

М. А. Великанов

ОШИБКИ ИЗМЕРЕНИЯ
И
ЭМПИРИЧЕСКИЕ ЗАВИСИМОСТИ



ГИДРОМЕТЕОРОЛОГИЧЕСКОЕ ИЗДАТЕЛЬСТВО

ЛЕНИНГРАД • 1962

АННОТАЦИЯ

В цикле наук метрического естествознания особое место занимают те науки, многие вопросы в которых решаются еще на основании эмпирических зависимостей, не сводимых к основным и непреложным физическим законам, а основанных на одних лишь измерениях. Возникающие в связи с этим неизбежные, как и для всякого измерения, ошибки сказываются не только на значениях параметров, но и, что гораздо важнее, на самой форме отыскиваемых зависимостей.

Неполнота информации, характерная вообще для неточных наук, придает нашему знанию об изучаемых явлениях вероятностный характер, распространяющийся и на возможности управления явлением и на надежность прогноза.

Предлагаемая книга построена на синтезе двух процессов: измерения величин и отыскания зависимостей между ними. Она предназначена для наблюдателей, экспериментаторов и прогнозистов в таких областях науки, как гидрология, метеорология, гидравлика, топография, сопротивление материалов, механика грунтов и др.

Книга рассчитана на читателя, знакомого с математикой лишь в объеме технических вузов, поэтому в ней приводится необходимый минимум знаний по теории вероятностей и математической статистике.

ПРЕДИСЛОВИЕ

Техника и методика измерений как в лабораторном эксперименте, так и при наблюдениях в природе за последние десятилетия в некоторых физических науках далеко продвинулись вперед. Для примера достаточно привести экспериментальные измерения по гидродинамике, при проведении которых в первой четверти настоящего века в основном применялась трубка Пито, гидрометрическая вертушка и поплавки, теперь же используются, кроме того, электротермоизмерители, высокочастотные кинематографы, а в самое последнее время началось освоение изотопов и ультразвука. То же наблюдается и во многих других отраслях физики и механики, но, конечно, далеко еще не во всех.

В таких условиях особенно ощущается как в нашей стране, так и за рубежом недостаток литературы, содержащей теорию ошибок, обработку измерений, способ наименьших квадратов, а также другие области математической статистики, касающиеся теории измерений. И даже в этом, сравнительно небольшом количестве литературы ярко выражен разрыв между углубленным критическим анализом ряда весьма ценных монографий [1, 3, 6, 8, 13, 15] и примитивностью изложения некоторых общедоступных курсов [12, 17, 18]. Приведенные первые литературные источники ввиду их математической сложности малодоступны экспериментаторам и наблюдателям обычного профиля. Когда же экспериментаторы используют литературу, в которой дано только примитивное освещение того или иного вопроса, то они рискуют получить одностороннее представление об основах теории ошибок.

В данной книге автор имеет в виду те науки, которые также принадлежат к физико-математическому циклу наук, но занимают особое место как в связи с преобладанием в них приближенных эмпирических зависимостей, получаемых из специально поставленных экспериментов, так и в связи с пониженной точностью тех измерительных приборов и методов, которые по многим причинам применяются еще в настоящее время. В этих науках экспериментатору зачастую приходится оперировать такими измерениями, в которых вероятная ошибка имеет тот же порядок, что и размер самой измеряемой величины, вследствие чего нарушается та самая симметричность распределения отклонений, на которой в значительной степени основана вся классическая теория ошибок, построенная на постулатах Гаусса.

С другой стороны, именно для этих «малоточных» наук наряду с использованием строгих физических законов специфичным является отыскание приближенных эмпирических зависимостей, в которых суммируются ошибки самих измерений и ошибки примененной формы зависимостей, расчленив которые возможно далеко не всегда.

Именно в «неточных» науках чаще всего и возникает необходимость в развитии теории ошибок, причем такой теории, которая рассматривала бы параллельно вопросы измерения величин и вопросы отыскания зависимостей между измеренными величинами, как две стороны одной и той же задачи.

Однако поставленная автором цель помочь лабораторным экспериментаторам и работникам наблюдательных сетей уяснить сущность совмещенной задачи побуждает уделить достаточно внимания также чисто теоретическому обоснованию многих вопросов, отнюдь не ограничиваясь практическими рекомендациями, как поступать в том или ином случае.

Классическая теория ошибок исходит из принятия непреложности и универсальности следующих трех теснейшим образом друг с другом связанных представлений и положений: [среднее арифметическое] → [нормальный закон распределения] → [способ наименьших квадратов].

Обычно принимается, по существу, без доказательств, что среднее арифметическое всех измерений дает «наивероятнейшее» значение измеряемой величины (постулат Гаусса), что вероятности отдельных ошибок зависят только от их величин (второй

постулат Гаусса) и, наконец, что наилучшее приближение к искомой зависимости дает та форма зависимости, при которой сумма квадратов отклонений является минимумом (принцип Лежандра, также развитый Гауссом).

Эти три одинаково недосказанных и недоказуемых положения практически почти всегда оправдываются, хотя это почти, подчеркивающее их ограниченность, побуждает излагать их в критическом разрезе. Тогда оказывается, что медиана имеет такое же, а иногда и большее право считаться центром распределения, как и среднее арифметическое. Мало того, оказывается, что постулат Гаусса о наибольшей вероятности среднего арифметического оправдывается лишь тогда, когда среднее совпадает с медианой, так как нормальное распределение может иметь силу лишь при достаточной малости отклонений по сравнению с измеряемой величиной. И, наконец, известно, что равноправный принципу Лежандра принцип Лапласа, ставящий на место суммы квадратов отклонений суммы их модулей, если и не получил большого развития, то только потому, что математические операции, связанные с ним, оказались более громоздкими. Это чисто практическое соображение, конечно, отнюдь не подрывает доверия к принципу Лапласа в теоретическом смысле.

Все изложенные соображения привели автора книги к необходимости, во-первых, совместить в одном аспекте как анализ ошибок, так и анализ приемов отыскания приближенных зависимостей, во-вторых, не посягая на приемы классической теории ошибок, связанной с упомянутой триадой недоказуемых положений, дать одновременно читателю возможность выйти из рамок классической теории, в частности, там, где постулируемая ею симметричность распределения отклонений явно не имеет места.

Автор надеется, что совмещенное изложение послужит на пользу работникам тех наук, в которых широко используются для научных выводов как эксперимент, так и наблюдения в природе.

Автор надеется также, что то критическое изложение основ теории ошибок, которое он проводит по всей книге, побудит экспериментаторов и наблюдателей природы не приписывать этим «основам» не присущую им «универсальность», а принимать

их с известными оговорками, иногда имеющими существенное значение.

С другой стороны, автор отнюдь не считает, что теоретическая углубленность неразрывно связана со строгостью и уточненностью математических доказательств. Существует ряд положений, имеющих глубокий смысл, как, например, закон больших чисел или «предельная теорема», и в то же время воспринимающихся читателем интуитивно как самоочевидные, а поэтому нет необходимости тратить силы и время читателя на изучение их строгих доказательств.

Например, известно, что изложение способа наименьших квадратов на матричной основе получается более стройным, тем не менее нет надобности утруждать читателя изучением теории матриц, если способ наименьших квадратов ему и без этого ясен.

Автор полагает, что максимальным упрощением математического аппарата можно достичь основной цели — дать экспериментаторам и наблюдателям природы со вузовскими, а не механико-математическими знаниями возможность хорошо разбираться в качестве тех измерений, на основе которых они делают свои, зачастую практически важные выводы.

Один из крупнейших физико-математиков XIX в. Анри Пуанкаре шутил сказал именно по поводу некоторых положений классической теории ошибок, что экспериментаторы считают их строго доказанными математиками, а математики считают их оправданными экспериментом.

Автор надеется, что читатели его книги не будут думать ни того, ни другого.

ВВЕДЕНИЕ

§ 1. Физические законы и эмпирические зависимости

Науки, принадлежащие к циклу метрического естествознания, т. е. применяющие количественные оценки своих объектов и отыскивающие количественные связи между ними, могут быть расположены в ряд по возрастанию в них степени участия строгих физических законов, образующих систему взаимосвязанных функциональных зависимостей. На первой ступени этого ряда находятся такие науки, как общая механика и некоторые разделы физики. В этих науках эксперимент используется главным образом в целях уточнения значений параметров, входящих в теоретические, функциональные зависимости.

На следующих ступенях ряда наук размещаются такие науки, как, например, гидродинамика и теория упругости, где наряду со строгими теоретическими зависимостями используются в качестве дополнения к наукам, стоящим на первой ступени ряда, а иногда как основные те эмпирические зависимости, которые получаются на основе специально поставленных экспериментов или наблюдений в природе.

Далее в ряду наук расположены науки, все более и более насыщенные экспериментальным материалом и использующие его уже в качестве основного орудия отыскания зависимостей между явлениями.

И, наконец, на последней ступени рассматриваемого ряда наук находятся те науки, которые сравнительно недавно поднялись над уровнем качественного или описательного естествознания и примкнули к метрическому естествознанию. К таким наукам можно отнести гидрологию суши, океанологию и особые разделы метеорологии, а также частично некоторые из приближенных технических наук: гидравлику, сопротивление материалов, механику грунтов и др. Последние отчасти дублируют некоторые более точные науки. Например, гидравлика дублирует

гидродинамику, сопротивление материалов — теорию упругости и т. д. Такое дублирование необходимо в тех вопросах и областях, в которых строгие решения еще не найдены, а на практике требуются решения хотя бы приближенные.

Поскольку содержание предлагаемой книги направлено на разрешение вопросов, возникающих именно в этом «нестрогом» звене метрического естествознания и соответствующих разделов техники, то полезно с самого начала охарактеризовать специфику этого звена.

Основным орудием познавательного процесса в науках этого звена является эксперимент и наблюдение. иначе говоря, измерения как в искусственно создаваемых лабораториях, так и в естественных условиях.

Известно, что каждое измерение связано с ошибками, так как безошибочных измерений вообще не может быть; поэтому проблема ошибок измерения займет в книге основное место. Постараемся осветить в проблеме ошибок измерения пока лишь один чрезвычайно важный вопрос, относящийся к природе ошибок. Возможны случаи, когда мы отыскиваем связь между двумя (или более) факторами и при этом знаем заранее из чисто теоретических соображений, что, кроме тех факторов, которые мы измеряем, никаких других, также влияющих на рассматриваемое явление, не существует.

Тогда неизбежные и в этом случае ошибки измерения будут зависеть лишь от качества прибора и качества примененной методики. Но возможен и иной случай, когда все факторы, влияющие на изучаемые явления, неизвестны, а известны лишь основные из них, влияние же всех прочих факторов мы оценить не в состоянии. Тогда независимо от чисто технических ошибок измерения мы будем иметь дело и с некоторой принципиальной ошибочностью производимого нами измерения. Если в первом из двух рассматриваемых случаев получаем в результате хотя и приближенную, но все же функциональную зависимость между измеряемыми величинами, то во втором случае получаемая связь будет не функциональной, а стохастической. Следовательно, определенному значению одной переменной соответствует целая совокупность значений другой переменной, распределенных по степени их вероятности.

Подчеркнем, что в обоих случаях отклонения измеренных значений от неизвестного нам, по существу, «истинного» значения всегда подчиняются какому-то вероятностному закону. Однако генезис этих отклонений, а следовательно, и закон их распределения будут в них существенно различны. Совершенно ясно, что при измерениях в точных науках существуют лишь отклонения первой категории, зависящие лишь от качества (точности) измерительного прибора. В неточных же науках мы часто встречаемся и с ошибками второй категории. Это особенно

важно отметить, так как именно в этих науках данные измерений служат основой для получения каких-то новых, теоретически невыводимых зависимостей. Задача эксперимента становится здесь шире, чем в точных науках. Именно эксперимент должен дать нам в этом случае основу для отыскания наилучшей формы еще не известной зависимости.

Попутно отметим ошибочность одного взгляда некоторых теоретиков, далеко стоящих от эксперимента и склонных видеть в нем лишь регистрацию наблюдаемых и измеряемых факторов изучаемого явления. Это далеко не так: именно на результатах эксперимента и строится дальнейший анализ, из которого мы черпаем сначала лишь качественные, а затем и количественные характеристики явлений природы. Построение затем на экспериментальном материале системы физических законов возможно потому, что в самом эксперименте, в его постановке элементы анализа в скрытом виде уже присутствуют.

Указанный Лениным путь науки — от конкретного к абстрактному и от абстрактного снова к конкретному — выражает связь между экспериментом и анализом в предельно ясной форме.

И нужно сказать, что в тех науках, которые еще недостаточно точны, ленинский принцип следует особенно строго соблюдать. Именно здесь мы проходим все стадии познавательного процесса, начиная от простого наблюдения и качественного анализа структуры изучаемого явления, через построение рабочих гипотез, служащих основой для постановки эксперимента, далее через обработку проведенных измерений к проверке, положенной в основу эксперимента рабочей гипотезы.

И, наконец, после многократного чередования конкретного и абстрактного приходим к такому овладению предметом, которое уже дает нам возможность управлять им с практической целью.

Таким образом, имеются четыре стадии познавательного процесса.

Первая стадия — качественное наблюдение, устанавливающее основные черты изучаемого явления и общий характер действующих на него факторов. На базе качественного анализа возможна рациональная постановка и проведение измерений, т. е. вторая стадия познавательного процесса.

Третья стадия охватывает все процессы, относящиеся к обработке измерений, при помощи которых можно проверить принятую рабочую гипотезу или установить новую систему связей между элементами явления.

И, наконец, наступает четвертая стадия — творческая работа по созданию тех искусственных условий, в которых изучаемое явление природы может быть использовано для практических целей.

Попутно упомянем, что в процессе использования природного явления мы почти всегда создаем для него такие условия, каких

в природе не бывает и в которых те же факторы проявляются иначе. Иногда изменение обстановки настолько велико, что приходится ставить дополнительные измерения в новых условиях, применяя, например, моделирование последних.

В виде примера достаточно взять изучение режима реки, которую намечено использовать для получения электрической энергии. Изучение режима реки начинается с большого комплекса измерений уровней, расходов, скоростей, уклонов и др. Но так как в дальнейшем реку преградит плотина и в ней будет создано водохранилище, в котором большая часть ранее измеренных факторов поведет себя иначе, чем в прежних естественных условиях, то для прогноза или расчета будущего искаженного режима приходится применять моделирование условий, в которых будет протекать эта река.

Далее, уже после постройки плотины иногда бывает интересно повторить в новых условиях, не встречающихся в самой природе, весь комплекс прежних измерений. В результате этого получают новые зависимости, дополняющие ранее определенные. Следовательно, можно сказать, что практика не только проверяет, но уточняет и расширяет теорию.

В связи с этим полезно еще подчеркнуть, что первое «конкретное» в схеме Ленина, т. е. то непосредственное созерцание природного явления, с которого начинается любое исследование, обладает всем богатством непосредственных образов, которых мы, переходя к последующей абстракции, частично лишаемся, так как всякой абстракции свойственна обедняющая ее схематичность.

Каждый исследователь природы понимает, что эта «обедняющая» схематизация в качестве переходного звена к более высокому уровню познания явления совершенно необходима, так как без нее нельзя выявить те закономерности, которые существуют как внутри самого явления, так и между явлением и окружающей его обстановкой.

В дальнейшем первичная схема познаний явлений постепенно обогащается новыми данными, получаемыми путем «возвращения от абстрактного снова к конкретному». Теоретическое объяснение явления становится все более и более полным, так как все большее и большее число факторов, первоначально принятых второстепенными, а затем отброшенных, вводится дополнительно в рассмотрение путем установления ряда новых эмпирических зависимостей.

Таким образом, эмпирическая зависимость, являющаяся на низком уровне знания лишь плохим заменителем физического закона, на более высоком уровне обогащает этот же закон учетом второстепенных факторов, которые в первоначальной форме закона не принимались во внимание.

На поверхности явлений эмпирические зависимости высту-

пают как проявление законов, выявленных в чистом виде анализом и абстракцией. Это общее и совершенно правильное положение как нельзя лучше освещает процесс обогащения абстрактного физического закона наложением на него совокупности эмпирических зависимостей.

Например, решения строгих дифференциальных уравнений гидродинамики далеко не полностью отражают структуру речного потока, но в соединении с эмпирически установленными свойствами турбулентности потока и с эмпирическими формулами размыва дна, берегов и движения наносов они дадут возможность экспериментаторам ближе подойти к истинной картине явления.

Приведем другой пример. Для учета испарения с поверхности естественных водоемов одного теоретически установленного закона Дальтона, очевидно, недостаточно. Но в соединении с эмпирическими исследованиями влияния ветра на испарение этот же закон становится более полным. Далее мы узнаем, что на испарение влияет, в сущности, не скорость ветра, а то турбулентное перемешивание, которое с этой скоростью связано. И перед нами открывается перспектива в установлении нового теоретического закона, в котором будут фигурировать дефицит влажности и турбулентность.

Таким образом, путем наложения на физические законы дополняющих эмпирических зависимостей и путем постепенного уточнения последних и преобразования их в новые физические законы менее строгие естественные науки постепенно будут становиться все более и более строгими.

Однако указанный процесс очень длителен. И так как жизнь не ждет и требует от науки хотя бы приближенных ответов там, где точные ответы еще заставляют себя ждать, то отыскание приближенных зависимостей еще долго будет сохранять свою актуальность.

§ 2. Измеримость величин

Отыскиваемые исследователем природы зависимости между факторами, или элементами, явления могут быть и чисто качественными, по крайней мере на низком уровне соответственной науки. Но на более высоком ее уровне, когда эти факторы, или элементы, явления могут рассматриваться как величины, т. е. как нечто измеримое, искомые зависимости могут и должны принимать количественный характер. И здесь независимо от установленной выше и совершенно бесспорной ошибочности всякого измерения появляется еще до нее вопрос о степени измеримости той или иной величины. Начнем с простого примера измерения длины какого-то предмета.

Если, например, измеряем тонкий металлический стержень,

микрометренным циркулем, то весь процесс измерения прост. Все ошибки обусловлены лишь качеством циркуля и внимательностью измеряющего лица.

Но если, например, измеряем срубленный ствол дерева стальной лентой, то будем ли мы вправе считать, что имеем какую-то определенную длину, при измерении которой допускаем лишь ту ошибку, которая связана с качеством измерителя, и не будет ли для нас слишком очевидно, что длины, измеряемые между различными точками двух крайних сечений ствола, могут различаться даже в большей степени, чем два измерения, проведенные между одними и теми же точками? Иными словами, надо сначала наметить эти точки, чтобы измерять расстояние между ними. Но если имеется в виду «длина ствола», а не длина между отдельными его точками, то надо сначала условиться, какую длину, между какими точками будем называть «длиной ствола». Если же, что весьма вероятно, такой выбор точек будет произвольным, то не окажется ли, что разброс измерений между различными точками превосходит разброс измерений, свойственный данному измерителю?

Приведем другой пример. Что называть шириной реки? Для прямолинейного канала призматического сечения ответ на этот вопрос прост и ясен. Но река всегда изгибается, непрерывно меняя и свою ширину, и свою кривизну. Предположив для простоты, что имеется точный план данного участка реки, ставится вопрос, как провести на этом плане прямую линию и какой отрезок этой прямой линии можно назвать «шириной реки»?

Допустим, что мы условились проводить линию створа везде нормально к динамической оси потока, и отрезок этой линии, отсекаемый двумя урезами, назовем шириной реки. Но разве динамическая ось потока может быть выбрана сколько-нибудь определенно при весьма неопределенном распределении скоростей в сечении потока и при столь же неопределенном переходе от скоростей в одном сечении к скоростям в другом?

В данном случае видим также, что разброс значений ширины при различном выборе динамической оси может значительно превзойти тот разброс, который свойственен мензульному измерению определенной длины.

Возможны и другие причины неопределенности измеряемой величины. Например, расплывчатость контура дождевой тучи придает измерению ее площади весьма условный характер, а между тем только знание размеров этой площади позволит, хотя бы грубо приближенно, перейти от показаний дождемера к подсчету объема осадков, выпавших на данной территории при таком-то ливне.

Не приводя большого количества аналогичных примеров, перейдем непосредственно к выводу из них.

Каждому конкретному измерению должно предшествовать

выяснение степени измеримости данной величины. Без такого предварительного качественного анализа трудно даже установить, о какой желательной точности может идти речь в каждом конкретном случае: нужно ли стремиться к большей точности или, наоборот, достаточно ограничиться меньшей точностью.

В данной книге вопрос об ошибках измерения занимает основное место, а поэтому имеется основание с самого начала связать ошибки измерения с вопросами качественного анализа, предшествующего всякому измерению.

ПРОСТЕЙШИЕ ВЕРОЯТНОСТНЫЕ ПРЕДСТАВЛЕНИЯ

§ 3. Частость и вероятность

Если какое-либо явление в природе или в искусственных условиях занимает относительно малый промежуток времени и о нем можно сказать, что оно «наступило», «проходит» или «прошло», то его можно назвать событием. Если известно, что в определенных и нами полностью учитываемых условиях это событие обязательно произойдет, то оно принадлежит к разряду достоверных. Если же, наоборот, известно, что при тех же условиях рассматриваемое событие не может произойти, то оно считается невозможным.

Между двумя крайними пределами — достоверности и невозможности — располагаются все однородные события, которые в одинаковых условиях могут произойти, а могут и не произойти. Отсюда вытекает возможность, приняв ожидаемость достоверного события за единицу, а невозможного — за нуль, охарактеризовать ожидаемость всех остальных правильными дробями

$$0 \leq p \leq 1.$$

Вся человеческая деятельность как в повседневной жизни, так и в производстве, технике и в исследованиях природы связана в каждый данный момент с ожиданием какого-то события или комплекса событий, способствующих или препятствующих совершению намеченного действия. Во многих случаях этот комплекс событий сам зависит от столь большого числа факторов, учет которых труден или даже невозможен, что предвидеть его мы не в состоянии, а можем лишь оценивать его ожидаемость и действовать в зависимости от этой ожидаемости, или, как говорят, в зависимости от его вероятности.

В некоторых сравнительно редких случаях эту вероятность мы можем непосредственно вычислить из известного нам отношения числа благоприятных случаев к общему числу равновоз-

возможных случаев, например, при бросании монеты вероятность того, что выпадет «орел», равна, очевидно, половине. А при бросании кости в игре «домино» вероятность выпадения любого числа очков от единицы до шестерки равна одной шестой.

Но в большинстве случаев теоретическая или «априорная» вероятность неизвестна, а оценивается эмпирически или «апостериорно» из наблюдаемой частоты появления благоприятных случаев. Отношение частоты к общему числу испытаний называется частотой.

Выясним соотношение между частотой, т. е. эмпирической вероятностью, и вероятностью априорной, теоретической.

В связи с этим приведем случай, когда отношение исходов неизвестно. Например, нам известно, что в урне заключены белые и черные шары, но в какой пропорции, мы не знаем.

Вынимая последовательно по одному шару и возвращая вынутый шар обратно, заметим его цвет. Предположим, что из n вынутых шаров m оказались белыми. Будем продолжать эту операцию сколь угодно долго. Опыт и интуиция показывают, что отношение $\frac{m}{n}$ при возрастании n стремится к постоянному числу

$$\lim \frac{m}{n} = p. \quad (1)$$

Однако указанное постоянство предела для конкретных частот имеет место лишь при условии однородности испытаний, иначе говоря, равновероятности для каждого из шаров, лежащих в урне, быть вынутым при любом испытании.

Равенство (1) приводит для двух частных случаев к следующим выводам:

1. Если в урне все шары белые, то отношение $\frac{m}{n}$ будет равно единице

2. Если в урне нет ни одного белого шара, то при всех испытаниях отношение $\frac{m}{n}$ будет равно нулю. Этим и подтверждается, что вероятность любого события удовлетворяет уже приведенному выше неравенству

$$0 \leq p \leq 1.$$

Перейдем к рассмотрению вероятностей двух (и более) событий A и B . Будем обозначать символом B/A условную вероятность события B , когда известно, что событие A произошло.

Возможны следующие соотношения:

1) $B/A = B$, когда событие A не влияет на появление события B . Если в этом случае имеет силу и обратное положение, $A/B = A$, то мы имеем независимые события.

2) Если имеем $B/A > B$, то событие A способствует появлению

события B . Если, наоборот, имеем $B/A < B$, то событие A затрудняет появление события B .

3) Если условная вероятность B/A равна нулю, то события B и A несовместны.

Если событие A заключается в наступлении хотя бы одного из попарно несовместимых событий B_1, B_2, \dots, B_n , то последние являются частными случаями события A . Если при этом хотя бы одно из событий B_j обязательно должно произойти, то вся система $B_1 - B_n$ является полной и вероятность появления любого из событий равна единице.

Теперь мы можем дать следующее более точное и более общее определение вероятности: вероятность события A есть отношение числа m всех благоприятствующих этому возможным исходов к общему числу равновозможных и несовместных исходов.

§ 4. Сложение и умножение вероятностей

Пусть событие A включает в себя наступление любого из m независимых и несовместных событий B_1, B_2, \dots, B_m . Какова вероятность этого сложного события?

Если каждому из событий B_j соответствует n_j благоприятных из n равновозможных исходов, то будем иметь

$$P(B_j) = \frac{n_j}{n}.$$

Поскольку все события B_j несовместны, то исходы, благоприятствующие одному из них, не могут благоприятствовать ни одному из прочих событий B_k ($k \neq j$).

Следовательно, число исходов, благоприятствующих всем событиям системы B_j , будет равно сумме исходов, благоприятствующих каждому в отдельности:

$$P(\Sigma B_j) = \sum \frac{n_j}{n} = \Sigma P(B_j). \quad (2)$$

В этом и заключается теорема сложения вероятностей независимых событий.

Поставим другую задачу — какова вероятность события A , заключающегося в одновременном наступлении всех событий системы C_1, C_2, \dots, C_k ? Рассмотрим сначала частный случай, когда все события системы независимы друг от друга, т. е. когда наступление любого из них не влияет на последующее наступление какого-либо другого. Совершенно очевидно, что допущение несовместности, принятое выше, при доказательстве теоремы сложения не имеет смысла, так как вероятность появления двух несовместных событий всегда равна нулю.

Примем, что каждому из событий C_j соответствует число благоприятных исходов n_j при полном общем для всех событий

числе исходов n . Вероятность события C_j будет, очевидно, равна $\frac{n_j}{n}$. Но если каждому благоприятному исходу события C_j поставим каждое из благоприятных исходов события C_k , равное n_k , при том же полном числе исходов, то мы, очевидно, получим произведение

$$P(C_j, C_k) = \frac{n_j}{n} \frac{n_k}{n} = P(C_j)P(C_k).$$

Этот вывод легко распространяется и на любое число независимых событий:

$$P(A) = \prod P(C_j). \quad (3)$$

Откажемся теперь от обоих ограничений — несовместности и независимости — и решим задачи сложения и умножения в более общем виде.

Как определить вероятность того, что произойдет одно из ряда событий $B_1, B_2, \dots, B_j, \dots, B_n$, когда мы знаем, что некоторые из этих событий попарно совместны? В таком случае мы должны вычесть из суммы вероятностей, соответствующих несовместным событиям, сумму вероятностей из совместного появления:

$$P(\Sigma B_j) = \Sigma P(B_j) - \Sigma P(B_j B_k). \quad (4)$$

Вычитаемое относится здесь, очевидно, ко всем парам совместных событий, причем если события B_j и B_k независимы, то по предыдущему

$$P(B_j, B_k) = P(B_j)P(B_k).$$

Если же они зависимы, т. е. появление одного влияет на появление другого, то, используя вышеприведенное выражение для условной вероятности, можем записать равенство

$$P(B_j, B_k) = P(B_j | B_k)P(B_k).$$

Поясним сказанное простейшим примером. Какова вероятность, что карта, случайно взятая из хорошо перетасованной колоды (52 карты), окажется или пиковой масти, или одним из тузов? Отдельно взятая вероятность вытянуть пиковую карту равна, очевидно, $\frac{1}{4}$, а отдельно взятая вероятность вынуть один

из тузов равна $\frac{1}{13}$. Но при их суммировании пиковый туз вошел бы в эту сумму дважды, что привело бы к ошибке, поэтому отдельную вероятность этого события, равную $\frac{1}{52}$, надо вычесть

из суммы. В результате получаем $\frac{1}{4} + \frac{1}{13} - \frac{1}{52} = \frac{4}{13}$.

Вышеприведенное выражение для вероятности последовательного появления двух зависимых событий можно расширить на любое их число. Представим себе ряд зависимых событий $A_1, A_2, \dots, A_j, \dots, A_n$, расположенных во временном порядке их возможного появления, и расширим понятие условной вероятности, установленное нами выше для двух лишь событий, на произвольное число.

Выражение

$$P\{A_k | A_1, A_2, \dots, A_{k-1}\}$$

означает вероятности появления события A_k , если стало известно, что все события от A до A_{k-1} , от которых оно зависит, уже наступили. Тогда имеем очевидно

$$P\{A_1, A_2, \dots, A_k\} = P(A_1)P(A_2|A_1), \dots, \dots, P(A_k|A_1, A_2, \dots, A_{k-1}). \quad (5)$$

Если для любого i -того события из этого ряда имеет место равенство

$$P(A_i | A_1, A_2, \dots, A_{i-1}) = P(A_i),$$

то все события этого ряда называются независимыми по совокупности. В этом и только в этом случае сохраняет силу правило (3).

§ 5. Закон больших чисел

В начале главы было введено понятие вероятности и установлено его отношение к эмпирической частоте при очень большом числе испытаний. Справедливость этого отношения подтверждается эмпирически. Вынимая последовательно шары из урны и возвращая их каждый раз обратно в урну, мы всегда получим, что отношение числа вынутых белых шаров к общему числу вынутых шаров будет тем ближе к отношению находящихся в урне белых шаров к общему их числу, чем больше число испытаний. И если мы по какой-либо причине не в состоянии непосредственно пересчитать шары в самой урне, то можем сделать это косвенно, путем многократных выниманий. Интуиция, основанная на обобщении опытов, показывает, что «иначе и быть не может». Этой интуицией постоянно руководствуются в повседневной жизни. Однако в научном исследовании нельзя полностью доверяться интуиции, а всегда надо стремиться подтвердить ее теоретическим анализом. Так возник в теории вероятностей закон больших чисел, содержащий в себе комплекс теорем таких авторов, как Бернулли, Пуассон, Чебышев, Марков и др.

Подробный анализ комплекса всех теорем с их доказательством представляет несомненный формально математический интерес. Однако профиль данной книги, а также ее объем дают

нам право не излагать чисто математические построения в тех случаях, когда результат их интуитивно очевиден.

Следуя указанному принципу, ограничимся изложением одной лишь теоремы Бернулли, поскольку в ней впервые показана сущность этого исключительно важного закона, который является мостом для перехода от случайности единичных явлений к их закономерности в массовом масштабе. Теорема Бернулли (XVIII в.) гласит, что если при n испытаниях событие произошло m раз, а теоретическая его вероятность равна p , то в пределе существует равенство

$$\lim P_{n \rightarrow \infty} \left\{ \left| \frac{m}{n} - p \right| < \varepsilon \right\} = 1, \quad (6)$$

где ε — сколь угодно малое число.

Не приводя формального анализа этой совершенно очевидной теоремы, строго доказанной Чебышевым, Марковым и другими учеными, остановимся лишь на ее практическом значении. Дело в том, что имеется, с одной стороны, теоретическая вероятность p для единичного явления и, с другой стороны, эмпирическое отношение $\frac{m}{n}$ при большом числе испытаний; в пределе получается равенство этих двух генетически различных величин. Отсюда вытекает, что подлежат рассмотрению две категории случаев:

1. Когда известна теоретическая вероятность события A и нужно определить число появления этого события m при достаточно большом n равенства их $m = np$.

2. Когда неизвестная вероятность p определяется из массовых испытаний, т. е. чисто эмпирически, с использованием того же предельного равенства $p = \frac{m}{n}$.

Назовем первую категорию исследований **прямым** применением закона больших чисел, а вторую категорию исследований, когда по массовым эмпирическим данным можно сделать заключение о теоретической вероятности события, назовем **обратным** применением того же закона. Но прежде надо уточнить смысл термина «испытание», который, кажется, указывает лишь на временную последовательность, что совсем необязательно. Мы можем, с одной стороны, в той же точке или в том же объеме последовательно создавать одни и те же условия, при которых может произойти событие A , с другой — можем создать однородные условия для возможности появления события A в очень многих местах некоторой области и в известный момент наблюдать, в скольких местах из всех имеющихся в наличии это событие в действительности произойдет. Это будет не временное, а пространственное испытание.

Например, имеется водный поток, протекающий над песчаным дном (для простоты примем песок однородным по размеру песчинок). Допустим, что из чисто гидродинамических соображений, подтвержденных также экспериментом, мы нашли хотя бы приближенное выражение для вероятности подъема со дна какой-либо произвольно выбранной песчинки в течение какого-то нами выбранного промежутка времени.

Закон больших чисел показывает, что число фактически поднявшихся песчинок с какой-то определенной (но достаточно большой) площадки и в течение того же промежутка времени будет равно вычисленной нами вероятности, умноженной на число песчинок, лежащих на этой площади. Такой путь применения закона больших чисел часто дает возможность получить новые статистические закономерности. В данном случае из вероятности подъема одной песчинки можно получить выражение для твердого расхода при заданной скорости потока.

Обратное применение закона больших чисел имеет место в тех случаях, когда входящие в задачи вероятности не могут быть вычислены теоретическим путем и их надо получить через эмпирические частоты при достаточно большом числе испытаний. Для иллюстрации приведем следующий пример. Некоторая территория покрыта сетью метеорологических станций обычной густоты. Расстояния между станциями превышают размеры летних ливневых туч, а поэтому две смежные станции могут показать «нуль» осадков, в то время как между ними прошел ливень.

Вполне очевидно, что строить по данным такой сети суточные изотермы совершенно недопустимо. Но если для каждой станции той же сети осреднить показания за достаточно большое число лет (например, за июль в течение 30—40 лет), то ввиду равновероятности прохождения ливневых туч над любой из полос, вытянутых по направлению господствующих ветров, средняя многолетняя цифра осадков за такое-то число июля будет уже достаточно репрезентативной для установления вероятной цифры осадков за этот день.

Здесь намеренно взяли из всех метеорологических характеристик именно осадки, выпадение которых носит пространственно-разрывной характер. Если заменить осадки, например, температурой, то окажется, что даже и при меньшей густоте станций можно провести достаточно надежные изотермы для данного месяца или даже для данных суток, не пользуясь многолетним осреднением. Следовательно, метод изолинии успешно применяется в тех случаях, когда изучаемая характеристика непрерывно распределена по какой-либо площади.

Широко применяются, как известно, иные формы изображения статистических зависимостей: таблицы, графики и эмпирические формулы. Везде, где отбрасывается анализ единичных явлений (когда он практически не представляет интереса), наблю-

дается переход к закономерностям массового статистического характера и использование в прямом или в обратном смысле закона больших чисел.

Теоретико-познавательный смысл закона, перекидывающего мост от хаоса эмпирических единичных явлений к стройной статистической закономерности, был особенно ясно выражен Пуассоном: «Совокупное действие большого числа случайных факторов при некоторых, весьма общих и весьма часто встречающихся условиях приводит к результату, почти не зависящему от случая».

ОДНОМЕРНЫЕ СОВОКУПНОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН

§ 6. Понятие о множествах

В предыдущей главе изложены вероятностные предположения в духе классической теории, согласно которой вероятность, как предел эмпирической частоты, выражается отношением числа равновозможных благоприятных исходов к общему числу также равновозможных исходов. Но нетрудно видеть, что допущение «равновозможности» исходов ограничивает применимость классической теории лишь к дискретной и конечной совокупности событий, на практике же зачастую мы должны эту совокупность считать бесконечной, например при вычислении вероятности поворота стрелки прибора на определенный угол или при вычислении вероятности попадания пули в мишень на таком-то расстоянии от центра и т. д.

Во всех аналогичных случаях дробь, определяющая вероятность в классическом смысле слова, или равна нулю, или не имеет определенного значения. Действительно, сумма бесконечного числа сколь угодно малых, но конечных величин всегда равна бесконечности, и при конечном числе благоприятных исходов наша дробь становится нулем, а при бесконечном их числе — неопределенностью.

В данном случае возможны два решения: 1) отказаться от постулата равновозможности и считать, как это и имело место во втором из вышеприведенных примеров, что попадания в мишень на разных расстояниях от центра мишени неравновозможны и что вероятность попадания с увеличением этого расстояния уменьшается и 2) независимо от выполнения этого постулата можно определять вероятность не для какого-либо определенного значения величины, могущей принимать бесконечное число значений, а лишь для какого-то интервала этих значений. Это будет иметь место и в отношении первого примера, когда вероятность такого-то угла для стрелки, вращающейся по кругу, составляет, очевидно, нуль. Однако вероятность одного из углов внутри определенного интервала при соблюдении постулата

равновозможности будет равна отношению числа градусов этого интервала к общему числу 360° .

Отметим, что первое решение применимо и к дискретному множеству событий, а второе — и к неравновозможности исходов.

Примером неравновозможности дискретных исходов, число которых бесконечно велико, будет решение следующей задачи.

Какова вероятность, что при бросании монеты «орел» выпадает n раз подряд ($n=1, 2, \dots, \infty$)? Эта вероятность равна, оче-

видно, $\frac{1}{2^n}$ ($n=1, 2, \dots, \infty$). Их сумма равна единице, а следовательно, основное требование классической теории вероятности выполнено именно благодаря убыванию вероятности с возрастанием числа n , т. е. в результате неравновозможности событий.

Приведем еще несколько примеров определения вероятности непрерывной совокупности событий, которая в дальнейшем изложении будет играть основную роль, поскольку ошибки измерения почти всегда имеют непрерывный характер.

1. Бросаем иглу на плоскость, на которой проведена система параллельных линий на равных расстояниях, меньших длины иглы. Какова вероятность, что упавшая игла пересечет какую-либо прямую? Это известная задача Бюффона, приводящая, как увидим ниже, к приближенному вычислению числа π .

2. Бросаем шарик так, чтобы он непременно упал на горизонтальную плоскость внутри определенного контура. Какова вероятность, что он упадет внутри другого контура, проведенного внутри первого?

3. Известна средняя суммарная продолжительность загрузки телефонной сети в такое-то время дня. Какова вероятность того, что вызванный мной в это время аппарат будет занят?

Во всех трех примерах как общее число исходов, так и число благоприятных исходов равны бесконечности. Это заставляет нас перейти от классического определения вероятности к какому-то другому, в котором могли бы фигурировать бесконечные числа исходов. Это возможно сделать путем замены понятия числа более общим понятием множества.

В современной математике со второй половины XIX в., а именно после работ Георга Кантора, было введено весьма общее понятие множества, играющего важную роль также и в перестройке теории вероятностей. Это весьма элементарное понятие, неопределимое по своей сущности, поясняется лишь примерами: множество лиц, живущих в данной стране, множество лиц, перенесших эпидемию такого года в таком-то городе и живущих в нем на первое января следующего года, множество телефонных абонентов такого-то города, множество автомашин, выпущенных таким-то заводом, множество целых положительных чисел от 0 до ∞ , множество рациональных функций непрерывной переменной, множество точек в отрезке прямой и т. д.

Каждое множество может иметь конечное (или бесконечное) число подмножеств, определяемых тем, что все его элементы входят в состав основного множества. Так, может быть подмножество машин, выпущенных с каким-либо дополнительным устройством, подмножество целых чисел, представляющих квадраты (или кубы) других целых чисел, и т. д. Для общности принимают, что каждое множество является одновременно и своим подмножеством, что оно может содержать в себе и один лишь элемент, и даже ни одного элемента. В последнем случае он называется «пустым множеством».

Вообще понятие множества обобщает понятие числа, снимая с него установившееся элементарное различие «именованных» и «отвлеченных» чисел, а также включает в себя, кроме конечных чисел, бесконечные числа. В этой связи необходимо добавить, что со времени работ Георга Кантора различаются бесконечности различных порядков, или, как принято говорить, различных «мощностей». Так как в дальнейшем придется считаться с этим различием, то здесь дается о нем качественное представление без углубления в строгие доказательства.

Существующие бесконечные множества, которые можно перенумеровать, называются счетными. Простейшим из них является натуральный ряд чисел. Строго доказывается (здесь это доказательство не приводится), что не только ряд целых чисел, но и совокупности всех рациональных дробей и даже совокупность всех чисел, которые получаются при решении целых алгебраических уравнений, все можно перенумеровать, т. е. все они представляют собой счетные множества.

Далее так же строго доказывается, что совокупность всех точек любого отрезка прямой нельзя перенумеровать, т. е. она не является «счетной», и что если с этого отрезка «снять» все точки, соответствующие дробям и алгебраическим числам, то еще останется бесконечно большое, более высокого порядка, множество трансцендентных точек, которое именно и придает линии качество непрерывности. Важно добавить, что мощность бесконечности числа точек как на площади, так и в объеме и даже в обобщенном «объеме» n -мерного пространства будет та же, что и для точек, составляющих отрезок прямой.

Итак, имеется три категории множеств: конечные, бесконечные — счетные и бесконечные — несчетные (непрерывные). Теперь на основе представлений о множествах вообще можно вернуться к вопросу обобщения теории вероятностей, поскольку классическая теория по своей структуре может относиться, очевидно, лишь к первой, простейшей категории множеств, а именно к конечным множествам.

Такое обобщение становится возможным на основе введенного французским математиком Лебегом понятия меры.

§ 7. Мера Лебега

Собственно говоря, «мера Лебега» имеет гораздо более широкое значение, но здесь ограничимся лишь показанием ее вероятностного смысла.

Ввиду универсальности понятия «множества» можно уверенно говорить о множестве исходов такого-то испытания или вообще о множестве каких-либо однородных событий, например о множестве возможных, но неизвестных «погод» завтрашнего дня и т. д. Каждое такое множество располагается своими отдельными «экземплярами» в каком-то «пространстве параметров». Введем здесь обобщенное понятие параметров, охватывающее не только координаты геометрического пространства, время, углы, расстояния, но даже возможные исходы какого-либо испытания. В задаче Бюффона (см. выше) такими параметрами будут расстояние центра иглы от ближайшей параллельной прямой и угол между иглой и прямой. В примере падения шарика на плоскости параметрами будут координаты точки касания шарика с плоскостью в момент падения, а в примере телефонного вызова — время.

В каждой вероятностной задаче рассматриваем множество возможных исходов какого-то испытания, из которых известная часть благоприятствует наступлению рассматриваемого события, а остальные ему не благоприятствуют. Назовем совокупность исходов, благоприятствующих наступлению ожидаемого события, областью какого-то множества, объединяющего совокупность всех возможных исходов.

Понятия «множества» и «области» позволяют выйти из ограниченных представлений классической теории вероятности и дать ей расширенное толкование. Действительно, в простейшем случае бросания шестигранной кости «множеством» будет число шесть, а областью благоприятствующих исходов — число «единица». Для получения вероятности того, что выпадет столько-то очков, надо лишь разделить область на множество; мы получим $\frac{1}{6}$. Далее надо ввести более общее понятие «вероятностной меры», которое дало бы возможность решить и «задачу Бюффона» и вообще любую непрерывную задачу, определяемую любым числом параметров.

Введем функцию n параметров $f(x_1, x_2, \dots, x_n)$, или сокращенно $f(x)$, которая для всех «точек», лежащих внутри области, принимает значение «единица», а для всех остальных точек — значение «нуль», и назовем «вероятностной мерой», или просто мерой, среднее значение этой функции во всем множестве. Сразу же увидим, что для всех задач классической теории мера совпадет с «вероятностью» в классическом ее определении. Покажем теперь на приведенных выше примерах, что и на всех мыслимых вероятностных задачах введенная нами мера

действительно отражает степень «ожидаемости» данного события в количественном ее выражении, т. е. вероятность.

Правда, в приведенных выше примерах было введено одно ограничение — равновозможности всех исходов производимого нами испытания. Более общий случай, когда разные исходы, имеющие разные вероятности, требуют введения нового понятия — распределения вероятностей, будет пока-

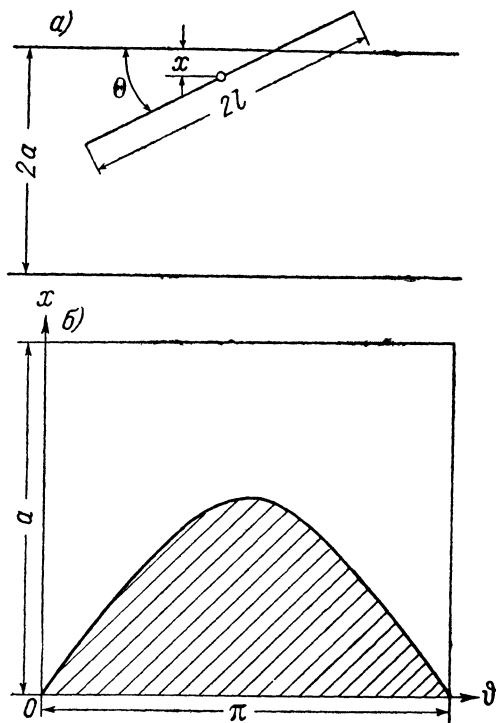


Рис. 1. Задача Бюффона.

a — пересечения иглой, *б* — схема отношения площадей

зан в следующих главах. Но для расширения понятия вероятности понятий непрерывных множеств с помощью «меры Лебега» достаточно пока ограничиться задачами с равновозможными исходами. Рассмотрим их в том же порядке, что и выше.

1. Обозначим через x расстояние от центра иглы до ближайшей прямой (рис. 1 *a*), а угол между иглой и прямой — через θ . Имеем два параметра, которые можно (условно) представить плоскостью (x, θ) . Параметр x изменяется в пределах от 0 до a , параметр θ — в пределах от 0 до π . Ожидаемое событие — пересечение иглой одной из параллельных прямых — может про-

изойти лишь в случае выполнения неравенства $x \leq l \sin \theta$. Следовательно, если множество представить в виде прямоугольника со сторонами a и π (рис. 1 б), то область попадания выразится заштрихованной частью этого прямоугольника.

Вероятностная мера, как указывалось ранее, выразится отношением той площади, для которой функция $f(x, \theta)$ равна единице, к площади всего множества, т. е. отношением

$$P(A) = \int_0^{\pi} \frac{l \sin \theta d\theta}{2\pi} = \frac{2l}{a\pi}.$$

Таким образом, бросая иглу достаточно большое число раз, можно, зная величину a и l , определить с некоторым приближением число π .

На рис. 2 изображена схема следующей задачи.

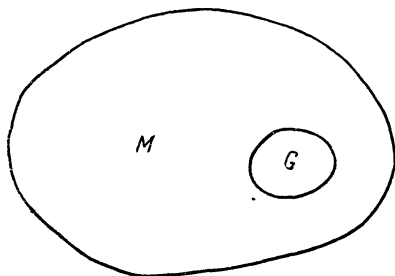


Рис. 2. К определению вероятности падения шарика внутри контура G .

Вероятность того, что шарик упадет внутри области G при постулируемой равномерности ему выпасть в любой точке множества M , равна, очевидно, отношению суммы интегралов

$$\int_{(G)} f(x, y) dx dy + \int_{(M-G)} f(x, y) dx dy$$

к интегралу

$$\int_{(M)} f(x, y) dx dy.$$

Обозначая площадь множества M через m , а площадь области G через g и учитывая, что второй интеграл суммы, очевидно, равен нулю, так как $f(x, y)$ для всех точек вне области G равна нулю, получаем

$$P(A) = \frac{g}{m}. \quad (7)$$

Этот тривиальный результат был нам заранее известен, но здесь приведено его подробное доказательство для того, чтобы показать самый смысл «лебеговой меры».

3. Если известно из непосредственного статистического подсчета за достаточно длительный период, что в таком-то временном интервале (например, между 13 и 14 час. рабочего дня) средняя загрузка телефонной сети составляет τ секунд, то $M=3600$ сек., а $G=\tau$ сек. Поэтому вероятность того, что телефонный вызов в течение этого часа получит сигнал занятости, выразится отношением

$$p = \frac{\tau}{3600}.$$

§ 8. Математическое ожидание

Одномерная совокупность характеризуется в первую очередь двумя статистическими свойствами: положением ее центра и рассеянием остальных ее элементов относительно этого центра.

Как увидим далее, возможны два определения центра и две соответствующие характеристики рассеяния. Но в этом разделе рассмотрим лишь один из двух возможных центров, носящий название математического ожидания, которое, как увидим, представляет собой также простое среднее арифметическое значение из всех элементов совокупности.

Термин «математическое ожидание», на первый взгляд не очень понятный, возник в теории вероятностей в связи с той областью азартных игр, в которой возникла и сама эта теория. Если игрок имеет вероятность p выиграть сумму a , то его математическое ожидание определяется произведением pa . Если эта вероятность соответственно относится к проигрышу, то его математическое ожидание становится отрицательной величиной. Если игрок имеет вероятность p_1 выиграть сумму a_1 , вероятность p_2 выиграть сумму a_2 и т. д., то его математическое ожидание должно составить сумму

$$p_1 a_1 + p_2 a_2 + \dots$$

Игра будет справедливой, если математические ожидания всех игроков в этой игре будут равны, иными словами, если в результате бесконечно продолжительной игры каждый останется с той суммой, которая у него была в самом начале.

При современном развитии теории вероятностей и распространении ее на всевозможные области этот первоначальный «игрецкий» термин получил распространение на все те области, в которых рассматриваются вероятности случайных величин. Но тождество этого термина с весьма привычным термином «среднего арифметического» привело к тому, что последний термин употребляется гораздо чаще.

Среди многочисленных и весьма разнообразных одномерных совокупностей для нас особое место занимает совокупность чисел, полученных при измерении какой-либо величины. И именно

этой совокупности, которую будем сокращенно называть «измерительной», посвящена значительная часть данной книги.

Измерительная совокупность всегда представляется ограниченной известными пределами, хотя внутри этих пределов она может мыслиться бесконечной по числу отдельных измерений. Далее, она характеризуется одним, и только одним, сгущением точек, обычно расположенных вблизи ее центра. И, наконец, расположение точек отдельных измерений обычно не очень сильно отклоняется от симметрии по отношению к ее центру.

Каждый экспериментатор интуитивно считает, что математическое ожидание полученной им измерительной совокупности должно находиться где-то вблизи ее центра. И это интуитивное представление большей частью оправдывается, о чем в дальнейшем будет сказано подробно. Это «истинное» значение в процессе измерения считается постоянным, но, по существу, оно, как из дальнейшего будет вполне ясно, никогда не может быть известно, так как полностью нельзя исключить все возможные и неизбежные ошибки, связанные со всяким измерением. Единственно, в чем можно быть уверенным, так это в том, что при некоторых ограничительных условиях, о которых будет сказано ниже, среднее арифметическое будет тем ближе к истинному значению, чем больше число измерений, однако совпасть с ним оно вообще не может.

Все результаты отдельных измерений могут быть приняты за «случайные» величины вследствие того, что они отклоняются от истинного значения под влиянием многочисленных факторов, которые учесть мы также не в состоянии. Будем рассматривать в самом общем виде сначала дискретную случайную величину X , представленную рядом значений $x_1, x_2, \dots, x_j, \dots, x_n$.

Среднее арифметическое этого ряда для сокращения записи будем обозначать горизонтальной чертой сверху:

$$\bar{x} = \frac{1}{n} \Sigma x_j. \quad (8)$$

В общем случае, когда величина x может принимать n значений, среднее арифметическое никакого иного смысла, кроме центра распределения, очевидно, не может иметь; следовательно, вопрос о повышении его «точности» с выражением числа n не может даже ставиться. Но в частном случае, когда рассматривается совокупность измерений какой-то определенной величины, это же среднее является каким-то приближением к истинному значению измеряемой величины и оценивается нами именно с этой точки зрения.

Будем исходить из рассмотрения однородных условий измерения, причем все ошибки как систематические, так и грубые,

выскакивающие, предполагаем исключенными. Учитываем только чисто случайные ошибки, которые характеризуют несовершенство процесса измерения и принципиальную невозможность производства измерительных отсчетов с абсолютной точностью.

Обозначим неизвестную, но по предположению постоянную измеряемую величину через a . Тогда каждое отдельное измерение даст какую-то, также нам неизвестную ошибку, определяемую равенством

$$\varepsilon_j = x_j - a.$$

Среднее из n измерений будет

$$\bar{\varepsilon} = \bar{x} - a. \quad (9)$$

Если бы можно было допустить, что при достаточном (очень большом) числе измерений все положительные и отрицательные ошибки друг друга компенсируют и дают в сумме нуль, то полученная нами средняя величина совпала бы с истинным значением измеряемой величины a . Но можно ли это допустить? Приведенный ниже анализ показывает, что такое допущение является весьма шатким.

А. Пуанкаре дал весьма глубокую критику понятия среднего арифметического и связанных с ним вероятностных представлений [3, гл. X]. Он отмечает, что «измеряя длину одного метра хотя бы миллион миллионов раз, мы никогда не получим данную длину с точностью до одного микрона» [3, стр. 187].

Это положение Пуанкаре может показаться практически несущественным, но это не так. Экспериментатор и наблюдатель должны понять, что даже при бесконечном числе измерений они не могут освободиться от какой-то доли ошибки, свойственной данному измерительному прибору. Если бы равенство $x - a = 0$ было справедливо при достаточно большом числе измерений, то можно было бы компенсировать неточность измерительного прибора соответственным повышением числа измерений. Такой вывод был бы очевидным абсурдом! Каждый экспериментатор интуитивно стремится применять в целях достижения большей точности более точные приборы. Дадим еще одно важное указание, также вытекающее из замечания А. Пуанкара.

Известно, что каждый измерительный прибор обладает присущим ему «порогом чувствительности», представляющим собой наименьшую возможную для него ступень отсчета измерения. И получаемое нами среднее из измерений также имеет ошибку порядка «порога чувствительности» прибора. Поэтому все знаки среднего из измерений, не меньше этой величины, должны быть отброшены, как фиктивные.

Вернемся к вопросу сравнения результатов получения среднего двумя изложенными выше способами и покажем, что тер-

мины «среднее арифметическое» и «математическое ожидание» в пределе при $n \rightarrow \infty$ тождественны.

При любом числе измерений (практически это удобнее делать при очень большом их числе) можно разбить их совокупность, расположенную в восходящем порядке значений, на равные интервалы. Пусть число интервалов будет k ; это число обычно гораздо меньше, чем n .

Возьмем внутри каждого i -того интервала его среднее значение x_i , полученное обычным арифметическим путем, и обозначим через m_i число членов внутри этого интервала. Тогда можно по-иному выразить то же среднее:

$$\bar{x} = \frac{1}{n} \sum_1^k m_i x_i. \quad (10)$$

Их тождество вытекает из того факта, что в любой сумме $x_1 + x_2 + \dots + x_j + \dots + x_n$ можно заменить любую ее часть произведением $m_i x_i$, от чего результат, очевидно, не изменится.

Введем теперь знаменатель n под знак суммы и будем увеличивать n до бесконечности, в результате чего при конечном k все m_i также возрастут до бесконечности. Пользуясь законом больших чисел, запишем

$$\lim \frac{m_i}{n} = p_i.$$

Следовательно, в пределе будем иметь

$$\bar{x} = \sum_1^k p_i x_i. \quad (11)$$

Это и есть математическое ожидание случайной величины x_1 , выраженное через математические ожидания ее частных значений по интервалам.

В пределе при $n \rightarrow \infty$ оба метода получения среднего должны, очевидно, привести к тождественным результатам, в связи с чем может быть поставлен вопрос лишь об их сравнительной трудоемкости. Общего решения такой вопрос не имеет, хотя, казалось бы, при очень большом числе n сумма операций по осреднениям сравнительно малого числа членов внутри каждого интервала и по суммированию полученных таким путем k средних займет меньше времени, чем сплошное суммирование, и, во всяком случае, будет более застраховано от случайных просчетов.

Но суть дела не в этом. При втором методе и при очень большом числе измерений мы попутно получаем (приближенное) выражение закона распределения данной совокупности по частотам ее значений.

Такой эмпирический закон распределения изображен на рис. 3. Полученная автором книги (в связи с одним экспериментальным исследованием) совокупность измерений распределена по ступенчатой кривой, которая в пределе при $n \rightarrow \infty$ должна будет совпадать с теоретическим законом.

Перейдем к вопросу осреднения непрерывной совокупности значений случайной величины. Такая непрерывность может иметь место при непрерывной записи измерения некоторой величины, значение которой колеблется около среднего, например при из-

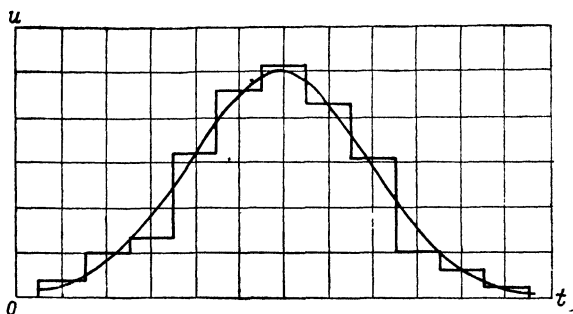


Рис. 3. Распределение частот при нормальном законе.

мерении пульсации скорости в какой-либо точке турбулентного потока. Рассмотрим именно этот случай, точность осреднения в котором будет зависеть от длительности измерения, так как «число» измерений при непрерывной записи всегда равно бесконечности. Самый простой прием получения среднего может быть изображен здесь обыкновенным интегралом Римана

$$\bar{u} = \frac{1}{T} \int_0^T u dt, \quad (12)$$

для вычисления которого достаточно или разделить график (рис. 3) вдоль оси времени на возможно более частые интервалы Δt , построить в средних точках интервалы ординаты, затем измерить их и просуммировать по формуле Симпсона, или спланиметрировать площадь, заключенную между кривой $u(t)$ и осью времени, чтобы получить приближенное значение интеграла (12). Деление полученного интеграла на период T даст искомое среднее значение.

Но можно поступить совершенно иначе. Нанесем на рис. 4 прямые, параллельные оси времени, через равные интервалы скорости Δu и подсчитаем число пересечений прямой, проведенной на высоте $\frac{u_i + u_{i+1}}{2}$. Пусть оно будет равно n_i ; общее число

пересечений всех параллельных прямых с кривой $u(t)$ обозначим через n . Следовательно, $\sum n_i = n$. Тогда дробь $\frac{n_i}{n}$ выразит эмпирическую частоту значений скоростей в интервале $\Delta u = u_{i+1} - u_i$.

Нетрудно видеть, что отношение $\frac{n_i}{n}$ является в то же время и мерой Лебега по определению, данному в предыдущем параграфе. Множеством здесь будет произведение $(n\Delta u)$, а областью — произведение $(n_i\Delta u)$, отношение которых

$$\frac{n_i \Delta u}{n \Delta u} = \frac{n_i}{n}$$

можно назвать мерой Лебега для i -того интервала. Перейдя за-

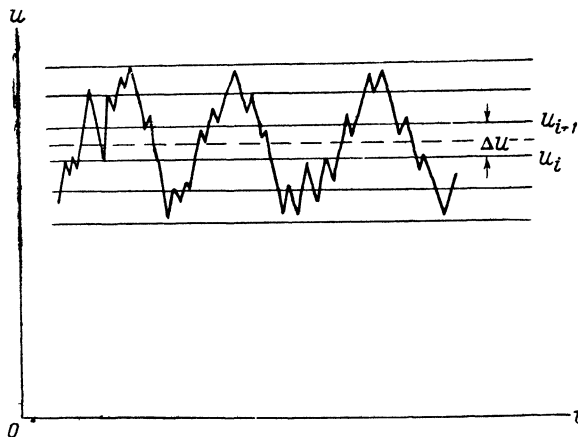


Рис. 4. Обработка хронограммы пульсации скоростей.

тем от конечного интервала к дифференциалу, получим в пределе равенство

$$\lim \frac{n_i \Delta u}{n \Delta u} = \left(\frac{n_i}{n du} \right) du,$$

в правой части которого в скобках стоит неопределенность, раскрывающаяся, очевидно, в плотность вероятности $f(u)$.

Затем от суммы $\sum u_i \frac{n_i}{n}$ перейдем к интегралу Лебега, который и даст точное выражение математического ожидания пульсирующей скорости

$$M.O. = \bar{u} = \int_{-\infty}^{+\infty} u p(u) du. \quad (13)$$

Здесь приведены оба осреднения в форме интегралов в целях большей наглядности, хотя на практике всегда придется пользоваться обыкновенным суммированием. Интеграл (10)

после деления хронограммы (рис. 4) вертикальными линиями на n интервалов Δt выразится через

$$\bar{u} = \frac{\sum u_i}{n}. \quad (14)$$

Интеграл (13) после деления той же хронограммы горизонтальными параллельными линиями на какое-то число интервалов Δu и обозначения числа пересечений в каждом интервале через n_i получит вид

$$\bar{u} = \sum \frac{n_i}{n} u_i. \quad (15)$$

Возвращаясь к изображению искомого среднего через два различных интеграла, отметим, что интеграл (12) представляет собой обычный интеграл Римана, а интеграл (13) — не что иное, как интеграл Лебега. В первом случае — исходным из размещения скоростей по времени, во втором — из размещения их же по вероятностям.

В математическом анализе доказывается, что интеграл (12) не всегда, т. е. не при всяком виде функций, существует, а интеграл (13) существует всегда. Но этот вопрос нас здесь не интересует, так как во всех практически мыслимых случаях измерений существуют оба интеграла.

Важно другое — вычисление среднего значения по вероятностной схеме может быть легче и точнее. В этом нетрудно убедиться, так как точное измерение ординат хронограммы гораздо труднее, чем точное нанесение на хронограмму параллельных горизонтальных прямых с последующим подсчетом чисел их пересечения с кривой.

§ 9. Рассеивание

Основными характеристиками всякой одномерной совокупности, в частности совокупности измерений, являются центр совокупности и ее рассеивание около этого центра.

Рассмотренное нами в предыдущем разделе математическое ожидание (или среднее арифметическое) представляет собой одно из двух возможных определений центра измерительной совокупности. Другим, также возможным центром будет медиана, которую рассмотрим в следующем параграфе.

Каждому из названных центров соответствует свое, т. е. ему свойственное, определение «рассеивания». В общем виде этот термин означает меру «разброса» элементов совокупности или каким-то способом осредненное расстояние их от центра. Сразу становится ясно, что алгебраическая сумма этих отклонений (деленная на их число) не может ни в какой мере охарактеризо-

вать этот разброс, так как даже очень большие отклонения разных знаков могут дать в сумме малую величину.

Следовательно, необходимо освободиться от знака, что возможно сделать двумя простыми способами: или путем возведения всех отклонений в квадрат, или определением их модулей.

Оба эти способа одинаково логичны. Можно даже сказать, что третьего способа не существует, если не идти на излишнее осложнение в связи с применением более высоких степеней.

Вопрос выбора способа рассмотрим позже. Сначала осветим подробнее первый путь — вычисление среднего квадратического отклонения — и докажем одно важное положение, устанавливающее тесную связь между рассмотренным выше средним арифметическим, принятым за центр распределения, и средним квадратическим отклонением как мерилom рассеяния. Это положение следующее: сумма квадратов отклонений от среднего арифметического меньше, чем от любого другого условно выбранного центра.

Для доказательства указанного выше положения возьмем произвольное значение центра и обозначим его через \tilde{x} . Квадратическое рассеяние относительно этого произвольного центра выразится через

$$\frac{\sum (x_k - \tilde{x})^2}{n}.$$

Зададимся целью выбрать \tilde{x} так, чтобы указанное выражение получило минимальное значение. Для этого, приравнявая нулю первую его производную по \tilde{x} , получим

$$\frac{\partial}{\partial \tilde{x}} \left\{ \frac{\sum (x_k - \tilde{x})^2}{n} \right\} = \frac{2}{n} \sum x_k - \frac{2\tilde{x}}{n} = 0,$$

откуда

$$\tilde{x} = \frac{\sum x_k}{n} = \bar{x}. \quad (16)$$

Выведенное положение (16) представляет собой принцип минимума суммы квадратов отклонений, установленный впервые Лежандром, а затем развитый в приложениях к различным вопросам Гауссом. Этот принцип, как увидим далее, также связан с законом нормального распределения и со способом наименьших квадратов, который играет доминирующую роль при отыскании формы эмпирических зависимостей и принадлежит также Гауссу. Среднее значение квадратов отклонений имеет большое значение в математической статистике и носит специальное название дисперсии.

Ниже, не доказывая самоочевидных положений, рассмотрим некоторые свойства дисперсии:

а) дисперсия постоянной величины равна нулю;

б) дисперсия случайной величины, умноженной на постоянную величину, равна квадрату этой постоянной, умноженной на дисперсию случайной величины:

$$D(kx) = k^2 D(x);$$

в) дисперсия суммы независимых случайных величин равна сумме их дисперсий. Докажем это положение для двух величин, так как распространение его на многие величины является слишком бесспорным.

Имеем

$$D(X + Y) = M.O. \{ (X + Y) - M.O.(X + Y) \}^2.$$

Раскрывая правую часть и замечая, что при независимости величин их отклонения от их средних также независимы (а потому их осредненное произведение равно нулю), напишем

$$D(X + Y) = M.O. (X - \bar{X})^2 + M.O. (Y - \bar{Y})^2 = D(X) + D(Y).$$

В большинстве случаев бывает удобнее вводить для характеристики рассеяния величину, однородную по размерности с рассматриваемой случайной величиной. Такая величина, получаемая извлечением квадратного корня из дисперсии, называется средним квадратичным, или, иначе, стандартным отклонением, или просто стандартом:

$$\sigma = \sqrt{D(X)}.$$

Перейдем к решению важного вопроса, освещенного выше лишь в качественном смысле: как зависит ошибка среднего арифметического от числа измерений; что она уменьшается это ясно интуитивно, но как именно, мы не знаем.

Имея n независимых и одинаково распределенных случайных величин X_1, X_2, \dots, X_n , каждая из которых представлена рядом ее значений, определим дисперсию их среднего:

$$D(\bar{X}) = D \left\{ \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} \right\} = \frac{1}{n} \frac{\Sigma D(X)}{n},$$

а так как все они одинаково распределены, т. е. имеют равные дисперсии, то получим

$$D(\bar{X}) = \frac{nD(X)}{n^2} = \frac{D(X)}{n},$$

откуда и следует, что стандартное отклонение среднего равно

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}. \quad (17)$$

Докажем далее, что дисперсия не может зависеть от начала отсчета всех величин данной совокупности. Введем

$$y_j = x_j + b,$$

откуда имеем

$$\sum (x_j - \bar{x})^2 = \sum \{(x_j + b) - (\bar{x} + b)\}^2 = \sum (y_j - \bar{y})^2.$$

Для дальнейшего необходимо установить некоторые общие положения, относящиеся к той категории одномерных совокупностей, которой главным образом будем заниматься, т. е. к категории измерительных совокупностей.

Принципиально можно измерять какую-либо определенную величину сколь угодно большое число раз, хотя и ясно, что, чем число измерений будет больше, тем тот или иной центр из всех измерений будет точнее, т. е. ближе к истинному значению измеряемой величины. Следовательно, если представить себе большое число последовательных серий измерений, то каждая из них будет дополнять и уточнять результаты всех предшествовавших серий. Если представить себе мысленно число этих серий бесконечно большим, то такая совокупность измерений будет наиболее полной. Назовем ее генеральной совокупностью всех возможных измерений той же величины. Все отдельные же серии измерений будут выборками из генеральной совокупности.

В дальнейшем увидим, что генеральная совокупность, являющаяся, очевидно, недостижимой, может быть успешно заменена теоретической, непрерывной кривой распределения, включающей в себя все те измерения, которые можно было бы осуществить, не имея ограничений во времени. Совершенно очевидно, что центр такой генеральной совокупности будет находиться ближе всего к «истинному» значению измеряемой величины, но утверждать их совпадение, строго говоря, все же не имеем права, поскольку ограничение, связанное с «порогом чувствительности», сохраняет свою силу при любом и даже бесконечном числе измерений. Известная теорема П. Чебышева, представляющая собой наиболее строгую формулировку закона больших чисел, доказывает упомянутое совпадение вполне строго [2]. Но эта теорема игнорирует наличие в измерителе порога чувствительности. Если условно допустить, что измеритель дает возможность записать все его показания с любым числом знаков, то, согласно этой теореме, строго доказывается, что центр генеральной совокупности всех измерений действительно совпадает с истинным значением измеряемой величины. А если это допущение приближенно оправдывается, т. е. порог чувствительности нашего измерителя пренебрежимо мал и математическое ожидание практически «совпадает» с истинным значением из-

меряемой величины, то можно получить истинную, генеральную дисперсию:

$$\sigma^2 = \frac{\sum (x_j - a)^2}{n},$$

которая по тем же соображениям, что и приведенные выше, будет меньше любой другой, отсчитанной от другого центра. Если же совокупность всех x_j является лишь конечной выборкой, то можно получить из нее лишь эмпирическую, выборочную дисперсию:

$$s^2 = \frac{\sum (x_j - \bar{x})^2}{n}.$$

Вопрос об отношении между генеральной и выборочной дисперсиями в общем виде не может быть решен. Известно только, что, чем меньше объем выборки, тем больше вероятность, что ее среднее существенно отличается от a , а следовательно, оно и менее надежно при непосредственном определении всех характеристик истинной, т. е. генеральной, совокупности. Именно в этом случае возникает необходимость опираться в своих выводах на структуру той теоретической функции распределения, которая по тем или иным соображениям считается наиболее подходящей для отображения неизвестной генеральной совокупности.

Вопрос о структуре функций распределения займет в следующих главах надлежащее место. Однако некоторые связанные с ним частные вопросы приходится осветить заранее.

Дело в том, что, переходя от вероятностей дискретных случайных величин к вероятностям непрерывных величин, как это уже было показано в предыдущем разделе, сразу же приходится сталкиваться с тем очевидным фактом, что вероятность какого-либо вполне определенного значения непрерывной случайной величины не может быть отличной от нуля, а потому можно говорить лишь о вероятности значений внутри какого-то интервала:

$$P(x_{k+1} - x_k) = \frac{P_{k+1} - P_k}{x_{k+1} - x_k}.$$

Но, с другой стороны, когда мы вправе считать интервал сколько угодно малым, тогда дробь, стоящая в правой части, при стремлении ее числителя и знаменателя к нулю даст в пределе производную функции p по переменной x :

$$\frac{dP}{dx} = p(x),$$

носящую название плотности вероятности случайной переменной x . Соответственно этому определению интегральная

функция $P(x)$ должна теперь трактоваться как характеристика изменения вероятности значений x , заключенных в пределе от нижней границы совокупности (например, от нуля или минус бесконечности) до x_1 . Следовательно,

$$P(x) = \int_0^x p(x) dx.$$

В дальнейшем придется применять к ряду размерных случайных величин следующие преобразования, ведущие к замене этого ряда другим безразмерным рядом.

1. Вычтя из каждого члена ряда его среднее значение и разделив полученную разность на среднеквадратическое отклонение, получим нормированный ряд

$$\xi_j = \frac{x_j - \bar{x}}{\sigma},$$

который обладает двумя свойствами: его среднее всегда равно нулю:

$$\bar{\xi}_j = \frac{\bar{x} - \bar{x}}{\sigma} = 0,$$

а дисперсия равна единице:

$$D(\xi) = \frac{\sum (\xi_j - \bar{\xi})^2}{n} = \frac{\bar{x}^2}{\sigma^2} - \frac{\bar{x}^2}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1.$$

2. Разделим каждый член ряда на отношение дисперсии к среднему:

$$\xi = \frac{x_j \bar{x}}{s^2}.$$

Ряд значений ξ будет обладать свойством, заключающимся в том, что его среднее и его дисперсия будут равны. Действительно, имеем

$$\bar{\xi} = \frac{\bar{x}^2}{s^2},$$

$$D(\xi) = \frac{1}{n} \sum \left(\frac{x_j \bar{x}}{s^2} - \frac{\bar{x}^2}{s^2} \right)^2 = \frac{\bar{x}^2 s^2}{s^4} = \bar{\xi}.$$

Первое преобразование (нормирование) широко применяется к нормальному (симметричному) закону распределения, играющему в теории ошибок доминирующую роль. Второе преобразование будем иногда применять к одному из асимметричных законов распределения.

§ 10. Медиана и мода

В предыдущих параграфах рассмотрены две тесно связанные характеристики совокупностей случайных величин: среднее и дисперсия. При этом установлено, что среднее значение представляет собой такой центр, относительно которого сумма квадратов отклонений имеет минимальное значение.

Но совершенно ясно, что возведение отклонений в квадрат отнюдь не является единственным возможным способом избавиться от знаков отклонений. Можно, кажется, поступить проще — суммировать модули (абсолютные значения) отклонений, что именно и было впервые предложено Лапласом ([1], стр. 56 и «Второе дополнение», стр. 41—50). Тогда по аналогии с принципом Лежандра, изложенным в предыдущем параграфе, возникает другой принцип и ставится подобная задача отыскания центра, от которого сумма модулей будет минимальной.

Расположим нашу измерительную совокупность в порядке возрастания значений ее элементов: $x_1, x_2, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n$.

Считая n нечетным, что при большом числе измерений не представляет ограничения, сделаем k равным $\frac{n-1}{2}$, что и будет номером элемента, равноотстоящего от обоих концов. Возьмем далее сумму модулей отклонений всех элементов от x_k , которую можно написать в виде

$$\sum_{i=1}^{i=k-1} (x_k - x_i) + \sum_{i=k+1}^{i=n} (x_i - x_k).$$

Здесь x_k — постоянная величина, а i принимает все значения, кроме $i=k$. Раскрывая это выражение, получим, что оно равно

$$\sum_{k+1}^n x_i - \sum_1^{k-1} x_i.$$

Легко доказать, что приведенная разность меньше тех двух разностей, которые получились бы, если бы мы отсчитывали отклонения от x_{k-1} или от x_{k+1} . Действительно, имеем

$$\sum_{k+1}^n x_i - \sum_1^{k-1} x_i < \sum_k^n x_i - \sum_1^{k-2} x_i.$$

Вычитая левую часть из правой, получаем

$$x_k - x_{k-2} > 0,$$

что, очевидно, является справедливым выражением. Следовательно,

$$\sum_{k+1}^n x_i - \sum_1^{k-1} x_i < \sum_{k+2}^n x_i - \sum_1^k x_i.$$

Вычитая правую часть неравенства из левой, также будем иметь, очевидно, справедливое неравенство

$$x_{k+1} - x_{k-1} > 0.$$

Перейдем теперь к непрерывному множеству, представленному плотностью вероятности.

Решаем задачу о минимуме интеграла:

$$\int_0^{\infty} |x - c| p(x) dx = \min, \quad (18)$$

который преобразуем к виду

$$\int_0^c (c - x) p(x) dx + \int_c^{\infty} (x - c) p(x) dx = \min.$$

Дифференцируем эту сумму по параметру c , означающему какой-то условно принятый центр.

Напомним известное в анализе правило дифференцирования по параметру:

$$\frac{d}{dc} \int_a^b f(x, c) dx = f(b, c) \frac{db}{dc} - f(a, c) \frac{da}{dc} + \int_a^b \frac{df}{dc} dx.$$

Не приводя здесь элементарных выкладок, покажем лишь их результат:

$$\int_0^c p(x) dx = \int_c^{\infty} p(x) dx, \quad (19)$$

из которого видно, что сумма модулей отклонений имеет минимальное значение тогда, и только тогда, когда наш условно принятый центр делит пополам площадь кривой распределения:

$$\int_0^{\infty} p(x) dx.$$

Но так как вся площадь распределения равна единице, то каждая сторона равенства (19) равна половине.

Имеем два (в пределе эквивалентных) определения того центра, который удовлетворяет принципу Лапласа и называется медианой. Для дискретного ряда медиана является средним членом совокупности, расположенной в возрастающем (или убывающем) порядке. Для генеральной совокупности измерений, представляющей собой всевозможные значения измерений с их вероятностями, медиана является абсциссой той точки, которая делит площадь кривой плотности вероятности пополам.

Нетрудно представить себе, что при возрастании числа измерений к бесконечности, когда, естественно, должны будем перейти от отдельных цифр к их частости, а следовательно, и к плотности вероятности, оба определения медианы совпадут.

До сих пор мы ограничивались дискретной совокупностью с нечетным числом измерений; при четном числе принято ставить медиану посередине между двумя центральными значениями

$$k = \frac{n}{2} \quad \text{и} \quad k + 1 = \frac{n}{2} + 1.$$

Вопрос выбора одного из двух равноусловных центров — среднего арифметического или медианы, — вообще говоря, довольно сложен, поэтому к нему еще не один раз вернемся. Однако основные соображения относительно выбора центра рационально высказать сейчас.

Во-первых, установим, что в тех случаях, когда ряд измерений, расположенных в порядке возрастающих значений, составляет простую арифметическую прогрессию $a, a+b, a+2b, \dots, a+nb$ с числом членов ряда $n+1$, их сумма равна $(n+1)(a+b\frac{n}{2})$. Отсюда получаем среднее арифметическое $a+b\frac{n}{2}$.

Медианой в случае n четного (т. е. нечетном числе членов) будет член $(1+\frac{n}{2})$, в случае же n нечетного медиана расположится посередине между членами $(1+\frac{n}{2})$ и $(2+\frac{n}{2})$. В обоих случаях она будет равна $a+b\frac{n}{2}$, т. е. совпадет со средним арифметическим и тем самым подтвердит тот очевидный факт, что в этом

ряду все отклонения от центра вправо и влево на равное число интервалов равны; следовательно, оба центра должны совпасть.

В дальнейшем будет видно, что симметрия отклонений, представленная в законе нормального распределения (а также и в некоторых других), неизбежно влечет за собой совпадение медианы со средним арифметическим. Однако указанное свойство, т. е. симметрия отклонений, хотя и встречается в большинстве случаев, но все же не является общим правилом, так как в некоторых случаях измерительная совокупность является асимметричной по существу. Такие случаи будут нами рассмотрены в одной из последующих глав.

Сейчас нам следует рассмотреть вопрос о выборе между средним арифметическим и медианой в тех случаях, когда они в пределе при бесконечно большом числе измерений должны совпасть, а при конечном, ограниченном числе измерений могут существенно разниться между собой, и выяснить, какому из этих центров надо отдать предпочтение.

Общего ответа, конечно, не может быть хотя бы потому, что каждый конкретный ряд измерений является, очевидно, лишь выборкой. Известно, что, чем меньше объем этой выборки, тем больше вероятность отклонения ее характеристик от характеристик генеральной совокупности.

В виде примера в табл. 1 приведены перестроенные от хронологического порядка к вариационному виду два ряда убывающих значений среднегодовых расходов рек Туры и Лобвы за 19-летний период.

Т а б л и ц а 1

Среднегодовой расход, л/сек. с 1 км ²		Среднегодовой расход, л/сек. с 1 км ²	
р. Тура	р. Лобва	р. Тура	р. Лобва
7,00	11,9	2,66	4,76
5,98	9,38	2,54	4,56
5,35	9,00	2,49	4,49
4,97	8,50	2,03	3,80
4,61	8,15	1,90	3,53
4,28	7,77	1,85	3,47
3,98	7,00	1,81	3,35
3,33	6,70	1,44	2,91
3,30	6,59		
2,73	5,30		
2,68	4,91		
		Среднее 3,42	6,11
		Медиана 2,73	5,30

Из данных табл. 1 видно, что медиана существенно меньше среднего значения, что вообще присуще всем асимметричным распределениям с положительной асимметрией, с которыми только и будем иметь дело в дальнейшем. Отрицательная асимметрия встречается чрезвычайно редко, а в измерительных совокупностях вообще не может иметь места.

