f(u_p)$	p
2,8	0,00792	0,00512	3,5	0,00087	0,00046
2,9	0,00595	0,00374	3,6	0,00061	0,00032
3,0	0,00443	0,00270	3,7	0,00042	0,00022
3,1	0,00327	0,00194	3,8	0,00029	0,00014
3,2	0,00238	0,00138	3,9	0,00020	0,00010
3,3	0,00172	0,00096	4,0	0,00013	0,00006
3,4	0,00123	0,00068			

Таблица III

Нормальное распределение (значения u_p при различных p)

p	1,00	0,95	0,90	0,85	0,80	0,75	0,70	0,65	0,60
u_p	0,00	0,0627	0,1257	0,1891	0,2533	0,3186	0,3853	0,4538	0,5244

p	0,55	0,50	0,45	0,40	0,35	0,30	0,25	0,20	0,15
u_p	0,5978	0,6745	0,7554	0,8416	0,9346	1,0364	1,1503	1,2816	1,4395

p	0,10	0,05	0,01	0,001	0,0001
u_p	1,6449	1,9600	2,5758	3,2905	3,8906

Таблица IV

Распределение χ^2

Значения χ^2_p и вероятности p того, что $\chi^2 \geqslant \chi^2_p$, при числе степеней свободы ν [13]

ν	p=0,99	0,98	0,95	0,90	0,80	0,70
1	0,000157	0,000628	0,00393	0,0158	0,0642	0,148
2	0,0201	0,0404	0,103	0,211	0,446	0,713
3	0,115	0,185	0,352	0,584	1,005	1,424
4	0,297	0,429	0,711	1,064	1,649	2,195
5	0,554	0;752	1,145	1,610	2,343	3,000
6	0,872	1,134	1,635	2,204	3,070	3,828
7	1,239	1,564	2,167	2,833	3,822	4,671
8	1,646	2,032	2,733	3,490	4,594	5,527
9	2,088	2,532	3,325	4,168	5,380	6,393
10	2,558	3,059	3,940	4,865	6,179	7,267
11	3,053	3,609	4,575	5,578	6,989	8,148
12	3,571	4,178	5,226	6,304	7,807	9,034
13	4,107	4,765	5,892	7,042	8,634	9,926
14	4,660	5,368	6,571	7,790	9,467	10,821
15	5,229	5,985	7,261	8,547	10,307	11,721
16	5,812	6,614	7,962	9,312	11,152	12,624
17	6,408	7,255	8,672	10,085	12,002	13,531
18	7,015	7,906	9,390	10,865	12,857	14,440
19	7,633	8,567	10,117	11,651	13,716	15,352
20	8,260	9,237	10,851	12,444	14,578	16,266
21	8,897	9,915	11,591	13,240	15,445	17,182
22	9,542	10,600	12,338	14,041	16,314	18,101
23	10,196	11,293	13,091	14,848	17,187	19,021
24	10,856	11,992	13,848	15,659	18,062	19,943
25	11,524	12,697	14,611	16,473	18,940	20,867
26	12,198	13,409	15,379	17,292	19,820	21,792
27	12,879	14,125	16,151	18,114	20,703	22,710
28	13,565	14,847	16,928	18,939	21,588	23,647
29	14,256	15,574	17,708	19,768	22,475	24,577
30	14,953	16,306	18,493	20,599	23,364	25,508

Продолжение табл. IV

ν	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	2,366	3,665	4,642	6,251	7,815	9,937	11,345
4	3,357	4,878	5,980	7,779	9,488	11,668	13,277
5	4,351	6,064	7,289	9,236	11,070	13,388	15,086
6	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	6,346	8,383	9,803	12,017	14,067	16,622	18,475
8	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	9,342	11,781	13,442	15,987	18,307	21,161	23,209
11	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	12,340	15,119	16,985	19,812	22,362	25,472	27,688
14	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	14,339	17,322	19,311	22,307	24,996	28,259	30,578
16	15,338	18,418	20,465	23,542	26,296	29,633	32,000
17	16,338	19,511	21,615	24,769	27,587	30,995	33,409
18	17,338	20,601	22,760	25,989	28,869	32,346	34,805
19	18,338	21,689	23,900	27,204	30,114	33,687	36,191
20	19,337	22,775	25,038	28,412	31,410	35,020	37,566
21	20,337	23,858	26,171	29,615	32,671	36,343	38,932
22	21,337	24,939	27,301	30,813	33,924	37,659	40,289
23	22,337	26,018	28,429	32,007	35,172	38,968	41,638
24	23,337	27,096	29,553	33,196	36,415	40,270	42,980
25	24,337	28,172	30,675	34,382	37,652	41,566	44,314
26	25,336	29,246	31,795	35,563	38,885	42,856	45,642
27	26,336	30,319	32,912	36,741	40,113	44,140	46,963
28	27,336	31,391	34,027	37,916	41,337	45,419	48,278
29	28,336	32,461	35,139	39,087	42,557	46,693	49,588
30	29,336	33,530	36,250	40,256	43,773	47,962	50,892

Таблица V

t-Распределение Стьюдента

Значения t_p и вероятности p того, что t отличается от нулевого среднего в любую сторону более чем на t_p, при числе степеней свободы ν [13]

ν	$p=0,9$	0,8	0,7	0,6	0,5	0,4
1	0,158	0,325	0,510	0,727	1,000	1,376
2	0,142	0,289	0,445	0,617	0,816	1,061
3	0,137	0,277	0,424	0,584	0,765	0,978
4	0,134	0,271	0,414	0,569	0,741	0,941
5	0,132	0,267	0,408	0,559	0,727	0,920
6	0,131	0,265	0,404	0,553	0,718	0,906
7	0,130	0,263	0,402	0,549	0,711	0,896
8	0,130	0,262	0,399	0,546	0,706	0,889
9	0,129	0,261	0,398	0,543	0,703	0,883
10	0,129	0,260	0,397	0,542	0,700	0,879
11	0,129	0,260	0,396	0,540	0,697	0,876
12	0,128	0,259	0,395	0,539	0,695	0,873
13	0,128	0,259	0,394	0,538	0,694	0,870
14	0,128	0,258	0,393	0,537	0,692	0,868
15	0,128	0,258	0,393	0,536	0,691	0,866
16	0,128	0,258	0,392	0,535	0,690	0,865
17	0,128	0,257	0,392	0,534	0,689	0,863
18	0,127	0,257	0,392	0,534	0,688	0,862
19	0,127	0,257	0,391	0,533	0,688	0,861
20	0,127	0,257	0,391	0,533	0,687	0,860
21	0,127	0,257	0,391	0,532	0,686	0,859
22	0,127	0,256	0,390	0,532	0,686	0,858
23	0,127	0,256	0,390	0,532	0,685	0,858
24	0,127	0,256	0,390	0,531	0,685	0,857
25	0,127	0,256	0,390	0,531	0,684	0,856
26	0,127	0,256	0,390	0,531	0,684	0,856
27	0,127	0,256	0,389	0,531	0,684	0,855
28	0,127	0,256	0,389	0,530	0,683	0,855
29	0,127	0,256	0,389	0,530	0,683	0,854
30	0,127	0,256	0,389	0,530	0,683	0,854
∞	0,12566	0,25335	0,38532	0,52440	0,67449	0,84162

Продолжение табл. V

ν	0,3	0,2	0,1	0,05	0,02	0,01
1	1,963	3,078	6,314	12,706	31,821	63,657
2	1,386	1,886	2,920	4,303	6,965	9,925
3	1,250	1,638	2,353	3,182	4,541	5,841
4	1,190	1,533	2,132	2,776	3,747	4,604
5	1,156	1,476	2,015	2,571	3,365	4,032
6	1,134	1,440	1,943	2,447	3,143	3,707
7	1,119	1,415	1,895	2,365	2,998	3,499
8	1,108	1,397	1,860	2,306	2,896	3,355
9	1,100	1,383	1,836	2,262	2,821	3,250
10	1,093	1,372	1,812	2,228	2,764	3,169
11	1,088	1,363	1,796	2,201	2,718	3,106
12	1,083	1,356	1,782	2,179	2,681	3,055
13	1,079	1,350	1,771	2,160	2,650	3,012
14	1,076	1,345	1,761	2,145	2,624	2,977
15	1,074	1,341	1,753	2,131	2,602	2,947
16	1,071	1,337	1,746	2,120	2,583	2,921
17	1,069	1,333	1,740	2,110	2,567	2,898
18	1,067	1,330	1,734	2,101	2,552	2,878
19	1,066	1,328	1,729	2,093	2,539	2,861
20	1,064	1,325	1,725	2,086	2,528	2,845
21	1,063	1,323	1,721	2,080	2,518	2,831
22	1,061	1,321	1,717	2,074	2,508	2,819
23	1,060	1,319	1,714	2,069	2,500	2,807
24	1,059	1,318	1,711	2,064	2,492	2,797
25	1,058	1,316	1,708	2,060	2,485	2,787
26	1,058	1,315	1,706	2,056	2,479	2,779
27	1,057	1,314	1,703	2,052	2,473	2,771
28	1,056	1,313	1,701	2,048	2,467	2,763
29	1,055	1,311	1,699	2,045	2,462	2,756
30	1,055	1,310	1,697	2,042	2,457	2,750
∞	1,03643	1,28155	1,64485	1,95996	2,32634	2,57582

Таблица VI

Биномиальное распределение [29]

В таблице приведены значения вероятности $\sum_{k=r}^{n} C_k^n p^k q^{n-k}$ наблюдать число «благоприятных» исходов, равное или большее r, в n независимых испытаниях, если вероятность благоприятного исхода в одном испытании равна p

p	$n=2,\ r=2$	$n=2,\ r=1$	$n=3,\ r=3$	$n=3,\ r=2$	$n=3,\ r=1$	$n=4,\ r=4$	p
0,01	0,0001000	0,0199000	0,0000010	0,0002980	0,0297010	0,0000000	0,01
0,02	0,0004000	0,0396000	0,0000080	0,0011840	0,0588080	0,0000002	0,02
0,03	0,0009000	0,0591000	0,0000270	0,0026460	0,0873270	0,0000008	0,03
0,04	0,0016000	0,0784000	0,0000640	0,0046720	0,1152640	0,0000026	0,04
0,05	0,0025000	0,0975000	0,0001250	0,0072500	0,1426250	0,0000062	0,05
0,06	0,0036000	0,1164000	0,0002160	0,0103680	0,1694160	0,0000130	0,06
0,07	0,0049000	0,1351000	0,0003430	0,0140140	0,1956430	0,0000240	0,07
0,08	0,0064000	0,1536000	0,0005120	0,0181760	0,2213120	0,0000410	0,08
0,09	0,0081000	0,1719000	0,0007290	0,0228420	0,2464290	0,0000656	0,09
0,10	0,0100000	0,1900000	0,0010000	0,0280000	0,2710000	0,0001000	0,10
0,11	0,0121000	0,2079000	0,0013310	0,0336380	0,2950310	0,0001464	0,11
0,12	0,0144000	0,2256000	0,0017280	0,0397440	0,3185280	0,0002074	0,12
0,13	0,0169000	0,2431000	0,0021970	0,0463060	0,3414970	0,0002856	0,13
0,14	0,0196000	0,2604000	0,0027440	0,0533120	0,3639440	0,0003842	0,14
0,15	0,0225000	0,2775000	0,0033750	0,0607500	0,3858750	0,0005062	0,15

0,16	0,0256000	0,2944000	0,0040960	0,0686080	0,4072960	0,0006554	0,16
0,17	0,0289000	0,3111000	0,0049130	0,0768740	0,4282130	0,0008352	0,17
0,18	0,0324000	0,3276000	0,0058320	0,0855360	0,4486320	0,0010498	0,18
0,19	0,0361000	0,3439000	0,0068590	0,0945820	0,4685590	0,0013032	0,19
0,20	0,0400000	0,3600000	0,0080000	0,1040000	0,4880000	0,0016000	0,20
0,21	0,0441000	0,3759000	0,0092610	0,1137780	0,5069610	0,0019448	0,21
0,22	0,0484000	0,3916000	0,0106480	0,1239040	0,5254480	0,0023426	0,22
0,23	0,0529000	0,4071000	0,0121670	0,1343660	0,5434670	0,0027984	0,23
0,24	0,0576000	0,4224000	0,0138240	0,1451520	0,5610240	0,0033178	0,24
0,25	0,0625000	0,4375000	0,0156250	0,1562500	0,5781250	0,0039062	0,25
0,26	0,0676000	0,4524000	0,0175760	0,1676480	0,5947760	0,0045698	0,26
0,27	0,0729000	0,4671000	0,0196830	0,1793340	0,6109830	0,0053144	0,27
0,28	0,0784000	0,4816000	0,0219520	0,1912960	0,6267520	0,0061466	0,28
0,29	0,0841000	0,4959000	0,0243890	0,2035220	0,6420890	0,0070728	0,29
0,30	0,0900000	0 5100000	0,0270000	0,2160000	0,6570000	0,0081000	0,30
0,31	0,0961000	0,5239000	0,0297910	0,2287180	0,6714910	0,0092352	0,31
0,32	0,1024000	0,5376000	0,0327680	0,2416640	0,6855680	0,0104858	0,32
0,33	0,1089000	0,5511000	0,0359370	0,2548260	0,6992370	0,0118592	0,33
0,34	0,1156000	0,5644000	0,0393040	0,2681920	0,7125040	0,0133634	0,34
0,35	0,1225000	0,5775000	0,0428750	0,2817500	0,7253750	0,0150063	0,35
0,36	0,1296000	0,5904000	0,0466560	0,2954880	0,7378560	0,0167962	0,36
0,37	0,1369000	0,6031000	0,0506530	0,3093940	0,7499530	0,0187416	0,37
0,38	0,1444000	0,6156000	0,0548720	0,3234560	0,7616720	0,0208514	0,38
0,39	0,1521000	0,6279000	0,0593190	0,3376620	0,7730190	0,0231344	0,39
0,40	0,1600000	0,6400000	0,0640000	0,3520000	0,7840000	0,0256000	0,40