Отметим попутно, если при исследовании какого-либо ряда измерений обнаружится, что медиана в нем располагается правее среднего значения, то это следует рассматривать как случайный эффект, связанный с непродолжительностью ряда. Если же исследуемый ряд по существу должен быть симметричен, то небольшие отклонения медианы от среднего в ту или иную сторону так же равновероятны, как и любые отклонения при симметричном распределении.

В частности, при симметричном распределении медиана имеет одно существенное преимущество перед средним, заключающееся в том, что измерения, сильно отклоняющиеся от смежных измерений, которые можно считать или маловероятными, или

просто ошибочными, на медиане совсем не отразятся, а на среднее значение, очевидно, повлияют. Этот вопрос будет рассмотрен в главе, посвященной природе ошибок.

Уместно указать еще на одно удобное свойство медианы: если нам по какой-либо причине надо построить для данного ряда соответственный ряд какой-нибудь монотонной функции от его членов, то нетрудно убедиться, что медиана нового ряда совпадет по своему положению с медианой основного ряда.

Изложенное имеет силу прежде всего в отношении высших моментов, о которых будет сказано в следующем разделе.

Сейчас надо сделать лишь одну оговорку. Ряд должен состоять только из положительных цифр, иначе всякая четная степень (или четная функция) обратит все отрицательные члены ряда в положительные, в результате чего положение медианы сильно изменится. Но такое ограничение незначительно суживает область применения медианы, так как результаты измерения почти всегда существенно положительны, за исключением измерений, отнесенных к их центру, приобретающих два направления.

В математической статистике, кроме двух изложенных нами возможных центров распределения, пользуются еще и третьим центром, а именно модой, определяемой равенством

$$\frac{\partial p(x)}{\partial x} = 0.$$

Мода есть абсцисса, соответствующая максимальной ординате кривой плотности вероятности.

Коренное отличие между тремя перечисленными центрами — средним арифметическим, медианой и модой — заключается в следующем: каждая кривая плотности всегда имеет одну-единственную медиану; среднее арифметическое она может не иметь, но если имеет, то одно; мод же кривая плотности может иметь сколько угодно, но может и не иметь ни одной.

В данной книге, посвященной исключительно совокупности измерений и выводам из них, мы не будем рассматривать многовершинные (полимодальные) кривые, но кривые, лишенные мод, будем рассматривать, так же как и кривые, лишенные среднего значения.

В тех случаях, когда кривая имеет все три указанных центра, последовательность их расположения обычно такова: среднее и мода занимают крайние положения, медиана располагается между ними.

§ 11. Моменты высших порядков

Рассмотрены основные характеристики ряда случайных величин — центр и рассеяние. Как увидим в дальнейшем, во многих случаях этого бывает достаточно, но не всегда. В теории ве-

роятности используют более общее понятие моментов, частными случаями которых являются вышеприведенные характеристики. Моментом называется математическое ожидание некоторой степени от разности членов ряда с какой-то условно взятой постоянной величиной

$$\nu_r = M.O. (x - a)^r. \quad (20)$$

Особый интерес представляют два случая: 1) когда a равно нулю, т. е. возводятся в степень сами члены ряда, моменты называются начальными; 2) когда $a = \bar{x}$, т. е. берется ряд отклонений членов ряда от их среднего, моменты называются центральными. Введем для этих моментов особое обозначение

$$\mu_r = M.O. (x - \bar{x})^r.$$

Между начальными и центральными моментами нетрудно установить следующие зависимости:

$$\mu_1 = M.O. (x - \bar{x}) = 0; \quad (21)$$

$$\mu_2 = M.O. (x - \bar{x})^2 = \nu_2 - \nu_1^2; \quad (22)$$

$$\mu_3 = M.O. (x - \bar{x})^3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3; \quad (23)$$

$$\mu_4 = M.O. (x - \bar{x})^4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4. \quad (24)$$

Приведенные переходные формулы имеют силу как для моментов конкретной совокупности (эмпирические моменты), так и для случая, когда от конкретной, выборочной, совокупности переходим к генеральной совокупности, представленной теоретическим законом распределения (теоретические моменты). В первом случае символ математического ожидания выразится знаком суммы, а во втором — знаком интеграла.

Но весьма важно отметить, что эмпирические моменты существуют для любой конкретной совокупности (это слишком очевидно!); теоретические же моменты для некоторых функций могут отсутствовать. В дальнейшем изложении этот сложный вопрос будет подробно рассмотрен. Здесь же полезно привести хотя бы один пример. В симметричном распределении Коши, плотность которого выражается через

$$p(x) = \frac{1}{\pi(1+x^2)},$$

обе ветви уходят в бесконечность, аналогично общеизвестному нормальному распределению

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Однако если второе распределение имеет моменты любого порядка, то первое распределение не имеет ни одного момента выше нулевого порядка (равного, очевидно, единице).

Действительно, имеем для $r=1$

$$\int_{-\infty}^{+\infty} \frac{xdx}{\pi(1+x^2)} = \frac{\ln(1+x^2)}{\pi} \Big|_0^{\infty} = \infty.$$

Можно без труда доказать то же самое и для $r=2, 3, \dots$

В математической статистике доказывается общее положение, что если r -й момент не существует, то все моменты выше r -го также не существуют. Это доказательство здесь не будет приведено.

С моментами выше четвертого порядка нам не придется иметь дела, но в первых четырех моментах необходимо разобраться. Согласно предыдущему, величина

$$\mu_2 = D(X) = \overline{(x-a)^2} = \sigma^2$$

является дисперсией нашей совокупности.

Если отклонения отдельных членов ряда расположены симметрично относительно центра распределения, то третий центральный момент (как и все нечетные центральные моменты) исчезает. Отсюда следует, что отличное от нуля значение третьего момента может служить мерой асимметрии. Из соображений размерности обычно вводят для характеристики асимметрии безразмерное отношение

$$S = C_s = \frac{\mu_3}{\frac{\mu_2^{\frac{3}{2}}}{\mu^2}} = \frac{\mu_3}{\mu^{\frac{3}{2}}}, \quad (25)$$

которое называют иногда коэффициентом асимметрии C_s , а иногда просто «асимметрией» S . Иногда бывает удобно пользоваться не стандартом, а его отношением к среднему

$$C_v = \frac{\sigma}{x},$$

называемым «коэффициентом вариации».

Точно так же момент четвертого порядка часто делят на квадрат дисперсии и вводят особую характеристику — эксцесс, имеющую вид

$$E = \frac{\mu_4}{\sigma^4} - 3$$

и характеризующую степень сглаженности кривой плотности вероятности. Но так как для нормального распределения дробь $\frac{\mu_4}{\sigma^4}$, как будет доказано в главе V, равна 3, то и вводят эту характеристику в указанной двучленной форме, при которой эксцесс нормального распределения оказывается равным нулю.

Выше рассмотрены лишь два рода моментов: моменты, отнесенные к началу распределения, — начальные, и моменты, отнесенные к среднему арифметическому, — центральные.

Но так как в дальнейшем будем иногда пользоваться другим «центром», а именно медианой, то важно для этих случаев дать соответствующие определения и обозначения, что нетрудно сделать, так как при использовании медианы моменты порядка выше первого особого интереса не представляют, а нулевой момент, как известно, не зависит от выбора положения центра. Поэтому в дальнейшем введем лишь следующие обозначения: абсциссу точки медианы будем обозначать нулевым индексом, например, для случайной величины x среднее равно \bar{x} , а медиана — x_0 . Первый абсолютный момент аналогично обозначим через σ_0 (или δ_0)

$$\sigma_0 = \int_0^{\infty} |x - x_0| p(x) dx. \quad (26)$$

Запишем тождество

$$\int_0^{\infty} |x - x_0| p(x) dx = \int_0^{x_0} (x_0 - x) p(x) dx + \int_{x_0}^{\infty} (x - x_0) p(x) dx.$$

Вынося постоянную x_0 за знак интеграла, будем иметь

$$\sigma_0 = x_0 \int_0^{x_0} p(x) dx - \int_0^{x_0} xp(x) dx + \int_{x_0}^{\infty} xp(x) dx - x_0 \int_{x_0}^{\infty} p(x) dx.$$

Однако первый и четвертый интегралы при нахождении медианы дадут в сумме нуль. Определив остальные, получим

$$\sigma_0 = \bar{x} - 2 \int_0^{x_0} xp(x) dx. \quad (27)$$

§ 12. Производящие функции

Прежде чем излагать в следующих главах теорию функций распределения, целесообразно дать читателю представление о производящих функциях, находящихся в неразрывной связи с теорией распределения.

Производящей функцией случайной переменной x или ее распределения называется математическое ожидание величины e^{xt} . Для дискретных распределений имеем

$$g(t) = M.O. e^{tx} = \sum P(x) e^{tx}, \quad (28)$$

а для непрерывных

$$g(t) = M.O. e^{tx} = \int_{-\infty}^{+\infty} e^{tx} p(x) dx. \quad (29)$$

Докажем, что производящая функция суммы независимых случайных величин равна произведению производящих функций слагаемых.

Имеем сумму величин

$$X = X_1 + X_2 + \dots + X_k + \dots + X_n = \Sigma X_k,$$

каждая из которых имеет производящую функцию $g_{x_k}(t)$.

Запишем по предыдущему выражение для производящей функции суммы:

$$g_x(t) = M.O. e^{t \Sigma x_k}.$$

Но поскольку величины x_k независимы, то математическое ожидание произведений равно произведению математических ожиданий. Следовательно,

$$g_x(t) = \Pi g_{x_k}(t). \quad (30)$$

Выражение (29) дает возможность определить производящую функцию по заданной плотности вероятности. Но известное преобразование Фурье¹ позволяет, наоборот, выразить плотность вероятности через характеристическую функцию, в результате чего будем иметь

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-tx} g(t) dt. \quad (31)$$

Таким образом, плотность вероятности и производящая функция взаимно и однозначно друг друга определяют, что очень важно, так как во многих случаях найти производящую функцию оказывается легче, чем плотность вероятности. Покажем теперь, что моменты любого порядка также легче находить из производящих функций.

Возьмем r -ую производную по t от производящей функции для дискретных и для непрерывных распределений:

$$\begin{aligned} \frac{d^r}{dt^r} g(t) &= \sum_1^{\infty} P(x_j) x_j^r e^{tx_j}, \\ \frac{d^r}{dt^r} g(t) &= \int_{-\infty}^{+\infty} x^r p(x) e^{tx} dx, \end{aligned}$$

приравняем нулю переменную t и получим в обоих случаях

$$\left\{ \frac{d^r}{dt^r} g(t) \right\}_{t=0} = \nu_r. \quad (32)$$

¹ См., например, Привалов И. И. Ряды Фурье, ГТТИ, М.—Л., 1934.

РАСПРЕДЕЛЕНИЕ ДИСКРЕТНОЙ СОВОКУПНОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН

§ 13. Биномиальное распределение

Решим следующую задачу. Известно, что в какой-то определенной комбинации условий событие A может произойти с вероятностью p . Вероятность того, что это событие не произойдет в этих условиях, будет $q=1-p$. Надо определить вероятность того, что при n испытаниях (т. е. при n раз повторенных однородных условиях) событие A повторится m раз. При этом будем считать, что порядок появления и не появления события A нас не интересует.

Если бы этот порядок представлял интерес, то вероятность, например, следующего порядка: $A, \bar{A}, \bar{A}, A, A, A, \bar{A}, \bar{A}, A$ была бы такой $p \cdot q \cdot q \cdot p \cdot p \cdot p \cdot q \cdot q \cdot p$ или в общем виде приняла бы выражение $p^m q^{n-m}$.

При неучете порядка эта вероятность будет, очевидно, больше в C_n^m раз, где $C_n^m = \frac{n!}{m!(n-m)!}$. Отсюда получаем окончательное выражение для биномиального распределения в виде

$$P_n^m = C_n^m p^m q^{n-m}. \quad (33)$$

Это распределение называется «биномиальным» потому, что представляет собой (в чем нетрудно убедиться) разложение бинома Ньютона $(p+q)^n$ на сумму из $(n+1)$ слагаемых, соответствующих значениям $m, m=0, 1, 2, \dots, n$.

Попутно заметим, что поскольку, согласно определениям, $p+q=1$, то и эта сумма равна единице, что и соответствует тривиальной истине — сумма вероятностей всех возможных событий равна достоверности, т. е. единице.

Отметим еще, что коэффициенты C_n^m , не зависящие от веро-

ятности p , всегда имеют определенную систему значений, составляющую так называемый «треугольник Паскаля»

$$\begin{array}{cccccc}
 & & & & & 1 \\
 & & & & 1 & & 1 \\
 & & 1 & & 2 & & 1 \\
 & 1 & & 3 & & 3 & & 1 \\
 1 & & 4 & & 6 & & 4 & & 1
 \end{array}$$

и т. д.

В частном случае, когда $p=q=\frac{1}{2}$, все распределение стано-

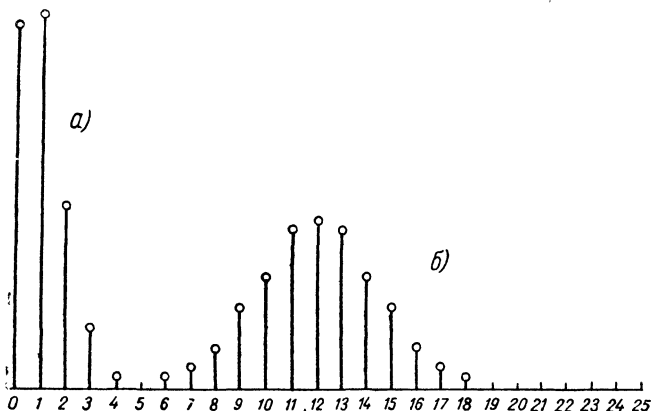


Рис. 5. Распределение по биномиальному закону.

а—при $n = 24$ $p = \frac{1}{2}$; б—при $n = 24$ $p = \frac{1}{24}$.

вится симметричным и выражается упрощенной формулой

$$P_n^m = C_n^m \frac{1}{2^n}.$$

В дальнейшем убедимся, что биномиальное распределение как в общем случае, так и в частном хорошо соответствует схеме накопления элементарных ошибок при измерениях физических величин.

Рассмотрим два крайних предела распределения (33).

Поскольку $p+q=1$, то высшим значением p можно считать $\frac{1}{2}$, так как при $p > \frac{1}{2}$ имеем $q < \frac{1}{2}$, и как только заменим p и q , снова получим $p \leq \frac{1}{2}$. При нижнем пределе для p , взятом сколь угодно близким к нулю, распределение (33) сохраняет

свое значение. В пределе же при $p=0$ и соответственно $q=1$ получаем достоверное непоявление события A , как это и следует из самого уравнения, когда при введении в него $p=0$ и соответственно $m=0$ имеем $C_n^0=1, 0^0=1$, а значит, $P_n^0=1$; а при $p=0$ и других m будем иметь $P_n^m=0$.

Для пояснения основных свойств биномиального распределения на рис. 5 приведен график этого распределения для двух случаев: а) при $n=24, p=\frac{1}{2}$ и б) при $n=24, p=\frac{1}{24}$.

В первом случае распределение будет симметричным и приближенным к закону Гаусса, рассмотренному в следующей главе. Во втором случае распределение приближается к закону Пуассона, который будет изложен в данной главе.

Для дальнейшего анализа биномиального распределения составим для него производящую функцию

$$\begin{aligned} g(t) &= M.O. e^{Xt} = \sum_{m=0}^{m=n} e^{mt} C_n^m p^m q^{n-m} = \\ &= \sum_0^n C_n^m (pe^t) q^{n-m} = (pe^t + q)^n. \end{aligned} \quad (34)$$

Для вычисления моментов берем, как изложено в § 10, значения производных от функции (28) при $t=0$:

$$\nu_1 = [n(pe^t + q)^{n-1} pe^t]_{t=0} = np; \quad (35)$$

$$\begin{aligned} \nu_2 &= [n(n-1)\{pe^t + q\} p^2 e^{2t} + n\{pe^t + q\}^{n-1}]_{t=0} = \\ &= n^2 p^2 + npq; \end{aligned} \quad (36)$$

$$\nu_3 = (np)^3 + 3q(np)^2 + npq(q-p); \quad (37)$$

$$\nu_4 = (np)^4 + 6n^3 p^3 q + n^2 p^2 q(7q-4p) - npq(1-6pq). \quad (38)$$

Пользуясь далее соотношениями (21) — (24), получим сначала значения центральных моментов, а из них значения среднего, среднеквадратического, асимметрии и эксцесса:

$$\mu_1 = 0; \quad (39)$$

$$\mu_2 = npq, \quad \sigma = \sqrt{npq}; \quad (40)$$

$$\mu_3 = npq(q-p); \quad (41)$$

$$S = \frac{q-p}{\sqrt{npq}} = \frac{1-2p}{\sqrt{npq}}; \quad (42)$$

$$\mu_4 = 3(npq)^2 + npq - 6n(pq)^2; \quad (43)$$

$$E = \frac{1-6pq}{npq}. \quad (44)$$

Рассмотрим снова два крайних случая, представленные в виде графиков на рис. 5, на котором в первом случае, когда $p = \frac{1}{2}$,

$$\sigma = \frac{\sqrt{n}}{2}, \quad S = 0, \quad E = -\frac{2}{n},$$

а во втором случае, когда p очень мало, $q \approx 1$,

$$\sigma = \sqrt{np}, \quad S = \frac{1}{\sqrt{np}}, \quad E = +\frac{1}{np}.$$

В первом случае при возрастании n к бесконечности полученные значения соответствуют закону Гаусса, при котором, как увидим в следующей главе, асимметрия и эксцесс равны нулю.

Во втором случае, если принять, что при p очень малом, а n очень большом произведение np порядка единицы, то полученные значения σ , S , E будут точно соответствовать закону Пуассона, к рассмотрению которого переходим ниже.

§ 14. Закон Пуассона

Рассмотрим частный случай биномиального закона, при котором p очень мало, а n очень велико, но их произведение представляет сравнительно небольшую величину:

$$\lambda = np. \quad (45)$$

Введенное ограничение предрешает применимость этого частного случая для определения вероятностей очень редких событий. В дальнейшем будем сохранять конечность значений λ даже в пределе, при $p \rightarrow 0$, $n \rightarrow \infty$.

Уравнение (33) после замены $p = \frac{\lambda}{n}$, $q = 1 - \frac{\lambda}{n}$ примет сначала вид

$$P_n^m = \frac{n!}{m!(n-m)!} \left(\frac{\lambda}{n}\right)^m \left(1 - \frac{\lambda}{n}\right)^{n-m},$$

а затем вид

$$P_n^m = \left[\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{m-1}{n}\right) \right] \left(1 - \frac{\lambda}{n}\right)^{n-m} \frac{\lambda^m}{m!}.$$

Но при сделанных предположениях все множители, заключенные в прямые скобки, стремятся к единице, а множитель $(1 - \frac{\lambda}{n})$, хотя и близкий к единице, но возведенный в чрезвычайно высокую степень $n - m \approx n$ приближается, согласно известному определению числа e , к $e^{-\lambda}$. В результате мы получаем

$$P(m) = \frac{e^{-\lambda} \lambda^m}{m!}. \quad (46)$$

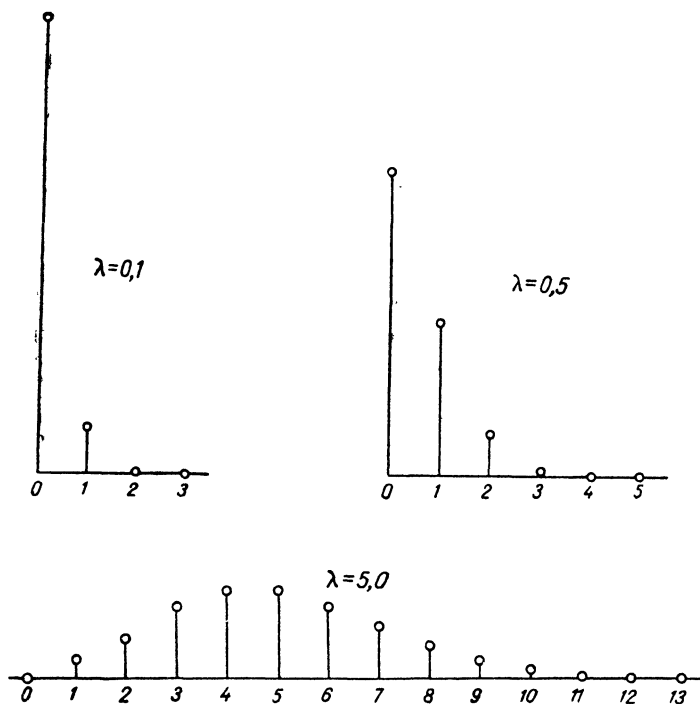


Рис. 6. Графики распределения Пуассона.

Это и есть закон Пуассона, в котором отсутствует число испытаний n , предполагаемое бесконечным (счетным), и вероятность p , предполагаемая исчезающе малой.

Отметим, что основное требование, предъявляемое к каждой функции распределения дискретных величин и заключающееся в том, чтобы сумма его членов равнялась единице именно в пределе, при $n \rightarrow \infty$, оправдывается

$$\sum_0^{\infty} \frac{e^{-\lambda} \lambda^m}{m!} = e^{-\lambda} \sum_0^{\infty} \frac{\lambda^m}{m!} = 1.$$

Огромная область плодотворного применения этого распределения к самым разнообразным «редким» событиям и экспериментальная его оправданность в тех случаях, когда оказывалось возможным осуществить сравнение теоремы с экспериментом, дают основание считать распределение Пуассона вполне закономерным.

Для иллюстрации расположения ординат выражения (46) на рис. 6 приведены три графика распределения, соответствующие значениям $\lambda=0,1$; $0,5$; $5,0$. На графиках видно, что при значениях $\lambda < 1$ наибольшая ордината (мода) имеет место в начальной точке распределения, т. е. при $m=0$. Далее, когда $\lambda=1$, ординаты моды и среднего ($m=0$ и $m=1$) равны. После этого начинается резкое падение ординат. И, наконец, при $\lambda > 1$ мода перемещается ко все большим и большим значениям. В частности, на графике при $\lambda=5,0$ начальная ордината составляет (с округлением) всего лишь $0,007$, а мода и среднее дают ординаты, равные $0,175^1$. Попутно отметим одно общее свойство распределения Пуассона, которое заключается в том, что, начиная со значения $\lambda=1$, ординаты моды и среднего всегда равны, а их абсциссы m всегда отличаются на единицу

$$P_{m-\lambda}(m) = P(m-1). \quad (47)$$

Для доказательства этого запишем равенство

$$P(m-1) = \frac{e^{-\lambda} \lambda^{m-1}}{(\lambda-1)!} = \frac{e^{-\lambda} \lambda^m}{\lambda(\lambda-1)!} = P(m).$$

Необходимо помнить ту абстракцию, которая была применена при выводе закона Пуассона, а именно $n \rightarrow \infty$, $p \rightarrow 0$, откуда следует, что закон может быть справедлив только при очень больших n и очень малых p . Анализ и проверка полностью подтверждают указанное ограничение.

А. Н. Колмогоровым [5, стр. 50] было дано более точное приближение, пригодное и без ограничения. Но здесь оно не приводится, так как в дальнейшем будем пользоваться законом Пуассона лишь при тех ограниченных условиях, которые полностью соответствуют его выводу.

Из производящей функции закона Пуассона, имеющей вид

$$g(t) = M.O. \left\{ \frac{e^{mt} e^{-\lambda} \lambda^m}{m!} \right\} = e^{\lambda e^t - \lambda}, \quad (48)$$

получаем начальные моменты:

$$v_1 = \lambda; \quad (49)$$

$$v_2 = \lambda^2 + \lambda; \quad (50)$$

$$v_3 = \lambda^3 + 3\lambda^2 + \lambda; \quad (51)$$

$$v_4 = \lambda^4 + 6\lambda^3 + \lambda^2 + \lambda. \quad (52)$$

¹ Все числовые данные заимствованы из работы [5].

Переходя к центральным моментам и от них к остальным характеристикам распределения, получим выражения:

$$\sigma = \sqrt{\lambda}; \quad S = \frac{1}{\sqrt{\lambda}}; \quad E = \frac{1}{\lambda}, \quad (53)$$

тождественные выражения биномиального распределения при очень малых значениях p и рассмотренные нами в предыдущем разделе. Это, между прочим, доказывает, что биномиальное распределение с уменьшением p очень быстро стремится к своему пределу — закону Пуассона, что увеличивает надежность пользования им в обычных условиях.

§ 15. Примеры применения закона Пуассона

Биномиальный закон, пределом которого при убывании p является закон Пуассона, отвечает общей схеме многократного повторения события при неизменной его вероятности. Разница между общей схемой и частным случаем закона Пуассона лишь в том, что в общей схеме число испытаний всегда конечно, а в частном случае оно условно считается сколь угодно большим: $n \rightarrow \infty$, $p \rightarrow 0$, $np = \text{const}$.

Далее будем считать, что параметр λ означает среднее число появления события A в течение периода времени t .

Наша основная предпосылка о постоянстве вероятности и о независимости этой вероятности от предшествующих появлений того же события дает право считать число его появлений пропорциональным времени. В этом случае, если обозначим через τ среднее значение интервала между двумя смежными появлениями события, получим

$$\lambda = \frac{t}{\tau} \quad (54)$$

(само событие условно считается «мгновенным»).

Числитель t — величина, зависящая от нашего выбора. Если, например, наблюдается последовательность вызовов на телефонной станции в течение определенной части дня, когда вероятность числа вызовов может считаться постоянной, то из достаточно длительного наблюдения (в те же часы, но за много суток) можно довольно точно определить значение параметра τ . Приняв произвольное значение периода t , можно получить соответствующее значение этого параметра. При этом очевидно, что в ночные часы, когда вызовов меньше, τ будет больше, чем днем, а значит $\frac{t}{\tau}$ будет меньше, в служебные же часы τ будет меньше, чем вечером. Во всяком случае, применимость закона

Пуассона при расчетах эксплуатации телефонной сети вполне установлена и успешно используется.

Другим аналогичным примером расположения во времени мгновенных и независимых событий может быть явление распада радиоактивного вещества. Испускание α -частиц при распаде атома происходит практически мгновенно. Расположение вылетов частиц во времени, как показывает наблюдение, носит вполне случайный характер. Если будем наблюдать это явление в течение достаточно длительного периода t , а затем разделим этот период на число вылетевших частиц, то получим значение

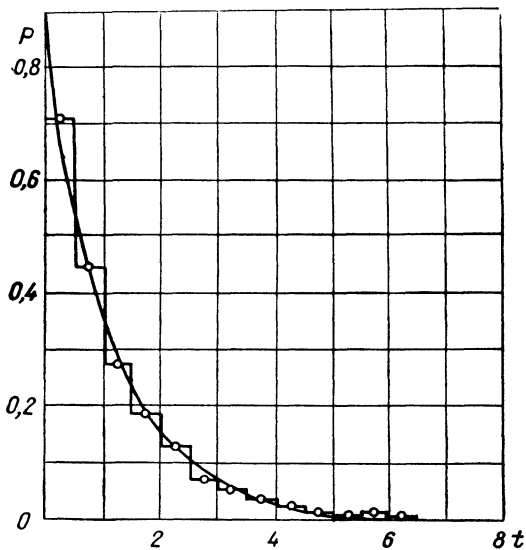


Рис. 7. Степенной закон распределения вылета альфа-частиц.

τ для данного вещества в данных условиях. Закон Пуассона, написанный теперь в виде

$$P_t(m) = \frac{\left(\frac{t}{\tau}\right)^m e^{-\frac{t}{\tau}}}{m!}, \quad (55)$$

будет отвечать на вопрос: какова вероятность того, что в течение периода t произойдет m вылетов α -частиц? В частности, можно задать и такой вопрос: какова вероятность того, что в течение периода t не вылетит ни одной частицы? Тогда при $m=0$ будем иметь более простое равенство

$$P_t(0) = e^{-\frac{t}{\tau}}, \quad (56)$$

где независимой переменной будет период t .

Теперь возьмем производную вероятности P по времени

$$\frac{dP}{dt} = -\frac{1}{\tau} e^{-\frac{t}{\tau}}, \quad (57)$$

которая должна быть отрицательной, так как с увеличением периода t указанная вероятность уменьшается. Надо отметить, что модуль этой производной с увеличением t также уменьшается, как это видно на рис. 7, где представлены результаты экспериментального подсчета, произведенного Хэллидеем [12]. Эмпирическая ступенчатая линия на графике показывает падение частот интервалов между последовательными вылетами α -частиц. Плавная кривая проведена, согласно уравнению (56), при вычисленном из эксперимента значении $\tau=1,169$ сек. Совпадение плавной и ступенчатой кривых достаточно хорошее. Кривая начинается на ординате $P=1$, что и соответствует достоверности того, что на нулевом отрезке времени ни одного вылета не может произойти. Теоретическая кривая падает до бесконечности, но практически нулевая ордината начинается уже на отрезке, равном 6 сек.

Некоторые примеры применимости закона Пуассона будут даны позже.

РАСПРЕДЕЛЕНИЕ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

§ 16. Плотность вероятности

В предыдущем изложении иногда, т. е. в тех случаях, когда необходимо было распространить рассуждения на непрерывные случайные величины, применялся термин «плотность вероятности». Переходя теперь к детальному анализу функций распределения непрерывных случайных величин, мы должны понятие плотности вероятности рассмотреть более подробно.

Назовем функцией распределения случайной непрерывной величины X вероятность того, что она примет любое значение, меньшее (или равное) x

$$P(X \leq x) = F(x).$$

Если функция $F(x)$ задана, то закон распределения будет известен. В частности, когда легко можно найти вероятность того, что эта величина принимает любое значение внутри интервала от x_1 до x_2 , то

$$P\{x_1 < X < x_2\} = F(x_2) - F(x_1).$$

Далее, уменьшая длину интервала $(x_2 - x_1) = \Delta x$, получим отношение разности значений функции распределения к длине интервала в пределе при уменьшении последнего до нуля. Это и будет плотность вероятности, имеющая вид

$$p(x) = \lim_{\Delta x \rightarrow 0} \left[\frac{F(x + \Delta x) - F(x)}{\Delta x} \right] \quad (58)$$

Так как правая часть равенства представляет собой не что иное, как производную функции распределения, то мы можем записать следующее соотношение:

$$p(x) = \frac{dF(x)}{dx}.$$

И обратно имеем

$$F(x) = \int_{-\infty}^x p(x) dx \quad (I)$$

или

$$F(x) = \int_0^x p(x) dx \quad (II)$$

в зависимости от того, принимает ли величина x все значения, как положительные, так и отрицательные (I), или только положительные (II).

Из вышеизложенного ясно, что функция $F(x)$ монотонна по существу и не может принимать отрицательных значений. Следовательно, плотность вероятности, так же как ее производная, всегда положительна.

Все изложенное относится к распределению одномерных случайных величин. Но так как придется уже в этой главе соприкасаться иногда и с двумерными совокупностями, поэтому оставим подробное рассмотрение вопросов, относящихся к двумерным и многомерным совокупностям, до изложения одной из следующих глав и осветим лишь частный вопрос о плотности вероятности двумерной совокупности при независимых ее компонентах.

Пусть x и y — две независимые случайные величины, представленные рядами, полученными по измерениям x_j, y_j . Функции их распределения даны в виде

$$P(X \leq x) = F_1(x); \quad P(Y \leq y) = F_2(y).$$

Переходя к бесконечно малым интервалам, получим

$$\frac{dF_1}{dx} = p_1(x)$$

и соответственно

$$\frac{dF_2}{dy} = p_2(y),$$

откуда будем иметь

$$\varphi(x_1 y) = p_1(x) p_2(y). \quad (59)$$

В дальнейшем придется иногда делать переход от плотности вероятности x к плотности вероятности какой-нибудь функции $f(x)$. Для такого перехода имеем очевидное соотношение

$$p_2(y) dy = p_1(x) dx$$

или

$$p_2(y) = p\{x(y)\} \frac{dx}{dy}. \quad (60)$$

§ 17. Понятие о гамма-функции

В теории распределения непрерывных случайных величин большую роль играет гамма-функция. В настоящем разделе дадим о ней самое элементарное представление, но достаточное для понимания дальнейшего изложения.

Нетрудно показать, что определенный интеграл

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt \quad (61)$$

при всех целых и положительных значениях x обращается в хорошо известный факториал от $(x-1)$. Применяя интегрирование по частям, непосредственно получаем

$$\int_0^{\infty} e^{-t} t^{x-1} dt = e^{-t} \frac{t^x}{x} \Big|_0^{\infty} + \frac{1}{x} \int_0^{\infty} e^{-t} t^x dt = \frac{1}{x} \Gamma(x+1),$$

откуда следует основное равенство

$$x\Gamma(x) = \Gamma(x+1), \quad (62)$$

которое для целых положительных x выражает определяющее свойство факториала

$$x! = x(x-1)!$$

Если теперь отказаться от ограничения целочисленности x , то уравнение (61) дает общее определение гамма-функции для любого положительного значения x .

Следует обратить внимание, что показатель при параметре t под знаком интеграла на единицу меньше, чем переменная в левой части, в связи с чем иногда более удобно отсчитывать независимую переменную не от нуля, а от единицы. В этом случае для целых x имеем

$$\Gamma(1+x) = x!$$

Из равенства (61) непосредственно получим $\Gamma(1) = 1$, а из равенства (62) также непосредственно будем иметь все остальные целочисленные значения $\Gamma(x)$:

$$\Gamma(2) = 1\Gamma(1) = 1;$$

$$\Gamma(3) = 2\Gamma(2) = 2$$

и т. д.

Из дробных значений x элементарным путем легко получить все значения, кратные половине. Действительно, имеем известный интеграл

$$\int_0^{\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}.$$

Подставляя в этот интеграл $x^2 = t$, преобразуем его к виду

$$\int_0^{\infty} e^{-t} t^{-\frac{1}{2}} dt = \sqrt{\pi}.$$

В левой части полученного интеграла по изложенному имеем гамма-функцию от $\frac{1}{2}$. Следовательно,

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Все остальные значения, кратные половине, можно очень легко получить из равенства (62):

$$\Gamma\left(\frac{3}{2}\right) = \Gamma\left(1 + \frac{1}{2}\right) = \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2};$$

$$\Gamma\left(\frac{5}{2}\right) = \frac{3}{4} \sqrt{\pi};$$

$$\Gamma\left(\frac{7}{2}\right) = \frac{5}{2} \Gamma\left(\frac{5}{2}\right) = \frac{15}{8} \sqrt{\pi}$$

и т. д.

Другие значения гамма-функции, кроме целочисленных и кратных половине, могут быть получены лишь числовыми методами. Ниже приведены значения x от 0 до 4 (табл. 2).

Таблица 2

x	$\Gamma(1+x)$	x	$\Gamma(1+x)$	x	$\Gamma(1+x)$	x	$\Gamma(1+x)$
0	1,000	1,0	1,000	2,0	2,000	3,0	6,000
0,1	0,951	1,1	1,046	2,1	2,197	3,1	6,811
0,2	0,918	1,2	1,102	2,2	2,424	3,2	7,757
0,3	0,897	1,3	1,167	2,3	2,684	3,3	8,857
0,4	0,887	1,4	1,242	2,4	2,981	3,4	10,135
0,5	0,886	1,5	1,329	2,5	3,323	3,5	11,631
0,6	0,893	1,6	1,427	2,6	3,710	3,6	13,350
0,7	0,909	1,7	1,545	2,7	4,172	3,7	15,436
0,8	0,931	1,8	1,676	2,8	4,693	3,8	17,833
0,9	0,962	1,9	1,828	2,9	5,301	3,9	20,674
						4,0	24,000

Для того чтобы продолжить эту таблицу, надо последовательно перемножать последующие $\Gamma(1+x)$ на предыдущие $(1+x)$. Общий вид кривой гамма-функции, построенной по данным таблицы, приведен на рис. 8.

В заключение дадим одно важное соотношение, относящееся к логарифму гамма-функции. Имея

$$\lg \Gamma(x+1) = \lg \Gamma(x) + \lg x$$

и взяв производные от обеих частей, получим

$$\frac{\Gamma'(x+1)}{\Gamma(x+1)} = \frac{\Gamma'(x)}{\Gamma(x)} + \frac{1}{x}. \quad (63)$$

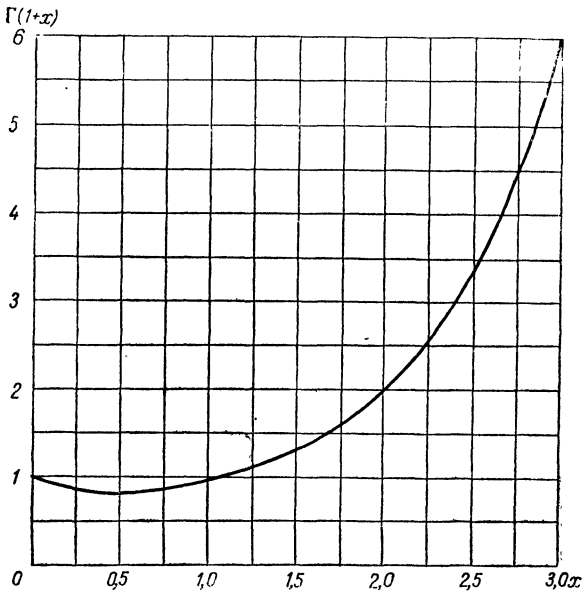


Рис. 8. График гамма-функции.

§ 18. Интерполяция биномиального распределения

В этом разделе математической статистики, т. е. в теории ошибок, непрерывные распределения имеют гораздо большее применение, чем прерывные, хотя в ряде случаев, которые позднее будут рассмотрены, измеряемая величина и отдельные измерения выражаются целыми числами, а само измерение сводится к подсчету. Однако основная и наиболее часто встречающаяся практика измерений показывает, что в результате измерений получается ряд дробей, число знаков которых зависит от точности измерителя и принципиально может быть сколь угодно большим, а два смежных измерения могут (тоже лишь принципиально!) отличаться на какую угодно малую величину. Поэтому при сколь угодно большом числе измерений мы вправе рассчитывать на получение последовательности чисел, заполняющих любой интервал, т. е. практически представляющих непрерывную совокупность.

В таком случае генеральная совокупность наших измерений, которая только мыслима, но технически недостижима, должна

изобразиться какой-то кривой распределения плотностей вероятности. В поисках такой кривой, естественно, нужно обратиться прежде всего к биномиальному закону, отвечающему как раз именно той схеме возвращаемых шаров, которая, как будет ясно из дальнейшего, лучше всего соответствует схеме накопления и суммирования положительных и отрицательных ошибок.

Следует задать себе вопрос: нельзя ли этот, по существу, дискретный закон интерполировать?

Если биномиальный закон распределения представить на графике, как это было нами сделано на рис. 5 для двух предельных случаев: вполне симметричного и сильно асимметричного распределения, то мы получим одновершинный ряд столбиков сперва возрастающей, а затем убывающей высоты, отстоящих друг от друга на интервалы, которые можно принять равными, хотя в дискретном распределении размер интервалов не имеет никакого значения. Совершенно очевидно даже при взгляде на два распределения, изображенные на рис. 5, что если будем увеличивать число испытаний n , но одновременно во столько же раз будем уменьшать длину промежуточных интервалов (обозначим их через ε), то сближение столбиков приведет в пределе при $n \rightarrow \infty$, $\varepsilon \rightarrow 0$ к непрерывной кривой, проходящей по их вершинам.

Примем, согласно сделанному выше предположению, что полная длина распределения от $m=0$ до $m=n$ при возрастании n (а значит, и m) остается постоянной и равной допустим l :

$$n\varepsilon = l.$$

Представим себе сначала ступенчатую кривую с очень малой постоянной шириной ступенек ε и тоже с малой, но с переменной высотой ступенек, равной, очевидно,

$$\Delta P = P_{m+1} - P_m.$$

Ищем выражение для отношения $\frac{\Delta P}{P}$, в котором за знаменатель можно принять среднее арифметическое из двух смежных ординат

$$P = \frac{P_m + P_{m+1}}{2},$$

откуда получим выражение

$$\frac{\Delta P}{P} = 2 \frac{P_{m+1} - P_m}{P_{m+1} + P_m},$$

в котором правую часть надо вычислить непосредственно из уравнения биномиального закона (33). Не приводя здесь весьма простых выкладок, напишем

$$\frac{P_{m+1}}{P_m} = \frac{p}{m+1} : \frac{q}{n-m},$$

откуда имеем

$$\frac{\Delta P}{P} = 2 \frac{pn - m - q}{pn + (1 - 2p)m + q}.$$

Перейдем затем от дискретной независимой переменной к другой, тоже дискретной, x с помощью равенства

$$x_m = m\varepsilon + \frac{\varepsilon}{2}.$$

Здесь индекс m при x означает последовательность расстояний средин ступенек от начала координат: $m=0$ до $m=n-1$.

Число ступенек будет, очевидно, на единицу меньше числа испытаний.

В таком изображении существует, конечно, известная условность, так как ни начало распределения, лежащее в начале координат, ни конец распределения, лежащий от него в расстоянии $n\varepsilon$, в этот ряд не войдут. Но по мере уменьшения ε (и соответственно возрастания n) эта условность будет постепенно исчезать. Подставляя значение m в равенство (64) и умножая числитель и знаменатель на ε , получаем

$$\frac{\Delta P}{P} = \frac{x_m - q\varepsilon - pn\varepsilon}{\frac{pn\varepsilon}{2} + \left(\frac{1}{2} - p\right)x_m + \frac{q\varepsilon}{2}}.$$

Обратим внимание на то, что разность ΔP , отнесенная нами при любом возрастании n к единице абсциссы, в пределе становится равной производной. Переходя к пределу, замечаем, что только величины $x = m\varepsilon$ и $n\varepsilon$ остаются конечными и в пределе при $\varepsilon \rightarrow 0$; величины же, образованные произведением ε на конечные величины, очевидно, исчезнут. Поэтому, переходя к непрерывному распределению от вероятности P к плотности вероятности $p(x)$, получим

$$\frac{1}{p(x)} \frac{dp}{dx} = - \frac{x - pn\varepsilon}{\left(p - \frac{1}{2}\right)x + \frac{pn\varepsilon}{2}}. \quad (64)$$

Отметим попутно, что множитель $\left(p - \frac{1}{2}\right)$, при симметричном распределении обращающийся в нуль, будет нами использован в следующем параграфе. Здесь же принимаем его отрицательным, что соответствует случаю $p < q$, т. е. положительной асимметрии.

Произведя деление в предыдущем уравнении и интегрируя, мы получим после переноса отсчета к началу координат и ряда простых операций следующее уравнение:

$$p(x) = Ce^{-\frac{x}{\beta}} x^\alpha,$$

в которое вводим для сокращения написания два параметра:

$$\alpha = \frac{n\varepsilon pq}{\left(p - \frac{1}{2}\right)^2}; \quad \beta = \frac{1}{2} - p.$$

Значение постоянной интегрирования получим из условия, что $\int_0^{\infty} p(x) dx = 1$, а также из известного определенного интеграла:

$$\int_0^{\infty} x^{\alpha} e^{-\frac{x}{\beta}} dx = \Gamma(\alpha + 1) \beta^{\alpha+1},$$

откуда имеем

$$\Gamma(\alpha + 1) \beta^{\alpha+1} = 1$$

и окончательно получаем так называемое гамма-распределение:

$$p(x) = \frac{x^{\alpha} e^{-\frac{x}{\beta}}}{\Gamma(\alpha + 1) \beta^{\alpha+1}}, \quad (65)$$

являющееся общей формой интерполяции биномиального закона, но с одной оговоркой: для тех ограничительных допущений, которые приводят биномиальный закон к закону Пуассона, эта форма оказывается неприменимой, так как одновременно стремятся к нулю оба множителя ε и p , входящие в состав параметра α .

Когда множитель p приводится к нулю даже при $n \rightarrow \infty$, то получаем лишь частный случай гамма-распределения (он же является и частным случаем распределения Пуассона), который рассмотрен в § 15.

$$p(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}. \quad (66)$$

Распределения (65) и (66) начинаются в точке $x=0$ и уходят в положительную бесконечность. Последнее обстоятельство находится как будто в противоречии с принятым с самого начала условием конечности произведения $n\varepsilon=l$. Но дело в том, что величина l , никогда далее не фигурирующая, может принимать любые (конечные) значения, а поэтому ничего не мешает приписать ей значение практически достаточно большое, чтобы охватить ту конкретную совокупность случайных величин, для которой это распределение применяется.

Впрочем, некоторое противоречие в распространении биномиального закона на совокупность случайных величин, прини-

мающих сколь угодно большие значения, все же остается (мы увидим это снова на выводе нормального закона!), однако ни какого практического значения это не имеет, так как вероятно сти очень больших значений ничтожно малы.

Составим теперь для распределения (65) производящую функцию

$$g(t) = M.O. e^{xt} = \frac{1}{\Gamma(\alpha + 1) \beta^{\alpha + 1}} \int_0^{\infty} x^{\alpha} e^{-\frac{x}{\beta} + xt} dx.$$

Заменяя под интегралом

$$x(1 - \beta t) = \xi; \quad dx = \frac{d\xi}{1 - \beta t},$$

получим

$$g(t) = \frac{\Gamma(\alpha + 1) \beta^{\alpha + 1}}{\Gamma(\alpha + 1) \beta^{\alpha + 1}} \frac{1}{(1 - \beta t)^{\alpha + 1}},$$

которое после всех сокращений будет иметь вид

$$g(t) = (1 - \beta t)^{-(\alpha + 1)}. \quad (67)$$

Последовательными производными этой функции будут

$$\frac{dg}{dt} = \beta(\alpha + 1)(1 - \beta t)^{-(\alpha + 2)};$$

$$\frac{d^2g}{dt^2} = \beta^2(\alpha + 1)(\alpha + 2)(1 - \beta t)^{-(\alpha + 3)};$$

$$\frac{d^3g}{dt^3} = \beta^3(\alpha + 1)(\alpha + 2)(\alpha + 3)(1 - \beta t)^{-(\alpha + 4)};$$

$$\frac{d^4g}{dt^4} = \beta^4(\alpha + 1)(\alpha + 2)(\alpha + 3)(\alpha + 4)(1 - \beta t)^{-(\alpha + 5)},$$

откуда получаем четыре первых начальных момента:

$$\left. \begin{aligned} \mu_1 &= \beta(\alpha + 1) \\ \mu_2 &= \beta^2(\alpha + 1)(\alpha + 2) \\ \mu_3 &= \beta^3(\alpha + 1)(\alpha + 2)(\alpha + 3) \\ \mu_4 &= \beta^4(\alpha + 1)(\alpha + 2)(\alpha + 3)(\alpha + 4) \end{aligned} \right\} \quad (68)$$

Соответственные центральные моменты получаем по формулам (21) — (24):

$$\left. \begin{aligned} \mu_1 &= 0 \\ \mu_2 &= \beta^2(\alpha + 1) \\ \mu_3 &= 2\beta^3(\alpha + 1) \\ \mu_4 &= 3\beta^4(\alpha + 1)(\alpha + 3) \end{aligned} \right\} \quad (69)$$

Отсюда имеем

$$\sigma_x = \beta \sqrt{\alpha + 1}, \quad C_v = \frac{1}{\sqrt{\alpha + 1}}, \quad -S = C_s = \frac{2}{\sqrt{\alpha + 1}},$$

$$E = \frac{\delta}{\alpha + 1} \quad (70)$$

или, решая два первых равенства относительно α и β , получим

$$\alpha = \frac{4}{S^2} - 1, \quad \beta = \frac{\delta S}{2}. \quad (71)$$

Отметим, что параметр α безразмерен, а параметр β имеет размерность величины x . Пределы значений параметра α легко определить из равенства (70). Действительно, при любой конечной асимметрии параметр α будет больше отрицательной единицы, которая является его нижним пределом, так как при $\alpha < -1$ мы получаем «мнимую асимметрию», что лишено смысла.

Что касается параметра β , входящего знаменателем при числителе x , то он по своему смыслу может принимать любые конечные и только положительные значения.

В частном случае $\alpha = 0$ мы имеем из тех же равенств $C_v = 1$, а само распределение, как уже было сказано выше, принимает тривиальный степенной вид.

Дадим для гамма-распределения положение моды. Приравняв к нулю производную

$$\frac{dp(x)}{dx} = \frac{1}{\Gamma(\alpha + 1) \beta^{\alpha + 1}} \left(\alpha x^{\alpha - 1} e^{-\frac{x}{\beta}} - \frac{x^\alpha}{\beta} e^{-\frac{x}{\beta}} \right),$$

получаем простое равенство

$$x_0 = \alpha \beta, \quad (72)$$

а так как среднее, как видно из равенств (68), равно $\beta(\alpha + 1)$, то мода расположится влево от среднего, что при положительной асимметрии вообще обязательно.

Рассмотренное распределение имеет два параметра: α и β , но нетрудно убедиться, что вид кривой зависит только от первого из них, второй же играет в нем лишь роль масштаба, в котором представлена распределенная величина. Это видно хотя бы из того, что в вышеприведенной таблице начальных и центральных моментов параметр β входит везде множителем в степени порядка момента.

Произведем замену переменной

$$\xi = \frac{x}{\beta} = \frac{x\bar{\sigma}_x}{\sigma_x^2}.$$

По общим правилам пересчета

$$p_1(x) dx = p_2(\xi) d(\xi)$$

составим выражение для распределения случайной величины ξ

$$p_2(\xi) = \frac{\xi^\alpha e^{-\xi}}{\Gamma(\alpha + 1)}. \quad (73)$$

Правая часть даст нам искомое распределение. Известно, что любой ряд случайных величин после указанного преобразования приобретает новое свойство — его дисперсия становится равной его математическому ожиданию. Это, очевидно, относится и к непрерывной случайной величине.

Пересчитаем моменты для полученного нового распределения:

$$\left. \begin{aligned} \nu_1 &= \alpha + 1 \\ \nu_2 &= (\alpha + 1)(\alpha + 2) \\ \nu_3 &= (\alpha + 1)(\alpha + 2)(\alpha + 3) \\ \mu_1 &= 0 \\ \mu_2 &= \alpha + 1 \\ \mu_3 &= 2(\alpha + 1) \end{aligned} \right\} \quad (74)$$

Интеграл правой части равенства (73) носит название «неполной гамма-функции»

$$\int_0^x \frac{x^\alpha e^{-x} dx}{\Gamma(\alpha + 1)},$$

которая, таким образом, является интегралом гамма-распределения при $\beta = 1$.

При $\alpha = 0$ это распределение, как известно, превращается в простую степенную зависимость, также имеющую практическое применение в некоторых специальных случаях теории ошибок.

§ 19. Нормальное распределение

Переходим теперь к выводу другого частного предельного случая биномиального распределения, принимая первым предельным случаем тот самый закон Пуассона, рассмотренный нами в предыдущей главе и относящийся к предельной асимметрии, когда вероятность p чрезвычайно мала, а противоположная ей вероятность q очень близка к единице. Распределение Пуассона сохраняет дискретность основного биномиального распределения, так как его интерполяция встречает значительные принципиальные затруднения. Но другой предельный случай, когда $p = q$ (а следовательно, обе они равны $\frac{1}{2}$), такую интерполяцию допускает и тем самым получает особое значение в тех слу-

чаях, когда случайная величина может принимать любые значения от $-\infty$ до $+\infty$, располагающиеся симметрично относительно центра распределения.

В следующей главе будет выведен тот же закон из основного свойства распределения случайных ошибок. Он носит название нормального. Сейчас представим его лишь как частный случай той интерполяции биномиального закона, которая в общем виде приводит к гамма-распределению.

Действительно, стоит лишь ввести в вывод, полученный в § 18, основное условие симметрии $p=q=\frac{1}{2}$, как весь вывод значительно упростится: в уравнении (64) исчезнет первое слагаемое, и можно будет получить очень простое уравнение

$$\frac{1}{y} \frac{\Delta y}{\Delta x} = - \frac{x - pn\varepsilon}{\frac{pn\varepsilon}{2}}.$$

Это распределение пока еще сохраняет свою дискретность. Известно, что в случае симметричности основного биномиального распределения его среднее арифметическое равно $pn = \frac{n}{2}$, а дисперсия равна $npq = \frac{n}{4}$. Следовательно, переходя к континууму, при котором мы полагаем $n \rightarrow \infty$, $\varepsilon \rightarrow 0$, а произведение $n\varepsilon$ равно конечной величине, можем записать

$$pn\varepsilon \rightarrow \frac{n}{2} \rightarrow \bar{x};$$

$$\frac{pn\varepsilon}{2} \rightarrow \frac{n\varepsilon}{4} \rightarrow \sigma_x^2,$$

в результате чего предыдущее равенство примет вид

$$\frac{1}{y} \frac{dy}{dx} = - \frac{(x - \bar{x})^2}{\sigma^2},$$

откуда после интегрирования будем иметь

$$y = Ce^{-\frac{(x - \bar{x})^2}{2\sigma^2}}.$$

Для определения постоянной C введем новую переменную $t = \frac{x - \bar{x}}{\sigma}$ и используем известный определенный интеграл¹

$$\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt = \sqrt{\pi}.$$

¹ Рыжик И. М. и Градштейн И. С. Таблицы интегралов, сумм рядов и произведений. ГТТИ, М.—Л., 1951.

Окончательно получаем выражение для нормального закона

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}, \quad (75)$$

по которому симметричное распределение уходит своими ветвями в положительную и отрицательную бесконечность. Центр распределения, в котором среднее арифметическое, очевидно совпадает с медианой и с модой, находится в абсциссе \bar{x} . Но здесь сталкиваемся с одним противоречием, которое для ясности дальнейшего изложения надо разъяснить заранее.

Нормальное распределение, так же как и выше рассмотренное гамма-распределение, получено путем интерполяции из одного и того же биномиального закона, в котором число благоприятных исходов m является, очевидно, существенно положительной величиной. Между тем в полученном симметричном распределении имеется отрицательная ветвь, которая также уходит в бесконечность. Следовательно, отклонения от центра влево, большие, чем x , должны попасть в область отрицательных x , что, очевидно, лишено смысла.

В исходном, биномиальном, распределении этого внутреннего противоречия не могло быть, так как при сколь угодно большом, но конечном числе испытаний и при среднем значении $M. O. m = np$ наибольшее отклонение от центра влево не могло перейти в отрицательную область значений: ее просто не существовало. Откуда же она могла появиться после интерполяции? Исчерпывающий ответ на этот вопрос можно дать лишь после того, как будет приведен самостоятельный вывод нормального закона. Сейчас покажем лишь, что указанное противоречие полностью снимается, если только допустить, что число испытаний n также возрастает к бесконечности, как это было принято для закона Пуассона. Но там вероятность p была принята убывающей в той же степени, что и возрастание n , а произведение np оставалось конечным. Здесь же вероятность p остается постоянной и равной $\frac{1}{2}$, а произведение np , равное $\bar{x} = \frac{n}{2}$, стремится к бесконечности. Отметим, что стандарт такого распределения также стремится к бесконечности:

$$\sigma = \sqrt{npq} = \frac{\sqrt{n}}{2}.$$

Таким образом, при теоретически беспредельном, а практически достаточно большом числе испытаний n получаем симметричные отклонения в обе стороны от чрезвычайно большого среднего, причем, несмотря на то что по-прежнему разность $|x - \bar{x}|$ при $|x| > \bar{x}$ становится отрицательной, но вероят-

ность таких значений x уже настолько исчезающе мала, что их наличием можно совершенно свободно пренебречь! Таким образом, приходим к практической возможности использования полученного интерполяцией биномиального закона симметричного распределения (75) во всех тех случаях, когда среднее значение ряда достаточно велико. При этом, правда, мы вынуждены отказаться от только что указанного требования, заключающегося в том, чтобы стандарт был равен половине среднего значения. Однако при замене фиктивного $n = \infty$ очень большим n такое условие уже становится практически приемлемым.

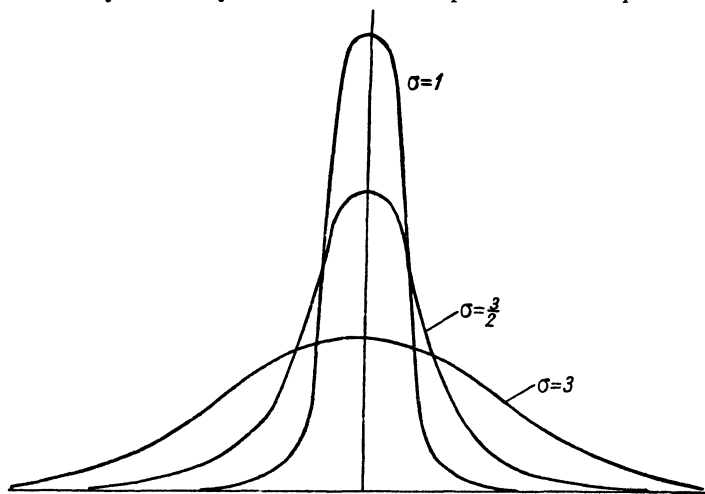


Рис. 9. График нормального распределения при σ , равной 1, $\frac{3}{2}$ и 3.

Приведенное рассуждение хорошо подтверждается графиком, изображенным на рис. 9. На графике видно, что на сравнительно небольшом расстоянии от центра распределения вероятность отклонений очень быстро уменьшается. Кроме того, простой расчет показывает, что вероятность отклонений, больших, чем при стандарте, будет меньше, чем $3 \cdot 10^{-3}$, что при практических расчетах вполне можно принять за «нуль».

Практически кривая Гаусса за пределами $x > 3\sigma$ как бы «обрывается». И если поставим условие, что истинное значение измеряемой величины, приблизительно совпадающее со средним значением многочисленных ее измерений, было больше тройного стандарта измерительного прибора, то такое условие практически очень легко выполнить. Можно даже сказать, что оно в большинстве случаев перевыполняется! Однако следует отметить, что тем самым попутно получаем один практически

весьма важный вывод о неприменимости нормального закона в тех случаях, когда значение среднего ненормально больше значения стандарта. На этот вывод в дальнейшем будем всегда опираться.

Структура показателя в уравнении (75) $\frac{x - \bar{x}}{\sigma}$ дает все основания к нормированию этого распределения методом, указанным в § 9. Вводя обозначение

$$z = \frac{x - \bar{x}}{\sigma_x}$$

и соответственно

$$dz = \frac{dx}{\sigma_x},$$

получаем нормированное выражение закона Гаусса

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad (76)$$

которое не имеет параметров и может быть представлено в виде таблицы. Такая таблица имеется в курсах теории вероятностей. В данной книге она приводится с удвоенными значениями (табл. 2 приложения I).

Выведем теперь производящую функцию для нормального распределения:

$$g(t) = M.O. e^{xt} = \int_{-\infty}^{+\infty} \frac{e^{xt}}{\sqrt{2\pi} \sigma} e^{-\frac{(x - \bar{x})^2}{2\sigma^2}} dx$$

и сделаем опять ту же замену:

$$z = \frac{x - \bar{x}}{\sigma}, \quad x = \sigma z + \bar{x},$$

после чего

$$g(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{(\sigma z + \bar{x})t - \frac{z^2}{2}} dz.$$

Далее, преобразуя показатель

$$(\sigma z + \bar{x})t - \frac{z^2}{2} = \bar{x}t + \frac{t^2\sigma^2}{2} - \frac{1}{2}(z - t\sigma)^2,$$

получаем выражение

$$g(t) = \frac{1}{\sqrt{2\pi}} e^{\bar{x}t + \frac{t^2\sigma^2}{2}} \int_{-\infty}^{+\infty} e^{-\frac{(z - t\sigma)^2}{2}} dz.$$

Но так как второй интеграл нам известен; он равен $\sqrt{2\pi}$, то производящая функция нормального закона выразится через

$$g(t) = e^{\bar{x}t + \frac{\sigma^2 t^2}{2}}, \quad (77)$$

откуда известным нам путем получим все начальные моменты:

$$\begin{aligned} \nu_1 &= \bar{x}; \\ \nu_2 &= \bar{x}^2 + \sigma^2; \\ \nu_3 &= \bar{x}^3 + 3\bar{x}\sigma^2; \\ \nu_4 &= \bar{x}^4 + 6\bar{x}^2\sigma^2 + 3\sigma^4. \end{aligned}$$

Применяя формулы перехода к центральным моментам, получим

$$\mu_1 = 0; \quad \mu_2 = \sigma^2; \quad \mu_3 = 0; \quad \mu_4 = 3\sigma^4, \quad (78)$$

откуда непосредственно

$$S = 0; \quad E = 0. \quad (79)$$

Таким образом, в нормальном распределении как асимметрия, так и эксцесс равны нулю. Во всех же других они могут быть и положительными, и отрицательными.

В теории вероятностей доказывается теорема, принадлежащая А. М. Ляпунову и носящая название центральной предельной, которая утверждает, что сумма большого числа очень малых величин при безграничном возрастании их числа, т. е. в пределе, стремится к нормальному распределению, если только их асимметрия стремится к нулю [5, стр. 129—130].

Здесь, не доказывая приведенную выше теорему, покажем, что во всех до сих пор рассмотренных распределениях асимметрия и эксцесс при безграничном возрастании числа измерений n действительно стремятся к нулю. Подчеркиваем заранее, что этот факт не эквивалентен тому, что эти распределения при $n \rightarrow \infty$ стремятся к нормальному закону, поскольку последний, помимо указанного свойства, обладает еще и другими свойствами, в частности симметрией всех отдельных значений. Но все же доказательство указанного положения хотя бы лишь частично, но совпадает с центральной предельной теоремой.

Прежде всего отметим, что асимметрия и эксцесс основного биномиального распределения, выраженные, как известно (см. § 13), через

$$S = \frac{1-2p}{\sqrt{npq}}; \quad E = \frac{1-6pq}{npq}, \quad (80)$$

при безграничном возрастании n стремятся к нулю, причем важно отметить, что это имеет место при любом значении p , т. е. при любом числителе дроби, выражающей S (хотя очевидно,

что при p , близком к половине, и $(1-2p)$, близком к нулю, это стремление проявляется более отчетливо).

В общем случае интерполированного биномиального закона, т. е. в гамма-распределении, мы имели

$$S = \frac{2}{\sqrt{\alpha+1}}; \quad E = \frac{6}{\alpha+1}.$$

Подставляя сюда полученное в § 13 развернутое выражение для α

$$\alpha = \frac{n\epsilon pq}{\left(p - \frac{1}{2}\right)^2},$$

в котором произведение $n\epsilon$ при любых значениях p и q может быть сделано больше любой наперед заданной величины, мы также получаем

$$S \rightarrow 0, \quad E \rightarrow 0.$$

Из дальнейшего будет видно, что и некоторые другие распределения из рассматриваемых в этой книге, например «хиквадрат-распределение», распределение Стьюдента, также в пределе приобретают указанные свойства нормального распределения. Но все же, как увидим далее, существуют и такие распределения, которые и в пределе этому требованию не удовлетворяют, поскольку в них не только третий момент, но и другие моменты равны бесконечности, следовательно условию Ляпунова удовлетворять никак не могут.

Этот важный вопрос будет изложен в главе VIII в связи с рассмотрением тех измерений, отклонение которых от среднего значения имеет определенную асимметричность. Сейчас же, возвращаясь к нормальному закону, мы должны еще добавить, что при обработке измерений пользуются еще одной характеристикой рассеивания, а именно «вероятной ошибкой», введенной также Гауссом и представляющей для нас, как это будет ясно из дальнейшего, особый интерес. Под этим термином понимают ту возможную ошибку, вероятности превышения и уменьшения которой одинаковы. Ее значение определяется равенством

$$\int_0^m e^{-\frac{x^2}{2\sigma^2}} dx = \int_m^\infty e^{-\frac{x^2}{2\sigma^2}} dx.$$

Расчет показывает, что $m=0,6745\sigma$. Рассмотрим статистический смысл величины m .

Пусть мы имеет очень большое число измерений n , за их центр примем их среднее арифметическое, и предположим, что отклонения от этого центра расположены вполне симметрично. Тогда ничего не изменится, если все эти отклонения независимо от их знака расположим на числовой прямой вправо от условно взятой начальной точки. Вероятной ошибкой

будет значение этого преобразованного ряда, лежащее как раз на его середине, т. е. вероятность отклонения от которого вправо и влево одна и та же. Однако известно, что такая точка называется медианой, основное свойство которой заключается в том, что среднее значение модулей отклонения от нее меньше, чем от любой другой точки. Следовательно, получили весьма интересный результат: вероятное отклонение представляет собой одновременно также медиану ряда модулей отклонений.

Это положение связывает в известной степени те две единственно возможные комбинации центра и рассеивания, которые были рассмотрены в § 9 и 10.

Можно легко показать, что стандарт равняется абсциссе точки перегиба нормальной кривой. Для этого достаточно приравнять к нулю вторую производную распределения

$$\frac{d^2p(x)}{dx^2} = \left\{ -\frac{1}{\sqrt{2\pi} \sigma^3} + \frac{x^2}{\sqrt{2\pi} \sigma^5} \right\} e^{-\frac{x^2}{\sigma^2}} = 0$$

или после всех сокращений принять

$$x_1 = \pm \sigma,$$

где x_1 — абсциссы точек перегиба.

Общий вид нормального распределения, отнесенный к его центру, приведен на рис. 9 для значений стандарта $\sigma=1$, $\sigma=3/2$, $\sigma=3$. В частности, кривая для $\sigma=1$ представляет собой, очевидно, также нормированный вид этого распределения.

§ 20. Хи-квадрат (χ^2)-распределение

Исходя из нормированного нормального распределения случайной величины X

$$\varphi(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

зададимся другой случайной величиной $Y=X^2$, принимающей, очевидно, только неотрицательные значения.

Имеем $x^2=y$, $2xdx=dy$, $dx = \frac{dy}{2\sqrt{y}}$. Поскольку значения случайной величины X по условию подчинены нормальному закону, то этому же закону должно быть подчинено и распределение величин $+\sqrt{y}$ и $-\sqrt{y}$. Но так как плотности вероятности этих величин равны,

$$\varphi(+\sqrt{y}) = \varphi(-\sqrt{y}),$$

то плотность вероятности величины y должна равняться сумме указанных плотностей

$$\begin{aligned}\varphi(y) &= \frac{1}{2\sqrt{y}} \left\{ \frac{1}{\sqrt{2\pi}} e^{-\frac{(+\sqrt{y})^2}{2}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{(-\sqrt{y})^2}{2}} \right\} = \\ &= \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}.\end{aligned}$$

Запишем полученное выражение в ином виде:

$$\varphi(x) = \frac{1}{\Gamma\left(\frac{1}{2}\right) 2^{1/2}} x^{-\frac{1}{2}} e^{-\frac{x}{2}} \quad (81)$$

и сравним теперь его с уравнением гамма-распределения.

Совершенно ясно, что полученное распределение (81) является частным случаем гамма-распределения при $\alpha = -\frac{1}{2}$, $\beta = 2$.

Берем производящую функцию гамма-распределения для названных значений параметров

$$g(t) = (1 - 2t)^{-\frac{1}{2}} \quad (82)$$

и представим себе совокупность значений случайной величины x_i ($i=1, 2, 3, \dots, k$), каждая из которых представлена совокупностью значений, подчиненных нормальному распределению, с одними и теми же значениями \bar{x} и σ^2 .

Рассмотрим сумму их квадратов:

$$\chi^2 = \sum_1^k x_i^2.$$

Известно, что производящая функция суммы равна произведению производящих функций слагаемых. Следовательно, имеем для распределения случайной величины χ^2 производящую функцию

$$g(t) = (1 - 2t)^{-\frac{k}{2}}. \quad (83)$$

Соответствующая кривая плотности вероятностей выразится через

$$p(x) = \frac{1}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}}} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \quad (84)$$

где $x = \chi^2$.

Дифференцируя производящую функцию и приравнявая к нулю параметр t , получим четыре начальных момента:

$$\left. \begin{aligned} \nu_1 &= k \\ \nu_2 &= k(2+k) \\ \nu_3 &= k(2+k)(4+k) \\ \nu_4 &= k(2+k)(4+k)(6+k) \end{aligned} \right\} \quad (85)$$

Отсюда по равенствам (21—24) вычисляем центральные моменты:

$$\left. \begin{aligned} \mu_1 &= 0 \\ \mu_2 &= 2k \\ \mu_3 &= 8k \\ \mu_4 &= 12k(k+3) \end{aligned} \right\} \quad (86)$$

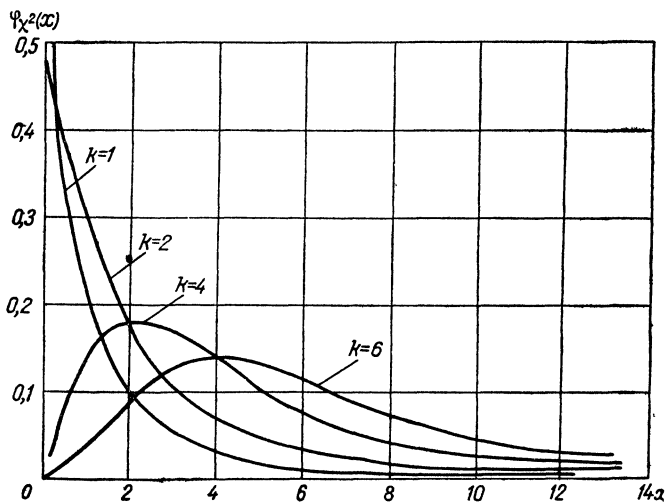


Рис. 10. График хикватрат-распределения при k , равном 1, 2, 4, 6.

а из них затем получаем выражения для стандарта, асимметрии и эксцесса:

$$\sigma = \sqrt{2k}; \quad S = \sqrt{\frac{8}{k}}; \quad E = \frac{12}{k}. \quad (87)$$

Форма кривой этого распределения зависит только от величины параметра k , который, как уже указывалось выше, является числом суммируемых квадратов, т. е. целым положительным числом, очевидно, отличным от нуля, так как значение $k=0$ не имеет смысла. При $k=1$, как показано на рис. 10, кривая, уходящая в начальной точке в бесконечность, опускается

сначала быстро (приблизительно до $x=4$), а затем медленно к нулю при $x \rightarrow \infty$.

При $k=2$ кривая вырождается в показательную функцию вида

$$p(x) = \frac{1}{2} e^{-\frac{x}{2}}, \quad (88)$$

начиная же со значения $k=3$, кривая исходит из начала координат и проходит через моду, абсциссу которой можно получить обычным путем, т. е. приравнявая к нулю первую производную

$$\frac{dp(x)}{dx} = 0.$$

Не приводя здесь простых выкладок, запишем непосредственно результат:

$$x_0 = k - 2.$$

После моды кривая стремится к нулю при $x \rightarrow \infty$.

Положение кривых для $k=1, 2, 4, 6$ представлено на рис. 10, заимствованном из работы [5], с добавлением кривой для $k=4$.

Это распределение, как и рассмотренное ранее гамма-распределение, относится, очевидно, к существенно положительным значениям величины x . Основное их отличие заключается в том, что в гамма-распределении имеются два параметра α и β , а в полученном распределении лишь один параметр k , притом целый численный. Этот параметр, который будет также встречаться при рассмотрении распределения Стьюдента, носит особое название — числа степеней свободы. Общий смысл этого термина будет объяснен несколько позже. В данном же случае значение параметра k будет на единицу меньше числа измерений, т. е. $k=n-1$.

Хи-квадрат-распределение имеет применение в вопросах оценки эмпирических дисперсий (см. главу VII).

Для понимания дальнейшего надо осветить следующий вопрос.

Имеем распределение некоторой случайной непрерывной величины x и знаем, что другая случайная величина y функционально зависит от переменной x . Какова будет плотность вероятности y ?

Рассуждаем следующим образом. Вероятность, что x принимает значения между x и $x+dx$, равна, как известно, $p(x) dx$. Но при функциональной связи между x и y такой же, очевидно, будет и вероятность того, что y принимает значения между y и $y+dy$. Отсюда имеем равенство

$$p_2(y) = p_1(x) \frac{dx}{dy}, \quad (89)$$

дающее ответ на поставленный вопрос.

Применим полученное равенство к решению следующей задачи.

Уравнение (84) было выведено для сумм квадратов нормированных величин. Найдем соответствующее распределение ненормированных величин.

Имеем по предыдущему

$$\chi^2 = \sum_1^k x_j^2,$$

откуда

$$y = \sigma^2 \chi^2 = \sum (\sigma x_j)^2.$$

Применяя далее уравнение (89) и учитывая соотношения между прежним x и новым y

$$x = \frac{y}{\sigma^2}; \quad \frac{dx}{dy} = \frac{1}{\sigma^2},$$

после подстановки вместо $p_1(x)$ распределения (84) получим следующее распределение, несколько от него отличное

$$p(x) = \frac{1}{2^{\frac{k}{2}} \sigma^k \Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{x}{2\sigma^2}}. \quad (90)$$

Таким образом мы вернулись к прежнему обозначению переменной распределения через x .

Далее, тем же путем найдем распределение среднего арифметического той же величины. Обозначим

$$y = \frac{\sigma^2 x}{k}; \quad \frac{dx}{dy} = \frac{k}{\sigma^2},$$

откуда после всех выкладок получаем

$$p(x) = \frac{\left(\frac{k}{2}\right)^{\frac{k}{2}}}{\sigma^k \Gamma\left(\frac{k}{2}\right)} x^{\frac{k}{2}-1} e^{-\frac{kx}{\sigma^2}}. \quad (91)$$

В дальнейшем нам понадобится распределение еще одной случайной величины

$$y = \frac{\sigma}{\sqrt{k}} \sqrt{x}; \quad x \equiv \chi^2.$$

Выразим x через y :

$$x = \frac{ky^2}{\sigma^2}; \quad \frac{dx}{dy} = \frac{2ky}{\sigma^2}.$$

Подставляя полученное значение x в равенство (91), умножая его на $\frac{dx}{dy}$, а затем меняя опять y на x , получаем искомое распределение в виде

$$p(x) = \frac{2 \left(\frac{k}{2}\right)^{\frac{k}{2}}}{\sigma^k \Gamma\left(\frac{k}{2}\right)} x^{k-1} e^{-\frac{kx^2}{2\sigma^2}}. \quad (92)$$

Уравнение (92) понадобится при выводе распределения Стьюдента.

§ 21. Распределение Стьюдента

Переходим к одному, весьма важному распределению, играющему основную роль в оценке среднего значения при нормальном распределении измерений.

Исходим из совокупности k независимых и нормально распределенных случайных величин, из которых образуем новую совокупность путем деления каждой величины этого ряда на рассмотренную в конце предыдущего параграфа величину

$$y = \frac{\sigma \sqrt{x}}{\sqrt{k}}; \quad x \equiv \chi^2,$$

распределение которой было получено (92).

Ищем закон распределения отношения

$$t = \frac{x}{y}. \quad (93)$$

Поскольку случайные величины x и y независимы, то можно утверждать, что их двумерная плотность вероятности равна произведению их плотностей вероятностей.

Величина x по условию распределена нормально; величина y распределена по закону (92). Следовательно, их двумерная плотность будет равна

$$\varphi(x, y) = C_n e^{-\frac{x}{2}} y^{k-1} e^{-\frac{ky^2}{2\sigma^2}}, \quad (94)$$

где коэффициент C_n содержит константы обоих распределений:

$$C_n = \sqrt{\frac{2}{\pi}} \frac{\left(\frac{k}{2}\right)^{\frac{k}{2}}}{\sigma^{k+1} \Gamma\left(\frac{k}{2}\right)}.$$

Полученная двумерная плотность представляет собой, очевидно, не кривую линию в плоскости, а кривую поверхность,

расположенную над плоскостью (x, y) и дающую для любого значения горизонтальных координат вертикальную ординату, выражающую отнесенную к единице площади вероятность попадания значения двумерной случайной величины в элементарный прямоугольник $(\delta x \delta y)$.

Но нас интересует не общая вероятность любых сочетаний переменных x и y , а вероятность распределения их дроби $t = \frac{x}{y}$.

Для этого из двумерной плотности мы должны получить одномерную плотность их отношения. Перейдем к новым переменным

$t = \frac{x}{y} > 0$, $u = y > 0$ и будем искать распределение t при любых значениях u . Для этого достаточно проинтегрировать найденное выражение, пересчитанное на новые переменные, т. е. на все возможные значения u от 0 до ∞ . Получим

$$p(t) = C_n \int_0^{\infty} e^{-\frac{u^2(k+t^2)}{2\sigma^2}} u^k du.$$

Известен определенный интеграл¹

$$\int_0^{\infty} e^{-px^2} x^{2a+1} dx = \frac{a!}{2p^{a+1}}.$$

В нашем случае надо приравнять

$$p = \frac{k+t^2}{2\sigma^2}; \quad 2a+1 = k.$$

Подставляя эти значения и заменяя, согласно определению гамма-функции

$$\left(\frac{k-1}{2}\right)! = \Gamma\left(\frac{k+1}{2}\right),$$

получим после очень простых выкладок следующее выражение:

$$S(t, k) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}, \quad (95)$$

являющееся распределением Стьюдента.

Обратим внимание на некоторые его свойства. Данное распределение симметрично при любых значениях k . Кроме того, можно доказать, что с возрастанием k оно асимптотически приближается к нормальному.

Действительно, из определения числа e имеем

$$\lim_{k \rightarrow \infty} \left[\left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}} \right] = e^{-\frac{t^2}{2}}.$$

¹ Рыжик И. М. и Градштейн И. С. Таблицы интегралов, сумм, рядов и произведений. ГТТИ, М.—Л., 1951, стр. 169.

Что же касается константы

$$\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}},$$

то из некоторых свойств гамма-функции можно было бы вывести вполне строго, что при $k \rightarrow \infty$ она стремится к $\frac{1}{\sqrt{2\pi}}$. Однако не будем проводить этого доказательства, так как и без него ясно,

что поскольку интеграл $\int_0^{\infty} \varphi(t) dt$ должен быть равен единице (по общим свойствам кривых распределения) и известно равен-

ство $\int_0^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{\frac{\pi}{2}}$, то вышеприведенная константа не мо-

жет не быть равной $\frac{1}{\sqrt{2\pi}}$.

Из других особенностей распределения Стьюдента следует особо подчеркнуть, что оно не содержит в явном виде ни среднего значения, ни стандартного отклонения основных величин, это связано с тем, что распределение было получено как отношение среднего значения и стандартного отклонения.

Общий вид кривой Стьюдента имеет сходство с нормальной кривой нормированного вида, и тем ближе к последней, чем больше k . Обычно считают, что при $k=30$ оба распределения практически совпадают. Но их расхождение при малых значениях k придает, как будет показано в следующей главе, особую ценность распределению Стьюдента именно при малом числе измерений. Для иллюстрации на рис. 11 приведен график (заимствованный у Г. Крамера) двух совмещенных кривых: нормированной нормальной и Стьюдента (при $k=3$). На графике видно, что вероятность малых отклонений по Стьюденту меньше, а больших отклонений больше, чем по Гауссу.

В дальнейшем нам может понадобиться рассмотрение нижнего предела распределения Стьюдента при $k=1$. Простой расчет даст из распределения (95) следующее выражение (напоми-

наем, что $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$):

$$S(t) = \frac{1}{\pi(1+t^2)}. \quad (96)$$

Поскольку отклонение Стьюдента от Гаусса растет с уменьшением параметра k , то наименьшее предельное выражение

Стьюдента, соответствующее значению $k=1$, должно давать также наибольшее отклонение от Гаусса.

Распределение (96) носит имя французского математика

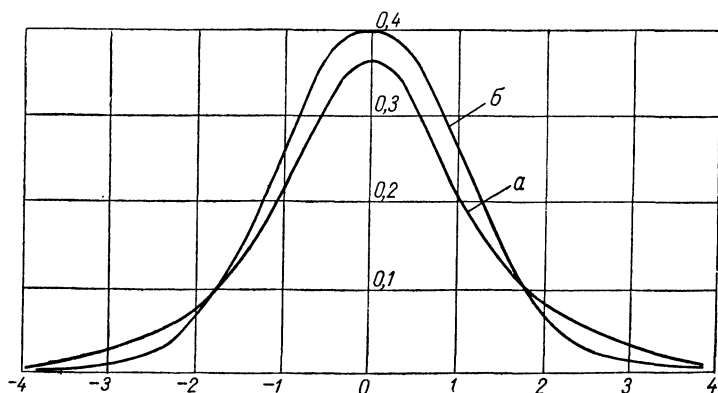


Рис. 11. График распределения Стьюдента при $k=3$ (а) и нормальное распределение (б).

Коши, потому что оно было получено им раньше, чем стало частным случаем распределения Стьюдента.

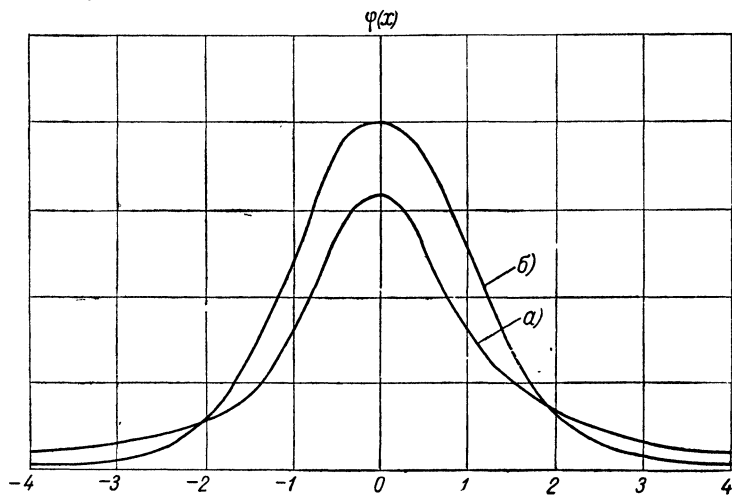


Рис. 12. График распределений Коши (а) и Гаусса (б).

Как видно, распределение Стьюдента имеет два предела: при $k=1$ оно превращается в распределение Коши, при $k=\infty$ — в распределение Гаусса.

Сравнение пределов приведено на рис. 12.

В заключение отметим (доказательство легко получит сам

читатель), что распределение Стьюдента имеет конечными только те моменты, порядок которых меньше k . Поэтому при $k=1$, т. е. для распределения Коши, как было показано, имеется лишь нулевой конечный момент (как всегда, равный единице); при $k=2$ имеется лишь первый момент (равный, очевидно, нулю), дисперсия же отсутствует. И только начиная с $k=3$ существует конечная дисперсия.

Указанное свойство распределения Стьюдента относится, очевидно, лишь к четным моментам порядка больше k -того, при которых они все равны бесконечности. Все же нечетные моменты любого порядка вследствие симметричности этого распределения должны быть равны нулю.

ОШИБКИ ИЗМЕРЕНИЯ

§ 22. Основные положения

Переходим к рассмотрению вопроса о природе ошибок, неизбежно сопутствующих любому измерению, а тем самым и любому эмпирическому исследованию. Последнее начинается, как правило, с измерения, проходит через стадию обработки полученных данных и заканчивается отысканием наилучшей формы зависимости между измеренными величинами. Рассмотрим процесс измерения какой-либо физической величины соответственным ей прибором. Для конкретности изложения приведем два простейших примера:

1. Измеряем длину стального тонкого стержня микрометренным циркулем. Для этого сначала раздвигаем ножки циркуля так, чтобы они по возможности совпали с концами стержня, а затем отсчитываем на отдельном масштабе (с нониусом или без него) длину между остриями циркуля.

2. Измеряем планиметром площадь фигуры, предварительно обведенной по ее контуру тонким карандашом. Для этого сначала обводим штифтом планиметра весь контур, а затем отсчитываем на валике угол суммарного его поворота.

Примеров измерений можно было бы привести любое количество, но так же, как и в вышеприведенных, всегда увидим две существенно различные операции: наложение измерителя на измеряемую величину и отсчет по шкале (связанной или не связанной с измерителем).

Проанализируем обе операции. Наложение измерителя никогда не может быть вполне точным: ножки циркуля даже при наличии микрометричного винта раздвинутся всегда или несколько больше, или несколько меньше, чем следовало бы по «истинному» размеру измеряемой величины. Получаются положительные и отрицательные отклонения, расположение которых

можно во многих случаях считать симметричным и, как увидим далее, приблизительно удовлетворяющим закону Гаусса.

Но для того чтобы из каждого отдельного измерения получить какую-то величину, выражаемую определенным числом, необходимо провести вторую операцию — наложить уже раздвинутый циркуль на масштаб. Здесь возникает затруднение — острие циркуля лишь для нашего невооруженного глаза представляется «точкой». Если же посмотрим на эту точку через лупу, то она окажется площадкой на искривленной и неправильной поверхности. В таком случае каждое отдельное измерение является, в сущности, не «величиной»; выражаемой числом, а каким-то не совсем определенным отрезком с концами, являющимися не точками, а «пятнами» или интервалами, внутри которых производить измерения уже невозможно. Эти интервалы представляют собой наименьшие величины, которые могут быть различимы при помощи данного измерителя.

Постараемся сформулировать все то, что мы вправе считать характеризующим понятие «интервала чувствительности». По определению концы того «отрезка», который выражает какое-либо отдельно взятое измерение, лежат внутри соответственных двух интервалов чувствительности. Но какое будет точное положение неизвестных «концов», мы не знаем и знать не можем. Мало того, у нас нет даже каких-либо оснований принять для вероятностей распределения возможных положений (неизвестной) точки внутри данного отрезка какую-либо кривую плотности вероятности. А предполагать эти вероятности равными вдоль всего отрезка так же мало оснований, как и принимать для них закон Гаусса. Единственное, о чем можно уверенно сказать, это то, что неизвестная точка лежит внутри такого-то отрезка.

Далее, мы имеем полное право считать, что размер этих отрезков, т. е. «интервалов чувствительности», одинаков для всех отдельных измерений, произведенных данным прибором. Иными словами, любое измерение данной серии представляет собой отрезок, концы которого находятся внутри интервалов одной и той же длины.

Принципиальная сторона введения в теорию ошибок понятия «интервала чувствительности» скажется только в том, что обычное утверждение (теорема Чебышева), будто бы при бесконечном числе измерений физической величины можно получить истинное ее значение, окажется не вполне точным. Строго говоря, при любом и даже бесконечно большом числе измерений всегда будут отклонения от истинного значения величины, размером не большим, но и не меньшим интервала чувствительности данного измерителя. Но каков этот размер, нам неизвестно.

Следовательно, можно сделать вывод, что если хотим получить серию более точных измерений, то надо перейти к измерителю, имеющему большую точность. Другого пути для получе-

ния более точных измерений нет. Увеличивать же с этой целью число измерений просто бессмысленно! Это, впрочем, слишком очевидно для любого экспериментатора или изыскателя, так как никому не придет в голову считать, например, что, измеряя какую-либо длину землемерной цепью бесконечное число раз, можно получить ее точное (!) значение. На практике мы всегда составляем таблицы измеренных величин с тем числом знаков, которое соответствует точности данного измерителя, причем знаем заранее, что то же самое число знаков будет иметь их среднее значение.

Перейдем теперь к рассмотрению названной таблицы измерений и будем считать ее цифры расположенными в возрастающем (или убывающем) порядке, так что, помимо среднего значения, также легко можем получить и медиану. По данным этой таблицы можно составить вторую таблицу, содержащую отклонения отдельных измерений от их центра. Не предвещая вопроса, примем ли за этот центр среднее арифметическое или медиану, будем сначала рассматривать эти отклонения в общем виде и установим для них два качественных постулата:

1. Меньшие отклонения должны иметь большую вероятность, чем отклонения более значительные.

2. Чем больше число измерений, тем их центр будет находиться ближе к истинному значению измеряемой величины (но, как отмечалось ранее, никогда с ним не совпадет).

Важно подчеркнуть, что любой закон, принятый в дальнейшем для распределения отклонений, обязан удовлетворять двум вышеприведенным качественным постулатам.

Вопрос, который надо рассмотреть хотя бы для общего случая, заключается в следующем. Совокупность измерений мы считаем совокупностью «случайных» величин, понимая, как всегда, под случайной физической величиной такую величину, значения которой зависят от огромного множества незначительных факторов, не учитываемых именно из-за своей незначительности. Можно ли утверждать, что все измерения всегда удовлетворяют этому определению случайных величин? Простые соображения приводят к отрицательному ответу.

Во-первых, в процессе измерения могут действовать также немалозначащие факторы: например, ошибочная регулировка прибора, влияние температуры на расширение измеряемого предмета, дальтонизм наблюдателя и многое другое. Отклонения, которые вызваны этими факторами, всегда возможно при известном влиянии обнаружить, исправить или учесть каким-либо способом. Такого рода ошибки принято называть систематическими. Совершенно очевидно, что никакому вероятностному закону распределения они подчиняться не будут, а поэтому надо их исключить раньше, чем переходить к отысканию закона распределения для остальных ошибок.

Во-вторых, в процессе измерения возможен простой недосмотр наблюдателя, неверный отсчет по шкале и т. п., а также иногда повреждение прибора, не замеченное наблюдателем. Такие факторы, хотя и несут тот же «случайный» характер (в несколько ином смысле), однако их эффект может выразиться в каких-либо чрезмерных, маловероятных, «выскакивающих» значениях измеряемой величины, они портят ту таблицу, в которую автоматически вписываются. Следовательно, их также надо обнаружить и исключить.

Наконец, во многих случаях, в особенности, когда путем измерения двух и более величин отыскивается эмпирическая зависимость между ними, приходится учитывать тот факт, что проведенное измерение соответствует не той «точке» пространства и не тому моменту «времени, к которым они относятся, а некоторой области пространства и некоторому промежутку времени. Последняя неточность сказывается, впрочем, не на самой обработке измерений, а лишь на оформлении тех зависимостей, которые получаются по измерениям двух и более взаимозависящих величин.

Что касается основной категории «случайных» ошибок, то для них должны быть установлены две качественно различные категории:

1. Когда вероятные ошибки измерительного прибора во много раз меньше размеров измеряемой величины, а следовательно, имеет место усиленное неравенство $\sigma \ll \bar{x}$.

2. Когда вероятные ошибки прибора имеют тот же порядок, что и размер измеряемой величины.

Поскольку соотношение между σ и x должно играть в обработке измерения доминирующую роль, то считаем целесообразным ввести для величин типа «а» термин «больших», а для типа «б» — «малых». Введенные к а в ы ч к и сохраним в дальнейшем, так как они должны подчеркивать условность терминов: величина может быть «большой» или «малой» не сама по себе, а лишь по отношению к стандарту измерителя, который применяется. Если при измерении той же величины перейдем от грубого измерителя к гораздо более точному, то она может в определенных условиях из категории «малых» перейти в категорию «больших».

Это важное различие лучше всего может быть показано на простейшем примере. Будем измерять длину какого-либо стержня (рис. 13). Точки измерений расположатся по какому-то закону по обе стороны от «истинного» значения длины. Теория (разовьем ее детально в следующей главе) доказывает, что при известных ограничительных условиях распределение отклонений будет удовлетворять нормальному закону Гаусса. Но дело в том, что этот закон допускает, как известно, отклонения любого размера в обе стороны: теоретически вплоть до плюс и ми-

нус бесконечности. Конечно, это является фикцией, так как известно, что практически в учете отклонений выходить за пределы «трех сигм», т. е. тройного стандарта, нет смысла. Но так как указанный стандарт зависит от качества измерительного прибора: чем прибор лучше, тем его стандарт меньше, то при измерителях невысокого качества длина трех стандартов может оказаться весьма значительной. Поэтому если длина измеряемого стержня значительно превышает трехстандартную длину, то ничто не мешает нам применять закон Гаусса, игнорируя те отклонения, вероятность которых практически ничтожна (рис. 13 а).

Но если размер стержня невелик, или измерительный прибор не очень высокого качества, или длина стержня меньше длины трех стандартов прибора, то отрицательная ветвь кривой Гаусса обрежется там, где она еще имела бы практическое значение. Следовательно, нормальный закон в этом случае не может быть применен: симметричность распределения становится невозможной, даже немислимой (рис. 13 б).

В этом случае надо искать для обработки измерений какой-то иной и обязательно асимметричный закон.

Но с другой стороны, по смыслу измерения требуется, чтобы суммарная вероятность всех превышений измеряемой величины равнялась суммарной вероятности всех ее занижений. Известно, что в асимметричном распределении среднее арифметическое такому условию не удовлетворяет; следовательно, при обработке измерений этой категории придется перейти от среднего к медиане, определяемой, как известно, равенством

$$\int_0^{x_0} p(x) dx = \int_{x_0}^{\infty} p(x) dx = \frac{1}{2},$$

которое по смыслу измерения мы обязаны удовлетворить.

В следующих двух главах будут последовательно рассмотрены принципиальные различия в обработке измерений «больших» и «малых» величин. Сейчас ограничимся лишь теми сторонами вопроса, которые являются общими для обеих указанных категорий.

В математической статистике различают генеральную совокупность, охватывающую полностью все рассматриваемые в данной конкретной задаче элементы, и выборку из этой

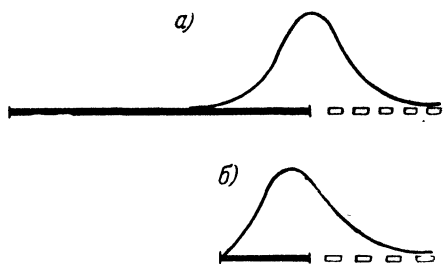


Рис. 13. Схема симметричного (а) и асимметричного (б) распределения ошибок.

генеральной совокупности, в которую входят лишь некоторые, случайно из нее выхваченные, элементы. Нам нет надобности излагать всю теорию выборочного метода и мы ограничимся общим утверждением, что в частном случае измерительной совокупности всегда имеем лишь выборку, репрезентативность которой зависит от ее объема. Чем большее число измерений произведено, тем ближе мы находимся к истинному значению измеряемой величины. Генеральной же совокупностью в данном случае является лишь абстрактная совокупность, включающая в себя бесконечное число (притом несчетное, непрерывное!) всех измерений, которые мы могли бы провести за бесконечный промежуток времени. Эту абстрактную генеральную совокупность можно представить себе лишь в виде теоретического закона распределения; и если бы (представим себе мысленно невозможный случай!) имелось в нашем распоряжении бесконечное число всех возможных измерений, то сама измеряемая величина (с точностью до «интервала чувствительности») могла бы быть вычислена.

Но имея всегда одну или хотя бы даже несколько выборок из генеральной совокупности, мы вынуждены сузить нашу задачу и вместо «истинного» значения измеряемой величины искать лишь те пределы, внутри которых она с данной «доверительной» вероятностью может находиться. Таким путем приходим к понятию оценки измеряемой величины, что равносильно нахождению доверительного интервала, внутри которого находится истинное значение измеряемой величины.

Практически принимают доверительную вероятность в каждой конкретной измерительной задаче в зависимости от качества измерительного прибора и строгости измерительной методики в пределах 0,90—0,99, но при очень грубых измерениях приходится идти и на более низкие ее значения.

Методы нахождения доверительного интервала по заданной доверительной вероятности будут изложены в соответствующих разделах книги. Здесь же мы ограничимся разбором одного вопроса, на котором все эти методы основаны: о взаимоотношении между стандартом выборки и стандартом генеральной совокупности.

Имеем n измерений величины:

$$x_1, x_2, x_3, \dots, x_j, \dots, x_n.$$

истинное значение которой равно a .

Среднее и выборочный (эмпирический) стандарт этих измерений будет

$$\bar{x} = \frac{\sum x_j}{n}; \quad s = \sqrt{\frac{\sum (x - x_j)^2}{n}}.$$

Составим две совокупности их отклонений: от a и от \bar{x} :

$$\varepsilon_j = x_j - a; \quad \xi_j = x_j - \bar{x}.$$

Суммируя оба ряда и вычитая вторую сумму из первой, получим

$$\Sigma \varepsilon_j - \Sigma \xi_j = n(\bar{x} - a).$$

Поскольку второе слагаемое левой части тождественно равно нулю, то

$$\bar{x} = a + \frac{\Sigma \varepsilon_j}{n}.$$

В результате сопоставления данного равенства с вышеприведенным равенством $\xi = \sqrt{\frac{\Sigma (x - x_j)^2}{n}}$ получим

$$\xi_j = x_j - a - \frac{\Sigma \varepsilon_j}{n} = \varepsilon_j - \frac{\Sigma \varepsilon_j}{n}$$

или если раскроем второе слагаемое, то запишем равенство

$$\xi_j = -\frac{\varepsilon_1}{n} - \frac{\varepsilon_2}{n} - \dots - \frac{\varepsilon_j(n-1)}{n} - \dots - \frac{\varepsilon_n}{n},$$

которое показывает, что эмпирические (выборочные) отклонения являются линейными функциями от истинных отклонений той же выборки.

Сравним теперь дисперсию конечного ряда измерений от истинного центра, т. е. от истинного значения измеряемой величины σ^2 , с эмпирической дисперсией того же ряда измерений от его среднего арифметического S^2 .

На основе вышеизложенного имеем n сумм, состоящих из n слагаемых:

$$- \xi_1 = \frac{n-1}{n} \varepsilon_1 + \frac{\varepsilon_2}{n} + \frac{\varepsilon_3}{n} + \dots + \frac{\varepsilon_n}{n};$$

$$- \xi_2 = \frac{\varepsilon_1}{n} + \frac{n-1}{n} \varepsilon_2 + \frac{\varepsilon_3}{n} + \dots + \frac{\varepsilon_n}{n};$$

$$\dots$$

$$- \xi_n = \frac{\varepsilon_1}{n} + \frac{\varepsilon_2}{n} + \frac{\varepsilon_3}{n} + \dots + \frac{n-1}{n} \varepsilon_n.$$

Все измерения считаются равнозначными (их дисперсии равны) и независимыми. Известно, что дисперсия линейной функции случайных величин равна их общей дисперсии, умноженной на сумму квадратов множителей при случайных величинах; поэтому из приведенных n сумм имеем

$$D(\xi) = s^2 = \sigma^2 \left\{ \left(\frac{n-1}{n} \right)^2 + (n-1) \left(\frac{1}{n} \right)^2 \right\},$$

откуда

$$s^2 = \sigma^2 \frac{n^2 - n}{n^2} = \sigma^2 \frac{n-1}{n}$$

и окончательно

$$\frac{s}{\sqrt{n-1}} = \frac{\sigma}{\sqrt{n}}. \quad (97)$$

Из равенства (97) следует, что при конечном числе измерений эмпирический стандарт всегда меньше теоретического стандарта, соответствующего распределению, которому подчинены эти измерения. Действительно, эмпирическая дисперсия, отсчитываемая от среднего арифметического, соответствует, как известно, принципу минимума суммы квадратов отклонений. Отклонения же отдельных измерений от истинного значения, которое при конечном числе измерений не является их средним арифметическим, этому принципу удовлетворять, очевидно, не могут. Необходимо подчеркнуть, что сказанное справедливо по отношению лишь конечного числа измерений; нормальный же закон справедлив, строго говоря, лишь по отношению бесконечного числа измерений. Следовательно, когда мы сопоставляем, как это всегда делается на практике, конечный ряд измерений с нормальным законом, то тем самым идем наперекор строгой логике.

Только в случае бесконечного ряда измерений (мыслимом, но неосуществимом) истинное значение измеряемой величины может быть одновременно также средним арифметическим всех произведенных измерений. Но тогда указанный принцип был бы снова оправдан, и в пределе

$$\lim \left(\frac{s}{\sigma} \right)_{n \rightarrow \infty} = \lim \sqrt{\frac{n-1}{n}} = 1.$$

Полученное чрезвычайно важное соотношение (97) показывает, что эмпирический стандарт s^2 надо относить не к полному числу измерений, а к уменьшенному на единицу. Это обстоятельство поможет нам понять, почему в тех вспомогательных распределениях, которые служат, как увидим далее, для статистических оценок истинных значений измеряемых величин, фигурирует не число n , а число $k = n - 1$, носящее название «числа степеней свободы» на основе соображений, которых мы для ясности вкратце коснемся.

Имеем, как всегда, n измерений x_j определенной физической величины. Составим из них линейную форму общего вида $L = a_1 x_1 + a_2 x_2 + \dots + a_j x_j + \dots + a_n x_n$. В частном случае при $a_1 = a_2 = \dots = \frac{1}{n}$ линейная форма L становится средним арифметическим, т. е. величиной, предположительно близкой к истин-

ному значению измеряемой величины. В общем случае таких линейных форм может быть несколько. Важно знать, сколько из них будет независимых, т. е. не выводимых одна из другой. Допустим, рассматриваем m линейных форм $L_i (i = 1, 2, 3, \dots, m)$. Если между формами существует линейная связь вида

$$\sum_{i=1}^{i=m} x_i L_i = 0$$

и если имеем, например, $\lambda_k \neq 0$, то, решая это уравнение относительно L_k , получим

$$L_k = - \sum \frac{\lambda_i}{\lambda_k} L_i,$$

где знак суммы распространяется, очевидно, на $(m - 1)$ слагаемых (так как одно слагаемое находится в левой части). Это значит, что имеем $(m - 1)$ независимых линейных форм, или, как говорят, $(m - 1)$ степеней свободы. Отсюда вытекает, что измерения, равные числу n , имеют не все n , а только $(n - 1)$ степеней свободы, иными словами, они не могут быть все «какими угодно», а связаны тем, что их среднее арифметическое выражает (хотя и приближенно) истинное значение измеряемой величины.

В математической статистике учение о «степенях свободы» измерительной совокупности занимает определенное место [5, стр. 307—308]. В данной книге будет изложено без доказательств из этого учения лишь то немногое, которое в дальнейшем потребуется.

Если рассматривать не линейную, а квадратическую форму, т. е. однородный многочлен второй степени, представляющий собой сумму квадратов вышеприведенной линейной формы, то число степеней свободы для нее можно считать равным числу степеней свободы исходных линейных форм. В этом случае доказывается особая теорема, сущность которой в том, что случайная величина χ^2 , представляющая сумму квадратов отклонений от центра, нормально распределенных с единичными дисперсиями случайных величин, будет распределена по ранее выведенному хиквадрат-распределению именно с тем числом степеней свободы k , которое характеризует исходную линейную форму.

Не приводя доказательства этой теоремы, здесь хотелось бы дать читателю о ней наглядное представление. Дело в том, что положение центра распределения обусловлено минимумом квадратов отклонений от этого центра. Такое условие создает одно ограничение как для исходной совокупности, так и для распределения квадратов отклонений от нее. Поэтому число степеней свободы должны будем принимать одинаковым для распределения

самих величин и для распределения квадратов их отклонений. Это положение будет использовано в § 27 при рассмотрении вопросов оценки среднего и стандарта.

§ 23. Систематические ошибки

В теории ошибок вопрос о систематических ошибках часто не рассматривается, поскольку действительно никакой «общей теории» для ошибок этого рода, очевидно, не может быть, так как все характеризуется здесь спецификой данной области науки и состоянием измерительной техники, используемой в этой области.

При рассмотрении вопроса о систематических ошибках мы также ограничимся лишь самыми общими положениями классификационного характера.

Систематические ошибки могут быть трех категорий: 1) зависящие от дефекта прибора (инструментальные ошибки), 2) ошибки, связанные с недоучетом влияния среды (ошибки среды), и 3) ошибки, обусловленные дефектом самого наблюдателя (личные ошибки).

Инструментальные ошибки в свою очередь делятся на три категории: 1) ошибки, обусловленные плохим качеством изготовления всего прибора или какой-либо его части, 2) ошибки, происходящие от погрешностей, связанных со сборкой и установкой прибора, и 3) ошибки, связанные с износом прибора в результате длительной его работы.

Поскольку дальнейшее изложение вопроса об ошибках построено на предположении, что систематические ошибки устранены, то казалось бы следует ответить сейчас лишь на два вопроса: 1) всегда ли возможно устранить ошибки и 2) если возможно, то как это сделать?

Однако здесь мы сталкиваемся с неопределенностью всей постановки, которая полностью лишает возможности четко ответить на первый вопрос и вынуждает ограничиваться тривиальными положениями о необходимости производить периодически тщательную проверку приборов всеми хорошо продуманными способами. Вообще вопрос об устранении инструментальных ошибок практически разрешается лишь с некоторой плохо определенной надежностью.

Ошибки, связанные с переучетом влияния среды, вероятно, можно устранить, если специально изучать влияния среды на показания прибора (например, путем измерения коэффициента теплового расширения базисного прибора) и составить таблицы поправок на эти систематические ошибки.

Наконец, что касается личных ошибок, то они зависят от наблюдателя, обладающего каким-либо дефектом. Для того чтобы избежать личных ошибок, наблюдателю надо изыскать

средства исправить дефект или отказаться от участия в тех измерениях, где этот дефект может играть существенную роль.

К вопросу систематических ошибок больше возвращаться не будем, так как считаем, что в каждом конкретном случае систематические ошибки могут быть «устранены», хотя, строго говоря, полной гарантии в их устранении никогда не может быть.

§ 24. Грубые (выскакивающие) ошибки

В ряду цифр, полученных измерением, иногда встречаются такие цифры, которые, отликаясь слишком сильно от всех остальных, вызывают особые подозрения в достоверности их получения. Такие «выскакивающие» цифры могут быть обусловлены двумя, качественно различными причинами.

Во-первых, тем, что данная серия измерений является, как известно, лишь ограниченной выборкой из генеральной совокупности.

В эту выборку всегда может случайно попасть такая цифра, частотность которой ничтожно мала. Очевидно, вероятность попадания такой цифры в данную серию измерений будет тем меньше, чем короче ряд измерений и чем резче эта цифра отличается от всех остальных. Однако события малой вероятности все-таки всегда возможны!

Во-вторых, в процессе измерения могло произойти случайное событие, повлиявшее на показание прибора. Например, при измерении скоростей в речном потоке гидрометрическая вертушка могла подвергнуться удару плывущей коряги или нивелировщик мог по рассеянности просчитать один метр на рейке. В обоих приведенных случаях выскакивающая ошибка обычно обнаруживается, а измерение или не учитывается, или исправляется. Но возможны случаи, когда обнаружить причину «выскакивания» очень трудно или даже невозможно, например когда между измерением и его обработкой прошел большой промежуток времени или сменился состав наблюдателей. В этих случаях возникает вопрос, что надежнее: включить такую цифру в подсчет среднего или ее выбросить. Причем из предыдущих рассуждений должно быть ясно, что выбрасываются не только измерения, на которые повлияла какая-либо случайная причина, но также и те измерения, теоретическая частотность которых настолько мала, что при данном числе измерений их появление в ряду измерений может считаться практически невероятным. Мы же, включая эту цифру в подсчет среднего, тем самым придаем ей больше доверия, чем она по своей маловероятности заслуживает.

По этому поводу будет уместно напомнить, что вообще оценку среднего, полученного из выборки, мы производим исходя лишь из числа измерений данной выборки, игнорируя при этом, что не исключена возможность попадания в эту выборку исключительно

малой или большой цифры, правда с очень малой вероятностью! Но вводить в оценку среднего эту маловероятную возможность его ошибочности мы не умеем, а поэтому должны просто решить, выбросить эту цифру или оставить. Для такого решения нужны какие-то более или менее надежные критерии. Подобные критерии в литературе имеются. Однако прежде чем перейти к их изложению, полезно напомнить, что существует такой центр совокупности, который а в т о м а т и ч е с к и исключает все выскакивающие как слишком малые, так и слишком большие цифры. Таким центром является хорошо известная нам м е д и а н а, использование которой необязательно связано с асимметричностью отклонений и которую мы вправе применять даже в тех случаях, когда генеральная совокупность наверное обладает симметрией. Но симметричность выборки вообще не очень обязательна, в отдельных случаях она может сильно нарушиться из-за присутствия хотя бы одной «выскакивающей» цифры. Во всяком случае, переход от среднего арифметического к медиане может сильно помочь при оценке среднего. Вопрос же о предпочтительности медианы при асимметричном, по существу, расположении отклонений будет рассмотрен в дальнейшем в не зависимости от наличия выскакивающих ошибок. Однако заранее ясно, что положительные свойства медианы скажутся и в этом отношении.

Попутно отметим, что в литературе имеется одно очень интересное предложение, принадлежащее известному геофизику Б. П. Вейнбергу. Он предложил особое среднее, названное им «средним серединной массы» и представляющее собой промежуточное решение в выборе между средним арифметическим и медианой.

Разделим нашу совокупность, расположенную в вариационный ряд (т. е. в порядке возрастания значений), на четыре к в а р т и л я и возьмем среднее арифметическое из той части, которая располагается во втором и в третьем квартилях, т. е. именно там, где содержатся более вероятные отклонения от центра.

Таким образом, сохраняются все преимущества среднего арифметического для средней массы значений и в то же время уменьшается вес всех тех значений, частотность которых вообще мала и поэтому представлена с меньшей надежностью.

Предложение Вейнберга не получило развития, однако полагаем, что в некоторых сомнительных случаях оно могло бы быть полезным. В табл. 3 приведен расчет Вейнберга к одному вариационному ряду проведенных измерений высоты одного предмета.

Медиана приведенного ряда (табл. 3) равна $\frac{31,68+31,72}{2} = 31,70$. Среднее арифметическое совокупности составляет 31,70,

№ измерений	1	2	3	4	5	6	7	8	9	10
Высота, мм	31,04	31,17	31,28	31,44	31,44	31,48	31,49	31,58	31,62	31,68
№ измерений	11	12	13	14	15	16	17	18	19	20
Высота, мм	31,72	31,73	31,87	31,87	31,88	31,88	31,96	31,98	32,12	32,22

т. е. оно совпадает с медианой. Нижняя граница второго квартиля 31,18, верхняя граница третьего квартиля 31,88. В этих пределах заключена «серединная масса». Ее среднее арифметическое 31,68. В данном примере отсутствуют «выскакивающие» значения, поэтому разница между тремя указанными средними ничтожна. Если представить себе, что в четвертом квартиле какая-нибудь цифра была бы сильно преувеличенной, то на медиану и на среднее серединной массы в данном случае она не повлияла бы, а оказала бы влияние только на среднее.

Перейдем к рассмотрению предложенных критериев для отбрасывания выскакивающих значений, не пользуясь медианой.

Будем считать, что ряд измерений равноточен и подчинен нормальному закону. Истинная его дисперсия, равная

$$\sigma^2 = \overline{(x_j - a)^2} = \frac{\sum (x_j - a)^2}{n},$$

нам так же неизвестна, как и истинное значение измеряемой величины a .

Эмпирическую дисперсию будем относить к первым n членам ряда, а член $(n+1)$ считаем «сомнительным». Чтобы ответить на вопрос, является ли этот член случайным или обусловленным какой-либо внешней причиной, найдем вероятность того, что модуль разности $|x_{n+1} - \bar{x}|$ будет равен или больше наперед заданной положительной величины λ .

Известно, что стандарт среднего убывает с извлечением корня квадратного из числа измерений. Следовательно, обозначая через s стандарт среднего из n членов, а через s' — стандарт полного ряда (т. е. включающего и максимальный член x_{n+1}), можем записать равенство

$$s' = \sqrt{\frac{n+1}{n}} s,$$

где

$$s = \sqrt{\frac{\sum (x_j - \bar{x})^2}{n}}$$

Мы видим теперь, что случайная величина

$$t = \frac{x_{n+1} - \bar{x}}{s'} = \frac{\lambda}{s'}$$

удовлетворяет условиям распределения Стьюдента и должна ему подчиняться. Следовательно,

$$P \{ |x_{n+1} - \bar{x}| \geq \lambda \} = 2 \int_{\frac{\lambda}{s'}}^{\infty} s(t) dt. \quad (98)$$

Появление цифры «2» перед интегралом правой части объясняется тем, что распределение Стьюдента симметрично, а в левой части стоит модуль разности; поэтому найденная нами вероятность относится как к превышению, так и к занижению среднего значения \bar{x} при учете максимального члена x_{n+1} . Обозначим эту вероятность для сокращения записи через

$$\beta = P \{ |\bar{x}_{n+1} - \bar{x}| \geq \lambda \}. \quad (99)$$

В табл. 1 приложения I, заимствованной из книги В. И. Романовского [10, стр. 28], приведены значения t_β для β , равной 0,05; 0,02; 0,01, и для различных значений n .

Используя данные этой таблицы, весь ход расчета проводим в следующем порядке. Для конкретного числа проведенных измерений n и для выбранной вероятности (например, $\beta=0,01$) получаем теоретическое значение t_β и сравниваем его с полученным из наших измерений эмпирическим значением $\frac{x_{n+1} - \bar{x}}{s'}$.

Если последняя величина больше первой (иначе говоря, если ее вероятность меньше принятой 0,01), то имеем право считать, что это «выскакивающее» измерение не могло быть получено в результате лишь случайной ошибки, а поэтому будет правильно, если его выбросим.

Кроме изложенного строгого метода расчета, существуют и другие, причем весьма распространен так называемый «критерий Шовене», гласящий, что одно измерение из общего числа

$(n+1)$ произведенных должно быть отброшено, если его вероятность, вычисленная по нормальному закону, не превышает $\frac{1}{2n}$.

Обозначая через $\pm \tilde{x}$ границы интервала, внутри которого считаем отклонения допустимыми, получим

$$P(\tilde{x}) = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\tilde{x}}^{+\tilde{x}} e^{-\frac{x^2}{2\sigma^2}} dx \leq \frac{1}{2n} \quad (100)$$

как вероятность того, что отклонение заключено в этом интервале. Тогда критерий Шовене выразится неравенством

$$1 - P(\tilde{x}) \geq \frac{1}{2n}. \quad (101)$$

В табл. 4 для различных n приводятся значения $\frac{x}{\sigma}$, при которых соответственное x должно быть отброшено.

Таблица 4

n	$\frac{x}{\sigma}$	n	$\frac{x}{\sigma}$
5	1,68	20	2,24
6	1,73	22	2,28
7	1,79	24	2,31
8	1,86	26	2,35
9	1,92	30	2,39
10	1,96	40	2,50
12	2,03	50	2,58
14	2,10	100	2,80
16	2,16	200	3,02
18	2,20	500	3,29

Так как критерий Шовене теоретического обоснования не имеет и является интуитивно эмпирическим, а метод, основанный на распределении Стьюдента, обладает известной строгостью, то рекомендовать применение в расчетах критерия Шовене можно лишь условно, ввиду его практической распространенности.

Сравним некоторые, довольно грубые подсчеты по обоим методам.

Примем приближенно, что дробь $\frac{x}{\sigma}$ можно приравнять к переменной t в распределении Стьюдента. В этом случае получим соотношения, приведенные в табл. 5.

n	β	По Стьюденту	По Шовене
20	0,01	2,9	2,2
	0,02	2,6	
40	0,01	2,7	2,5
	0,02	2,4	
60	0,01	2,7	2,6
	0,02	2,4	

Из данных табл. 5 видно, что расчетные цифры по Шовене сравнительно не очень сильно отличаются от расчетных цифр по Стьюденту, особенно для $n > 30$, если последние цифры отнести к вероятностям порядка $\beta = 0,01 - 0,02$.

Приведенный пример расчета как будто оправдывает практическую распространенность критерия Шовене, но все же надо отдать предпочтение предыдущему теоретически обоснованному методу, предложенному В. И. Романовским, тем более что данные таблицы весьма облегчают такой расчет.

В заключение отметим, что существует еще один метод расчета высказывающих измерений теоретически столь же строгий, как и метод В. И. Романовского. Этот метод принадлежит Н. В. Смирнову [13]. Но мы сочли достаточным ограничиться изложением метода Романовского.

§ 25. Случайные ошибки

Рассмотрены две категории ошибок, которые при измерении какой-либо величины практически возможно считать исключенными, хотя, повторяем, полной гарантии того, что они найдены и исключены, иметь невозможно.

Будем теперь условно считать, что все остальные отклонения измерителя от какого-то нам неизвестного центра зависят только от случайных факторов. В литературе встречается утверждение, что чисто случайные отклонения в показаниях измерительного прибора должны располагаться (хотя бы в пределе) с обязательной симметрией в отношении их центра. Такое отождествление при обработке измерений тесно связано с весьма обычным и часто оправдываемым применением нормального закона Гаусса. Но в § 22 показано с бесспорной очевидностью, что возможны случаи, когда это отождествление теряет свою силу. Это имеет место именно в тех случаях, когда измеряемая вели-

чина относится к категории «малых» величин, т. е. когда усиленное неравенство $\bar{x} \gg \sigma_x$ не выполняется.

Действительно, измеряя какую-либо существенно положительную величину, получаем некоторую кривую распределения возможных отклонений в обе стороны от центра, получающегося где-то вблизи истинного размера измеряемой величины. Но совершенно очевидно, что если отклонения вправо (т. е. в сторону превышения) в принципе не имеют никакого поставленного им предела, то отклонения влево ясно ограничены размером самой величины, так как абсолютно немислимо при измерении существенно положительной величины получить какую-то отрицательную величину! Вероятность того, что отрицательное отклонение превысит по модулю величину, конечно, равна нулю. Отрицательное же отклонение по Гауссу имеет конечное значение.

В этом коренном дефекте «закона» Гаусса необходимо серьезно разобраться, так как бывают случаи, когда подобное противоречие, может привести к существенным ошибкам.

Важным, с нашей точки зрения, является то, что в большинстве случаев — практически почти всегда — возможные ошибки измерительного прибора гораздо меньше измеряемой величины. Такое соотношение между тем, что измеряется, и тем, что измеряет, лежит в основе всей техники измерения. Расстояния между двумя пунктами при проведении железной дороги измеряются многократной передвижкой стальной ленты, а длина небольшого стального стержня измеряется циркулем. В обоих случаях вероятная ошибка гораздо меньше измеряемой величины, поэтому мы смело в обоих случаях применяем кривую Гаусса, уходящую обоими концами в положительную и отрицательную бесконечности, мысленно «отрезая» у этой кривой обе ее бесконечные ветви, как лишённые смысла!

Но если это основное условие не выполняется и стандарт отклонения имеет тот же порядок, что и измеряемая величина, то и вопрос о возможности применения закона Гаусса снимается, и мы должны для анализа измерений перейти к какому-то асимметричному распределению. Все случайные отклонения и в этом случае останутся, очевидно, случайными. Однако если в нормальном законе среднее, медиана и мода совпадают, то в асимметричном распределении этого уже не будет. Именно при положительной асимметрии среднее окажется правее моды и медианы, а медиана расположится между средним и модой.

Возникает вопрос: какой из трех (прежде совпадающих) центров мы теперь выберем? Нам кажется, что мы выберем не среднее и, конечно, не моду, а именно тот центр, который сохраняет принцип симметрии хотя бы в суммарном смысле: вероятность всех отклонений от него влево равна вероятности всех отклонений от него вправо, т. е. центром будет медиана.

Следующие две главы мы посвятим более детальному разбору

двух указанных качественно различных измерений. Но прежде чем перейти к их анализу, рассмотрим еще один принципиальный вопрос, характерный именно для неточных наук, когда зачастую приходится измерять величины, которые сами под влиянием многочисленных разнообразных факторов варьируют во времени около некоторого среднего значения. Очень часто бывает, что только это среднее значение представляет интерес и только оно оказывается тесно связанным с другими измеримыми величинами. Отдельные же измеренные нами значения представляют собой лишь случайные отклонения от интересующего нас закономерного среднего. Тогда ничто не мешает рассматривать отклонения этих отдельных значений от среднего как «ошибки» в измерениях этого среднего, хотя здесь имеются ошибки не измерительного прибора, а, так сказать, ошибки измеряемой величины, если считать, что измеряется ее среднее значение.

Здесь будут две существенно отличные категории отклонений: первая зависит, как известно, от точности измерительного прибора и от субъективных качеств наблюдателя, вторая охватывает реальные отклонения. Соответственно первая категория должна подчиняться основным условиям и выводам теории ошибок. Относительно же второй категории этот вопрос нуждается в специальной проверке.

Поставим его сначала в чисто практическую плоскость: интересуют ли нас эти реальные отклонения сами по себе или важно знать только то среднее, около которого они флюктуируются? Если нас интересует среднее, то следует выяснить с помощью статистической обработки многих рядов аналогичных наблюдений, а также путем теоретического анализа, подчиняются ли реальные отклонения тем же условиям, каким подчиняются измерительные отклонения, и в первую очередь являются ли они независимыми? Получив утвердительный ответ, мы этот вопрос решаем просто и можем при анализе рядов измерений не различать указанные две категории, а брать их только суммарно. Если же у нас нет уверенности в том, что реальные отклонения обладают теми же свойствами, что и измерительные отклонения, то надо стараться их разъединить, если, конечно, это возможно.

Если реальная величина флюктуирует только во времени, то выделить из суммарных отклонений их измерительную часть можно, но лишь в тех случаях, когда реальные отклонения имеют частоту, гораздо меньшую частоты измерительных отклонений, например когда производим учащенные измерения температуры воды в реке или температуры воздуха точным термометром, то такое разделение вполне возможно.

Но если измеряем, например, местную скорость в турбулентном потоке, пульсирующую несколько раз в секунду, то вынуждены просто игнорировать влияние неточности прибора на измерение мгновенных скоростей.

Поставим задачу — получить среднюю многолетнюю цифру стока какой-либо реки по измерениям ее годовых значений. Мы не можем, очевидно, дважды повторить измерение стока отдельного года, а поэтому вынуждены игнорировать ошибки, связанные с его измерением.

На практике интересным является случай, когда измеряемая величина флюктуирует и в пространстве и во времени. Такой величиной может быть высота снежного покрова. Обычно нас интересует только средняя величина запаса воды в снеге к началу снеготаяния по всему речному бассейну; поэтому мы вправе считать, что измерения высоты и плотности снега в отдельных его точках являются случайными величинами, отклоняющимися от их среднего значения.

Из трех приведенных примеров к «большим» величинам, удовлетворяющим усиленному неравенству $\bar{x} \gg \sigma_x$, нужно отнести, очевидно, лишь скорости в турбулентном потоке. К «малым» величинам, безусловно, относится годовой сток реки, колебание которого обычно имеет тот же порядок, что и сама величина стока, особенно это имеет место в засушливых районах. Что касается высоты снежного покрова, то при большой его мощности (т. е. преимущественно в северных районах) ее колебания обычно невелики по сравнению со средней высотой, при малой же мощности (т. е. в южных районах) эти колебания бывают относительно велики. В соответствии с этим высоту снежного покрова можно относить иногда к «большим», а иногда к «малым» величинам.

Если оставить в стороне вопрос о выделении из суммарных отклонений их измерительной части, которая, очевидно, должна занимать в них сравнительно незначительный объем, и сосредоточить внимание на суммарных отклонениях, условно отождествляя их с одними лишь реальными, то мы должны прежде всего отказаться от постулата их независимости. Ошибкам измерения не присущ какой-либо порядок: их можно переставлять произвольно. Реальные же отклонения связаны с существом флюктуирующей величины, а поэтому мы не вправе заранее принимать отсутствие связи между смежными их значениями, хотя в некоторых случаях это и может иметь место.

Отсюда следует вывод, что при статистической обработке ряда наблюдений реально флюктуирующей величины не имеем права, во-первых, преобразовать ее ряд в вариационный и, во-вторых, обязаны ввести, помимо центра распределения и рассеяния, еще и коэффициент корреляции между близлежащими значениями ряда, хотя последний в некоторых случаях может быть равен нулю. Это достигается введением особой структурной функции, которая может быть или пространственной, или временной.

Примем интервал длины λ для двух смежных по длине на-

блюдений и интервал τ для двух наблюдений, последовательных во времени. В первом случае структурная функция для величины f будет

$$b_f(\lambda) = \overline{\{f(x + \lambda) - f(x)\}^2},$$

во втором случае

$$b_f(\tau) = \overline{\{f(t + \tau) - f(t)\}^2}.$$

Вводя простейшие преобразования и принимая постоянство дисперсии и корреляции как по длине, так и по времени, получим

$$b_f(\lambda) = 2\sigma_f^2 \{1 - r_f^2(\lambda)\};$$

$$b_f(\tau) = 2\sigma_f^2 \{1 - r_f^2(\tau)\}.$$

В обоих случаях имеем

$$\text{при } \lambda = 0 \text{ или } \tau = 0 \quad r = 1, \quad b_f = 0;$$

$$\text{при } \lambda = \infty \text{ или } \tau = \infty \quad r = 0, \quad b_f = 2\sigma_f^2$$

(практически b становится нулем на конечных интервалах). В этом выводе мы принимали корреляцию постоянной по длине (или по времени) и зависящей только от величины интервала (λ или τ). Известно, что при любом законе изменения коэффициента корреляции он всегда убывает с длиной интервала. Закон этого убывания в общем виде нам неизвестен и должен быть определен, в частности, для каждого объекта исследования. Например, на основании детальных исследований точности снего съемок в 1954 г. И. С. Шпак пришел к выводу, что при интервалах λ от 10 до 20 м показания в отдельных точках становятся практически независимыми и к ним можно применять теорию ошибок.

С другой стороны, статистические исследования ряда годовых стоков р. Волги у Куйбышева показали, что для смежных лет коэффициент корреляции, достигнув величины 0,42, быстро уменьшается и переходит через нуль в отрицательную область, а затем снова возрастает до незначительного предела и т. д.

В последнее время вопрос об измерении флюктуирующих величин в применении к снегомерным съемкам был детально исследован группой ученых: Д. Л. Лайхтманом, Р. Л. Коганом и др.¹ В результате исследований этими авторами получены весьма интересные данные по кривым структурных функций, величинам погрешностей среднего значения и т. д. В настоящей книге, посвященной главным образом измерительным ошибкам, указанные выше вопросы подробно не рассматриваются.

¹ См. Тр. ГГО, вып. 108, 1960.

ИЗМЕРЕНИЕ «БОЛЬШИХ» ВЕЛИЧИН

§ 26. Симметричное распределение случайных ошибок

Исключая из дальнейшего изложения рассмотрение систематических и «выскакивающих» ошибок, сосредоточим наше внимание на одних лишь случайных ошибках, причем в этой главе рассмотрим распределение их при измерении «больших» величин, вероятные отклонения которых, согласно изложенным выше соображениям, могут считаться расположенными симметрично относительно их центра. Закон такого симметричного распределения уже неоднократно упоминался в предыдущих главах и был выведен в § 19 как частный случай интерполяции биномиального закона, показывающий симметричное распределение при интерполяции. Далее рассмотрим непосредственную связь этого закона с процессом измерения.

Ищем количественное выражение того закона, которому должна удовлетворять совокупность отклонений от центра при измерении «большой», в нашем понимании, величины. Качественная сторона вопроса в общем виде была освещена в предыдущей главе двумя постулатами. Ограничение же вопроса «большими» величинами предreshает выбор центра измерений, за который мы принимаем их среднеарифметическое.

В нашем выводе будем исходить из большого числа измерений, которое далее, переходя к пределу, примем равным бесконечности. Относя все измерения к их центру, т. е. по изложенному выше к их среднему арифметическому, примем как постулат, который, как нам кажется, весьма правдоподобен и заключается в том, что плотности вероятностей этих отклонений как в сторону их превышения, так и в сторону занижения равны. Важно отметить, что этот постулат (принадлежащий Гауссу) заранее предreshает совпадение среднего арифметического с модой и медианой в том законе, который мы, в конце концов, получим.

Второе вводимое нами допущение относится к независимости

всех измерений, а значит, и всех отклонений от центра. Интуитивно оно представляется почти очевидным.

И, наконец, весьма важное третье допущение, вскрытое в выводе Гаусса французскими математиками Бертраном и Пуанкаре, заключается в том, что вероятность каждого измерения предполагается зависящей только от его отклонения от центра, а не от размера измеряемой величины. На этом допущении нам надо особо остановиться.

Если, например, имеется очень длинный ряд из n измерений: $y_1, y_2, \dots, y_j, \dots, y_n$, по которому построен ряд отклонений от среднего: $x_1, x_2, \dots, x_j, \dots, x_n$, то исследователь обязан допустить, что плотность их распределения зависит не только от размера отклонений, но и от размера самой величины, от которой это отклонение отсчитывается, иными словами, мы должны были бы искать не $p(x)$, а $\varphi(x, y)$. Но в этом случае, как показал А. Пуанкаре [3, стр. 167], получается совсем не закон Гаусса, а другой гораздо более сложный!

Чем же оправдывается принятие отнюдь не самоочевидного допущения? Согласно Пуанкаре, такое допущение оправдывается только тем, что все реальные, т. е. не слишком маловероятные, отклонения обычно бывают очень малы по сравнению с измеряемой величиной. Но такая трактовка тождественна с вводимым нами определением «больших» по сравнению с возможными отклонениями величин, к которым только и применим, как было изложено в предыдущей главе, нормальный закон Гаусса!

Переходим к выводу распределения случайных ошибок и прием давно установившуюся в теории ошибок схему, согласно которой все случайные ошибки, зависящие от очень большого числа малозначащих факторов, можно представить суммами чрезвычайно малых элементарных ошибок. Поскольку малость этих гипотетических элементов чисто условная и ничем не может быть ограничена, то в дальнейшем мы имеем право считать их бесконечно малыми, а в пределе в надлежащих случаях даже приравнивать к нулю. Кроме того, ничто не мешает нам принимать эти ошибки равными. Такое дополнительное упрощение к схеме элементарных ошибок, введенное Хагеном, делает весь вывод более наглядным и более простым, а поэтому в дальнейшем мы его используем.

Элементарная ошибка ε может быть и положительной, и отрицательной, следовательно, по закону симметрии будем иметь для них

$$p = q = \frac{1}{2}.$$

Рассмотрим, каково будет возможное распределение этих

элементарных ошибок. При одной ошибке могут быть два случая: $-\varepsilon$ и $+\varepsilon$, каждая с вероятностью $\frac{1}{2}$. При двух ошибках возможны четыре комбинации:

$$\begin{aligned} &-\varepsilon \quad -\varepsilon; \\ &-\varepsilon \quad +\varepsilon; \\ &+\varepsilon \quad -\varepsilon; \\ &+\varepsilon \quad +\varepsilon. \end{aligned}$$

Если не будем придерживаться порядка расположения положительных и отрицательных ошибок, то получим в этом случае не четыре, а лишь три комбинации: -2ε , 0 , $+2\varepsilon$ с соответствующими вероятностями $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$.

При трех ошибках, каждая из которых может быть с вероятностью $\frac{1}{2}$ положительной или отрицательной, будем иметь восемь возможных порядков:

$$\left| \begin{array}{c|c|c|c|c|c|c|c} -\varepsilon & -\varepsilon & -\varepsilon & -\varepsilon & +\varepsilon & +\varepsilon & +\varepsilon & +\varepsilon \\ -\varepsilon & -\varepsilon & +\varepsilon & +\varepsilon & -\varepsilon & -\varepsilon & +\varepsilon & +\varepsilon \\ -\varepsilon & +\varepsilon & -\varepsilon & +\varepsilon & -\varepsilon & +\varepsilon & -\varepsilon & +\varepsilon \end{array} \right|$$

А если игнорировать порядок и учитывать лишь числа ошибок каждого знака, то число возможных комбинаций окажется равным четырем:

$$-3\varepsilon; \quad -\varepsilon; \quad +\varepsilon; \quad +3\varepsilon,$$

с соответственными вероятностями $\frac{1}{8}$, $\frac{3}{8}$, $\frac{3}{8}$, $\frac{1}{8}$.

Перейдем к обобщенной формулировке этого положения. Каждую конкретную ошибку будем считать алгебраической суммой из n элементарных ошибок, из которых каждая может быть с вероятностью $\frac{1}{2}$ или положительной, или отрицательной. Такая схема равнозначна схеме урны, из которой вынимаются и тотчас возвращаются белые и черные шары. Если представим положительную ошибку белым шаром, а отрицательную черным, а число тех и других в урне равными, то получим классическое распределение Бернулли при $p=q=\frac{1}{2}$, с той лишь разницей, что там каждое испытание давало m белых и $(n-m)$ черных шаров, здесь же мы, суммируя получаемое каждый раз распределение ошибок, находим очень просто их сумму:

$$x_m = m(+\varepsilon) + (n-m)(-\varepsilon) = (2m-n)\varepsilon. \quad (102)$$

Вероятность каждой такой суммы будет, очевидно, та же, что и комбинации из m положительных и $(n-m)$ отрицательных

ошибок, т. е. при симметричном их распределении она будет равна

$$P_n^m = \frac{n!}{m!(n-m)!} \frac{1}{2^n}. \quad (103)$$

Таким образом мы получили дискретное распределение суммарных ошибок.

Две предельные, максимальные по модулю, ошибки будут, очевидно, $-n\varepsilon$ и $+n\varepsilon$. Вероятности их одинаковы и равны $\frac{1}{2^n}$, т. е. очень малы. Большие вероятности имеют члены, расположенные в средней части полученного распределения, там, где алгебраическая сумма элементарных ошибок по модулю мала.

Далее увидим, что и после интерполяции характер распределения хотя и сохранится, но будет иметь существенное отличие, заключающееся в том, что у полученного нами дискретного распределения два конечных предела $\pm n\varepsilon$, а у непрерывного распределения после интерполяции вследствие убывания ε к нулю оба конца будут стремиться к $\pm\infty$. Это, конечно, является фикцией, ухудшающей теоретическую постановку вопроса, но по ряду соображений, которые станут убедительными в дальнейшем изложении, мы эту фикцию примем как вполне законную.

Формальная сторона перехода от дискретного биномиального распределения к интерполированному подробно была изложена в § 18 и 19 сначала в общем случае, а затем и для частного случая симметрии отклонений. Следовательно, нет надобности повторять все, что было ранее изложено. Общее уравнение для отношения $\frac{\Delta P}{P}$, полученное в § 18,

$$\frac{\Delta P}{P} = 2 \frac{pn - m - q}{pn - (2p - 1)m + q}$$

преобразуем теперь на основе нового равенства (102) из схемы суммирования элементарных ошибок и будем иметь равенство, вначале несколько отличное от полученного в § 19, а именно

$$\frac{\Delta P}{P} = \frac{2pn - \frac{x_m}{\varepsilon} - n - 2q}{pn - (2p - 1)\left(\frac{x_m}{2\varepsilon} + \frac{n}{2}\right) + q}.$$

Но далее, вводя условие $p = q = \frac{1}{2}$, умножая числитель и знаменатель на ε и отбрасывая величины, убывающие к нулю вместе с ε , получим

$$\frac{\Delta P}{P} = - \frac{2x_m}{n\varepsilon}.$$

Приращение ΔP вычислялось нами для разности двух смежных значений m , т. е. для единицы, поэтому

$$\Delta P = \frac{\Delta P}{\Delta m},$$

где $\Delta m = 1$. Переходя к переменной x_m , для которой имеем

$$\Delta x_m = x_{m+1} - x_m = 2\varepsilon,$$

запишем соответственно

$$\frac{1}{P} \frac{\Delta P}{\Delta x_m} = - \frac{x_m}{n\varepsilon^2}.$$

Но нам известно, что ε^2 — дисперсия элементарной ошибки, следовательно $n\varepsilon^2$ является дисперсией конкретных ошибок, т. е. $n\varepsilon^2 = \sigma^2$. Переходя к пределу, получим

$$\frac{dy}{y} = - \frac{xdx}{\sigma^2}. \quad (104)$$

Интегрируя, освобождаясь от логарифмов и вводя хорошо известное нам значение произвольной постоянной, придем, очевидно, к полученной ранее форме нормального закона:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \quad (105)$$

в которой осредненное значение измерений не может фигурировать, поскольку весь вывод проведен относительно одних лишь отклонений от центра.

Закон Гаусса, как, впрочем, и всякий закон распределения плотностей вероятности, относится, очевидно, лишь к бесконечному числу измерений, иначе говоря, к их генеральной совокупности. На практике же обычно имеется весьма ограниченное число измерений, из которых можно получить лишь эмпирическое среднее и эмпирический стандарт. Переход от эмпирических параметров к истинному среднему и истинному стандарту возможен, очевидно, лишь «вероятностный», т. е. такой при котором каждому интервалу значений, внутри которого может находиться «истинное» значение данного параметра, соответствует определенная вероятность этого нахождения. Этот вопрос относится к теории «оценок».

§ 27. Оценка среднего и стандарта

Понятие оценки является одним из основных в теории ошибок. Это понятие тесно связано с сопоставлением генеральной (неизвестной) совокупности с той конечной из нее выборкой, которую мы конкретно получаем по измерениям.

Оценкой в общем случае называется определение степени

приближения значений параметров, полученных нами из эмпирических данных, к той генеральной совокупности, выборку из которой эмпирические данные представляют.

Рассмотрим, каким условиям должна удовлетворять рационально построенная оценка. Во-первых, она, очевидно, должна сходиться по вероятности к истинному значению оцениваемой величины при возрастании числа измерений к бесконечности. Такое условие называется *состоятельностью* оценки. Во-вторых, оценка не должна содержать систематических погрешностей, «смещающих» ее значение, т. е. она должна быть *несмещенной*. Смещение оценки может быть или положительным, т. е. завышающим, или отрицательным, т. е. занижающим истинное значение оцениваемого параметра.

Этому требованию удовлетворяет, например, согласно теореме Чебышева, среднее арифметическое, поскольку

$$M.O. \bar{x}_k = a,$$

где \bar{x} — среднее арифметическое k -той выборки.

Эмпирический же стандарт этому требованию не удовлетворяет, как известно из ранее доказанного равенства

$$\frac{s_k^2}{n-1} = \frac{\sigma^2}{n},$$

из которого следует, что оценка σ по эмпирическому (выборочному) стандарту s_k дает отрицательное смещение: ¹

$$s_k - \sigma = \sqrt{\frac{n-1}{n}} - 1 \approx \frac{1}{2n}.$$

Из существующих методов составления оценок остановимся лишь на одном — на методе *доверительных интервалов*, идейная сущность которого была ранее затронута, а именно мы отыскиваем те пределы, внутри которых находится с заданной вероятностью истинное значение оцениваемого параметра. Отрезок, ограниченный этими пределами, называется *доверительным интервалом*, а та вероятность, которая требуется по условиям задачи, называется *доверительной вероятностью*.

Рассмотрим вопрос сначала в самом общем виде. Имеется конечное число произведенных измерений, т. е. какая-то выборка из неизвестной генеральной совокупности. Относительно последней имеются основания предполагать, что она может быть выражена таким-то определенным законом распределения.

Представим этот закон сначала в общем виде уравнения

¹ Мы пользуемся здесь приближенным равенством $\sqrt{1 \pm \frac{1}{n}} \approx 1 \pm \frac{1}{2n}$.

плотности вероятностей $p(x)$ с таким-то числом подлежащих определению параметров. Имеем

$$p(x) = f(x, \theta_1, \theta_2, \dots, \theta_m). \quad (106)$$

Для m параметров (обычно их число не превышает двух) могут быть найдены приближенные значения из нашей конкретной выборки:

$$\theta_i = F_i(x_1, x_2, \dots, x_n).$$

Из генеральной совокупности (если бы она была нам доступна) можно было бы определить точные значения m параметров. Образует новую безразмерную переменную, в которую войдут, кроме нашей исходной переменной x , также и приближенные значения указанных параметров. Для новой переменной t ищем закон распределения, не содержащий никаких новых параметров. Тогда для любого интервала $t_1 \leq t \leq t_2$ с любыми пределами t_1 и t_2 сможем вычислить доверительную вероятность.

Иллюстрируем приведенное общее определение простейшим примером распределения нормального закона, который имеет, как известно, два параметра: математическое ожидание, в пределе совпадающее с истинным значением измеряемой величины a , и стандартное отклонение измерительного прибора σ . Наша конкретная выборка вместо этих двух точных параметров дает лишь два приближенных параметра \bar{x} и s . Нам надо оценить первые из вторых параметров.

Для начала упростим задачу и предположим, что стандарт измерителя σ известен (например, из других измерений) и нам нужно оценить лишь параметр a .

Дисперсия среднего арифметического нашего конкретного ряда, взятая от центра генеральной совокупности, будет равна $\frac{\sigma^2}{n}$. Среднее же, полученное непосредственно из выборочного ряда \bar{x} ; является лишь случайной величиной, зависящей от точности и числа измерений. Считая, что отклонения выборочных средних от генерального среднего должны подчиняться нормальному закону, образуем нормированную величину

$$z = \frac{\bar{x} - a}{\frac{\sigma}{\sqrt{n}}}, \quad (107)$$

плотность распределения которой будет подчиняться нормальному закону в нормированном виде, т. е. будет свободна от каких-либо параметров

$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Отсюда нетрудно определить и тот доверительный интервал (z_1, z_2) , внутри которого с заданной вероятностью помещается

постоянная, но неизвестная величина a . Обозначая заданную доверительную вероятность через δ , запишем равенство

$$P(z_1 < z < z_2) = \delta. \quad (108)$$

Лучше всего расположить концы доверительного интервала симметрично относительно центра распределения $z_1 = -z$; $z_2 = +z$. Тогда длина доверительного интервала будет равна $2z \frac{\sigma}{\sqrt{n}}$.

Смысл равенства (108) заключается в том, что если бы мы сделали многочисленные выборки, то каждый соответственный интервал, длиной $2z \frac{\sigma}{\sqrt{n}}$, перекрывал бы неизвестную постоянную величину a с частотой, приближающейся в пределе к вероятности δ . При этом надо помнить, что величины a и σ постоянные (в данной задаче), а x — случайная величина, принимающая для каждой выборки различные значения. Абсциссы концов интервала и вероятность δ — взаимообуславливающиеся величины. Зависимость между ними получается очень просто: если δ — вероятность того, что отрезок с абсциссами $-z \frac{\sigma}{\sqrt{n}}$, $+z \frac{\sigma}{\sqrt{n}}$ перекрывает постоянную a , то эта вероятность должна быть равна части площади, нормированной кривой Гаусса, заключенной между абсциссами $-z$ и $+z$. Следовательно,

$$\delta = \int_{-z}^{+z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 2 \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \quad (109)$$

В приложении I приведена табл. 2, по данным которой непосредственно можно найти правую часть выражения (109), носящую название «удвоенной функции Лапласа». Полученное по таблице значение z надо умножить на $\frac{\sigma}{\sqrt{n}}$, чтобы получить значения абсцисс доверительного интервала в единицах измеряемой величины.

Для пояснения приведем конкретный пример. Измерена в поле какая-то длина. Из 16 измерений получено среднее арифметическое, равное 21,45 м. Эта величина — приближенное значение измеряемой длины. Для определения возможного отклонения измеряемой длины от «истинной» учтем, что из опыта аналогичных измерений среднеквадратическое отклонение примененного метода должно составлять 0,35 м. Для суждения о вероятном интервале, внутри которого помещается истинное значение той же длины, зададимся доверительной вероятностью $\delta = 0,90$. По табл. 2 приложения I получаем соответственное значение $z = 1,60$, откуда $z \frac{\sigma}{\sqrt{n}} = 0,15$ м. Следовательно,

истинное значение измеряемой длины находится между значениями 21,30 и 21,60.

При повышении доверительной вероятности, например, до 0,95 будем иметь: $z = 1,96$, $z \frac{\sigma}{\sqrt{n}} = 0,17$. Следовательно, доверительный интервал имеет теперь длину 0,34 м, истинное же значение измеряемой длины находится между значениями 21,28 и 21,62.

В общем случае длина интервала должна, очевидно, возрастать с σ и убывать с n .

Переходим теперь к более сложному, но практически более интересному случаю, когда мы не знаем не только истинного значения измеряемой величины, но и истинного значения стандарта измерений. И ничего, кроме одной (или нескольких) серий произведенных измерений и соответственно одной (или нескольких) пар значений параметров \bar{x} и s_k , у нас не имеется! Это равносильно тому, что о генеральной совокупности вообще ничего неизвестно, а имеется лишь ряд выборок, которые считаются подчиненными одному и тому же нормальному закону, с одними и теми же истинными значениями параметров a и σ . Причем, хотя a неизвестно, мы знаем, согласно закону больших чисел, что в пределе, при бесконечно большом числе наблюдений, было бы выполнено равенство

$$M.O. x_k = a.$$

Что же касается генерального стандарта σ , то никакой зависимости между ним и эмпирическим стандартом не существует, а поэтому, если заменим σ на s , то получим, хотя также нормированную, но случайную величину иного вида

$$t = \frac{\bar{x}_k - a}{\frac{s_k}{\sqrt{n}}},$$

где $\frac{s_k}{\sqrt{n}}$ — стандарт эмпирического среднего k -той выборки.

Мысленно представим какое угодно большое число аналогичных выборок с различными \bar{x}_k и s_k . Получим распределение новой случайной величины t , которое будет, очевидно, уже иным, чем в предыдущем случае, так как в нем от выборки к выборке изменялось только \bar{x}_k , теперь же изменяется и \bar{x}_k и s_k (а значит, и $s_{\bar{x}}$).

В главе V было получено нужное нам сейчас распределение случайной величины, а именно распределение Стьюдента. Запишем его вторично

$$s(t, k) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}. \quad (110)$$

Это распределение также нормировано, но, в отличие от нормированного закона Гаусса, не имеющего никаких параметров, оно имеет один целочисленный параметр k . Причем, как было показано в § 29, значение этого параметра равно $k=n-1$, а распределение (110) приближается к нормальному закону и в пределе с ним совпадает.

Дальнейший ход рассуждений аналогичен предыдущему: также назначается доверительная вероятность, для которой составляется доверительный интервал из очевидного равенства

$$\delta = P\{-z < t_{\delta} < +z\} = 2 \int_0^z s(t, k) dt \quad (111)$$

(последняя часть равенства следует из симметричности распределения).

Расстояния от центра распределения до концов интервала будут также равны $\pm z \frac{s_x}{\sqrt{n}}$, длина доверительного интервала

составит $2z \frac{s_x}{\sqrt{n}}$.

В табл. 3 приложения I приведены значения t_{δ} по Стьюденту — Фишеру, удовлетворяющие равенству (111).

Для иллюстрации расчета по этой таблице приведем тот же численный пример, какой был дан в первом расчете.

1. При доверительной вероятности $\delta=0,90$, эмпирическом стандарте 0,47 и значении $k=n-1=15$ получаем по табл. 3 $z=1,75$; $z \frac{0,47}{\sqrt{n}}=0,21$, откуда по изложенному пределы доверительного интервала будут 21,45—0,21; 21,45+0,21, иными словами, от 21,24 до 21,66.

2. При доверительной вероятности $\delta=0,95$ по табл. 3 $z=2,13$, остальные величины те же, что и в предыдущем расчете. Следовательно, $z \frac{0,27}{\sqrt{n}}=0,26$ и доверительный интервал имеет пределы от 21,19 до 21,71.

По второму расчету того же примера для тех же доверительных вероятностей получены более широкие пределы значений истинного размера измеренной длины. Это следует из того, что в первом расчете было принято устойчивое значение стандарта на основе других измерений, якобы более точных. Во втором же расчете взят эмпирический стандарт по тем же измерениям. В первом расчете соответственно использован нормированный закон Гаусса, а во втором — распределение Стьюдента.

Переходим далее к оценке генерального стандарта σ . Если бы истинное значение a было нам надо (или получено теоретически), то, имея достаточно длинный ряд измерений x_j , легко

получили бы приближенное значение генерального стандарта из равенства

$$\sigma = \sqrt{\frac{\sum (x_j - a)^2}{n}}.$$

Но так как значение a дано лишь в виде доверительного интервала, то и значение σ также можно получить в том же виде.

Запишем отношение выборочной дисперсии к генеральной $\frac{s^2}{\sigma^2}$. Сумма k таких отношений даст случайную величину, обозначенную через χ^2 :

$$\chi^2 = \frac{ks^2}{\sigma^2}.$$

Распределение этой случайной величины было получено ранее, но так как мы хотим дать оценку не дисперсии, а квадратному корню из дисперсии, т. е. стандарту, то нам удобнее оперировать не с χ^2 , а с χ . Исходя из общего правила перехода от распределения одной случайной величины к распределению другой, функционально с ней связанной, запишем

$$f_1(x) dx = f_2(y) dy.$$

Преобразуем уравнение (84) от случайной величины $x = \chi^2$ к случайной величине $y = \chi$. Имеем

$$\frac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}}} dx = f(y) dy,$$

а так как $dx = 2y dy$, то получаем (заменяя y через χ)

$$\varphi(\chi) = \frac{\chi^{k-1} e^{-\frac{\chi^2}{2}}}{\Gamma\left(\frac{k}{2}\right) 2^{\frac{k}{2}-1}}. \quad (112)$$

Полученное равенство дает нам возможность произвести оценку приближенного равенства $\sigma \approx s$. Искомый доверительный интервал в общем виде характеризуется неравенством

$$s - \varepsilon < \sigma < s + \varepsilon,$$

где $\pm \varepsilon$ — его концы.

Приведенное равенство, согласно введенному определению

$$\chi = \frac{s\sqrt{k}}{\sigma},$$

легко (путем деления на него величины $s\sqrt{k}$) переписывается в следующее неравенство:

$$\frac{s}{s + \varepsilon} \sqrt{k} < \frac{s\sqrt{k}}{\sigma} = \chi < \frac{s}{s - \varepsilon} \sqrt{k}. \quad (113)$$

Обозначая

$$\chi_1 = \frac{s\sqrt{k}}{s+\varepsilon}; \quad \chi_2 = \frac{s\sqrt{k}}{s-\varepsilon},$$

будем иметь выражение для вероятности неравенства (113) в виде

$$\delta = \int_{\chi_1}^{\chi_2} P(\chi, k) d\chi. \quad (114)$$

Неравенство (113) можно привести к более удобному виду, введя относительную точность интервала

$$q = \frac{\varepsilon}{s}.$$

Тогда получим

$$P \left\{ \frac{\sqrt{k}}{1+q} < \chi < \frac{\sqrt{k}}{1-q} \right\} = \int_{\chi_1}^{\chi_2} p(\chi_1 k) d\chi, \quad (115)$$

где

$$\chi_1 = \frac{\sqrt{k}}{1+q}; \quad \chi_2 = \frac{\sqrt{k}}{1-q}.$$

Зависимость между δ , q и k представлена данными табл. 4 приложения I, заимствованной из монографии В. П. Романовского [10, стр. 114]. Доверительная вероятность в этой таблице дана только тремя практически наиболее интересными цифрами: 0,95, 0,99, 0,999. Число степеней свободы $k=n-1$ выражено значениями от 1 до 500.

При $k=1$ (т. е. при $n=2$) получаются непомерно высокие значения q , что совершенно естественно, так как при двух измерениях значение стандарта становится малоопределенным, а поэтому и длина доверительного интервала сильно возрастает, в особенности при высоких значениях вероятности δ . Здесь указываем на это обстоятельство лишь с иллюстративными целями; практического значения оно не имеет, так как реальное число измерений всегда значительно больше двух.

В заключение скажем несколько слов о возможности проверки подчиненности экспериментальных измерений нормальному закону.

Непосредственная проверка путем совмещения экспериментального ступенчатого графика с теоретической кривой возможна лишь в тех исключительных случаях, когда число измерений чрезвычайно велико. Если же число измерений составляет сотни, то, разбив его по интервалам протяженностью, например, по $\frac{\sigma}{2}$, получим всего лишь шесть интервалов, из которых средние два будут содержать несколько десятков измерений или

даже свыше сотни каждый, два вторых интервала (считая от центра) будут гораздо меньше, а третьи от центра — порядка немногих единиц. Сравнение полученной ломаной линии с теоретической кривой, как это показано на рис. 3, может, конечно, дать некоторое представление (но скорее качественное) о близости распределения проведенных измерений к нормальному закону. Однако небольшое количество точек в удаленных от центра интервалах делает такое сравнение малоубедительным. И если это имеет место при нескольких сотнях измерений, как показано выше, то при обычных нескольких десятках измерений о нем нечего и думать. Поэтому практически приходится ограничиваться вычислением асимметрии и эксцесса, которые теоретически, как было показано в § 19, должны быть равны нулю; следовательно, их выборочные (эмпирические) значения должны быть относительно малы. Обычно этим на практике и ограничиваются.

§ 28. Неравноточные измерения

До сих пор мы ограничивались разбором лишь тех случаев, когда все измерения имеют одинаковую точность. На практике это обычно имеет место, но все же не всегда. Иногда приходится обрабатывать серию неравноточных измерений; причем различная их точность может быть результатом или применения различных приборов, или сменой наблюдателей, или, наконец, изменениями условий среды.

В таком случае имеем две, три (и более) выборки разного качества, но относящиеся, по существу, к измерению одной и той же величины.

Задача заключается в том, чтобы привести эти различные выборки к сравнимому виду и чтобы далее вести обработку вплоть до выводов для совокупности всех выборок.

Начнем с простейшего и, по существу, тривиального случая, когда все измерения производились одним прибором и одним наблюдателем в однородных условиях, а разные серии различаются лишь числом измерений. Иными словами, имеется m серий с различными числами измерений, дающих m средних арифметических. Как найти для всех этих серий общую арифметическую середину? Слишком очевидно, что простое среднее из всех m средних не будет истинным средним для всех серий, так как серии с малым числом измерений и серии с большим числом измерений не могут считаться равноценными: средние квадратичные ошибки первых будут больше, чем средние квадратичные ошибки вторых. Точное среднее запишется в виде

$$\bar{x} = \frac{\sum_1^{n_1} x_{j, 1} + \sum_1^{n_2} x_{k, 2} + \dots}{n_1 + n_2 + \dots}.$$

Но так как каждое из слагаемых числителя может быть представлено

$$\sum x_{e, j} = n_e \bar{x}_e,$$

то предыдущее равенство переписывается в виде

$$\bar{x} = \frac{\sum n_e \bar{x}_e}{\sum n_e}.$$

Множители n_e при выборочных средних x_e являются поправочными коэффициентами для перехода от выборочных средних к общему среднему. Разделив теперь числитель и знаменатель на среднее квадратическое отклонение (ошибку) отдельного измерения, которое, как сказано ранее, должно быть одинаковым для всех серий, получим

$$\bar{x} = \frac{\sum \frac{n_e}{\sigma^2} x_e}{\sum \frac{n_e}{\sigma^2}}.$$

Множители $\frac{n_e}{\sigma^2}$ назовем весами средних выборочных серий, а множитель $\frac{\sum n_e}{\sigma^2} = \frac{n}{\sigma^2}$ будет весом общего среднего. Введя теперь для весов новое обозначение $\frac{n_e}{\sigma^2} = p_e$, $\frac{n}{\sigma^2} = p$, получим равенство в окончательном виде:

$$\bar{x} = \frac{\sum p_e x_e}{p}. \quad (116)$$

Равенство (116) является, очевидно, вполне строгим для рассматриваемого тривиального случая, но с помощью одного весьма приемлемого допущения оно может быть принято и для более общего случая, когда серии измерений отличаются друг от друга не только числами измерений, но и их качеством. Качество, т. е. в положительном смысле слова «надежность», отдельных измерений можно охарактеризовать величиной, обратной дисперсии. Тогда для i -той серии, характеризуемой дисперсией σ_i^2 и числом измерений n_i , дисперсия частного среднего будет равна $\frac{\sigma_i^2}{n_i}$, а следовательно, вес p_i этого среднего будет

$$p_i = \lambda \frac{n_i}{\sigma_i^2},$$

где λ — общий для всех серий и совершенно произвольный коэффициент пропорциональности. Аналогично для совокупности всех серий

$$p = \lambda \sum_{i=1}^{i=m} \frac{n_i}{\sigma_i^2}.$$

Так как множитель λ входит в числитель и в знаменатель дроби, то он может быть взят совершенно произвольно. Для удобства можно, например, принять «вес» одной какой-либо i -той серии за единицу:

$$p_i = \lambda \frac{n_i}{\sigma_i^2} = 1,$$

а веса всех остальных серий, а также совокупности серий вычислять из равенства

$$p_e = \frac{n_e}{n_i} \frac{\sigma_i^2}{\sigma_e^2}. \quad (117)$$

Перейдем к решению задачи об оценке среднего при неравноточных измерениях. Уточним постановку задачи. Имеется n измерений величины a с помощью m серий, по n_i измерений в каждой серии с дисперсией s_i^2 .

Принимаем теперь, что каждая серия подчинена своему нормальному закону с параметрами: a — постоянным для всех серий и σ_i^2 — характерным для каждой серии.

Стандарт среднего каждой серии равен $\frac{s_i}{n_i - 1}$, вес этого среднего $p_i = \frac{n_i}{\sigma_i^2}$. За приближенное значение истинного значения примем общее среднее:

$$\bar{x} = \frac{\sum p_i x_i}{p}$$

и дадим оценку приближенного равенства $a \approx \bar{x}$. Поступаем аналогично предыдущему, т. е. образуем нормированную величину вида $t = \frac{\bar{x} - a}{s_x}$. Но знаменатель этой дроби в случае неравноточных измерений приобретает иной вид. Разделив взвешенную сумму квадратов отклонений частных средних от общего среднего $\sum p_i (\bar{x}_i - \bar{x})^2$ на общий вес и на число серий без единицы, получим

$$s_x = \sqrt{\frac{\sum p_i (\bar{x}_i - \bar{x})^2}{p(m-1)}}.$$

Окончательное выражение для нормированной величины t примет вид

$$t = \frac{\sqrt{p(m-1)} (\bar{x} - a)}{\sqrt{\sum p_i (\bar{x}_i - \bar{x})^2}}. \quad (118)$$

Распределение величины t должно быть, очевидно, то же, что и соответственной величины для равноточных измерений, так как все изменения, связанные с неравноточностью в ее структуре, уже учтены.

Имеем, следовательно, аналогичное и столь же строгое равенство

$$P \{ \bar{x} - t_{\delta} s_{\bar{x}} < a < \bar{x} + t_{\delta} s_{\bar{x}} \} = 2 \int_0^{t_{\delta}} s(t, k) dt. \quad (119)$$

Пользуясь данными табл. 4 приложения I, получим доверительный интервал длиной $2t_{\delta} s_{\bar{x}}$, внутри которого с вероятностью δ помещается истинная величина a .