Продолжение табл. VI

p	$n=2,\ r=2$	$n=2,\ r=1$	$n=3,\ r=3$	$n=3,\ r=2$	$n=3,\ r=1$	$n=4,\ r=4$	p
0,41	0,1681000	0,6519000	0,0689210	0,3664580	0,7946210	0,0282576	0,41
0,42	0,1764000	0,6636000	0,0740880	0,3810240	0,8048880	0,0311170	0,42
0,43	0,1849000	0,6751000	0,0795070	0,3956860	0,8148070	0,0341880	0,43
0,44	0,1936000	0,6864000	0,0851840	0,4104320	0,8243840	0,0374810	0,44
0,45	0,2025000	0,6975000	0,0911250	0,4252500	0,8336250	0,0410062	0 45
0,46	0,2116000	0,7084000	0,0973360	0,4401280	0,8425360	0,0447746	0,46
0,47	0,2209000	0,7191000	0,1038230	0,4550540	0,8511230	0,0487968	0,47
0,48	0,2304000	0,7296000	0,1105920	0,4700160	0,8593920	0,0530842	0,48
0,49	0,2401000	0,7399000	0,1176490	0,4850020	0,8673490	0,0576480	0,49
0,50	0,2500000	0,7500000	0,1250000	0,5000000	0,8750000	0,0625000	0,50

p	$n=4,\ r=3$	$n=4,\ r=2$	$n=4,\ r=1$	$n=5,\ r=5$	$n=5,\ r=4$	$n=5,\ r=3$	p
0,01	0,0000040	0,0005920	0,0394040	0,0000000	0,0000000	0,0000099	0,01
0,02	0,0000315	0,0023365	0,0776318	0,0000000	0,0000008	0,0000776	0,02
0,03	0,0001056	0,0051864	0,1147072	0,0000000	0,0000040	0,0002580	0,03
0,04	0,0002483	0,0090957	0,1506534	0,0000001	0,0000124	0,0006022	0,04
0,05	0,0004812	0,0140188	0,1854938	0,0000003	0,0000300	0,0011581	0,05
0,06	0,0008251	0,0199109	0,2192510	0,0000008	0,0000617	0,0019703	0,06
0,07	0,0013000	0,0267280	0,2519480	0,0000017	0,0001133	0,0030799	0,07

0,08	0,0019251	0,0344269	0,2836070	0,0000033	0,0001917	0,0045253	0,08
0,09	0,0027192	0,0429648	0,3142504	0,0000059	0,0003044	0,0063413	0,09
0,10	0,0037000	0,0523000	0,3439000	0,0000100	0,0004600	0,0085600	0,10
0,11	0,0048848	0,0623912	0,3725776	0,0000161	0,0006676	0,0112105—	0,11
0,12	0,0062899	0,0731981	0,4003046	0,0000249	0,0009373	0,0143189	0,12
0,13	0,0079312	0,0846808	0,4271024	0,0000371	0,0012795	0,0179086	0,13
0,14	0,0098235	0,0968005—	0,4529918	0,0000538	0,0017057	0,0220003	0,14
0,15	0,0119812	0,1095188	0,4779938	0,0000759	0,0022275	0,0266119	0,15
0,16	0,0144179	0,1227981	0,5021286	0,0001049	0,0028574	0,0317587	0,16
0,17	0,0171464	0,1366016	0,5254168	0,0001420	0,0036081	0,0374538	0,17
0,18	0,0201787	0,1508933	0,5478782	0,0001890	0,0044930	0,0437073	0,18
0,19	0,0235264	0,1656376	0,5695328	0,0002476	0,0055256	0,0505275	0,19
0,20	0,0272000	0,1808000	0,5904000	0,0003200	0,0067200	0,0579200	0,20
0,21	0,0312096	0,1963464	0,6104992	0,0004084	0,0080904	0,0658883	0,21
0,22	0,0355643	0,2122437	0,6298494	0,0005154	0,0096513	0,0744338	0,22
0,23	0,04027228	0,2284592	0,6484696	0,0006436	0,0114175	0,0835557	0,23
0,24	0,0453427	0,2449613	0,6663782	0,0007963	0,0134038	0,0932512	0,24
0,25	0,0507812	0,2617188	0,6835938	0,0009766	0,0156250	0,1035156	0,25
0,26	0,0565947	0,2787013	0,7001342	0,0011881	0,0180962	0,1143424	0,26
0,27	0,0627888	0,2958792	0,7160176	0,0014349	0,0208325—	0,1257232	0,27
0,28	0,0693683	0,3132237	0,7312614	0,0017210	0,0238487	0,1376478	0,28
0,29	0,0763376	0,3307064	0,7458832	0,0020511	0,0271596	0,1501045	0,29
0,30	0,0837000	0,3483000	0,7599000	0,0024300	0,0307800	0,1630800	0,30
0,31	0,0914584	0,3659776	0,7733288	0,0028629	0,0347244	0,1765593	0,31
0,32	0,0996147	0,3837133	0,7861862	0,0033554	0,0390070	0,1905263	0,32

Продолжение табл. VI

p	$n=4,\ r=3$	$n=4,\ r=2$	$n=4,\ r=1$	$n=5,\ r=5$	$n=5,\ r=4$	$n=5,\ r=3$	p
0,33	0,1081704	0,4014816	0,7984888	0,0039135	0,0436419	0,2049631	0,33
0,34	0,1171259	0,4192581	0,8102526	0,0045435	0,0486426	0,2198509	0,34
0,35	0,1264812	0,4370188	0,8214938	0,0052522	0,0540225	0,2351694	0,35
0,36	0,1362355	0,4547405—	0,8322278	0,0060466	0,0597943	0,2508973	0,36
0,37	0,1463872	0,4724008	0,8424704	0,0069344	0,0659705—	0,2670122	0,37
0,38	0,1569339	0,4899781	0,8522366	0,0079235	0,0725627	0,2834907	0,38
0,39	0,1678728	0,5074512	0,8615416	0,0090224	0,0795824	0,3003084	0,39
0,40	0,1792000	0,5248000	0,8704000	0,0102400	0,0870400	0,3174400	0,40
0,41	0,1909112	0,5420048	0,8788264	0,0115856	0,0949456	0,3348596	0,41
0,42	0,2030011	0,5590469	0 8868350	0,0130691	0,1033083	0,3525403	0,42
0,43	0,2154640	0,5759080	0,8944400	0,0147008	0,1121367	0,3704549	0,43
0,44	0,2282931	0,5925709	0,9016550	0,0164916	0,1214383	0,3885753	0,44
0,45	0,2414812	0,6090188	0,9084937	0,0184528	0,1312200	0,4068731	0,45
0,46	0,2550203	0,6252357	0,9149694	0,0205963	0,1414876	0,4253194	0,46
0,47	0,2689016	0,6412064	0,9210952	0,0229345	0,1522460	0,4438849	0,47
0,48	0,2831155	0,6569165—	0,9268838	0,0254804	0,1634992	0,4625400	0,48
0,49	0,2976520	0,6723520	0,9323480	0,0282475	0,1752500—	0,4812550—	0,49
0,50	0,3125000	0,6875000	0,9375000	0,0312500	0,1875000	0,5000000	0,50

p	$n=5,\ r=2$	$n=5,\ r=1$	$n=6,\ r=6$	$n=6,\ r=5$	$n=6,\ r=4$	$n=6,\ r=3$	p
0,01	0,0009801	0,0490100	0,0000000	0,0000000	0,0000001	0,0000196	0,01

0,02	0,0038424	0,0960792	0,0000000	0,0000000	0,0000023	0,0001529	0,02
0,03	0,0084721	0,1412660	0,0000000	0,0000001	0,0000116	0,0005044	0,03
0,04	0,0147580	0,1846273	0,0000000	0,0000006	0,0000360	0,0011684	0,04
0,05	0,0225925	0,2261191	0,0000000	0,0000018	0,0000864	0,0022298	0,05
0,06	0,0318713	0,2660960	0,0000000	0,0000044	0,0001762	0,0037643	0,06
0,07	0,0424934	0,3043116	0,0000001	0,0000095—	0,0003210	0,0058389	0,07
0,08	0,0543613	0,3409185—	0,0000003	0,0000184	0,0005384	0,0085121	0,08
0,09	0,0673805	0,3759679	0,0000005	0,0000328	0,0008477	0,0118348	0,09
0,10	0,0814600	0,4095100	0,0000010	0,0000550	0,0012700	0,0158500	0,10
0,11	0,0965117	0,4415941	0,0000018	0,0000878	0,0018273	0,0205936	0,11
0,12	0,1124509	0,4722681	0,0000030	0,0001344	0,0025431	0,0260947	0,12
0,13	0,1291956	0,5015791	0,0000048	0,0001986	0,0034413	0,0323759	0,13
0,14	0,1466673	0,5295730	0,0000075	0,0002850	0,0045469	0,0394537	0,14
0,15	0,1647900	0,5562947	0,0000114	0,0003987	0,0058852	0,0473386	0,15
0,16	0,1834910	0,5817881	0,0000168	0,0005453	0,0074816	0,0560359	0,16
0,17	0,2027002	0,6060959	0,0000241	0,0007312	0,0093619	0,0655457	0,17
0,18	0,2223506	0,6292602	0,0000340	0,0009637	0,0115516	0,0758631	0,18
0,19	0,2423777	0,6513216	0,0000470	0,0012504	0,0140760	0,0869790	0,19
0,20	0,2627200	0,6723200	0,0000640	0,0016000	0,0169600	0,0988800	0,20
0,21	0,2833185	0,6922944	0,0000858	0,0020216	0,0202280	0,1115487	0,21
0,22	0,3041169	0,7112826	0,0001134	0,0025253	0,0239035—	0,1249641	0,22
0,23	0,3250616	0,7293216	0,0001480	0,0031216	0,0280093	0,1391020	0,23
0,24	0,3461014	0,7464475—	0,0001911	0,0038221	0,0325671	0,1539352	0,24
0,25	0,3671875	0,7626953	0,0002441	0,0046387	0,0375977	0,1694336	0,25
0,26	0,3882738	0,7780993	0,0003089	0,0055842	0,0431203	0,1855646	0,26

Продолжение табл. VI

p	$n=5,\ r=2$	$n=5,\ r=1$	$n=6,\ r=6$	$n=6,\ r=5$	$n=6,\ r=4$	$n=6,\ r=3$	p
0,27	0,4093166	0,7926928	0,0003874	0,0066722	0,0491530	0,2022934	0,27
0,28	0,4302743	0,8065082	0,0004819	0,0079168	0,0557124	0,2195832	0,28
0,29	0,4511077	0,8195771	0,0005948	0,0093326	0,0628136	0,2373955—	0,29
0,30	0,4717800	0,8319300	0,0007290	0,0109350	0,0704700	0,2556900	0,30
0,31	0,4922565—	0,8435969	0,0008875	0,0127400	0,0786932	0,2744255—	0,31
0,32	0,5125046	0,8546066	0,0010737	0,0147640	0,0874932	0,2935593	0,32
0,33	0,5324940	0,8649875—	0,0012915—	0,0170239	0,0968779	0,3130483	0,33
0,34	0,5521962	0,8747667	0,0015448	0,0195372	0,1068534	0,3328483	0,34
0,35	0,5715850	0,8839709	0,0018383	0,0223218	0,1174239	0,3529148	0,35
0,36	0,5906359	0,8926258	0,0021768	0,0253958	0,1285914	0,3732032	0,36
0,37	0,6093266	0,9007563	0,0025657	0,0287777	0,1403559	0,3936685	0,37
0,38	0,6276363	0,9083867	0,0030109	0,0324864	0,1527154	0,4142660	0,38
0,39	0,6455465—	0,9155404	0,0035187	0,0365408	0,1656655	0,4349512	0,39
0,40	0,6630400	0,9222400	0,0040960	0,0409600	0,1792000	0,4556800	0,40
0,41	0,6801017	0,9285076	0,0047501	0,0457632	0,1933103	0,4764088	0,41
0,42	0,6967179	0,9343643	0,0054890	0,0509696	0,2079858	0,4970949	0,42
0,43	0,7128768	0,9398308	0,0063214	0,0565983	0,2232135	0,5176963	0,43
0,44	0,7285679	0,9449268	0,0072563	0,0626682	0,2389786	0,5381721	0,44
0,45	0,7437825	0,9496716	0,0083038	0,0691980	0,2552639	0,5584823	0,45
0,46	0,7585132	0,9540835—	0,0094743	0,0762063	0,2720502	0,5785885	0,46
0,47	0,7727541	0,9581805—	0,0107792	0,0837109	0,2893163	0,5984534	0,47

0,48	0,7865008	0,9619796	0,0122306	0,0917294	0,3070388	0,6180412	0,48
0,49	0,7997501	0,9654975—	0,0138413	0,1002787	0,3251924	0,6373176	0,49
0,50	0,8125000	0,9687500	0,0156250	0,1093750	0,3437500	0,6562500	0,50

p	$n=6,\ r=2$	$n=6,\ r=1$	$n=7,\ r=7$	$n=7,\ r=6$	$n=7,\ r=5$	$n=7,\ r=4$	p
0,01	0,0014604	0,0585199	0,0000000	0,0000000	0,0000000	0,0000003	0,01
0,02	0,0056871	0,1141576	0,0000000	0,0000000	0,0000001	0,0000053	0,02
0,03	0,0124559	0,1670280	0,0000000	0,0000000	0,0000005—	0,0000264	0,03
0,04	0,0215528	0,2172422	0,0000000	0,0000000	0,0000020	0,0000813	0.04
0,05	0,0327738	0,2649081	0,0000000	0,0000001	0,0000060	0,0001936	0,05
0,06	0,0459248	0,3101302	0,0000000	0,0000003	0,0000147	0,0003915—	0,06
0,07	0,0608207	0,3530098	0,0000000	0,0000008	0,0000313	0,0007072	0,07
0,08	0,0772859	0,3936450—	0,0000000	0,0000017	0,0000600	0,0011763	0,08
0,09	0,0951534	0,4321307	0,0000000	0,0000034	0,0001061	0,0018366	0,09
0,10	0,1142650	0,4685590	0,0000001	0,0000064	0,0001765	0,0027280	0,10
0,11	0,1344708	0,5030187	0,0000002	0,0000112	0,0002791	0,0038916	0,11
0,12	0,1556289	0,5355959	0,0000004	0,0000188	0,0004234	0,0053693	0,12
0,13	0,1776055—	0,5663738	0,0000006	0,0000300	0,0006202	0,0072028	0,13
0,14	0,2002741	0,5954328	0,0000011	0,0000464	0,0008817	0,0094339	0,14
0,15	0,2235157	0,6228505—	0,0000017	0,0000695	0,0012216	0,0121032	0,15
0,16	0,2472185	0,6487020	0,0000027	0,0001013	0,0016551	0,0152503	0,16
0,17	0,2712775—	0,6730596	0,0000041	0,0001443	0,0021984	0,0189131	0,17
0,18	0,2955943	0,6959933	0,0000061	0,0002014	0,0028695—	0,0231276	0,18

p	$n=6,\ r=2$	$n=6,\ r=1$	$n=7,\ r=7$	$n=7,\ r=6$	$n=7,\ r=5$	$n=7,\ r=4$	p
0,19	0,3200770	0,7175705—	0,0000089	0,0002757	0,0036873	0,0279276	0,19
0,20	0,3446400	0,7378560	0,0000128	0,0003712	0,0046720	0,0333440	0,20
0,21	0,3692034	0,7569125	0,0000180	0,0004923	0,0058450—	0,0394053	0,21
0,22	0,3936934	0,7748004	0,0000249	0,0006440	0,0072285—	0,0461368	0,22
0,23	0,4180414	0,7915776	0,0000340	0,0008320	0,0088458	0,0535606	0,23
0,24	0,4421844	0,8073001	0,0000459	0,0010625	0,0107209	0,0616955—	0,24
0,25	0,4660645—	0,8220215—	0,0000610	0,0013428	0,0128784	0,0705566	0,25
0,26	0,4896285—	0,8357935	0,0000803	0,0016805	0,0153436	0,0801558	0,26
0,27	0,5128282	0,8486658	0,0001046	0,0020843	0,0181420	0,0905009	0,27
0,28	0,5356198	0,8606859	0,0001349	0,0025637	0,0212996	0,1015962	0,28
0,29	0,5579638	0,8718997	0,0001725	0,0031288	0,0248421	0,1134424	0,29
0,30	0,5798250	0,8823510	0,0002187	0,0037908	0,0287955—	0,1260360	0,30
0,31	0,6011720	0,8920818	0,0002751	0,0045618	0,0331855	0,1393702	0,31
0,32	0,6219773	0,9011325	0,0003436	0,0054546	0,0380373	0,1534344	0,32
0,33	0,6422168	0,9095416	0,0004262	0,0064832	0,0433757	0,1682141	0,33
0,34	0,6618702	0,9173460	0,0005252	0,0076622	0,0492247	0,1836917	0,34
0,35	0,6809201	0,9245811	0,0006434	0,0090075	0,0556075	0,1998457	0,35
0,36	0,6993523	0,9312805	0,0007836	0,0105356	0,0625462	0,2166517	0,36
0,37	0,7171556	0,9374765	0,0009493	0,0122642	0,0700617	0,2340816	0,37
0,38	0,7343215	0,9431998	0,0011442	0,0142116	0,0781734	0,2521046	0,38
0,39	0,7508441	0,9484796	0,0013723	0,0163973	0,0868994	0,2706869	0,39
0,40	0,7667200	0,9533440	0,0016384	0,0188416	0,0962560	0,2897920	0,40

0,41	0,7819481	0,9578195—	0,0019475	0,0215655	0,1062575—	0,3093807	0,41
0,42	0,7965294	0,9619313	0,0023054	0,0245909	0,1169164	0,3294116	0,42
0,43	0,8104670	0,9657036	0,0027182	0,0279404	0,1282428	0,3498411	0,43
0,44	0,8237658	0,9691590	0,0031928	0,0316375	0,1402448	0,3706237	0,44
0,45	0,8364326	0,9723194	0,0037367	0,0357062	0,1529277	0,3917122	0,45
0,46	0,8484755	0,9752051	0,0043582	0,0401710	0,1662945	0,4130579	0,46
0,47	0,8599045	0,9778356	0,0050662	0,0450571	0,1803454	0,4346107	0,47
0,48	0,8707306	0,9802294	0,0058707	0,0503900	0,1950779	0,4563199	0,48
0,49	0,8809663	0,9824037	0,0067822	0,0561956	0,2104864	0,4781337	0,49
0,50	0,8906250	0,9843750	0,0078125	0,0625000	0,2265625	0,5000000	0,50

p	$n=7,\ r=3$	$n=7,\ r=2$	$n=7,\ r=1$	$n=8,\ r=8$	$n=8,\ r=7$	$n=8,\ r=6$	p
0,01	0,0000340	0,0020310	0,0679347	0,0000000	0,0000000	0,0000000	0,01
0,02	0,0002636	0,0078565	0,1318745—	0,0000000	0,0000000	0,0000000	0,02
0,03	0,0008630	0,0170930	0,1920172	0,0000000	0,0000000	0,0000000	0,03
0,04	0,0019838	0,0293803	0,2485525	0,0000000	0,0000000	0,0000001	0,04
0,05	0,0037570	0,0443805	0,3016627	0,0000000	0,0000000	0,0000004	0,05
0,06	0,0062940	0,0617771	0,3515224	0,0000000	0,0000000	0,0000012	0,06
0,07	0,0096876	0,0812739	0,3982991	0,0000000	0,0000001	0,0000029	0,07
0,08	0,0140140	0,1025946	0,4421534	0,0000000	0,0000002	0,0000064	0,08
0,09	0,0193335—	0,1254814	0,4832390	0,0000000	0,0000004	0,0000127	0,09
0,10	0,0256915	0,1496944	0,5217031	0,0000000	0,0000007	0,0000234	0,10
0,11	0,0331201	0,1750111	0,5576867	0,0000000	0,0000014	0,0000407	0,11

p	$n=7,\ r=3$	$n=7,\ r=2$	$n=7,\ r=1$	$n=8,\ r=8$	$n=8,\ r=7$	$n=8,\ r=6$	p
0,12	0,0416388	0,2012250—	0,5913244	0,0000000	0,0000026	0,0000673	0,12
0,13	0,0512558	0,2281454	0,6227452	0,0000001	0,0000044	0,0001067	0,13
0,14	0,0619685	0,2555963	0,6520722	0,0000001	0,0000074	0,0001633	0,14
0,15	0,0737652	0,2834159	0,6794229	0,0000003	0,0000119	0,0002423	0,15
0,16	0,0866251	0,3114559	0,7049097	0,0000004	0,0000185—	0,0003499	0,16
0,17	0,1005201	0,3395804	0,7286395—	0,0000007	0,0000279	0,0004891	0,17
0,18	0,1154147	0,3676661	0,7507145	0,0000011	0,0000413	0,0006816	0,18
0,19	0,1312677	0,3956008	0,7712321	0,0000017	0,0000596	0,0009239	0,19
0,20	0,1480320	0,4232832	0,7902848	0,0000026	0,0000845—	0,0012314	0,20
0,21	0,1656562	0,4506224	0,8079609	0,0000038	0,0001176	0,0016164	0,21
0,22	0,1840845	0,4775369	0,8243443	0,0000055—	0,0001611	0,0020926	0,22
0,23	0,2032581	0,5039547	0,8395148	0,0000078	0,0002176	0,0026751	0,23
0,24	0,2231150	0,5298122	0,8535481	0,0000110	0,0002899	0,0033805	0,24
0,25	0,2435913	0,5550537	0,8665161	0,0000153	0,0003815—	0,0042267	0,25
0,26	0,2646212	0,5796314	0,8784872	0,0000209	0,0004964	0,0052329	0,26
0,27	0,2861378	0,6035043	0,8895260	0,0000282	0,0006391	0,0064109	0,27
0,28	0,3080735—	0,6266383	0,8996939	0,0000378	0,0008150—	0,0078097	0,28
0,29	0,3303603	0,6490052	0,9090488	0,0000500	0,0010298	0,0094256	0,29
0,30	0,3529305	0,6705828	0,9176457	0,0000656	0,0012903	0,0112922	0,30
0,31	0,3757169	0,6913541	0,9255365—	0,0000853	0,0016040	0,0134351	0,31
0,32	0,3986531	0,7113070	0,9327701	0,0001100	0,0019791	0,0158811	0,32
0,33	0,4216739	0,7304340	0,9393929	0,0001406	0,0024250—	0,0186577	0,33

0,34	0,4447157	0,7487320	0,9454484	0,0001786	0,0029518	0,0217935	0,34
0,35	0,4677167	0,7662014	0,9509777	0,0002252	0,0035708	0,0253175	0,35
0,36	0,4906169	0,7828465—	0,9560195	0,0002821	0,0042944	0,0292594	0,36
0,37	0,5133587	0,7986743	0,9606102	0,0003512	0,0051358	0,0336492	0,37
0,38	0,5358871	0,8136952	0,9647839	0,0004348	0,0061098	0,0385171	0,38
0,39	0,5581494	0,8279219	0,9685726	0,0005352	0,0072321	0,0438932	0,39
0,40	0,5800960	0,8413696	0,9720064	0,0006554	0,0085197	0,0498074	0,40
0,41	0,6016799	0,8540554	0,9751135—	0,0007985—	0,0099909	0,0562892	0,41
0,42	0,6228574	0,8659982	0,9779202	0,0009683	0,0116653	0,0633676	0,42
0,43	0,6435877	0,8772187	0,9804510	0,0011688	0,0135637	0,0710705—	0,43
0,44	0,6638333	0,8877388	0,9827291	0,0014048	0,0157085—	0,0794247	0,44
0,45	0,6835599	0,8975816	0,9847756	0,0016815	0,0181230	0,0884559	0,45
0,46	0,7027366	0,9067711	0,9866107	0,0020048	0,0208321	0,0981878	0,46
0,47	0,7213354	0,9153321	0,9882529	0,0023811	0,0238619	0,1086426	0,47
0,48	0,7393321	0,9232900	0,9897193	0,0028179	0,0272400	0,1918402	0,48
0,49	0,7567054	0,9306706	0,9910259	0,0033233	0,0309948	0,1317981	0,49
0,50	0,7934375	0,9375000	0,9921875	0,0039062	0 0351562	0,1445312	0,50

p	$n = 8,\ r = 5$	$n = 8,\ r = 4$	$n = 8,\ r = 3$	$n = 8,\ r = 2$	$n = 8,\ r = 1$	$n = 9,\ r = 9$	p
0,01	0,0000000	0,0000007	0,0000539	0,0026901	0,0772553	0,0000000	0,01
0,02	0,0000002	0,0000105	0,0004155—	0,0103369	0,1492370	0,0000000	0,02
0,03	0,0000013	0,0000515—	0,0013499	0,0223408	0,2162566	0,0000000	0,03
0,04	0,0000052	0,0001574	0,0030797	0,0381472	0,2786104	0,0000000	0,04

Продолжение табл. VI

p	$n=8,\ r=5$	$n=8,\ r=4$	$n=8,\ r=3$	$n=8,\ r=2$	$n=8,\ r=1$	$n=9,\ r=9$	p
0,05	0,0000154	0,0003718	0,0057882	0,0572447	0,3365796	0,0000000	0,05
0,06	0,0000373	0,0007456	0,0096229	0,0791618	0,3904311	0,0000000	0,06
0,07	0,0000786	0,0013359	0,0146987	0,1034657	0,4404182	0,0000000	0,07
0,08	0,0001493	0,0022033	0,0211005—	0,1297593	0,4867811	0,0000000	0,08
0,09	0,0002619	0,0034113	0,0288868	0,1576795	0,5297475—	0,0000000	0,09
0,10	0,0004316	0,0050244	0,0380918	0,1868953	0,5695328	0,0000000	0,10
0,11	0,0006765	0,0071068	0,0487281	0,2171054	0,6063411	0,0000000	0,11
0,12	0,0010169	0,0097216	0,0607892	0,2480369	0,6403655—	0,0000000	0,12
0,13	0,0014759	0,0129297	0,0742514	0,2794433	0,6717883	0,0000000	0,13
0,14	0,0020790	0,0167887	0,0890764	0,3111029	0,7007821	0,0000000	0,14
0,15	0,0028539	0,0213525—	0,1052128	0,3428170	0,7275095—	0,0000000	0,15
0,16	0,0038303	0,0266703	0,1225980	0,3744085—	0,7521241	0,0000001	0,16
0,17	0,0050399	0,0327863	0,1411603	0,4057205—	0,7747708	0,0000001	0,17
0,18	0,0065160	0,0397393	0,1608200	0,4366148	0,7955859	0,0000002	0,18
0,19	0,0082929	0,0475622	0,1814910	0,4669707	0,8146980	0,0000003	0,19
0,20	0,0104064	0,0562816	0,2030822	0,4966835	0,8322278	0,0000005	0,20
0,21	0,0128926	0,0659180	0,2254991	0,5256634	0,8482891	0,0000008	0,21
0,22	0,0157883	0,0764853	0,2486441	0,5538346	0,8629886	0,0000012	0,22
0,23	0,0191302	0,0879910	0,2724183	0,5811335	0,8764264	0,0000018	0,23
0,24	0,0229548	0,1004362	0,2967223	0,6075088	0,8886965	0,0000026	0,24
0,25	0,0272980	0,1138153	0,3214569	0,6329193	0,8998871	0,0000038	0,25
0,26	0,0321948	0,1281168	0,3465239	0,6573339	0,9100805	0,0000054	0,26