Глава VIII

ИЗМЕРЕНИЕ МАЛЫХ (СУЩЕСТВЕННО ПОЛОЖИТЕЛЬНЫХ) ВЕЛИЧИН

§ 29. Асимметрия распределения случайных ошибок

Переходим ко второму из двух качественно различных случаев измерения: когда среднее значение и стандарт измерений — величины одного порядка, и усиленное неравенство $\bar{x} \gg \delta_x$, необходимое для применимости нормального закона Гаусса, не выполняется. Как указывалось в главе VII, симметрия распределения в этом случае даже противоречит самой сущности измерения. Однако другое положение теории ошибок, отнюдь не связанное с условием симметрии и показывающее, что каждую конкретную случайную ошибку можно рассматривать как сумму огромного числа элементарных ошибок, может остаться в силе с одним лишь изменением, что вероятности положительного и отрицательного значений той же по модулю элементарной ошибки ε не равны. В этом случае мы получим из того же положения какое-то другое, очевидно асимметричное, распределение.

Советская школа математиков во главе с акад. А. Н. Колмогоровым идет именно по этому пути. В монографии Б. В. Гнеденко и А. Н. Колмогорова [15] вопрос о предельных распределениях ставится в общем виде. Путем некоторой сложной комбинации законов Гаусса и Пуассона авторы получают такой вид характеристичных функций, который лишь при некоторых ограничительных условиях дает (при пересчете к функциям распределения) нормальный закон Гаусса. В общем же случае он приводит в пределе к системе асимметричных распределений.

Ю. В. Линник [9] дал некоторое развитие теории указанных распределений, а Н. В. Смирнов исследовал весьма интересный случай, соответствующий частным значениям входящих в рас-

пределения параметров $\alpha = -1$ и $\beta = 2$ и имеющий плотность вероятности вида

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2x}} x^{-\frac{3}{2}}. \quad (120)$$

Рассмотрим указанный случай, являющийся наиболее простым и одновременно достаточно характерным для всей системы распределений Гнеденко и Колмогорова.

Естественно, напрашивается замена

$$y = \frac{1}{2x}; \quad dx = -\frac{dy}{2y^2},$$

при помощи которой уравнение (120) преобразуется в следующее уравнение:

$$p(x) = \frac{1}{\sqrt{\pi}} x^{-\frac{1}{2}} e^{-x}, \quad (121)$$

где y снова заменен на x .

Теперь непосредственно убеждаемся, что распределение Смирнова после указанного преобразования представляет собой хорошо известное гамма-распределение (73) при $\alpha = -\frac{1}{2}$ ¹. Но, с другой стороны, видно, что между распределением Смирнова до его преобразования и гамма-распределением существует весьма существенное различие: последнее, т. е. гамма-распределение, имеет моменты любого порядка, как было показано в § 18, а основное распределение (120), наоборот, никаких моментов, кроме нулевого, не имеет. Для доказательства запишем общее выражение r -го начального момента распределения (120):

$$\nu_r = \int_0^{\infty} p(x) x^r dx = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2x}} x^{r-\frac{3}{2}} dx.$$

В определенном интеграле²

$$\int_0^{\infty} e^{-y^q} y^p dy = \frac{1}{q} \Gamma\left(\frac{p+1}{q}\right)$$

произведем замену:

$$y = 2x; \quad q = -1; \quad p = r - \frac{3}{2},$$

¹ Напоминаем, что $\Gamma(\alpha+1)$ при $\alpha = -\frac{1}{2}$ равна $\sqrt{\pi}$.

² См. Рыжик И. М. Таблицы интегралов, сумм, рядов и произведений. ГТТИ, М.—Л., 1943, стр. 134.

в результате которой легко преобразуем его к виду

$$\int_0^{\infty} e^{-\frac{1}{2x}} x^{r-\frac{3}{2}} dx = 2^{\frac{1}{2}-r} \Gamma\left(\frac{1}{2}-r\right).$$

Поскольку r — целое неотрицательное число, то становится очевидным, что лишь при $r=0$ выражение под знаком гамма-функции будет положительным; и лишь в этом случае выражение для ν_r будет конечным. Все же моменты выше нулевого порядка обращаются в бесконечность, иными словами, не существуют.

Значение нулевого момента (выражающего, как нам известно, площадь кривой плотности распределения), очевидно, равно единице.

Мы знаем, что при измерениях малых, существенно положительных величин, среднее арифметическое и стандартное отклонение не представляют интереса, а лишь медиана и среднее абсолютное от нее отклонение. И то, и другое здесь, очевидно, существует, причем можем воспользоваться известным свойством медианы, заключающемся в том, что медиана ряда, образованного любой монотонной функцией основного ряда, равна этой же функции от медианы основного ряда, и вычислить медиану распределения Смирнова по данным таблицы, составленной Е. Е. Слуцким, для гамма-распределения (табл. 6 приложения I). В частности, при $\alpha = -\frac{1}{2}$ абсцисса медианы, вычисленная по данным таблицы Е. Е. Слуцкого, оказывается равной 0,231, что в распределении Смирнова соответствует $x_0 = 2,16$.

Ввиду того что распределение Смирнова в первоначальном и в преобразованном видах весьма различно, то целесообразно сравнить медианы обоих видов распределения на графике (рис. 14).

Как известно, нами рассмотрено два распределения, не имеющих конечных моментов выше нулевого: симметричное распределение Коши и асимметричное распределение Смирнова. В действительности же таких распределений, конечно, очень много. Здесь уместно показать, что отсутствие конечных моментов у какого-либо распределения не исключает возможности его использования в теории ошибок. Только обычный и подробно нами изложенный метод моментов здесь уже, очевидно, не может быть использован по той простой причине, что хотя для любого конкретного ряда измерений, рассматриваемого как «выборку» из неизвестной генеральной совокупности, все моменты конечны, но если известно, что для генеральной совокупности, представленной кривой плотности вероятностей, их нет, то и оценку параметров этой кривой при помощи эмпирического ряда произвести нельзя. Следовательно, необходимо искать иной метод. В мате-

матической статистике такой метод разрабатывается и носит название метода квантилей. Хотя степень разработанности данного метода еще недостаточна для уверенного и надежного использования в практике, но ввиду его теоретического своеобразия и несомненной плодотворности в будущем при развитии теории измерения «малых» величин считаем полезным изложить этот метод в общем виде.

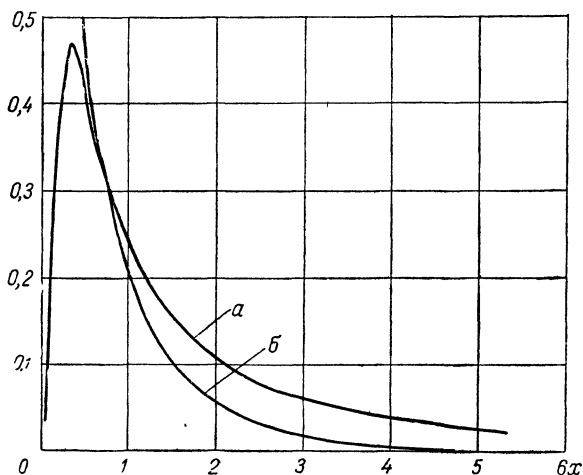


Рис. 14. График распределения Н. В. Смирнова (а) и обратного ему гамма-распределения (б).

Для всякой плотности распределения $p(x)$ существует и ее интегральная форма

$$P(x) = \int_{-\infty}^x p(x) dx,$$

причем

$$0 \leq P(x) \leq 1.$$

Решим приведенное уравнение в обратном направлении, т. е. найдем такое значение x , для которого P равно заданному числу. Это и будет квантиль x_p . Один из квантилей, а именно медиану распределения, мы рассматривали выше. Довольно часто пользуются нижним и верхним квартилями, для которых $P(x)$ равно соответственно $\frac{1}{4}$ и $\frac{3}{4}$. Вообще же могут быть квантили любого уровня вероятности x_p . Нахождение теоретических квантилей аналитически из выражения для плотности вероятности возможно лишь для немногих простейших распределений. Но числовым методом, т. е. путем табулирования функции $P(x)$, их можно найти всегда.

Что касается эмпирических квантилей, то схема их отыскания с помощью ступенчатого графика выборочного ряда, расположенного в порядке возрастания значения x и разбитого на ряд интервалов, приведена на рис. 15¹. Схема построена следующим образом. Совокупность из 200 измерений однородных объектов была разбита на 12 интервалов по величинам отклонений от среднего (принятого за нуль). Над каждым интервалом проведена горизонтальная площадка на высоте «накопленной частоты». Под этим термином понимаем сумму частот — от крайней (слева) частоты до данного интервала включительно.

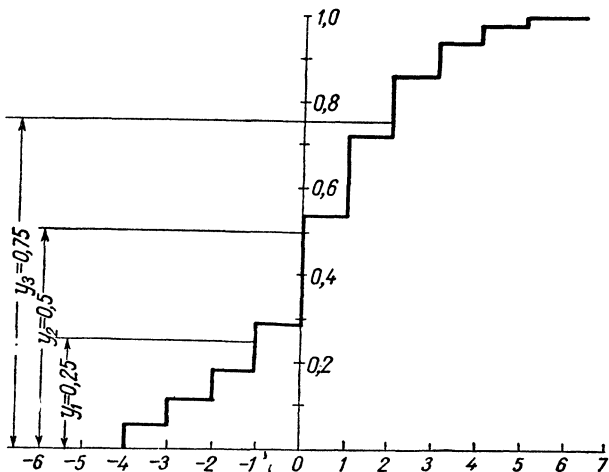


Рис. 15. Схема нахождения эмпирических квантилей.

На отдельной вертикали отложены накопленные частоты 0,25, 0,50 и 0,75. Проведенные на этих высотах горизонтальные прямые отсекают на ступенчатом графике соответственно абсциссы медианы и двух квантилей.

Когда число измерений значительно меньше и нет смысла разбивать измерения по интервалам, то график строят иначе: по оси абсцисс откладывают непосредственно измеренные значения x_j , а по оси ординат — те же накопленные частоты, но не в виде ступенчатой кривой, а в виде полигона.

Таким путем при достаточном объеме выборки можно определить значения нескольких квантилей, а затем использовать эти значения для нахождения параметров теоретического распределения совершенно так же, как ранее были использованы для этой цели эмпирические моменты.

Более подробное изложение указанного метода можно найти в работе [5, стр. 234—246]. Повторяем, что техническая разра-

¹ Указанная схема заимствована из работы [5, стр. 235].

ботка метода еще не находится на той высоте, которая необходима для свободного его применения, а потому в настоящее время практически удобнее исходить из такого закона распределения, при котором все моменты являются конечными. Это отнюдь не исключает использования указанного метода в отношении основного квантиля — медианы, которая при асимметричном распределении вообще должна быть предпочтена среднему арифметическому.

Поэтому в дальнейшем изложении уделим внимание двум хорошо известным распределениям: непрерывному гамма-распределению и дискретному распределению Пуассона. Мы упо-

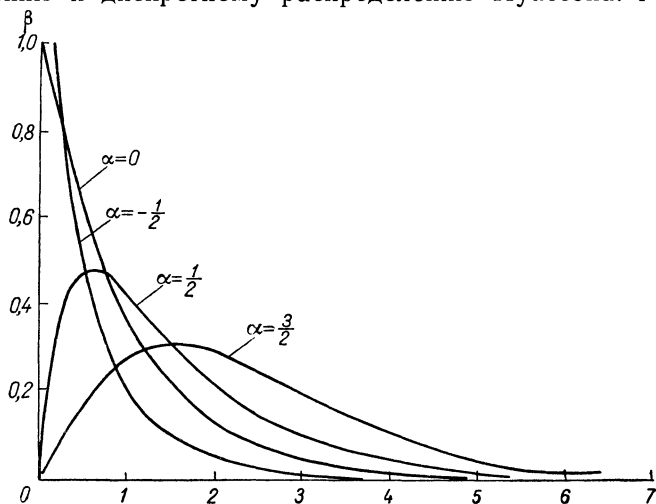


Рис. 16. Кривые гамма-распределения для $\beta=1$ и α , равной $-\frac{1}{2}$, 0 , $\frac{1}{2}$, $\frac{3}{2}$.

минаем и о последнем, хотя в теории ошибок гораздо чаще наблюдаются отклонения, принимающие в известных пределах любые значения, т. е. отклонения непрерывные, чем отклонения целочисленные, когда все измерения сводятся к простому подсчету одинаковых событий, например подсчет полного числа оборотов гидрометрической вертушки, подсчет числа песчинок в единице объема потока или числа альфа-частиц, вылетающих из радиоактивного вещества, и т. д. Именно при таких измерениях закон Пуассона (а иногда и более общий биномиальный закон) может найти нужное применение.

Однако все же главную роль в дальнейшем изложении данной главы будет играть гамма-распределение, график которого приведен на рис. 16.

Но прежде необходимо показать возможность асимметрии отклонений на нескольких примерах.

Представим себе, что измеряем штангенциркулем диаметр какого-либо вала. Примем вполне естественное допущение, что вал достаточно жесткий, чтобы не подвергаться сжатию от воздействия измерительного прибора. Совершенно очевидно, что любое отклонение штангенциркуля от диаметральной плоскости вала на какой-то угол даст всегда завышенное измерение его диаметра и никогда не даст заниженное. Если даже принять, что все остальные ошибки, зависящие от других причин, расположатся симметрично, то все же ряд измерений, суммирующий в се ошибки, будет непременно асимметричным, причем обязательно с положительной асимметрией.

Приведем пример из геодезии. При нивелировке равнинной местности возможны ошибки двух категорий: от ошибочной установки уровня нивелира и от невертикального положения рейки. Первая категория ошибок может иметь симметричное распределение, но вторая ни в каком случае, так как при любом невертикальном положении рейки всегда получится преувеличенное показание высоты.

Приведем пример из электрических измерений. Представим себе, что ищем величину удельного сопротивления химически чистой меди, которое, как известно, меньше, чем при ее сплаве с любым другим металлом. Совершенно ясно, что все образцы меди, измеряемые нами, не могут содержать вполне чистую медь. Следовательно, все измерения будут величинами, несколько бльшими по сравнению с той истинной величиной, которую ищем. Поэтому для всех измерений существует неравенство $x_j > a$.

Добавляя к ошибкам, обусловленным посторонними примесями в образцах меди, ошибки измерительного прибора (вероятно, симметричные), все же получим асимметрию суммарных ошибок.

Изучая сток какой-либо небольшой реки в засушливом районе, зададимся целью определить то максимальное количество годовых осадков, которое, полностью испаряясь, дает в результате нулевой сток. Аналогично предыдущему примеру можем рассматривать любое конкретно измеренное количество годовых осадков, дающее очень малый сток, как ошибочное (в сторону превышения) измерение искомого количества осадков нулевого стока. Что же касается тех измерений, которые дадут значения еще ниже искомого предела и, очевидно, также дадут нулевой сток, то они для нас интереса не представляют.

Следующий пример возьмем из области исследования появления первичных ледяных кристалликов при разных степенях охлаждения воды. Для этого произведем серии измерений температуры воды при различных отрицательных ее значениях (на-

пример, при $-0,001^\circ$, $-0,002^\circ$ и т. д.). В каждой серии будем записывать продолжительности не появления ледяных кристалликов. Каждая серия измерений даст, несомненно, то степенное измерение, которое выведено нами из закона Пуассона в главе IV. Действительно, обозначая через τ среднее значение интервала времени, внутри которого ледяной кристаллик еще не появлялся, и рассуждая так же, как в § 15, получим равенство аналогичное равенству (56):

$$p(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}},$$

которое и даст нам закон распределения продолжительностей не появления ледяных кристалликов. При нулевой продолжительности ($t=0$) $p = \frac{1}{\tau}$.

Нулевая же вероятность, естественно, относится к $t = \infty$. Вычисляя моменты, получим

$$\nu_1 = \tau; \quad \nu_2 = 2\tau^2; \quad \nu_3 = 6\tau^3,$$

откуда по формулам преобразования

$$\mu_2 = \tau^2; \quad \mu_3 = 2\tau^3; \quad \sigma = \tau.$$

Определим положение медианы:

$$\int_0^{t_0} \frac{1}{\tau} e^{-\frac{t}{\tau}} dt = \frac{1}{2},$$

откуда

$$t_0 = \tau \ln 2 = 0,695 \tau.$$

Как и должно быть, медиана и в данном случае меньше среднего арифметического.

Приведем пример из области измерения объемной концентрации взвешенных наносов. Если принять приближенно, что все наносы, проходящие через некоторый неподвижный объем, взятый внутри равномерного турбулентного потока, имеют одинаковый геометрический и гидравлический размер, то измерение объемной концентрации в лабораторных лотках можно легко осуществить с помощью высокочастотного кинематографа и фотоэлектрическим методом. Обычная концентрация наносов в реках, каналах и в лабораторных лотках очень мала. Среднее ее значение по высоте выражается тысячными долями, причем внизу, над дном потока, концентрация больше, чем вверху. На некоторой высоте имеется «потолок наносов», выше которого концентрация данной крупности равна нулю. Несколько ниже потолка наносов последовательные числа песчинок, одно-

временно находящихся внутри некоторого небольшого объема, будут выражаться рядом примерно следующего вида: 0, 0, 0, 1, 0, 2, 0, 0, 0, 1, 0, 0, 0, 0. Совершенно очевидно, что такое чередование должно давать дискретное распределение, ближе всего подходящее к закону Пуассона.

Если провести указанное выше исследование другим методом, например фотоэлектрическим, основанным на том, что просвечивание стеклянного лотка дает фотоэлектрический эффект, зависящий от прозрачности воды потока, а последняя зависит от среднего количества проносимых наносов, то в результате измерений получим непрерывное изменение индикатора прозрачности.

Поскольку оба метода должны дать равнозначные результаты, то естественно предположить, что распределение данных измерений по второму методу подчинится интерполированному закону Пуассона. Но так как в математической статистике неизвестно ни одной удачной попытки интерполяции этого дискретного закона, то не остается иного выхода, кроме применения наиболее близкого к нему гамма-распределения.

Общий вывод из всех предшествовавших рассуждений приводит к необходимости в дальнейшем изложении подробно рассмотреть вопрос о применимости гамма-распределения к теории ошибок при измерении «малых» величин.

§ 30. Применимость гамма-распределения

Это распределение получено в § 18 путем интерполяции биномиального закона в общем случае неравенства вероятностей p и q , т. е. в случае отсутствия симметрии. Теперь покажем, что оно может быть также получено из схемы суммирования элементарных ошибок, изложенной в § 26, лишь при ограничительном условии $p=q$. Кажется естественным предположить, что из той же схемы суммирования элементарных ошибок при условии отказа от указанного ограничения мы непременно получим более общее и асимметричное распределение, которое не может быть не чем иным, как хорошо известным нам гамма-распределением, что мы сейчас и покажем.

Будем рассуждать так же, как в § 26, причем отказ от предпосылки симметрии заставляет нас вместо распределения (103) принять общее выражение биномиального закона

$$P_n^m = C_n^m p^m q^{n-m}. \quad (122)$$

Что же касается выражения для алгебраической суммы элементарных ошибок, то оно, очевидно, не зависит от ограничения симметрией и будет тождественно с прежним:

$$x_m = (2m - n) \varepsilon. \quad (123)$$

Найдем выражение для отношения $\frac{\Delta P}{P}$, причем, так же как и в § 18, примем

$$\begin{aligned} \Delta P &= P_{m+1} - P_m; \\ P &= \frac{P_{m+1} + P_m}{2}. \end{aligned}$$

Не повторяя выкладок, изложенных в § 18, приведем лишь полученный результат:

$$\frac{\Delta P}{P} = 2 \frac{pn - m - q}{pn + (1 - 2p)m + q},$$

подставляя в который значение x_m из выражения (123), получим сначала

$$\frac{\Delta P}{P} = \frac{pn - \frac{x_m}{2\varepsilon} - \frac{n}{2} - q}{\frac{pn}{2} + \left(\frac{1}{2} - p\right)\left(\frac{x_m}{2\varepsilon} + \frac{n}{2}\right) + \frac{q^2}{2}},$$

а после умножения числителя и знаменателя на 2ε и отбрасывания всех произведений ε на конечные величины получим

$$\frac{\Delta P}{P} = - \frac{x_m - 2n\varepsilon\left(p - \frac{1}{2}\right)}{\frac{n\varepsilon}{2} - \left(p - \frac{1}{2}\right)x_m}. \quad (124)$$

По ранее обоснованным соображениям, заменяя дискретную переменную x_m непрерывной x , приращение ΔP производной $\frac{dy}{dx}$ (где y — плотность вероятности), получаем уравнение, мало отличающееся от уравнения (64) в § 18, а формально с ним тождественное и приводящее после интегрирования к известному гамма-распределению (65):

$$p(x) = \frac{x^\alpha e^{-\frac{x}{\beta}}}{\Gamma(\alpha + 1) \beta^{\alpha + 1}}. \quad (125)$$

Все дальнейшие выводы, приведенные в § 18 в виде уравнений (66) — (74), также будем давать без доказательств. К этим выводам относится частный случай при $\beta = 1$:

$$p(x) = \frac{x^\alpha e^{-x}}{\Gamma(\alpha + 1)} \quad (126)$$

и частный случай при $\alpha = 0$:

$$p(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}. \quad (127)$$

Напоминаем, что распределение (127) является в то же время частным случаем распределения Пуассона при $m=0$.

Таким образом, в нашем распоряжении для обработки измерений «малых» величин, когда симметрия отклонений явно отсутствует, имеется четыре распределения:

- 1) распределение Пуассона (46);
- 2) гамма-распределение общего вида (125);
- 3) то же, при $\beta=1$ (126);
- 4) то же, при $\alpha=0$ (127).

Приведем для сравнения таблицу их характеристик.

Таблица 6

№ уравнений	μ_1	μ_2	σ	μ_3	C_v	C_s
46	λ	λ	$\sqrt{\lambda}$	λ	$\frac{1}{\sqrt{\lambda}}$	$\frac{1}{\sqrt{\lambda}}$
125	$\beta(\alpha+1)$	$\beta^2(\alpha+1)$	$\beta\sqrt{\alpha+1}$	$2\beta^3(\alpha+1)$	$\frac{1}{\sqrt{\alpha+1}}$	$\frac{2}{\sqrt{\alpha+1}}$
126	$\alpha+1$	$\alpha+1$	$\sqrt{\alpha+1}$	$2\sqrt{\alpha+1}$	$\frac{1}{\sqrt{\alpha+1}}$	$\frac{2}{\sqrt{\alpha+1}}$
127	β	β^2	β	$2\beta^3$	1	2

Считая вопрос о применимости закона Пуассона в области дискретных измерений достаточно ясным из изложенного в § 13, а также из некоторых приведенных выше примеров, перейдем к более детальному анализу общего вида гамма-распределения и его частного случая при $\alpha=0$.

Прежде всего напомним, что из двух параметров α и β только первый определяет форму кривой распределения, второй же связан лишь с масштабом записи измерений. Известно, что любой ряд измерений можно путем преобразования:

$$y_j = \frac{x_j \bar{x}}{\sigma_x^2}$$

заменить таким рядом, у которого среднее и дисперсия равны, что и будет свойством упрощенного вида гамма-распределения (126).

Поэтому мы вправе ограничиться для непрерывных измерений применением уравнений (126) и (127).

Напомним, что параметр α может принимать любые положительные значения, а отрицательные значения — лишь в пределах от 0 до -1 . Причем нижний предел дает, как видно из дан-

ных табл. 6, коэффициент вариации, равный бесконечности. Левее этого предела, при $\alpha < -1$, последний становится даже мнимым, что лишено смысла.

Нетрудно подсчитать, что значения α , немногим бóльшие отрицательной единицы, дадут непомерно высокие значения C_v , например, при $\alpha = -\frac{3}{4}$ имеем $C_v = 2$; иными словами, стандартное отклонение вдвое превышает измеряемую величину, что неправдоподобно! Поэтому условно примем (в запас) за нижний предел $\alpha = -\frac{1}{2}$, при котором $C_v = 1,41$.

Верхний предел параметра α установить труднее, так как теоретически он может быть каким угодно большим, а практически его возрастание соответствует повышению точности измерений, которая для «малых» величин не может быть высокой. Если задаться, например, значением $C_v = 0,25$, когда измеряемая величина в четыре раза больше стандарта, то получим соответственное значение $\alpha = 15$.

Практически полученная величина будет более чем достаточна, так как при значениях α порядка 15 распределение отклонений уже настолько приближается к симметричному, что нет смысла отступать от обычного применения закона Гаусса. К тому же известно, что отклонения свыше трех стандартов при нормальном законе имеют вероятность исчезающе малую, т. е. меньше, чем 0,003. Поэтому мы вправе считать, что при $\alpha > 15$ кривая плотностей вероятностей гамма-распределения и нормального закона практически совпадут.

Таким образом, для наших целей достаточно использовать гамма-функцию в пределах значений параметра α от $-\frac{1}{2}$ до $+15$. В этих же пределах, очевидно, необходимо вычислить и следующие характеристики: 1) коэффициент вариации, 2) коэффициент асимметрии, 3) медиану.

Для характеристики медианы нам придется обратиться к данным таблиц, составленных Е. Е. Слуцким.¹ В этих таблицах приведены числовые значения интегралов

$$p(x) = \int_x^{\infty} \frac{x^\alpha e^{-x}}{\Gamma(\alpha+1)} dx,$$

носящих название «неполных гамма-функций».

Приведенный интеграл в пределе, когда $x=0$, обращается в единицу. При некотором же значении x , соответствующем медиане распределения, этот интеграл дает $\frac{1}{2}$. По данным таблиц

¹ Слуцкий Е. Е. Таблицы неполной гамма-функции. Изд. АН СССР, М., 1950.

Слущкого нетрудно определить значения этого интеграла при различных величинах параметра α . Подобные определения выполнены нами ниже.

Далее, по данным указанных таблиц можно определить теоретическое рассеяние, за которое теперь следует принимать значение абсолютного момента отклонений от медианы σ_0 . Это рассеяние вычисляется следующим путем:

$$\begin{aligned}\sigma_0 &= \int_0^{\infty} |x - x_0| p(x) dx = \int_0^{x_0} (x_0 - x) p(x) dx + \int_{x_0}^{\infty} (x - x_0) p(x) dx = \\ &= x_0 \int_0^{x_0} p(x) dx - x_0 \int_{x_0}^{\infty} p(x) dx - \int_0^{x_0} xp(x) dx + \int_{x_0}^{\infty} xp(x) dx.\end{aligned}$$

Первые два слагаемых правой части равенства, согласно определению медианы, дадут в сумме нуль, остальные два преобразуются следующим образом:

$$xp(x) = \frac{x^{\alpha+1} e^{-x}}{\Gamma(\alpha+1)} = (\alpha+1) \frac{x^{\alpha+1} e^{-x}}{\Gamma(\alpha+2)}.$$

Присоединяя к предыдущей сумме два слагаемых:

$$\pm \int_{x_0}^{\infty} xp(x) dx$$

и имея в виду равенство

$$\nu_1 = \int_0^{\infty} xp(x) dx = \alpha + 1,$$

получим

$$\sigma_0 = -\nu_1 + 2 \int_{x_0}^{\infty} xp(x) dx = (\alpha+1) \left\{ 2 \int_{x_0}^{\infty} \frac{x^{\alpha+1} e^{-x}}{\Gamma(\alpha+2)} dx - 1 \right\}. \quad (128)$$

Стоящий в фигурных скобках интеграл содержит отсчитанную от медианы обеспеченность гамма-распределения с показателем на единицу больше данного интеграла, т. е. не x^α , а $x^{\alpha+1}$.

Из таблиц Слущкого видно, что интеграл

$$\int_{x_0}^{\infty} \frac{x^{\alpha+1} e^{-x}}{\Gamma(\alpha+2)} dx > \frac{1}{2},$$

а следовательно, выражение в фигурных скобках равенства (128) всегда положительно, т. е. $\sigma_0 > 0$ (как и должно быть).

Для обработки измерений с помощью гамма-распределения приведенных характеристик — медианы x_0 и первого абсолютного

момента σ_0 — вполне достаточно. Однако для наглядности всего метода обработки полезно привести вычисленные нами по данным указанных таблиц обеспеченности среднего арифметического. В распределениях с положительной асимметрией они должны быть, очевидно, больше обеспеченности медианы, т. е. больше половины. Однако разность должна уменьшаться с возрастанием параметра α . При значениях α , близких к принятому нами верхнему пределу $\alpha = 15$, превышение обеспеченности среднего над обеспеченностью медианы, т. е. над $P(x_0)$, становится практически пренебрежимым, что мы и предполагали, выбирая указанный предел.

§ 31. Оценка медианы

При оценке результатов измерения «больших» величин мы имели дело с двумя практически независимыми величинами: средним (т. е. математическим ожиданием) и дисперсией (т. е. квадратом стандартного отклонения). Напоминаем о сказанном в § 26 и вторично поясняем, что в основном биномиальном распределении среднее и дисперсия отнюдь не независимы: первое равно np , второе — npq , следовательно, при симметрии отклонений дисперсия всегда равна половине среднего. Но, переходя от дискретного биномиального распределения к непрерывному и вводя в качестве дополнительной предпосылки схему суммирования огромного числа очень малых элементарных ошибок, мы тем самым как бы «отрываем» дисперсию от среднего и делаем ее независимой от него.

Несколько иначе обстоит дело при измерении «малых» величин с применением к их обработке медианы и гамма-распределения в однопараметрической его форме (126).

В этом случае имеется единственный параметр

$$v_1 = \sigma^2 = \alpha + 1,$$

определяющий и среднее и дисперсию.

Кроме того, используя данные таблиц Е. Е. Слуцкого, нами получено для каждой величины параметра α соответственные значения медианы (табл. 6 приложения I), по которым построен график изменения медианы (рис. 17). Значения медианы не могли быть получены аналитически. Но все же взятые из точных таблиц они являются теоретически вполне надежными.

Кроме того, непосредственно из ряда произведенных измерений, расположенного в возрастающем порядке, легко найти также медиану¹ и значение абсолютного первого момента, отнесенного к медиане. С другой стороны, среднее арифметическое того же ряда наблюдений дает нам, согласно только что приведенному

¹ Мы не будем особо останавливаться на случае четного числа измерений.

равенству, значение параметра α , из которого, пользуясь данными указанных таблиц или графиком (рис. 17), получаем значения медианы и абсолютного момента, соответствующие данной величине параметра α . Считать этот метод строго теоретическим нельзя, так как значение параметра α получено в этом случае из того же ряда наблюдений, т. е. эмпирически.

Таким образом, имеем два значения медианы, полученные из одной и той же совокупности измерений. В математической ста-

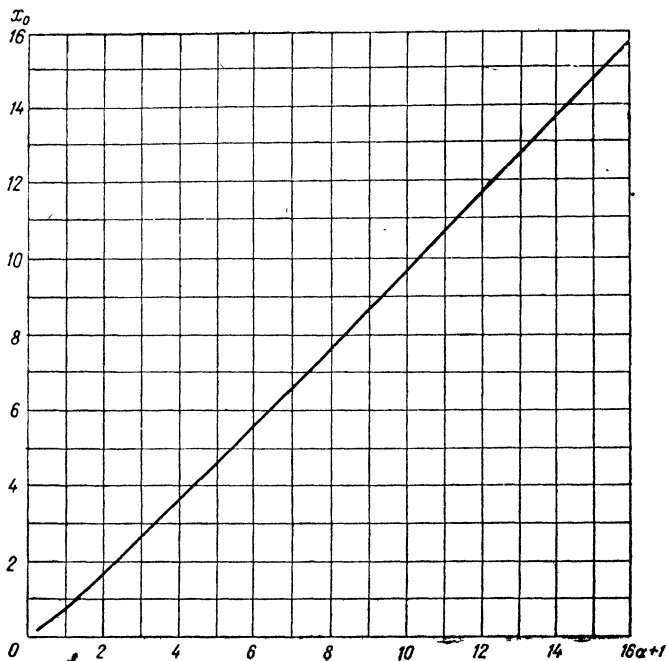


Рис. 17. График изменения медианы в гамма-распределении.

тистике не решается в общем случае вопрос о сравнительной надежности вычисления значений медианы по обоим изложенным методам. Можно лишь утверждать, что чем больше число измерений (и чем они точнее), тем расхождение между значениями медианы, вычисленными по обоим методам, будет меньше.

Начнем с оценки медианы, полученной по первому методу, специфичность которого в том, что нельзя предположить теоретический закон распределения, а нужно исходить исключительно из эмпирического ряда или (в лучшем случае) из многих эмпирических рядов. Для простоты будем считать, что все эти ряды или выборки имеют одинаковое число измерений. Будем рассматривать одну из выборок, преобразовав ее в вариацион-

ный ряд (т. е. расположив все измерения в возрастающем порядке их значений), причем в ряд, пересчитанный в безразмерный по известной формуле перехода

$$x_j = \frac{y_j \bar{y}}{\sigma_y^2}.$$

В дальнейшем будем рассматривать только безразмерный ряд $x_1, x_2, x_3, \dots, x_j, \dots, x_n$.

Медиана, полученная обычным методом нахождения того члена, который отстоит от концов ряда на равное число членов, не будет, очевидно, «истинной» медианой всей возможной совокупности измерений, т. е. медианой генеральной совокупности, включающей бесконечное число n -членных выборок. Но мы можем определить тот интервал, внутри которого с определенной вероятностью находится истинное значение медианы.¹

Зададимся такими пределами: пусть x_r — член ряда, левее которого расположены члены заведомо меньше медианы, а x_s — член, правее которого расположены все члены заведомо больше медианы. Отметим для ясности, что в каждой выборке эти концы будут иметь различные значения.

Поскольку в рассматриваемой совокупности измерений нас интересует лишь один признак, будет ли любой данный член больше или меньше медианы, то мы вправе применить к этой совокупности биномиальный закон (считая, например, что белый шар, вынутый из урны, соответствует члену, большему медианы, а черный — меньшему). Но так как по смыслу медианы вероятности быть больше или меньше ее равны, т. е. мы должны принять $p = q = \frac{1}{2}$, то общее выражение вероятности того, что из n произведенных измерений ровно k измерений дадут значения, меньшие (или большие) медианы, запишется в виде

$$P_n^k = C_n^k \frac{1}{2^n}.$$

Следовательно, применяя теорему сложения вероятностей, получим вероятности того, что все первые $(r-1)$ члены будут меньше медианы, а все последние члены от $(s+1)$ до n будут больше медианы соответственно в виде

$$\frac{1}{2^n} \sum_{k=1}^{k=r-1} C_n^k \quad \text{и} \quad \frac{1}{2^n} \sum_{k=s+1}^{k=n} C_n^k.$$

¹ Приведенный ниже расчет основан на соображениях, изложенных Н. В. Смирновым в работе [5, стр. 245—246].

Получаем вероятность того, что медиана расположена не левее x_r и не правее x_s :

$$P(x_{r-1} \leq x_0 \leq x_{s+1}) = \frac{1}{2^n} \sum_{k=r}^{k=s} C_n^k. \quad (129)$$

Изложенный метод дает возможность, задавшись числовым интервалом, определить вероятность того, что истинная медиана расположена не левее и не правее его концов. Априори ясно, что чем больше число измерений, тем при меньшем числе членов числового интервала его доверительная вероятность будет удовлетворять наперед заданной величине. При бесконечном числе измерений медиана, полученная прямым подсчетом, была бы истинной, т. е. имела бы вероятность, равную единице.

Переходим к рассмотрению второго метода оценки медианы, основанного на предположении, что бесконечное число измерений должно удовлетворять закону гамма-распределения. Как было сказано выше, в этом случае медиана в точности соответствует зависимости от параметра α . Однако само значение α , полученное из ограниченной выборки, не может быть, очевидно, «точным», а будет находиться с заданной вероятностью внутри какого-то доверительного интервала.

Использовать с этой целью распределение Стьюдента, как это делалось при измерении «больших» величин, очевидно, нельзя ввиду его симметричности при всех степенях свободы. В этом случае можно применить хиквадрат-распределение, которое при малых степенях свободы обладает ясно выраженной асимметрией. Однако при измерении «малых» величин, когда относительная ошибка сравнительно велика, стремиться к большому числу измерений вообще не имеет смысла, так как это может дать лишь фиктивное повышение точности.

Приравниваем

$$\chi^2 = \frac{ks^2}{\sigma^2},$$

где $k=n-1$, s^2 — эмпирическая (выборочная) дисперсия, определяемая равенством

$$s^2 = \frac{1}{n} \sum (x_j - \bar{x})^2,$$

а σ — теоретическая дисперсия генеральной совокупности, в рассматриваемом случае равная $(\alpha+1)$. Запишем неравенство, определяющее доверительный интервал

$$s^2 - \varepsilon_1 \leq \alpha + 1 \leq s^2 + \varepsilon_2.$$

Здесь ε_1 и ε_2 — расстояния от эмпирической дисперсии до концов интервала. Мы обозначим их различно, так как вследст-

вие асимметричности распределения они не могут быть равны при одной и той же доверительной вероятности.

Разделим теперь величину ks^2 на все части приведенного неравенства и введем сокращенные обозначения:

$$\frac{\varepsilon_1}{s^2} = q_1; \quad \frac{\varepsilon_2}{s^2} = q_2.$$

Получим, имея в виду равенство

$$\sigma^2 = \alpha + 1,$$

следующее неравенство:

$$\frac{k}{1+q_2} \leq \chi^2 \leq \frac{k}{1-q_1}.$$

Или, наконец, вводя

$$\frac{k}{1+q_2} = \chi_1; \quad \frac{k}{1-q_1} = \chi_2,$$

запишем окончательное равенство в виде

$$P \{ \chi_1^2 \leq \chi^2 \leq \chi_2^2 \} = \delta. \quad (130)$$

Распределение случайной величины

$$\chi^2 = \frac{ks^2}{\sigma^2},$$

представляющей собой отношение суммы эмпирических дисперсий Σs_j^2 к теоретической дисперсии, нам известно из § 20 и уравнения (74). В табл. 6 приложения I, являющейся частью таблицы, помещенной в работе [5 стр. 503—504], приведены значения интеграла

$$P(\chi^2 > \chi_q^2) = \int_{\chi_q^2}^{+\infty} p(k, \chi^2) d\chi^2 = \delta.$$

Данные табл. 6 приложения I дают возможность быстро решить уравнение (130). Например, при $k=6$ и $\delta=0,70$ условие $\chi^2 > \chi_1^2$ дает $\chi_1^2=3,8$, а условие $\chi^2 < \chi_2^2$ надо заменить равнозначным ему условием

$$P(\chi^2 < \chi_2^2) = 1 - P(\chi^2 > \chi_2^2) = 0,30,$$

после чего получаем по табл. 6 $\chi_2^2 = 7,2$.

Таким образом, мы пришли к выводу, что соответствующее истинному значению параметра $(\alpha+1)$ значение χ^2 располагается в интервале между 3,8 и 7,2. Отсюда уже нетрудно определить границы и для самого параметра $(\alpha+1)$, согласно неравенству

$$s^2 - \varepsilon_1 < \alpha + 1 < s^2 + \varepsilon_2.$$

Переход к интересующему нас доверительному интервалу для медианы находим по графической зависимости (рис. 17), откладывая абсциссы $s^2 - \varepsilon_1$ и $s^2 + \varepsilon_2$ и обозначая соответственные ординаты через a_1 и a_2 .

Для облегчения расчетов по обработке измерений «малых» величин, когда отклонения от центра не удовлетворяют условиям симметрии и возникает предположение, что придется перейти от закона Гаусса к гамма-распределению, нами составлена на основе данных вышеупомянутых таблиц Е. Е. Слуцкого табл. 6 приложения I, в которой приведены основные характеристики этого распределения, а также и некоторые другие данные, непосредственно не используемые в расчетах, но полезные для лучшего понимания сущности этого распределения. Именно в этой же табл. 6 даны для значений α от $-\frac{1}{2}$ до $+15$ (при более высоких значениях α практически допустим возврат к нормальному закону) следующие характеристики: среднее арифметическое равное дисперсии и равное $(\alpha + 1)$, коэффициент вариации C_v , (коэффициент асимметрии $C_s = 2C_v$ не приводится), медиана x_0 , превышение среднего над медианой $\bar{x} - x_0$ и, наконец, обеспеченность среднего $P(\bar{x})$. Из данных табл. 6 видно довольно быстрое возрастание медианы со средним $(\alpha + 1)$ и очень медленное возрастание превышения среднего над медианой. Это превышение при $\alpha + 1 = 16$, по-видимому, становится постоянной величиной. Данные последней графы табл. 6 приложения I показывают, что превышение обеспеченности среднего над медианой продолжает расти до максимального значения $\alpha = 15$. Но поскольку теоретически ясно, что совпасть эти обеспеченности могут только в пределе, при $\alpha = \infty$, то нужно предполагать в области значений

$$P(\bar{x}) - P(x_0) < 0,33$$

какую-то асимптоту.

Для интерполяции значений медианы в тех же пределах значений α на рис. 17 приведен соответствующий график. Далее покажем на примерах, как надо производить обработку измерений в случаях, когда возникает предположение или имеется уверенность в асимметричном распределении отклонений.

В качестве первого примера возьмем измерения годового стока на какой-либо стоковой станции. Поскольку сток не может быть «отрицательным», то отклонения от среднего значения стока влево ограничены, а, следовательно, асимметрия, по крайней мере, существенных отклонений неизбежна. Но в какой мере асимметрия может повлиять на положение центра, следует в каждом отдельном случае выяснить. Напомним, что в данном примере «отклонения» от среднего не являются полностью

«ошибками измерения», но суммируются с реальными отклонениями самой величины.

В тех случаях, когда эти отклонения могут считаться вполне «случайными», их обработку формально надо вести теми же методами, что и «ошибки измерения» в собственном смысле (т. е. зависящие только от измерительного прибора).

Имеем измерения стока в л/сек. с 1 км² за 9 лет на р. Заказе (бассейн р. Москвы). Составляем вариационный ряд измерений: 3,55; 3,67; 3,72; 3,75; 3,92; 4,95; 6,40; 6,70; 6,80. Среднее ряда $\bar{x}=4,82$, медиана $x_0=3,92$, дисперсия $s^2=1,79$, первый абсолютный момент, отсчитанный от медианы, $s=1,23$. Абсолютный момент приведен здесь лишь для сравнения его со стандартом, который равен $\sqrt{1,79}=1,34$. Задаемся (произвольно) доверительным интервалом для медианы, а именно будем считать, что ее «истинное» значение находится между третьим и седьмым (т. е. от 3,72 до 6,40) членами ряда.

По формуле (129) соответственная вероятность выразится следующим образом:

$$\frac{9!}{2^9} \left\{ \frac{1}{3!6!} + \frac{1}{4!5!} + \frac{1}{5!4!} + \frac{1}{6!3!} + \frac{1}{7!2!} \right\} \approx 0,8.$$

Сделаем расчет по второму способу, т. е. исходя из предположения, что наш ряд подчиняется гамма-распределению. Преобразуем его к безразмерному виду умножением каждого члена ряда на отношение $\frac{y}{\sigma_y^2} = 2,7$. Попутно отметим, что такое преобразование не является обязательным, так как можно оперировать и с гамма-распределением в его двухпараметрической форме и оценивать оба его параметра α и β . Но там, где такой переход возможен, лучше им пользоваться, потому что он облегчает весь расчет.

В данном случае, как увидим далее, такое преобразование приводит к численным значениям, не выходящим за пределы тех значений параметра α , которые имеются в табл. 6 приложения I.

После указанного пересчета получаем следующий безразмерный вариационный ряд: 9,58; 9,91; 10,04; 10,12; 10,58; 13,37; 17,28; 18,09; 18,36, в котором $\bar{x}=s^2=13,00$; $x_0=10,58$; $s_0=3,82$.

Но для полученного значения параметра $\alpha=x-1=12$ значение медианы, взятое из табл. 6 приложения I, составляет 12,67, что значительно превышает приведенное выше значение 10,58, взятое непосредственно из ряда. Впрочем, следует помнить, что значение параметра α получено из того же (короткого) ряда и, следовательно, также содержит ошибку. Поэтому мы должны

прежде всего оценить истинное значение параметра α . Используем с этой целью случайную величину

$$\chi^2 = \frac{ks^2}{\sigma^2},$$

где k — число степеней свободы, равное 8; s^2 — эмпирическая дисперсия нашего ряда, равная 13; σ^2 — неизвестная дисперсия генеральной совокупности.

В табл. 5 приложения I даны для разных значений k и δ значения χ_q^2 , удовлетворяющие равенству

$$\delta = P\{\chi^2 > \chi_q^2\} = \int_{\chi_q^2}^{\infty} p(\chi^2, k) d\chi^2.$$

В нашем случае имеем $k=8$, а δ принимаем равным тому значению, которое получилось у нас по первому способу, т. е. 0,8. Для нижнего предела по табл. 5 приложения I $\chi_1^2=4,59$. Для верхнего предела используем очевидное равенство

$$P\{\chi^2 < \chi_q^2\} = 1 - P\{\chi^2 > \chi_q^2\} = 0,2.$$

Из этой же таблицы получаем $\chi_2^2 = 11,00$. Из равенства

$$\chi_2^2 = \frac{ks^2}{\sigma^2} = \frac{ks^2}{\alpha + 1}$$

получаем пределы для $(\alpha + 1)$:

$$9 < \alpha + 1 < 22,$$

соответствующие пределам для медианы 8,7 и 21,3. Из полученного видно, что второй способ превысил верхний предел медианы, который по первому способу был равен 17,3. Нижний же предел получился весьма близкий в обоих способах. Такое соотношение пределов вообще характерно для случаев асимметрии отклонений: нижняя граница доверительного интервала в условиях положительной асимметрии должна быть более устойчивой, чем верхняя, так как вероятность занижающих отклонений меньше, чем завышающих. Из приведенного примера видно весьма неустойчивое значение медианы, объясняющееся совпадением двух факторов: при малой асимметрии и краткости ряда измерений ошибки и среднего и медианы могут быть значительны. Так как указанные ошибки независимы, то их совпадение или несовпадение носит случайный характер. В данном случае верхняя граница доверительного интервала для медианы превышает среднее, что вообще противоречит условиям распределения с положительной асимметрией. Несколько лучшее соотношение в аналогичных случаях мы должны получить для бассейнов в засушливом районе, где разность между средним и медианой

относительно больше, чем во влажных районах. В качестве второго примера в табл. 7 приведены количества осадков и величины стока по Дубовской стоковой станции.

Таблица 7

Год	Осадки, мм	Сток, мм	Год	Осадки, мм	Сток, мм
1950	350	0,0	1955	455	0,5
1951	370	29,6	1956	477	102
1952	461	15,2	1957	250	12
1953	306	66,5	1958	546	6,1
1954	313	2,3	1959	274	6,2

Из данных табл. 7 прежде всего видно явное отсутствие какой-либо четкой связи между осадками и стоком: при осадках 30 мм сток имеет значение 66 мм, а при бóльших осадках, равных 350 мм, сток почему-то уменьшается до нуля! Это может быть обусловлено неучетом больших колебаний в водообмене грунтовыми водами с соседними бассейнами, что при малой площади рассматриваемого бассейна (20 км²) играет большую роль. Поэтому, оставляя в данном случае в стороне вопрос о связи осадков со стоком, будем рассматривать лишь сток, причисляя к ошибкам его измерения колебания водообмена. Перестроив ряд стока (табл. 3) в вариационный ряд, получим 0,0; 0,5; 2,3; 6,1; 6,2; 12,0; 15,2; 29,6; 66,5; 102. Последнюю цифру 102, как сильно завышенную, разумно исключить. Тогда медиана будет равна 6,2, среднее составит 15,3, а дисперсия — 42,9, т. е. $\sigma=6,5$. Для сравнения приведем также значение первого абсолютного момента, отсчитанного от медианы $\sigma_0=13,0$.

Вычислим по формуле (129) аналогично предыдущему примеру вероятность того, что «истинная» медиана находится между третьим и седьмым членами ряда (т. е. между 2,3 и 15,2). Получаем, очевидно, ее значение, равное (с округлением) 0,8.

Для расчета по второму способу преобразуем вариационный ряд умножением каждого его члена на $\frac{y}{\sigma y^2} = 0,035$. Получим 0,0; 0,011; 0,081; 0,213; 0,217; 0,420; 0,632; 1,036; 2,324. Данный ряд по предположению должен удовлетворять однопараметрическому гамма-распределению. Следовательно, его среднее и дисперсия должны удовлетворять равенству

$$\bar{x} = \sigma_x^2 = \alpha + 1 = 0,536.$$

Округляя значение 0,536, получим $\alpha = -0,5$, что, согласно данным табл. 6 приложения I, соответствует значению медианы, равному 0,231, т. е. достаточно близкому к медиане преобразо-

ванного ряда, составляющей 0,217. Доверительный интервал для истинной медианы находится, как и в предыдущем примере, с помощью хиквадрат-распределения. Имеем

$$\chi^2 = \frac{ks^2}{\sigma^2},$$

где s^2 — эмпирическая дисперсия нашей совокупности, равная 0,43 (дисперсию первоначального ряда мы умножим на 0,035²). Число степеней свободы равно 8; доверительную вероятность примем равной ее значению по первому способу, т. е. 0,8. Отсюда, согласно данным табл. 5 приложения I, пределы значений $(\alpha+1)$ будут $0,38 < \alpha+1 < 0,91$. Соответственные пределы для медианы по табл. 6 приложения I (берем их с округлением) будут равны 0,23 и 0,69. Совпадение получилось гораздо лучше, чем в предыдущем примере, но характерно, что, при точном совпадении нижних пределов по обоим способам, верхний предел по второму способу значительно превышает значение по первому способу.

Следующий пример приведем из области изменения вылета альфа-частиц из радиоактивного вещества. Эти весьма точные измерения дают исключительно отчетливое представление об асимметричных отклонениях. Заимствуем данные измерений, произведенных английским физиком Хэллидеем, из монографии А. Уорсинга и Дж. Геффнера [12, стр. 193]. График, построенный по этим измерениям, был приведен в § 15 в связи с вопросом применимости распределения Пуассона (см. рис. 7). Число измерений вылета альфа-частиц равнялось 3133, при этом точно фиксировалось время вылета каждой частицы. Первоначальная обработка измерений заключалась в подсчете числа вылетов в секунду за каждый интервал в полсекунды. Результаты измерений приведены в табл. 8.

Таблица 8

Интервалы	Частоты	Интервалы	Частоты
0—0,5	0,71	3,0—3,5	0,40
0,5—1,0	0,45	3,5—4,0	0,30
1,0—1,5	0,29	4,0—4,5	0,20
1,5—2,0	0,19	4,5—5,0	0,15
2,0—2,5	0,14	5,0—5,5	0,10
2,5—3,0	0,60	5,5—6,0	0,00

Среднее значение τ промежутка между двумя вылетами в результате обработки измерений получилось равным 1,17. Теоретическое значение медианы соответственно равно $t_0=0,69$. $\tau = =0,81$.

Применим к оценке медианы второй из изложенных выше способов (первый, как более грубый, а в данном случае и более громоздкий, здесь малоприменим). Означенные измерения должны, как указано в § 15, подчиняться степенному закону

$$p(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}},$$

иначе говоря, гамма-распределению двухпараметрической формы при $\alpha=0$, $\beta=\tau$. Дисперсия этого распределения равна, как известно, $\beta^2=\tau^2$, т. е. квадрату среднего. Оценим эту дисперсию, пользуясь опять хиквадрат-распределением и принимая

$$\chi^2 = \frac{k\tau^2}{\sigma^2}.$$

Здесь τ^2 — эмпирическая дисперсия, равная $1,17^2=1,37$; σ^2 — истинная неизвестная дисперсия. Число степеней свободы k ввиду очень большого числа измерений примем максимальным по табл. 6 приложения I, т. е. равным 30. Доверительную вероятность также возьмем высокой, равной 0,99. Будем иметь $\chi_1^2 = 15,0$; $\chi_2^2 = 54,0$, откуда, составив равенство

$$\sigma^2 = \frac{k\tau^2}{\chi^2} = \frac{41,10}{\chi^2},$$

получим границы доверительного интервала для дисперсии:

$$0,76 \leq \sigma^2 \leq 2,73.$$

Если для сравнения уменьшим доверительную вероятность до 0,9, то получим соответственно

$$\chi_1^2 = 20,6; \quad \chi_2^2 = 40,3$$

и соответственно

$$1,02 \leq \sigma^2 \leq 1,99.$$

Извлекая корень квадратный и умножая на 0,69, получим границы доверительного интервала для медианы:

$$0,69 \leq t_0 \leq 0,97.$$

В данном случае мы видим, во-первых, что экспериментальная медиана, равная 0,81, расположена приблизительно в середине доверительного интервала и, во-вторых, что среднее арифметическое, полученное непосредственно из измерений и составляющее 1,17, расположено далеко от верхней границы этого интервала, следовательно, асимметрия отклонений выражена здесь весьма отчетливо.

Третий пример приведем также из области неточных измерений. На некоторой реке Средне-Русской возвышенности было произведено в конце зимы большое количество замеров толщины льда и высоты снежного покрова над ним. Одно из измерений толщины льда, произведенное в разных точках поперечного створа, приводится ниже в форме вариационного ряда (в мет-

рах) 0,82; 0,83; 0,83; 0,85; 0,85; 0,87; 0,87; 0,96; 1,07; 1,24; 1,24; 1,25; 1,30. Здесь $x_0=0,87$; $\bar{x}=1,00$; $s^2=0,030$; $s_0=0,17$.

Зададимся для медианы интервалом вероятности от пятого до девятого члена ряда (т. е. от 0,85 до 1,07). По формуле (129) эта вероятность будет равна

$$\frac{13!}{2^{13}} \left\{ \frac{2}{518!} + \frac{2}{617!} + \frac{2}{914!} \right\} \approx 0,87.$$

Переходя к расчету по второму способу, основанному на допущении, что наш ряд подчиняется гамма-распределению, заметим, что пересчет ряда к безразмерному виду дал бы среднее, равное 33, иначе говоря, $\alpha=32$. Но нам известно, что при таких высоких значениях этого параметра роль асимметрии становится практически пренебрежимой, и мы поступим правильно, если в данном случае сделаем расчет по нормальному закону, приняв за среднее $\bar{x}=1,00$, а за эмпирический стандарт $s=0,17$.

Методика указанного расчета изложена в § 27. Поэтому не будем на ней подробно останавливаться, а дадим краткий расчет, исходя из того, что оценке подлежит как «истинное» значение, так и «истинный» стандарт. Для оценки первого воспользуемся таблицей Стьюдента — Фишера (табл. 3 приложения I) и примем за доверительную вероятность полученное нами выше значение, округляя его до 0,90. При $k=12$ $z=1,78$, откуда расстояния концов доверительного интервала от эмпирического среднего получаются равными

$$\pm z \frac{s}{\sqrt{n}} = \pm 0,084.$$

Так как исходные значения в примере, приведенном в § 27, даны в метрах, то длина интервала с вероятностью 0,9 получается равной 16 см.

Для оценки генерального стандарта используем, согласно методике В. И. Романовского, хи-распределение (112) и данные табл. 4 приложения I. Имеем

$$\chi = \frac{s \sqrt{k}}{\sigma}; \quad \chi_1 = \frac{\sqrt{k}}{1+q}; \quad \chi_2 = \frac{\sqrt{k}}{1-q}.$$

По данным табл. 4 для $k=12$ получаем $q = \frac{\varepsilon}{s} = 0,51$, откуда концы доверительного интервала будут иметь значения $\varepsilon = \pm 0,087$.

Последний пример возьмем из области измерений коэффициента вязкости методом определения скорости падения шара. При очень малых числах Рейнольдса $Re = \frac{vr}{\nu}$ должен иметь силу закон Стокса

$$v = \frac{2gr^2 \left(\frac{\rho'}{\rho} - 1 \right)}{g\nu}.$$

С возрастанием числа Рейнольдса начнет сказываться влияние сил инерции обтекающих струй, отчето суммарное сопротивление движению падающего шара будет, очевидно, возрастать. Иначе говоря, если при не очень малых числах Рейнольдса определять коэффициент вязкости по формуле Стокса, то получим неверное, а именно завышенное его значение, включающее в себя в скрытом виде влияние тех самых сил инерции, которые, как известно, отбрасывались при выводе формулы Стокса. Ра-венство, связывающее фиктивный, т. е. завышенный, коэффициент вязкости с его истинным значением¹, приведено без доказательства:

$$\nu' = \nu + 0,021 (\nu r).$$

Перейдем далее к конкретному примеру.

Определяем коэффициент вязкости масла «Велосит Л». По измерениям падения нескольких шариков получены следующие семь его значений (мы расположили их в порядке возрастающих значений): 0,22; 0,23; 0,24; 0,29; 0,36; 0,46; 0,54. Все значения, очевидно, завышены, так как ошибка от недоучета сил инерции не может быть отрицательной. Следует найти то значение ν' , которое совпадает с ν , т. е. будет соответствовать падению «нулевого шарика». Задача упрощается тем, что наименьшие значения, полученные при измерении самых малых шариков, довольно хорошо совместились с прямой линией. Интерполируя прямую линию до пересечения с осью ординат, получим значение 0,18, которое должны принять за истинное значение ν .

§ 32. Оценка прогнозов

К измерениям «малых» величин следует отнести и прогнозирование величин, связанных с явлениями, которые еще не наступили, но, согласно ряду предшествующих ему признаков, должны наступить.

Прогнозы бывают двух категорий: функциональные и стохастические. Первые определяют характеристики ожидаемого явления с вероятностью «единица», например прогнозы солнечных затмений и других астрономических явлений. Их мы касаться не будем. Вторые прогнозы определяют характеристики ожидаемого явления с некоторой вероятностью. В этих прогнозах утверждается лишь то, что прогнозируемая величина будет с такой-то вероятностью находиться в таком-то интервале значений.

К стохастическим прогнозам относятся в первую очередь прогнозы погоды, или речных паводков, или вскрытия ледяного покрова на реке, т. е. прогнозы метеорологические и гидрологи-

¹ См. Великанов М. А. и Зегжда А. П. Равномерное движение шара в жидкости. Изв. Науч. мелиорат. ин-та, вып. XVIII, 1929.

ческие. Оценка стохастического прогноза (если он имеет количественный характер) всегда сводится к предвычислению интервала значений какой-то величины с какой-то доверительной вероятностью. Чем этот интервал меньше или чем отвечающая ему вероятность больше, тем прогноз считается «надежнее».

Общее представление о стохастических связях будет дано в следующей главе. Сейчас нам достаточно знать, что эти связи определяют не величины, а законы распределения случайных величин, так как они могут принимать с различными вероятностями различные значения.

Общность прогнозов с измерением малых величин состоит в том, что как в прогнозах, так при измерении ищем границы доверительного интервала значений какой-то величины. Различия между ними в том, что при измерениях различные значения обуславливаются неточностью измерителя, а в прогнозе они реально возможны и реально осуществимы (с какой-то вероятностью).

Функциональные прогнозы в данной книге рассматривать не будем. Что касается стохастических прогнозов, то приведем методику их оценки.

Нам нет надобности проводить сколько-нибудь четкое различие между гидрологическими и метеорологическими прогнозами, так как в большинстве случаев они методически объединяются. Действительно, прогноз снегового половодья является, в сущности, прогнозом снеготаяния, зависящего в основном от хода температуры воздуха, а прогноз ливневого паводка сводится к прогнозу ливня. Прогноз вскрытия льда на реке является прогнозом повышения температуры (при известной толщине льда), а прогноз меженного расхода реки — прогнозом летних осадков.

Таким образом, составление прогнозов производится в следующем порядке:

1. Сбор данных о факторах, предшествующих ожидаемому явлению, влияющему на него, а также измерение этих факторов.

2. Составление расчетной схемы стохастической связи между факторами, предшествующими ожидаемому явлению, и самим явлением.

3. Составление конкретного прогноза в данных условиях и его оценка.

Далее рассмотрим лишь последнее звено конкретного прогноза, считая, что первые два детально излагаются в специальной литературе по метеорологии и гидрологии.

Для конкретизации изложенного рассмотрим подробно процесс формирования снегового половодья. Для того чтобы охватить его полностью, необходимо знать:

а) распределение снежного покрова по площади бассейна с учетом плотности снега;

б) распределение по площади и ход изменения во времени температуры воздуха;

в) рельеф бассейна и распределение в нем растительного покрова и почво-грунтов.

Нами перечислены все главнейшие факторы, обуславливающие весь временный процесс таяния снега, его испарение и инфильтрацию, стекание в реку талой воды и формирование паводочной волны. Но если бы даже мы могли в конкретном случае количественно знать все эти факторы (а это немыслимо!), то и тогда не смогли бы дать сколько-нибудь точный расчет объема и движения паводочной волны по той простой причине, что все перечисленные составные процессы: таяние, испарение, инфильтрация и стекание — являются, по существу, случайными, так как они зависят от множества мельчайших причин, учесть которые полностью невозможно.

Аналогичное нагромождение большого количества мельчайших причин мы имеем и в случае ливневого паводка, когда наибольшая неопределенность заключается во временном и пространственном распределении ливневых осадков, а также в потерях воды на инфильтрацию из всего объема выпавших осадков. В обоих примерах прогнозируемая величина представляет собой сумму многочисленных мельчайших слагаемых, и, так же как в теории ошибок, эта сумма не может рассматриваться иначе, как находящаяся с какой-то вероятностью внутри каких-то границ.

Нам кажется, что приведенных двух примеров вполне достаточно, чтобы построить общую схему стохастического прогноза.

В общем виде прогнозируемая величина зависит от m факторов, из которых одни измеряются, а другие прогнозируются. Свяжем эти факторы с величинами $X_1, X_2, \dots, X_j, \dots, X_m$. Каждая из приведенных величин представлена рядом измерений и имеет среднее значение \bar{x}_j и дисперсию s_j^2 . Прежде всего зададим вопрос, какова будет дисперсия прогнозируемой величины

$$U = F\{X_1, X_2, \dots, X_j, \dots, X_m\},$$

если известен вид функции U и дисперсия всех величин X_j . Ответ на данный вопрос будет приведен в § 36 [см. уравнение (163)]. А пока читатель пусть примет на веру следующее выражение:

$$\sigma_u^2 = \sum_1^m \left(\frac{\partial u}{\partial x_j}\right)^2 \sigma_j^2 + 2 \sum \left(\frac{\partial u}{\partial x_j}\right) \left(\frac{\partial u}{\partial x_k}\right) \overline{(x_j - \bar{x}_j)(x_k - \bar{x}_k)}.$$

В случае, когда все факторы являются независимыми, второе слагаемое равно нулю. В тех случаях, когда лишь некоторые факторы взаимосвязаны, суммирование во втором слагаемом распространяется только на них.

Таким образом, зная дисперсии всех величин, характеризующих факторы, от которых зависит прогнозируемое явление, а также значение ковариаций между ними, принципиально возможно вычислить дисперсию прогнозируемой величины. Зависимость, о которой идет здесь речь, имеет форму функциональной связи (иначе ее нельзя было бы дифференцировать). По существу, эта связь всегда эмпирическая и приближенная, т. е. дает лишь приближенные значения прогнозируемой величины при заданных значениях обуславливающих факторов.

Однако знание дисперсии прогнозируемой величины отнюдь не характеризует полностью качества и надежности прогноза. Последняя определяется границами того интервала, внутри которого с заданной вероятностью она находится. Для определения указанного интервала необходимо знать или задаться плотностью вероятности прогнозируемой величины. И здесь, имея в виду неизбежную грубость прогноза, ошибка которого всегда имеет тот же порядок, что и сама прогнозируемая величина, нам приходится, в соответствии со всем вышеизложенным, сразу же отказаться как от традиционного закона Гаусса, так и от столь же традиционного среднего арифметического и перейти на медиану и на простейшее из асимметричных распределений, т. е. на гамма-распределение.

Будем считать, что прогноз данной категории, относящийся к данным конкретным условиям (например, прогноз пика половодья р. Оки у г. Калуги в 1961 г.), составляется не впервые, а имеет многолетнюю историю предшествующих прогнозов, для которых в свое время измерялись все факторы, обуславливающие прогнозируемое явление, а факторы, которые не измерялись, прогнозировались. Кроме того, необходимо принять, что теоретическая схема прогноза разработана достаточно для того, чтобы дать хотя бы в весьма приближенном виде структуру функции U . Без какого-либо представления об этой структуре прогноз вообще невозможен.

Поэтому, не вдаваясь в подробности, будем считать, что дисперсия σ_U^2 нами вычислена. Если к этому добавить вышесказанное допущение, что распределение величины U подчиняется закону гамма-распределения, то из равенства $s_U^2 = \alpha + 1$ получим эмпирическое значение параметра α .

Оценка истинного значения параметра α производится методом, изложенным выше в этом же параграфе, с применением хиквадрат-распределения. Методика перехода от оценки $(\alpha + 1)$ к оценке медианы также изложена выше. Здесь неуместно подробно излагать вопросы прогнозов; мы хотели лишь показать идейное сродство прогнозов с измерением «малых» величин. Тот факт, что при составлении прогноза мы измеряем не то, что есть, а то, что еще будет, не меняет дела, так как это является лишь частным случаем косвенных измерений. Гораздо суще-

ственнее то, что при составлении прогнозов часто имеем слишком малое число измерений обуславливающих факторов, иногда даже лишь одно измерение. Поясним для этого случая, что для получения дисперсии необходимо иметь хотя бы два измерения x_1 и x_2 . Тогда

$$D(x) = s^2 = \frac{\left(\frac{x_1 + x_2}{2} - x_1\right)^2 + \left(\frac{x_1 + x_2}{2} - x_2\right)^2}{2} = \frac{(x_1 - x_2)^2}{4}.$$

При одном же измерении никакой дисперсии вообще быть не может.

В связи с этим полезно напомнить, что оба используемые при оценках результатов измерений распределения: Стьюдента и «хиквадрат» имеют смысл лишь при $k \geq 1$, при одном же измерении $k = n - 1 = 0$, что лишено смысла.

Поэтому вышеизложенное применимо лишь в случаях по меньшей мере двукратного измерения всех обуславливающих факторов (включая и прогнозируемые факторы).

Разработанная автором книги приближенная теория снегового половодья приводит к равенству

$$Q_{\max} = CHT^{-1}, \quad (131)$$

где C — величина приблизительно постоянная для данного бассейна, H — снегозапас в слое воды для данного года, T — продолжительность снеготаяния данного года, T^{-1} — соответствующий индикатор интенсивности снеготаяния.

Подсчеты, проведенные для некоторых бассейнов, показали, что корреляционная связь между H и T^{-1} очень низка (порядка $r = 0,40$). Поэтому грубо будем считать их «независимыми». В таком случае дисперсия s^2 выразится приближенно очень просто через сумму дисперсий обоих факторов

$$s_Q^2 = \left(\frac{\partial Q}{\partial H}\right)^2 s_H^2 + \left(\frac{\partial Q}{\partial T^{-1}}\right)^2 s_{T^{-1}}^2 = C^2 \left\{ \frac{s_H^2}{T^2} + \frac{s_{T^{-1}}^2 H^2}{T^2} \right\}. \quad (132)$$

Предложенный автором метод еще не нашел оперативного применения при составлении гидрологических прогнозов главным образом потому, что входящий в него прогноз температуры воздуха в период снеготаяния еще недостаточно хорошо разработан.

Не будучи в состоянии поэтому привести конкретный пример, иллюстрирующий применение предложенного метода, автор ограничивается расчетом, в котором цифры подобраны так, чтобы они были правдоподобны.

Представим себе, что на некоторой равнинной реке сравнительно небольшого бассейна (в несколько тысяч квадратных километров) производились в течение нескольких лет измерения: 1) для построения гидрографа весеннего половодья, 2) средней

по бассейну высоты снега с пересчетом на слой воды и 3) температуры воздуха за период снеготаяния.

По этим измерениям можно легко установить (для данного бассейна) среднее количество стаявшего снега на один градус положительной температуры и, кроме того, получить величину коэффициента C в формуле (131) по средним значениям за годы наблюдений.

Предположим далее, что весной такого-то года незадолго до снеготаяния была получена по многим измерениям средняя для всего бассейна высота H с ошибкой, которую легко можно вычислить. Предположим далее, что ход температуры в период перехода ее через нуль мог быть прогнозирован по какому-то методу с вероятной ошибкой, оцениваемой по теоретическим соображениям. По этим данным нетрудно вычислить и вероятный срок стаивания измеренного снежного покрова, после чего вероятное значение Q_{\max} определяется из вышеприведенного уравнения, причем оценка прогнозируемой величины производится аналогично примерам измерения «малых» величин.

Приведем для конкретности правдоподобные цифры. В измеренном снежном покрове содержится 240 мм слоя воды. Из измерений, произведенных ранее, известно, что каждому градусу положительной температуры в течение суток соответствует в данных условиях 5 мм стаявшего снега, а прогноз хода температуры показывает, что та сумма положительных температур, которая необходима для стаивания всего снеготаяния, может быть рассчитана следующим образом:

$$\frac{240}{5} = 48^\circ \text{ тепла,}$$

Такое количество тепла может быть получено за 8 суток, следовательно, величина T^{-1} равна $\frac{1}{8}$. Значение C по ранее произведенным измерениям (считая Q в м³/сек., H в мм, а T в сутках) получилось равным 54. Следовательно, прогнозируемый максимальный расход составит около 1600 м³/сек. Считая далее вероятную ошибку в измерении H равной ± 20 мм, а вероятную ошибку в прогнозе числа суток, в течение которых будет получена необходимая сумма положительных температур, равной ± 3 суток, можно произвести оценку вероятной ошибки самого расхода. Из уравнения (132)

$$s_Q = \frac{C}{T} \sqrt{s_H^2 + \frac{s_T^2 H^2}{T^2}} = 620 \text{ м}^3/\text{сек.}$$

Такая на первый взгляд большая ошибка, равная 37%, обусловлена главным образом принятой нами значительной ошибкой прогноза температуры и связанной с ней продолжительности снеготаяния. Можно было бы игнорировать ошибку в измерении снежного покрова, однако от этого суммарная ошибка

прогноза очень мало уменьшилась бы. Но если можно точность прогноза температуры значительно повысить и довести расчетную ошибку в определении продолжительности снеготаяния до одних суток, то вероятная суммарная ошибка прогноза расхода снизилась бы до $280 \text{ м}^3/\text{сек.}$, т. е. до 17%.

Само собой разумеется, что во всем приведенном методическом построении нами полностью игнорирована ошибка самой теории, которая, конечно, всегда имеет место. Однако ее размер нужно определить частично теоретическим анализом и главным образом проверкой теории на практике. Этот вопрос мы здесь рассматривать не будем.

Третий пример оценки прогноза приведем из области измерения глубин на перекатах. Еще в 90-х годах прошлого столетия была обнаружена В. А. Макаровым коррелятивная связь между глубиной на перекате и уровнем воды на ближайшем водомерном посту, заключающаяся в том, что при подъеме уровня воды перекат частично заносится наносами, а при понижении уровня размывается. Детальному исследованию это явление в то время подвергнуто не было. Лишь в 20-х годах настоящего столетия Государственным гидрологическим институтом с этой целью на Кумовском перекате р. Дона проведены регулярные наблюдения за глубинами h и уровнями воды H на Калачевской гидрометрической станции.

За период с 1928 по 1934 г. было проведено около 120 измерений течений на глубинах h при соответствующих уровнях H . Выяснилось, что глубина h_0 в переходный момент от намыва к размыву за 7-летний период наблюдений оставалась почти постоянной. Связь между h и H получилась в виде ломаной линии, состоящей из двух прямых с различными для разных лет положениями переходной точки, что, конечно, для разработки методики прогноза было весьма неудобно. Построенная Г. И. Шамовым по тем же данным зависимость между глубиной и скоростью на перекате оказалась более отчетливой. Указанную зависимость Шамов выразил формулой

$$h = h_0 \left(\frac{u}{u_0} \right)^{3/2},$$

в которой u и u_0 вычисляются по уравнению Шези для расходов, соответствующих уровням H . Как видно, эта зависимость дает связь между h и H (при $h_0 = \text{const}$), а поэтому прогноз глубин h сводится к прогнозу уровней H . Не касаясь вопроса о методах прогноза уровня, ограничимся указанием, что если метод, используемый на практике, дает доверительный интервал для значений H , то из вышеприведенного уравнения легко получается соответственный доверительный интервал и для h .

СВЯЗЬ МЕЖДУ ИЗМЕРЕННЫМИ ВЕЛИЧИНАМИ

§ 33. Двумерные и многомерные совокупности случайных величин

В изложении вопросов измерения отдельных величин не было надобности вводить в рассмотрение наличие или отсутствие связей между ними. Но, вообще говоря, эти измерения часто производятся с целью обнаружения связи и придания ей формы, наилучшим образом отвечающей как действительной связи измеренных величин (если, конечно, она существует), так и точности их измерения.

Каждая из измеренных величин представлена каким-то рядом случайных величин, выражающих отдельные измерения. В общем виде можно представить совокупность из m измеряемых величин как случайный m -мерный вектор, а каждое отдельное из n измерений m величин как одно из его значений.

Вполне естественно будет, если начнем с рассмотрения простейшего случая, т. е. двумерного вектора, идущего из начала координат и изображающего при заданных значениях его компонент — точку на плоскости, а при заданной совокупности значений — совокупность точек, отвечающих различным измерениям.

При двумерной дискретной совокупности нас может интересовать вопрос об условных вероятностях вида $P(A|B)$ или $P(B|A)$. Для многомерных совокупностей надо вводить соответственно условные вероятности вида $P(A|B, C, D, \dots)$. Связь между компонентами многомерной совокупности отразится и на плотности вероятностей ее компонентов.

Если взять для простоты двумерную совокупность с компонентами X, Y , то вероятность того, что какое-то двухкомпонентное ее значение расположится влево и вниз от заданной величины x, y выразится двойным неравенством

$$P\{X < x; Y < y\}$$

или соответственно равенством

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y \varphi(x, y) dx dy, \quad (133)$$

из которого хорошо видно, что функция $\varphi(x, y)$ выражает собой вероятность нахождения точки (x, y) внутри элементарного прямоугольника dx, dy , отнесенную к единице его площади, иначе говоря, двумерную плотность вероятности.

Из теоремы умножения вероятностей следует, что в случае, когда компоненты x и y независимы, двумерная плотность вероятности тождественно равна произведению одномерных плотностей:

$$\varphi(x, y) = p_1(x) \cdot p_2(y).$$

В общем случае, как следствие основных положений теории вероятностей, имеем следующее равенство:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x, y) dx dy = 1. \quad (134)$$

Рассмотрим вопрос об основных характеристиках случайных величин — математическом ожидании и дисперсии — применительно к сумме компонентов многомерных совокупностей. Такая сумма $X_1 + X_2 + \dots + X_j + \dots + X_m$ состоит из слагаемых, каждое из которых представлено измерениями $x_{j,1}, x_{j,2}, \dots, x_{j,k}, \dots, x_{j,n}$.

Совершенно ясно, что математическое ожидание, являясь линейной операцией, даст один и тот же результат как при зависимых, так и при независимых компонентах. В обоих случаях имеем одно и то же равенство:

$$M. O. \sum_1^m X_j = \sum_1^m M. O. X_j.$$

Но иначе дело обстоит с дисперсией. Имеем

$$D \sum X_j = M. O. \{ \sum X_j - \sum \bar{X}_j \}^2,$$

если раскроем правую часть и осредним, то получим

$$D \sum X_j = \sum (X_j - \bar{X}_j)^2 + 2 \sum \overline{(X_j - \bar{X}_j)(X_i - \bar{X}_i)}.$$

Индексы j и i относятся здесь к суммируемым случайным величинам, число которых равно m . Дальше, суммируя по отдельным измерениям, число которых равно n , получим

$$\bar{X}_j = \frac{1}{n} \sum x_{k,l}.$$

Введем сокращенные обозначения

$$\sigma_j^2 = \frac{\sum_1^n (X_{j,l} - \bar{X}_j)^2}{n};$$

$$\mu_{j,k} = \frac{1}{n} \sum (X_{j,l} - \bar{X}_j) (X_{k,l} - \bar{X}_k).$$

Тогда предыдущее равенство примет вид

$$\sigma_{\Sigma}^2 X_j = \sum \sigma_j^2 + 2 \sum \mu_{j,k}. \quad (135)$$

В частном случае, когда все X_j независимы, второе слагаемое отпадает; если же только часть их зависима, то сумму вторых слагаемых надо взять лишь для этих зависимых случайных величин.

Вводимые нами члены вида $\mu_{j,k}$ носят название ковариантов. Они положительны, если рост величины X_j сопутствует росту величины X_k , и отрицательны, если росту X_j соответствует убывание X_k , и, наконец, равны нулю при полном отсутствии связи между X_j и X_k , т. е. когда частость одинаковых и различных знаков разностей одна и та же. Вернее сказать, они тем ближе к нулю, чем при большем числе n это условие ближе к истине. В дальнейшем нам придется иметь дело главным образом с двумя зависимыми величинами. В этом случае равенство (135) принимает вид

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2\mu_{x,y}. \quad (136)$$

Предположим теперь, что переменные x и y нормированы, т. е. получены из других переменных ξ и η путем преобразований:

$$x = \frac{\xi - \bar{\xi}}{\sigma_{\xi}}; \quad y = \frac{\eta - \bar{\eta}}{\sigma_{\eta}}.$$

Тогда

$$M.O. x = 0; \quad M.O. y = 0;$$

$$D(x) = 1; \quad D(y) = 1,$$

откуда непосредственно

$$\sigma_{x+y}^2 = 2 \left\{ 1 + \frac{(\xi - \bar{\xi})(\eta - \bar{\eta})}{\sigma_{\xi} \sigma_{\eta}} \right\}.$$

Из вышеприведенных соотношений между (x, y) и (ξ, η) вытекает тождество

$$\mu_{x,y} = \mu_{\xi,\eta} = \frac{(\xi - \bar{\xi})(\eta - \bar{\eta})}{\sigma_{\xi} \sigma_{\eta}}.$$

Выражения вида правой части равенства носят название коэффициентов корреляции:

$$r_{x,y} = \frac{(x - \bar{x})(y - \bar{y})}{\sigma_x \sigma_y}. \quad (137)$$

Подробный анализ этой весьма важной характеристики приведем после изложения основ теории корреляции.

§ 34. Функциональная и стохастическая связи

Наблюдая в природе (и в лаборатории) временный ход какого-либо явления, мы всегда стараемся выделить из наблюдаемых в нем факторов те, связь которых с предшествовавшей во времени его формой представляется наиболее понятной, а в некоторых случаях — даже очевидной.

Считаем, что каждое явление природы и каждая стадия в процессе его изменения зависят полностью от совокупности тех факторов, которые являются его реальными причинами. Но можно ли всегда быть вполне уверенным, что действительно учтены все эти причинные факторы и что ни один из них не упущен? Очевидно, нет! В особенности это относится к тем явлениям, которые происходят в весьма сложной обстановке многочисленных факторов, влияние каждого из которых сравнительно ничтожно, а число их велико. В этих случаях изучаемая нами связь теряет свою строгую функциональность и наше исследование попадает в область случайностей, в которой из-за неучета этих второстепенных, но многочисленных факторов изучаемая система переходит, как нам кажется, не в определенное для нее состояние, а в одно из возможных для нее состояний. Такая связь носит название стохастической: в ней совокупность учитываемых нами факторов еще недостаточна для точной фиксации изучаемого нами перехода, и можно рассчитать лишь ту область, внутри которой с известным распределением вероятностей возможны многие, а иногда даже бесконечно многие переходы.

Внесем здесь одну оговорку. Не следует думать, что недоучет второстепенных факторов является лишь результатом технического несовершенства современной техники и методики измерения и что с повышением последней постепенно все стохастические связи будут преобразовываться в функциональные. В некоторых областях знания это отчасти и возможно, но в общем виде такое предположение было бы ошибочным. Дело в том, что полный и исчерпывающий учет всех мельчайших и кратковременных факторов (если и представить его себе мысленно возможным в будущем) не имеет зачастую никакого практического смысла. Если взять к примеру турбулентное движение жидкости и задать вопрос: кому и для чего может понадобиться измерять то сочетание гидродинамических сил (давления и вязкости), которое обусловило в данном конкретном случае мгновенный поворот движения частицы жидкости направо, а не налево? Не будет ли разумнее просто рассчитать вероятности всех этих поворотов и оперировать с какими-то осредненными их значениями.

Приведем другой пример. Возможен ли, а если даже возможен, то нужен ли точный расчет траектории приближения циклона, как известно, зависящей от весьма многочисленных факторов: давления, температуры, рельефа местности, растительного покрова и т. д., которые при сравнительно быстром перемещении циклона ни наблюдать, ни, тем более, измерять нет возможности. Если бы эта возможность и была, то извлечь из нее какую бы то ни было практическую пользу, очевидно, нельзя.

Необходимо добавить, что измерение и обработка измерений быстро протекающих процессов встречает чрезвычайные трудности, по крайней мере, в настоящее время, а поэтому разработка методов отыскания наилучшей формы для стохастических связей в целом ряде наук еще долгое время будет актуальной.

Приведем некоторые примеры связей, которые не могут быть не стохастическими.

Наблюдаем два явления — грозу и ливень, часто сопутствующие друг другу. Известно, что иногда выпадает ливень из грозовой тучи, но без электрических разрядов. Бывают и сухие грозы. Связь между этими явлениями несомненная, но отнюдь не функциональная, а лишь стохастическая.

Резкое падение барометра в жаркое летнее время почти всегда предвещает грозу, однако грозовая туча может пройти и вдали от наблюдателя. Большое накопление снега за зиму часто предвещает интенсивное половодье, но зачастую бывает и прерывистое весеннее таяние, когда периоды таяния чередуются с легкими заморозками, и при этом даже мощный снежный покров постепенно исчезает.

Стохастичность несколько иного рода имеет место, например, при вылете альфа-частиц из радиоактивного вещества. Измерения, приведенные в § 15, ясно показывают, что интервалы времени между двумя смежными вылетами распределяются по закону Пуассона, т. е. являются, бесспорно, случайными. Точно так же можно считать, что компоненты скорости в точке турбулентного потока¹ распределяются по закону Гаусса, т. е. опять-таки чисто случайно. В данном случае получается даже какая-то аналогия: если бы эта скорость была постоянной, а ее измеряли бы каким-нибудь прибором, то результаты наших измерений также распределялись бы по закону Гаусса. Аналогичных примеров, когда физическая величина реально колеблется по тому же закону, по которому распределялись бы результаты ее измерения, если бы она была постоянной, можно привести большое количество. Например, высота снежного покрова в разных точках на плоской местности изменяется также по закону Гаусса.

¹ Кроме точек, очень близких к стенкам потока.

В этой главе ограничимся рассмотрением лишь той стохастичности, которая обуславливается недоучетом всех факторов, влияющих на данное явление.

В соответствии с известными нам положениями об условных вероятностях скажем, что стохастическая связь существует тогда, и только тогда, когда имеет место неравенство

$$P(A/B) \neq P(A).$$

При равенстве указанных вероятностей никакой связи между X и Y , очевидно, не существует.

§ 35. Основы теории корреляции

Среди всех возможных видов стохастической связи особый практический интерес представляет тот частный случай, когда одна случайная величина влияет не на весь закон распределения другой величины, а лишь на ее математическое ожидание и дисперсию. Этот частный случай носит название корреляционной связи. Приходится признать, что только он один в математической статистике достаточно хорошо разработан.

Начнем с простейшего случая двумерной случайной величины, компоненты которой X и Y будем вообще считать зависящими друг от друга (хотя эта зависимость иногда может и отсутствовать). И поскольку мы намерены рассматривать корреляционную связь, в которой фигурируют лишь средние значения и дисперсии, то естественно начать с постановки вопроса, как выражается влияние одной величины на среднее и дисперсию другой величины. Ответ на этот вопрос требует введения двух новых понятий: условного математического ожидания и условной дисперсии.

Будем считать, что компоненты двумерной случайной величины X и Y представлены двумя рядами соответственно измеренных значений с достаточно большим числом членов, что дает возможность принимать частоту их значений за их вероятность. Оговариваем, что столь большого числа измерений на практике обычно не бывает, но степень строгости ниже следующего доказательства зависит от того, в какой мере это допущение выполняется.

Итак, имеем два длинных синхронных (или, вообще говоря, соответствующих друг другу) ряда:

$$\begin{aligned} x_1, x_2, \dots, x_j, \dots, x_n, \\ y_1, y_2, \dots, y_j, \dots, y_n, \end{aligned}$$

в которых общность индексов означает одновременность (или соответственность) измерений; вопрос же о связи между ними остается пока не решенным. Но стоит нанести все результаты измерений графически в плоскости x, y , как сразу же обнаружится один из трех возможных случаев, представленных на рис. 18: а) точки поднимаются вверх — прямая корреляции;

б) точки группируются около какой-то горизонтальной прямой — нулевая корреляция и в) точки опускаются вниз вдоль горизонтальной оси — обратная корреляция.

Придерживаясь сделанного допущения о достаточно большом числе произведенных измерений, можем перестроить их ряд, рас-

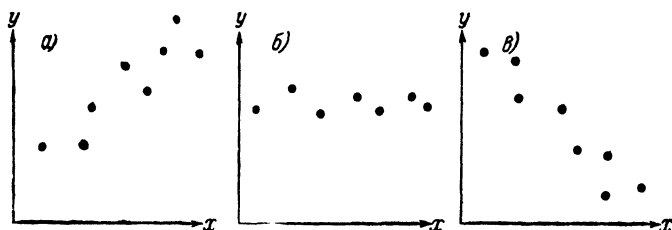


Рис. 18. Схема расположения точек при прямой (а), нулевой (б) и обратной (в) корреляции.

полагая все пары измерений x_j, y_j в возрастающем порядке одной из переменных, например x . После этого разобьем весь ряд на равные интервалы значений той же переменной x (рис. 19) и

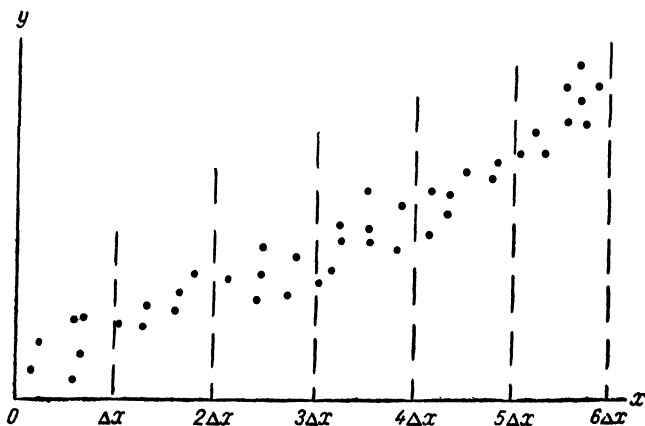


Рис. 19. Схема разбивки точек по интервалам x .

для k -того интервала определим число его значений n_k и их среднее значение \bar{x}_k . Математические ожидания значений y обозначим через \bar{y}_k . Получим снова двойную таблицу, но с другим значением индексов и с числом пар значений, равным теперь не числу измерений, а числу интервалов m :

$$\begin{array}{c} \bar{x}_1, \bar{x}_2, \dots, \bar{x}_k, \dots, \bar{x}_m, \\ \bar{y}_1, \bar{y}_2, \dots, \bar{y}_k, \dots, \bar{y}_m. \end{array}$$

Если переменная X приняла значения k -того интервала \bar{x}_k , то соответствующее ей условное математическое ожидание или условное среднее величины Y будет равно

$$\bar{y}(x_k) = \sum y p(y/x_k).$$

Или, отказавшись от индексов, указывающих номер интервала, запишем в общем виде

$$M.O. \{ Y/x \} = \sum_y y p(y/x). \quad (138)$$

Соответственно вычисляя условное среднее x по y , получим аналогичное выражение:

$$M.O. \{ X/y \} = \sum_x x p(x/y). \quad (139)$$

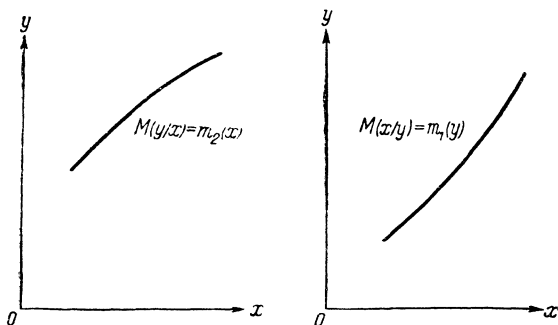


Рис. 20. Схема расположения линии регрессии.

Вместо индексов здесь указан лишь тот аргумент, по которому производится суммирование.

Далее представим, что для каждого из значений x мы нашли по уравнению (138) соответственное условное среднее и провели по полученным точкам (путем их интерполяции) соответственные кривые. Они называются линиями регрессии и схематически изображены на графиках (рис. 20) как для y по x , так и для x по y (обращаем внимание на противоположное направление кривизны на двух графиках, которая и должна быть).

Для дальнейшего изложения нам удобнее представить уравнения линии регрессии в несколько ином виде:

$$\bar{y}(x) = \frac{\sum_y y p(x, y)}{\sum_y p(x, y)}; \quad (140)$$

$$\bar{x}(y) = \frac{\sum_x x p(x, y)}{\sum_x p(x, y)}. \quad (141)$$

Тождество приведенных уравнений с выражением (138) и (139) вытекает из очевидных соотношений:

$$p(x) = \sum_y p(x, y),$$

$$p(y) = \sum_x p(x, y),$$

а также из определения условных вероятностей:

$$p(y|x) = \frac{p(y, x)}{p(x)};$$

$$p(x|y) = \frac{p(x, y)}{p(y)}.$$

Если бы связь между x и y была функциональной, то каждому значению x соответствовало бы одно определенное значение y и линия регрессии обратилась бы в простую кривую функциональной зависимости. Но так как этого нет, то отдельные точки (x_j, y_j) отклоняются от линии регрессии, т. е. имеется какое-то рассеяние, которое может быть оценено аналогично рассеянию при измерении одиночной величины с помощью дисперсии, которая в этом случае носит название условной дисперсии (условной в том смысле, что надо условиться, по отношению к какой из двух переменных она отсчитывается). Выражения для условных дисперсий получаются следующие:

$$\sigma_{y|x}^2 = \sum \{y - \bar{y}(x)\}^2 p(y|x); \quad (142)$$

$$\sigma_{x|y}^2 = \sum \{x - \bar{x}(y)\}^2 p(x|y). \quad (143)$$

Докажем теперь одно простое, но очень важное положение, которое указывает на то, что общая дисперсия величины Y равна сумме двух дисперсий: дисперсии ее условных средних от общего среднего значения и дисперсии отдельных значений от линии регрессии. Доказывается это положение очень просто. Запишем тождество

$$\{\bar{y}(x) - \bar{y}\} + \{y_j - \bar{y}(x)\} = (y_j - \bar{y}),$$

возведем его в квадрат, просуммируем на все n членов первоначального ряда и разделим на n :

$$\frac{\sum (y_j - \bar{y})^2}{n} = \frac{\sum \{\bar{y}(x) - \bar{y}\}^2}{n} + \frac{\sum \{y_j - \bar{y}(x)\}^2}{n} +$$

$$+ \frac{2 \sum \{\bar{y}(x) - \bar{y}\} \{y_j - \bar{y}(x)\}}{n}.$$

Но третье слагаемое правой части при достаточно большом числе должно быть близко к нулю, так как знаки его двух множителей совершенно независимы и при одних и тех же модулях

могут одинаково часто встречаться и равными и противоположными. Следовательно, окончательно

$$\sigma_Y^2 = \sigma_{Y|X}^2 + \sigma_{Y|X}^2. \quad (144)$$

Введем теперь основную характеристику корреляционной зависимости, носящую название корреляционного отношения. Определим это отношение условием, чтобы при полной независимости между X и Y оно равнялось нулю, а при функциональной зависимости между ними — единице. Тогда, введя определяющее равенство

$$\eta_{Y|X}^2 = 1 - \frac{\sigma_{Y|X}^2}{\sigma_Y^2} = \frac{\sigma_{Y(x)}^2}{\sigma_Y^2}, \quad (145)$$

тем самым удовлетворим обоим требованиям. Действительно, если величины X и Y связаны однозначной функциональной зависимостью, то рассеяние около линии регрессии отсутствует: $\sigma_{Y|X}^2 = 0$ и, следовательно, $\eta = 1$. Если же между X и Y нет никакой зависимости, то условная дисперсия равна полной $\sigma_{Y|X}^2 = \sigma_Y^2$ и, следовательно, $\eta = 0$.

Нетрудно убедиться также в справедливости обратных положений. Если $\eta = 1$, то рассеяние около линии регрессии отсутствует, если же $\eta = 0$, то рассеяние равнозначно с общей дисперсией.

Переходим теперь к наиболее интересному частному случаю — линейности тех зависимостей, которые были представлены ранее на общем графике рис. 18. Запишем эту линейную зависимость в виде

$$y(x) = a + bx. \quad (146)$$

Возьмем математическое ожидание от всех трех слагаемых:

$$M.O. y(x) = \bar{y};$$

$$M.O. a = a;$$

$$M.O. x = \bar{x}.$$

Получим равенство

$$\bar{y} = a + b\bar{x}, \quad (147)$$

из которого следует, что прямая регрессии проходит через центр распределения x, y .

Вычитая равенство (147) из равенства (146), будем иметь

$$y(x) - \bar{y} = b(x - \bar{x}). \quad (148)$$

Умножая обе части на $(x - \bar{x})$ и осредняя, получим

$$\overline{(y - \bar{y})(x - \bar{x})} = b\sigma_x^2,$$

откуда значение параметра

$$b = \frac{\mu_{x,y}}{\sigma_x^2}. \quad (149)$$

С другой стороны, нам известно из определения коэффициента корреляции (137) равенство

$$r_{x,y} = \frac{\mu_{x,y}}{\sigma_x \sigma_y}. \quad (150)$$

Следовательно, линейное уравнение регрессии (рис. 20) имеет вид

$$y - \bar{y} = r_{x,y} \frac{\sigma_y}{\sigma_x} (x - \bar{x}). \quad (151)$$

Аналогично могли бы получить для линейной регрессии x по y уравнение

$$x - \bar{x} = \frac{\mu_{x,y}}{\sigma_y^2} (y - \bar{y}).$$

Если бы в этом уравнении заменили

$$\frac{\mu_{x,y}}{\sigma_y^2} = r_{x,y} \frac{\sigma_x}{\sigma_y},$$

то имели бы равенство

$$(x - \bar{x}) = r_{x,y} \frac{\sigma_x}{\sigma_y} (y - \bar{y}). \quad (152)$$

Из сопоставления уравнений (151) и (152) видно, что они отнюдь не тождественны. Действительно, в выводе уравнения (151) мы исходим из условного среднего y , а в выводе уравнения (152) — из условного среднего x , что при функциональной связи было бы равноценным, но при корреляционной связи это отнюдь не одно и то же. Указанное положение доказывается делением одного уравнения на другое, в результате которого получается равенство

$$r_{x,y}^2 = 1,$$

соответствующее функциональной зависимости.

Между корреляционным отношением и коэффициентом корреляции существует определенная связь.

Имеем, согласно равенству (145),

$$\eta_{Y|X}^2 = \frac{\sigma_{y(x)}^2}{\sigma_y^2}; \quad \eta_{X|Y}^2 = \frac{\sigma_{x(y)}^2}{\sigma_x^2}.$$

Числители правых частей, согласно вышеизложенному, равны

$$\sigma_{y(x)}^2 = \sum_x \{ \bar{y}(x) - \bar{y} \}^2 p(x);$$

$$\sigma_{x(y)}^2 = \sum_y \{ \bar{x}(y) - \bar{x} \}^2 p(y).$$

Но в случае линейных зависимостей эти равенства, согласно уравнениям (151) и равенству (152), преобразуются в следующие:

$$\sigma_{y(x)}^2 = \left[r_{x,y} \frac{\sigma_y}{\sigma_x} \right]^2 \sigma_x^2 = r_{x,y}^2 \sigma_y^2,$$

$$\sigma_{x(y)}^2 = \left[r_{x,y} \frac{\sigma_x}{\sigma_y} \right]^2 \sigma_y^2 = r_{x,y}^2 \sigma_x^2,$$

откуда окончательно

$$\eta_{Y|X}^2 = \eta_{X|Y}^2 = r_{x,y}^2. \quad (153)$$

Таким образом, корреляционное отношение, которое в общем случае зависит от выбора независимой переменной, в частном случае линейной зависимости отождествляется с коэффициентом корреляции, который от этого выбора не зависит.

Все вышеизложенные свойства корреляционного отношения, естественно, сохраняют силу и для коэффициента корреляции, как его частного случая. Эти свойства надо сравнивать лишь для того, чтобы отличить линейную зависимость от более общего случая, т. е. от нелинейной зависимости.

Здесь выясняется прежде всего, что в случае линейной зависимости для данной совокупности измеренных значений (x, y) существуют между ними две, и только две, формы зависимости:

$$y = f_1(x);$$

$$x = f_2(y).$$

В случае нелинейной связи таких зависимостей может быть бесконечное множество, так как корреляционное отношение указывает лишь на наличие какой-то связи между компонентами рассматриваемой совокупности, но никаких указаний на характер формы этой связи дать не может.

Следовательно, вопрос о выборе между двумя приведенными характеристиками надо рассматривать в иной плоскости.

В связи с этим остановимся на следующем вопросе, чисто принципиальном и относящемся непосредственно к структуре коэффициента корреляции. Этот коэффициент является дробью, числитель которой представляет собой смешанный второй момент, характеризующий, как известно, тесноту связи между x и y . В знаменателе же имеется произведение стандартов измерения, которое может характеризовать лишь точность (т. е. размер возможной ошибки) обоих измерений.

Возникает вопрос: нельзя ли эти два столь разнородных фактора разделить более отчетливо?

Отнеся отсчет величин x и y к центрам их распределения, получим

$$M.O. x = 0; \quad M.O. y = 0; \quad r_{x,y} = \frac{\overline{xy}}{\sigma_x \sigma_y}.$$

Предположим, что измерения тех же величин искажены добавочными ошибками δ_x и δ_y , в результате которых

$$x_1 = x + \delta_x; \quad y_1 = y + \delta_y.$$

Считая, что эти добавочные ошибки также случайны, а число измерений достаточно велико, получим

$$M.O. \delta x = 0; \quad M.O. \delta y = 0.$$

Таким же образом имеем право принять, что случайные ошибки δ_x и δ_y взаимно независимы и некоррелированы с основными отклонениями от центра (\bar{x} , \bar{y}). Следовательно,

$$M.O. (x \delta x) = 0; \quad M.O. (y \delta y) = 0; \quad M.O. (\delta x \delta x) = 0.$$

Соответственно

$$D(x_1) = \sigma_{x_1}^2 = \sigma_x^2 + \sigma_{\delta x}^2; \quad D(y_1) = \sigma_{y_1}^2 = \sigma_y^2 + \sigma_{\delta y}^2;$$

$$M.O. (x_1 y_1) = M.O. (xy).$$

Найдем теперь значение коэффициента корреляции между x_1 и y_1 :

$$r_{x_1 y_1} = \frac{\overline{x_1 y_1}}{\sigma_{x_1} \sigma_{y_1}} = \frac{\overline{xy}}{\sigma_x \sigma_y} \frac{\sigma_x \sigma_y}{\sigma_{x_1} \sigma_{y_1}} = r_{xy} \frac{\sigma_x \sigma_y}{\sigma_{x_1} \sigma_{y_1}}. \quad (154)$$

Как видно, полученный коэффициент корреляции меньше прежнего на отношение произведения прежних стандартов к произведению новых стандартов.

Приведенное рассуждение (сообщенное нам Н. В. Смирновым) не дает, конечно, полного выделения в коэффициенте корреляции его стохастической части, что, по существу, и невозможно, но все же отчетливо выявляет роль точности измерения в вычислении значения этого коэффициента и попутно доказывает, что значение числителя в коэффициенте корреляции от точности измерений не зависит (вернее, почти не зависит).

Вообще надо сказать, что практическая роль коэффициента корреляции очень велика и его часто используют даже там, где линейность связи явно отсутствует. Дело в том, что коэффициент корреляции как мера связи между величинами отнюдь не представляется ни надежным, ни бесспорным, в особенности при низких его значениях. Но как индикатор наличия какой-то связи он совершенно бесспорен, в особенности при высоких его значениях. Поэтому там, где предварительный графический анализ дает указание на тесноту связи, весьма полезно вычислить коэффициент корреляции с целью подтверждения этого предварительного указания. В случае получения для него значения порядка 0,80—0,90 можно с уверенностью считать, что, независимо от линейности или нелинейности этой связи, она достаточно тесна, чтобы оправдать применение точных методов ее оформления, как, например, способа наименьших квадратов, о котором будет сказано ниже.

Попутно дадим указание на наименее трудоемкий метод вычисления коэффициента корреляции, а именно из определяющего его равенства (143)

$$r = \frac{\overline{(x-\bar{x})(y-\bar{y})}}{\sqrt{\overline{(x-\bar{x})^2}} \sqrt{\overline{(y-\bar{y})^2}}},$$

из которого нетрудно получить следующее:

$$r = \frac{\Sigma xy - n\bar{x}\bar{y}}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}.$$

Число и объем арифметических действий во втором равенстве, очевидно, меньше, чем в первом.

Закончим данный параграф составлением исчерпывающего

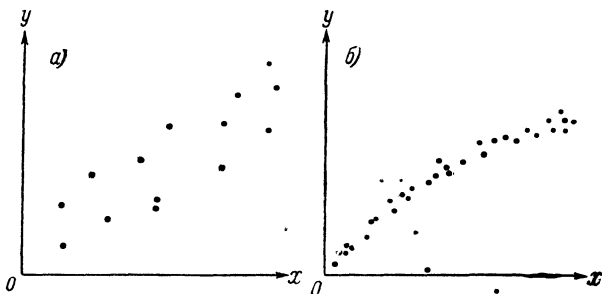


Рис. 21. Схема различия коэффициента корреляции и корреляционного отношения.

а — большой разброс, б — малый разброс.

ответа на вопрос о выборе между коэффициентом корреляции и корреляционным отношением.

Начнем с нанесения на график (рис. 21) системы точек x_j, y_j и посмотрим, имеется ли определенная тенденция к их группировке и каков разброс нанесенных точек.

Совершенно очевидно, что чем разброс точек больше, тем и тенденция к их группировке будет проявляться менее отчетливо. При очень большом разбросе только значительное отклонение от линейности способно проявиться достаточно выпукло, чтобы с ним уже нельзя было не считаться. В противном же случае практически удобнее пренебречь этой слабовыраженной нелинейностью и остановиться на более простом исходе, т. е. на коэффициенте корреляции. Нужно всегда помнить, что ни тот, ни другой индикатор связи не являются, строго говоря, вполне обоснованными, так как оба имеют в известной степени условный характер и лишь при очень высоких значениях (т. е. близких к положительной или отрицательной единице) они действительно отражают тесную связь двух величин. Но при большом разбросе точек столь высокие их значения неосуществимы; и потому

ошибка от игнорирования слабовыраженной нелинейности, вероятно, окажется того же порядка, что и их принципиальная ошибочность в оценке степени тесноты корреляционной связи (рис. 21 а).

Возьмем противоположный случай: малый разброс точек, когда форма линии регрессии представляется более отчетливо. Тогда нег основания игнорировать эту нелинейность (рис. 21 б) и идти на заведомо более грубое спрямление. Но тогда возникает вопрос, каким методом надо руководствоваться, чтобы провести эту кривую не слишком субъективно и так, чтобы она была в каком-то смысле «лучше» всякой другой? Этому сложному вопросу будет посвящена следующая глава. Из нее будет видно, что вполне общего и строгого решения этого вопроса получить вообще невозможно. Но если задаться какой-либо определенной формой кривой (например, параболой третьей степени), то ее параметры можно определить так, чтобы они были для этой формы кривой и для данного исходного материала действительно наилучшими.

§ 36. Нормальная корреляция

В этом параграфе осветим вопрос о связи между корреляцией и нормальным распределением. Для этого выведем выражение, полученное впервые С. Н. Бернштейном, для нормального распределения двумерной случайной величины.

Известно, что при нормальном распределении одномерной случайной величины наиболее вероятным ее значением является ее среднее арифметическое (совпадающее в этом случае и с модой, и с медианой). Естественно принять, что то же условие сохранится и для двумерной случайной величины, т. е. мода $x = \bar{x}$, мода $y = \bar{y}$.

Нам известно, что величины x и y могут рассматриваться как координаты точки на плоскости, а поэтому имеем право считать величины x_j и y_k при разных значениях индексов j и k независимыми. При равных же значениях этих индексов величины x_j и y_j могут находиться в любой зависимости: коррелятивной и в частном случае даже функциональной.

По изложенному выше наибольшей вероятностью (при очень большом числе точек) обладает точка с координатами (\bar{x}, \bar{y}) , иначе говоря, центр тяжести всех точек. И поскольку вероятность остальных точек должна, согласно основному нормальному закону, непрерывно убывать от этого максимума, то ограничим искомую двумерную нормальную функцию по аналогии с одномерной степенным видом

$$F(x, y) = \text{const } e^{\varphi(x, y)}.$$

Далее, пользуясь также методом аналогии, поставим условие,

чтобы функция $\varphi(x, y)$ была какой-то квадратической формой от разностей $(x - \bar{x})$ и $(y - \bar{y})$. Причем для сокращения записи выберем начало отсчета в центре тяжести всех точек. Тогда, сохраняя и при новом начале отсчета те же обозначения x, y , получим (доказательства давать не будем) следующее равенство, выражающее закон нормального распределения двумерной случайной величины:

$$F(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp \left\{ \frac{1}{2(1-r^2)} \left[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - 2r \frac{xy}{\sigma_x\sigma_y} \right] \right\}. \quad (155)$$

Покажем, что функция $F(x, y)$ приводит к следующим выводам:

1. При $r=0$, т. е. при взаимной независимости переменных (x, y) , уравнение (155) вырождается в произведение плотностей вероятностей двух независимых, нормально распределенных величин.

Действительно, при $r=0$

$$F(x, y) = \left[\frac{1}{\sqrt{2\pi}\sigma_x} \exp \left(-\frac{x^2}{2\sigma_x^2} \right) \right] \left[\frac{1}{\sqrt{2\pi}\sigma_y} \exp \left(-\frac{y^2}{2\sigma_y^2} \right) \right].$$

2. Интегрирование функции $F(x, y)$ по одной из переменных дает плотность одномерного распределения другой переменной. Для доказательства сделаем подстановку:

$$u = \frac{x}{\sigma_x}; \quad v = \frac{y}{\sigma_y},$$

в результате получим

$$f(x) = \frac{1}{2\pi\sigma_x\sqrt{1-r^2}} \int_{-\infty}^{+\infty} \exp \left\{ -\frac{1}{2(1-r^2)} (u^2 + v^2 - 2ruv) \right\} dv.$$

Добавим в показателе сумму, тождественно равную нулю:

$$-\frac{u^2}{2(1-r^2)} + \frac{u^2 r^2}{2(1-r^2)} - \frac{u^2 r^2}{2(1-r^2)} - \frac{v^2}{2(1-r^2)} + \\ + \frac{2ruv}{2(1-r^2)},$$

после чего получим

$$f(x) = \frac{e^{-\frac{u^2}{2}}}{2\pi\sigma_x\sqrt{1-r^2}} \int_{-\infty}^{+\infty} e^{-\frac{(v-ru)^2}{2(1-r^2)}} dv.$$

Если теперь произведем еще одну подстановку

$$\frac{v - ru}{\sqrt{1 - r^2}} = z,$$

то, учитывая, что ru от v не зависит, будем иметь

$$dv = \sqrt{1 - r^2} dz,$$

а так как интеграл

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz,$$

тождественно равен единице, то предыдущее выражение приводится к следующему виду:

$$f(x) = \frac{e^{-\frac{x^2}{2\sigma_x^2}}}{\sqrt{2\pi}\sigma_x} \quad (156)$$

Иными словами, интегрирование по y приводит к плотности нормального распределения x . Симметричность функции $F(x, y)$ делает излишним доказательство того же положения и для переменной y .

Постараемся пояснить смысл полученных нами выводов.

Первый из них равнозначен утверждению, что при нулевом значении коэффициента корреляции между двумя случайными величинами они тем самым являются независимыми. Необходимо подчеркнуть, что такой вывод имеет силу лишь при нормальной корреляции. В общем же виде при любой зависимости между случайными величинами некоррелированность отнюдь не тождественна с независимостью. Можно представить себе такую связь между двумя случайными величинами, когда одна из них влияет, например, лишь на стандарт другой, оставляя без изменения ее среднее значение. Тогда при $r=0$ эти величины все же будут зависимыми.

Второй из двух приведенных выводов должен быть истолкован в том смысле, что каково бы ни было значение коэффициента корреляции между двумя нормально коррелированными величинами, каждая из них подчиняется нормальному закону независимо от другой.

Оба указанных свойства нормальной корреляции тесно связаны с тем фактом, что регрессия между нормально коррелированными величинами может быть только линейной. Доказательство этого в важного положения приведем, заранее оговаривая, что обратное положение отнюдь необязательно, так как линейная регрессия возможна и при отсутствии между величинами нормальной корреляции.

Если дадим выражение для условного нормального распреде-

ления, причем в случае линейной регрессии, которая в наших те- перешних обозначениях для централизованного отсчета x и y запи- шется в виде

$$y = r \frac{\sigma_y}{\sigma_x} x,$$

то сразу же увидим, что регрессия как y по x , так и x по y при нормальной их корреляции будет всегда линейной.

Видно также, что условная дисперсия y по x не зависит от x и соответственно дисперсия x по y не зависит от y . Иными словами, среднеквадратическое отклонение одной переменной при всех значениях другой одинаково и равно соответственно

$$\sigma_{y/x} = \sigma_y \sqrt{1 - r^2}; \quad \sigma_{x/y} = \sigma_x \sqrt{1 - r^2}.$$

Исходя из общего определения условного распределения

$$\varphi(x, y) = \varphi(y | x) \varphi(x),$$

запишем, в частности, для условного нормального распределения очевидное отношение (пользуясь также обозначениями u и v)

$$\begin{aligned} \varphi(y | x) &= \frac{\exp \left[-\frac{1}{2(1-r^2)} (u^2 + v^2 - 2ruv) \right]}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} : \frac{\exp -\frac{u^2}{2}}{\sqrt{2\pi}\sigma_x} = \\ &= \frac{\exp \left[-\frac{y - r \frac{\sigma_x}{\sigma_y} x}{2\sigma_y(1-r^2)} \right]^2}{\sqrt{2\pi}\sigma_y\sqrt{1-r^2}}. \end{aligned} \quad (157)$$

Мы действительно получили плотность условного распреде- ления случайной величины Y при заданном значении $X=x$. Оно является нормальным, и его условная дисперсия, как видно из уравнения (157), равна

$$D(Y | X) = \sigma_y^2 (1 - r^2).$$

Непосредственно видно также, что при независимости X и Y , т. е. при $r=0$, уравнение (157) дает хорошо нам известное нор- мальное распределение для y .

§ 37. Множественная корреляция

На практике встречаются случаи, когда надо выяснить, суще- ствует ли какая-либо зависимость между тремя (и более) факто- рами, доступными измерению. Метод корреляции позволяет дать ответ и на такой вопрос. Иными словами, имея ряд одновремен- ных (или соответственных) измерений трех (и более) факторов, возможно констатировать наличие (или отсутствие) влияния одного из них на остальные.

Ввиду сложности вопроса ограничимся здесь его простейшим

случае — линейной нормальной корреляцией между тремя нормально распределенными случайными величинами, полученными путем измерения, т. е. содержащими случайные ошибки.

Без нарушения общности можно также принять, что три величины: x, y, z , заданные рядами соответственных измерений x_k, y_k, z_k ; $k=1, 2, 3, \dots, n$, пересчитаны к их центрам, в результате чего

$$\bar{x} = \bar{y} = \bar{z} = 0.$$

Следовательно, по аналогии с выведенным в предыдущем параграфе распределением двух переменных можем записать $F(x, y, z) = K \exp - [ax^2 + by^2 + cz^2 + 2fyz + 2gxz + 2hxy]$.

Постоянные K, a, b, c, f, g, h подлежат определению. В показателе будет однородная квадратическая форма, которую обозначим через φ . Ее определитель равен

$$\Delta = abc + 2fgh - af^2 - bg^2 - ch^2. \quad (158)$$

Из анализа известно, что

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\varphi} dx dy dz = \frac{\pi^{\frac{3}{2}}}{\Delta^{\frac{1}{2}}}.$$

Но так как по общим свойствам всех кривых распределения тот же интеграл, умноженный на k , должен быть равен единице, то

$$k = \frac{\sqrt{\Delta}}{\pi^{\frac{3}{2}}}.$$

Дальше доказательство пойдет следующим путем. Будем поочередно игнорировать изменение одной из переменных: то x , то y , то z . Каждый раз распределение двух других переменных должно, очевидно, удовлетворять выведенному в предыдущем параграфе уравнению (155). После трех таких исключений, путем сравнения неопределенных коэффициентов тройной корреляции a, b, c, f, g, h с ранее найденными коэффициентами двойной корреляции можем легко найти первые.

Введем определитель, составленный из коэффициентов корреляции, которые теперь обозначим для удобства индексами $r_{1,2}, r_{2,3}, r_{3,1}$ (коэффициенты с равными индексами, очевидно, равны единицам),

$$R = \begin{vmatrix} 1 & r_{1,2} & r_{1,3} \\ r_{2,1} & 1 & r_{2,3} \\ r_{3,1} & r_{3,2} & 1 \end{vmatrix} \quad (159)$$

Это есть симметричный определитель третьего порядка. Он равен

$$R = 1 + 2r_{1,2} r_{2,3} r_{1,3} - r_{1,2}^2 - r_{2,3}^2 - r_{3,1}^2.$$

Составим к нему миноры¹:

$$\begin{aligned} R_{1,1} &= 1 - r_{2,3}^2; & R_{2,2} &= 1 - r_{1,3}^2, & R_{3,3} &= 1 - r_{1,2}^2; \\ R_{1,2} &= r_{1,2} - r_{2,3} r_{1,3}; & R_{2,3} &= r_{2,3} - r_{1,3} r_{1,2}; \\ R_{1,3} &= r_{1,3} - r_{2,3} r_{2,1}. \end{aligned}$$

Теперь, опуская все простые, но громоздкие выкладки, запишем непосредственно окончательный результат

$$F(x, y, z) = \frac{1}{(2\pi)^{3/2} \sigma_x \sigma_y \sigma_z R^{1/2}} \exp \left\{ \frac{R_{1,1} x^2}{\sigma_x^2} + \frac{R_{2,2} y^2}{\sigma_y^2} + \frac{R_{3,3} z^2}{\sigma_z^2} + \frac{2R_{2,3} yz}{\sigma_y \sigma_z} + \frac{2R_{3,1} zx}{\sigma_z \sigma_x} + \frac{2R_{1,2} xy}{\sigma_x \sigma_y} \right\}. \quad (160)$$

Читатель легко убедится, что если в этом равенстве приравняем нулю все слагаемые с индексом 3, а также содержащие множитель z и понизим показатель при 2π : вместо $\frac{3}{2}$ поставим 1, то получим в точности уравнение (155).

По тому же принципу можно построить закон нормального распределения для четырех (и более) связанных величин, если распределение каждой из них нормальное. Но практически такое расширение не может понадобиться, кроме исключительно редких случаев. Но тогда решение, хотя очень громоздкое и трудоемкое, может быть получено аналогичным путем.

§ 38. Среднее и дисперсия функции измеренных величин

Далее рассмотрим величину, функционально зависящую от нескольких измеренных величин. Поскольку последние представлены рядами значений и известны нам с некоторыми ошибками, то и величина, функционально от них зависящая, также должна отражать в себе их неточность и потому также должна быть представлена каким-то рядом значений (а не одной определенной величиной), который будет иметь свое среднее и свою дисперсию. Ответим на вопрос, как найти среднее и дисперсию функции, зная средние и дисперсии тех величин, от которых она зависит. Причем очевидно, что в случае если исходные величины зависят друг от друга, то нам необходимо знать и ковариации между ними.

Возьмем функцию от трех переменных

$$u = F(x, y, z). \quad (161)$$

Ограничимся случаем, когда зависимость имеет гладкий

¹ Напоминаем, что минорами называются определители низшего (в данном случае второго) порядка, получаемые для любого элемента основного определителя.

характер. Это значит, что в некоторых небольших интервалах значений переменных x, y, z вторыми производными $\frac{\partial^2 u}{\partial x^2}$ и т. д. можно пренебречь. Примем, кроме того, что разброс точек, полученных при измерении этих переменных, сосредоточился в основном в сравнительно небольшом диапазоне их значений и что размер этого диапазона не превышает того, внутри которого приняли функцию достаточно «гладкой». Тогда после разложения функции (161) в ряд Тэйлора имеем право, отыскивая лишь приближенные значения среднего и дисперсии, отбросить все нелинейные члены. Получим

$$u = F_0(x, y, z) + (x - \bar{x}) \left(\frac{\partial F}{\partial x} \right)_0 + (y - \bar{y}) \left(\frac{\partial F}{\partial y} \right)_0 + (z - \bar{z}) \left(\frac{\partial F}{\partial z} \right)_0. \quad (162)$$

Нулевой индекс при производных означает, что производные берутся при средних значениях x, y, z .

Если теперь осредним обе части равенства, то ввиду постоянства производных и очевидности равенств:

$$M.O. (x - \bar{x}) = 0; \quad M.O. (y - \bar{y}) = 0; \quad M.O. (z - \bar{z}) = 0$$

сразу получаем приближенное выражение среднего значения $u(x, y, z)$ в виде

$$\bar{u} \approx F(\bar{x}, \bar{y}, \bar{z}).$$

Для получения приближенного значения дисперсии запишем определяющее дисперсию равенство

$$D(u) \approx M.O. (u - \bar{u})^2 = M.O. \left\{ \left(\frac{\partial F}{\partial x} \right)_0 (x - \bar{x}) + \left(\frac{\partial F}{\partial y} \right)_0 (y - \bar{y}) + \left(\frac{\partial F}{\partial z} \right)_0 (z - \bar{z}) \right\}^2 = \left(\frac{\partial F}{\partial x} \right)_0^2 \sigma_x^2 + \left(\frac{\partial F}{\partial y} \right)_0^2 \sigma_y^2 + \left(\frac{\partial F}{\partial z} \right)_0^2 \sigma_z^2 + 2 \left(\frac{\partial F}{\partial x} \right)_0 \left(\frac{\partial F}{\partial y} \right)_0 \mu_{xy} + 2 \left(\frac{\partial F}{\partial x} \right)_0 \left(\frac{\partial F}{\partial z} \right)_0 \mu_{xz} + 2 \left(\frac{\partial F}{\partial y} \right)_0 \left(\frac{\partial F}{\partial z} \right)_0 \mu_{yz}. \quad (163)$$

В частном случае независимости переменных x, y, z три последних слагаемых отпадают, и мы получаем простое равенство

$$D(u) = \sum \left(\frac{\partial u}{\partial x_j} \right)_0^2 \sigma_{x_j}^2, \quad (164)$$

где индекс 0 при квадратах производных показывает, что значения этих производных взяты вблизи среднего значения функции u .

Таким образом, приближенные уравнения (163) и (164) будут

тем точнее, чем меньше изменчивость этих производных вблизи среднего значения функции u , иными словами, чем эта функция имеет более «гладкий» характер, или чем ближе к нулю ее вторые производные.

Отсюда непосредственно следует, что при линейности функции u , когда все ее производные становятся постоянными, указанные зависимости становятся точными.

Но можно указать случаи, когда и при нелинейности функции $u(x, y, z)$ и при взаимной независимости переменных и среднее, и дисперсия этой функции могут быть выражены точно. Например, для произведения независимых величин имеем, очевидно, для среднего

$$\bar{u} = \bar{x}\bar{y}\bar{z}.$$

а ее дисперсия получается после логарифмирования равной

$$\frac{\sigma_u^2}{\bar{u}^2} = \frac{\sigma_x^2}{\bar{x}^2} + \frac{\sigma_y^2}{\bar{y}^2} + \frac{\sigma_z^2}{\bar{z}^2}.$$

ОТЫСКАНИЕ ФОРМЫ СВЯЗИ МЕЖДУ ДВУМЯ ИЗМЕРЕННЫМИ ВЕЛИЧИНАМИ

§ 39. Постановка задачи

После рассмотрения вопросов, связанных непосредственно с измерением величин, мы в предыдущей главе перешли к вопросу отыскания наличия (или отсутствия) связи между ними. Ответ на этот вопрос дает теория корреляции. Следующий, возникающий затем вопрос: если эта связь между измеренными величинами существует, то какова ее форма? Такой вопрос в общем виде находится за пределами теории корреляции. Только в частном случае, когда эта связь заведомо линейная, можно определить ее параметры с помощью одного лишь корреляционного метода, не прибегая к каким-либо дополнительным суждениям или принципам. Но это возможно потому, что сама линейность является уже той формой, которую на основании каких-то соображений постулируем. Если же точки, расположенные на графике, явно противоречат такому предположению, то для отыскания формы нелинейной зависимости, наиболее вероятной или наиболее близкой к истинной, нам надо найти новые специальные руководящие принципы.

Замечательно то, что для отыскания этих принципов достаточно идти путем расширения тех принципов, которыми мы руководствовались в более простой задаче обработки измерения изолированных величин, т. е. не связанных друг с другом.

Таких принципов, как известно из предыдущего изложения, два. Подчеркиваем, что принципов только два и не может быть больше, так как упразднить влияние знака при тех случайных отклонениях, которые неизбежны при всяком измерении, можно только двумя путями: или брать модули этих отклонений, или возводить последние в квадрат (обращение для этой цели к более высоким степеням, чем квадрат, не имеет никакого смысла).

Переходим к расширенной трактовке этих двух, нам уже известных, принципов.

Принцип Лежандра требует, чтобы сумма квадратов отклонений отдельных измерений от их центра (или от «истинного» размера измеряемой величины) была минимальной.

Принцип Лапласа предъявляет аналогичное требование к сумме абсолютных значений (модулей) этих отклонений.

Нам известно, что центром, удовлетворяющим первому требованию, является среднее арифметическое, а центром, удовлетворяющим второму требованию, — медиана.

Известно также, что в тех случаях, когда все отклонения отдельных измерений располагаются симметрично, среднее арифметическое совпадает с медианой (так же, как и с модой), и вопрос о выборе между двумя принципами переходит в этих случаях в чисто практическую плоскость — что легче (или что точнее) вычислить: среднее или медиану. В более же общем случае, допускающем возможность и асимметричного распределения отклонений, тот же вопрос выбора имеет принципиальное значение.

Переход от измерения одной величины к измерению двух взаимосвязанных величин требует некоторого расширения указанных принципов, так как если прежде речь шла об отклонениях проведенных измерений от точки, изображающей или центр рассеяния измерений, или истинное значение измеряемой величины, то теперь в расширенном использовании тех же принципов никакого центра рассеяния нет, а есть какая-то кривая (в частном случае прямая), предположительно принятая нами. Эту кривую надо расположить таким образом (т. е. дать ей такие параметры), чтобы сумма квадратов (или сумма модулей) отклонений этой предположительной кривой от заданных значений, полученных измерениями, была наименьшей.

В общем случае имеем измеренную величину u , зависящую от нескольких величин, также измеримых. Форма этой зависимости иногда бывает задана теорией вопроса. В этом случае нам остается лишь на основе измерения подыскать значения параметров этой теоретической зависимости (если сама теория этих параметров не дает). Бывает другой случай, более интересующий нас, когда форму зависимости должны отыскать на основе проведенных измерений. Сейчас покажем, что вопрос о форме искомой зависимости можно при допустимых ограничениях также свести к вопросу числа тех параметров, которые надо определить.

Действительно, стоит нам лишь представить искомую зависимость между двумя измеримыми величинами расположенной в ряд Тэйлора, который, не нарушая общности, имеем право заменить степенным рядом

$$y = a_0 + a_1x + a_2x^2 + \dots + a_mx^m,$$

и сразу станет ясно, что число слагаемых будет определять также форму той зависимости, на которой мы остановимся.

В этом уравнении число параметров на единицу больше порядка параболы. Эти параметры должны быть в искомой зависимости постоянными, переменными же будут неизвестные x , y . Но в процессе нахождения искомой зависимости имеет место обратное положение: вместо неизвестных x и y будем иметь n пар их измерений, следовательно, это известные величины. Значения же параметров a_0 , a_1 и т. д. мы должны еще искать, следовательно, они будут пока еще неизвестными. Если подставить в это уравнение измеренные значения x_j , y_j , то можно получить систему из n уравнений (по числу проведенных измерений) вида

$$y_k = \sum_{i=0}^{i=m} a_i x_{k,i}; \quad k = 1, 2, 3, \dots, n,$$

неизвестными в которых будут m параметров. Эти уравнения носят название условных. Так как их число обычно больше (даже гораздо больше!) числа неизвестных параметров, т. е. $m \ll n$, поэтому имеем здесь систему с избыточным числом уравнений. Решается эта система с помощью применения к ней одного из двух указанных принципов.

Принцип Лежандра выразится тогда равенством

$$\sum_{k=0}^{k=n} \{ a_0 + a_1 x_k + a_2 x_k^2 + \dots - y_k \}^2 = \min, \quad (165)$$

а принцип Лапласа равенством

$$\sum_{k=0}^{k=n} | a_0 + a_1 x_k + a_2 x_k^2 + \dots - y_k | = \min. \quad (166)$$

В обоих случаях раскрытие условия минимальности проводится путем дифференцирования по неизвестным параметрам с приравниванием полученных производных нулю. В результате получится новая система уравнений, число которых равно, очевидно, числу параметров. Эти уравнения носят название нормальных.

Под эту общую схему можно подвести также простейший частный случай, когда $m=0$ и имеется лишь один искомый параметр a_0 , который будет равен среднему значению, согласно принципу Лежандра, и медиане, согласно принципу Лапласа.

Следующим расширением этой общей схемы будет отыскание двух первых параметров уравнения, иными словами, нахождение линейной зависимости y от x .

Вторая задача также решена нами методами линейной корреляции. Но теперь рассмотрим ее с другой стороны. Идя далее, мы можем трактовать задачу отыскания параболы второго порядка как определение трех параметров: a_0 , a_1 , a_2 ,

а отыскания параболы третьей степени как определение четырех параметров: a_0, a_1, a_2, a_3 и т. д.

Рассмотрим последовательность этих решений. Одно число обозначает при заданном начале отсчета точку на числовой прямой. Два числа можно трактовать двояко: или как координаты точки на плоскости, или как параметры, определяющие положение прямой линии в той же плоскости.

Три числа могут трактоваться или как координаты точки в пространстве, или как параметры, определяющие положение плоскости, и т. д.

Обобщая предыдущее рассуждение, придем к тому, что отыскание на основании n пар измерений параметров параболы m -го порядка, выражающей связь между этими двумя величинами, формально тождественно с нахождением координат точки в m -мерном пространстве.

Нами приведена эта чисто геометрическая абстрактная схема, конечно, не для практических целей расчета, для которых она слишком громоздка, а лишь в виде иллюстрации к той проводимой нами аналогии между процессом измерения и процессом отыскания зависимостей между измеренными величинами. Практически эта схема может быть полезна лишь в тех случаях, когда геометрическое построение технически осуществимо, а именно на плоскости и отчасти в трехмерном пространстве, где можно применить методы начертательной геометрии. Дальше идти этим путем, очевидно, уже нельзя. В дальнейшем нами будут показаны другие пути для решения изложенных выше задач, но уже без рассмотрения «многомерных пространств». Однако думаем, что приведенное (хотя и абстрактное) построение не будет бесполезно для уяснения сущности вопроса.

Переходя к практической стороне, нужно прежде всего отметить, что принцип Лежандра, детально разработанный главным образом Гауссом, получил в литературе гораздо большее развитие, чем принцип Лапласа. Можно отметить, что даже сам Лаплас, подчеркивая идейное преимущество медианы перед средним арифметическим, констатировал наличие известных формальных трудностей при использовании медианы в случае измерения двух и более величин.

Поэтому в дальнейшем будем вынуждены гораздо подробнее излагать метод Лежандра—Гаусса. Что же касается принципа Лапласа, то здесь ограничимся изложением графо-аналитического метода, предложенного английским статистиком Эджуорсом и малоизвестного в литературе. Опубликованный в конце XIX в. [19], этот метод, по-видимому, не заслужил большого внимания у практиков, занимающихся обработкой измерений. Однако, по нашему мнению, метод представляет несомненный интерес хотя бы при отыскании одних лишь линейных зависимостей.

Такие крупнейшие математики, работавшие в этой области,

как Лаплас, Бертран и Пуанкаре, всегда подчеркивали условность «постулата среднего арифметического», указывая, что «постулат медианы» имеет не меньше теоретических прав на практическое использование. Соответственно этому связанный с первым постулатом принцип наименьшей суммы квадратов не имеет никаких теоретических преимуществ перед принципом наименьшей суммы модулей, связанным с медианой. Поэтому их сравнение должно быть рассмотрено исключительно с практической точки зрения, т. е. нужно выяснить, когда трудоемкость вычислений меньше и когда точность результатов больше.

Здесь приходится подчеркнуть, что для простейшего случая измерения одной величины найти медиану гораздо легче, чем среднее арифметическое. Что же касается вероятной ошибки, то, по мнению ряда авторов, она имеет в обоих случаях одинаковый порядок, хотя этот вопрос еще недостаточно освещен в литературе. Следовательно, есть основания отдавать предпочтение медиане даже при симметричном расположении отклонений, когда теоретически она должна совпасть со средним арифметическим. Что же касается расположений явно асимметричных, то здесь медиана имеет не только практические, но и теоретические преимущества. Эта сторона вопроса нами подробно освещена в связи с измерениями «малых» величин.

В заключении данного параграфа рассмотрим весьма интересную попытку выйти за пределы способа наименьших квадратов, приобретенного со времени Гаусса устойчивую репутацию в литературе по отысканию формы коррелятивных связей между случайными величинами. Попытка, сделанная акад. Б. И. Срезневским в 1889 г., состоит в своеобразном использовании квантилей, о которых кратко изложено в § 29. Позже, в 1914 г., Срезневский назвал свой метод «методом равных повторяемостей», который в последнее время получил развитие и оформление в работах Г. А. Алексеева. Его специальная статья, посвященная этому методу, впервые публикуется в данной книге. Отсылая к ней читателя, дадим здесь лишь краткую ее характеристику.

Будем исходить из условия монотонности коррелятивной зависимости между двумя случайными величинами X и Y и расположим оба ряда измерений взаимосвязанных величин в порядке возрастания (или убывания) их значений. Тогда при достаточном числе измерений сможем определить для каждого порядкового номера m пары значений $(x_m$ и $y_m)$ его обеспеченность, согласно известной эмпирической формуле¹

$$p = \frac{m - 0,3}{n + 0,4},$$

¹ Эта эмпирическая дробь заменяет в коротких рядах ту теоретическую дробь $\frac{m}{n}$, которая справедлива лишь при очень большом значении числа n .

и получить ряд значений x_p, y_p . Тогда доказываемся (см. приложение II), что кривая связи между равнообеспеченными значениями $y_p = \varphi(x_p)$ даст наименьшие отклонения от нее эмпирических точек, причем как по абсциссе, так и по ординате. Сопоставляя этот результат с традиционным методом «способа наименьших квадратов», который дает кривую с наименьшими отклонениями или по абсциссам, или по ординатам, мы приходим к выводу, что метод Срезневского для монотонных зависимостей имеет свои преимущества.

В частности, для линейных зависимостей, для которых теория корреляции и способ наименьших квадратов дают две формы зависимостей:

$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x}), \quad (167)$$

$$y - \bar{y} = \frac{1}{r} \frac{\sigma_y}{\sigma_x} (x - \bar{x}), \quad (168)$$

метод Срезневского, разработанный Алексеевым, дает лишь одну зависимость

$$y - \bar{y} = \frac{\sigma_y}{\sigma_x} (x - \bar{x}). \quad (169)$$

§ 40. Применение принципа Лежандра для вычисления поправок при недостаточном числе измерений

Этот принцип получил, как было сказано, очень широкое практическое применение не только в основной задаче: отыскания формы эмпирических зависимостей, но и в некоторых побочных, в частности в задаче введения поправок в измеренные значения параметров в теоретически установленной зависимости.

С этой простейшей задачи мы и начнем изложение.

Предположим, нам известно, что некоторые три величины x, y, z связаны строгой линейной зависимостью вида

$$ax + by + cz = A. \quad (170)$$

Однако в результате произведенных нами измерений получены для этих величин такие значения x', y', z' , при которых левая часть уравнения (170) дает в сумме не A , как следовало бы, а $(A + h)$. Превышение экспериментально измеренной величины над теоретически установленной является, очевидно, результатом ошибок измерения. Поскольку имеем основание больше верить равенству (170), чем равенству

$$ax' + by' + cz' = A + h, \quad (171)$$

то мы вынуждены искать таких поправок к этим измеренным значениям, которые восстановили бы теоретическое равенство. Запишем условно

$$a(x' + \alpha) + b(y' + \beta) + c(z' + \gamma) = A \quad (172)$$

и будем искать значения параметров на основе принципа Лежандра, т. е. придадим им такие значения, при которых сумма их квадратов будет минимальной.

Вычтя из равенства (172) равенство (171), получим уравнение поправок

$$a\alpha + b\beta + c\gamma + h = 0, \quad (173)$$

к которому надо добавить введенное нами условие

$$\alpha^2 + \beta^2 + \gamma^2 = \min. \quad (174)$$

Вычитая левую часть уравнения (173), умноженную на неопределенный множитель $2k$, из левой части равенства (174), получим многочлен

$$\alpha^2 + \beta^2 + \gamma^2 - 2k(a\alpha + b\beta + c\gamma + h),$$

представляющий собой величину существенно положительную и тождественно равную сумме квадратов поправок вследствие равенства нулю выражения в скобках.

Следовательно, дифференцируя эту функцию поправок по α , β , γ и приравнявая эти производные нулю, получим выражение для поправок через известные коэффициенты a , b , c и через неопределенный коэффициент k . После всех операций

$$\alpha = k a; \quad \beta = k b; \quad \gamma = k c.$$

Умножая эти равенства соответственно на a , b , c , складывая и используя равенство (174), получим

$$k = - \frac{h}{a^2 + b^2 + c^2},$$

а подставляя это значение k в вышеприведенные три равенства, получим окончательно требуемые условия:

$$\alpha = - \frac{ah}{a^2 + b^2 + c^2}; \quad \beta = - \frac{bh}{a^2 + b^2 + c^2}; \quad \gamma = - \frac{ch}{a^2 + b^2 + c^2}. \quad (175)$$

Совершенно очевидно, что то же самое мы получили бы и при большем числе слагаемых.

Приведем конкретный пример, соответствующий изложенному. Применяя триангуляцию к съемке местности, нами измерены для некоторого треугольника его углы и получена их сумма, равная не 180° , как следовало бы, а на 15 мин. больше. Как разложить это превышение, очевидно обусловленное неточностью прибора, на измеренные углы?

В этой задаче

$$a = b = c = 1; \quad A = 180^\circ; \quad h = 15',$$

следовательно, по уравнению (175)

$$\alpha = \beta = \gamma = - \frac{h}{3} = - 5'.$$

§ 41. Применение принципа Лежандра при избыточном числе измерений

Особое развитие получил принцип Лежандра при решении обратной задачи, когда форма зависимости между измеренными величинами неизвестна и мы ищем ее из измерений. Совершенно очевидно, что это возможно лишь при избыточном числе измерений, когда ошибки отдельных измерений гасятся их численностью. Методика решения таких задач была в основном разработана Гауссом и в дальнейшем доработана некоторыми русскими математиками, прежде всего П. Л. Чебышевым.

Исторически эта методика получила не совсем удачное наименование «способа наименьших квадратов»¹.

Сущность этого способа нами в основном освещена в начале этой главы. Она сводится к тому, чтобы решать линейные уравнения относительно параметров тех уравнений, форма которых или заранее установлена по каким-либо соображениям по существу изучаемого явления, или была нами намечена в виде попытки. Предположим, что эта форма может быть сведена к виду

$$y = \Sigma a_k \psi_k(x), \quad (176)$$

где ψ_k — известные функции от x , не содержащие никаких неизвестных параметров, а a_k — неизвестные параметры, подлежащие определению из совокупности измерений (x_j, y_j) .

Подчеркиваем, что вид функции ψ_k принципиально может быть какой угодно: степенной, логарифмической, тригонометрический и т. д., хотя практические методы решения используются главным образом в частном случае, когда имеется простое равенство

$$\psi_k = x^k,$$

сводящее уравнение (176) к более простому виду:

$$y = a_0 + a_1x + a_2x^2 + \dots + a_kx^k + \dots + a_mx^m. \quad (177)$$

Такое упрощение оправдывается, как указано ранее, тем обстоятельством, что любая функция, разложенная в ряд Тэйлора, может быть сведена к сумме степеней независимой переменной. Но это упрощение совсем не обязательно для способа наименьших квадратов. В дальнейшем покажем, как отыскивать параметры и при более сложных формах ψ_k . Теперь же ограничимся рассмотрением лишь упрощенной формы (177). Итак, всякая задача способа наименьших квадратов является линейной по

¹ Не совсем удачное хотя бы потому, что речь идет здесь не о «наименьших квадратах», а о наименьшей их сумме. Ведь если переместить условный центр в сторону от среднего арифметического, то одни квадраты отклонений станут больше, другие меньше, но их сумма всегда будет больше.

отношению параметров. Но по отношению измеряемых величин, между которыми ищем связь, она может быть и линейной и нелинейной.

§ 42. Линейная задача

Покажем прежде всего, что при корреляционных связях между x и y общее линейное уравнение

$$Ax + By + C = 0$$

имеет три качественно различные формы. Действительно, разделив это уравнение на один из трех постоянных коэффициентов, получим соответственно

$$\frac{x}{a} + \frac{y}{b} = 1; \quad (178)$$

$$y = a_0 + a_1x; \quad (179)$$

$$x = b_0 + b_1y. \quad (180)$$

Если бы связь между x и y была функциональной, то все три уравнения были бы, очевидно, тождественны, так как их параметры легко пересчитываются. Но способ наименьших квадратов применим только к корреляционным связям (к функциональным применение его лишено смысла), и по отношению к ним все три уравнения (178), (179) и (180) по существу различны.

Для большей наглядности покажем справедливость этого, как будто «парадоксального» положения на совсем простом примере.

Два следующих равенства:

$$y = ax; \quad x = \frac{1}{a}y$$

представляются как будто бы очевидно тождественными. И если бы x и y были определенными величинами, то это и было бы справедливо! Но надо вспомнить, что каждая из них нам известна лишь по ряду произведенных и, очевидно, ошибочных измерений. Все известные результаты x_j, y_j как бы блуждают около истинных и нам неизвестных значений x и y . Но если применим к вышеприведенным равенствам способ наименьших квадратов, то получим, как будет показано, совсем различные результаты.

Вычтем из обоих равенств их осредненные значения:

$$\bar{x} = \frac{1}{a} \bar{y}; \quad \bar{y} = a\bar{x}$$

и применим к полученным разностям принцип Лежандра:

$$\sum (x_j^1 - \frac{1}{a} y_j^1)^2 = \min; \quad \sum (y_j^1 - ax_j^1)^2 = \min.$$

Приравнивая к нулю в первом случае производную по $\frac{1}{a}$, а во втором — производную по a , получим соответственно из первого уравнения

$$a = \frac{\sigma_y^2}{x'y'},$$

а из второго

$$a = \frac{\overline{x'y'}}{\sigma_x^2}.$$

Тождество обоих выражений возможно лишь в случае, когда

$$\frac{\overline{x'y'}}{\sigma_x \sigma_y} = r_{x,y} = \pm 1,$$

т. е. когда связь между x и y функциональная.

Переходим к более сложной задаче решения уравнений (179) и (180).

В первом из этих уравнений после вычитания из него его же среднего значения получим

$$y - \bar{y} = a(x - \bar{x})$$

или иначе

$$y'_j = ax'_j.$$

Принцип Лежандра даст нам

$$\sum (y'_j - ax'_j)^2 = \min,$$

а дифференцируя по a и приравнивая нулю производную, получим

$$a = \frac{\overline{x'y'}}{\sigma_x^2} = r_{x,y} \frac{\sigma_y}{\sigma_x},$$

что и приводит к окончательному виду решения уравнения (179):

$$y - \bar{y} = r_{x,y} \frac{\sigma_y}{\sigma_x} (x - \bar{x}). \quad (181)$$

Здесь нет надобности давать полностью решение аналогичного уравнения (180). Можно, не производя никаких операций, по аналогии записать

$$x - \bar{x} = r_{x,y} \frac{\sigma_x}{\sigma_y} (y - \bar{y}), \quad (182)$$

откуда видно, что, несмотря на совершенное, казалось бы, подобие двух решений, они существенно отличаются друг от друга и лишь в одном частном случае, когда $r_{x,y} = \pm 1$, оказываются тождественными.

Докажем теперь то же положение и по отношению линейного уравнения, представленного в форме (178). Поскольку искомые параметры в ней стоят знаменателями, то перейдем на этот раз к более удобным обозначениям:

$$a = \xi^{-1}; \quad b = \eta^{-1},$$

после чего система условных уравнений примет вид

$$\xi x_k + \eta y_k = 1 \quad (183)$$

(где все x_k и y_k известны по изложенному ранее).

Согласно принципу Лежандра, мы должны записать

$$\sum \{ \xi x_k + \eta y_k - 1 \}^2 = \min. \quad (184)$$

Берем производные от левой части по переменным ξ и η и приравняем их нулю. Получим

$$\left. \begin{aligned} \xi \sum x_j^2 + \eta \sum x_j y_j - \sum x_j &= 0 \\ \xi \sum x_j y_j + \eta \sum y_j^2 - \sum y_j &= 0 \end{aligned} \right\} \quad (185)$$

Решение обоих равенств в отношении переменных ξ и η настолько просто, что, не приводя всех выкладок, запишем непосредственно

$$\begin{aligned} \xi &= \frac{\sum x_j \sum y_j^2 - \sum x_j y_j \sum y_j}{\sum x_j^2 \sum y_j^2 - (\sum x_j y_j)^2}; \\ \eta &= \frac{\sum x_j \sum x_j y_j - \sum y_j \sum x_j^2}{(\sum x_j y_j)^2 - \sum y_j^2 \sum x_j^2}. \end{aligned}$$

Разделяя в обоих равенствах числитель и знаменатель на n^2 , а затем используя известные из теории корреляции равенства

$$\begin{aligned} \overline{xy} &= \bar{x}\bar{y} + \overline{x'y'}, \\ \overline{x^2} &= \bar{x}^2 + \sigma_x^2, \\ \overline{y^2} &= \bar{y}^2 + \sigma_y^2 \end{aligned}$$

и заменяя ξ и η их обратными величинами a и b , получим окончательную форму решения поставленной задачи

$$x \{ \bar{x} \sigma_y^2 - \bar{y} \overline{x'y'} \} + y \{ \bar{x} \overline{x'y'} - \bar{y} \sigma_x^2 \} = \overline{x'y'^2} - \sigma_x^2 \sigma_y^2. \quad (186)$$

Вычитая из этого уравнения его осредненное значение, после простых выкладок приведем его к следующей форме:

$$\frac{x - \bar{x}}{\bar{x} \sigma_y' - \bar{y} \sigma_x^2} + \frac{y - \bar{y}}{\bar{x} \sigma_y^2 - \bar{y} \sigma_x' y'} = 0. \quad (187)$$

Далее нетрудно показать, что при условии

$$r_{x,y} = \pm 1$$

уравнение (187) окажется тождественным с уравнениями (181) и (182).

Действительно, разделим оба знаменателя уравнения (187) на произведение $\sigma_x \sigma_y$ и приравняем единице полученные выражения $\frac{\overline{x'y'}}{\sigma_x \sigma_y}$. В этом случае первый знаменатель даст произведение $\sigma_x \sigma_y$

$$\sigma_x \left(\frac{\overline{x}}{\sigma_x} - \frac{\overline{y}}{\sigma_y} \right),$$

а второй —

$$\sigma_y \left(\frac{\overline{x}}{\sigma_x} - \frac{\overline{y}}{\sigma_y} \right).$$

После сокращения обоих произведений на общий знаменатель получим

$$\frac{x - \overline{x}}{\sigma_x} = \frac{y - \overline{y}}{\sigma_y}, \quad (188)$$

т. е. тот же самый результат, к которому приводятся также два других уравнения: (181) и (182) — при $r_{x,y} = \pm 1$, т. е. при функциональной связи между x и y .

Здесь сочли необходимым подробнее остановиться на важном и интересном факте, что уравнения, явно тождественные в функциональном смысле, оказываются различными, если их применить к коррелятивным связям между измеренными величинами. В этом факте следует разобраться и выяснить, в чем причина такого расхождения. Обратим прежде всего внимание, что все три указанные уравнения удовлетворяются при значениях $x = \overline{x}$ и $y = \overline{y}$. А так как точка с координатами $(\overline{x}, \overline{y})$ является центром тяжести всей системы точек (x_j, y_j) , то значит (и это очень важно) все три прямые проходят через одну эту точку и разница между ними только в угле их наклона к координатным осям. Если все три уравнения преобразовать так, чтобы в левой части уравнений было отношение

$$\frac{y - \overline{y}}{x - \overline{x}},$$

то в правой части получим различные тангенсы наклона прямых, а именно в уравнении (181)

$$\operatorname{tg} \alpha = r_{x,y} \frac{\sigma_x}{\sigma_y}, \quad (189)$$

в уравнении (182)

$$\operatorname{tg} \alpha = r_{x,y}^{-1} \frac{\sigma_y}{\sigma_x}, \quad (190)$$

в уравнении (187)

$$\operatorname{tg} \alpha = \frac{\left(r_{x, y} \frac{\bar{y}}{\sigma_y} - \frac{\bar{x}}{\sigma_x} \right) \sigma_y}{\left(r_{x, y} \frac{\bar{x}}{\sigma_x} - \frac{\bar{y}}{\sigma_y} \right) \sigma_x}. \quad (191)$$

Из приведенного сопоставления становится совершенно ясным, что причина расхождения в наклонах трех прямых заключается не в различии стандартов отклонений σ_x и σ_y , как это можно было принять по интуиции, а только в степени коррелятивности этих двух переменных. Действительно, при $r_{x, y} = \pm 1$ все три прямые совпадают, и чем меньше единицы модуль $r_{x, y}$, тем больше будет расхождение в наклонах прямых, вращающихся все время около своего общего центра (\bar{x}, \bar{y}) . Если представить себе предел, правда, не имеющий смысла, когда $r_{x, y} = 0$, то первая прямая даст нулевой наклон, вторая — бесконечность (т. е. угол $\alpha = \frac{\pi}{2}$), а третья — конечный тангенс, равный

$$\operatorname{tg} \alpha = \frac{\bar{x}}{y} \frac{\sigma_x^2}{\sigma_y^2}. \quad (192)$$

Этот предел приведен здесь скорее как математический курьез, лишний раз свидетельствующий о несовершенстве коэффициента корреляции в роли «индикатора связи», в особенности при низких его значениях.

При более же высоких значениях $r_{x, y}$, порядка 0,8—0,9, расхождение указанных трех прямых будет сравнительно незначительным, но все же таким, с которым надо считаться.

С другой стороны, следует обратить внимание на полную тождественность уравнений (181) и (182) с ранее выведенными из одной лишь теории корреляции. Такое совпадение получается потому, что сама теория корреляции исходит из того же самого среднего арифметического, который связан с принципом Лежандра.

Но такая тождественность ограничивается лишь областью линейных зависимостей. При переходе к зависимостям явно нелинейного характера теория корреляции может, самое большое, обнаружить наличие связи, а все оформление последней может быть получено лишь на основе одного из двух указанных принципов. Но прежде чем перейти к разбору нелинейных решений, дадим еще одно линейное решение, которое представляет особый интерес, а в известном смысле является и наиболее логичным так как оно отвечает интуитивному стремлению каждого экспериментатора провести искомую прямую возможно ближе к экспериментальным точкам.

§ 43. Вероятнейшая прямая

Поставим себе задачу провести искомую прямую так, чтобы минимальной была не сумма квадратов отрезков, параллельных осям x или y , а сумма квадратов расстояний от точек до искомой прямой.

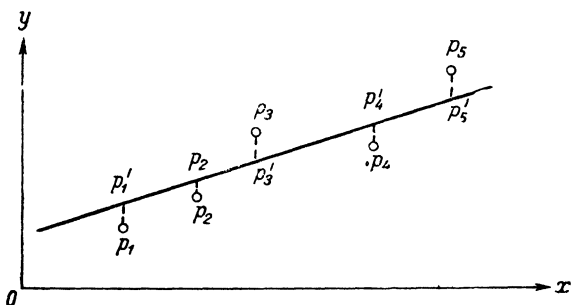


Рис. 22. Схема проведения прямой $y(x)$.

Для наглядности представим все три решения на графиках: 1) решение уравнения (181) изображено на рис. 22, 2) решение

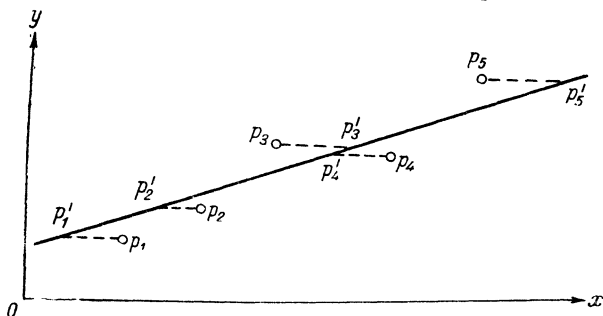


Рис. 23. Схема проведения прямой $x(y)$.

уравнения (182) — на рис. 23, 3) искомое решение — на рис. 24.

Для нахождения искомого решения придадим уравнению (179) несколько иную форму. Поскольку ясно, что параметр a_1 равен тангенсу угла наклона прямой к оси x , то имеем право записать

$$y - a_0 - a_1 x = \frac{-x \sin \alpha + y \cos \alpha - a_0 \cos \alpha}{\cos \alpha} = 0. \quad (193)$$

Если исключить тривиальный случай, когда $\cos \alpha = 0$ ($\alpha = \frac{\pi}{2}$), то это равенство заменится следующим:

$$x \sin \alpha - y \cos \alpha - a_0 \cos \alpha = 0. \quad (194)$$

Из аналитической геометрии известно, что трехчлен левой части представляет собой расстояние от точки (x, y) до данной прямой.

Применяя далее к левой части уравнения (194) принцип Лежандра, запишем

$$\sum \{ x_j \sin \alpha - y_j \cos \alpha + a_0 \cos \alpha \}^2 = \min. \quad (195)$$

Это уравнение, хотя и нелинейное по отношению к параметру α , но легко решается.

Берем сначала производную по a_0 и приравниваем ее к нулю:

$$\sum x_j \sin \alpha \cos \alpha - \sum y_j \cos^2 \alpha + a_0 \cos^2 \alpha = 0,$$

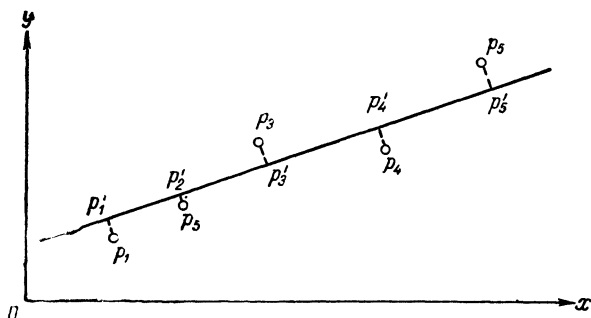


Рис. 24. Схема проведения вероятнейшей прямой.

откуда получим

$$a_0 = \frac{\sum y_j \cos^2 \alpha - \sum x_j \sin \alpha \cos \alpha}{n \cos^2 \alpha} = \bar{y} - \operatorname{tg} \alpha \bar{x}.$$

Подставляя значение a_0 в общее уравнение, получим

$$y - \bar{y} = \operatorname{tg} \alpha (x - \bar{x}), \quad (196)$$

откуда видим, что вероятнейшая прямая также проходит через ту же центральную точку (\bar{x}, \bar{y}) , через которую проходят и все три ранее полученные прямые, но угол ее наклона, как покажем далее, опять будет иным.

Теперь нам удобнее перейти к координатной системе, повернутой на угол α , с началом в центральной точке (\bar{x}, \bar{y}) . По общему правилу преобразования координат

$$\begin{aligned} \xi'_j &= x'_j \cos \alpha + y'_j \sin \alpha; \\ \eta'_j &= -x'_j \sin \alpha + y'_j \cos \alpha, \end{aligned}$$

где

$$x'_j = x_j - \bar{x}; \quad y'_j = y_j - \bar{y}.$$

В этой новой системе координат имеем очень простое соотношение

$$\frac{d\eta_j}{d\alpha} = -\xi_j. \quad (197)$$

Так как теперь все расстояния точек до прямой выражаются ординатами η_j , то принцип Лежандра представится равенством

$$\sum \eta_j^2 = \min; \quad \sum \eta_j \frac{d\eta_j}{d\alpha} = 0$$

или по предыдущему

$$\sum \xi_j \eta_j = 0. \quad (198)$$

Это интересное свойство вероятнейшей прямой сближает ее с известными в механике главными осями инерции, так как оказывается, что вероятнейшая прямая является той из главных осей инерции данной совокупности точек, по отношению которой момент инерции имеет наименьшее значение. Попутно отметим, что можно было получить такой же результат, если бы решали пространственную задачу $(x_j y_j z_j)$.

Подставляя далее в равенство (194) выражение новых координат через прежние, получим после простых операций равенство

$$\sum \{x_j^2 - y_j^2\} \sin \alpha \cos \alpha - \sum x_j^1 y_j^1 \{ \sin^2 \alpha - \cos^2 \alpha \} = 0,$$

откуда выражение для угла α

$$\operatorname{tg} 2\alpha = \frac{2 \sum x'_j y'_j}{n \{ \sigma_x^2 - \sigma_y^2 \}} = \frac{2r_{x,y}}{\frac{\sigma_x}{\sigma_y} - \frac{\sigma_y}{\sigma_x}}. \quad (199)$$

Если $\sigma_x > \sigma_y$ и $r_{x,y} > 0$, а также если $\sigma_x < \sigma_y$ и $r_{x,y} < 0$, то имеем в обоих случаях прямую с подъемом вдоль оси x . Когда же знак знаменателя в выражении (199) противоположен знаку $r_{x,y}$, прямая опускается вдоль оси x . Хорошо видно, что в обоих случаях угол (подъема или спада) $\alpha < 45^\circ$.

§ 44. Принцип Лапласа и метод Эджуорса

Покажем далее решение линейной задачи согласно принципу Лапласа, т. е. с заменой требования минимальной суммы квадратов отклонений требованием минимальной суммы их модулей. При измерении одной величины такое требование приводило, как мы знаем, к медиане совокупности измерений. При измерении двух связанных величин мы должны будем построить соответственное обобщение, аналогичное переходу от среднего арифме-

тического к «способу наименьших квадратов». Такое обобщение было дано Эджуорсом [19], но в применении лишь к линейной задаче, т. е. к двум неизвестным в системе условных уравнений. Ограничимся здесь изложением практических приемов этого метода. Аналогичный переход к параболе второго порядка, т. е. к трем переменным в системе условных уравнений, также осуществим. Однако поскольку он является громоздким и практически малоудобным, рассматривать его не будем.

Начнем с ответа на вопрос, что называть «медианой» при измерении двух величин. Нам известно, что даже при измерении одной величины медиана не всегда обладает полной определенностью. Так, при четном числе измерений приходится располагать медиану условно посередине между двумя смежными значениями. Для двух же взаимосвязанных величин дело значительно осложняется, так как определению подлежит известное число параметров (не менее двух), а принцип медианы может относиться лишь к рядам проведенных измерений, т. е. в конечном счете не к параметрам, а к самим переменным. Весь ход дальнейших рассуждений представим, как у Эджуорса, графическим построением. Считая, что расположение точек на графике дает основание предполагать эту связь линейной, примем искомую зависимость в виде

$$y = a_0 + a_1 x.$$

После подстановки в нее данных измерений будем обозначать искомые параметры через

$$a_0 = \xi; \quad a_1 = \eta.$$

Для измеренных величин примем вместо x_j и y_j постоянные a_j и b_j . Тогда n условных уравнений будут иметь вид

$$\xi + a_j \eta - b_j = 0. \quad (200)$$

Далее надо найти медианы для рядов значений a_j и b_j и определить из них те частные значения переменных параметров ξ и η , для которых принцип Лапласа будет иметь силу:

$$\sum_1^n |\xi + a_j \eta - b_j| \equiv R = \min. \quad (201)$$

Нанесем на график (рис. 25) всю систему условных уравнений (200). Направления и положения условных прямых будут, конечно, различаться, и тем больше, чем менее точны были измерения основных переменных x_j , y_j . Но все же условные прямые в основном расположатся сравнительно близко одна от другой. Область, обведенная пунктиром, в которой пересечения отдельных прямых встречаются чаще, является той областью, в которой должна также находиться точка медианы с координатами: ξ_0 , η_0 , дающими наиболее вероятные значения двух искоемых параметров. Положение точки медианы в семействе условных прямых

должно удовлетворять следующему условию: отрезок, на котором она расположена, должен иметь по обе стороны равные числа отрезков условных прямых.

На рис. 25 изображено семь условных прямых. По их отрезкам от одного пересечения до другого проведена утолщенная ломаная линия с таким расчетом, что по обе стороны от любого ее отрезка находится по равному числу условных прямых. Такую ломаную линию будем в дальнейшем называть медианной линией. Совершенно очевидно, что искомая точка ξ_0, η_0 должна

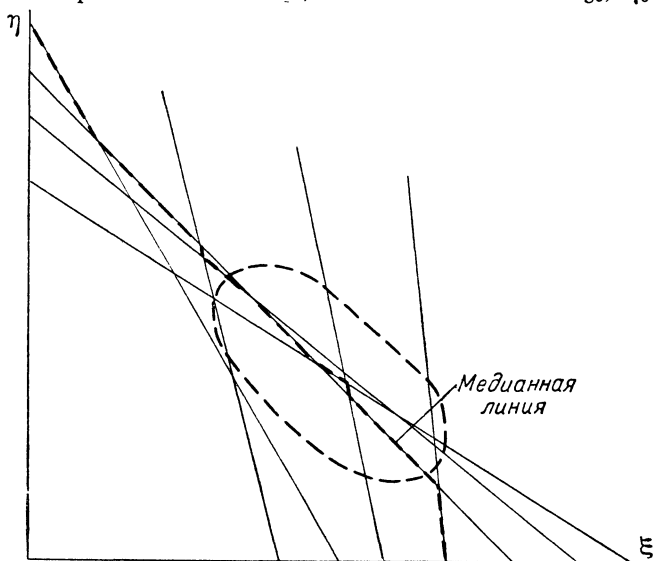


Рис. 25. Схема расположения условных прямых.