0,27	0,0376789	0,1433229	0,3718268	0,6807302	0,9193540	0,0000076	0,27
0,28	0,0437826	0,1594099	0,3972716	0,7030939	0,9277796	0,0000106	0,28
0,29	0,0505362	0,1763486	0,4227673	0,7244179	0,9354246	0,0000145	0,29
0,30	0,0579676	0,1941044	0,4482262	0,7447017	0,9423520	0,0000197	0,30
0,31	0,0661027	0,2126377	0,4735644	0,7639506	0,9486202	0,0000264	0,31
0,32	0,0749644	0,2319043	0,4987023	0,7821752	0,9542837	0,0000352	0,32
0,33	0,0845724	0,2518558	0,5235647	0,7993904	0,9593932	0,0000464	0,33
0,34	0,0949435—	0,2724399	0,5480813	0,8156156	0,9639959	0,0000607	0,34
0,35	0,1060909	0,2936006	0,5721863	0,8308731	0,9681355	0,0000788	0,35
0,36	0,1180242	0,3152791	0,5958195	0,8451888	0,9718525	0,0001016	0,36
0,37	0,1307490	0,3374141	0,6189255	0,8585906	0,9751844	0,0001300	0,37
0,38	0,1442673	0,3599420	0,6414542	0,8711089	0,9781660	0,0001652	0,38
0,39	0,1585766	0,3827973	0,6633607	0,8827757	0,9808293	0,0002087	0,39
0,40	0,1736704	0,4059136	0,6846054	0,8936243	0,9832038	0,0002621	0,40
0,41	0,1895380	0,4292234	0,7051539	0,9036892	0,9853170	0,0003274	0,41
0,42	0,2061644	0,4526588	0,7249765	0,9130054	0,9871937	0,0004067	0,42
0,43	0,2235301	0,4761522	0,7440490	0,9216086	0,9888571	0,0005026	0,43
0,44	0,2416115	0,4996359	0,7623517	0,9295345	0,9903283	0,0006181	0,44
0,45	0,2603807	0,5230437	0,7798697	0,9368189	0,9916266	0,0007567	0,45
0,46	0,2798056	0,5463101	0,7965925—	0,9434974	0,9927698	0,0009222	0,46
0,47	0,2998501	0,5693713	0,8125139	0,9496049	0,9937740	0,0011191	0,47
0,48	0,3204741	0,5921658	0,8276319	0,9551761	0,9946540	0,0013526	0,48
0,49	0,3416336	0,6146339	0,8419484	0,9602447	0,9954232	0,0016284	0,49
0,50	0,3632812	0,6367188	0,8554688	0,9648438	0,9960938	0,0019531	0,50

Продолжение табл. VI

p	$n=9,\ r=8$	$n=9,\ r=7$	$n=9,\ r=6$	$n=9,\ r=5$	$n=9,\ r=4$	$n=9,\ r=3$	p
0,01	0,0000000	0,0000000	0,0000000	0,0000000	0,0000012	0,0000803	0,01
0,02	0,0000000	0,0000000	0,0000000	0,0000004	0,0000186	0,0006139	0,02
0,03	0,0000000	0,0000000	0,0000001	0,0000028	0,0000904	0,0019796	0,03
0,04	0,0000000	0,0000000	0,0000003	0,0000113	0,0002743	0,0044824	0,04
0,05	0,0000000	0,0000000	0,0000012	0,0000332	0,0006426	0,0083610	0,05
0,06	0,0000000	0,0000001	0,0000033	0,0000798	0,0012783	0,0137953	0,06
0,07	0,0000000	0,0000003	0,0000082	0,0001666	0,0022713	0,0209123	0,07
0,08	0,0000000	0,0000007	0,0000178	0,0003136	0,0037151	0,0297932	0,08
0,09	0,0000000	0,0000015—	0,0000351	0,0005453	0,0057041	0,0404781	0,09
0,10	0,0000001	0,0000030	0,0000642	0,0008909	0,0083311	0,0529721	0,10
0,11	0,0000002	0,0000057	0,0001106	0,0013838	0,0116851	0,0672496	0,11
0,12	0,0000003	0,0000103	0,0001813	0,0020615—	0,0158497	0,0832589	0,12
0,13	0,0000006	0,0000177	0,0002847	0,0029649	0,0209015	0,1009264	0,13
0,14	0,0000012	0,0000292	0,0004315	0,0041384	0,0269090	0,1201601	0,14
0,15	0,0000020	0,0000464	0,0006340	0,0056287	0,0339315	0,1408534	0,15
0,16	0,0000033	0,0000715	0,0009068	0,0074847	0,0420187	0,1628877	0,16
0,17	0,0000053	0,0001071	0,0012664	0,0097568	0,0512099	0,1861356	0,17
0,18	0,0000083	0,0001565	0,0017318	0,0124962	0,0615338	0,2104631	0,18
0,19	0,0000127	0,0002238	0,0023240	0,0157541	0,0730086	0,2357321	0,19
0,20	0,0000189	0,0003139	0,0030664	0,0195814	0,0856417	0,2618025—	0,20
0,21	0,0000277	0,0004323	0,0039844	0,0240280	0,0994300	0,2885336	0,21
0,22	0,0000397	0,0005861	0,0051056	0,0291417	0,1143602	0,3157860	0,22

0,23	0,0000561	0,0007828	0,0064598	0,0349682	0,1304093	0,3434228	0,23
0,24	0,0000779	0,0010316	0,0080784	0,0415503	0,1475448	0,3713111	0,24
0,25	0,0001068	0,0013428	0,0099945	0,0489273	0,1657257	0,3993225	0,25
0,26	0,0001445	0,0017279	0,0122430	0,0571345—	0,1849026	0,4273345—	0,26
0,27	0,0001932	0,0021999	0,0148598	0,0662028	0,2050189	0,4552307	0,27
0,28	0,0002554	0,0027735—	0,0178821	0,0761583	0,2260112	0,4829018	0,28
0,29	0,0003342	0,0034646	0,0213477	0,0870218	0,2478100	0,5102460	0,29
0,30	0,0004330	0,0042909	0,0252948	0,0988087	0,2703409	0,5371688	0,30
0,31	0,0005561	0,0052716	0,0297621	0,1115286	0,2935250—	0,5635841	0,31
0,32	0,0007081	0,0064277	0,0347877	0,1251852	0,3172797	0,5894136	0,32
0,33	0,0008945—	0,0077818	0,0404096	0,1397759	0,3415198	0,6145872	0,33
0,34	0,0011215—	0,0093580	0,0466645—	0,1552923	0,3661579	0,6390429	0,34
0,35	0,0013962	0,0111822	0,0535882	0,1717193	0,3911056	0,6627267	0,35
0,36	0,0017265	0,0132818	0,0612147	0,1890360	0,4162737	0,6855925—	0,36
0,37	0,0021215	0,0156858	0,0695762	0,2072151	0,4415733	0,7076016	0,37
0,38	0,0025913	0,0184246	0,0787022	0,2262237	0,4669166	0,7287230	0,38
0,39	0,0031470	0,0215299	0,0886197	0,2460227	0,4922170	0,7489326	0,39
0,40	0,0038011	0,0250348	0,0993526	0,2665677	0,5173903	0,7682130	0,40
0,41	0,0045674	0,0289732	0,1109212	0,2878090	0,5423549	0,7865533	0,41
0,42	0,0054610	0,0333803	0,1233422	0,3096920	0,5670323	0,8039487	0,42
0,43	0,0064986	0,0382916	0,1366281	0,3321576	0,5913478	0,8203997	0,43
0,44	0,0076984	0,0437436	0,1507869	0,3551423	0,6152309	0,8359122	0,44
0,45	0,0090802	0,0497728	0,1658220	0,3785791	0,6386154	0,8504969	0,45
0,46	0,0106653	0,0564157	0,1817320	0,4023977	0,6614400	0,8641687	0,46
0,47	0,0124771	0,0637089	0,1985102	0,4265251	0,6836483	0,8769467	0,47

p	$n=9,\ r=8$	$n=9,\ r=7$	$n=9,\ r=6$	$n=9,\ r=5$	$n=9,\ r=4$	$n=9,\ r=3$	p
0,48	0,0145405	0,0716881	0,2161445	0,4508861	0,7051895	0,8888531	0,48
0,49	0,0168823	0,0803884	0,2346175	0,4754037	0,7260180	0,8999136	0,49
0,50	0,0195312	0,0898437	0,2539062	0,5000000	0,7460938	0,9101563	0,50
p	$n=9,\ r=2$	$n=9,\ r=1$	$n=10,\ r=10$	$n=10,\ r=9$	$n=10,\ r=8$	$n=10,\ r=7$	p
0,01	0,0034357	0,0864828	0,0000000	0,0000000	0,0000000	0,0000000	0,01
0,02	0,0131149	0,1662522	0,0000000	0,0000000	0,0000000	0,0000000	0,02
0,03	0,0281582	0,2397689	0,0000000	0,0000000	0,0000000	0,0000000	0,03
0,04	0,0477658	0,3074660	0,0000000	0,0000000	0,0000000	0,0000000	0,04
0,05	0,0712114	0,3697506	0,0000000	0,0000000	0,0000000	0,0000001	0,05
0,06	0,0978380	0,4370052	0,0000000	0,0000000	0,0000000	0,0000003	0,06
0,07	0,1270524	0,4795889	0,0000000	0,0000000	0,0000000	0,0000008	0,07
0,08	0,1583210	0,5278386	0,0000000	0,0000000	0,0000001	0,0000020	0,08
0,09	0,1911657	0,5720702	0,0000000	0,0000000	0,0000002	0,0000045—	0,09
0,10	0,2251590	0,6125795	0,0000000	0,0000000	0,0000004	0,0000091	0,10
0,11	0,2599213	0,6496436	0,0000000	0,0000000	0,0000008	0,0000173	0,11
0,12	0,2951163	0,6835216	0,0000000	0,0000000	0,0000015	0,0000308	0,12
0,13	0,3304482	0,7144558	0,0000000	0,0000001	0,0000029	0,0000525—	0,13
0,14	0,3656580	0,7426726	0,0000000	0,0000002	0,0000051	0,0000856	0,14

0,15	0,4005208	0,7683831	0,0000000	0,0000003	0,0000087	0,0001346	0,15
0,16	0,4348430	0,7917843	0,0000000	0,0000006	0,0000142	0,0002051	0,16
0,17	0,4684590	0,8130597	0,0000000	0,0000010	0,0000226	0,0003042	0,17
0,18	0,5012296	0,8323804	0,0000000	0,0000017	0,0000350	0,0004401	0,18
0,19	0,5330389	0,8499054	0,0000001	0,0000027	0,0000528	0,0006229	0,19
0,20	0,5637924	0,8657823	0,0000001	0,0000042	0,0000779	0,0008644	0,20
0,21	0,5934148	0,8801484	0,0000002	0,0000064	0,0001127	0,0011783	0,21
0,22	0,6218484	0,8931311	0,0000003	0,0000097	0,0001599	0,0015804	0,22
0,23	0,6490509	0,9048483	0,0000004	0,0000143	0,0002232	0,0020885	0,23
0,24	0,6749938	0,9154094	0,0000006	0,0000207	0,0003068	0,0027228	0,24
0,25	0,6996613	0,9249153	0,0000010	0,0000296	0,0004158	0,0035057	0,25
0,26	0,7230480	0,9334596	0,0000014	0,0000416	0,0005362	0,0044618	0,26
0,27	0,7451586	0,9411284	0,0000021	0,0000577	0,0007350	0,0056181	0,27
0,28	0,7660059	0,9480013	0,0000030	0,0000791	0,0009605—	0,0070039	0,28
0,29	0,7856098	0,9541515—	0,0000042	0,0001072	0,0012420	0,0086507	0,29
0,30	0,8039968	0,9596464	0,0000059	0,0001437	0,0015904	0,0105921	0,30
0,31	0,8211982	0,9645479	0,0000082	0,0001906	0,0020179	0,0128637	0,31
0,32	0,8372499	0,9689129	0,0000113	0,0002505	0,0025384	0,0155029	0,32
0,33	0,8521914	0,9727935—	0,0000153	0,0003263	0,0031673	0,0185489	0,33
0,34	0,8660649	0,9762373	0,0000206	0,0004214	0,0039219	0,0220422	0,34
0,35	0,8789150—	0,9792881	0,0000276	0,0005399	0,0048213	0,0260243	0,35
0,36	0,8907877	0,9819856	0,0000366	0,0006865	0,0058864	0,0305376	0,36
0,37	0,9017303	0,9843662	0,0000481	0,0008668	0,0071403	0,0356252	0,37
0,38	0,9117906	0,9864629	0,0000628	0,0010871	0,0086079	0,0413301	0,38
0,39	0,9210166	0,9883059	0,0000814	0,0013546	0,0103163	0,0476949	0,39
0,40	0,9294561	0,9899223	0,0001049	0,0016777	0,0122946	0,0547619	0,40

p	$n=9,\ r=2$	$n=9,\ r=1$	$n=10, r=10$	$n=10,\ r=9$	$n=10,\ r=8$	$n=10,\ r=7$	p
0,41	0,9371566	0,9913370	0,0001342	0,0020658	0,0145738	0,0625719	0,41
0,42	0,9441645—	0,9925723	0,0001708	0,0025295—	0,0171871	0,0711643	0,42
0,43	0,9505255—	0,9936485	0,0002161	0,0030809	0,0201696	0,0805763	0,43
0,44	0,9562838	0,9945838	0,0002720	0,0037335—	0,0235583	0,0908427	0,44
0,45	0,9614824	0,9953946	0 0003405	0,0045022	0,0273918	0,1019949	0,45
0,46	0,9661627	0,9960957	0,0004242	0,0054040	0,0317105	0,1140612	0,46
0,47	0,9703644	0,9967002	0,0005260	0,0064574	0,0365560	0,1270655—	0,47
0,48	0,9741255	0,9972201	0,0006493	0,0076828	0,0419713	0,1410272	0,48
0,49	0,9774822	0,9976658	0,0007979	0,0091028	0,0480003	0,1559607	0,49
0,50	0,9804688	0,9980469	0,0009766	0,0107422	0,0546875	0,1718750	0,50

p	$n=10,\ r=6$	$n=10,\ r=5$	$n=10,\ r=4$	$n=10,\ r=3$	$n=10,\ r=2$	$n=10,\ r=1$	p
0,01	0,0000000	0,0000000	0,0000020	0,0001138	0,0042662	0,0956179	0,01
0,02	0,0000000	0,0000007	0,0000305	0,0008639	0,0161776	0,1829272	0,02
0,03	0,0000001	0,0000054	0,0001471	0,0027650—	0,0345066	0,2625759	0,03
0,04	0,0000007	0,0000218	0,0004426	0,0062137	0,0581538	0,3351674	0,04
0,05	0,0000028	0,0000637	0,0010285—	0,0115036	0,0861384	0,4012631	0,05
0,06	0,0000079	0,0001517	0,0020293	0,0188378	0,1175880	0,4613849	0,06
0,07	0,0000193	0,0003139	0,0035761	0,0283421	0,1517299	0,5160177	0,07