находиться на одном из отрезков медианной линии, провести которую при нечетном числе измерений (т. е. нечетном числе условных прямых) не представляет затруднений. Случай четного числа измерений здесь рассматриваться не будет, так как он вносит в расчет несущественные осложнения. Но практически всегда легче выбросить одно из измерений и оперировать далее с нечетным их числом.

Используем далее основное условие (201), в котором $R = \min$. Поскольку точка минимума R должна по изложенному располагаться на медианной линии, то условие минимума выразится через

$$\frac{dR}{ds} = 0,$$

где $ds = \sqrt{d\xi^2 + d\eta^2}$, а соотношение между дифференциалами для каждого отрезка медианной прямой дается уравнением той ус-

ловной прямой, к которой принадлежит этот отрезок. Например, для прямой $\xi + a_j \eta - b_j = 0$ имеем $d\xi = -a_j d\eta$, следовательно, $ds = d\eta \sqrt{1 + a_j^2}$.

Вдоль медианной линии слева направо (т. е. на графике в сторону убывания ординаты η) будем брать сумму соответственных a_j для условных прямых, которые располагаются вправо от медианной линии. Обозначим эту сумму через A , а сумму a_j для условных прямых, которые располагаются влево от медианной линии, — соответственно через A' .

Так как при нашем написании условных уравнений все коэффициенты при ξ равны единице, то равенство нулю производной $\frac{dR}{ds}$ зависит только от равенства сумм коэффициентов при η . Поэтому

$$\frac{dR}{ds} = \frac{(A' - A) d\eta}{ds} = \frac{A' - A}{\sqrt{1 + a_j^2}} = 0.$$

Приведенное условие справедливо, очевидно, лишь для той точки (или для того отрезка) медианной линии, для которой выполняется условие $A' = A$.

В вышеизложенном заключается весь ход расчета по методу Эджуорса, который на первый взгляд кажется громоздким и сложным. Однако это не так! Ведь все графическое построение сводится к нанесению прямых по заданным их отрезкам на координатных осях. Проведение по ним медианной линии чрезвычайно просто, а дальнейшие арифметические операции, безусловно, проще, чем в способе наименьших квадратов.

Для иллюстрации воспользуемся тем самым примером, который приводит Эджуорс в работе [19].

Для определения искомого параметра было проведено пять измерений соответственных значений x_j и y_j . Используя ту же форму условных уравнений, запишем

$$\begin{aligned} \xi + a_j \eta - b_j &= 0; \quad j = 1, 2, 3, 4, 5; \\ a_j &= 1, 1, 1, 0,5, 0,5; \\ b_j &= 4, 5, 6, 3, 3,5. \end{aligned}$$

Все пять условных прямых и медианная линия $ABCD$ изображены на рис. 26. Вычисляем последовательные значения производной $\frac{dR}{ds}$ по отрезкам:

$$AB, \quad dR = -0,4 d\eta, \quad ds = \sqrt{2} d\eta; \quad \frac{dR}{ds} = -0,28$$

$$BC, \quad dR = 0;$$

$$CD, \quad dR = 0,5 d\eta, \quad \frac{dR}{ds} = +0,35.$$

Здесь приведен весь ход изменения производной $\frac{dR}{ds}$ вдоль

всей медианной линии. Но в нашу задачу входило лишь нахождение того отрезка, для которого $\frac{dR}{ds} = 0$. Это также можно было легко сделать непосредственно из графика (рис. 26), так как только для отрезка BC имеется по обе его стороны равное количество прямых разного направления. Следовательно, только здесь возможно равенство $A = A'$, а это все, что нам надо было узнать. По обе же стороны от этого отрезка производная $\frac{dR}{ds} > 0$, т. е. R возрастает. Следовательно, лишь на этом отрезке имеем

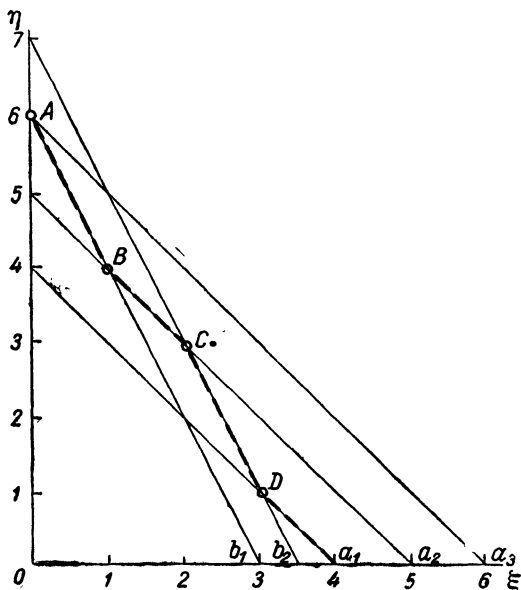


Рис. 26. Пример линейной задачи по Эджуорсу.

минимальное значение R . По общему правилу точку медианы надо взять в середине этого отрезка. Ее координаты, взятые в масштабе графика, будут равны

$$\xi_0 = 1,5; \quad \eta_0 = 3,5.$$

Заимствованный нами у Эджуорса пример, конечно, не является показательным, так как число измерений в нем очень мало. Обычно при большем числе измерений объем всех вычислений будет соответственно больше. Однако все же этот объем гораздо меньше того, который требуется способом наименьших квадратов. Что же касается графических построений, т. е. нанесения в плоскости (ξ, η) семейства условных прямых и проведения по этому семейству медианной линии, то все это элементарно просто.

Кроме того, следует учесть, что хотя при графических построениях имеют место неточности чисто чертежного происхождения, но они, вообще говоря, малы, зато грубые и значительные ошибки при графических методах вообще исключаются.

Если же ко всему этому добавить, что применение медианы вместо среднего автоматически устраняет или только ослабляет влияние «выскакивающих» подозрительных по своей маловероятности измерений, то станет ясно, что метод Эджуорса, основанный на принципе Лапласа, заслуживает, хотя бы в линейных задачах, равноправного положения со способом наименьших квадратов, а в некоторых случаях его можно даже предпочесть последнему.

§ 45. Нелинейные задачи

При отыскании вида эмпирической зависимости линейный случай встречается сравнительно часто. В то же время именно этот случай получил полное методическое развитие, дающее возможность иметь всегда однозначное решение и давать ему должную оценку. Переходя к общему случаю «нелинейных» задач, прежде всего должны подчеркнуть отрицательный характер этого термина, отрицающего лишь признак линейности, но ничего сколько-нибудь определенного не выдвигающего взамен его. Действительно, морфология эмпирических кривых (ее осветим в последней главе книги) настолько сложна и многообразна, что даже трудно говорить о едином методе выбора между различными нелинейными формами.

Но принципиальная трудность отчасти смягчается тем обстоятельством, что эмпирические зависимости являются в то же время и приближенными как в смысле неточности измерений, на которых они основаны, т. е. разброса экспериментальных точек каждой из измеряемых величин, так и в смысле малой определенности той связи, которая между этими величинами отыскивается.

Такое «отыскивание» начинается обычно с нанесения на клетчатке точек, координаты которых соответствуют одному измерению двух взаимосвязанных величин: x и y . Такой график до некоторой степени характеризует обе указанные стороны вопроса: тесноту связи, т. е. степень разброса точек, и четкость выраженности этой связи. Соотношение между двумя названными характеристиками может быть весьма различным, но, вообще говоря, чем теснота больше (т. е. разброс меньше), тем больше четкость в вырисовке формы связи. Здесь возможны четыре случая:

1. При очень большом разбросе точек обычно теряется четкость представления о различии между линейной и нелинейной связью.

2. Но если при этом отклонение от линейности очень велико, то даже при значительном разбросе точек все же оказывается возможным отказаться от прямой линии и перейти к параболе, хотя и без уверенности в ее порядке.

3. Могут быть случаи, когда разброс точек сравнительно мал, но сама зависимость настолько малоотчетлива, что выбор между прямой линией и параболой или между параболой второго и третьего порядков очень труден.

4. Хорошее совпадение большой тесноты с большой четкостью дает возможность ближе подойти к «истинной» (нам, по существу, неизвестной) форме зависимости.

Обычно именно благодаря наглядности графика получаем первую возможность установить форму связи. Дальнейший же ход процесса отыскания этой формы зависит от принятия определенного руководящего принципа. Поэтому предыдущее изложение вынуждает нас в дальнейшем остановиться на одном лишь принципе Лежандра, поскольку принцип Лапласа, при всех его достоинствах, а иногда и преимуществах, практически далее линейных зависимостей до сих пор не был продвинут.

Зато принцип Лежандра благодаря в первую очередь трудам Гаусса, а затем выдающихся русских математиков Чебышева и Маркова получил настолько обстоятельное развитие, что его применение в огромном диапазоне практических задач может считаться вполне освещенным.

Ограничиваясь в этой главе лишь зависимостью между двумя величинами:

$$y = f(x),$$

каждая из которых известна нам по ряду произведенных измерений:

$$\begin{array}{c} x_1, x_2, \dots, x_j, \dots, x_n \\ y_1, y_2, \dots, y_j, \dots, y_n \end{array}$$

Мы отыскиваем такую форму этой связи, которая при достаточной общности давала бы наилучшие возможности для применения к ней принципа Лежандра.

Поскольку известно, что, действуя согласно принципу Лежандра, мы должны будем брать производные величины:

$$\{y_j - f(x_j)\}^2$$

по тем неизвестным параметрам, которые скрыты в условной форме функции $f(x)$, то естественно в этом случае искать такую форму функции, которая была бы в отношении этих параметров **линейной**.

Этому требованию, как было сказано, удовлетворяет одна, и только одна, форма связи:

$$y = \sum_0^m \xi_k \psi_k(x), \quad (202)$$

где ξ_k — неизвестные параметры, которые мы должны найти из совокупности значений $(x_j y_j)$, а ψ_k — любые известные функции от переменной x .

В дальнейшем нам придется суммировать величины как по индексу j , обозначающему номер измерения основных величин, так и по индексу k , обозначающему номер параметра в разложении искомой функции. Чтобы избежать возможного смешения, будем применять в первом суммировании символы, введенные впервые Гауссом и получившие в способе наименьших квадратов широкое применение, а именно:

$$\sum x_j = [x]; \quad \sum x_j y_j = [xy]; \quad \sum x_j^2 = [xx],$$

$$\sum \psi_k(x_j) \psi_l(x_j) = [\psi_k \psi_l]$$

и т. д.

Применяя теперь к условным уравнениям (200) принцип Лежандра, получим сначала равенство

$$\sum_{j=1}^{j=n} \{y_j - \sum \xi_k \psi_k(x_j)\}^2 = \min.$$

Затем, дифференцируя это равенство последовательно по всем параметрам ξ_k ($k=0, 1, 2, \dots, m$) и приравнявая все производные к нулю, получим систему из m нормальных уравнений вида

$$\xi_0 [\psi_0 \psi_0] + \xi_1 [\psi_0 \psi_1] + \dots = [\psi_0 y];$$

$$\xi_0 [\psi_1 \psi_0] + \xi_1 [\psi_1 \psi_1] + \dots = [\psi_1 y]$$

и т. д. или в сокращенном виде

$$\sum_{l=0}^{l=m} \xi_k [\psi_k \psi_l] - [\psi_k y] = 0. \quad (203)$$

В этом уравнении все x и y получены измерениями, следовательно, все ψ_k для всех значений x_j также известны, а поэтому и все скобки Гаусса представляют собой известные числа. В результате имеем $(m+1)$ уравнений первой степени с $(m+1)$ неизвестными ξ_k ($k=0, 1, 2, 3, \dots, m$).

Задача может считаться принципиально решенной при любом виде функции ψ_k .

Далее введем одно практическое упрощение, т. е. будем считать, что все функции ψ_k есть просто различные степени переменного x :

$$\psi_0 = 1, \quad \psi_1 = x, \quad \psi_2 = x^2$$

и т. д.

Так как в большинстве практически важных случаев мы имеем право представить приближенно любую функцию в форме степенного многочлена, то такое ограничение не является жестким. Повышая порядок параболы m , мы принципиально в со-

стоянии достичь любого приближения ее к совокупности наших измерений.

Но здесь возникает вопрос, связанный опять-таки с теснотой связи, т. е. с разбросом точек. Если разброс велик и как следствие этого форма связи имеет малую четкость, то имеет ли смысл повышать порядок параболы в целях «уточнения» формы зависимости, когда исходный материал в этом отношении недостаточно точен? Не рискуем ли мы получить в результате чистое фиктивное уточнение?

Д. И. Менделеев сказал: «Точность вычислений не должна превосходить точности измерений». Это мудрое практическое правило в полной мере относится также к решению задачи отыскания и оформления зависимости между измеренными величинами, так как нет смысла уточнять зависимости между неточными величинами. Перефразируя отчасти Менделеева, можно сказать, что точность связи между измеренными величинами не может превосходить точности измерения самих величин.

Именно с этой точки зрения мы и должны подходить к вопросу перехода от линейной связи $m=1$ к параболе второго порядка $m=2$ или от второго порядка к третьему и т. д. Но для того, чтобы последовательно осуществить такой переход с постепенным повышением порядкового индекса m , необходимо построить все решение так, чтобы при переходе к следующему порядку не приходилось заново подсчитывать все параметры ξ_n , ранее вычисленные для предыдущего порядка. Это достигается особым способом, разработанным П. Л. Чебышевым и изложенным нами в следующем параграфе. Сейчас же дадим схему решения уравнения (202), правая часть которого является суммой первых $(m+1)$ членов степенного ряда.

Для сокращения записи введем упрощенные обозначения (которые сохраним в дальнейшем, так как они являются удобными):

$$a = 1; \quad b = x; \quad c = x^2; \quad d = x^3; \quad l = y.$$

В обозначениях Гаусса получим соответственно:

$$\begin{aligned} [aa] &= n, & [ab] &= n\bar{x}, & [bb] &= [ac] = nx^2; \\ [al] &= n\bar{y}, & [bl] &= n\bar{xy} \end{aligned}$$

и т. д.

В этих обозначениях система нормальных уравнений, соответствующих основному уравнению (202), примет вид

$$\left. \begin{aligned} [aa] \xi_0 + [ab] \xi_1 + [ac] \xi_2 + \dots - [al] &= 0 \\ [ba] \xi_0 + [bb] \xi_1 + [bc] \xi_2 + \dots - [bl] &= 0 \\ [ca] \xi_0 + [cb] \xi_1 + [cc] \xi_2 + \dots - [cl] &= 0 \end{aligned} \right\} \quad (204)$$

Последовательное исключение переменных $\xi_0, \xi_1, \xi_2, \dots$ будем вести по следующему способу, также данному Гауссом.

Умножим первое уравнение на $\frac{[ab]}{[aa]}$ и вычтем его из второго. Будем иметь

$$\left\{ [bb] - \frac{[ab][ab]}{[aa]} \right\} \xi_1 + \left\{ [bc] - \frac{[ab][ac]}{[aa]} \right\} \xi_2 + \\ + \dots - \left\{ [bl] - \frac{[ab][al]}{[aa]} \right\} = 0.$$

Затем, снова умножив первое уравнение на $\frac{[ac]}{[aa]}$ и вычтя его из третьего уравнения, получим

$$\left\{ [bc] - \frac{[ab][ac]}{[aa]} \right\} \xi_1 + \left\{ [cc] - \frac{[ac][ac]}{[aa]} \right\} \xi_2 + \\ + \dots - \left\{ [cl] - \frac{[ac][al]}{[aa]} \right\} = 0.$$

В целях установления компактности и симметрии решений обычно пользуются сокращенными обозначениями, также введенными Гауссом:

$$[bb] - \frac{[ab][ab]}{[aa]} = [bb. 1];$$

$$[bl] - \frac{[ab][al]}{[aa]} = [bl. 1]$$

и т. д.

Единица в скобках правой части означает замену в числителе первой буквой, стоящей в знаменателе вычитаемого, т. е. буквой a , последовательно первой и второй буквы уменьшаемого. После введения этих обозначений система (204) примет после исключения из нее первой переменной ξ_0 следующую форму:

$$[bb. 1] \xi_1 + [bc. 1] \xi_2 + \dots - [bl. 1] = 0;$$

$$[bc. 1] \xi_1 + [cc. 1] \xi_2 + \dots - [cl. 1] = 0.$$

В том же порядке исключаем вторую переменную, пользуясь аналогичными сокращениями:

$$[cc. 1] - \frac{[bc. 1][bc. 1]}{[bb. 1]} = [cc. 2],$$

$$[cl. 1] - \frac{[bc. 1][bl. 1]}{[bb. 1]} = [cl. 2],$$

где цифры 2 в скобках означают букву b в знаменателе. В результате получим вторую преобразованную форму нормальных уравнений:

$$[cc. 2] \xi_2 + \dots - [cl. 2] = 0,$$

по которой в системе из трех нормальных уравнений непосредственно определяется третья переменная ξ_2 . Далее, подставляя полученное значение в первое из уравнений первого преобразования, получим значение ξ_1 , а подставив оба эти значения в первое из уравнений основной системы (204), будем иметь значение первой переменной ξ_0 .

Таким образом из первоначальной системы нормальных уравнений получим следующую систему так называемых элиминационных уравнений:

$$\left. \begin{aligned} [aa] \xi_0 + [ab] \xi_1 + [ac] \xi_2 + \dots - [al] &= 0 \\ [bb \cdot 1] \xi_1 + [bc \cdot 1] \xi_2 + \dots - [bl \cdot 1] &= 0 \\ [cc \cdot 2] \xi_2 + \dots - [cl \cdot 2] &= 0 \end{aligned} \right\} \cdot (205)$$

Мы ограничились для конкретности тремя параметрами в уравнении (202), иначе говоря, параболой второго порядка. Однако простота конструкции элиминационных уравнений позволяет без особого труда распространить эту же форму решения также на любое число переменных (применяя и далее аналогичные сокращенные обозначения Гаусса). Неудобным является лишь то, что трудоемкость вычисления основных сумм вида $[aa]$; $[ab]$; $[bc]$; $[bc \cdot 1]$ и т. д. чрезвычайно быстро возрастает со степенью параболы. Из структуры уравнений хорошо видно, что для параболы третьего порядка требуется возводить полученные измерением числа в пятую и шестую степени, а для параболы четвертой степени — соответственно в седьмую и восьмую степени. При большом числе измерений дальнейшее суммирование всех этих степеней, даже при использовании арифмометра, требует огромной затраты времени. Опыт показал, что, пользуясь (при возведении степеней) таблицами логарифмов, мы незначительно экономим время, а поэтому применять эти таблицы на практике не рекомендуется. Все эти соображения побуждают проявлять большую «скупость» при повышении степени той параболы, которая должна применяться для интерполирования поля измеренных точек.

Ввиду того что обозначениями Гаусса, как весьма удобными и компактными, мы будем пользоваться в следующей главе, посвященной изложению способа Чебышева, то для удобства приведем в табл. 9 обозначения, совмещенные с параллельными обозначениями через основные переменные x , y .

Обращаем внимание на существенный недостаток изложенного метода непосредственного подсчета параметров ξ_k из степенного ряда: каждый раз при переходе к более высокому порядку параболы мы обязаны вновь пересчитывать все ранее подсчитанные параметры. Например, если рассчитываем ограни-

$$a = 1; \quad b = x; \quad c = x^2; \quad d = x^3; \quad e = x^4; \quad l = y$$

$$[aa] = n; \quad [ab] = n\bar{x}; \quad [ac] = n\bar{x}^2;$$

$$[bc] = [ad] = n\bar{x}^3; \quad [ae] = n\bar{x}^4;$$

$$[bb] = n\bar{x}^2; \quad [bc] = n\bar{x}^3; \quad [bd] = n\bar{x}^4$$

$$[al] = n\bar{y}; \quad [bl] = n\bar{x}y; \quad [cl] = n\bar{x}^2y;$$

$$[bb. 1] = [bb] - \frac{[ab][ab]}{[aa]} = n \{ \bar{x}^2 - \bar{x}^2; \} = n\sigma_x^2;$$

$$[bc. 1] = [bc] - \frac{[ab][ac]}{[aa]} = n \{ \bar{x}^3 - \bar{x} \bar{x}^2 \}; = n\bar{x}'y';$$

$$[bl. 1] = [bl] - \frac{[ab][al]}{[aa]} = n \{ \bar{x}y - \bar{x}^2\bar{y} \};$$

$$[cl. 1] = [cl] - \frac{[ac][al]}{[aa]} = n \{ \bar{x}^2y - \bar{x}^2y \};$$

$$[cc. 1] = [cc] - \frac{[ac][ac]}{[aa]} = n \{ \bar{x}^4 - \bar{x}^2 \};$$

$$[cc. 2] = [cc. 1] - \frac{[cb. 1][cb. 1]}{[bb]} = n \left\{ \bar{x}^4 - \bar{x}^2 - \frac{(\bar{x}^3 - \bar{x} \bar{x}^2)^2}{\sigma_x^2} \right\};$$

$$[cl. 2] = [cl. 1] - \frac{[bl. 1][cb. 1]}{[bb]} = n \left\{ \bar{x}^2y - \bar{x}^2y - \frac{(\bar{x}y - \bar{x} \bar{y})(\bar{x}^3 - \bar{x} \bar{x}^2)}{\bar{x}^2} \right\};$$

$$[cl. 3] = [cl. 2] - \frac{[cl. 2][cc. 2]}{[cc]} \text{ и т. д.}$$

читься линейной зависимостью, то значение ξ_1 окажется равным отношению

$$\frac{[bl. 1]}{[bb. 1]}.$$

Если же, найдя линейную зависимость неудовлетворительной, решили перейти к параболе второго порядка, то тот же параметр ξ_1 получает, как это нетрудно видеть из вышеизложенной схемы расчета, новое значение:

$$\xi_1 = \frac{[bl. 1]}{[bb. 1]} + \frac{[bc. 1][cl. 2]}{[cc. 2]}.$$

В это значение входят уже величины c , равные x^2 , которые в линейной зависимости не могли иметь места.

Покажем на примере, как надо использовать данные табл. 5 для расчета параболы второго порядка. В целях некоторого технического упрощения приведем пример, в котором измерения переменной, зависящей от x , производились симметрично от центра через равные интервалы переменной x . На практике такие случаи бывают. Например, измеряется температура при нагревании какого-либо слабопроводящего стержня в разных его точках или измеряются высоты на поверхности земли вдоль некоторой прямой через равные расстояния от данной точки в обе стороны и т. д. Отметим, что в этих случаях измерения x считаются «точными», т. е. их ошибки игнорируются, а расстояние x относится уже не к точности измерений, а лишь к последовательности длин отрезков. В табл. 10 приводится ряд измерений и произведения, которые нам нужны.

Таблица 10

x_j	x_j^2	x_j^3	x_j^4	y_j	$x_j y_j$	$x_j^2 y_j$	y_j^2
-6	36	-216	2 196	2	-12	72	4
-4	16	-64	256	10	-40	160	100
-3	9	-27	81	17	-51	153	289
-1	1	-1	1	33	-33	33	1 089
0	0	0	0	51	0	0	2 601
+1	1	+1	1	66	+66	66	4 356
+3	9	+27	81	96	+288	864	9 216
+4	16	+64	256	120	+480	1 920	14 400
+6	36	+216	1 296	172	+1 032	6 192	29 584
Сумма 0	124	0	3 268	567	1 730	9 460	61 639

По суммам в таблице подсчитываем средние значения указанных произведений, а затем по приведенной ранее табл. 9 находим значения скобок Гаусса и решаем систему элиминационных уравнений. Не приводя здесь простейших операций, запишем непосредственный результат искомой параболы:

$$y = 48,44 + 13,952x + 1,06x^2.$$

Если произвести подсчет суммарных отклонений этой параболы от измеренных точек, то он покажет ее практически достаточную точность.

§ 46. Способ Чебышева

Идея этого способа сводится к ортогонализации функций ψ_h , входящих в состав условных и нормальных уравнений. Поясним смысл этого термина.

Система функций $\psi_0, \psi_1, \psi_2, \dots, \psi_h, \dots, \psi_m$, зависящих от переменной x , представленной рядом ее измеренных значений $x_1,$

$x_2, \dots, x_j, \dots, x_n$ называется ортогональной, если произведение двух ее значений $\psi_l \psi_k$ при $l \neq k$, просуммированное на все значения x_j ($j=1, 2, 3, \dots, n$), равно нулю:

$$[\psi_l \psi_k] = 0.$$

В применениях способа наименьших квадратов к нахождению эмпирических зависимостей (иначе говоря, к интерполяции эмпирических данных), мы имеем в общем случае систему из n условных уравнений вида

$$y_j = \sum_{k=0}^{k=m} \xi_k \psi_k,$$

содержащую m неизвестных параметров и столько же известных функций от измеренных значений переменной x . Выводимая из этой системы другая система из $(m+1)$ нормальных уравнений имеет вид

$$\sum_{j=1}^{j=n} \psi_l \{ \xi_k \psi_k - y_j \} = 0.$$

В предыдущем изложении всюду введено упрощающее допущение $\psi_k = x^k$. Нетрудно видеть, что такая структура функций ψ_k условию ортогональности не удовлетворяет. Но так как это условие, как будет видно в дальнейшем, вносит значительные упрощения в решение нормальных уравнений и, кроме того, что еще важнее, дает возможность разрешить вопрос о выборе порядка интерполирующей параболы, то надо попытаться систему ψ_k преобразовать к ортогональному виду.

Этого можно было бы добиться различными путями; но проще всего этого добиться, оставаясь в плоскости степенного многочлена. Представим каждое из его слагаемых также в виде степенного многочлена, причем так, чтобы вторичная система удовлетворяла условию ортогональности.

Действительно, если придадим k -той функции вид суммы k -той степени переменной x и многочлена $(k-1)$ степени, то получим

$$\psi_k = x^k + \sum_0^{k-1} a_l \psi_l.$$

Умножим обе части равенства на ψ_l и просуммируем на все значения x . В левой части, где $k > l$, будет, очевидно, нуль, а в правой части останутся лишь произведения вида $[\psi_l \psi_l]$, т. е. с равными индексами, остальные также исчезнут. Получим

$$[x^k \psi_l] + a_l [\psi_l \psi_l] = 0. \quad (206)$$

Из этого равенства можем последовательно определить функции ψ_l .

Многочлен ψ_0 , как независимый от x (о чем говорит нулевой индекс), примем за единицу.

Многочлен ψ_1 должен быть линейной функцией от x и от предыдущей функции ψ_0 . Напишем

$$\psi_1 = x + b\psi_0.$$

По условию ортогональности

$$[\psi_0\psi_1] = [x] + [\psi_0\psi_0]b = 0,$$

откуда

$$b = -\frac{[x]}{n} = -\bar{x}$$

и, следовательно,

$$\psi_1 = x - \bar{x}.$$

Для многочлена ψ_2 естественно принять параболу второго порядка вида

$$\psi_2 = x^2 + b_1\psi_1 + b_0\psi_0.$$

Подставляя далее известные значения ψ_0 и ψ_1 , получим

$$\psi_2 = x^2 + b_1(x - \bar{x}) + b_0.$$

Умножая обе части равенства сначала на ψ_0 , а затем на ψ_1 , получим после суммирования по x

$$[\psi_2\psi_0] = 0 = [x^2] + b_0n;$$

$$[\psi_2\psi_1] = 0 = [x^2(x - \bar{x})] + b_1[(x - \bar{x})^2],$$

откуда будем иметь

$$b_0 = \frac{[x^2]}{n}; \quad b_1 = -\frac{[x^2(x - \bar{x})]}{[(x - \bar{x})^2]}.$$

Следовательно,

$$\psi_2 = x^2 - \frac{[x^2(x - \bar{x})]}{[(x - \bar{x})^2]}(x - \bar{x}) - \frac{[x^2]}{n}.$$

Обращаем внимание читателя, что в обоих равенствах (для ψ_1 и для ψ_2), так же как и в следующих, величины, стоящие в скобках Гаусса или имеющие черту осреднения, являются постоянными, полученными из совокупности измеренных значений x .

Следующая функция ψ_3 находится тем же путем. Принимаем для нее

$$\psi_3 = x^3 + b_2\psi_2 + b_1\psi_1 + b_0\psi_0.$$

Подставляя известные значения ψ_0 , ψ_1 , ψ_2 , получим

$$\begin{aligned} \psi_3 = x^3 + b_2 \left\{ x^2 - \frac{[x^2(x - \bar{x})]}{[(x - \bar{x})^2]}(x - \bar{x}) + \frac{[x^2]}{n} \right\} + \\ + b_1(x - \bar{x}) + b_0. \end{aligned}$$

Умножая последовательно на ψ_0 , ψ_1 и ψ_2 , суммируя по x и применяя условие ортогональности, будем иметь

$$\begin{aligned} [\psi_3\psi_0] &= 0 = [x^3] + b_0n; \\ [\psi_3\psi_1] &= 0 = [x^3\psi_1] + b_1[\psi_1\psi_1]; \\ [\psi_3\psi_2] &= 0 = [x^3\psi_2] + b_2[\psi_2\psi_2]. \end{aligned}$$

Чтобы избежать слишком громоздких выражений, на этот раз сокращенно обозначим приведенное выше значение ψ_2 через $A(x)$. Тогда коэффициенты b_0 , b_1 и b_2 получат следующий вид:

$$\begin{aligned} b_0 &= \overline{x^3}; \\ b_1 &= -\frac{[x^3(x-\overline{x})]}{(x-\overline{x})^2}; \\ b_2 &= -\frac{[x^3A(x)]}{[A(x)A(x)]}. \end{aligned}$$

Перестроим теперь полученные выше выражения для ψ_0 , ψ_1 , ψ_2 с обозначениями Гаусса (стр. 201), согласно которым сделаем заключение о следующих тождествах:

$$\begin{aligned} \frac{[ab]}{[aa]} &= \overline{x}; & \frac{[ac]}{[aa]} &= \overline{x^2}; \\ \frac{[bc.1]}{[bb.1]} &= \frac{\overline{x^3 - x^2\overline{x}}}{\overline{x^2 - \overline{x}^2}}, \end{aligned}$$

которые дают нам право записать ранее выведенные выражения в иной форме:

$$\begin{aligned} \psi_0 &= 1; \\ \psi_1 &= x - \frac{[ab]}{[aa]}\psi_0; \\ \psi_2 &= x^2 - \frac{[bc.1]}{[bb.1]}\psi_1 - \frac{[ac]}{[aa]}\psi_0. \end{aligned}$$

Используя симметрию, которая свойственна вообще обозначениям Гаусса, можем по аналогии записать также следующие равенства:

$$\begin{aligned} \psi_3 &= x^3 - \frac{[cd.2]}{[cc.2]}\psi_2 - \frac{[bd.1]}{[bb.1]}\psi_1 - \frac{[ad]}{[aa]}\psi_0; \\ \psi_4 &= x^4 - \frac{[de.3]}{[dd.3]}\psi_3 - \frac{[ce.2]}{[cc.2]}\psi_2 - \frac{[be.1]}{[bb.1]}\psi_1 - \frac{[ae]}{[aa]}\psi_0 \end{aligned}$$

и т. д.

Впрочем, на практике очень редко вычисляют четвертую степень.

Предоставляем читателю самому проверить тождественность данных нами по аналогии выражений для ψ_3 и ψ_4 с теми, которые

получаются по вышеизложенному методу постепенного повышения индекса k .

В дальнейшем нам придется пользоваться также выражениями вида

$$[\psi_k \psi_k],$$

иначе говоря, суммами квадратов многочленов Чебышева, распространенными на всю совокупность измерений. В этом случае имеем

$$\begin{aligned} [\psi_0 \psi_0] &= [aa] = n; \\ [\psi_1 \psi_1] &= [bb. 1] = n(\bar{x}^2 - \bar{x}^2) = n\sigma_x^2. \end{aligned}$$

Для следующего значения k получим выражение путем довольно простых операций. Во-первых, имеем очевидное равенство

$$[x^2] = [ac] = \frac{[ac]}{[aa]} \psi_0,$$

во-вторых, из табл. 9 нам известны равенства

$$\begin{aligned} [bc. 1] &= [(x - \bar{x}) x^2]; \\ [cc. 1] &= [cc] - \frac{[ac]}{[aa]}; \end{aligned}$$

$$[cc. 2] = [cc. 1] - \frac{[bc. 1][bc. 1]}{[bb. 1]}.$$

Возводя в квадрат полученное выше выражение для ψ_2 и суммируя на все значения x , легко получим следующие равенства:

$$\begin{aligned} [\psi_2 \psi_2] &= \frac{[bc. 1][bc. 1]}{[bb. 1]^2} [bb. 1] + \frac{[ac]^2}{[aa]^2} [aa] + [cc] - \\ &- 2 \frac{[bc. 1]}{[bb. 1]} [(x - \bar{x}) x^2] - 2 \frac{[ac]^2}{[aa]} = [cc] - \frac{[bc. 1]^2}{[bb. 1]} - \\ &- \frac{[ac]^2}{[aa]} = [cc. 2]. \end{aligned}$$

Далее, пользуясь также симметрией обозначений Гаусса, запишем

$$\left. \begin{aligned} [\psi_0 \psi_0] &= [aa]; & [\psi_1 \psi_1] &= [bb. 1]; & [\psi_2 \psi_2] &= [cc. 2] \\ [\psi_3 \psi_3] &= [dd. 3]; & [\psi_4 \psi_4] &= [ee. 4] \end{aligned} \right\} \quad (207)$$

и т. д.

Переходим теперь к составлению системы нормальных уравнений, которая в условиях ортогонализации многочленов значительно упрощается. Раскрывая общую систему уравнений (205), действительно получаем очень простую систему уравнений:

$$\xi_k [\psi_k \psi_k] = [y_j \psi_k]; \quad j = 1, 2, 3, \dots, n,$$

откуда

$$\xi_k = \frac{[y_j \psi_k]}{[\psi_k \psi_k]} . \quad (208)$$

Из приведенного выше видно, что при использовании способа Чебышева нет надобности выводить дополнительную систему элиминационных уравнений, так как система нормальных уравнений решается автоматически.

Преобразуем систему уравнений (208) к обозначениям Гаусса. Пользуясь также данными табл. 9, сразу запишем

$$\xi_0 = \frac{[al]}{[aa]} ; \quad \xi_1 = \frac{[bl.1]}{[bb.1]} ; \quad \xi_2 = \frac{[cl.2]}{[cc.2]} . \quad (209)$$

Распространяя обнаруженную аналогию на дальнейшие значения, получим

$$\xi_3 = \frac{[dl.3]}{[dd.3]} ; \quad \xi_4 = \frac{[el.4]}{[ee.4]} \quad (209')$$

и т. д.

Мы изложили метод подсчета ортогональных многочленов Чебышева, а также метод вычисления параметров в основном уравнении

$$y = \sum_0^m \xi_k \psi_k(x).$$

Из предыдущего ясно, что как многочлены, так и параметры вычисляются последовательно, переходя от низших значений параметра k к более высоким. Нам остается показать, как определяется суммарная погрешность при каждом значении k .

Запишем для j -того измерения очевидное равенство

$$\sum_{k=0}^{k=m} \{ \xi_k \psi_{k,j} - y_j \} = \varepsilon_j,$$

где через ε_j обозначена неизвестная ошибка.

Возьмем сумму квадратов левых и правых частей этого равенства:

$$\sum_{j=1}^{j=n} \sum_{k=0}^{k=m} (\xi_k \psi_{k,j} - y_j)^2 = [\varepsilon\varepsilon]$$

и раскроем в левой части осреднение по индексу j :

$$[\varepsilon\varepsilon] = \sum_{k=0}^{k=m} \xi_k^2 [\psi_k \psi_k] + [yy] - 2 \sum_{k=0}^{k=m} \xi_k [\psi_k y_j].$$

Далее, используя равенство (208), после простых выкладок и

сокращений получим искомую зависимость в форме

$$[\epsilon\epsilon] = [yy] - \sum_{k=0}^{k=m} \xi_k^2 [\psi_k \psi_k]. \quad (210)$$

Поскольку вычитаемое из правой части при любом k есть величина существенно положительная, то ясно, что с каждым повышением индекса k среднее значение суммарных отклонений уменьшается.

Раскрывая выражение (210) с помощью уравнений (208) и (209), сразу же получим общий вид суммарных ошибок:

$$[\epsilon\epsilon] = [ll] - [aa] \xi_0^2 - [bb. 1] \xi_1^2 - [cc. 2] \xi_2^2 - \dots \quad (211)$$

Рассмотрим частные значения сумм $[\epsilon\epsilon]$ для $k=0$ и $k=1$. По предыдущему

$$[\epsilon\epsilon]_{k=0} = n \{ \overline{y^2} - \bar{y}^2 \} = n\sigma_y^2; \quad [\epsilon\epsilon]_{k=1} = \frac{\overline{x'y'^2}}{\sigma_x^2}. \quad (212)$$

В сумме получаем для линейной зависимости выражение

$$M.O. = \frac{[\epsilon\epsilon]}{n} = \sigma_y^2 - \frac{\overline{x'y'^2}}{\sigma_x^2} = \sigma_y^2 (1 - r_{x,y}^2). \quad (213)$$

Выражение (213) служит хорошим дополнением к данным, полученным ранее для линейной зависимости, так как в нем отчетливо видна выраженная связь между суммарным отклонением точек от прямой и коэффициентом корреляции между переменными x и y .

Следуя Идельсону [16, стр. 172—175], возьмем, так же как и он, для иллюстрации способа Чебышева пример, который приводится в трудах П. Л. Чебышева (т. 1, стр. 493).

Некоторая пара взаимосвязанных величин была измерена 11 раз. На основании этих измерений отыскивалось выражение корреляционной связи между x и y . Не приводя здесь значения непосредственных измерений, дадим лишь выражения сумм Гаусса, которые затем будут использованы:

$$\begin{aligned} [aa] &= 11; & [ab] &= 5,403; & [bb] &= 3,683; \\ [bc] &= 3,087; & [cc] &= 2,896; & [al] &= 303,210; \\ [bl] &= 156,678; & [cl] &= 107,595; & [ll] &= 8590,8. \end{aligned}$$

По этим данным читатель также легко подсчитает следующие выражения Гаусса:

$$\begin{aligned} [bb. 1] &= 1,029; & [bc. 1] &= 1,280; & [bl. 1] &= 7,978; \\ [cc. 1] &= 1,592; & [cl. 1] &= 6,590. \end{aligned}$$

Отсюда вычисляем по формулам (207) выражение для первых трех многочленов Чебышева:

$$\begin{aligned}\psi_0 &= 1; \\ \psi_1 &= x - 0,491; \\ \psi_2 &= x^2 - 1,242x + 0,275.\end{aligned}$$

Соответственные значения параметров, вычисленные по формулам (209), будут равны:

$$\xi_0 = 27,564; \quad \xi_1 = 7,532; \quad \xi_2 = -47,270.$$

Подставляя значения ψ_k и ξ_k в основное уравнение

$$y = \sum_{k=0}^{k=m} \xi_k \psi_k,$$

получим после простых действий для трех значений m :

$$\begin{aligned}m = 0; \quad y &= 27,564; \\ m = 1; \quad y &= 23,865 + 7,532x; \\ m = 2; \quad y &= 10,847 + 66,256x - 47,27x^2.\end{aligned}$$

Первое выражение соответствует игнорированию зависимости y от x , второе — линейной зависимости между ними, третье — параболе второго порядка.

Если вычислим по формулам (207) значение $(\psi_k \psi_k)$, то легко сможем по формулам (212) определить также величины сумм погрешностей:

$$\begin{aligned}\text{при } m = 0 \quad [e\epsilon] &= 233,0, \\ \text{при } m = 1 \quad [e\epsilon] &= 174,6, \\ \text{при } m = 2 \quad [e\epsilon] &= 7,0.\end{aligned}$$

Резкое падение сумм погрешностей при переходе от линейной зависимости к параболе и сравнительно малое значение их при параболе второго порядка показывают, что на этой стадии мы можем остановиться.

ОТЫСКИВАНИЕ ФОРМЫ СВЯЗИ МЕЖДУ ТРЕМЯ ИЗМЕРЕННЫМИ ВЕЛИЧИНАМИ

§ 47. Расширение способа наименьших квадратов на многие переменные

В науках эмпирического характера часто приходится отыскивать связи между тремя (и даже более) измеренными величинами.

Рассмотрим этот вопрос сначала в общем виде. Мы имеем основание предполагать или из теоретических соображений, или из корреляционных расчетов, что некоторая величина u , являющаяся одной из характеристик изучаемого явления, зависит от ряда факторов, выраженных какими-то величинами $u = F(x, y, z)$. Если к тому же мы вправе предполагать эту зависимость линейной, то после n -кратных измерений всех входящих в нее величин мы получим систему из n условных уравнений вида

$$\left. \begin{aligned} u_1 &= A + Bx_1 + Cy_1 + Dz_1 \\ u_2 &= A + Bx_2 + Cy_2 + Dz_2 \\ \dots & \dots \dots \dots \dots \dots \dots \dots \\ u_n &= A + Bx_n + Cy_n + Dz_n \end{aligned} \right\} \quad (214)$$

где A, B, C, D — искомые параметры, а величины с индексами — измеренные, а стало быть, известные.

Непосредственно бросается в глаза, что различие между приведенной системой и системой (202) весьма несущественное: там, где прежде имелись различные степени одной переменной x , теперь имеются различные переменные, к которым мы вправе применить те же обозначения, а именно:

$$a = 1; \quad b = x; \quad c = y; \quad d = z.$$

Возвращаясь соответственно к прежним обозначениям для неизвестных параметров, т. е. заменяя $A = \xi_0; B = \xi_1; C = \xi_2; D = \xi_3$, пользуясь, как прежде, обозначениями сумм по Гауссу и при-

меня принцип Лежандра, получим в результате нормальную систему уравнений, в точности повторяющую систему (203).

Совершенно очевидно, что то различие в содержании символов a, b, c, \dots , на которое указывалось выше, ни в коей мере не может повлиять на ход решения этой, по существу иной, системы нормальных уравнений. Следуя методу последовательного исключения неизвестных параметров Гаусса, получим опять ту же систему элиминационных уравнений (204). Иными словами, наша линейная система (214) со многими независимыми переменными x, y, z решается при помощи того же порядка операций, что и нелинейная система, в которую входит одна переменная x в степенях $1, x, x^2, \dots$ потому, что в обоих случаях мы получаем линейную систему относительно неизвестных параметров $\xi_0, \xi_1, \xi_2, \dots$.

Но тогда возникает вопрос, нельзя ли и в системе (214) заменить измеренные значения основных переменных x_j, y_j, z_j через какие-то, заранее известные, функции от них: $\varphi(x_j); \psi(y_j)$ и т. д.

Все изложенное в предыдущей главе дает основание ответить и на этот вопрос утвердительно. В простейшем случае двух независимых переменных $u=f(x, y)$ система условных уравнений примет вид

$$\begin{aligned} \xi_0 + \xi_1 \varphi(x_j) + \xi_2 \psi(y_j) - u_j &= 0; \\ j &= 1, 2, 3, \dots, n. \end{aligned}$$

Соответствующая система из трех нормальных уравнений после перехода к тем же символам $a=1; \varphi(x)=b; \psi(y)=c$ будет иметь хорошо известный нам вид:

$$\left. \begin{aligned} [aa] \xi_0 + [ab] \xi_1 + [ac] \xi_2 - [al] &= 0 \\ [ba] \xi_0 + [bb] \xi_1 + [bc] \xi_2 - [bl] &= 0 \\ [ca] \xi_0 + [cb] \xi_1 + [cc] \xi_2 - [cl] &= 0 \end{aligned} \right\} \quad (215)$$

Поскольку далее мы не будем выходить за пределы зависимости одной измеренной величины u от двух измеренных величин x, y , то приведенной системой (215) в дальнейшем можно ограничиться. Поясним еще раз новые значения прежних символов a, b, c . Символ a по-прежнему обозначает единицу, следовательно, соответственно $[aa]=n$. Символ b означает функцию от x , тогда соответственно

$$[ab] = \overline{n\varphi(x)}; \quad [bb] = [\varphi(x)\varphi(x)] = n\overline{\varphi(x)^2}.$$

Точно так же символ c означает известную функцию от y , следовательно,

$$[ac] = n\overline{\varphi(x)\psi(y)}; \quad [bc] = [\varphi(x)\psi(y)]$$

и т. д.

В рассмотренном выше частном случае линейной зависимости u от x и y вместо $\varphi(x)$ записывалось просто x , а вместо $\psi(y)$ — просто y . Весь ход решения оставался тем же.

В заключение отметим, что переход от переменных x и y к каким-то функциям от них может иметь место лишь тогда, когда вид этих функций теоретически известен, что, вообще говоря, на практике встречается сравнительно редко. Но все же приведем для примера один из возможных подобных случаев.

Допустим, что есть основание предполагать линейную зависимость u от x и параболическую зависимость u от y . Основное уравнение запишется тогда в виде

$$u = A + Bx + Cy + Dy^2.$$

Вводя соответственные обозначения

$$a = 1; \quad b = x; \quad c = y; \quad d = y^2; \quad l = u;$$

$$\xi_0 = A; \quad \xi_1 = B; \quad \xi_2 = C; \quad \xi_3 = D,$$

получим, в конце концов, следующую систему из четырех нормальных уравнений:

$$\left. \begin{aligned} [aa] \xi_0 + [ab] \xi_1 + [ac] \xi_2 + [ad] \xi_3 - [al] &= 0 \\ [ba] \xi_0 + [bb] \xi_1 + [bc] \xi_2 + [bd] \xi_3 - [bl] &= 0 \\ [ca] \xi_0 + [cb] \xi_1 + [cc] \xi_2 + [cd] \xi_3 - [cl] &= 0 \\ [da] \xi_0 + [db] \xi_1 + [dc] \xi_2 + [dd] \xi_3 - [dl] &= 0 \end{aligned} \right\} \quad (216)$$

которая решается при помощи общего метода, изложенного в § 44.

Отметим для ясности, что изложенный в § 45 способ Чебышева для отыскания зависимостей от двух (и более) переменных неприменим, так как принцип ортогонализации относится лишь к многочленам с одной переменной.

§ 48. Семейства эмпирических кривых

При отыскании эмпирических зависимостей графические методы по крайней мере в начальной стадии этого процесса являются основными. Но даже и в следующей стадии, заключающейся в выяснении характера разброса эмпирических точек около намеченной кривой, они могут быть полезны, так как всегда легче увидеть отклонения на графике, чем определять их вычислением по способу наименьших квадратов.

При отыскании эмпирической зависимости между двумя измеренными величинами предварительное нанесение их на график не представляет никаких затруднений и весьма облегчает всю дальнейшую вычислительную работу, так как если разброс точек на графике оказывается даже на глаз слишком велик, то едва ли имеет смысл находить его численное выражение. Во

многих случаях исследователи, как правило, ограничиваются первой стадией и проводят искомую зависимость непосредственно на графике, не прибегая к трудоемкому способу наименьших квадратов.

Однако при отыскании связи между тремя величинами:

$$F(x, y, z) = 0 \quad (217)$$

изображать эту связь на плоском графике становится уже невозможным, а поэтому приходится применять один из двух возможных в таких случаях методов: метод построения семейства кривых или метод проведения изолиний.

Рассмотрим первый из приведенных методов, который удобно применять в тех случаях, когда при сравнительно большом числе измерений одна из трех переменных делится на несколько групп с осреднением внутри каждой группы. Пусть это будет переменная z . Принимаем ее осредненные значения за зависимые переменные:

$$z_k = F_k(x, y); \quad k = 1, 2, 3, \dots, m \quad (218)$$

(k — номер группы, m — число групп). Каждое осредненное значение приписывается всем точкам, для которых оно было осреднено.

Этот метод особенно удобен, когда значения третьей переменной заранее задаются условиями эксперимента (тогда и осреднение отпадает). Например, в случаях:

- 1) измерения скоростей потока при различных значениях глубины и уклона и при заданной шероховатости дна;
- 2) измерения скоростей фильтрации через грунты различной крупности при заданных температурах;
- 3) измерения скорости осаждения наносов разной крупности при заданной солености воды и т. д.

Во всех аналогичных случаях бывает удобно считать переменную параметром, а уравнение (218) записывать в виде

$$y = \Phi(x; \delta) \quad (219)$$

и искать его форму и значение входящих в него коэффициентов или графически, или по способу наименьших квадратов. Последний, конечно, применим также непосредственно к трем переменным, как показано в предыдущем параграфе, без выделения одной из них. Но можно с уверенностью сказать, что если вышеуказанные условия для такого выделения удовлетворены, то расчет с выделением параметра будет и легче и надежнее. В виде примера рассмотренной операции на рис. 27 приведен график зависимости коэффициента сопротивления λ от числа Рейнольдса Re при различных относительных шероховатостях $\frac{\delta}{h}$, полученной А. П. Зегждой (на графике даны их обратные значения — коэффициенты «гладкости» $\frac{h}{\delta}$). В исследовании

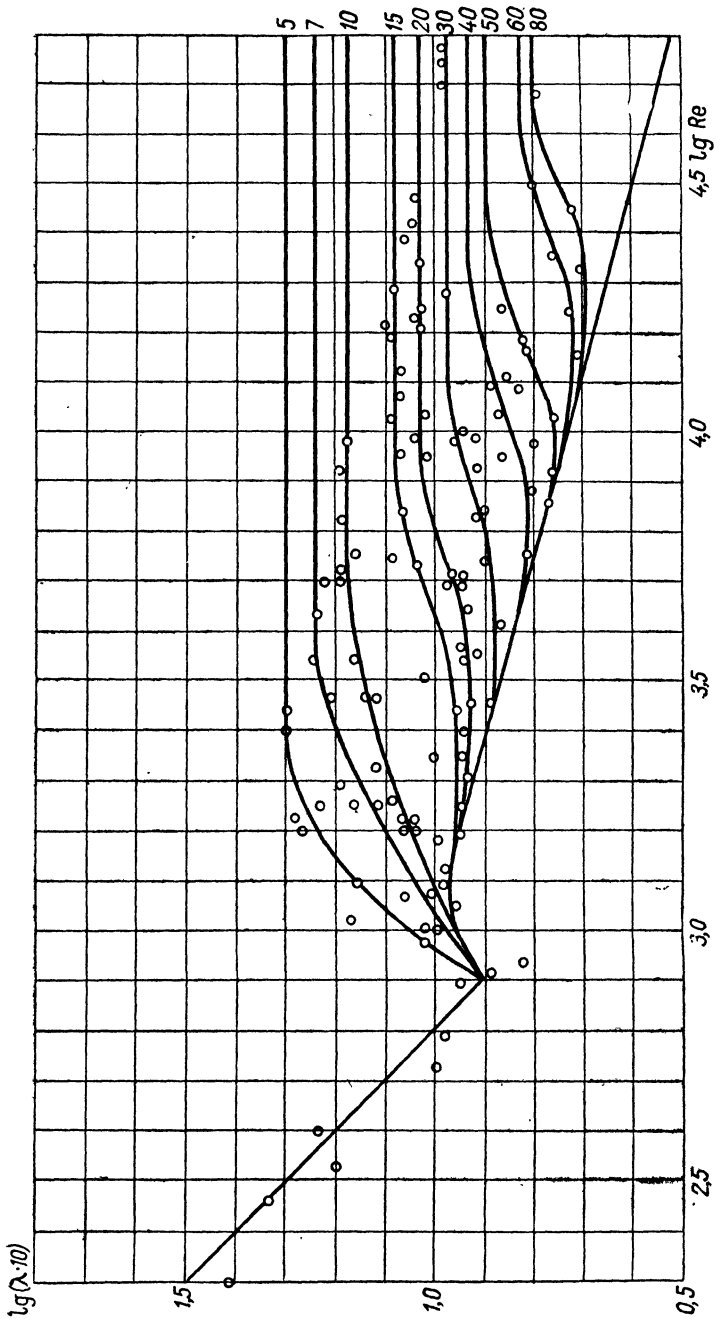


Рис. 27. График зависимости коэффициента сопротивления λ от числа Рейнольдса Re , построенный по опытам А. П. Зегжды.

Зегжды вся задача отчасти облегчалась тем, что форма зависимости для предельных случаев теоретически была известна: при очень малых $Re \lambda Re = \text{const}$, при очень больших $Re \lambda = \text{const}$. В общем случае изложенный выше метод также весьма удобен и широко применяется.

§ 49. Метод изолиний

Кроме метода, изложенного выше, существует и другой метод, заключающийся в нанесении системы точек в координатах x, y с указанием цифрой на каждой из них значения z . Эту систему можно взять за основу для проведения плавных линий равного значения переменной z , т. е. изолиний z . Случаев, когда этот метод оказывается особенно удобным, можно привести большое количество. Так, например, в гидравлике распределение скоростей по сечению канала (или реки) часто изображают в виде изотак. Но особенно широко используются разного рода изолинии в науках, изучающих распределение какого-либо признака или фактора на территории (или акватории). Изображения, полученные в виде изолиний, интерполирующих распределение данного признака в измеренных точках, здесь особенно удобны.

К таким изолиниям относятся горизонтали (изогипсы), применяемые в геодезии и топографии, изотермы и изобары в климатологии, изобаты в океанографии и т. д. Если измеряемая величина меняется во времени, как, например, температура, то получается целая серия карт: изотермы летнего и зимнего времени и т. д. В этом методе основную роль играет интерполяция на плоскости (на картах, охватывающих большие пространства, на проекциях сферической поверхности), так как само измерение величин в отдельных точках пространства ничем не отличается здесь от других измерений. Основные два вопроса, которые встают здесь перед исследователем, заключаются, во-первых, в том, как производить интерполяцию исходной дискретной системы измерений в точках данного фактора, и, во-вторых, при каком числе и каком расположении этих точек, применяя интерполяцию, получим наиболее близкое к истинному распределение интересующих нас величин по изучаемой территории.

Оба вопроса, как увидим далее, взаимно связаны и не имеют вполне удовлетворительного решения. Но на практике требуется хотя бы приближенное решение этих вопросов.

Представим себе поле точек с измеренными для них значениями величины z (рис. 28). Как провести между ними линии равного значения z ?

Самым простым предположением, из которого обычно и исходят, является предположение о линейности измерения вели-

чины между двумя близлежащими точками. Тогда, если это допущение близко к истине, между двумя точками можно найти одну изоточку (иногда более), в которой значения величины z легко вычисляются, обычно с округлением. Например, если в двух точках, расположенных близко, имеются измеренные температуры $7,2$ и $9,6^\circ$, то можно с полной уверенностью утверждать, что между ними где-то проходят изотермы 8 и 9° , пересекая соединяющую их прямую линию в как-то расположенных на ней изоточках. Справедливость этого вывода в общем виде не вызывает сомнения, но вопрос, заключающийся

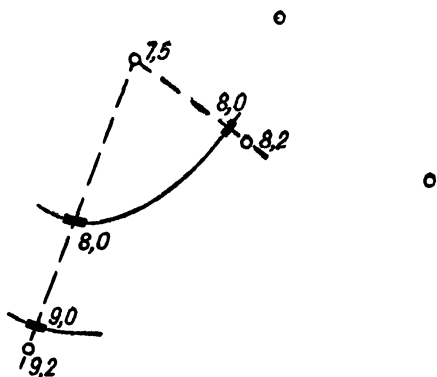


Рис. 28. Схема отыскания изоточек.

в том, расположены ли эти изоточки по линейному закону, остается открытым. Не имея, допустим, никакого представления о расположении изотерм, мы вынуждены искать изоточки между каждой парой точек исходного поля, как это изображено на рис. 28. Но такое допущение равносильно утверждению постоянства a и b в следующих равенствах:

$$\frac{\partial z}{\partial x} = a; \quad \frac{\partial z}{\partial y} = b,$$

что справедливо лишь в случае распределения z по уравнению

$$z = ax + by + c,$$

выражающему плоскость, проходящую, по крайней мере, между точкой A и всеми точками, расположенными близко от нее.

В общем случае такое допущение, конечно, неприемлемо. Но если исходные точки расположены достаточно густо, а поверхность распределения высоты z над плоскостью (x, y) имеет плавный характер, то такая замена хотя бы в порядке приближенности становится приемлемой. Переходим к рассмотрению второго из поставленных выше вопросов — какое число исход-

ных точек измерения для территории данного размера будет достаточным для проведения по ним изолиний, действительно отражающих истинный характер распределения рассматриваемого свойства.

Необходимо различать два случая: если мы хотим составить план местности в горизонталях (изогипсах), то от нас зависит, как расположить магистрали и поперечники нивелировки; также при мензульной высотомерной съемке от нас зависит, в каких местах и как часто будут расставлены реечники. Так как рельеф местности находится перед глазами топографа, то поставленный вопрос решается на месте им же самим. Но когда речь идет о составлении климатографической карты, то обычно имеется очень мало оснований для выбора как местоположения метеорологических станций, так и желательной густоты их сети. Поэтому размещение пунктов метеорологической сети на малоизученной территории производится обычно по аналогии с их фактическим расположением на соседних или на сходных территориях, где они работают уже несколько лет.

Вопрос здесь ставится следующим образом. Изучаемый фактор распределяется по данной неизученной территории, образуя некоторую, нам неизвестную поверхность $z(x, y)$. Сколько и как нужно расположить на этой территории измерительные пункты, чтобы по их данным охарактеризовать эту поверхность системой изолиний с достаточной полнотой?

Поставленный вопрос слабо освещен в науке, хотя частично он был затронут советским геофизиком Б. П. Вейнбергом [20], сформулировавшим следующий постулат: «Зависимость между функцией и аргументом можно в первом приближении считать известной, если выявлены все области, где эта функция достигает максимальных и минимальных значений, и если места и размеры этих экстремумов также известны».

Примем это требование как минимальное и добавим к нему другое, вполне естественное требование: разности в показаниях двух смежных пунктов не могут обуславливаться ошибками измерений, а должны иметь объективную ценность (без чего теряется даже смысл проведения изолиний!). Тогда грубоприближенное решение этого вопроса может быть дано на основе известного в теории вероятностей критерия Бьенэмэ. Этот критерий относится вообще к ряду последовательных во времени или вдоль какой-либо линии измерений. Он утверждает, что если число экстремумов в этой последовательности значительно меньше двух третей числа членов ряда, то имеется большая вероятность того, что этот ряд правильно описывает характер изменения (во времени или по длине) изучаемого фактора.

Очень простое доказательство приведенной теоремы было

дано французским математиком Бертраном. Это доказательство заключалось в том, что любой член такого ряда может быть или больше, или меньше своих соседей, или, наконец, его значение может находиться в промежутке между ними. Отсюда следует вывод, что в чисто случайном ряду вероятность для любого члена быть экстремумом равна двум третям. Если же число экстремумов значительно меньше двух третей числа членов, то маловероятно, чтобы ряд был случайным.

Для того чтобы применить этот критерий к плану какой-либо территории, покрытой сетью наблюдательных пунктов с нанесенной на каждом пункте цифрой полученного измерения, достаточно провести на этом плане ряд линий, соединяющих эти пункты (рис. 29), и выписать полученные таким путем ряды зна-

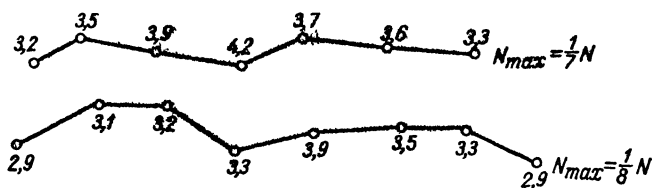


Рис. 29. Схема использования критерия Бьенэмэ.

чений. Если числа экстремумов этих рядов будут значительно меньше двух третей их членов, то по ним можно провести изолинии.

Мы остановились более подробно на менее исследованной задаче проведения изолиний того фактора, распределение которого по площади нам малоизвестно или даже совершенно неизвестно, и когда, кроме измерений величины этого фактора в отдельных более или менее произвольно выбранных точках, у нас ничего нет!

Случай, когда форма поверхности распределения этого фактора частично известна, хотя бы из теоретических соображений (например, при нанесении изотак для определения сечения водовода), или когда число исходных точек можно произвольно увеличивать (как, например, при нивелировке местности), как более простые, мы рассматривать не будем.

МЕТОД РАЗМЕРНОСТЕЙ

§ 50. Основные положения

Любая физическая зависимость при математическом (т. е. количественном) оформлении должна удовлетворять основному требованию — однородности в размерном отношении. Это значит, что все слагаемые указанной зависимости обязаны выражать физические величины одной и той же природы, т. е. быть или скоростями, или ускорениями, или количествами тепла и т. д. Это основное требование, очевидно, в одинаковой степени относится как к строгим физическим законам, так и к приближенным эмпирическим зависимостям. В первом случае требование удовлетворяется автоматически, так как при выводе физического закона из других, ранее известных законов возможность получить сумму физически неоднородных величин совершенно исключается! Но что касается эмпирических формул, то при выводе их из ряда измерений такая ошибка вполне возможна.

Следует отметить, что в литературе, посвященной эмпирическим наукам, ошибка в размерности часто встречается. Бывают даже случаи, когда она сохраняется долгое время, и к ней постепенно привыкают авторы работ и читатели. Достаточно указать на вопиющий пример — так называемую «формулу Шези» (для средней скорости потока), одну из основных в гидравлике, которая во всех работах записана в виде уравнения

$$u_{\text{cp}} = c \sqrt{RI},$$

в правой части которого R — гидравлический радиус, т. е. длина, I — уклон, т. е. отвлеченное число, а в левой же части стоит скорость! Замечательно то, что в той же гидравлике скорость потока в отверстии или на водосливе выражается правильно, через $\sqrt{2gH}$, что является величиной, имеющей размерность скорости. А в отношении скорости в трубах и каналах совершенно необходимая величина g почему-то не учитывается.

Такого же рода ошибки или другие подобные встречаются часто в эмпирических формулах, выводимых по данным измерений с помощью л о г а р и ф м и ч е с к и х а н а м о р ф о з (вопрос, который будет освещен в следующей главе). Примеров нарушения размерности у формул, широко используемых в практических целях, можно привести большое количество. По нашему мнению, пора уделить должное внимание вопросу о недопустимости использования таких формул для практических целей.

Например, один автор ставит в свою формулу расход в степени 1,02 и получает $Q^{1,02}$. Ясно, что эта лишняя дробь 0,02 получена путем «добросовестного» использования логарифмической анаморфозы! Но автор не задумался, во-первых, над тем, что ошибка его измерений гораздо больше, чем разница между Q и $Q^{1,02}$, во-вторых, что $Q^{1,02}$ — выражение, не имеющее смысла.

Другим автором выражена зависимость расхода наносов от скорости в виде

$$p = Au^n,$$

где n — показатель, значение которого „устанавливается экспериментом“. Какую же размерность в таком случае будет иметь коэффициент A ? Очевидно, переменную, что опять-таки лишено смысла!

Еще один автор ставит в формулу сумму

$$H + 0,14\Delta^{0,2}H^{0,6},$$

(где H — глубина, а Δ — высота выступов шероховатости) и суммирует длину с длиной в степени 0,8, что заведомо бессмысленно.

Мы ограничились немногими примерами, но их, к сожалению, большое количество как в советской, так и в зарубежной литературе по гидрологии, гидравлике и гидротехнике.

Следует установить непреложность принципа физической строгости не только для точных наук, в которых он в основном соблюдается, но и для неточных наук: пускай величины в них будут грубо приближенными, но структура величин должна быть безупречной! Можно было бы привести много примеров, когда структурная порочность приближенных зависимостей наносит вред при их техническом использовании, не говоря уже о том, что всякая ошибка в принципе размерностей вносит затруднения в технику расчетов.

§ 51. П-теорема

Значение принципа размерностей отнюдь не ограничивается тем, что он упорядочивает структуру математических оформлений эмпирических зависимостей. Во многих случаях принцип размерностей даже помогает отыскивать эти зависимости.

Такая эвристическая роль принципа размерностей вытекает из того, что он дает возможность путем перехода от размерных переменных величин, входящих в состав зависимости, перейти к безразмерным их комбинациям, называемым критериями, число которых всегда меньше на два или на три числа размерных переменных. Надо только знать заранее все те размерные величины, которые входят в отыскиваемую связь. Если какая-либо величина нами упущена или, наоборот, введена лишняя, то анализ размерностей или обнаружит нашу ошибку, или приведет к явно ошибочному результату.

Физические величины, рассматриваемые с точки зрения их размерности, могут быть первичными (иначе основными) и вторичными. Система первичных величин, или система единиц, с помощью которых можно выразить и измерить любую из интересующих нас величин, имеет несколько условный характер, но число этих единиц всегда равно трем, следовательно, из системы масса—длина—время можно исходить так же, как из системы масса—длина—скорость и из многих других мыслимых, но не представляющих практического интереса. Число единиц всегда будет равно трем. Существуют задачи, в которых одна из величин системы не рассматривается, например масса не участвует в задаче маятника, так же как она не участвует в задаче волны на поверхности жидкости, а время не входит в статические задачи теории упругости. В этих случаях на практике приходится оперировать лишь двумя основными величинами.

Система первичных величин должна удовлетворять единственному требованию, заключающемуся в том, чтобы все остальные, вторичные, величины могли быть выражены какими-либо произведениями первичных величин.

Предположим теперь, что мы ищем зависимость между какими-либо измеренными или измеримыми величинами. Запишем искомую зависимость символически в виде

$$\varphi(\alpha, \beta, \gamma, \dots) = 0.$$

Примем по предыдущему, что в скобках поставлена исчерпывающая совокупность взаимосвязанных величин. Будем считать, что их размерности нам известны. И, наконец, поставим одно естественное и тривиальное требование, чтобы от изменения размера первичных единиц (футы или метры, секунды или часы) форма искомой зависимости не изменялась. Уравнения, удовлетворяющие этому требованию, называются полными, и только к ним применим анализ размерностей.

По изложенному выше имеются три первичные величины, поэтому можем выразить размерность всех входящих в наше уравнение величин через эти три величины. Будем считать, что

величина α содержит первую из основных величин в степени α_1 , вторую — в степени α_2 , третью — в степени α_3 . Соответственно величина β имеет степени первичных величин $\beta_1, \beta_2, \beta_3$ и т. д.

Увеличим теперь размеры первичных величин: первой в x_1 , второй в x_2 , третьей в x_3 раз. Так как наше уравнение полное, то и в новых единицах оно будет тождественно прежнему. Однако если при переходе к новым единицам величину α заменим величиной

$$\alpha' = \alpha (x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3}),$$

а затем аналогично произведем замену остальных величин, то получим уравнение той же структуры, но с новыми единицами:

$$\varphi(\alpha' \beta' \gamma' \dots) = 0$$

или иначе

$$\varphi \{ \alpha x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3}; \beta x_1^{\beta_1} x_2^{\beta_2} x_3^{\beta_3}, \dots \} = 0.$$

Это уравнение справедливо, очевидно, при любых значениях множителей x_1, x_2, x_3 . Продифференцируем его относительно x_1 и после этого примем $x_1 = x_2 = x_3 = 1$. Будем иметь

$$\frac{\partial \varphi}{\partial \alpha} \alpha \alpha_1 + \frac{\partial \varphi}{\partial \beta} \beta \beta_1 + \dots = 0.$$

Если ввести новые переменные:

$$\alpha'' = \alpha^{\frac{1}{\alpha_1}}; \quad \beta'' = \beta^{\frac{1}{\beta_1}}; \quad \gamma'' = \gamma^{\frac{1}{\gamma_1}}$$

и т. д., то предыдущее уравнение примет вид

$$\alpha'' \frac{\partial \varphi}{\partial \alpha''} + \beta'' \frac{\partial \varphi}{\partial \beta''} + \gamma'' \frac{\partial \varphi}{\partial \gamma''} + \dots = 0.$$

Отметим, что новые переменные зависят лишь от степеней первой из основных (первичных) величин.

Пусть число величин, входящих в уравнение $\varphi = 0$, равно n . Обозначим последнюю из n величин α'', β'' и т. д. через ζ'' и разделим на нее все остальные величины: α'', β'', \dots . Получим $(n - 1)$ отношений, которые обозначим через z_1, z_2, z_3, \dots . После подстановки этих отношений в наше уравнение получим

$$\varphi(\alpha'', \beta'', \dots, \zeta'') \equiv \varphi(z_1 \zeta'', z_2 \zeta'', \dots).$$

Если взять частную производную по ζ'' от правой части, то, как нетрудно видеть, получим нуль. Иными словами, эта правая часть от ζ'' не зависит, следовательно, число величин, входящих в наше уравнение, сокращено на единицу. Но совершенно очевидно, что эту же операцию можно провести еще два раза, т. е. со второй и с третьей из первичных величин. Таким образом, в результате всех операций получим, во-первых, число величин, входящих в уравнение, уменьшенным на три и, во-вто-

рых, вместо тех размерных величин, которые числом n входили в наше исходное уравнение, будем теперь иметь $(n-3)$ безразмерных отношений, или критериев. В этом и заключается сущность II-теоремы, названной так потому, что все безразмерные отношения, или критерии, обычно обозначаются этой буквой. Следовательно, вместо размерного уравнения

$$\varphi(\alpha, \beta, \gamma, \dots) = 0$$

получим безразмерное, иначе критериальное, уравнение

$$\Phi\{\Pi_1, \Pi_2, \dots\} = 0. \quad (220)$$

§ 52. Составление критериев

Переход от размерных уравнений к безразмерным (критериальным) наиболее удобен как при обработке (а иногда и при организации эксперимента), так и в особенности при отыскании вида эмпирических зависимостей между измеренными величинами. Покажем это на нескольких примерах.

Предположим, что мы поставили задачу найти экспериментальным путем общий вид зависимости между следующими величинами: 1) скоростью (средней по сечению) потока, текущего в наклонном лотке, достаточно широком, чтобы влиянием боковых стенок можно было пренебречь; 2) глубиной потока, считая его продольно однородным (т. е. равномерным); 3) продольным компонентом действующей на него силы тяжести $g \sin \alpha = gi$ и 4) плотностью и вязкостью жидкости.

Прежде всего отметим, что размерность массы входит из всех приведенных величин только в плотность ρ и вязкость μ . Следовательно, эти величины могут входить в любой критерий лишь в форме их отношения

$$\frac{\mu}{\rho} = \nu,$$

носящего название кинематического коэффициента вязкости, размерность которого, легко получаемая из размерностей μ и ρ , равна

$$|\nu| = \frac{L^2}{T}.$$

Итак, ищем зависимость вида

$$F(\nu, h, gi, v) = 0. \quad (221)$$

Поскольку во все величины, стоящие в уравнении, масса не входит, то, согласно II-теореме, число величин, которые останутся в уравнении после приведения его к критериальному виду, будет на два меньше, т. е. равно двум.

Будем составлять все возможные безразмерные комбинации из четырех стоящих в уравнении (221) величин. Прежде всего

бросается в глаза, что из какой-либо пары этих величин нельзя составить никакую безразмерную величину. Следовательно, их надо комбинировать по три. Число возможных комбинаций из четырех элементов по три, согласно элементарной теории сочетаний, равно четырем. Обозначая через x , y , z неизвестные показатели (их значения во всех комбинациях различны), запишем

$$[v^x h^y \nu^z]; [v^x (gi)^y h^z]; [h^x ((gi)^y \nu^z)]; [v^x (gi)^y \nu^z].$$

Во всех четырех произведениях надо найти такие значения показателей, при которых эти комбинации станут безразмерными.

Напомним размерность всех множителей:

$$|v| = LT^{-1}; |h| = L; [gi] = LT^{-2}; [\nu] = L^2 T^{-1}.$$

Для первой комбинации условие безразмерности даст

$$x + y + 2z = 0; \quad -x - z = 0.$$

При двух уравнениях и трех неизвестных одно из неизвестных становится произвольным. Приняв $y=1$, получаем $x=1$, $z=-1$. Следовательно, первая комбинация дает критерий

$$\Pi_1 = \frac{vh}{\nu}.$$

Остальные три критерия получаем аналогичным путем

$$\Pi_2 = \frac{ghi}{\nu^2}; \quad \Pi_3 = \frac{gh^3i}{\nu^2}; \quad \Pi_4 = \frac{\nu^3}{\nu gi}.$$

Сразу убеждаемся, что последние два числа получаются из первых двух простым их перемножением (в чем и выражается сущность П-теоремы)

$$\Pi_3 = \Pi_2 \Pi_1^2, \quad \Pi_4 = \Pi_1 : \Pi_2.$$

Следовательно, имеем только два независимых критерия, хорошо известных в гидродинамике и гидравлике. Первым из них является число Рейнольдса Re , вторым — коэффициент сопротивления λ :

$$Re = \frac{vh}{\nu}; \quad \lambda = \frac{ghi}{\nu^2}. \quad (222)$$

Таким образом, нами получен исключительно важный результат, позволяющий непосредственно по данным измерений четырех величин (v , h , gi , ν) определять графически или иным путем зависимость вида

$$\lambda = F(Re). \quad (223)$$

Если бы мы захотели ввести в рассмотрение фактор шероховатости, то к ранее полученным критериям должны были бы

добавить еще отношение двух линейных размеров: глубины потока и высоты выступов шероховатости $\frac{\delta}{h}$ ¹. Тогда для трех критериев получили бы уже не кривую, а семейство кривых:

$$\lambda = F\left(\text{Re}; \frac{\delta}{h}\right). \quad (224)$$

Такое семейство кривых, изображенное на рис. 27, было получено в результате серии экспериментов А. П. Зегжды (см. § 47).

В двух только что приведенных примерах мы получили два и три критерия², зависимость между которыми определялась графическим путем. В других случаях эта зависимость может отыскиваться способом наименьших квадратов, а иногда с применением логарифмической анаморфозы. Но все аналогичные примеры охватывают область отыскания по данным измерений эмпирических (всегда тем самым и приближенных) зависимостей между измеряемыми величинами.

Хотя данная книга посвящена именно этому вопросу, однако в целях более полного освещения всех возможностей, представляемых II-теоремой, приведем два примера, когда с помощью того же анализа размерностей получается также строгая физическая зависимость. Это бывает тогда, и только тогда, когда решение задачи приводит к образованию одного лишь критерия. Такое уравнение, очевидно, может быть решено в отношении этого критерия, и мы получим простое равенство:

$$\Pi = \text{const.}$$

Следует отметить, что значение этой постоянной не может быть получено с помощью принципа размерностей. Оно определяется или экспериментом, или путем теоретических построений. Приведем два примера.

В первом примере ищем зависимость между длиной маятника и периодом его качания. Простое рассуждение показывает, что к двум переменным величинам необходимо добавить еще одну, которая в каждом конкретном случае является постоянной (хотя, вообще говоря, для поверхности земного шара незначительно изменяется), это ускорение силы тяжести.

Из величин l , T , g можно, очевидно, составить лишь один критерий:

$$\Pi = l^x T^y g^z.$$

¹ Оно носит название критерия Мизеса.

² Критерии, представляющие отношения двух величин одинаковой размерности, иногда называют симплексами.

Имеем два условия безразмерности:

$$x + y = 0;$$

$$y - 2z = 0,$$

откуда, принимая $y=1$, получаем

$$z = \frac{1}{2}; \quad x = -\frac{1}{2}.$$

Следовательно, искомый критерий равен

$$\Pi = \frac{T\sqrt{g}}{\sqrt{l}} = \text{const},$$

или иначе

$$T = \text{const} \sqrt{\frac{l}{g}}. \quad (225)$$

Если бы мы решали эту же задачу с помощью интегрирования дифференциальных уравнений механики, то получили бы и значение постоянной, равное π .

Второй пример заимствуем у Бриджмена [21, стр. 74—76]. Определяем давление газа при заданной температуре и числе атомов в единице объема. Простые соображения по кинетической теории газов приводят к введению в расчет трех физических постоянных: массы атома m , абсолютной температуры θ , газовой постоянной k .

Находим значения показателей, обращающих произведение

$$pm^x N^y \theta^z k^t$$

в безразмерное, в котором показатель при p принимаем равным единице, потому что мы ищем именно зависимость p от остальных величин. Так как какой-либо из показателей нам все равно придется принять за единицу, то, очевидно, удобнее принять за единицу именно упомянутый показатель.

Размерность давления равна размерности силы, отнесенной к площади, т. е. $ML^{-1}T^{-2}$. Размерность температуры равна размерности кинетической энергии молекул, т. е. ML^2T^{-2} . Размерность N числа молекул в единице объема равна L^{-3} и, наконец, газовая постоянная является безразмерной величиной.

Условия безразмерности запишутся в следующем виде:

$$1 + x + z = 0;$$

$$-1 - 3y + 2z = 0;$$

$$-2 - 2z = 0,$$

откуда получаем $x=0$, $y=-1$, $z=-1$. Искомый критерий

$$\Pi = \frac{p}{Nk\theta} = \text{const},$$

а искомая зависимость

$$p = \text{const} Nk\theta. \quad (226)$$

Отметим, что условно введенная нами вначале масса атома после всех операций с размерностями оказалась исключенной из искомой зависимости, т. е. получилось, что давление газа от массы молекул не зависит. Это положение, известное в кинетической теории газов, выявлено здесь в результате анализа размерностей, при помощи которого во многих случаях (но не всегда) можно действительно установить, зависит ли данная величина от какой-то другой или нет.

Рассмотрим теперь пример, относящийся к тем неточным наукам, которым посвящена настоящая книга, а именно к гидрологии.

В. М. Лохтин, один из основоположников учения о русле рек, высказал в качестве постулата, что размеры и форма речного русла в среднем для значительного и однородного участка реки зависят от трех факторов: водоносности реки, ее уклона и крупности тех наносов, которые она способна передвигать. Попробуем представить совокупность трех основных факторов в виде критерия.

Обозначая руслоформирующий расход через q , продольный компонент силы тяжести, пропорциональный уклону, через gi и, наконец, средний диаметр руслообразующих наносов через d , получим, что произведение $q^x (gi)^y d^z$ будет безразмерным лишь в случае, когда $x=1$, $y=-\frac{1}{2}$, $z=-\frac{5}{2}$ (напоминаем, что один из показателей должен быть принят равным единице).

Таким образом, мы получили критерий Лохтина в виде

$$Л = \frac{q}{d^2 \sqrt{gdi}}. \quad (227)$$

Однако можно бы несколько видоизменить наши основные предпосылки, исходя из того, что подвижность наносов связана с действием на них силы тяжести. Известно, например, что наименьшая скорость течения, при которой наносы данного диаметра сдвигаются с места, пропорциональна квадратному корню из произведения gd . Введем соответственно в исходное произведение вместо d произведение gd . Не приводя простых и аналогичных выкладок, хорошо известных из предыдущих примеров, получим несколько иное выражение для критерия Лохтина:

$$Л = \frac{qi^2}{d^2 \sqrt{gd}}. \quad (228)$$

Оказалось, что второй вариант искомого критерия отличается от первого лишь добавлением безразмерного множителя $i^{\frac{5}{2}}$, который также является критерием (или симплексом), поскольку представляет собой отношение падения уровня реки к длине ее участка.