0,08	0,0000415—	0,0005857	0,0058013	0,0400754	0,1878825—	0,5656115	0,08
0,09	0,0000810	0,0010096	0,0088338	0,0540400	0,2254471	0,6105839	0,09
0,10	0,0001469	0,0016349	0,0127952	0,0701908	0,2639011	0,6513216	0,10
0,11	0,0002507	0,0025170	0,0177972	0,0884435—	0,3027908	0,6881828	0,11
0,12	0,0004069	0,0037161	0,0239388	0,1086818	0,3417250—	0,7214990	0,12
0,13	0,0006332	0,0052967	0,0313048	0,1307642	0,3803692	0,7515766	0,13
0,14	0,0009505—	0,0073263	0,0399642	0,1545298	0,4184400	0,7786984	0,14
0,15	0,0013832	0,0098741	0,0499698	0,1798035	0,4557002	0,8031256	0,15
0,16	0,0019593	0,0130101	0,0613577	0,2064005	0,4919536	0,8250988	0,16
0,17	0,0027098	0,0168038	0,0741472	0,2341305	0,5270412	0,8448396	0,17
0,18	0,0036694	0,0213229	0,0883411	0,2628010	0,5608368	0,8625520	0,18
0,19	0,0048757	0,0266325—	0,1039261	0,2922204	0,5932435	0,8784233	0,19
0,20	0,0063694	0,0327935—	0,1208739	0,3222005—	0,6241904	0,8926258	0,20
0,21	0,0081935	0,0398624	0,1391418	0,3525586	0,6536289	0,9053172	0,21
0,22	0,0103936	0,0478897	0,1586739	0,3831197	0,6815306	0,9166422	0,22
0,23	0,0130167	0,0569196	0,1794024	0,4137173	0,7078843	0,9267332	0,23
0,24	0,0161116	0,0669890	0,2012487	0,4441949	0,7326936	0,9357111	0,24
0,25	0,0197277	0,0781269	0,2241249	0,4744072	0,7559748	0,9436865—	0,25
0,26	0,0239148	0,0903542	0,2479349	0,5042200	0,7777550	0,9507601	0,26
0,27	0,0287224	0,1036831	0,2725761	0,5335112	0,7980705—	0,9570237	0,27
0,28	0,0341994	0,1181171	0,2979405	0,5621710	0,8169646	0,9625609	0,28
0,29	0,0403932	0,1336503	0,3239164	0,5901015—	0,8344869	0,9674476	0,29
0,30	0,0473490	0,1502683	0,3503893	0,6172172	0,8506917	0,9717525—	0,30
0,31	0,0551097	0,1679475—	0,3772433	0,6434445—	0,8656366	0,9755381	0,31
0,32	0,0637149	0,1866554	0,4043626	0,6687212	0,8793821	0,9788608	0,32

p	$n=10,\ r=6$	$n=10,\ r=5$	$n=10,\ r=4$	$n=10,\ r=3$	$n=10,\ r=2$	$n=10,\ r=1$	p
0,33	0,0732005—	0,2063514	0,4316320	0,6929966	0,8919901	0,9817716	0,33
0,34	0,0835979	0,2269866	0,4589388	0,7162304	0,9035235	0,9843166	0,34
0,35	0,0949341	0,2485045—	0,4861730	0,7383926	0,9140456	0,9865373	0,35
0,36	0,1072304	0,2708415	0,5132284	0,7594627	0,9236190	0,9884708	0,36
0,37	0,1205026	0,2939277	0,5400038	0,7794292	0,9323056	0,9901507	0,37
0,38	0,1347603	0,3176870	0,5664030	0,7982887	0,9401661	0,9916070	0,38
0,39	0,1500068	0,3420385—	0,5923361	0,8160453	0,9472594	0,9928666	0,39
0,40	0,1662386	0,3668967	0,6177194	0,8327102	0,9536426	0,9939534	0,40
0,41	0,1834452	0,3921728	0,6424762	0,8483007	0,9593705	0,9948888	0,41
0,42	0,2016092	0,4177749	0,6665372	0,8628393	0,9644958	0,9956920	0,42
0,43	0,2207058	0,4436094	0,6898401	0,8763538	0,9690684	0,9963797	0,43
0,44	0,2407033	0,4695813	0,7123307	0,8888757	0,9731358	0,9969669	0,44
0,45	0,2615627	0,4955954	0,7339621	0,9004403	0,9767429	0,9974670	0,45
0,46	0,2832382	0,5215571	0,7546952	0,9110859	0,9799319	0,9978917	0,46
0,47	0,3056772	0,5473730	0,7744985	0,9208530	0,9827422	0,9982511	0,47
0,48	0,3288205—	0,5729517	0,7933480	0,9297839	0,9852109	0,9985544	0,48
0,49	0,3526028	0,5982047	0,8112268	0,9379222	0,9873722	0,9988096	0,49
0,50	0,3769531	0,6230469	0,8281250	0,9453125	0,9892578	0,9990234	0,50

Таблица VII

Распределение $\mathcal{F}$

В таблице приведены значения $\mathcal{F}_{\nu_1, \nu_2, \alpha}$, где ν_1—число степеней свободы числителя, ν_2— число степеней свободы знаменателя и $\alpha = 1 - \int_0^{\mathcal{F}_{\nu_1, \nu_2, \alpha}} \Phi(\mathcal{F})\, d\mathcal{F}$. Соответствующие значения $\mathcal{F}$ для $1-\alpha$ могут быть найдены из соотношения $\mathcal{F}_{\nu_1, \nu_2, 1-\alpha} = (\mathcal{F}_{\nu_1, \nu_2, 1-\alpha})^{-1}$

$\alpha =$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	16 211	20 000	21 615	22 500	23 056	23 437	23 715	23 925	24 091
2	198,50	199,00	199,17	199,25	199,30	199,33	199,36	199,37	199,39
3	55,552	49,799	47,467	46,195	45,392	44,838	44,434	44,126	43,882
4	31,333	26,284	24,259	23,155	22,456	21,975	21,622	21,352	21,139
5	22,785	18,314	16,530	15,556	14,940	14,513	14,200	13,961	13,772
6	18,635	14,544	12,917	12,028	11,464	11,073	10,786	10,566	10,391
7	16,236	12,404	10,882	10,050	9,5221	9,1554	8,8854	8,6781	8,5138
8	14,688	11,042	9,5965	8,8051	8,3018	7,9520	7,6942	7,4960	7,3386
9	13,614	10,107	8,7171	7,9559	7,4711	7,1338	6,8849	6,6933	6,5411
10	12,826	9,4270	8,0807	7,3428	6,8723	6,5446	6,3025	6,1159	5,9676
11	12,226	8,9122	7,6004	6,8809	6,4217	6,1015	5,8648	5,6821	5,5368
12	11,754	8,5096	7,2258	6,5211	6,0711	5,7570	5,5245	5,3451	5,2021
13	11,374	8,1865	6,9257	6,2335	5,7910	5,4819	5,2529	5,0761	4,9351
14	11,060	7,9217	6,6803	5,9984	5,5623	5,2574	5,0313	4,8566	4,7173
15	10,798	7,7008	6,4760	5,8029	5,3721	5,0708	4,8473	4,6743	4,5364
16	10,575	7,5138	6,3034	5,6378	5,2117	4,9134	4,6920	4,5207	4,3838
17	10,384	7,3536	6,1556	5,4967	5,0746	4,7789	4,5594	4,3893	4,2535
18	10,218	7,2148	6,0277	5,3746	4,9560	4,6627	4,4448	4,2759	4,1410
19	10,073	7,0935	5,9161	5,2681	4,8526	4,5614	4,3448	4,1770	4,0428
20	9,9439	6,9865	5,8177	5,1743	4,7616	4,4721	4,2569	4,0900	3,9564
21	9,8295	6,8914	5,7304	5,0911	4,6808	4,3931	4,1789	4,0128	3,8799
22	9,7271	6,8064	5,6524	5,0168	4,6088	4,3225	4,1094	3,9440	3,8116
23	9,6348	6,7300	5,5823	4,9500	4,5441	4,2591	4,0469	3,8822	3,7502
24	9,5513	6,6610	5,5190	4,8898	4,4857	4,2019	3,9905	3,8264	3,6949
25	9,4753	6,5982	5,4615	4,8351	4,4327	4,1500	3,9394	3,7758	3,6447
26	9,4059	6,5409	5,4091	4,7852	4,3844	4,1027	3,8928	3,7297	3,5989
27	9,3423	6,4885	5,3611	4,7396	4,3402	4,0594	3,8501	3,6875	3,5571
28	9,2838	6,4403	5,3170	4,6977	4,2996	4,0197	3,8110	3,6487	3,5186
29	9,2297	6,3958	5,2764	4,6591	4,2622	3,9830	3,7749	3,6130	3,4832
30	9,1797	6,3547	5,2388	4,6233	4,2276	3,9492	3,7416	3,5801	3,4505
40	8,8278	6,0664	4,9759	4,3738	3,9860	3,7129	3,5088	3,3498	3,2220
60	8,4946	5,7950	4,7290	4,1399	3,7600	3,4918	3,2911	3,1344	3,0083
120	8,1790	5,5393	4,4973	3,9207	3,5482	3,2849	3,0874	2,9330	2,8083
∞	7,8794	5,2983	4,2794	3,7151	3,3499	3,0913	2,8968	2,7444	2,6210

0,005

10	12	15	20	24	30	40	60	120	∞
24 224	24 426	24 630	24 836	24 940	25 044	25 148	25 253	25 359	25 465
199,40	199,42	199,43	199,45	199,46	199,47	199,47	199,48	199,49	199,51
43,686	43,387	43,085	42,778	42,622	42,466	42,308	42,149	41,989	41,829
20,967	20,705	20,438	20,167	20,030	19,892	19,752	19,611	19,468	19,325
13,618	13,384	13,146	12,903	12,780	12,656	12,530	12,402	12,274	12,144
10,250	10,034	9,8140	9,5888	9,4741	9,3583	9,2408	9,1219	9,0015	8,8793
8,3803	8,1764	7,9678	7,7540	7,6450	7,5345	7,4225	7,3088	7,1933	7,0760
7,2107	7,0149	6,8143	6,6082	6,5029	6,3961	6,2875	6,1772	6,0649	5,9505
6,4171	6,2274	6,0325	5,8318	5,7292	5,6248	5,5186	5,4104	5,3001	5,1875
5,8467	5,6613	5,4707	5,2740	5,1732	5,0705	4,9659	4,8592	4,7501	4,6385
5,4182	5,2363	5,0489	4,8552	4,7557	4,6543	4,5508	4,4450	4,3367	4,2256
5,0855	4,9063	4,7214	4,5299	4,4315	4,3309	4,2282	4,1229	4,0149	3,9039
4,8199	4,6429	4,4600	4,2703	4,1726	4,0727	3,9704	3,8655	3,7577	3,6465
4,6034	4,4281	4,2468	4,0585	3,9614	3,8619	3,7600	3,6553	3,5473	3,4359
4,4236	4,2498	4,0698	3,8826	3,7859	3,6867	3,5850	3,4803	3,3722	3,2602
4,2719	4,0994	3,9205	3,7342	3,6378	3,5388	3,4372	3,3324	3,2240	3,1115
4,1423	3,9709	3,7929	3,6073	3,5112	3,4124	3,3107	3,2058	3,0971	2,9839
4,0305	3,8599	3,6827	3,4977	3,4017	3,3030	3,2014	3,0962	2,9871	2,8732
3,9329	3,7631	3,5866	3,4020	3,3062	3,2075	3,1058	3,0004	2,8908	2,7762
3,8470	3,6779	3,5020	3,3178	3,2220	3,1234	3,0215	2,9159	2,8058	2,6904
3,7709	3,6024	3,4270	3,2431	3,1474	3,0488	2,9467	2,8408	2,7302	2,6140
3,7030	3,5350	3,3600	3,1764	3,0807	2,9821	2,8799	2,7736	2,6625	2,5455
3,6420	3,4745	3,2999	3,1165	3,0208	2,9221	2,8198	2,7132	2,6016	2,4837
3,5870	3,4199	3,2456	3,0624	2,9667	2,8679	2,7654	2,6585	2,5463	2,4276
3,5370	3,3704	3,1963	3,0133	2,9176	2,8187	2,7160	2,6088	2,4960	2,3765
3,4916	3,3252	3,1515	2,9685	2,8728	2,7738	2,6709	2,5633	2,4501	2,3297
3,4499	3,2839	3,1104	2,9275	2,8318	2,7327	2,6296	2,5217	2,4078	2,2867
3,4117	3,2460	3,0727	2,8899	2,7941	2,6949	2,5916	2,4834	2,3689	2,2469
3,3765	3,2111	3,0379	2,8551	2,7594	2,6601	2,5565	2,4479	2,3330	2,2102
3,3440	3,1787	3,0057	2,8230	2,7272	2,6278	2,5241	2,4151	2,2997	2,1760
3,1167	2,9531	2,7811	2,5984	2,5020	2,4015	2,2958	2,1838	2,0635	1,9318
2,9042	2,7419	2,5705	2,3872	2,2898	2,1874	2,0789	1,9622	1,8341	1,6885
2,7052	2,5439	2,3727	2,1881	2,0890	1,9839	1,8709	1,7469	1,6055	1,4311
2,5188	2,3583	2,1868	1,9998	1,8983	1,7891	1,6691	1,5325	1,3637	1,0000