Нами приведен пример образования критерия, вытекающего из постулата Лохтина. Рассматривать же общий вопрос (находящийся еще в процессе разработки) о совокупности всех тех критериев, которые дадут полное освещение задачи руслового процесса, мы здесь не будем. Достаточно сказать, что уже установлено наличие зависимостей между этим критерием и некоторыми симплексами $\frac{b}{a}$ и $\frac{h}{a}$. Следовательно, его роль и место в теории руслового процесса установлены именно путем анализа размерностей.

Для полного охвата всего значения этого принципа надо еще учесть, что число всех возможных критериев в какой-либо конкретной задаче зависит только от числа входящих в нее размерных величин. Это число всегда меньше последнего на число основных единиц.

Те же безразмерные комбинации, которые содержат в себе лишь постоянные размерные величины, в решении критериальных уравнений роли не играют (кроме удобства сокращения числа символов).

Отметим еще, что число размерных постоянных в исходном размерном уравнении не может быть больше числа физических переменных величин. Это следует из того, что каждое размерное уравнение можно сделать полным, если каждой физической переменной величине придать соответствующую размерную постоянную. Исходя из этих соображений, рассмотрим возможное соотношение между числом размерных величин и числом критериев.

Если полное число всех физических размерных величин (и постоянных и переменных) равно числу основных единиц, то, согласно П-теореме, число критериев будет равно нулю, т. е. никакой связи между величинами не существует, и никакого критериального уравнения построить для данной задачи как-будто бы нельзя! Однако это немыслимо. Следовательно, в постановку задачи вкралась какая-то ошибка: нами забыта какая-то физическая величина, которая также входит в искомую зависимость!

Если число физических размерных величин на единицу больше числа основных (первичных) величин, то получается лишь один критерий. Но уравнение с одним неизвестным имеет определенное решение $\Pi = \text{const}$. Этот случай соответствует строгой физической закономерности, а поэтому в данной книге не излагается.

При двух переменных критериях имеем кривую на плоскости, при трех — семейство кривых (или систему изолиний). Свыше трех переменных критериев обычно не встречается, и та задача, в которой это имело бы место, может считаться практически неразрешимой. Но так как, повторяем, число критериев меньше числа размерных величин на число первичных единиц, которое

равно двум или трем, то, если число физических величин в исходном размерном уравнении составляет пять или шесть, тогда из них не меньше трех являются переменными размерными величинами.

Перейдем к рассмотрению процесса составления критериальных уравнений в том случае, когда никакого исходного размерного уравнения нет, и мы непосредственно от обработки проведенных измерений переходим к составлению эмпирических зависимостей сразу же в критериальной форме.

Имеем совокупность измерений величин, из которых некоторые сохраняли постоянное значение, а другие изменялись экспериментатором по какому-то плану. Наконец, некоторые величины берутся непосредственно из справочников, без измерения их в данной серии опытов. Факторы, сохранившие постоянное значение в данной серии опытов, но подвергшиеся изменению, входят в состав критериев как средние значения. Факторы, подвергшиеся изменению по плану эксперимента или сами изменявшиеся под влиянием учитываемых причин, включаются в состав переменных критериев всеми своими значениями. В результате получается два или три переменных критерия, связанные между собой определенными зависимостями. Анализ размерностей сокращает в данном случае работу экспериментатора, уменьшая число тех переменных, между которыми ему надлежит отыскивать зависимость. И, кроме того, такой анализ подводит экспериментатора к тем комбинированным величинам (критериям), зависимость между которыми охватывает большее число факторов в более отчетливой форме.

Приведенный пример опытов А. П. Зегжды (см. рис. 27) является для изложенного положения хорошей иллюстрацией: вместо ряда зависимостей между многими величинами мы имеем на основе применения принципа размерностей четкую связь между критериями λ и Re при различных значениях третьего $\frac{\delta}{h}$. Указанный третий фактор в данном случае имеет заданные по условиям опыта значения. Однако это совсем не обязательно, так как можно легко представить себе отыскание такой зависимости между тремя величинами, когда ни одна из них не задана условиями опыта, а все три измеряются в каждом опыте. Тогда может возникнуть вопрос, какую из трех величин удобнее принять за параметр семейства кривых. Общего ответа на этот вопрос дать нельзя, так как решается он в каждом отдельном случае по интуиции.

ЛОГАРИФМИЧЕСКАЯ АНАМОРФОЗА

§ 53. Общие свойства логарифмических изображений

Основное свойство процесса логарифмирования, заключающееся в превращении произведения в сумму, иногда может быть удачно использовано при отыскании формы эмпирической зависимости, особенно между двумя переменными, когда графическое ее изображение легко воспроизводимо.

Например, если есть основание предполагать, что кривая вида

$$y = Ax^n \quad (229)$$

или вида

$$y = Ba^x \quad (230)$$

может при надлежащем подборе параметров расположиться вблизи экспериментальных точек, то весьма удобно представить эти кривые соответственно в форме

$$\lg y = \lg A + n \lg x;$$

$$\lg y = \lg B + x \lg a$$

и представить в виде прямых. Первую прямую можно нанести в логарифмическом масштабе, вторую — в полулогарифмическом. В часто встречающихся случаях, когда при логарифмировании не меняется знак переменной, иными словами, когда логарифмируемая переменная больше единицы, применение логарифмической анаморфозы не встречает никаких практических затруднений, а применение логарифмической или полулогарифмической клетчатки делает всю операцию перехода совсем легкой.

Но в тех случаях, когда логарифмируемая переменная частично (или полностью) принимает значения меньше или даже гораздо меньше единицы, вся операция логарифмирования измеренных величин осложняется.

Трудность возникает здесь в том, что короткий интервал основного ряда $0-1$ при переходе к логарифмическому изображению вырастает в бесконечный интервал. Иными словами, сразу получаем для существенно положительной величины диапазон возможных значений от $-\infty$ до $+\infty$.

Когда имеем дело с величинами, измеряемыми настолько точно, что их ошибками можно практически не интересоваться, то операция логарифмирования, по существу, ничего не изменяет в математической форме зависимости между двумя (и более) величинами. Но когда эти ошибки велики и могут влиять (и даже сильно) на форму отыскиваемой зависимости, то логарифмическая анаморфоза может значительно исказить все соотношение между размером самой измеряемой величины и размерами ошибок ее измерения.

Сравним логарифмическую обработку измерений «больших» и «малых» величин, как они были определены в главах VII и VIII.

Имеем ряд измерений величины, истинное значение которой равно $x_1, x_2, \dots, x_j, \dots, x_n$. Будем считать все систематические и все высказывающиеся ошибки исключенными, а ошибки отсчета, связанные с порогом чувствительности прибора, учтем с помощью соответственного ограничения числа знаков.

Оговариваем, что в изложении этого вопроса будем пользоваться натуральными логарифмами, как более удобными. Имеем

$$\ln x_j = \ln (\bar{x} + x'_j) = \ln \bar{x} + \ln \left(1 + \frac{x'_j}{\bar{x}} \right).$$

В случае «больших» величин, т. е. при $x'_j \ll \bar{x}$, правая часть преобразуется в следующий вид:

$$\ln x_j = \ln \bar{x} + \frac{x'_j}{\bar{x}},$$

или иначе

$$\ln x_j - \ln \bar{x} = \frac{x'_j}{\bar{x}}. \quad (231)$$

Однако если имеем для «больших» величин симметрию действительных отклонений x'_j , то, согласно равенству (231), отклонения $\ln x_j$ от $\ln \bar{x}$ также будут симметричными.

Это очень важный вывод, дающий нам право при логарифмировании зависимостей между «большими» величинами пользоваться как нормальным законом Гаусса, так и способом наименьших квадратов.

Совершенно иное получается для «малых» величин, когда усиленное неравенство $x_j \ll \bar{x}$ теряет смысл, а следовательно, от-

падает и равенство (231). Тогда, возвращаясь к общему равенству

$$\ln x_j - \ln \bar{x} = \ln \left(1 + \frac{x'_j}{\bar{x}} \right),$$

в правой части которого дробь $\frac{x'_j}{\bar{x}}$ уже приобретает порядок единицы, а потому приближенное равенство, справедливое для малых значений ξ ,

$$\ln(1 + \xi) = \xi$$

теряет смысл, мы вынуждены искать другой путь для обработки измерений. Но ведь тогда и измерения, как показано в главе VIII, могут дать отклонения, значительно отходящие от симметрии. Распределение отклонений должно уже следовать асимметричному закону, принятому нами ранее как гамма-распределение. В этом случае возникает вопрос, как найти закон распределения логарифма той величины, которая сама подчиняется гамма-распределению.

Следуя общему правилу перехода от одной плотности вероятности к другой, запишем

$$\frac{x^{\alpha e - x}}{\Gamma(\alpha + 1)} dx = p(\xi) d\xi,$$

где $\xi = \ln x$, а следовательно, $\frac{dx}{d\xi} = e^\xi$.

Получаем после простейших преобразований

$$p(\xi) = \frac{e^{(\alpha + 1)\xi} - e^\xi}{\Gamma(\alpha + 1)}. \quad (232)$$

Это распределение так же асимметрично, как исходное гамма-распределение, но в то же время подобно нормальному распределению уходит обоими концами в бесконечности.

Легко показать, что полученное распределение не имеет моментов выше нулевого, которые нам сейчас не нужны, так как достаточно знать медиану распределения. Последнюю очень легко определить, если опереться на известное ее свойство (см. § 10). Поскольку логарифм переменной является ее монотонной функцией, то

$$\xi_0 = \ln x_0.$$

В главе VIII показано, что для обработки данных измерений, подчиняющихся гамма-распределению, нет надобности пользоваться всеми значениями последнего, а достаточно для каждого значения параметра α знать медиану. Кроме того, может представлять интерес сравнение теоретического значения первого абсолютного момента с его эмпирическим значением.

В табл. 11 приведены значения медиан распределения (232), полученные с помощью простого соотношения

$$\xi_0 = \ln x_0.$$

Но для того чтобы читатель получил наглядное представление о виде кривой плотности вероятности логарифмов величин, подчиненных гамма-распределителю (232), на рис. 30 приводим сравнение распределений x и $\ln x$ для одного частного значения параметра $\alpha=1$. На графике (рис. 30) видно, что мода этого распределения (при $\alpha=1$) расположена на абсциссе $+0,693$, от которой плотности вероятностей, уменьшаясь в обе стороны,

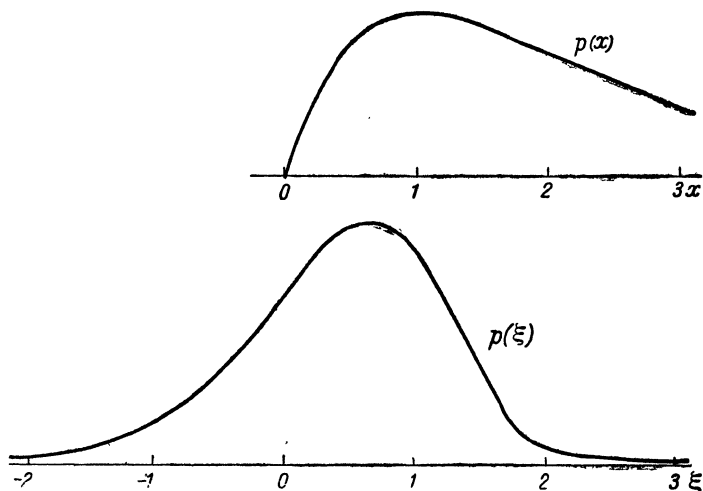


Рис. 30. График распределения x и $\xi = \ln x$.

асимптотически приближаются к нулю в обеих бесконечностях и явно отклоняются от симметрии. Что касается медианы, то при данном значении параметра (и при всех высших) она располо-

Таблица 11

$\alpha + 1$	ξ_0	$\alpha + 1$	ξ_0	$\alpha + 1$	ξ_0	$\alpha + 1$	ξ_0
0,5	-1,465	3,0	0,983	8,0	2,037	13,0	2,539
1,0	-0,365	4,0	1,301	9,0	2,160	14,0	2,615
1,5	+0,169	5,0	1,541	10,0	2,269	15,0	2,686
2,0	0,518	6,0	1,735	11,0	2,367	16,0	2,752
2,5	0,777	7,0	1,897	12,0	2,457		

жена в положительной области (из данных таблицы видно, что ее положительные значения начинаются с $\alpha = \frac{1}{2}$).

Используем далее приведенный график для того, чтобы ясно представить, как влияет логарифмирование измеренных величин на их рассеяние.

Предположим, что измеряем некоторую весьма малую, но существенно положительную величину, например концентрацию наносов в верхних слоях потока (в области «потолка» наносов). Именно такие измерения были произведены в 1954 г. на кафедре физики руслового потока МГУ под руководством автора книги. Измерения производились дважды: 1) высокочастотным кинематографом и 2) фотоэлектрическим методом. Коснемся лишь измерений, произведенных кинематографом, и рассмотрим, например, запись кинематографа в фильме № 4. Высота потока при измерениях кинематографом была разделена на восемь зон, каждая высотой 1 мм.

В табл. 12 представлена последовательность чисел заснятых песчинок, пересчитанная в вариационный ряд, в самой верхней, восьмой, зоне.

Таблица 12

Число песчинок	0	1	2	3	4	5	6	7	8
Его частость	60	6	20	5	4	1	1	2	1

Совершенно очевидно, что при столь малом числе песчинок их целочисленная регистрация должна была подчиняться закону Пуассона, для которого, как известно, среднее равно второму центральному моменту. Из табл. 12 путем простых операций получаем среднее, равное 1,1, и медиану, равную нулю.

При положительной асимметрии медиана всегда меньше среднего. В данном же случае при целочисленной регистрации меньше единицы может быть только нуль, а поэтому ясно, что переход к логарифмической анаморфозе вообще не имеет смысла, так как медиана распределения уходит в минус бесконечность.

В седьмой зоне того же фильма концентрация была несколько больше, чем в восьмой зоне. Здесь наблюдалось следующее распределение по частотам (табл. 13)

Таблица 13

Числа	0	1	2	3	4	5	6	7	8	9	10
Частость	20	14	18	9	21	8	3	2	3	1	1

В результате обработки измерений получили среднее, равное 2,8, медиану, равную 2,0.

Переход к логарифмической анаморфозе в седьмой зоне,

вообще говоря, не исключается. Но неизвестно, как такой переход отразится на рассеянии измерений.

Если отсчитывать все отклонения от среднего, то для восьмой зоны большая их часть даст отрицательную бесконечность в логарифмической анаморфозе и очень малые величины в обычном изображении. Если же отсчитывать отклонения от медианы, то в логарифмической анаморфозе часть их даст неопределенность вида $\infty - \infty$, а остальные выразятся положительной бесконечностью.

В седьмой же зоне даже после логарифмической анаморфозы сохранится конечная медиана. Некоторые отклонения от этой медианы все же дадут так же отрицательную бесконечность, как и отклонения от среднего.

Приведенный пример представляется в известном смысле крайним, так как измерения, дающие нуль, а следовательно, и отрицательную бесконечность после логарифмирования, могут иметь место лишь при целочисленной регистрации, т. е. в дискретных измерениях, подчиняющихся закону Пуассона. Но и в непрерывной регистрации измерения «малых», существенно положительных величин размеры отклонений от очень малых значений медианы или среднего могут быть также очень малы и, следовательно, при переходе к их логарифмам могут дать очень большие (по модулю) значения как самого центра, так и в особенности отклонений от него. В результате этого картина рассеяния весьма исказится.

Логарифмирование эмпирической зависимости весьма удобно в тех случаях, когда эта зависимость является одночленной и имеет степенной, или показательный, вид. После ее логарифмирования получаем линейную зависимость, а линейность расположения или самих измеренных величин, или хотя бы каких-либо монотонных функций от этих величин чрезвычайно облегчает все дальнейшие операции по нахождению параметров этих зависимостей. Это весьма убедительно показано при изложении способа наименьших квадратов. Но практически встречаются случаи, когда процесс логарифмирования создает как раз обратный эффект, превращая очень простую линейную зависимость в нелинейную. Например, проще всего записать

$$y = a + bx,$$

но если мы обе части этого равенства прологарифмируем, то между $\lg x$ и $\lg y$ получится криволинейная зависимость, обращающаяся в линейную только в одном случае, когда $a=0$. Могут нам задать вопрос, зачем же логарифмировать ту зависимость, которая и сама по себе линейна. Конечно, незачем. Однако если эта сумма возведена в степень, которую можем определить лишь на основе данных измерений

$$y = (a + bx)^m,$$

то возникает трудность, непреодолимая для способа наименьших квадратов, при помощи которого, как известно, нельзя определить показатели. Для того чтобы превратить показатель в коэффициент, надо логарифмировать в данном случае сумму, в которую входит параметр, также нуждающийся в определении из тех же измерений.

В целях формального упрощения возникшей перед нами задачи заметим, что один из параметров, стоящих в скобках, можем отбросить на основе тождества

$$\lg(a + bx)^m = m \lg b + m \lg \left(x + \frac{a}{b}\right).$$

Поэтому в дальнейшем будем записывать наше равенство в виде

$$y = (a + x)^m \quad (233)$$

и соответственно

$$\lg y = m \lg(a + x). \quad (233')$$

Задача линеаризации уравнения (233) возникла сначала в гидрометрии в связи с построением эмпирических кривых зависимости расхода воды в реке от ее горизонта:

$$Q = F(H).$$

Здесь Q — совершенно определенная и измеримая величина, а H — условная величина, зависящая от того, какой горизонт принимается за нуль отсчета. Если задаться нулем отсчета так, чтобы он соответствовал нулю расхода (что в большей части возможно), то искомая зависимость обычно (при правильном русле без поймы) может быть представлена приближенно в конечном виде:

$$Q = AH^m.$$

После логарифмирования снова получаем линейное уравнение, доступное для способа наименьших квадратов. Но во многих случаях отыскание такого нуля отсчета для H или невозможно, или неудобно. В этом случае приходится использовать уравнение вида

$$Q = a(b + H)^m,$$

непосредственно неразрешимое.

Ниже в этом же параграфе нами будет изложен простой и приближенный метод линеаризации последнего уравнения, предложенный В. Г. Глушковым, а следующий параграф посвятим более строгому анализу того же вопроса, имеющего, кстати сказать, гораздо более общее значение, чем отыскание формы кривых расхода. Но сначала установим некоторые, вполне строгие и бесспорные положения, от которых затем перейдем к приближенному методу Глушкова.

Теорема. Если между измеренными величинами x и y существует зависимость вида (233)

$$y = (a + x)^m,$$

после логарифмирования дающая линейную связь между $\lg y$ и $\lg(a+x)$ при определенном значении $a = a_0$, то при другой, большей или меньшей, величине a зависимость между ними даст кривизну, обусловленную разностью $a - a_0 = c$.

Обозначим $a + x = z$ и найдем вторую производную от логарифмированной зависимости вида $y = (z + c)^m$. Имеем¹

$$\ln y = m \ln(z + c);$$
$$\frac{d^2 \ln y}{d(\ln z)^2} = \frac{mcz}{(z + c)^2}.$$

Следовательно, лишь при $c = 0$ логарифмированная зависимость становится линейной, а при $c \neq 0$ знак кривизны совпадает со знаком c .

Это положение вполне строгое, но, повторяем, лишь при условии, что зависимость (233) между переменными x и y действительно существует. Но так как, строго говоря, в применении к зависимости расхода от горизонта оно выполняется в лучшем случае лишь с приближением, иногда даже весьма грубым, то и сам метод имеет, конечно, лишь приближенный характер. Впрочем, к крупным ошибкам применение этого метода не может привести, так как при графическом построении мы от них застрахованы.

Чтобы наглядно проиллюстрировать изменение кривизны логарифмированной зависимости при изменении величины параметра b , приводим на рис. 31 график, полученный Глушковым для реки Средне-Русской возвышенности.² На этом графике совершенно отчетливо видно выпрямление кривой с повышением значения параметра b от $-0,3$ до $+0,23$.

Возьмем на первоначально полученной кривой три произвольно выбранные точки. Если найдем такое значение добавка b , при котором в деформированной кривой эти три точки расположатся на одной прямой, то можем считать, что и вся кривая будет близка к прямой. Такой вывод является, конечно, весьма приближенным, но для пологих кривых, встречающихся на практике, он оправдывается.

Введем для приближенной зависимости

$$\lg Q = \lg A + m \lg(a + x)$$

¹ Здесь удобнее брать натуральные логарифмы.

² Мы не сочли нужным сделать пересчет графика, построенного по данным 1915 г., в метрические меры.

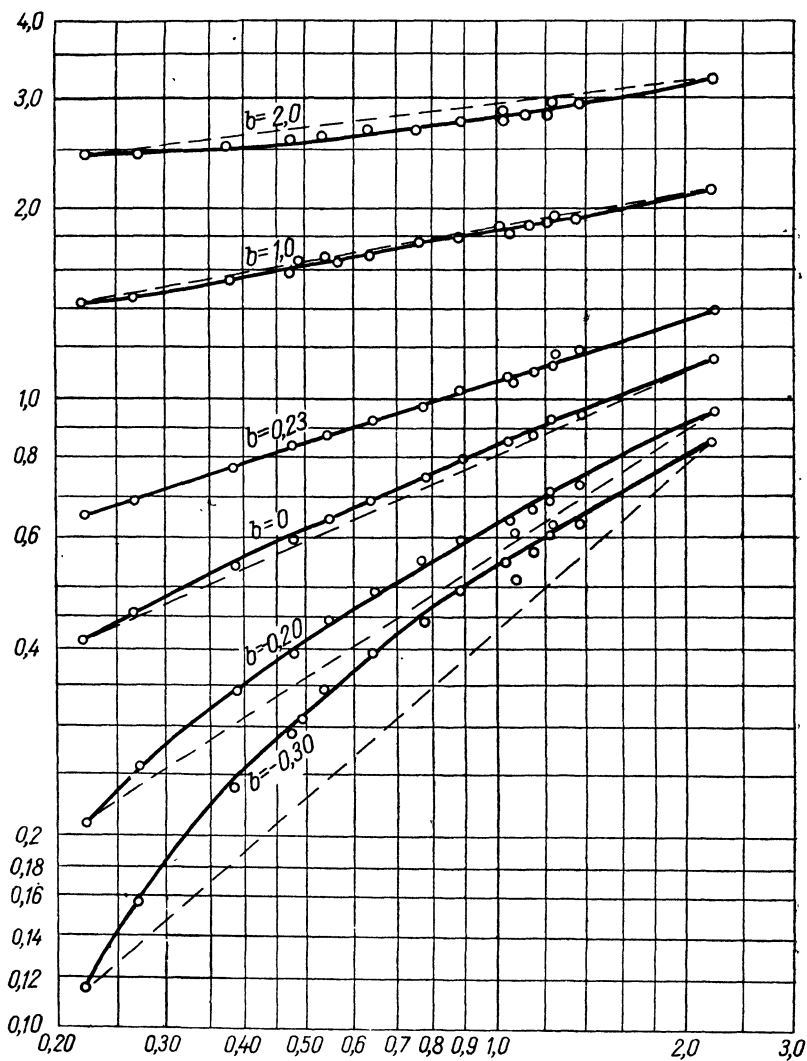


Рис. 31. Изменение кривизны логарифмики суммы в зависимости от параметра b .

сокращенные обозначения

$$X = \lg(a + x); \quad Y = \lg y$$

и запишем условие, при котором три точки: X_1Y_1 , X_2Y_2 , X_3Y_3 расположатся на прямой:

$$\frac{Y_1 - Y_2}{Y_2 - Y_3} = \frac{X_1 - X_2}{X_2 - X_3}.$$

Поскольку приведенное условие является единственным, то мы вправе приравнять оба равенства к любому числу, например к единице. Подставляя вместо X и Y их значения, запишем

$$\lg \frac{Q_1}{Q_2} = \lg \frac{Q_2}{Q_3};$$

$$\frac{\lg(H_1 + a)}{\lg(H_2 + a)} = \frac{\lg(H_2 + a)}{\lg(H_3 + a)}.$$

Из первого равенства следует

$$Q_2 = \sqrt{Q_1 Q_3},$$

т. е. ордината средней точки должна быть средней геометрической двух крайних точек (в логарифмическом масштабе это даст среднюю точку).

Второе отношение дает значение самого добавка:

$$a = \frac{H_2^2 - H_1 H_3}{(H_1 + H_3) - 2 H_2}.$$

Дальнейший ход решения очень прост. Вычисляем значение a из трех наблюдаемых горизонтов (H_1 , H_2 , H_3), добавляем (или вычитаем сообразно знаку) к ним величину a , а затем в точках $\lg(H_1 + a)$, $\lg(H_2 + a)$, $\lg(H_3 + a)$ воздвигаем ординаты:

$$Y_1 = \lg Q_1; \quad Y_2 = \lg Q_2; \quad Y_3 = \lg Q_3.$$

Согласно сказанному, точка ординаты Y_2 должна располагаться на прямой посередине между точками Y_1 и Y_3 , но с какой точностью это условие выполнится, сказать заранее трудно. Дело в том, что, во-первых, все величины H и Q измерены с ошибками, а, во-вторых, весь метод построен на предположении (это и является главным), что существует такое значение добавка a , при котором логарифмированная зависимость

$$\lg Q = \lg A + m \lg(H + a)$$

есть прямая линия. По-видимому, это бывает часто, но, конечно, не всегда, а при пойменном сечении реки даже невозможно, чтобы это было.

На рис. 32 приведен график, построенный Глушковым для той же реки, что и на рис. 31. Здесь спрямление зависимости расходов от уровней может считаться весьма удовлетворительным. Но в тех случаях, когда таким путем спрямления достичь

не удастся, это доказывает только ошибочность исходного предположения о существовании для данного створа реки зависимости вида $y = (a+x)^m$.

§ 54. Логарифмики сложения и вычитания

Весьма строгий анализ поведения кривых, получающихся при логарифмировании суммы (или разности), был сделан А. А. Саткевичем [25]. И хотя разработанный им метод имеет сравнительно ограниченную область применения, мы здесь сочли полезным дать о нем читателю ясное представление, по возмож-

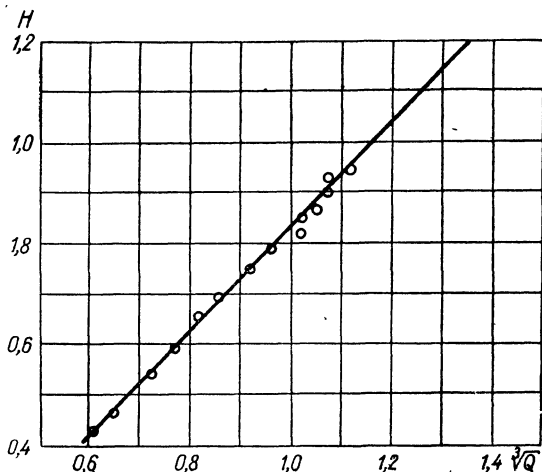


Рис. 32. Спрявление кривой по методу В. Г. Глушкова.

ности сократив и упростив его изложение.

Из предыдущего параграфа известно, что изменение величины параметра b в уравнении

$$\lg y = \lg (b + x)$$

влияет на кривизну этой кривой в переменных $\lg x$ и $\lg y$. Но так как это изменение явно связано с переносом начала отсчета переменной x , то естественно построить весь анализ именно на рассмотрении последствий переноса начала координат переменных x , y или их логарифмов:

$$X = \lg x; \quad Y = \lg y.$$

Но прежде всего условимся в тех обозначениях, которыми мы, следуя за Саткевичем, будем пользоваться в дальнейшем.

Будем откладывать в обычной прямоугольной системе координат как основные переменные x , y , так и их логарифмы X и Y .

Основанием логарифма может быть любое число a , хотя из

чисто практических соображений мы не будем отходить от обычного десятичного логарифмирования и, ставя везде для общности основание a , будем под ним разумеать «10». Следовательно, $x = a^X$; $y = a^Y$.

Но если на тех же осях откладываются и основные переменные x , y и их логарифмы X , Y , то масштабы, в которых эти разнородные величины будут откладываться, должны быть, очевидно, различными. Оказывается наиболее удобным откладывать логарифмы величин X и Y в равномерном масштабе, а основные величины x и y — в логарифмическом.

Тогда каждая точка будет отстоять от начала координат на величины X и Y , измеренные в равномерном масштабе, а приписанные ей основные величины x и y будут измеряться в логарифмическом масштабе. При этом подчеркиваем, что числа x и y , как аргументы логарифмов, могут быть только положительными. Если же основной материал, полученный измерениями, состоит частично из положительных, частично из отрицательных величин (что бывает сравнительно редко), то необходимо ко всем измерениям добавить соответственную положительную величину, чтобы избавиться от величин отрицательных. Если же все результаты измерений отрицательны, то достаточно у всех их переменить знак на обратный, что очень просто. Теперь нам надо ответить на вопрос, как изменяются логарифмические координаты точки X и Y при переносе начала отсчета основных величин x , y . Для простоты ограничимся параллельным переносом начала. Пусть координаты старого начала в новой системе координат будут:

$$X_0 = \lg x_0; \quad Y_0 = \lg y_0,$$

а координаты точки старого начала обозначим заключением их в скобки. Тогда:

$$\lg x = \lg(x) + \lg x_0;$$

$$\lg y = \lg(y) + \lg y_0$$

или иначе

$$x = x_0(x); \quad y = y_0(y).$$

Следовательно, если полученная нами зависимость, записанная в основных величинах (x) и (y), переносится в новое начало, то для получения соответственных новых значений аргументов мы должны умножить старые значения аргументов на координаты старого начала в новых осях.

При пользовании полулогарифмическими координатами указанная операция должна применяться, очевидно, лишь к логарифмируемой координате, т. е., имея старые координаты (x) и

$\lg(y)$, а координаты старого начала в новых осях x_0 и $\lg y_0$, получим в новых осях

$$\left. \begin{aligned} x &= (x) + x_0 \\ y &= (y) + y_0 \end{aligned} \right\} \quad (234)$$

Переходим теперь к выводу основных выражений для логарифмик сложения и вычитания.

На графике, изображенном на рис. 33, прямая линия OC — биссектриса прямого угла между положительными координатами X и Y — связывает переменные x и y , а также X и Y в их простейшей зависимости:

$$x = y; \quad X = Y$$

(еще раз напоминаем, что числа x и y снимаем с графика, поль-

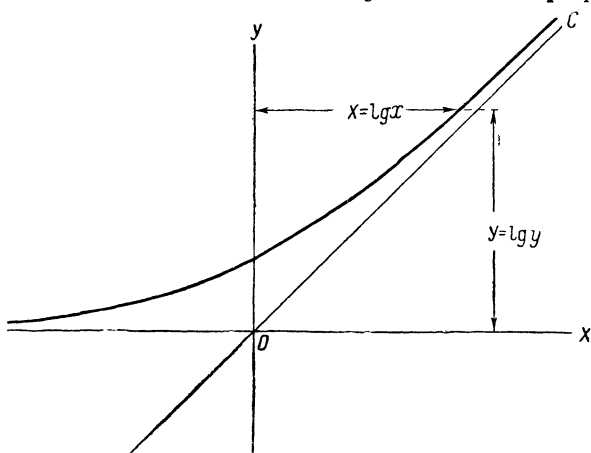


Рис. 33. Логарифмика сложения.

зуюсь логарифмическим масштабом, а числа X и Y берем в равномерном масштабе).

Проведем теперь следующую деформацию этой прямой. Сохраняя все ординаты Y , сдвинем соответствующие абсциссы на единицу. Получим кривую, которая при равномерном изображении выразится уравнением

$$a^x + 1 = a^y, \quad (235)$$

а в логарифмическом

$$x + 1 = y. \quad (235')$$

Это и есть логарифмика сложения.

Как видно на графике (рис. 33) и как легко вывести из уравнения (235), кривая имеет две асимптоты: отрицательную ветвь оси X и положительное направление биссектрисы. Действительно, при $X = +\infty$ имеем $Y = +\infty$, а при $X = -\infty$ $Y = 0$.

Все конечные точки кривой легко вычисляются из уравнения (235). Ниже приводим их таблицу.

Таблица 14

X	-3,0	-2,0	-1,0	-0,5	0,0	+0,5	+1,0	+2,0	+3,0
Y	0,000	0,004	0,043	0,119	0,301	0,619	1,041	2,004	3,000

Из данных табл. 14 видно, что на абсциссах $X = \pm 3$ ординаты кривой практически совпадают с ординатами асимптот.

Иной тип кривой бывает, когда знаки при x и y различны.

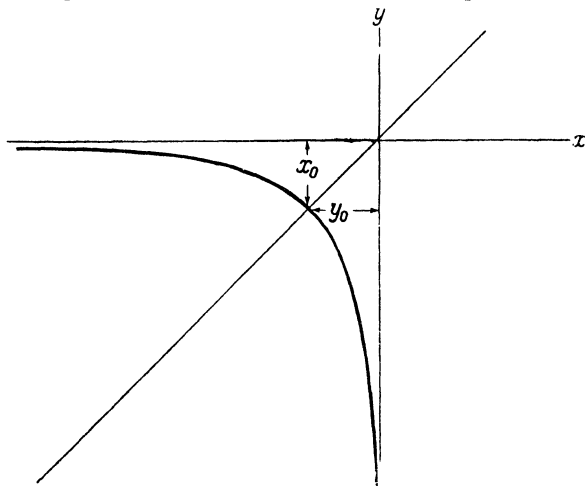


Рис. 34. Логарифмика вычитания.

В этом случае получаем логарифмику вычитания (рис. 34). Ее уравнение запишется в виде

$$1 - a^X = a^Y \quad (236)$$

или в логарифмическом масштабе

$$1 - x = y. \quad (236')$$

При переносе обеих переменных в одну сторону получим соответственно

$$a^X + a^Y = 1,$$

что показывает симметричность этой кривой в отношении осей X и Y . Далее, при $X = -\infty$ $Y = 0$, при $Y = -\infty$ $X = 0$. Положительных же значений X и Y не существует. Это значит, что вся кривая помещается в третьем квадранте и ее асимптотами служат отрицательные ветви X и Y .

Вычисление ординат кривой по точкам из уравнения приведено в табл. 15.

X	-0,301	-0,477	-0,602	-1,0	-2,0	-3,0
Y	-0,301	-0,176	-0,125	-0,046	-0,0044	-0,00043

В этой таблице дана лишь часть кривой, лежащая влево от вершины кривой, расположенной в точке:

$$-X = -\lg 2 = -0,301;$$

$$-Y = -\lg 2 = -0,301.$$

Вправо от этой точки вся кривая имеет тождественный вид и может быть построена по аналогии с первой кривой.

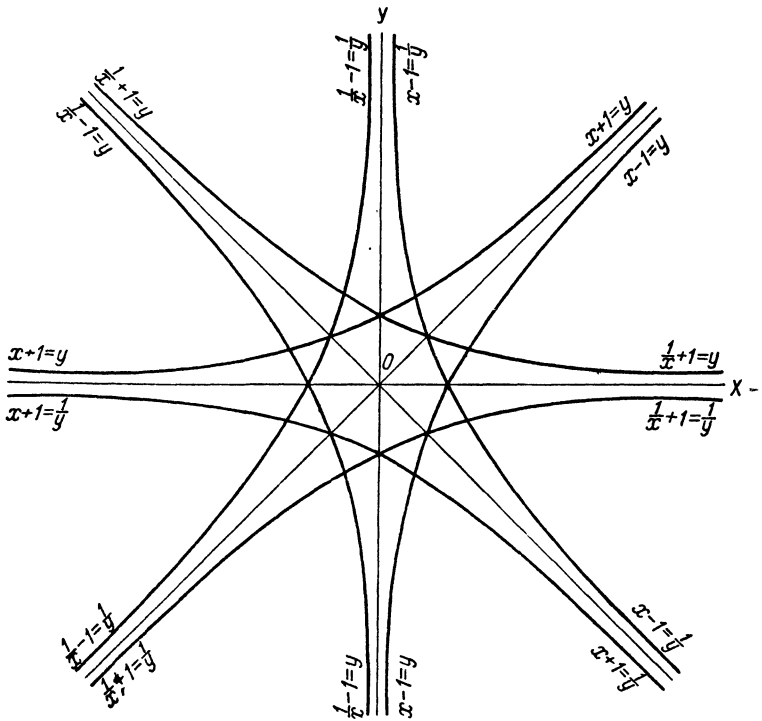


Рис. 35. Звезда логарифмик сложения.

Покажем теперь, что логарифмика сложения имеет восемь вариантов расположения, а логарифмика вычитания — четыре варианта.

Действительно, в уравнении первой кривой

$$a^X + 1 = a^Y$$

можно, во-первых, заменить X через $-X$ и Y через $-Y$, что соответствует замене x на $\frac{1}{x}$ и соответственно y на $\frac{1}{y}$, а во-вторых, можно взаимно переместить X и Y .

Для логарифмики же вычитания

$$a^y + a^x = 1$$

имеет смысл лишь первая замена, вторая же, ввиду симметрии кривой относительно осей координат, ничего нового дать не может.

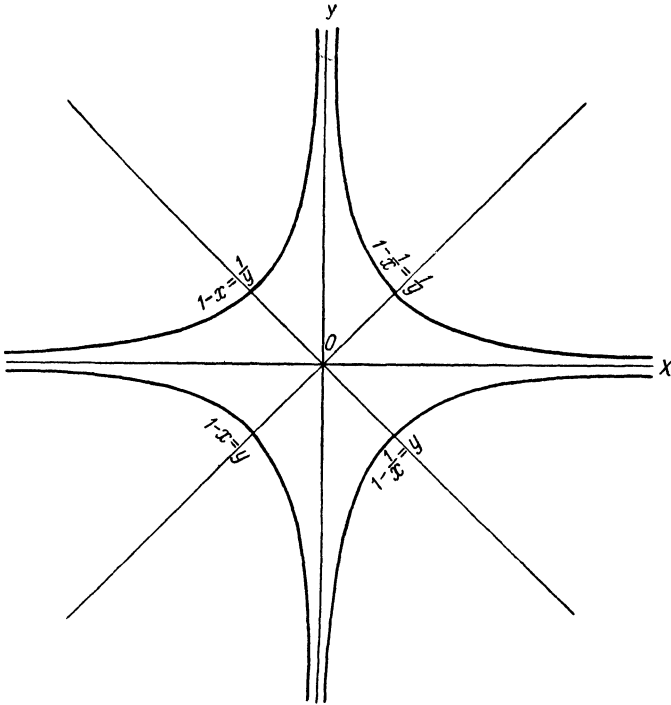


Рис. 36. Звезда логарифмик вычитания.

Полученные две «звезды» — одна для логарифмики сложения (рис. 35), другая для логарифмики вычитания (рис. 36) — исчерпывают всю совокупность основных кривых, которые могут быть использованы в целях линеаризации уравнений более общего вида.

На графиках звезд на каждой кривой указаны в логарифмическом масштабе соответственные уравнения:

$$x + 1 = y,$$

$$\frac{1}{x} + 1 = y$$

и т. д. (рис. 35),

$$1 - x = y,$$
$$1 - \frac{1}{x} = y$$

и т. д. (рис. 36).

Читатель, внимательно ознакомившийся с изложенным в начале параграфа, сможет без труда переписать те же двенадцать уравнений в равномерном масштабе. Например, вместо

$$\frac{1}{x} + 1 = y$$

он запишет

$$a^{-x} + 1 = a^y.$$

При этом вычислять координаты точек следует, очевидно, лишь из уравнений второго типа.

Обратим теперь внимание на то, что все двенадцать кривых, нанесенных на двух звездах, морфологически сводятся к трем; все же остальные могут быть получены из первых путем передвижения по плоскости. Уравнения трех кривых следующие:

$$x + 1 = y; \quad (237)$$

$$x - 1 = y; \quad (238)$$

$$-x + 1 = y. \quad (239)$$

Из этих кривых первые две соответствуют двум основным логарифмикам сложения, а третья представляет общий вид всех логарифмик вычитания. Отсюда возникает естественное и вполне разумное предложение — вырезать одно общее лекала¹, очерченное тремя вышеприведенными кривыми, с общим центром, соответствующим точке пересечения их асимптот, а также и общему началу координат (рис. 37).

Если совокупность точек, полученных логарифмированием измеренных величин, достаточно хорошо располагается вдоль одной из кривых лекала, то центр лекала (отмеченный кружочком) и даст нам соответствующие значения начал отсчета обеих переменных: $X_0 = \lg x_0$, $Y_0 = \lg y_0$.

Нам остается теперь применить то правило переноса начал отсчета, которое непосредственно вытекает из сказанного в начале параграфа. Оно может быть сформулировано следующим образом. Если в некоторую зависимость между измеренными величинами x и y вводится переменная начала их отсчета, то формально это окажется умножением прежних координат то-

¹ Его удобно изготовить из прозрачной целлулоидной пластинки.

чек на логарифмические координаты прежнего начала в новых осях:

$$x = x_0(x); \quad y = y_0(y).$$

Иначе говоря, прежние координаты выражаются через новые отношения:

$$(x) = \frac{x}{x_0}; \quad (y) = \frac{y}{y_0}.$$

Это дает нам право преобразовывать любое уравнение того же типа, что и предыдущие, но содержащее неизвестные

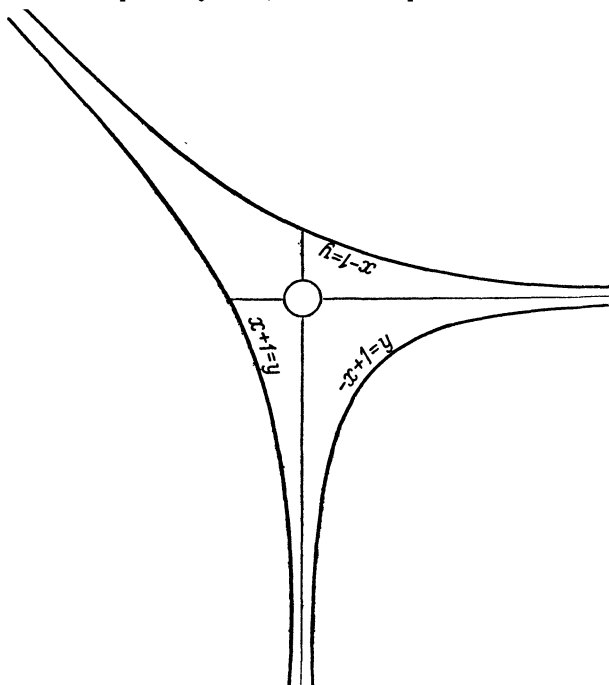


Рис. 37. Лекало для логарифмик.

п а р а м е т р ы, к тому каноническому виду, в котором оно фигурирует на звездах (рис. 35 и 36). Например, уравнение

$$(x) + a = b(y)$$

путем замены

$$a = x_0, \quad \frac{a}{b} = y_0$$

приводится к виду

$$\frac{x}{x_0} + 1 = \frac{y}{y_0}.$$

Уравнение

$$(x)(y + a) = b$$

путем замены

$$\frac{b}{a} = x_0, \quad a = y_0$$

приводится к виду

$$\frac{x_0}{x} - 1 = \frac{y}{y_0}.$$

То же самое читатель легко проделает со всеми двенадцатью уравнениями. Таким образом, испытательное наложение всех трех сторон лекала на совокупность измеренных точек $(x_j y_j)$ установит принадлежность этой совокупности к такому-то из рассмотренных типов. или, наоборот, отвергнет такое предположение. В первом утвердительном случае при помощи этого же лекала мы получим (графически) значение параметров x_0, y_0 . Последнее звено всей операции — замена уравнения в канонической форме соответственным уравнением прямой — не представляет уже никаких затруднений, поскольку параметры, выражающие добавки к аргументам, известны и мы в состоянии их добавлять уже чисто арифметически.

Применим изложенные соображения к конкретному примеру, опять-таки подчеркивая, что весь изложенный метод годится лишь при малом разбросе точек.

Имеем совокупность измерений взаимосвязанных величин x и y . Логарифмируем их и наносим в виде точек с координатами $X = \lg x$ и $Y = \lg y$, придерживаясь такого масштаба, при котором заготовленное лекало могло бы охватить все точки (рис. 38). Над каждой измеренной точкой надписываем ее основные координаты (аргументы) x и y . Передвигая затем лекало по плоскости, находим такое его положение, при котором измеренные точки меньше всего отклоняются от одной из трех ограничивающих его кривых. Определяем для этого положения координаты центра лекала X_0, Y_0 и соответственно x_0, y_0 .

Выбираем на звезде (сложения или вычитания) наиболее подходящую кривую и ставим в ее уравнении $\frac{x}{x_0}$ вместо x и $\frac{y}{y_0}$ вместо y , где x_0, y_0 — величины, полученные графически (с помощью лекала). В результате получим, например, уравнение

$$\frac{x}{x_0} - 1 = \frac{y}{y_0}$$

или иначе

$$(x - x_0) y_0 = x_0 y.$$

Это и есть искомое уравнение прямой линии. Пересчитанное в переменные X и Y , оно даст

$$\frac{a^X - a^{X_0}}{a^Y} = \frac{X_0}{Y_0} = \text{const.}$$

Нам представляется, что изложенного примера вполне достаточно для того, чтобы читатель смог сам применить к измеренным точкам одно из двенадцати уравнений, написанных на обеих звездах, если (повторяем!) наша совокупность измерений вообще может быть интерполирована одним из этих уравнений. А это, очевидно, далеко не всегда имеет место.

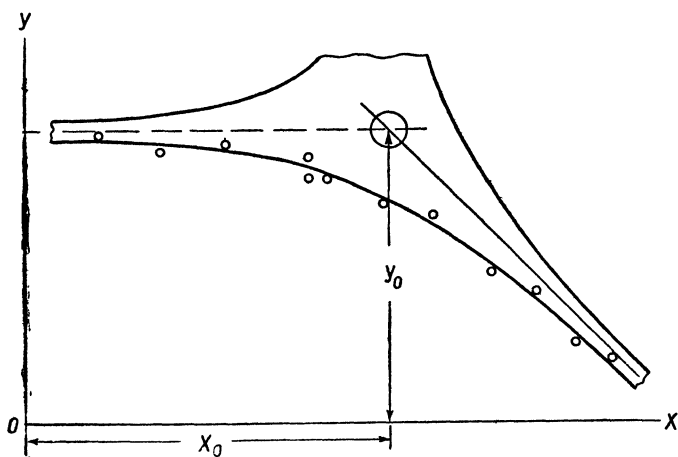


Рис. 38. Пример расчета с помощью лекала.

В заключение данного параграфа рассмотрим, в какой мере тот же метод можно применить и к более сложному случаю, когда в каком-либо из типовых уравнений обе части возведены в разные степени. Возьмем к примеру уравнение (235) и деформируем его следующим образом:

$$(x + 1)^m = y^n$$

или иначе

$$x + 1 = y^p, \quad (240)$$

где $p = \frac{n}{m}$. Применить здесь для отыскания начала координат x_0, y_0 одно общее лекало, очевидно, нельзя, так как для каждого значения показателя p и для каждого типа основного уравнения оно будет различным. И если в нашей задаче значение p неизвестно и мы должны его отыскать, то это возможно лишь путем последовательного испытания разных лекал, составленных для разных типов основного уравнения и разных значений p . Работа весьма трудоемкая и рекомендовать ее для целей реше-

ния практических задач можно лишь в некоторых частных случаях. Например, предположим, что характер задачи позволяет заранее остановиться хотя бы на вышеприведенном уравнении (240) или на другом аналогичном и что приблизительное значение показателя p нам из каких-либо соображений известно. Тогда будет достаточно сделать пять-шесть лекал со значениями p , близкими к этому приближенному, и путем последовательного испытания найти координаты x_0, y_0 . Уравнение (240) примет в этом случае вид

$$\frac{x}{x_0} + 1 = \left(\frac{y}{y_0}\right)^p$$

и после элементарных преобразований может быть приведено к виду уравнения между X и Y :

$$a^{pY} = \frac{a^{pY_0}}{a^{X_0}} (a^X + a^{X_0}), \quad (241)$$

иначе говоря, к искомой линейной зависимости.

Теперь вернемся к задаче о зависимости расхода воды в реке от показания рейки и увидим, что ей действительно соответствует типовое уравнение (235) и, следовательно, к ней можно применить только что приведенную схему расчета (240), в которой переменная y — расход Q , т. е. величина размерности L^3T^{-1} , а переменная x — высота уровня H по рейке, имеющая размерность длины. Показатель p в этом уравнении не может сильно отличаться от $\frac{1}{3}$, и поставленная задача может быть решена методом Саткевича, по-видимому, без особых затруднений. Следует отметить, что метод Глушкова при решении указанной задачи потребует гораздо меньше труда и времени, хотя, правда, результат решения может быть не очень точным.

В заключение отметим, что основная часть метода Саткевича, разработанная им применительно к задачам, охватываемым вышеприведенными двенадцатью уравнениями, имеет несомненную ценность. Практическое ее применение при пользовании предложенным им лекалом (см. рис. 37) не представляет никаких затруднений. Ранее был приведен пример решения таких задач. Сейчас же рассмотрим их общую схему решения.

Во все двенадцать уравнений, надписанных на звездах сложения и вычитания (см. рис. 35 и 36), можно ввести неопределенные параметры p и q , после чего все они могут быть представлены символически в одной форме:

$$px^m + qy^n = 1, \quad (242)$$

где показатели m и n могут принимать лишь одно из двух значений: или $+1$, или -1 . Коэффициенты p и q могут быть и положительными и отрицательными. Передвигая упомянутое лекало по полю экспериментальных точек, представленных их ло-

гарифмами, можно определить, удовлетворяет ли распределение этих точек одной из трех представленных на лекале кривых. В этом случае получаем значение показателей m и n , а из координат центра лекала получаем значения x_0 и y_0 . Из последних легко находим значения неизвестных параметров p и q . Преобразуя затем найденное уравнение к новым переменным:

$$(x) = \frac{x}{x_0}; \quad (y) = \frac{y}{y_0}$$

(где x_0 и y_0 теперь известны), приходим к одной из кривых, представленных на двух звездах. Асимптота этой кривой и даст нам линеаризованную форму искомой зависимости.

ОСНОВНЫЕ ФОРМЫ ЭМПИРИЧЕСКИХ КРИВЫХ

§ 55. Типы монотонных кривых

При отыскании приближенных зависимостей между измеренными взаимосвязанными величинами вопрос о выборе математической формы для этой зависимости, иначе говоря, выбор того приближенного уравнения, которому искомая связь должна удовлетворять, решается совокупно как на основе интуитивной оценки расположения экспериментальных точек (x_k, y_k) , так и с помощью учета тех теоретических соображений, которые даже в малоисследованных областях все же не могут полностью отсутствовать. Например, будучи уверенными, что переменная y с возрастанием переменной x также возрастает, можем из каких-либо теоретических соображений считать, что это возрастание не может быть безграничным. При повышении количества осадков и при слабой изменчивости температуры испарение хотя и возрастает, но в определенных географических условиях оно не может превзойти известного предела. При возрастании скорости потока размыв русла будет также возрастать, но это возрастание ограничено определенным пределом. Подобных примеров можно привести большое количество.

Аналогичные соображения могут побудить нас выбрать уравнение с горизонтальной асимптотой для переменной y . В других случаях какие-то теоретические соображения могут привести к мысли о наличии в искомой зависимости экстремального значения для переменной y в таком-то интервале значений x . Могут быть и такие случаи, когда по существу изучаемого явления нулевому значению x должно обязательно соответствовать нулевое же значение y , следовательно, искомая кривая обязана выходить из начала координат. Сюда относится опять-таки зависимость между годовыми осадками и годовым испарением, которая, очевидно, должна выражаться кривой, проходящей через начало координат. Пользуясь примером, приведен-

ным выше, можно утверждать, что эта кривая должна иметь в начальной точке производную, равную единице, так как малое количество выпавших осадков полностью испаряется.

Вообще говоря, почти всегда имеются некоторые предварительные указания хотя бы и качественного характера для выбора вида искомой кривой, а поэтому последняя обычно имеет форму, частично уже определенную теорией. Если этого нет и если мы для нахождения искомой зависимости ничего, кроме совокупности измерений, не имеем (что иногда бывает), то вся задача сводится к одной лишь интерполяции эмпирических точек, т. е. технически задача упрощается, хотя это и снижает ценность результата.

В заключительной главе, посвященной форме отыскиваемых кривых, обратим особое внимание именно на те кривые, форма которых частично диктуется какими-либо соображениями, не зависящими от качества (точности) проведенных измерений. Учитывая такие соображения, можно в некоторых случаях прийти к обоснованному отбрасыванию отдельных измерений, если они дают точки, положения которых противоречат тому, что нам заранее известно.

Ограничимся в этой главе разбором двух наиболее обычных категорий эмпирических кривых: 1) монотонных кривых и 2) кривых с одним экстремумом.

В данном параграфе рассмотрим первую категорию кривых, причем линейные зависимости, о которых в прежних главах было сказано вполне достаточно, здесь не будем рассматривать.

Начнем с параболы второго порядка

$$y = a + bx + cx^2, \quad (243)$$

характеризующейся тремя параметрами, геометрический смысл которых очевиден.

1. Параметр a дает длину отрезка на оси y при $x=0$.

2. Параметр b дает в той же точке тангенс угла наклона кривой к оси x .

3. Параметр c , как нетрудно видеть из выражения для кривизны кривой

$$\frac{d^2y}{dx^2} : \left\{ 1 + \left(\frac{dy}{dx} \right)^2 \right\}^{3/2} = \frac{2c}{\{1 + (b + 2cx)^2\}^{3/2}},$$

определяет изменение кривизны вдоль оси x . Таким образом, при $x=0$ имеем максимальную кривизну кривой, равную $\frac{2c}{(1+b^2)^{3/2}}$,

которая затем убывает и в пределе при $x \rightarrow \infty$ составляет нуль. Знак кривизны кривой совпадает со знаком c ; при $c > 0$ кривая непрерывно поднимается, обращаясь выпуклостью вниз (или

вправо), и наоборот. При $c < 0$ условие наличия экстремума определяется из очевидного равенства

$$b + 2cx = 0,$$

из которого следует, что в области положительных значений x один из двух параметров — или b , или c — должен быть отрицательным. В этом случае расстояние от начала до экстремума

$$x = -\frac{b}{2c} > 0.$$

Если значение x меньше наименьшего из измеренных значений x_j или больше наибольшего из них, то полученная кривая, рассматриваемая лишь как интерполяционная, может считаться практически монотонной. Если же кривую экстраполируем (на

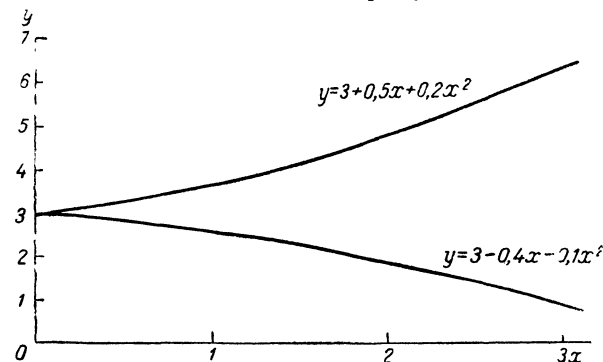


Рис. 39. Вид возрастающей и убывающей зависимости параболического типа.

основании каких-либо теоретических соображений), то для ее монотонности во всей области положительных значений x необходимо, чтобы оба параметра (b и c) были либо положительными, либо отрицательными, а это значит, что экстремум вообще будет отсутствовать. В первом случае мы получим возрастающую монотонную зависимость, во втором — убывающую, как это показано на рис. 39.

Таким образом, характер расположения экспериментальных точек на графике определяет общий вид параболы второго порядка. Значение же параметров a , b , c надо определять, очевидно, способом наименьших квадратов, что для параболы второго порядка требует затраты небольшого труда. Но может случиться, что отклонения точек от кривой получатся большими, чем следовало ожидать, или, что еще важнее, расположения отклонений в обе стороны от кривой дадут явную асимметрию. Тогда рекомендуется воспользоваться способом Чебышева и подсчитать, какие отклонения дадут те же точки при параболе

третьего порядка. Если новые отклонения станут меньше и симметричнее, то лучше перейти от второго порядка к третьему. Дальше третьего порядка можно идти лишь при очень точных измерениях, дающих основание отыскивать для них более точную форму зависимости.

Рассмотренная параболическая форма относится к той категории монотонных зависимостей, в которых при безграничном возрастании одной переменной другая также безгранично возрастает (или убывает). Сюда же можно отнести также следующие формы:

1. Показательную зависимость вида $y = A \cdot 10^{\pm ax}$, где два

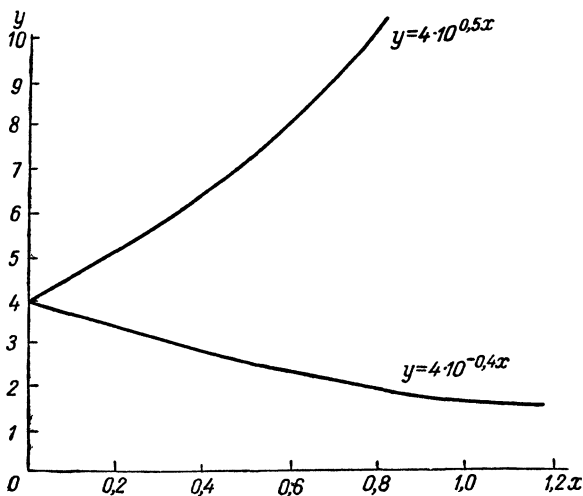


Рис. 40. Вид возрастающей и убывающей зависимости показательного типа.

знака дают для y соответственно или возрастание, или убывание. Более общая форма получится при замене показателя ax через какую-либо монотонную функцию от x , например через рассмотренную выше параболу второго порядка с одинаковыми знаками при b и c . Тогда будем иметь

$$y = 10^{a + bx + cx^2}. \quad (244)$$

Уравнение (244) после логарифмирования можно легко решить способом наименьших квадратов.

Вид простой зависимости с одночленным показателем для положительных и отрицательных значений параметра b приведен на рис. 40. Более сложная монотонная зависимость вида (244) изображена на рис. 41 для случаев положительных и отрицательных значений b и c .

2. При решении задач можно использовать степенную параболу с дробным показателем, который должен быть найден из совокупности измеренных точек способом наименьших квадратов. В этом случае

$$y = bx^m. \quad (245)$$

Если же параболы второго и даже третьего порядка не дают удовлетворительной картины отклонений, то, применяя дробную параболу, можно найти лучшее решение. Из анализа уравнения (245) легко получить, что при $m > 1$ кривая в начальной точке касается оси y , а при $m < 1$ — оси x . В обоих случаях кривизна ее падает вдоль осей x и y .

Кривая (245) имеет тот недостаток, что она обязана выходить

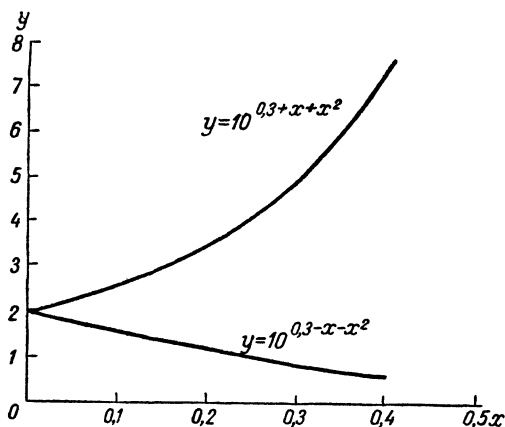


Рис. 41. Вид кривой $y = 10^{a+bx+cx^2}$.

из начала координат. Предыдущая кривая (244) была в этом отношении свободнее: при $x=0$ она пересекала ось y на расстоянии 10^a от начала, а параметр a находился по расположениям измеренных точек и мог иметь любое, соответствующее им значение. В кривой же (245) это может быть достигнуто лишь путем присоединения свободного числа

$$y = a + bx^m, \quad (246)$$

что затрудняет пользование способом наименьших квадратов. Действительно, преобразуя (246) к виду

$$x = \frac{1}{b^{1/m}} (y - a)^{1/m} \quad (247)$$

или после логарифмирования к виду

$$\lg x = \frac{1}{m} \lg b + \frac{1}{m} \lg (y - a),$$

получаем выражение, которое не поддается ни строгому методу Саткевича, ни приближенному методу Глушкова (см. предыдущую главу). И, по-видимому, единственным способом решения этого уравнения будет предварительное приближенное определение параметра a , которое равно y при $x=0$.

Если скопление экспериментальных точек вблизи оси y достаточно густое, то имеем право, в частности, приняв для этого скопления гипотезу линейной зависимости вида

$$y = a + bx,$$

вычислить из нее один лишь параметр a (считая значение b в линейризированной зависимости менее надежным), после этого, вводя новую переменную

$$z = y - a,$$

найти значение параметров b и m способом наименьших квадратов из уравнения

$$\lg x = \frac{1}{m} \lg z - \frac{1}{m} \lg b.$$

Рассмотрим теперь те монотонные зависимости, в которых переменная y возрастая не может превысить известного предела или убывая не может быть ниже известного предела, иными словами, кривая должна иметь горизонтальную асимптоту или сверху, или снизу. Простейший вид такого рода кривых дается уравнением

$$y = \frac{a + bx}{1 + cx}. \quad (248)$$

В зависимости от значений параметров a, b, c производная

$$\frac{dy}{dx} = \frac{b - ac}{(1 + cx)^2}$$

дает при положительном числителе $b - ac$ возрастание, а при отрицательном убывание к асимптоте. Высота асимптоты над осью x в обоих случаях выражается через $\frac{b}{c}$, как это представлено на рис. 42 для двух систем значений параметров: 1) $a=2, b=5, c=1$ и 2) $a=4, b=3, c=2$.

Далее нам остается только показать, как применить способ наименьших квадратов к решению уравнения (248), явно нелинейному в отношении параметров a, b, c . Преобразуем это уравнение к виду

$$y - \frac{b}{c} = \frac{a - \frac{b}{c}}{cx + 1}$$

и зададимся целью найти хотя бы в первом приближении значение дроби $\frac{b}{c}$, которая выражает, как известно, высоту асимптоты

над (или под) осью x . Если в поле измеренных точек имеется несколько точек с достаточно большими значениями x_j , то, выделяя эту группу из общей совокупности точек, имеем право принять одно из двух крайних предположений: 1) или искомая асимптота находится на средней для всех точек высоте, 2) или ее высота совпадает с самой высокой (соответственно самой низкой) из точек. Среднее из этих двух предположений представляется как

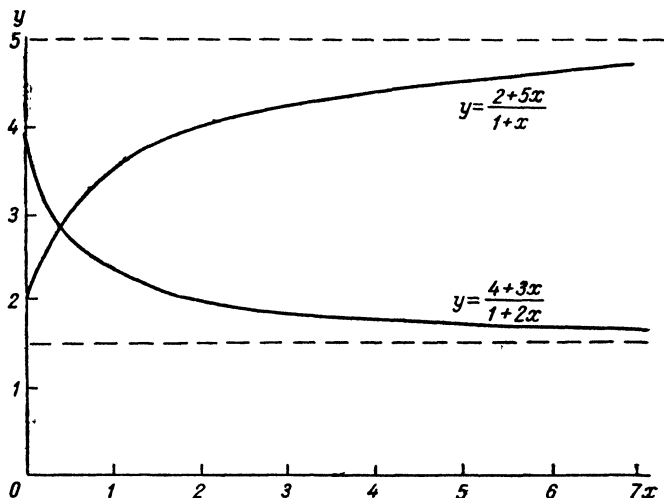


Рис. 42. Вид кривой $y = \frac{a + bx}{1 + cx}$.

будто наиболее правдоподобным. Приняв последнее и вводя обозначения:

$$\frac{1}{y_j - \frac{b}{c}} = z_j; \quad a - \frac{b}{c} = d,$$

получим простое уравнение:

$$dz_j - cx_j - 1 = 0,$$

решая которое способом наименьших квадратов найдем значение параметров c и d , а затем и параметра a , равного $d + \frac{b}{c}$.

Кривые с асимптотами можно также получить путем использования некоторых простейших трансцендентных функций: три-

гонометрических, гиперболических и показательных. Например, все следующие выражения:

$$y = a \cdot 10^{-\frac{b}{x}}; \quad (249)$$

$$y = \frac{2a}{\pi} \operatorname{arc} \operatorname{tg} \frac{x}{b}; \quad (250)$$

$$y = a \operatorname{th} \frac{x}{b} \quad (251)$$

дают при $x = \infty$ одну и ту же асимптоту $y = a$, как это представлено на совмещенном графике (рис. 43).

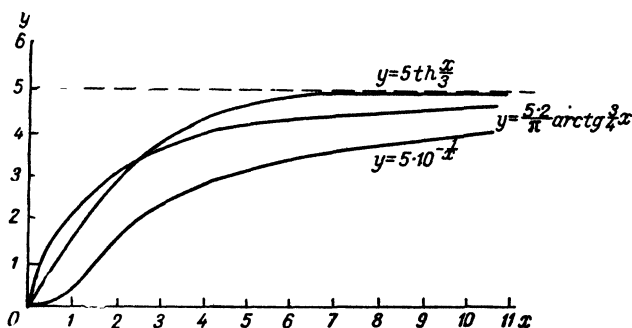


Рис. 43. Сравнение трех кривых, ограниченных сверху одной асимптотой.

Но так как все три кривые, выходя из начала координат, совпадают, то следует выяснить способом наименьших квадратов лишь следующий вопрос, при какой кривой разброс опытных точек будет меньше. Что касается определения их параметров a и b , то при решении уравнения (249) это делается очень просто: логарифмируя его, получим

$$\lg y = \lg a - \frac{b}{x}, \quad (249')$$

а затем, заменяя $\frac{1}{x}$ через z , будем иметь простейший случай линейной зависимости между двумя параметрами. Анализ остальных уравнений: (250) и (251) — показывает, что зависимости между параметрами получаются нелинейными, и, следовательно, способ наименьших квадратов непосредственно применить к ним нельзя. Тогда надо поступить иначе: определить параметр a , т. е. высоту асимптоты по уравнению (249). Поскольку значение этого

параметра по доказанному выше будет одно и то же во всех трех кривых, то, заменяя $\frac{y}{a}$ через z , преобразуем уравнения (250) и (251) в следующие уравнения:

$$x = b \operatorname{tg} \frac{z\pi}{2}; \quad (250')$$

$$x = b \operatorname{Arth} z. \quad (251')$$

Эти последние уравнения теперь легко решаются обычными приемами, заключающимися в том, что надо лишь подсчитать из ряда измеренных значений y_j два других:

$$z_j = \operatorname{tg} \frac{y_j\pi}{2a}; \quad u_j = \operatorname{Arth} \frac{y_j}{a},$$

после чего решать линейные уравнения относительно параметра b .

Приведенный пример соответствует задаче в гидрологии речного стока — выяснению вида эмпирической зависимости между количеством годовых осадков и суммарным испарением с речного бассейна. Аналогичные задачи возможны, конечно, и в других областях.

Что касается других, более простых кривых, убывающих к горизонтальной асимптоте, отличной от оси x , то наиболее простое решение получается путем использования уже рассмотренного уравнения (248), которое допускает любое начало кривой и любое положение асимптоты. В частном же случае, когда по смыслу задачи асимптотой должна быть ось x , дело значительно упрощается и можно применить, например, уравнение (244), приняв в нем $b < 0$, $c = 0$. В тех случаях, когда решение сокращенного уравнения

$$\lg y = a - bx$$

способом наименьших квадратов даст завышенные и явно асимметричные отклонения, можно перейти к параболическому уравнению вида

$$\lg y = a - bx - cx^2$$

(коэффициент c должен быть, очевидно, отрицательным).

§ 56. Кривые с одним экстремумом

В начале предыдущего параграфа показано, что парабола второго порядка

$$y = a + bx + cx^2$$

имеет экстремум в положительной области x лишь в том случае, когда знаки параметров b и c различны, причем из очевидных соображений явствует, что если $b > 0$, $c < 0$, то будет максимум, если же $b < 0$, $c > 0$ — минимум (рис. 44). Вторично оговари-

ваем, что этот экстремум может оказаться и за пределами измеренных точек.

То же соотношение будет иметь место и при показательной зависимости вида (244).

Зависимости с экстремумом для y вдоль оси x можно получить вообще из многих кривых второго порядка, но не из всех. Достаточно указать на уравнение (248) как на пример тех уравнений второго порядка, которые не имеют экстремума. В этом

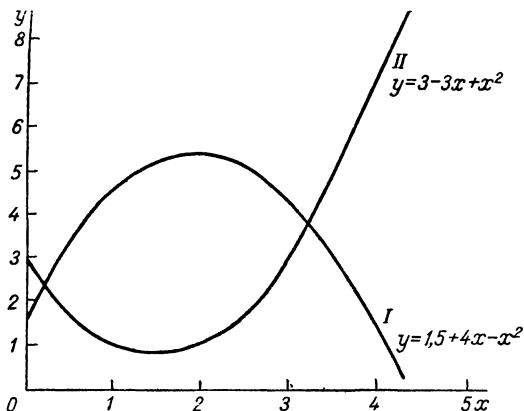


Рис. 44. Вид параболы с максимумом (I) и с минимумом (II).

уравнении числитель и знаменатель в правой части суть линейные функции от x и производная $\frac{dy}{dx}$ не может равняться нулю, т. е. экстремум в ней отсутствует. Но стоит только в том же уравнении добавить или в числителе, или в знаменателе (или и в числителе, и в знаменателе) слагаемое, содержащее x^2 , как у кривой появится экстремум. Возьмем, например, уравнение

$$y = \frac{x}{a + bx + cx^2}. \quad (252)$$

Оно будет иметь экстремум (рис. 45) в том случае, если параметры a и c или оба положительны, или оба отрицательны.

Абсцисса экстремума в обоих случаях будет равна $\sqrt{\frac{a}{c}}$.

В первом случае будем иметь максимум, а во втором — минимум, как это видно из знака второй производной:

$$\frac{d^2y}{dx^2} = \frac{2cx}{(a + bx + cx^2)^4}.$$

Существенное различие в формах кривой (рис. 44) и кривой (рис. 45) заключается в том, что первая после максимума (или минимума) опускается (или соответственно поднимается) в отрицательную (или положительную) бесконечность для $x = \infty$. Вторая кривая после экстремума асимптотически приближается к оси x . Поэтому выбор одной из них диктуется визуальной оценкой расположения экспериментальных точек.

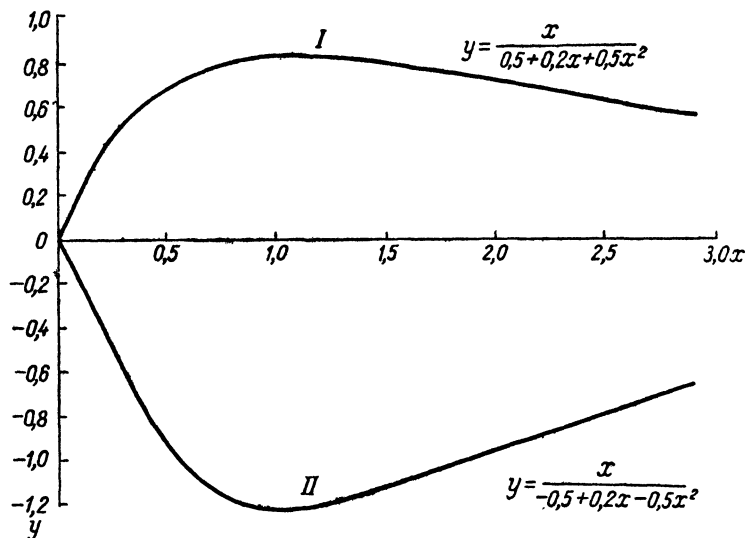


Рис. 45. Вид кривой $y = \frac{x}{a + bx + cx^2}$ с максимумом (I) и с минимумом (II).

Весьма удобным выразителем этого второго типа расположения точек может служить кривая, получаемая из произведения степенной и показательной функций:

$$y = Ax^a \cdot 10^{-bx}. \quad (253)$$

Эта кривая, в сущности, представляет собой сходство с хорошо нам известным гамма-распределением, так как в последнем заменили число e через основание десятичных логарифмов и поставили вместо гамма-функции неопределенный коэффициент A , который находится, как и два других параметра, способом наименьших квадратов из логарифмирования (253):

$$\lg y = \lg A + a \lg x - bx.$$

На этом кратком обзоре лишь наиболее употребительных форм эмпирических кривых заканчивается заключительная глава данной книги.

ЛИТЕРАТУРА

1. Laplace P. Theorie analytique des probabilités. Paris, 3-ème edition, 1820.
2. Марков А. А. Исчисление вероятностей. Гос. изд-во, М., 1924.
3. Poisson A. Calcul des probabilités. Paris, 1912.
4. Гливенко В. И. Курс теории вероятностей ГОНТИ, М.—Л., 1939.
5. Дунин-Барковский И. В. и Смирнов Н. В. Теория вероятностей и математическая статистика в технике (Общая часть). Гостехиздат, М., 1955.
6. Крамер Г. Математические методы статистики. Гос. изд-во иностр. лит., М., 1948.
7. Ван дер Варден Б. Л. Математическая статистика. Изд-во иностр. лит., М., 1960.
8. Колмогоров А. Н. Определение центра рассеивания и меры точности по ограниченному числу наблюдений. Изв. АН СССР, сер. матем., т. 6, 1942.
9. Линник Ю. В. Об устойчивости вероятностных законов с показателем меньшим единицы. ДАН СССР, т. XXIV, № 4, 1954.
10. Романовский В. И. Основные задачи теории ошибок, Гостехиздат, М.—Л., 1947.
11. Линник Ю. В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. Физматгиз, М., 1958.
12. Уорсинг А. и Геффнер Дж. Методы обработки экспериментальных данных. Изд-во иностр. лит., М., 1949.
13. Смирнов Н. В. Об оценке максимального члена в ряду наблюдений. ДАН СССР, т. 33, № 5, 1941.
14. Уиттекер Э. и Робинсон Г. Математическая обработка результатов наблюдений. ГТТИ, М.—Л., 1933.
15. Гнеденко Б. В. и Колмогоров А. Н. Предельные распределения для сумм независимых случайных величин. ГТТИ, М.—Л., 1949.
16. Идельсон Н. И. Способ наименьших квадратов, изд. 2-е. Кучуб, Л., 1932.
17. Чеботарев А. С. Способ наименьших квадратов с основами теории вероятностей. ОНТИ, М.—Л., 1936.
18. Шилов П. И. Способ наименьших квадратов. Геодезиздат, М., 1941.
19. Edgeworth F. Y. A new method of reducing observations.. Phil. Mag., 25 (1888), p. 184.
20. Вейнберг Б. П. Критерий достаточной малости расстояния между точками наблюдений при микросъемках. Изв. ГГО, № 1—2, Л., 1932.
21. Бриджмен П. В. Анализ размерностей. ОНТИ, М., 1934.
22. Седов Л. И. Методы подобия и размерности в механике. Гостехиздат, М., 1957.
23. Цукерман М. Л. Эмпирические формулы неперiodического характера. Методы их отыскания по данным экспериментальных исследований. Энергоиздат, М.—Л., 1932.
24. Семендяев К. А. Эмпирические формулы. ГТТИ, М., 1933.
25. Саткевич А. А. Приемы исследования эмпирических кривых. II. Координатные изображения в логарифмических масштабах как средство анализа числовых результатов наблюдения. НКЗ, М., 1922.