$\alpha =$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	4052,2	4999,5	5403,3	5624,6	5763,7	5859,0	5928,3	5981,6	6022,5
2	98,503	99,000	99,166	99,249	99,299	99,332	99,356	99,374	99,388
3	34,116	30,817	29,457	28,710	28,237	27,911	27,672	27,489	27,345
4	21,198	18,000	16,694	15,977	15,522	15,207	14,976	14,799	14,659
5	16,258	13,274	12,060	11,392	10,967	10,672	10,456	10,289	10,158
6	13,745	10,925	9,7795	9,1483	8,7459	8,4661	8,2600	8,1016	7,9761
7	12,246	9,5466	8,4513	7,8467	7,4604	7,1914	6,9928	6,8401	6,7188
8	11,259	8,6491	7,5910	7,0060	6,6318	6,3707	6,1776	6,0289	5,9106
9	10,561	8,0215	6,9919	6,4221	6,0569	5,8018	5,6129	5,4671	5,3511
10	10,044	7,5594	6,5523	5,9943	5,6363	5,3858	5,2001	5,0567	4,9424
11	9,6460	7,2057	6,2167	5,6683	5,3160	5,0692	4,8861	4,7445	4,6315
12	9,3302	6,9266	5,9526	5,4119	5,0643	4,8206	4,6395	4,4994	4,3875
13	9,0738	6,7010	5,7394	5,2053	4,8616	4,6204	4,4410	4,3021	4,1911
14	8,8616	6,5149	5,5639	5,0354	4,6950	4,4558	4,2779	4,1399	4,0297
15	8,6831	6,3589	5,4170	4,8932	4,5556	4,3183	4,1415	4,0045	3,8948
16	8,5310	6,2262	5,2922	4,7726	4,4374	4,2016	4,0259	3,8896	3,7804
17	8,3997	6,1121	5,1850	4,6690	4,3359	4,1015	3,9267	3,7910	3,6822
18	8,2854	6,0129	5,0919	4,5790	4,2479	4,0146	3,8406	3,7054	3,5971
19	8,1850	5,9259	5,0103	4,5003	4,1708	3,9386	3,7653	3,6305	3,5225
20	8,0960	5,8489	4,9382	4,4307	4,1027	3,8714	3,6987	3,5644	3,4567
21	8,0166	5,7804	4,8740	4,3688	4,0421	3,8117	3,6396	3,5056	3,3981
22	7,9454	5,7190	4,8166	4,3134	3,9880	3,7583	3,5867	3,4530	3,3458
23	7,8811	5,6637	4,7649	4,2635	3,9392	3,7102	3,5390	3,4057	3,2986
24	7,8229	5,6136	4,7181	4,2184	3,8951	3,6667	3,4959	3,3629	3,2560
25	7,7698	5,5680	4,6755	4,1774	3,8550	3,6272	3,4568	3,3239	3,2172
26	7,7213	5,5263	4,6366	4,1400	3,8183	3,5911	3,4210	3,2884	3,1818
27	7,6767	5,4881	4,6009	4,1056	3,7848	3,5580	3,3882	3,2558	3,1494
28	7,6356	5,4529	4,5681	4,0740	3,7539	3,5276	3,3581	3,2259	3,1195
29	7,5976	5,4205	4,5378	4,0449	3,7254	3,4995	3,3302	3,1982	3,0920
30	7,5625	5,3904	4,5097	4,0179	3,6990	3,4735	3,3045	3,1726	3,0665
40	7,3141	5,1785	4,3126	3,8283	3,5138	3,2910	3,1238	2,9930	2,8876
60	7,0771	4,9774	4,1259	3,6491	3,3389	3,1187	2,9530	2,8233	2,7185
120	6,8510	4,7865	3,9493	3,4796	3,1735	2,9559	2,7918	2,6629	2,5586
∞	6,6349	4,6052	3,7816	3,3192	3,0173	2,8020	2,6393	2,5113	2,4073

0,01 *Продолжение табл. VII*

10	12	15	20	24	30	40	60	120	∞
6055,8	6106,3	6157,3	6208,7	6234,6	6260,7	6286,8	6313,0	6339,4	6366,0
99,399	99,416	99,432	99,449	99,458	99,466	99,474	99,483	99,491	99,501
27,229	27,052	26,872	26,690	26,598	26,505	26,411	26,316	26,221	26,125
14,546	14,374	14,198	14,020	13,929	13,838	13,745	13,652	13,558	13,463
10,051	9,8883	9,7222	9,5527	9,4665	9,3793	9,2912	9,2020	9,1118	9,0204
7,8741	7,7183	7,5590	7,3958	7,3127	7,2285	7,1432	7,0568	6,9690	6,8801
6,6201	6,4691	6,3143	6,1554	6,0743	5,9921	5,9084	5,8236	5,7372	5,6495
5,8143	5,6668	5,5151	5,3591	5,2793	5,1981	5,1156	5,0316	4,9460	4,8588
5,2565	5,1114	4,9621	4,8080	4,7290	4,6486	4,5667	4,4831	4,3978	4,3105
4,8492	4,7059	4,5582	4,4054	4,3269	4,2469	4,1653	4,0819	3,9965	3,9090
4,5393	4,3974	4,2509	4,0990	4,0209	3,9411	3,8596	3,7761	3,6904	3,6025
4,2961	4,1553	4,0096	3,8584	3,7805	3,7008	3,6192	3,5355	3,4494	3,3608
4,1003	3,9603	3,8154	3,6646	3,5868	3,5070	3,4253	3,3413	3,2548	3,1654
3,9394	3,8001	3,6557	3,5052	3,4274	3,3476	3,2656	3,1813	3,0942	3,0040
3,8049	3,6662	3,5222	3,3719	3,2940	3,2141	3,1319	3,0471	2,9595	2,8684
3,6909	3,5527	3,4089	3,2588	3,1808	3,1007	3,0182	2,9330	2,8447	2,7528
3,5931	3,4552	3,3117	3,1615	3,0835	3,0032	2,9205	2,8348	2,7459	2,6530
3,5082	3,3706	3,2273	3,0771	2,9990	2,9185	2,8354	2,7493	2,6597	2,5660
3,4338	3,2965	3,1533	3,0031	2,9249	2,8442	2,7608	2,6742	2,5839	2,4893
3,3682	3,2311	3,0880	2,9377	2,8594	2,7785	2,6947	2,6077	2,5168	2,4212
3,3098	3,1729	3,0299	2,8796	2,8011	2,7200	2,6359	2,5484	2,4568	2,3603
3,2576	3,1209	2,9780	2,8274	2,7488	2,6675	2,5831	2,4951	2,4029	2,3055
3,2106	3,0740	2,9311	2,7805	2,7017	2,6202	2,5355	2,4471	2,3542	2,2559
3,1681	3,0316	2,8887	2,7380	2,6591	2,5773	2,4923	2,4035	2,3099	2,2107
3,1294	2,9931	2,8502	2,6993	2,6203	2,5383	2,4530	2,3637	2,2695	2,1694
3,0941	2,9579	2,8150	2,6640	2,5848	2,5026	2,4170	2,3273	2,2325	2,1315
3,0618	2,9256	2,7827	2,6316	2,5522	2,4699	2,3840	2,2938	2,1984	2,0965
3,0320	2,8959	2,7530	2,6017	2,5223	2,4397	2,3535	2,2629	2,1670	2,0642
3,0045	2,8685	2,7256	2,5742	2,4946	2,4118	2,3253	2,2344	2,1378	2,0342
2,9791	2,8431	2,7002	2,5487	2,4689	2,3860	2,2992	2,2079	2,1107	2,0062
2,8005	2,6648	2,5216	2,3689	2,2880	2,2034	2,1142	2,0194	1,9172	1,8047
2,6318	2,4961	2,3523	2,1978	2,1154	2,0285	1,9360	1,8363	1,7263	1,6006
2,4721	2,3363	2,1915	2,0346	1,9500	1,8600	1,7628	1,6557	1,5330	1,3805
2,3209	2,1848	2,0385	1,8783	1,7908	1,6964	1,5923	1,4730	1,3246	1,0000

$\alpha =$

v_2 \ v_1	1	2	3	4	5	6	7	8	9
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385
3	10,128	9,5521	9,2766	9,1172	9,0135	8,9406	8,8868	8,8452	8,8123
4	7,7086	6,9443	6,5914	6,3883	6,2560	6,1631	6,0942	6,0410	5,9988
5	6,6079	5,7861	5,4095	5,1922	5,0503	4,9503	4,8759	4,8183	4,7725
6	5,9874	5,1433	4,7571	4,5337	4,3874	4,2839	4,2066	4,1468	4,0990
7	5,5914	4,7374	4,3468	4,1203	3,9715	3,8660	3,7870	3,7257	3,6767
8	5,3177	4,4590	4,0662	3,8378	3,6875	3,5806	3,5005	3,4381	3,3881
9	5,1174	4,2565	3,8626	3,6331	3,4817	3,3738	3,2927	3,2296	3,1789
10	4,9646	4,1028	3,7083	3,4780	3,3258	3,2172	3,1355	3,0717	3,0204
11	4,8443	3,9823	3,5874	3,3567	3,2039	3,0946	3,0123	2,9480	2,8962
12	4,7472	3,8853	3,4903	3,2592	3,1059	2,9961	2,9134	2,8486	2,7964
13	4,6672	3,8056	3,4105	3,1791	3,0254	2,9153	2,8321	2,7669	2,7144
14	4,6001	3,7389	3,3439	3,1122	2,9582	2,8477	2,7642	2,6987	2,6458
15	4,5431	3,6823	3,2874	3,0566	2,9013	2,7905	2,7066	2,6408	2,5876
16	4,4940	3,6337	3,2389	3,0069	2,8524	2,7413	2,6572	2,5911	2,5377
17	4,4513	3,5915	3,1968	2,9647	2,8100	2,6987	2,6143	2,5480	2,4943
18	4,4139	3,5546	3,1599	2,9277	2,7729	2,6613	2,5767	2,5102	2,4563
19	4,3808	3,5219	3,1274	2,8951	2,7401	2,6283	2,5435	2,4768	2,4227
20	4,3513	3,4928	3,0984	2,8661	2,7109	2,5990	2,5140	2,4471	2,3928
21	4,3248	3,4668	3,0725	2,8401	2,6848	2,5727	2,4876	2,4205	2,3661
22	4,3009	3,4434	3,0491	2,8167	2,6613	2,5491	2,4638	2,3965	2,3419
23	4,2793	3,4221	3,0280	2,7955	2,6400	2,5277	2,4422	2,3748	2,3201
24	4,2597	3,4028	3,0088	2,7763	2,6207	2,5082	2,4226	2,3551	2,3002
25	4,2417	3,3852	2,9912	2,7587	2,6030	2,4904	2,4047	2,3371	2,2821
26	4,2252	3,3690	2,9751	2,7426	2,5868	2,4741	2,3883	2,3205	2,2655
27	4,2100	3,3541	2,9604	2,7278	2,5719	2,4591	2,3732	2,3053	2,2501
28	4,1960	3,3404	2,9467	2,7141	2,5581	2,4453	2,3593	2,2913	2,2360
29	4,1830	3,3277	2,9340	2,7014	2,5454	2,4324	2,3463	2,2782	2,2229
30	4,1709	3,3158	2,9223	2,6896	2,5336	2,4205	2,3343	2,2662	2,2107
40	4,0848	3,2317	2,8387	2,6060	2,4495	2,3359	2,2490	2,1802	2,1240
60	4,0012	3,1504	2,7581	2,5252	2,3683	2,2540	2,1665	2,0970	2,0401
120	3,9201	3,0718	2,6802	2,4472	2,2900	2,1750	2,0867	2,0164	1,9588
∞	3,8415	2,9957	2,6049	2,3719	2,2141	2,0986	2,0096	1,9384	1,8799

0,05

Продолжение табл. VII

10	12	15	20	24	30	40	60	120	∞
241,88	243,91	245,95	248,01	249,05	250,09	251,14	252,20	253,25	254,32
19,396	19,413	19,429	19,446	19,454	19,462	19,471	19,479	19,487	19,496
8,7855	8,7446	8,7029	8,6602	8,6385	8,6166	8,5944	8,5720	8,5494	8,5265
5,9644	5,9117	5,8578	5,8025	5,7744	5,7459	5,7170	5,6878	5,6581	5,6281
4,7351	4,6777	4,6188	4,5581	4,5272	4,4957	4,4638	4,4314	4,3984	4,3650
4,0600	3,9999	3,9381	3,8743	3,8415	3,8082	3,7743	3,7398	3,7047	3,6688
3,6365	3,5747	3,5108	3,4445	3,4105	3,3758	3,3404	3,3043	3,2674	3,2298
3,3472	3,2840	3,2184	3,1503	3,1152	3,0794	3,0428	3,0053	2,9669	2,9276
3,1373	3,0729	3,0061	2,9365	2,9005	2,8637	2,8259	2,7872	2,7475	2,7067
2,9782	2,9130	2,8450	2,7740	2,7372	2,6996	2,6609	2,6211	2,5801	2,5379
2,8536	2,7876	2,7186	2,6464	2,6090	2,5705	2,5309	2,4901	2,4480	2,4045
2,7534	2,6866	2,6169	2,5436	2,5055	2,4663	2,4259	2,3842	2,3410	2,2962
2,6710	2,6037	2,5331	2,4589	2,4202	2,3803	2,3392	2,2966	2,2524	2,2064
2,6021	2,5342	2,4630	2,3879	2,3487	2,3082	2,2664	2,2230	2,1778	2,1307
2,5437	2,4753	2,4035	2,3275	2,2878	2,2468	2,2043	2,1601	2,1141	2,0658
2,4935	2,4247	2,3522	2,2756	2,2354	2,1938	2,1507	2,1058	2,0589	2,0096
2,4499	2,3807	2,3077	2,2304	2,1898	2,1477	2,1040	2,0584	2,0107	1,9604
2,4117	2,3421	2,2686	2,1906	2,1497	2,1071	2,0629	2,0166	1,9681	1,9168
2,3779	2,3080	2,2341	2,1555	2,1141	2,0712	2,0264	1,9796	1,9302	1,8780
2,3479	2,2776	2,2033	2,1242	2,0825	2,0391	1,9938	1,9464	1,8963	1,8432
2,3210	2,2504	2,1757	2,0960	2,0540	2,0102	1,9645	1,9165	1,8657	1,8117
2,2967	2,2258	2,1508	2,0707	2,0283	1,9842	1,9380	1,8895	1,8380	1,7831
2,2747	2,2036	2,1282	2,0476	2,0050	1,9605	1,9139	1,8649	1,8128	1,7570
2,2547	2,1834	2,1077	2,0267	1,9838	1,9390	1,8920	1,8424	1,7897	1,7331
2,2365	2,1649	2,0889	2,0075	1,9643	1,9192	1,8718	1,8217	1,7684	1,7110
2,2197	2,1479	2,0716	1,9898	1,9464	1,9010	1,8533	1,8027	1,7488	1,6906
2,2043	2,1323	2,0558	1,9736	1,9299	1,8842	1,8361	1,7851	1,7307	1,6717
2,1900	2,1179	2,0411	1,9586	1,9147	1,8687	1,8203	1,7689	1,7138	1,6541
2,1768	2,1045	2,0275	1,9446	1,9005	1,8543	1,8055	1,7537	1,6981	1,6377
2,1646	2,0921	2,0148	1,9317	1,8874	1,8409	1,7918	1,7396	1,6835	1,6223
2,0772	2,0035	1,9245	1,8389	1,7929	1,7444	1,6928	1,6373	1,5766	1,5089
1,9926	1,9174	1,8364	1,7480	1,7001	1,6491	1,5943	1,5343	1,4673	1,3893
1,9105	1,8337	1,7505	1,6587	1,6084	1,5543	1,4952	1,4290	1,3519	1,2539
1,8307	1,7522	1,6664	1,5705	1,5173	1,4591	1,3940	1,3180	1,2214	1,0000

$\alpha =$

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9
1	1,0000	1,5000	1,7092	1,8227	1,8937	1,9422	1,9774	2,0041	2,0250
2	0,66667	1,0000	1,1349	1,2071	1,2519	1,2824	1,3045	1,3213	1,3344
3	0,58506	0,88110	1,0000	1,0632	1,1024	1,1289	1,1482	1,1627	1,1741
4	0,54863	0,82843	0,94054	1,0000	1,0367	1,0617	1,0797	1,0933	1,1040
5	0,52807	0,79877	0,90715	0,96456	1,0000	1,0240	1,0414	1,0545	1,0648
6	0,51489	0,77976	0,88578	0,94191	0,97654	1,0000	1,0169	1,0298	1,0398
7	0,50572	0,76655	0,87095	0,92619	0,96026	0,98334	1,0000	1,0126	1,0224
8	0,49898	0,75683	0,86004	0,91464	0,94831	0,97111	0,98757	1,0000	1,0097
9	0,49382	0,74938	0,85168	0,90580	0,93916	0,96175	0,97805	0,99037	1,0000
10	0,48973	0,74349	0,84508	0,89882	0,93193	0,95436	0,97054	0,98276	0,99232
11	0,48644	0,73872	0,83973	0,89316	0,92608	0,94837	0,96445	0,97661	0,98610
12	0,48369	0,73477	0,83530	0,88848	0,92124	0,94342	0,95943	0,97152	0,98097
13	0,48141	0,73145	0,83159	0,88454	0,91718	0,93926	0,95520	0,96724	0,97665
14	0,47944	0,72862	0,82842	0,88119	0,91371	0,93573	0,95161	0,96360	0,97298
15	0,47775	0,72619	0,82569	0,87830	0,91073	0,93267	0,94850	0,96046	0,96981
16	0,47628	0,72406	0,82330	0,87578	0,90812	0,93001	0,94580	0,95773	0,96705
17	0,47499	0,72219	0,82121	0,87357	0,90584	0,92767	0,94342	0,95532	0,96462
18	0,47385	0,72053	0,81936	0,87161	0,90381	0,92560	0,94132	0,95319	0,96247
19	0,47284	0,71906	0,81771	0,86987	0,90200	0,92375	0,93944	0,95129	0,96056
20	0,47192	0,71773	0,81621	0,86830	0,90038	0,92210	0,93776	0,94959	0,95884
21	0,47108	0,71653	0,81487	0,86688	0,89891	0,92060	0,93624	0,94805	0,95728
22	0,47033	0,71545	0,81365	0,86559	0,89759	0,91924	0,93486	0,94665	0,95588
23	0,46965	0,71446	0,81255	0,86442	0,89638	0,91800	0,93360	0,94538	0,95459
24	0,46902	0,71356	0,81153	0,86335	0,89527	0,91687	0,93245	0,94422	0,95342
25	0,46844	0,71272	0,81061	0,86236	0,89425	0,91583	0,93140	0,94315	0,95234
26	0,46793	0,71195	0,80975	0,86145	0,89331	0,91487	0,93042	0,94217	0,95135
27	0,46744	0,71124	0,80894	0,86061	0,89244	0,91399	0,92952	0,94126	0,95044
28	0,46697	0,71059	0,80820	0,85983	0,89164	0,91317	0,92869	0,94041	0,94958
29	0,46654	0,70999	0,80753	0,85911	0,89089	0,91241	0,92791	0,93963	0,94879
30	0,46616	0,70941	0,80689	0,85844	0,89019	0,91169	0,92719	0,93889	0,94805
40	0,46330	0,70531	0,80228	0,85357	0,88516	0,90654	0,92197	0,93361	0,94272
60	0,46053	0,70122	0,79770	0,84873	0,88017	0,90144	0,91679	0,92838	0,93743
120	0,45774	0,69717	0,79314	0,84392	0,87521	0,89637	0,91164	0,92318	0,93218
∞	0,45494	0,69315	0,78866	0,83918	0,87029	0,89135	0,90654	0,91802	0,92698

0,50 *Продолжение табл. VII*

10	12	15	20	24	30	40	60	120	∞
2,0419	2,0674	2,0931	2,1190	2,1321	2,1452	2,1584	2,1716	2,1848	2,1981
1,3450	1,3610	1,3771	1,3933	1,4014	1,4096	1,4178	1,4261	1,4344	1,4427
1,1833	1,1972	1,2111	1,2252	1,2322	1,2393	1,2464	1,2536	1,2608	1,2680
1,1126	1,1255	1,1386	1,1517	1,1583	1,1649	1,1716	1,1782	1,1849	1,1916
1,0730	1,0855	1,0980	1,1106	1,1170	1,1234	1,1297	1,1361	1,1426	1,1490
1,0478	1,0600	1,0722	1,0845	1,0907	1,0969	1,1031	1,1093	1,1156	1,1219
1,0304	1,0423	1,0543	1,0664	1,0724	1,0785	1,0846	1,0908	1,0969	1,1031
1,0175	1,0293	1,0412	1,0531	1,0591	1,0651	1,0711	1,0771	1,0832	1,0893
1,0077	1,0194	1,0311	1,0429	1,0489	1,0548	1,0608	1,0667	1,0727	1,0788
1,0000	1,0116	1,0232	1,0349	1,0408	1,0467	1,0526	1,0585	1,0645	1,0705
0,99373	1,0052	1,0168	1,0284	1,0343	1,0401	1,0460	1,0519	1,0578	1,0637
0,98856	1,0000	1,0115	1,0231	1,0289	1,0347	1,0405	1,0464	1,0523	1,0582
0,98431	0,99560	1,0071	1,0186	1,0243	1,0301	1,0360	1,0418	1,0476	1,0535
0,98051	0,99186	1,0033	1,0147	1,0205	1,0263	1,0321	1,0379	1,0437	1,0495
0,97732	0,98863	1,0000	1,0114	1,0172	1,0229	1,0287	1,0345	1,0403	1,0461
0,97454	0,98582	0,99716	1,0086	1,0143	1,0200	1,0258	1,0315	1,0373	1,0431
0,97209	0,98334	0,99466	1,0060	1,0117	1,0174	1,0232	1,0289	1,0347	1,0405
0,96993	0,98116	0,99245	1,0038	1,0095	1,0152	1,0209	1,0267	1,0324	1,0382
0,96800	0,97920	0,99047	1,0018	1,0075	1,0132	1,0189	1,0246	1,0304	1,0361
0,96626	0,97746	0,98870	1,0000	1,0057	1,0114	1,0171	1,0228	1,0285	1,0343
0,96470	0,97587	0,98710	0,99838	1,0040	1,0097	1,0154	1,0211	1,0268	1,0326
0,96328	0,97444	0,98565	0,99692	1,0026	1,0082	1,0139	1,0196	1,0253	1,0311
0,96199	0,97313	0,98433	0,99558	1,0012	1,0069	1,0126	1,0183	1,0240	1,0297
0,96081	0,97194	0,98312	0,99436	1,0000	1,0057	1,0113	1,0170	1,0227	1,0284
0,95972	0,97084	0,98201	0,99324	0,99887	1,0045	1,0102	1,0159	1,0215	1,0273
0,95872	0,96983	0,98099	0,99220	0,99783	1,0035	1,0091	1,0148	1,0205	1,0262
0,95779	0,96889	0,98004	0,99125	0,99687	1,0025	1,0082	1,0138	1,0195	1,0252
0,95694	0,96802	0,97917	0,99036	0,99598	1,0016	1,0073	1,0129	1,0186	1,0243
0,95614	0,96722	0,97835	0,98954	0,99515	1,0008	1,0064	1,0121	1,0177	1,0234
0,95540	0,96647	0,97759	0,98877	0,99438	1,0000	1,0056	1,0113	1,0170	1,0226
0,95003	0,96104	0,97211	0,98323	0,98880	0,99440	1,0000	1,0056	1,0113	1,0169
0,94471	0,95566	0,96667	0,97773	0,98328	0,98884	0,99441	1,0000	1,0056	1,0112
0,93943	0,95032	0,96128	0,97228	0,97780	0,98333	0,98887	0,99443	1,0000	1,0056
0,93418	0,94503	0,95593	0,96687	0,97236	0,97787	0,98339	0,98891	0,99445	1,0000

ЛИТЕРАТУРА

1. ЦИТИРОВАННАЯ

1. Arley N., Buch K., Introduction to the theory of probability and statistics, New York, 1957 (есть перевод: Н. Арлей, К. Бух, Введение в теорию вероятностей и математическую статистику, ИЛ, 1951).
2. Barnard G., Jenkins G., Winsten C., Journ. Roy. Stat. Soc., A125, 321, 569 (1962).
3. Bartlett M., Phil. Mag., 44, 249 (1953).

3а. Beale E. M. L., Journ. Roy. Stat. Soc., 22B, 41 (1960).

4. Bortkiewicz L., Das Gesetz der kleinen Zahlen, Leipzig, 1898.

4а. Box G. E. P., Lucas H. L., Biometrica, 46, 77 (1959).

5. Fundamental formulas of physics, Ed. by D. Menzel, Chap. 2, New York, 1955 (есть перевод: Д. Мензел, Основные формулы физики, гл. 2, ИЛ, 1957).
6. Cramer H., Mathematical methods of statistics, Princeton, 1954 (есть перевод: Г. Крамер, Математические методы статистики, ИЛ, 1948).
7. Crow E., Davis F., Maxfield M., Statistical manual, Dover, 1960.
8. Davies O., Statistical methods in research and production, New York, 1957.
9. Dixon W., Massey F., Introduction to statistical analysis, New York, 1957.

9а. DuMond J. W. M., Cohen E. R., Rev. Mod. Phys., 25, 691 (1953).

10. Feller W., An introduction to probability theory and its application, Second ed., vol. I, New York, 1957 (есть перевод: В. Феллер, Введение в теорию вероятностей и ее приложения, изд-во «Мир», 1964).
11. Fisher R., Statistical methods and scientific inference, New York, 1956 (есть перевод: Р. Фишер, Статистические методы для исследователей, М., Госстатиздат, 1958).
12. Fisher R., On an absolute criterion for fitting frecuency curves, Messenger of Math., 41.
13. Fisher R., Yates F., Statistical tables for biological, agricultural and medical research, New York, 1957.
14. Forsythe G., Journ. Soc. Ind. Appl. Math., 5, 74 (1957).
15. Gauss K. F., Werke, vol. 4, 1880.
16. Glasser R., Data redaction in physics; NRL 5667, Washington, 1961.
17. Гнеденко Б., Хинчин А., Элементарное введение в теорию вероятностей, изд-во «Наука», 1964.
18. Grenander U., Rosenblatt M., Statistical analysis of stationary time series, New York, 1957.

18а. Hartley H. O., Technometrics, 3, 269 (1961).

19. Jauneau J., Morellet D., Proceeding of the 1964 Easter Shool for Physicists, vol. 1, 1964, CERN.
20. Jeffreys H., Theory of probability, Oxford, 1948.
21. Kendall M. G., Stuart A., The advanced theory of statistics, vol. 1, 2, New York, 1958.

22. Клепиков Н. П., Соколов С. Н., Анализ и планирование экспериментов методом максимума правдоподобия, М., 1964.
23. Колмогоров А., Основные понятия теории вероятностей, М., ОНТИ, 1936.
24. Linder A., Handliche Sammlung Mathematischstatistischer Tafeln, Birkhäuser, 1961.
25. Merriman M., Transaction of the Connecticute Academy of Arts and Science, 4, 1877.
26. Molina E., Poisson's exponential binomial limit, Amsterdam, 1947.
27. Mood A., Introduction to the theory of statistics, New York, 1950.
28. Mood A., Graybill F., Introduction to the theory of statistics, New York, 1963.
29. National Bureau of Standards, «Tables of the binomial probability distribution», Washington, 1952.
30. National Bureau of Standards, «Tables of the bivariate normal distribution function», Washington, 1959.
31. Orear J., Notes on statistics for physicists, Berkeley, 1958 (есть перевод: Д. Орир, Заметки о статистике для физиков, Дубна, Препринт ОИЯИ, 292, 1959).
32. Pearson E., Hartley H., Biometrika tables for statisticians, Cambridge, 1958.
33. Plackett R., Principles of regression analysis, Oxford, 1960.
34. Savage L. et al., The foundation of statistical inference, London, 1961.
35. Stein C., Journ. Roy. Stat. Soc., A125, 565 (1962).
36. van der Waerden B., Mathematische Statistik, Berlin, 1957 (есть перевод: Б. Ван дер Варден, Математическая статистика, ИЛ, 1960).
36а. Watson G. S., Biometrica, 42, 327 (1955).
36б. Watson G. S., Hannan E. J., Biometrica, 43, 436 (1955).
37. Weatherburn C., A first course in mathematical statistics, Cambridge, 1952.
38. Wilks S., Mathematical statistics, Princeton, 1947 (есть перевод: С. Уилкс, Математическая статистика, изд-во «Наука», 1968).
39. Williams E., Regression analysis, New York, 1959.

2. РЕКОМЕНДОВАННАЯ

40. Anderson T., An introduction to multivariate statistical analysis, New York, 1958 (есть перевод: Т. Андерсон, Введение в многомерный статистический анализ, М., Физматгиз, 1963).
41. Aitken A., Statistical mathematics, New York, 1957.
42. Дунин-Барковский И., Смирнов Н., Теория вероятности и математическая статистика в технике, М., Гостехиздат, 1955.
43. Fischer R., The desing of experiments, New York, 1951.

44. Hoel R., Introduction to mathematical statistics, New York, 1954.
45. Lehmann E., Testing statistical hypotheses, New York, 1959 (есть перевод: Э. Леман, Проверка статистических гипотез, М., изд-во «Наука», 1964).
46. Vajda S., The Theory of games and linear programming, London, 1956.
47. Wald A., Sequental analysis, New York, 1951 (есть перевод: А. Вальд, Последовательный анализ, М., Физматгиз, 1960).

3. ДОБАВЛЕННАЯ РЕДАКТОРОМ ПЕРЕВОДА

48. Бернштейн С. Н., Теория вероятностей, изд. 4, М., Гостехиздат, 1946.
49. Breitenberger E., Nuclear physics, 4, 56 (1955).
50. Cziffra P., Moravcsik M., A practical guide to the method of least squares, UCRL 8523 Rev., 1959.
51. Гнеденко Б. В., Курс теории вероятностей, изд. 3, М., Физматгиз, 1959.
52. Гольданский В. Н., Куценко А. В., Подгорецкий М. И., Статистика отсчет в при регистрации ядерных частиц, М., Физматгиз, 1959.
53. Hamilton W., Statistics in physical science, New York, 1964.
54. Johnson N., Leone F., Statistics and experimental desing in engineering and the physical science, vol. 1, 2, New York, 1964.
55. Линник Ю. В., Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений, М., Физматгиз, 1962.
56. Mandel J., The statistical analysis of experimental data, New York, 1964.
57. Смирнов Н. В., Дунин-Барковский И. В., Краткий курс математической статистики для технических приложений, М., Физматгиз, 1959.
58. Scheffe H., The analysis of variance, New York, 1959 (есть перевод: Г. Шеффе, Дисперсионный анализ, М., 1963).
59. Харрис Т., Теория ветвящихся случайных процессов, изд-во «Мир», 1966.
60. Янко Я., Математико-статистические таблицы, М., 1961.
61. Janossy L., Theory and Practice of the evaluation of measurements (есть перевод: Л. Янош, Теория и практика обработки результатов измерений, изд-во «Мир», 1965).
62. Бартлетт М. С., Введение в теорию случайных процессов, ИЛ, 1958.
63. Böck R. K., в книге «Международная школа по физике высоких энергий», Дубна, 1967.
64. Прохоров Ю. В., Розанов Ю. А., Теория вероятностей, изд-во «Наука», 1967.
65. Рао С. Р., Линейные статистические методы и их применение, изд-во «Наука», 1968.
66. Большев Л. Н., Смирнов Н. В., Таблицы математической статистики, М., 1968.

Оглавление

Предисловие редактора перевода к первому изданию . . . 5
Предисловие редактора перевода ко второму изданию . . . 9
Из предисловия автора 11

Глава первая. Введение в теорию вероятностей 13
§ 1. Определение вероятности 14
Комбинаторное определение 15
Частотное определение 16
Современное определение, основанное на теории меры 16
«Субъективное» определение 17
§ 2. Основные законы теории вероятностей 17
Сложение вероятностей 17
Условная вероятность 20
Умножение вероятностей 20
Независимость событий 21
§ 3. Дискретные распределения 22
Биномиальное распределение ($\nu = 2$) 22
Предельные формы биномиального распределения . . 23
Распределение Пуассона ($\nu = \aleph_0$) 23
§ 4. Непрерывные распределения 27
Парадокс нулевой вероятности 27
Равномерное (прямоугольное) распределение 28
Нормальное распределение 30
§ 5. Математическое ожидание и дисперсия случайной величины 32
Биномиальное распределение 33
Распределение Пуассона 35
Равномерное распределение 36
Нормальное распределение 37
Распределение Коши 40
§ 6. Моменты случайной величины. Производящие функции моментов 42
Производящие функции моментов (ПФМ) 43
Биномиальное распределение 46
Распределение Пуассона 46
Нормальное распределение 46
Распределение $y = ax + b$, где a и b — постоянные величины 47
Гамма-распределение 47
Глава вторая. Совместные распределения вероятностей 49
§ 7. Функции распределения вероятностей двух или нескольких случайных величин 49
Сумма случайных величин 50
Выборочное среднее 52
§ 8. Закон больших чисел 52
Выборочное среднее 54
Центральная предельная теорема 55
Использование таблиц нормального распределения для вычисления биномиального распределения . . . 57
Пример, относящийся к распределению Коши . . . 58
§ 9. Распределение вероятности для функции случайной величины 59

Перенос ошибок 60
Распределение вероятности для функции дискретной величины 62
Распределение вероятности для функции непрерывной величины 64
Распределение χ^2 67
Случай неединичной дисперсии 69
Преобразование нескольких случайных величин . . . 70
Сумма квадратов отклонений от среднего 71
Выборочная дисперсия 72

Глава третья. Проверка гипотез 74

§ 10. Критерий согласия χ^2 74
Число степеней свободы 76
§ 11. Распределение t Стьюдента и его применения . . . 80
Различие между двумя выборочными средними . . . 83
Доверительный интервал 88
§ 12. Анализ сделанных предположений 90
Использование вероятностной бумаги для проверки распределения внутри выборки 92
Линеаризация кривой 93
§ 13. Распределение $\mathcal{F}$ Фишера 96
Применение критерия $\mathcal{F}$ к решению задачи о проведении кривой по точкам 100

Глава четвертая. Принцип максимального правдоподобия 102

§ 14. Функция правдоподобия 102
Биномиальное распределение 102
Непрерывный параметр 104
Параметры непрерывного распределения 105
Достаточные статистики 109
§ 15. Графический анализ функции правдоподобия 111
Случай дискретного параметра 112
Случай непрерывного параметра 113
Принцип правдоподобия 116
Свойства оценок максимального правдоподобия в случаях выборки малого и большого объема . . . 119
Пример использования функции правдоподобия . . . 123
Случайная выборка малого объема 124
Эффективность оценки $\hat{\sigma}^2$ 127
Случайная выборка большого объема 127
Предположение относительно нормальности распределения 130
§ 16. Двумерная функция правдоподобия 130
§ 17. Теорема Бейеса 138
Применение априорной вероятности 141
Случай, когда априорная вероятность неизвестна . . . 142
Библиография 145

Глава пятая. Метод наименьших квадратов 146

§ 18. История развития метода 146
§ 19. Анализ регрессий 147
Введение 147
Аппроксимация полиномом 149
Остаточная сумма квадратов 152
Не зависящие от вида распределения свойства оценок наименьших квадратов 153
Случай исходных данных с неодинаковыми дисперсиями 156
§ 20. Ортогональные полиномы 157
Рекуррентное соотношение Форсайта 161
Увеличение степени полинома 164
§ 21. Нормальный регрессионный анализ 166
Мера предосторожности 168
Библиография 171

Приложение I. Математическое ожидание остаточной суммы квадратов 173

Приложение II. Получение выборки случайных величин с заданной плотностью вероятности . . . 175

ДОПОЛНЕНИЯ 177

I. Дж. Малви. Статистические методы обработки экспериментальных данных 177

1. Введение 177
§ 1. Предварительные замечания 177
§ 2. Определения и обозначения 177
2. Определения характеристик выборки 180
§ 1. Меры положения 180
§ 2. Меры рассеяния 180
§ 3. Коэффициент корреляции 181
3. Оценки параметров генеральной совокупности по характеристикам выборки 181
4. Специальные распределения 185
§ 1. Нормальное (гауссово) распределение 185
§ 2. Распределение χ^2 186
§ 3. Распределение t (Стьюдента) 187
§ 4. Равномерное (прямоугольное) распределение 188
§ 5. Биномиальное распределение 188
§ 6. Распределение Пуассона 189
5. Доверительные пределы 190
§ 1. Выборки из нормального распределения . . . 190
§ 2. Большие выборки и приближенно нормальные оценки 190
6. Метод максимального правдоподобия 191
§ 1. Оценка максимального правдоподобия 191
§ 2. Свойства оценок максимального правдоподобия; среднеквадратичное отклонение . 192

§ 3. S-Функция Бартлетта 194
7. Перенос ошибок 195
§ 1. Матрица ошибок (ковариационная матрица) . . 195
§ 2. Линейные функции 195
§ 3. Нелинейные функции 196
§ 4. Среднеквадратичное отклонение отношения двух величин 197
§ 5. Среднеквадратичное отклонение произведения двух величин 197
§ 6. Общая формула переноса ошибок для независимых переменных 198
8. Метод наименьших квадратов. Аппроксимация полиномом 198
§ 1. Постановка задачи 198
§ 2. Линейные функции 200
§ 3. Аппроксимация прямой линией 200
§ 4. Обобщенный метод наименьших квадратов . . 201
9. Проверка гипотез 201
§ 1. Общие замечания 201
§ 2. Критерий χ^2 203
§ 3. Оценка параметров с помощью критерия χ^2 204
§ 4. Проверка положения среднего значения нормального распределения 205

II. У. К. Гамильтон. Метод наименьших квадратов и проверка линейных гипотез 206

§ 1. Метод наименьших квадратов 206
§ 2. Геометрическая интерпретация матрицы ошибок 216
§ 3. Многомерное нормальное распределение и свойства распределений квадратичных форм 218
§ 4. Распределение квадратичных форм в методе наименьших квадратов 220
§ 5. Метод наименьших квадратов при наличии линейной связи параметров 222
§ 6. Многомерные линейные гипотезы 224
§ 7. Мощность критерия $\mathcal{F}$ для многомерной линейной гипотезы 228
§ 8. Выбор весов в методе наименьших квадратов 229
§ 9. Перенос ошибок 232
§ 10. Метод наименьших квадратов в случае нелинейной зависимости от параметров 233
§ 11. Задача планирования эксперимента 244

СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ 253

Литература 290

1. Цитированная 290
2. Рекомендованная 291
3. Добавленная редактором перевода 292

УДК 530:519.2

Перевод
В. Ф. ГРУШИНА

Под редакцией
Е. М. ЛЕЙКИНА

Редакция литературы по физике

Инд. 2-3-2
57-70

Д. Худсон

СТАТИСТИКА ДЛЯ ФИЗИКОВ

Редакторы *Л. В. Гессен и Е. С. Куранский*
Художник *А. Б. Шкловская*
Художественный редактор *П. Ф. Некунда*
Технический редактор *Н. Д. Толстякова*
Корректор *К. Л. Водяницкая*

Сдано в производство 11/V 1970 г. Подписано к печати 24/VIII 1970 г.
Бумага матт. мел. $84 \times 108^1/_{32} = 4{,}63$ бум. л. Усл. печ. л. 15,54.
Уч.-изд. л. 13,01. Изд. № 2/5422. Цена 1 р. 14 к. Зак. 254

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1-й Рижский пер., 2

Московская типография № 16 Главполиграфпрома
Комитета по печати при Совете Министров СССР
Москва, Трехпрудный пер., 9