Вероятности „выскакивающих“ измерений (по Стьюденту)

n	$\beta = 0,05$	$\beta = 0,02$	$\beta = 0,01$
2	15,561	38,973	77,964
3	4,969	8,042	11,460
4	3,558	5,077	6,530
5	3,041	4,105	5,043
6	2,777	3,635	4,355
7	2,616	3,360	3,963
8	2,508	3,180	3,711
9	2,431	3,053	3,536
10	2,372	2,959	3,409
11	2,327	2,887	3,310
12	2,291	2,829	3,233
13	2,261	2,782	3,170
14	2,236	2,743	3,118
15	2,215	2,710	3,075
16	2,197	2,683	3,038
17	2,181	2,658	3,006
18	2,168	2,637	2,997
19	2,156	2,618	2,953
20	2,145	2,602	2,932
21	2,135	2,587	2,912
22	2,127	2,575	2,895
23	2,119	2,562	2,880
24	2,112	2,552	2,865
25	2,105	2,541	2,852
26	2,099	2,532	2,840
27	2,094	2,524	2,830
28	2,088	2,517	2,820
29	2,083	2,509	2,810
30	2,079	2,503	2,802
40	2,048	2,456	2,742
60	2,018	2,411	2,683
120	1,988	2,368	2,628
∞	1,960	2,326	2,576

Удвоенная функция Лапласа $\delta = 2 \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

z	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0,0797	0,0876	0,0955	0,1034	0,1113	0,1192	0,1271	0,1350	0,1428	0,1507
0,2	0,1585	0,1663	0,1741	0,1819	0,1897	0,1974	0,2051	0,2128	0,2205	0,2282
0,3	0,2358	0,2434	0,2510	0,2586	0,2661	0,2737	0,2812	0,2886	0,2960	0,3035
0,4	0,3108	0,3182	0,3255	0,3328	0,3401	0,3473	0,3545	0,3616	0,3688	0,3759
0,5	0,3829	0,3900	0,3969	0,4039	0,4108	0,4177	0,4245	0,4313	0,4381	0,4448
0,6	0,4515	0,4581	0,4647	0,4713	0,4778	0,4843	0,4908	0,4971	0,5035	0,5098
0,7	0,5161	0,5223	0,5285	0,5346	0,5407	0,5468	0,5528	0,5587	0,5616	0,5705
0,8	0,5763	0,5821	0,5878	0,5935	0,5991	0,6047	0,6102	0,6157	0,6211	0,6265
0,9	0,6319	0,6372	0,6424	0,6476	0,6528	0,6579	0,6629	0,6680	0,6729	0,6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	0,7287	0,7330	0,7373	0,7415	0,7457	0,7499	0,7540	0,7580	0,7620	0,7660
1,2	0,7699	0,7737	0,7775	0,7813	0,7850	0,7887	0,7923	0,7959	0,7994	0,8030
1,3	0,8064	0,8098	0,8132	0,8165	0,8198	0,8230	0,8262	0,8293	0,8324	0,8355
1,4	0,8385	0,8415	0,8444	0,8473	0,8501	0,8529	0,8557	0,8584	0,8613	0,8638
1,5	0,8664	0,8670	0,8715	0,8740	0,8764	0,8789	0,8812	0,8836	0,8859	0,8882
1,6	0,8904	0,8926	0,8948	0,8969	0,8990	0,9011	0,9031	0,9051	0,9070	0,9090
1,7	0,9109	0,9127	0,9146	0,9164	0,9181	0,9199	0,9216	0,9233	0,9249	0,9266
1,8	0,9281	0,9297	0,9312	0,9328	0,9342	0,9357	0,9371	0,9385	0,9399	0,9412
1,9	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9534
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	0,9643	0,9651	0,9660	0,9668	0,9676	0,9684	0,9692	0,9700	0,9707	0,9715
2,2	0,9722	0,9729	0,9736	0,9742	0,9749	0,9756	0,9762	0,9768	0,9774	0,9780
2,3	0,9786	0,9791	0,9797	0,9802	0,9807	0,9812	0,9817	0,9822	0,9827	0,9832
2,4	0,9836	0,9840	0,9845	0,9849	0,9853	0,9857	0,9861	0,9865	0,9869	0,9872
2,5	0,9876	0,9879	0,9883	0,9886	0,9889	0,9892	0,9895	0,9898	0,9901	0,9904
2,6	0,9907	0,9910	0,9912	0,9915	0,9917	0,9920	0,9922	0,9924	0,9926	0,9928
2,7	0,9931	0,9933	0,9935	0,9937	0,9939	0,9940	0,9942	0,9944	0,9946	0,9947
2,8	0,9949	0,9950	0,9952	0,9954	0,9955	0,9956	0,9958	0,9959	0,9960	0,9962
2,9	0,9963	0,9964	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	0,9981	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986
3,2	0,9986	0,9987	0,9987	0,9988	0,9988	0,9988	0,9989	0,9989	0,9990	0,9990

Значения t_{δ} в равенстве $\delta = 2 \int_0^{t_{\delta}} \delta(t, k) dt$

(по Стьюденту — Фишеру)

k	$\alpha = 0,4$	$\alpha = 0,5$	$\alpha = 0,6$	$\alpha = 0,7$	$\alpha = 0,8$
1	0,727	1,000	1,376	1,963	3,078
2	617	0,816	1,061	1,336	1,886
3	584	765	0,978	1,250	2,638
4	569	741	941	1,190	1,533
5	559	727	920	1,156	1,476
6	553	718	906	1,134	1,440
7	549	711	896	1,119	1,415
8	546	706	889	1,108	1,397
9	543	703	883	1,100	1,383
10	542	700	879	1,093	1,372
11	540	697	876	1,088	1,363
12	539	695	873	1,083	1,356
13	538	694	870	1,079	1,350
14	537	692	868	1,076	1,345
15	536	691	866	1,074	1,341
16	535	690	865	1,071	1,337
17	534	689	863	1,069	1,333
18	534	688	862	1,067	1,330
19	533	688	861	1,066	1,328
20	533	687	860	1,064	1,325
21	532	686	859	1,063	1,323
22	532	686	858	1,061	1,321
23	532	685	858	1,060	1,319
24	531	685	857	1,059	1,318
25	531	684	856	1,058	1,316
26	531	684	856	1,058	1,315
27	531	684	855	1,057	1,314
28	530	683	855	1,056	1,313
29	530	683	854	1,055	1,311
30	530	683	854	1,055	1,310
40	529	681	851	1,050	1,303
60	527	679	848	1,046	1,296
120	526	677	845	1,041	1,289
∞	524	674	842	1,036	1,282

k	$\alpha = 0,90$	$\alpha = 0,95$	$\alpha = 0,98$	$\alpha = 0,99$	$\alpha = 0,999$
1	6,314	12,706	31,821	63,657	636,619
2	2,920	4,303	6,965	9,925	31,598
3	2,353	3,182	4,541	5,841	12,941
4	2,132	2,776	3,747	4,604	8,610
5	2,015	2,571	3,365	4,032	6,859
6	1,943	2,447	3,143	3,707	5,959
7	1,895	2,365	2,998	3,409	5,405
8	1,860	2,306	2,896	3,355	5,041
9	1,833	2,262	2,821	3,250	4,781
10	1,812	2,228	2,764	3,169	4,587
11	1,796	2,201	2,718	3,106	4,487
12	1,782	2,179	2,681	3,055	4,318
13	1,771	2,160	2,650	3,012	4,221
14	1,761	2,145	2,624	2,977	4,140
15	1,753	2,131	2,602	2,947	4,073
16	1,746	2,120	2,583	2,921	4,015
17	1,740	2,110	2,567	2,898	3,965
18	1,734	2,103	2,552	2,878	3,922
19	1,729	2,093	2,539	2,861	3,883
20	1,725	2,086	2,528	2,845	3,850
21	1,721	2,080	2,518	2,831	3,819
22	1,717	2,074	2,508	2,819	3,792
23	1,714	2,069	2,500	2,807	3,767
24	1,711	2,064	2,492	2,797	3,745
25	1,708	2,060	2,485	2,787	3,725
26	1,706	2,056	2,479	2,779	3,707
27	1,703	2,052	2,473	2,771	3,690
28	1,701	2,048	2,467	2,763	3,674
29	1,699	2,045	2,462	2,756	3,659
30	1,697	2,042	2,457	2,750	3,646
40	1,684	2,021	2,423	2,704	3,551
60	1,671	2,000	2,390	2,660	3,460
120	1,658	1,980	2,358	2,617	3,373
∞	1,645	1,960	2,326	2,576	3,291

Значения q , удовлетворяющие неравенству $P(s - \varepsilon < \sigma < s + \varepsilon) = \alpha$

k	$\alpha = 0,95$	$\alpha = 0,99$	$\alpha = 0,999$
1	14,947	77,786	397,947
2	3,415	9,978	30,802
3	1,914	4,111	10,112
4	1,372	2,670	5,637
5	1,091	2,006	3,877
6	0,915	1,624	2,977
7	0,797	1,377	2,424
8	0,711	1,203	2,055
9	0,645	1,076	1,796
10	0,593	0,977	1,599
12	0,515	0,833	1,327
14	0,460	0,733	1,145
16	0,418	0,659	1,013
18	0,385	0,602	0,916
20	0,358	0,556	0,838
25	0,310	0,473	0,700
30	0,276	0,416	0,609
35	0,253	0,375	0,544
40	0,234	0,343	0,494
45	0,219	0,318	0,455
50	0,209	0,297	0,424
60	0,186	0,266	0,375
70	0,172	0,242	0,339
80	0,160	0,224	0,311
90	0,150	0,209	0,289
100	0,142	0,196	0,271
150	0,115	0,159	0,220
200	0,099	0,135	0,184
250	0,089	0,120	0,162
300	0,081	0,109	0,146
350	0,075	0,100	0,134
400	0,070	0,093	0,124
450	0,066	0,088	0,116
500	0,062	0,083	0,110

Значения χ^2 , удовлетворяющие равенству $\delta = \int_0^{\chi^2} p(\chi^2, k) d\chi^2$

Число степеней свободы	Вероятность $P(\chi^2 > \chi_q^2)$							
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30
1	0,00016	0,006	0,0039	0,016	0,064	0,148	0,455	1,07
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,66
4	0,30	0,43	0,71	1,06	1,65	2,19	3,36	4,9
5	0,55	0,75	1,14	1,61	2,34	3,00	4,35	6,1
6	0,87	1,13	1,63	2,20	3,07	3,83	5,35	7,2
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,4
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,5
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,7
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8
11	3,1	3,6	4,6	5,6	7,0	8,1	10,3	12,9
12	3,6	4,2	5,2	6,3	7,8	9,0	11,3	14,0
13	4,1	4,8	5,9	7,0	8,6	9,9	12,3	15,1
14	4,7	5,4	6,6	7,8	9,5	10,8	13,3	16,2
15	5,2	6,0	7,3	8,5	10,3	11,7	14,3	17,3
16	5,8	6,6	8,0	9,3	11,2	12,6	15,3	18,4
17	6,4	7,3	8,7	10,1	12,0	13,5	16,3	19,5
18	7,0	7,9	9,4	10,9	12,9	14,4	17,3	20,6
19	7,6	8,6	10,1	11,7	13,7	15,4	18,3	21,7
20	8,3	9,2	10,9	12,4	14,6	16,3	19,3	22,8

Число степеней свободы	Вероятность $P(\chi^2 > \chi_q^2)$				
	0,20	0,10	0,05	0,02	0,01
1	1,64	2,7	3,8	5,4	6,6
2	3,22	4,6	6,0	7,8	9,2
3	4,64	6,3	7,8	9,8	11,3
4	6,0	7,8	9,5	11,7	13,3
5	7,3	9,2	11,1	13,4	15,1
6	8,6	10,6	12,6	15,0	16,8
7	9,8	12,0	14,1	16,6	18,5
8	11,0	13,4	15,5	18,2	20,1
9	12,2	14,7	16,9	19,7	21,7
10	13,4	16,0	18,3	21,2	23,2
11	14,6	17,3	19,7	22,6	24,7
12	15,8	18,5	21,0	24,1	26,2
13	17,0	19,8	22,4	25,5	27,7
14	18,2	21,1	23,7	26,9	29,1
15	19,3	22,3	25,0	28,3	30,6
16	20,5	23,5	26,3	29,6	32,0
17	21,6	24,8	27,6	31,0	33,4
18	22,8	26,0	28,9	32,3	34,8
19	23,9	27,2	30,1	33,7	36,2
20	25,0	28,4	31,4	35,0	37,6

Основные характеристики гамма-распределения для

$$\alpha = -\frac{1}{2}, 0, +\frac{1}{2}, 1, 2, \dots, 15$$

α	$\bar{x} = \sigma_x^2 = \alpha + 1$	C_v	x_0	$\bar{x} - x_0$	$P(\bar{x})$
-0,5	0,5	1,41	0,231	0,269	0,683
0,0	1,0	1,00	0,694	0,306	0,632
+0,5	1,5	0,82	1,184	0,316	0,608
1,0	2,0	0,71	1,678	0,322	0,594
1,5	2,5	0,63	2,174	0,326	0,584
2,0	3,0	0,57	2,673	0,327	0,577
3,0	4,0	0,50	3,671	0,329	0,567
4,0	5,0	0,45	4,670	0,330	0,560
5,0	6,0	0,41	5,669	0,331	0,554
6,0	7,0	0,37	6,668	0,332	0,550
7,0	8,0	0,35	7,667	0,333	0,547
8,0	9,0	0,33	8,667	0,333	0,544
9,0	10,0	0,32	9,666	0,334	0,542
10,0	11,0	0,30	10,666	0,334	0,540
11,0	12,0	0,28	11,666	0,344	0,538
12,0	13,0	0,28	12,666	0,344	0,537
13,0	14,0	0,27	13,666	0,344	0,536
14,0	15,0	0,26	14,666	0,344	0,534
15,0	16,0	0,25	15,666	0,344	0,533

**ОТЫСКАНИЕ ФОРМЫ СВЯЗИ МЕЖДУ
СЛУЧАЙНЫМИ ВЕЛИЧИНАМИ МЕТОДОМ КВАНТИЛЕЙ**

Коррелятивную зависимость между двумя случайными величинами x и y обычно характеризуют с помощью кривых регрессии y по x :

$$y = \bar{y}_x = \varphi(x; a, b, \dots) \quad (1)$$

или x по y :

$$x = \bar{x}_y = \psi(y; a', b', \dots), \quad (2)$$

параметры которых a, b, \dots или a', b', \dots определяются по методу наименьших квадратов, т. е. из условия минимума суммы или среднего значения квадратов отклонений кривой регрессии от наблюдаемых точек

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \quad (3)$$

по ординатам

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n [y_i - \varphi(x_i; a, b, \dots)]^2 = \min \quad (4_1)$$

или по абсциссам

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n [x_i - \psi(y_i; a', b', \dots)]^2 = \min. \quad (4_2)$$

Кривые регрессии (1) и (2) характеризуют зависимость условных средних значений одной из величин (\bar{y}_x или \bar{x}_y) от данного значения другой величины (x или y).

Теснота связи (или разброс эмпирических точек) относительно кривых регрессии оценивается по величине корреляционного отношения

$$\rho_y = \sqrt{1 - \frac{s_y^2}{\sigma_y^2}} \quad \text{или} \quad \rho_x = \sqrt{1 - \frac{s_x^2}{\sigma_x^2}}, \quad (5)$$

где

$$\sigma_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2; \quad \sigma_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (6)$$

являются дисперсиями или средними значениями квадратов отклонений y и x относительно их средних значений

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i; \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (7)$$

s_y^2 и s_x^2 — условные дисперсии или средние значения квадратов отклонений y и x относительно кривых регрессии (1) и (2), характеризующиеся выражениями (4₁) и (4₂).

При наличии однозначной функциональной зависимости $y = \varphi(x)$ между x и y все эмпирические точки (x_i, y_i) располагаются вдоль кривой зависимости $y = \varphi(x)$, и поэтому

$$s_x^2 = s_y^2 = 0; \quad \rho_x = \rho_y = 1.$$

При отсутствии зависимости между x и y кривые регрессии (1) и (2) представляют взаимно перпендикулярные прямые линии, параллельные осям координат $\overline{y_x} = \overline{y}$, $\overline{x_y} = \overline{x}$, и поэтому

$$s_y^2 = \sigma_y^2; \quad s_x^2 = \sigma_x^2; \quad \rho_y = \rho_x = 0.$$

Таким образом, корреляционное отношение может изменяться в пределах от нуля до единицы.

При криволинейной корреляции корреляционное отношение ρ_y и ρ_x всегда больше коэффициента корреляции r_{xy} , и только в случае прямолинейной корреляции имеет место равенство $\rho_y = \rho_x = |r_{xy}|$.

Характеристика коррелятивной зависимости с помощью кривых регрессии имеет ряд существенных недостатков.

1. Вместо единой кривой связи, скрытой за корреляцией, определяются две кривые связи (1) или (2).

2. Приходится заранее задаваться тем или иным видом уравнения кривой регрессии, ориентируясь в большинстве случаев на глазомерную оценку формы связи по расположению эмпирических точек на графике коррелятивной зависимости.

3. Определение параметров кривых регрессии (a, b, \dots или a', b', \dots) по методу наименьших квадратов сводится к решению сложной системы уравнений:

$$\frac{\partial s_y^2}{\partial a} = 0; \quad \frac{\partial s_y^2}{\partial b} = 0, \dots$$

или

$$\frac{\partial s_x^2}{\partial a'} = 0; \quad \frac{\partial s_x^2}{\partial b'} = 0, \dots,$$

которая решается сравнительно просто только в тех случаях, когда все уравнения системы линейные относительно искомых параметров.

Поэтому на практике (в гидрологии, климатологии и других прикладных науках) часто прибегают к графическому глазомерному построению кривой связи $y = \varphi(x)$ посередине поля эмпирических точек, как медианной линии связи, выше и ниже которой располагается примерно по 50% от всех эмпирических точек на графике связи.

Очевидно, что такой способ построения кривой связи субъек-

тивен, особенно при малой тесноте связи, т. е. при большом разбросе эмпирических точек относительно искомой кривой связи.

Учитывая принципиальные и практические недостатки метода регрессий, акад. Б. И. Срезневский в 1889 г. предложил определять единую кривую связи между двумя коррелятивно зависимыми случайными величинами из условия равенства повторяемости обеих величин [7, 8], или, другими словами, по связи между равнообеспеченными значениями, или квантилями, обеих величин.

Эта прогрессивная идея, к сожалению, долгое время не получала должного конструктивного развития и практического применения и была забыта. Свидетельством этого может служить приближенный способ удлинения статистических рядов по равнообеспеченным значениям двух коррелятивно зависимых величин, предложенный в 1935 г. Г. П. Ивановым [6] без ссылок на предложения Б. И. Срезневского [7, 8] и без теоретического обоснования.

Позднее, в работах А. А. Шепелевского [10] и О. А. Дроздова [5], делается попытка дать теоретико-вероятностную схему и приемы разыскания единой функциональной зависимости, скрытой за корреляцией.

В работе автора [1] рассмотрены уравнения и свойства кривых связи равнообеспеченных значений двух случайных величин и предложен простой графоаналитический способ приведения к длительному периоду наблюдений параметров кривых распределения с помощью кривых связи равнообеспеченных значений.

Напомним, что кривую связи равнообеспеченных значений двух случайных величин x и y можно записать двумя символическими выражениями:

1) при наличии прямой связи между x и y

$$p(y) = p(x) \text{ или } y_p = \varphi(x_p); \quad (8)$$

2) при обратной связи между x и y

$$p(y) = 1 - p(x) = p^*(x) \text{ или } y_p = \varphi(x_{p^*}) = \varphi(x_{1-p}), \quad (8')$$

где $p(x)$ и $p(y)$ — вероятности превышения (или обеспеченности) данных значений x и y ; $p^*(x) = 1 - p(x)$ — вероятность не превышения данного значения x ; x_p и y_p — значения x и y , вероятность превышения (или обеспеченность) которых равна p . Значения x_{p^*} или y_{p^*} в математической статистике принято называть квантилями.

При наличии однозначной возрастающей или убывающей функциональной зависимости $y = \varphi(x)$ между x и y кривая связи равнообеспеченных значений $y_p = \varphi(x_p)$ или $y_p = \varphi(x_{p^*})$ в точности отражает функциональную зависимость $y = \varphi(x)$, и наоборот.

Во многих случаях коррелятивно зависимые величины x и y

таковы, что их нормированные отклонения от средних значений характеризуются одинаковой кривой распределения, т. е.

1) при наличии прямой связи между x и y

$$t_p = \frac{y_p - \bar{y}}{\sigma_y} = \frac{x_p - \bar{x}}{\sigma_x} = f_1(p); \quad (9)$$

2) при обратной связи между x и y

$$t_p = \frac{y_p - \bar{y}}{\sigma_y} = - \frac{x_{1-p} - \bar{x}}{\sigma_x} = f_2(p), \quad (9')$$

где \bar{x} и \bar{y} — средние значения x и y ; σ_x и σ_y — средние квадратические отклонения x и y от средних значений.

Если коррелятивно зависимые величины x и y характеризуются одним и тем же видом кривой распределения (например, биномиальным) с тремя стандартными параметрами \bar{x} , σ_x , C_{sx} и \bar{y} , σ_y , C_{sy} , то их нормированные отклонения от средних значений являются функциями от вероятности превышения p и от коэффициентов асимметрии C_{sx} и C_{sy} (см., например, таблицу Фостера-Рыбкина):

$$\begin{aligned} \frac{y_p - \bar{y}}{\sigma_y} &= \Phi(p, C_{sy}) \quad \text{при } C_{sy} > 0; \\ \frac{x_p - \bar{x}}{\sigma_x} &= \Phi(p, C_{sx}) \quad \text{при } C_{sx} > 0; \\ \frac{x_{1-p} - \bar{x}}{\sigma_x} &= -\Phi(p, |C_{sx}|) \quad \text{при } C_{sx} < 0. \end{aligned}$$

Отсюда вытекают уравнения кривых связи равнообеспеченных значений:

1) при $C_{sy} > 0$ и $C_{sx} > 0$, т. е. при прямой связи между x и y ,

$$y_p - \bar{y} = \frac{\sigma_y}{\sigma_x} \frac{\Phi(p, C_{sy})}{\Phi(p, C_{sx})} (x_p - \bar{x}); \quad (10)$$

2) при $C_{sy} > 0$ и $C_{sx} < 0$, т. е. при обратной связи между x и y ,

$$y_p - \bar{y} = - \frac{\sigma_y}{\sigma_x} \frac{\Phi(p, C_{sy})}{\Phi(p, |C_{sx}|)} (x_{1-p} - \bar{x}). \quad (10')$$

В частности, при равенстве абсолютных величин коэффициентов асимметрии (т. е. при $|C_{sy}| = |C_{sx}|$), и только при этом условии, кривые связи равнообеспеченных значений (10) и (10') превращаются в прямые линии вида (9) и (9'):

$$y_p - \bar{y} = \frac{\sigma_y}{\sigma_x} (x_p - \bar{x}) \quad \text{при } C_{sy} = C_{sx}; \quad (11)$$

$$y_p - \bar{y} = - \frac{\sigma_y}{\sigma_x} (x_{1-p} - \bar{x}) \quad \text{при } C_{sy} = -C_{sx}; \quad (11')$$

проходящие через центр распределения (\bar{x}, \bar{y}) , с тангенсами углов наклона к оси x , равными соответственно

$$\operatorname{tg} \alpha = \frac{y_p - \bar{y}}{x_p - \bar{x}} = \frac{\sigma_y}{\sigma_x}, \quad (12)$$

$$\operatorname{tg} \beta = \frac{y_p - \bar{y}}{x_{1-p} - \bar{x}} = -\frac{\sigma_y}{\sigma_x}. \quad (12')$$

Докажем, что кривые связи равнообеспеченных значений

$$y_p = \varphi(x_p) \quad \text{или} \quad x_p = \varphi^{-1}(y_p)$$

наименее уклоняются от наблюдаемых точек (x_i, y_i) как по орди-

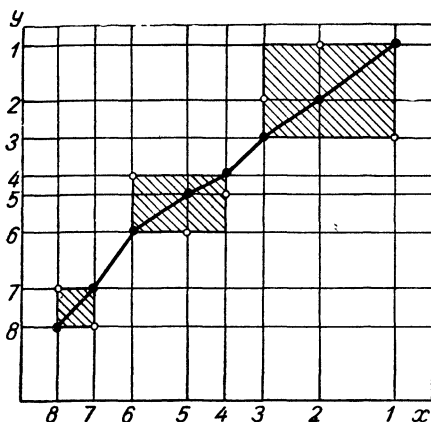


Рис. 1. Отклонение наблюдаемых точек x_i, y_i по абсциссам и ординатам от кривой связи равнообеспеченных величин $y_p = \varphi(x_p)$ (характеризуется заштрихованными площадями элементарных прямоугольников).

натам $[y_i - \varphi(x_i)]$, так и по абсциссам $[\varphi^{-1}(y_i) - x_i]$, т. е. среднее значение произведений уклонений по ординатам и абсциссам является минимальным:

$$\bar{s}_{xy} = \frac{1}{n-1} \sum_{i=1}^n [y_i - \varphi(x_i)] [\varphi^{-1}(y_i) - x_i] = \min. \quad (13)$$

Геометрически величина \bar{s}_{xy} характеризуется средним значением площадей всех элементарных «прямоугольников отклонений»:

$$s_i = [y_i - \varphi(x_i)] [\varphi^{-1}(y_i) - x_i] \quad (14)$$

(заштрихованных на рис. 1) с вершинами в наблюдаемых точках (x_i, y_i) и диагоналями вдоль кривой связи.

Из выражения (14) и на рис. 1 видно, что при любом местоположении точки (x_i, y_i) относительно кривой связи $y = \varphi(x)$ (т. е. над кривой или под кривой) площадь каждого элементарного прямоугольника s_i сохраняет положительный знак при прямой связи и отрицательный при обратной связи между x и y . Благодаря этому важному свойству сохранения знака среднее значение площадей элементарных прямоугольников \bar{s}_{xy} может служить мерой уклонения кривой связи от наблюдаемых точек по ординатам и абсциссам, или, короче говоря, мерой разброса связи. Следовательно, из условия минимума меры разброса, т. е. среднего значения площадей всех элементарных прямоугольников отклонений

$$\bar{s}_{xy} = \frac{1}{n-1} \sum_{i=1}^n s_i = \min$$

можно определять параметры (a, b, \dots) искомой кривой связи $y = \varphi(x; a, b, \dots)$, наименее уклоняющейся от наблюдаемых точек как по ординате, так и по абсциссе.

По аналогии с «методом наименьших квадратов», в котором параметры искомой кривой связи $y = \varphi(x; a, b, \dots)$ определяются из условия минимума среднего значения квадратов уклонений кривой связи от наблюдаемых точек только по ординатам (4₁) или только по абсциссам (4₂), метод определения параметров искомой кривой связи $y = \varphi(x; a, b, \dots)$ из условия ее наилучшего приближения к наблюдаемым точкам как по ординатам, так и по абсциссам (13) будем называть «методом наименьших прямоугольников».

Рассмотрим сначала наиболее простой и часто встречающийся на практике случай линейного сглаживания коррелятивной связи между наблюдаемыми значениями x_i и y_i в виде прямой линии

$$y = \varphi(x; a, b) = ax + b \quad (15)$$

или

$$x = \varphi^{-1}(y; a, b) = \frac{1}{a} y - \frac{b}{a}.$$

Определим параметры a и b сглаживающей прямой по методу наименьших прямоугольников, т. е. из условия минимума меры разброса

$$\bar{s}_{xy} = \frac{1}{n-1} \sum_{i=1}^n (y_i - ax_i - b) \left(\frac{1}{a} y_i - \frac{b}{a} - x_i \right) = \min \quad (16)$$

или в упрощенном виде

$$\bar{s}_{xy} = \frac{1}{n-1} \cdot \frac{1}{a} \sum_{i=1}^n (y_i - ax_i - b)^2 = \min. \quad (16')$$

Для этого находим частные производные от \bar{s}_{xy} по a и b и приравниваем их к нулю:

$$\begin{aligned} \frac{\partial \bar{s}_{xy}}{\partial a} = & -\frac{1}{n-1} \left[\frac{1}{a^2} \sum_{i=1}^n (y_i - ax_i - b)^2 + \right. \\ & \left. + \frac{2}{a} \sum_{i=1}^n (y_i - ax_i - b) x_i \right] = 0; \end{aligned} \quad (17)$$

$$\frac{\partial \bar{s}_{xy}}{\partial b} = -\frac{1}{n-1} \cdot \frac{2}{a} \sum_{i=1}^n (y_i - ax_i - b) = 0. \quad (17')$$

Из последнего уравнения находим

$$\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i - nb = 0,$$

откуда

$$b = \frac{1}{n} \sum_{i=1}^n y_i - a \frac{1}{n} \sum_{i=1}^n x_i = \bar{y} - a\bar{x}. \quad (18)$$

Подставляя полученное выражение для b в уравнение (17), имеем

$$\sum_{i=1}^n [y_i - \bar{y} - a(x_i - \bar{x})]^2 + 2a \sum_{i=1}^n [y_i - \bar{y} - a(x_i - \bar{x})] x_i = 0.$$

Из этого уравнения находим величину искомого параметра:

$$a^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma_y^2}{\sigma_x^2}.$$

или

$$a = \pm \frac{\sigma_y}{\sigma_x}, \quad (19)$$

где знак плюс соответствует прямой связи, а знак минус — обратной связи между x и y .

Далее, подставляя выражение (19) в уравнение (18), получаем формулу для определения второго параметра:

$$b = \bar{y} \pm \frac{\sigma_y}{\sigma_x} \bar{x}, \quad (20)$$

где знак минус берется при $a > 0$ и знак плюс — при $a < 0$.

Наконец, подставляя полученные выражения (19) и (20) для параметров a и b в уравнения (15) и (16'), имеем уравнение искомой прямой, наименее уклоняющейся от наблюдаемых точек по ординате и по абсциссе:

$$y - \bar{y} = \pm \frac{\sigma_y}{\sigma_x} (x - \bar{x}), \quad (21)$$

а также явное выражение для минимума меры разброса, т. е. среднего значения произведений отклонений прямой (21) от наблюдаемых точек по ординате и по абсциссе:

$$\bar{s}_{xy} = \pm \frac{\sigma_y}{\sigma_x} \frac{1}{n-1} \sum_{i=1}^n \left[y_i - \bar{y} \mp \frac{\sigma_y}{\sigma_x} (x_i - \bar{x}) \right]^2, \quad (22)$$

где верхние знаки соответствуют прямой связи между x и y , а нижние — обратной связи.

Раскрывая в выражении (22) квадратные скобки и производя почленное суммирование с учетом выражений (6) для σ_x^2 , σ_y^2 , а также известного выражения для коэффициента корреляции между x и y

$$r_{xy} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y}, \quad (23)$$

получаем простое соотношение между мерой разброса \bar{s}_{xy} и коэффициентом корреляции:

$$\bar{s}_{xy} = 2 \sigma_x \sigma_y (\pm 1 - r_{xy}) \quad (24)$$

или

$$r_{xy} = \pm 1 - \frac{\bar{s}_{xy}}{2 \sigma_x \sigma_y}, \quad (25)$$

где знак плюс соответствует прямой связи между x и y , а знак минус — обратной связи.

В частности, из выражений (24) и (25) следует, что

при $r_{xy} = \pm 1 \quad \bar{s}_{xy} = 0,$ (26)

при $r_{xy} = 0 \quad \bar{s}_{xy} = \pm 2 \sigma_x \sigma_y.$ (27)

Из приведенного выше вывода уравнения (21) видно, что, какова бы ни была коррелятивная зависимость между x и y (пря-

молинейная или криволинейная), прямая линия (21) по сравнению со всякой другой прямой наименее уклоняется от наблюдаемых точек (x_i, y_i) как по абсциссам, так и по ординатам.

Заметим для сравнения, что прямая регрессии y по x

$$y - \bar{y} = r_{xy} \frac{\sigma_y}{\sigma_x} (x - \bar{x}) \quad (28)$$

и прямая регрессии x по y

$$y - \bar{y} = \frac{1}{r_{xy}} \frac{\sigma_y}{\sigma_x} (x - \bar{x}) \quad (29)$$

наименее уклоняются от наблюдаемых точек (x_i, y_i) соответственно только по ординатам или только по абсциссам.

При этом условные дисперсии s_y^2 и s_x^2 , т. е. средние значения квадратов уклонения наблюдаемых точек от линий регрессии (28) и (29) соответственно по ординате и абсциссе равны

$$s_y^2 = \frac{1}{n-1} \sum_{i=1}^n \left[y_i - \bar{y} - r_{xy} \frac{\sigma_y}{\sigma_x} (x_i - \bar{x}) \right]^2,$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n \left[x_i - \bar{x} - r_{xy} \frac{\sigma_x}{\sigma_y} (y_i - \bar{y}) \right]^2.$$

Раскрывая в этих выражениях квадратные скобки и производя почленное суммирование с учетом выражений (6) и (23) для σ_x^2 , σ_y^2 и r_{xy} , получаем простые соотношения между условными и безусловными дисперсиями (s_y^2 , s_x^2 и σ_y^2 , σ_x^2) при данном значении коэффициента корреляции r_{xy} :

$$s_y^2 = \sigma_y^2 (1 - r_{xy}^2), \quad s_x^2 = \sigma_x^2 (1 - r_{xy}^2). \quad (30)$$

Отсюда вытекают выражения вида (5):

$$r_{xy} = \pm \sqrt{1 - \frac{s_y^2}{\sigma_y^2}} = \pm \sqrt{1 - \frac{s_x^2}{\sigma_x^2}}. \quad (31)$$

Выражения (30) и (31), как видим, отличаются по своей структуре от похожих по смыслу выражений (24) и (25).

Уравнение (21) в точности совпадает с уравнением линии связи равнообеспеченных значений (11) или (11'). Это свидетельствует о том, что линия связи равнообеспеченных значений (21) по сравнению со всякой другой прямой наименее уклоняется от наблюдаемых точек как по ординатам, так и по абсциссам.

Из сопоставления уравнения (21) с уравнениями прямых регрессий (28) и (29) видно, что все три прямые проходят через

центр распределения (\bar{x}, \bar{y}) с тангенсами углов наклона к оси x , равными соответственно

$$\operatorname{tg} \alpha = \pm \frac{\sigma_y}{\sigma_x}, \quad \operatorname{tg} \alpha_1 = r_{xy} \frac{\sigma_y}{\sigma_x}, \quad \operatorname{tg} \alpha_2 = \frac{1}{r_{xy}} \frac{\sigma_y}{\sigma_x}. \quad (32)$$

Из этих выражений вытекает простое соотношение между тангенсами углов наклона прямых (21), (28) и (29):

$$\operatorname{tg} \alpha = \pm \sqrt{\operatorname{tg} \alpha_1 \operatorname{tg} \alpha_2}, \quad (33)$$

т. е. тангенс угла наклона линии связи равнообеспеченных значений равен среднему геометрическому значению тангенсов углов наклона линий регрессии.

Перейдем теперь к общему решению задачи по отысканию формы связи между случайными величинами в случае криволинейной зависимости.

Будем рассматривать в дальнейшем коррелятивную зависимость между случайными величинами x и y на плоскости новых координат:

$$v = p(y), \quad u = p(x) \quad \text{или} \quad u^* = p^*(x) = 1 - p(x), \quad (34)$$

т. е. в квадрате вероятностей

$$0 \leq v \leq 1, \quad 0 \leq u \leq 1 \quad \text{или} \quad 0 \leq u^* \leq 1. \quad (35)$$

При этом эмпирические точки

$$v_i = p(y_i), \quad u_i = p(x_i) \quad \text{или} \quad u_i^* = p^*(x_i) = 1 - p(x_i),$$

определяемые с помощью кривых обеспеченности $p(x)$ и $p(y)$ по соответственным наблюдаемым значениям x_i и y_i , располагаются равномерно по каждой оси координат u, v или u^*, v в пределах от нуля до единицы.

Кроме того, любая монотонная (т. е. прямая или обратная) криволинейная коррелятивная зависимость между x и y превращается в прямую линейную коррелятивную зависимость между новыми величинами u и v или u^* и v .

Отсюда видно, что преобразование координат с помощью кривых обеспеченности вида (34) имеет преимущество перед другими видами преобразования, например, перед логарифмическим преобразованием, в результате которого достигается спрямление зависимостей только степенного или показательного вида: $y = ax^n$ или $y = ae^{\beta x}$.

В дальнейшем эмпирические точки (u_i, v_i) будем определять непосредственно по эмпирическим значениям обеспеченности наблюдаемых соответственных значений [3, 4], т. е. по формулам:

$$p(x_m) = \frac{m - 0,3}{n + 0,4}; \quad p(y_{m'}) = \frac{m' - 0,3}{n + 0,4}, \quad (36)$$

где n — общее число наблюдаемых точек (x_i, y_i) ; m и m' — порядковые номера наблюдаемых соответственных значений x_i, y_i

после расположения членов каждого ряда x_1, x_2, \dots, x_n и y_1, y_2, \dots, y_n в убывающем (или возрастающем) порядке.

Формула вида (36) вытекает из следующих теоретических соображений.

Будем рассматривать явление превышения или непревышения данного значения случайной величины $x = \text{const}$ в большой серии N случайных выборок, объемом n каждая:

$$\left. \begin{array}{l} x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)} \\ x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)} \\ \vdots \\ x_1^{(N)}, x_2^{(N)}, \dots, x_n^{(N)} \end{array} \right\} \quad (37)$$

Будем предполагать, что в каждой данной выборке объемом n при каждом испытании вероятность превышения данного значения $x = \text{const}$ постоянна и равна p , но может случайно изменяться для разных N выборок в пределах от нуля до единицы.

Такую серию случайных выборок можно получить, например, если рассматривать максимальные в году расходы воды, наблюдаемые в одном и том же расчетном створе реки за N смежные или произвольно выбранные n -летние периоды.

Для каждой данной выборки объемом n по теореме умножения вероятностей превышения ($p = \text{const}$) и вероятностей непревышения ($1-p$), а также по теореме сложения вероятностей всех возможных сочетаний из i превышений и $n-i$ непревышений находим вероятность превышения данного значения x при n испытаниях ровно i раз:

$$\varphi_i(p) = \frac{n(n-1)\dots(n-i+1)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot i} p^i (1-p)^{n-i}. \quad (38)$$

Эта вероятность представляет $i+1$ -й член в разложении бинома Ньютона

$$[(1-p) + p]^n = (1-p)^n + n(1-p)^{n-1}p + \dots + p^n = 1, \quad (39)$$

состоящего из $n+1$ членов вида (38) для значений $i = 0, 1, 2, \dots, n$.

Суммируя вероятности $\varphi_i(p)$ для значений $i = 0, 1, 2, \dots, m-1$, мы получим, очевидно, суммарную вероятность

$$P_m(p) = \varphi_0(p) + \varphi_1(p) + \dots + \varphi_{m-1}(p), \quad (40)$$

характеризующую вероятность превышения данного значения x менее m раз среди n элементов выборки x_1, x_2, \dots, x_n .

В частности, при значениях $m = 1$ и $m = n$, т. е. для наибольшего и наименьшего членов данного ряда x_1 и x_n , функции распределения вероятностей всех возможных наибольших и наименьших значений $x_1^1, x_1^2, \dots, x_1^{(N)}$ и $x_n^1, x_n^2, \dots, x_n^{(N)}$ характеризуются соответственно выражениями

$$P_1(p) = (1-p)^n; \quad P_n(p) = 1 - p^n. \quad (41)$$

Для разных выборок с изменением вероятности превышения p в пределах от нуля до единицы функции распределения вероятностей $P_m(p)$ изменяются, наоборот, в пределах от единицы до нуля, т. е. представляют убывающие функции от вероятности превышения p .

При наличии только одной выборки объемом n (или, другими словами, при наличии наблюдений только за один n -летний период) наиболее естественно предполагать, что наблюдаемые значения x_1, x_2, \dots, x_n занимают наиболее положение среди значений x_m того же порядкового номера всех других выборок (или всех других n -летних периодов наблюдений), т. е. полагать

$$P_1(p) = P_2(p) = \dots = P_n(p) = 0,5 \quad (42)$$

или в явном виде

$$P_1(p_1) = (1 - p_1)^n = 0,5; \quad (42_1)$$

$$P_2(p_2) = (1 - p_2)^n + n(1 - p_2)^{n-1}p_2 = 0,5; \quad (42_2)$$

.....

$$P_n(p_n) = 1 - p_n^n = 0,5. \quad (42_n)$$

Нетрудно убедиться на конкретных числовых примерах, что n уравнений от (42₁) до (42 _{n}) с n неизвестными p_1, p_2, \dots, p_n можно заменить (по предложению Н. Н. Чегодаева) с достаточной для практики точностью приближенной интерполяционной формулой вида (36).

Таким образом, формула (36) основана на предположении, что члены данного вариационного ряда x_1, x_2, \dots, x_n , расположенные в убывающем порядке, занимают медианное положение среди членов того же порядкового номера всех других вариационных рядов, получаемых в результате большой серии случайных выборок, объемом n каждая (например, в результате большого числа неперекрывающихся, т. е. независимых n -летних периодов наблюдений).

Заметим попутно, что среднее значение вероятностей превышения m -го члена вариационного ряда в большой серии N случайных выборок, объемом n каждая, характеризуется выражением

$$\bar{p}_m = \frac{1}{N} \sum_{i=1}^N p(x_m^{(i)}) = - \int_0^1 p dP_m(p), \quad (43)$$

где знак минус перед интегралом поставлен потому, что дифференциал $dP_m(p)$ от убывающей функции $P_m(p)$ имеет отрицательный знак.

Применяя к интегралу (43) правило интегрирования по частям и учитывая, что произведение $pP_m(p)$ равно нулю при значениях $p = 0$ и $p = 1$, получаем простую формулу:

$$\bar{p}_m = \int_0^1 P_m(p) dp = \frac{m}{n+1}. \quad (44)$$

В частности,

$$\bar{p}_1 = \int_0^1 (1-p)^n dp = \frac{1}{n+1}; \quad (44_1)$$

$$\bar{p}_2 = \int_0^1 [(1-p)^n + n(1-p)^{n-1}p] dp = \frac{2}{n+1}; \quad (44_2)$$

.....

$$\bar{p}_n = \int_0^1 (1-p^n) dp = \frac{n}{n+1}. \quad (44_n)$$

Согласно выражениям (41) и (44₁), обеспеченность наибольшего члена ряда x_1 среди наибольших членов $x_1^1, x_1^2, \dots, x_1^{(N)}$ в N случайных выборках равна

$$P_1(\bar{p}_1) = (1 - \bar{p}_1)^n = \left(1 - \frac{1}{n+1}\right)^n \approx 0,38, \quad (45)$$

т. е. меньше медианы 0,5.

Наоборот, согласно выражениям (41) и (44_n), обеспеченность наименьшего члена ряда x среди наименьших членов $x_n^1, x_n^2, \dots, x_n^{(N)}$ в N случайных выборках равна

$$P_n(\bar{p}_n) = 1 - (\bar{p}_n)^n = 1 - \left(\frac{n}{n+1}\right)^n \approx 0,62, \quad (46)$$

т. е. больше медианы 0,5.

Именно поэтому формула вида (36), основанная на предположении (42), заслуживает предпочтения перед формулой вида (44).

Теперь докажем, что кривая связи равнообеспеченных значений, характеризуемая диагоналями квадрата вероятностей

$$v = u \quad \text{или} \quad v = u^* = 1 - u,$$

т. е.

$$p(y) = p(x) \quad \text{или} \quad p(y) = p^*(x) = 1 - p(x), \quad (47)$$

наименее уклоняется от наблюдаемых точек (u_i, v_i) по обеим координатам u и v по сравнению со всякой другой прямой.

Действительно, искомая прямая, наименее уклоняющаяся от наблюдаемых точек по ординатам и абсциссам, выражается уравнением вида (21):

$$v - \bar{v} = \pm \frac{\sigma_v}{\sigma_u} (u - \bar{u}), \quad (48)$$

где $\bar{v} = \bar{u}$ и $\sigma_v = \sigma_u$ — средние и среднеквадратические значения эмпирических обеспеченностей, т. е.

$$\bar{v} = \bar{u} = \frac{1}{n} \sum_{m=1}^n u_m = \frac{1}{n} \sum_{m=1}^n p(x_m) = \frac{1}{n} \sum_{m=1}^n \frac{m-0,3}{n+0,4} = 0,5; \quad (49)$$

$$\begin{aligned} \sigma_v^2 = \sigma_u^2 &= \frac{\sum_{m=1}^n (u_m - \bar{u})^2}{n-1} = \frac{\sum_{m=1}^n \left(\frac{m-0,3}{n+0,4} - 0,5\right)^2}{n-1} = \\ &= \frac{n(n+1)}{12(n+0,4)^2} \approx \frac{1}{12}. \end{aligned} \quad (50)$$

Заметим, что теоретическая плотность распределения обеспеченности в интервале $(0, \underline{1})$ равна единице, поэтому теоретические значения параметров $v = u$ и $\sigma_v^2 = \sigma_u^2$ характеризуются следующими интегралами:

$$\bar{v} = \bar{u} = \int_0^1 u \, du = 0,5, \quad (49')$$

$$\sigma_v^2 = \sigma_u^2 = \int_0^1 (u - 0,5)^2 \, du = \frac{1}{12}. \quad (50')$$

Подставляя найденные численные значения параметров $\bar{u} = \bar{v} = 0,5$ и $\sigma_u = \sigma_v = \frac{1}{\sqrt{12}}$ в равенство (48), получаем уравнения искомой прямой, наименее уклоняющиеся от наблюдаемых точек (u_i, v_i) по обеим координатам:

при наличии прямой связи между x и y

$$v = u, \quad \text{т. е. } p(y) = p(x),$$

при обратной связи между x и y

$$v = 1 - u, \quad \text{т. е. } p(y) = 1 - p(x),$$

что и требовалось доказать.

Перейдем теперь к вопросу об оценке тесноты связи между величинами $u = p(x)$ и $v = p(y)$.

Нетрудно видеть, что отклонения любой наблюдаемой точки (u_i, v_i) от диагонали квадрата вероятностей $v = u$ или $v = 1 - u$ по обеим координатам равны друг другу. Это значит, что каждый элементарный прямоугольник отклонений (14) при переходе к новой системе координат u, v преобразуется в элементарный квадрат отклонений с площадью

$$s_i = (v_i - u_i)^2 \quad \text{или} \quad s_i^* = -(v_i - u_i^*)^2 = -(v_i + u_i - 1)^2. \quad (51)$$

Среднее значение площадей всех таких элементарных квадратов отклонений

$$\bar{s}_{uv} = \frac{1}{n-1} \sum_{i=1}^n (v_i - u_i)^2 \quad \text{или} \quad \bar{s}_{uv}^* = -\frac{1}{n-1} \sum_{i=1}^n (v_i - u_i^*)^2 \quad (52)$$

может служить мерой разброса связи между величинами $u = p(x)$ и $v = p(y)$, а следовательно, и между исходными случайными величинами x и y , поскольку функции $u = p(x)$ и $v = p(y)$ монотонные.

Действительно, при наличии однозначной возрастающей или убывающей функциональной зависимости между x и y все наблюдаемые точки $v_i = p(y_i)$, $u_i = p(x_i)$ или $u_i^* = 1 - p(x_i)$ лежат на одной из диагоналей квадрата вероятностей и все отклонения $v_i - u_i$ или $v_i - u_i^*$ равны нулю, т. е. в этом случае мера разброса связи равна нулю: $s_{uv} = s_{uv}^* = 0$.

В другом крайнем случае, т. е. для независимых случайных величин x и y , сочетание любых значений u и v в пределах квадрата вероятностей равновозможно и наблюдаемые точки (u_i, v_i) располагаются равномерно по всей площади квадрата вероятностей. При этом теоретическая плотность распределения сочетаний независимых величин $u = p(x)$ и $v = p(y)$ в пределах всего квадрата вероятностей (35) одинаковая и равна единице. Это позволяет для независимых величин u и v вычислить меру разброса связи \bar{s}_{uv} , т. е. среднее значение квадратов

уклонений точек (u, v) от линии равнообеспеченных значений $v = u$ или $v = 1 - u$ как двойной интеграл:

$$\bar{s}_{uv} = -\bar{s}_{uv}^* = \int_0^1 \int_0^1 (v - u)^2 du dv = \frac{1}{6}. \quad (53)$$

По ограниченному числу эмпирических точек (u_i, v_i) , равномерно расположенных в квадрате вероятностей (рис. 2), согласно формуле (52), среднее значение квадратов уклонения точек от диагонали $v = u$ или $v = 1 - u$ также равно $\frac{1}{6}$; например, по четырем точкам (рис. 2 а)

$$\bar{s}_{uv} = \frac{1}{4-1} 2 \left(\frac{3}{4} - \frac{1}{4} \right)^2 = \frac{1}{6},$$

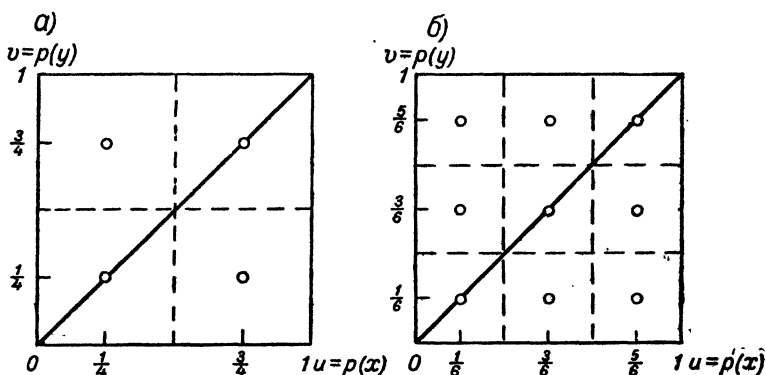


Рис. 2. Схемы равномерного расположения эмпирических точек (u_i, v_i) в квадрате вероятностей.

а — для четырех точек, б — для девяти точек.

по девяти точкам (рис. 2 б)

$$\bar{s}_{uv} = \frac{1}{9-1} 2 \left[\left(\frac{5}{6} - \frac{1}{6} \right)^2 + \left(\frac{5}{6} - \frac{3}{6} \right)^2 + \left(\frac{3}{6} - \frac{1}{6} \right)^2 \right] = \frac{1}{6}.$$

Очевидно, что при наличии коррелятивной зависимости между x и y мера разброса связи \bar{s}_{uv} или \bar{s}_{uv}^* , характеризуемая выражениями (52), заключается в пределах

$$0 \leq \bar{s}_{uv} \leq \frac{1}{6}; \quad -\frac{1}{6} \leq \bar{s}_{uv}^* \leq 0. \quad (54)$$

Эти неравенства вытекают также непосредственно из выражений вида (26) и (27):

$$0 \leq \bar{s}_{uv} \leq 2\sigma_u\sigma_v; \quad -2\sigma_u\sigma_v \leq \bar{s}_{uv}^* \leq 0, \quad (54')$$

где $2\sigma_u\sigma_v = \frac{1}{6}$ согласно равенству (50).

Коэффициент корреляции r_{uv} , отражающий тесноту связи между величинами $u=p(x)$ и $v=p(y)$, характеризуется выражением вида (25):

при прямой связи между x и y

$$r_{uv} = 1 - \frac{\bar{s}_{uv}}{2\sigma_u\sigma_v} = 1 - 6\bar{s}_{uv} = 1 - \frac{6}{n-1} \sum_{i=1}^n [p(y_i) - p(x_i)]^2, \quad (55)$$

при обратной связи между x и y

$$\begin{aligned} r_{uv} &= -1 - \frac{\bar{s}_{uv}^*}{2\sigma_u\sigma_v} = -1 - 6\bar{s}_{uv}^* = \\ &= -1 + \frac{6}{n-1} \sum_{i=1}^n [p(y_i) - p^*(x_i)]^2, \end{aligned} \quad (55')$$

где \bar{s}_{uv}^* — отрицательная величина, а эмпирические обеспеченности $p(y_i)$, $p(x_i)$ или $p^*(x_i) = 1 - p(x_i)$ наблюдаемых соответственных значений x_i и y_i определяются по формулам вида (36).

В качестве примера в табл. 1 приведена схема вычисления коэффициента корреляции r_{uv} между обеспеченностями модулей годового стока р. Туры у г. Тюмень (x л/сек. с 1 км²) и р. Лобвы у р. п. Лобва (y л/сек. с 1 км²) за совместный 19-летний период гидрометрических наблюдений (1935—1953 гг.).

Графики коррелятивной зависимости между эмпирическими значениями x_i и y_i , а также между эмпирическими обеспеченностями $u_i=p(x_i)$ и $v_i=p(y_i)$ представлены на рис. 3 а, б.

По вычисленной в табл. 1 сумме квадратов уклонов

$$S_{uv} = \sum_{i=1}^{i=19} [p(x_i) - p(y_i)]^2 = 1,152.$$

Согласно формуле (55), находим коэффициент корреляции:

$$r_{uv} = 1 - \frac{6S_{uv}}{n-1} = 1 - \frac{6 \cdot 1,152}{18} = 0,62.$$

Полученный коэффициент корреляции $r_{uv}=0,62$ близок по величине к коэффициенту корреляции $r_{xy}=0,66$ [1, стр. 122].

Аналогичные примеры вычисления коэффициента корреляции r_{uv} по формуле (55) или (55') показывают, что при наличии прямолинейной коррелятивной зависимости между x и y коэффициент корреляции r_{uv} всегда близок по величине к коэффициенту корреляции r_{xy} , т. е. $r_{uv} \approx r_{xy}$.

В случае криволинейной коррелятивной зависимости между x и y , как известно, коэффициент корреляции r_{xy} не может служить мерой тесноты связи между x и y . Коэффициент же корреляции

Таблица 1

№ п/п (m)	Год	Модули годового стока, л/сек. с 1 км ²		Модули в убывающем порядке, л/сек. с 1 км ²		$\% \frac{v'0 + u}{v'0 - u} = d$	$\% (1x) d$	$\% (1K) d$	$\% (1K) d - (1x) d$	$\cdot z[(1K) d - (1x) d]$
		Р. Тура — г. Тюмень x_i	Р. Лобва — р. п. Лобва y_i	x_m	y_m					
1	1935	2,73	9,00	7,00	11,9	3,6	50,0	13,9	36,1	1303
2	1936	1,90	5,30	5,98	9,38	8,8	80,9	50,0	30,9	956
3	1937	2,66	8,50	5,35	9,00	13,9	60,3	19,1	41,2	1697
4	1938	2,03	2,91	4,97	8,50	19,1	75,8	96,4	-20,6	424
5	1939	1,44	4,49	4,61	8,15	24,2	96,4	70,6	25,8	666
6	1940	2,49	4,56	4,28	7,77	29,4	70,6	65,5	5,1	26
7	1941	3,33	4,91	3,98	7,00	34,5	39,7	55,2	-15,5	240
8	1942	3,98	6,70	3,33	6,70	39,7	34,5	39,7	-5,2	27
9	1943	4,97	9,38	3,30	6,59	44,8	19,1	8,8	10,3	106
10	1944	3,30	7,00	2,73	5,30	50,0	44,8	34,5	10,3	106
11	1945	2,68	3,35	2,68	4,91	55,2	55,2	91,2	-36,0	1296
12	1946	4,61	6,59	2,66	4,76	60,3	24,2	44,8	-20,6	424
13	1947	7,00	8,15	2,54	4,56	65,5	3,6	24,2	-20,6	424
14	1948	5,35	7,77	2,49	4,49	70,6	13,9	29,4	-15,5	240
15	1949	4,28	3,80	2,03	3,80	75,8	29,4	75,8	-46,4	2153
16	1950	5,98	11,9	1,90	3,53	80,9	8,8	3,6	5,2	27
17	1951	2,54	3,47	1,85	3,47	86,1	65,5	86,1	-20,6	424
18	1952	1,85	3,53	1,81	3,35	91,2	86,1	80,9	5,2	27
19	1953	1,81	4,76	1,44	2,91	96,4	91,2	60,3	30,9	956

Сумма $S_{\text{ит}} = 1,1522$

ляции r_{uv} , характеризуемый выражением (55) или (55'), дает оценку тесноты связи при любой монотонной криволинейной зависимости между двумя случайными величинами x и y .

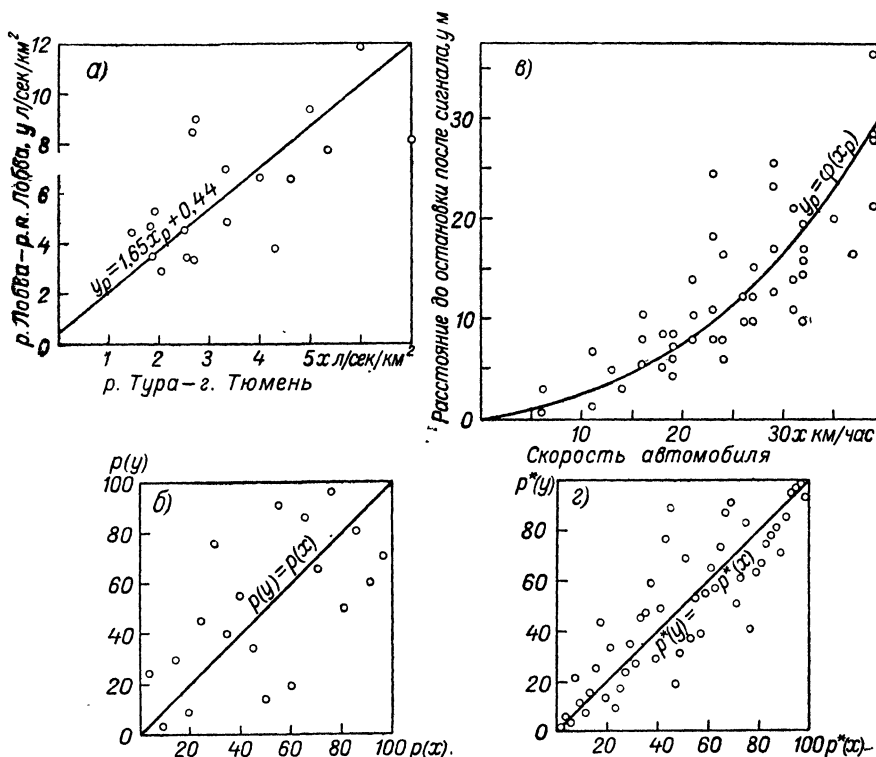


Рис. 3. Графики коррелятивной связи.

а — между модулями годового стока р. Туры у г. Тюмень и р. Лобва у р. п. Лобва по наблюдениям за 1935—1953 гг. б — то же на квадрате вероятностей, в — между скоростью автомобиля x км/час и расстоянием до остановки после сигнала y м для 50 различных автомобилей; г — то же на квадрате вероятностей.

В качестве примера в табл. 2 приведено вычисление коэффициента корреляции r_{uv} между обеспеченностью скорости автомобиля (x_i км/час) и обеспеченностью соответствующего расстояния до остановки после подачи сигнала (y_i м) для 50 различных автомобилей [9, стр. 483].

График коррелятивной зависимости между эмпирическими значениями x_i , y_i , а также между их эмпирическими вероятностями неперевышения $u_i^* = p^*(x_i)$, $v_i^* = p^*(y_i)$ представлены на рис. 3 в, г.

№ п/п (т)	Скорость автомобиля x_i км/час	Расстояние до остановки после сигнала y_i м	Расстояние в возрастном тающем порядке (y_m) м	Вероятности непревышения		$p^*(x_i) - p^*(y_i)$	$[p^*(x_i) - p^*(y_i)]^2$
				$p_m^* = \frac{m - 0,3}{n + 0,4} 100\%$			
				$p^*(x_i)$	$p^*(y_i)$		
1	6	0,6	0,6	1,4	1,4	0	0
2	6	3	1,2	3,4	5,4	-2	4
3	11	1,2	3	5,4	3,4	2	4
4	11	6,7	3	7,3	21,2	-13,9	193
5	13	4,9	4,3	9,3	11,3	-2	4
6	14	3	4,9	11,3	7,3	4	16
7	16	5,5	5,2	13,3	15,3	-2	4
8	16	7,9	5,5	15,3	25,2	-9,9	98
9	16	10,4	6,1	17,2	43,1	-25,9	670
10	18	5,2	6,1	19,2	13,3	5,9	35
11	18	8,5	6,7	21,2	33,2	-12	144
12	19	4,3	7,3	23,2	9,3	13,9	193
13	19	6,1	7,9	25,2	17,2	8	64
14	19	7,3	7,9	27,2	23,2	4	16
15	19	8,5	7,9	29,2	35,1	-5,9	35
16	21	7,9	7,9	31,2	27,2	4	16
17	21	10,4	8,5	33,2	45	-11,8	139
18	21	10,4	8,5	35,1	47	-11,9	142
19	21	14	9,8	37,1	58,9	-21,8	475
20	23	7,9	9,8	39,1	29,2	9,9	98
21	23	11	9,8	41,1	49	-7,9	62
22	23	18,3	10,4	43,1	76,8	-33,7	1135
23	23	24,4	10,4	45	88,7	-43,7	1908
24	24	6,1	10,4	47	19,2	-27,8	770
25	24	7,9	11	49	31,2	17,8	316
26	24	16,5	11	51	68,8	-17,8	316
27	26	9,8	12,2	53	37,1	12,9	166
28	26	12,2	12,2	55	53	2	4
29	27	9,8	12,8	57	39,1	17,9	320
30	27	12,2	14	58,9	55	3,9	15
31	27	15,2	14	60,9	64,9	-4	16
32	29	12,8	14,6	62,9	57	5,9	35
33	29	17,1	15,2	64,9	72,8	-7,9	62
34	29	23,2	15,9	66,8	86,7	-19,9	395
35	29	25,6	16,5	68,8	90,7	-21,9	480
36	31	11	16,5	70,8	51	19,8	391
37	31	14	17,1	72,8	60,9	11,9	142
38	31	20,8	17,1	74,8	82,7	-7,9	62
39	32	9,8	18,3	76,8	41,1	35,7	1270
40	32	14,6	19,5	78,8	62,9	15,9	252
41	32	15,9	20,1	80,8	66,8	14	196
42	32	17,1	20,8	82,7	74,8	7,9	62
43	32	19,5	21,3	84,7	78,8	5,9	35
44	35	20,1	23,2	86,7	80,8	5,9	35
45	37	16,5	24,4	88,7	70,8	17,9	320
46	39	21,3	25,6	90,7	84,7	6	36
47	39	28	25,9	92,6	94,6	-2	4
48	39	28,4	28,0	94,6	96,6	-2	4
49	39	36,6	28,4	96,6	98,4	-1,8	3
50	40	25,9	36,6	98,4	92,6	5,8	34

Сумма $S_{uv} = 1,1196$

По вычисленной в табл. 2 сумме квадратов уклонений

$$S_{uv} = \sum_{i=1}^{i=50} [p^*(x_i) - p^*(y_i)]^2 = 1,12.$$

Согласно формуле (55), находим коэффициент корреляции:

$$r_{uv} = 1 - \frac{6 S_{uv}}{n-1} = 1 - \frac{6 \cdot 1,12}{49} = 0,86.$$

По данным табл. 2 были построены на клетчатке вероятностей эмпирические кривые обеспеченности $p(x_i) = 100 - p^*(x_i)$, $p(y_i) = 100 - p^*(y_i)$ и по трем опорным ординатам (или квантилям) каждой кривой

$$x_5 = 38,8 \text{ км/час}, \quad x_{50} = 24,8 \text{ км/час}, \quad x_{95} = 10,6 \text{ км/час};$$

$$y_5 = 28,8 \text{ м}, \quad y_{50} = 11,2 \text{ м}, \quad y_{95} = 2,5 \text{ м}$$

были определены по методу автора [1] стандартные статистические параметры биномиальной кривой распределения:

$$\bar{x} = 24,8 \text{ км/час}, \quad \sigma_x = 8,6 \text{ км/час}, \quad C_{sx} = 0;$$

$$\bar{y} = 12,8 \text{ м}, \quad \sigma_y = 8,3 \text{ м}, \quad C_{sy} = 1,2.$$

По этим параметрам с помощью известной таблицы вероятностей нормированных уклонений от среднего значения ординат биномиальной кривой распределения (таблица Фостера-Рыбкина)

$$t_p = \frac{x_p - \bar{x}}{\sigma_x} = \Phi(p, C_{sx})$$

вычислены равнообеспеченные значения:

$$x_p = \bar{x} + \sigma_x \Phi(p, C_{sx}) = 24,8 + 8,6 \Phi(p; 0);$$

$$y_p = \bar{y} + \sigma_y \Phi(p, C_{sy}) = 12,8 + 8,3 \Phi(p; 1,2),$$

которые представлены в табл. 3.

Таблица 3

Величины	Обеспеченность p %						
	1	5	20	50	80	95	99
x_p км/час	44,8	38,8	32	24,8	17,6	10,6	4,8
y_p м	38,9	28,8	18,9	11,2	5,9	2,5	0,8

По данным табл. 3 на рис. 2 в построена кривая связи равнообеспеченных значений $y_{p^*} = \varphi(x_{p^*})$, которая наименее уклоняется от наблюдаемых точек (x_i, y_i) как по абсциссе, так и по ординате. Относительно этой кривой теснота связи характеризуется коэффициентом корреляции $r_{uv} = 0,86$, который был вычислен выше непосредственно по эмпирическим значениям вероятностей неперевышения $p^*(x_i)$ и $p^*(y_i)$, т. е. не прибегая к построению кривой связи равнообеспеченных значений.

Существующие методы оценки достоверности выборочного или эмпирического коэффициента корреляции r_{xy} (см., например, [11, стр. 294] и [12, стр. 394]) основаны на предположении, что исходные случайные величины x и y подчиняются нормальному (или близкому к нему) закону распределения. Поэтому существующие методы, вообще говоря, нельзя применять для оценки достоверности эмпирического коэффициента корреляции r_{uv} между случайными величинами $u = p(x)$ и $v = p(y)$ с равномерным распределением в интервале от нуля до единицы.

Ниже приводится вывод формул для оценки достоверности эмпирического коэффициента корреляции r_{uv} , вычисляемого по формуле (55) или (55').

Из выражения (55) следует, что среднее квадратическое σ_r коэффициента корреляции r_{uv} и среднее квадратическое σ_s меры разброса \bar{s}_{uv} связаны соотношением

$$\sigma_r = 6 \sigma_s. \quad (56)$$

В свою очередь среднее квадратическое σ_s среднего значения площадей элементарных квадратов отклонений

$$\bar{s} = \frac{1}{n-1} \sum_{i=1}^n s_i = \frac{1}{n-1} \sum_{i=1}^n [p(x_i) - p(y_i)]^2$$

связано со средним квадратическим σ_s площадей отдельных квадратов

$$\sigma_s = \sqrt{\frac{\sum_{i=1}^n (s_i - \bar{s})^2}{n-1}} = \sqrt{s^2 - (\bar{s})^2} \quad (57)$$

известным соотношением

$$\sigma_s = \frac{\sigma_s}{\sqrt{n}} = \sqrt{\frac{s^2 - (\bar{s})^2}{n}}. \quad (58)$$

Выборочная мера разброса связи \bar{s} , согласно выражению (55) или (55'), равна

$$\bar{s} = \frac{1 - |r|}{6}, \quad (59)$$

т. е. убывает с увеличением абсолютной величины выборочного коэффициента корреляции $|r|$ по линейному закону от значения $\bar{s}_{\max} = \frac{1}{6}$ при $r=0$ до $\bar{s}=0$ при $|r|=1$.

Величина

$$\bar{s}^2 = \frac{1}{n-1} \sum_{i=1}^n s_i^2 = \frac{1}{n-1} \sum_{i=1}^n [p(x_i) - p(y_i)]^2,$$

согласно выражению (57), всегда удовлетворяет неравенству

$$\bar{s}^2 = (\bar{s})^2 + \sigma_s^2 > (\bar{s})^2 = \frac{(1-|r|)^2}{36}$$

и убывает с увеличением абсолютной величины выборочного коэффициента корреляции по квадратическому закону

$$\bar{s}^2 = \bar{s}_{\max}^2 (1-|r|)^2 = \frac{1}{15} (1-|r|)^2. \quad (60)$$

Здесь

$$\bar{s}_{\max}^2 = \int_0^1 \int_0^1 (u-v)^4 du dv = \frac{1}{15}$$

является максимальным значением величины \bar{s}^2 при $r=0$, т. е. для независимых величин x и y , когда наблюдаемые точки $u_i = p(x_i)$, $v_i = p(y_i)$ располагаются равномерно внутри квадрата вероятностей $0 \leq u \leq 1$, $0 \leq v \leq 1$ и теоретическая плотность распределения вероятности различных сочетаний (u, v) постоянная и равна единице.

Подставляя выражения (59) и (60) для \bar{s} и \bar{s}^2 в равенство (58), получаем расчетную формулу для среднего квадратического отклонения σ_s выборочной меры разброса связи \bar{s} от ее истинного значения

$$\sigma_s = (1-|r|) \sqrt{\frac{\frac{1}{15} - \frac{1}{36}}{n}} = \frac{0,1972(1-|r|)}{\sqrt{n}}. \quad (61)$$

Отсюда с учетом соотношения (56) получаем также расчетную формулу для среднего квадратического отклонения σ_r выборочного коэффициента корреляции r от его истинного значения

$$\sigma_r = 6 \sigma_s = \frac{1,183(1-|r|)}{\sqrt{n}}. \quad (62)$$

Точный закон распределения выборочного коэффициента корреляции r в пределах от -1 до $+1$ весьма сложен, поэтому

Р. Фишер показал, что в результате преобразования величины r с помощью гиперболического арка-тангенса

$$r = \operatorname{th} z, \quad z = \operatorname{Arth} r = \frac{1}{2} \ln \frac{1+r}{1-r} \quad (63)$$

распределение новой случайной величины z в пределах от $-\infty$ до $+\infty$ хорошо приближается к нормальному закону даже при небольших объемах выборки.

Среднее квадратическое отклонение σ_z выборочной величины $z = f(r) = \operatorname{Arth} r$ от ее истинного значения z^* при $|r| < 1$ связано со средним квадратическим отклонением σ_r известной приближенной формулой

$$\sigma_z = |f'(r)| \sigma_r = \frac{\sigma_r}{1-r^2}, \quad (64)$$

где $f'(r) = \frac{1}{1-r^2}$ есть производная от $\operatorname{Arth} r$ при выборочном значении r .

Подставляя в правую часть равенства (64) выражение (62) для σ_r , получаем приближенную расчетную формулу для среднего квадратического отклонения σ_z выборочного значения вспомогательной величины $z = \frac{1}{2} \ln \frac{1+r}{1-r}$ от ее истинного значения

$$\sigma_z = \frac{1,183}{(1+|r|) \sqrt{n}}, \quad (65)$$

где выборочный или эмпирический коэффициент корреляции r определяется по формуле (55) или (55').

Следовательно, по закону нормального распределения истинное значение вспомогательной величины z^* с доверительной вероятностью P заключается в интервале

$$z_1 = z - \sigma_z t_p < z^* < z + \sigma_z t_p = z_2, \quad (66)$$

где $t_p = \frac{|z-z^*|}{\sigma_z}$ — абсолютная величина нормированного отклонения, определяемая в зависимости от принятой доверительной вероятности P по таблице интеграла вероятности Гаусса (см., например, [13, стр. 81]):

$$P = \frac{2}{\sqrt{2\pi}} \int_0^{t_p} e^{-\frac{t^2}{2}} dt = f(t_p).$$

По найденным границам доверительного интервала

$$z_1 = z - \sigma_z t_p, \quad z_2 = z + \sigma_z t_p \quad (67)$$

с помощью таблицы значений гиперболического тангенса [13, стр. 52] или таблицы значений арка-тангенса (табл. 4) легко

Значения $z = \frac{1}{2} \ln \frac{1+r}{1-r}$, отвечающие значениям r

r	Сотые доли ¹									
	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0,010	0,020	0,030	0,040	0,050	0,060	0,070	0,080	0,090
0,1	0,100	0,110	0,121	0,131	0,141	0,151	0,161	0,172	0,182	0,192
0,2	0,203	0,213	0,224	0,234	0,245	0,255	0,266	0,277	0,288	0,299
0,3	0,310	0,320	0,332	0,343	0,354	0,365	0,377	0,388	0,400	0,412
0,4	0,424	0,436	0,448	0,460	0,472	0,485	0,497	0,510	0,523	0,536
0,5	0,549	0,563	0,576	0,590	0,604	0,618	0,633	0,648	0,662	0,678
0,6	0,693	0,709	0,725	0,741	0,758	0,775	0,793	0,811	0,829	0,848
0,7	0,867	0,887	0,908	0,929	0,950	0,973	0,996	1,020	1,045	1,071
0,8	1,099	1,127	1,157	1,188	1,221	1,256	1,293	1,333	1,376	1,422
0,9	1,472	1,528	1,589	1,658	1,738	1,832	1,946	2,092	2,298	2,647
0,99	2,647	2,700	2,759	2,826	2,903	2,994	3,106	3,250	3,453	3,800

Примеры: 1. Требуется определить значение z при $r = 0,84$. Имеем $z = 1,221$.
 2. Требуется определить значение z при $r = -0,84$. Имеем $z = -1,221$.

¹ Для последней строки это тысячные доли.

определяются соответствующие границы доверительного интервала для истинного значения коэффициента корреляции:

$$r_1 = \text{th } z_1, \quad r_2 = \text{th } z_2, \quad r_1 < r^* < r_2, \quad (68)$$

в пределах которого с принятой вероятностью P заключается коэффициент корреляции.

При заданной доверительной вероятности P критерием наличия (или отсутствия) корреляционной связи может служить выполнение (или невыполнение) неравенства

$$|z| > \sigma_z t_P, \quad (69)$$

т. е. абсолютная величина выборочного значения $z = \frac{1}{2} \ln \frac{1+z}{1-r}$ должна быть больше ее возможной доверительной погрешности $\sigma_z t_P$.

Известно, что при больших объемах выборки (практически при $n > 30$) выборочный коэффициент корреляции распределяется по закону, близкому к нормальному. Благодаря этому при объемах выборки $n > 30$ критерием наличия (или отсутствия) корреляционной связи может служить выполнение (или невыполнение) неравенства

$$|r| > \sigma_r t_P, \quad (70)$$

т. е. абсолютная величина выборочного коэффициента корреляции должна быть больше ее возможной доверительной погрешности $\sigma_r t_P$ при данной доверительной вероятности P .

Истинное значение коэффициента корреляции r^* с вероятностью P заключается в доверительном интервале

$$r_1 = r - \sigma_r t_P < r^* < r + \sigma_r t_P = r_2, \quad (71)$$

где нормированное уклонение t_P , как и прежде, определяется в зависимости от принятой доверительной вероятности P по таблице интеграла вероятности Гаусса, а σ_r — по формуле (62).

Для иллюстрации изложенного выше метода оценим достоверность эмпирического коэффициента корреляции $r_{uv} = 0,62$, вычисленного по данным табл. 1, т. е. при объеме выборки $n = 19$.

Согласно выражению (63), с помощью табл. 4 находим значение вспомогательной величины z , соответствующее эмпирическому коэффициенту корреляции

$$z = \frac{1}{2} \ln \frac{1+0,62}{1-0,62} = 0,725.$$

По формуле (65) находим среднее квадратическое отклонение z от истинного значения z^* :

$$\sigma_z = \frac{1,183}{(1+0,62) \sqrt{19}} = 0,167.$$

Принимая доверительную вероятность равной $P = 0,95$, нахо-

дим по таблице интеграла вероятности Гаусса соответствующее нормированное уклонение $t_p = t_{0,95} = 1,96$.

Отсюда следует, что критерий наличия корреляционной связи (69) выполняется:

$$z = 0,725 > 0,167 \cdot 1,96 = 0,328.$$

Согласно неравенству (66), истинное значение вспомогательной величины z^* с доверительной вероятностью $P=0,95$ заключается в интервале

$$0,397 = 0,725 - 0,328 < z^* < 0,725 + 0,328 = 1,053.$$

Отсюда с помощью табл. 4 находим границы соответствующего доверительного интервала для коэффициента корреляции:

$$r_1 = \text{th}(0,397) = 0,38, \quad r_2 = \text{th}(1,053) = 0,78.$$

Таким образом, истинное значение коэффициента корреляции r^* с вероятностью $P=0,95$ заключается в интервале

$$0,38 < r^* < 0,78.$$

Аналогичным образом при той же доверительной вероятности $P=0,95$, оценивая достоверность эмпирического коэффициента корреляции $r_{uv}=0,86$, вычисленного по данным табл. 2, т. е. при объеме выборки $n=50$, последовательно получаем

$$z = \frac{1}{2} \ln \frac{1+0,86}{1-0,86} = 1,293;$$

$$\sigma_z = \frac{1,183}{(1+0,86) \sqrt{50}} = 0,09;$$

$$z = 1,293 > 0,09 \cdot 1,96 = 0,176;$$

$$1,117 = 1,293 - 0,176 < z^* < 1,293 + 0,176 = 1,469;$$

$$r_1 = \text{th}(1,117) = 0,81, \quad r_2 = \text{th}(1,469) = 0,90;$$

$$0,81 < r^* < 0,90.$$

Из этого примера (при $n=50$, $r=0,86$) видно, что по сравнению с предыдущим примером (при $n=19$, $r=0,62$) доверительный интервал для коэффициента корреляции оказался более узким.

Применяя (в целях сравнения) для оценки эмпирического коэффициента корреляции $r_{uv}=0,86$ при $n=50$ формулы (62), (70) и (71), последовательно получаем

$$\sigma_r = \frac{1,183(1-0,86)}{\sqrt{50}} = 0,0234;$$

$$r = 0,86 > 0,0234 \cdot 1,96 = 0,046;$$

$$0,814 = 0,86 - 0,046 < r^* < 0,86 + 0,046 = 0,906.$$

Полученный доверительный интервал, как и следовало ожидать, практически совпадает с прежним $0,81 < r^* < 0,90$.

Наконец, если применим формулы (62), (70) и (71) для оценки эмпирического коэффициента корреляции $r_{uv} = 0,62$ при $n = 19$, то при той же доверительной вероятности $P = 0,95$ получим

$$\sigma_r = \frac{1,183(1 - 0,62)}{\sqrt{19}} = 0,103;$$

$$r = 0,62 > 0,103 \cdot 1,96 = 0,20;$$

$$0,42 = 0,62 - 0,20 < r^* < 0,62 + 0,20 = 0,82.$$

Полученный доверительный интервал, как и следовало ожидать, несколько отличается от прежнего более точного интервала $0,38 < r^* < 0,78$.

В заключение следует подчеркнуть, что важнейшее свойство кривых связи равнообеспеченных значений, свойство их наименьшего отклонения от наблюдаемых точек как по абсциссе, так и по ординате, позволяет использовать кривые связи равнообеспеченных значений в качестве общего и весьма простого математического аппарата для выявления формы связи, скрытой за корреляцией.

Важно заметить, что форма кривой связи равнообеспеченных значений $y_p = \varphi(x_p)$ выявляется автоматически, в то время как в методах наименьших квадратов или прямоугольников приходится принимать заранее (априори) тот или иной общий вид уравнения связи, содержащий неизвестные искомые параметры связи.

В результате преобразования исходных случайных величин x, y с помощью кривых обеспеченности в новые случайные величины $u = p(x), v = p(y)$ с равномерным распределением в пределах от нуля до единицы любая монотонная криволинейная коррелятивная зависимость между x_i и y_i превращается в прямолинейную зависимость между новыми случайными величинами $u_i = p(x_i)$ и $v_i = p(y_i)$. При этом кривая связи равнообеспеченных значений $p(y) = p(x)$ или $p(y) = 1 - p(x)$ графически изображается в виде одной из диагоналей $v = u$ или $v = 1 - u$ в квадрате распределения вероятностей $0 \leq u \leq 1, 0 \leq v \leq 1$ новых случайных величин.

Благодаря этому предлагаемые простые формулы для определения эмпирического коэффициента корреляции r_{uv} между случайными величинами $u = p(x), v = p(y)$ и формулы для оценки его достоверности применимы при любой монотонной форме кор-

релятивной связи и при любых законах распределения исходных случайных величин x и y .

Коэффициент корреляции r_{uv} между величинами обеспеченностей $u=p(x)$, $v=p(y)$ инвариантен по отношению к любым монотонным преобразованиям $\varphi(x)$ и $\psi(x)$ исходных случайных величин x и y , т. е. остается неизменным, если обе функции преобразования $\varphi(x)$ и $\psi(y)$ возрастающие либо убывающие. Если одна из них возрастающая, а другая убывающая, то коэффициент корреляции r_{uv} только изменяет знак. Известно, что коэффициент корреляции r_{xy} между исходными случайными величинами x и y инвариантен только по отношению к линейным преобразованиям.

ЛИТЕРАТУРА

1. Алексеев Г. А. Графоаналитические способы определения и приведения к длительному периоду наблюдений параметров кривых распределения. Тр. ГГИ, вып. 73, 1960.
2. Алексеев Г. А. Приближенный метод определения вероятных значений гидрологических величин. Тр. ГГИ, вып. 61, 1957.
3. Алексеев Г. А. О формулах для вычисления обеспеченности гидрологических величин. Метеорология и гидрология, № 6, 1955.
4. Алексеев Г. А. К вопросу об оценке вероятной повторяемости гидрологических величин. Метеорология и гидрология, № 3, 1958.
5. Дроздов О. А. Некоторые приемы разыскания функциональной зависимости, скрытой за корреляцией, и приведение повторяемостей метеорологических величин к одному периоду. Тр. НИУ ГУГМС, сер. 1, вып. 15. Гидрометеиздат, 1945.
6. Зайков Б. Д. и Белинков С. Ю. Норма годового стока, погрешность ее и приведение коротких рядов к длительному периоду. Исследования рек СССР, вып. 8 — Вопросы стока. Редиздат ЦУЕГМС, 1935.
7. Срезневский Б. И. О силе ветра в Санкт-Петербурге и Кронштадте. О зависимости между силой ветра и барометрическим градиентом. Записки по гидрографии, вып. 11, 1889.
8. Срезневский Б. И. Решение вопроса о корреляции между двумя переменными. Метеорол. вестн., т. XXIV, 1914.
9. Хальд А. Математическая статистика с техническими приложениями. Изд-во иностр. лит., 1956.
10. Шепелевский А. А. О вычислении функциональной связи, скрытой за корреляцией. Тр. ГГО, вып. 15 — Климатология, 1937.
11. Смирнов Н. В. и Дунин-Барковский И. В. Краткий курс математической статистики для технических приложений. Физматгиз, 1959.
12. Дунин-Барковский И. В. и Смирнов Н. В. Теория вероятностей и математическая статистика в технике (общая часть). Гостехиздат, М., 1955.
13. Бронштейн И. Н. и Семендяев К. А. Справочник по математике для инженеров и учащихся втузов. Гостехиздат, 1957.

ОГЛАВЛЕНИЕ

	Стр.
Предисловие	3
Глава I. Введение	7
§ 1. Физические законы и эмпирические зависимости	7
§ 2. Измеримость величин	11
Глава II. Простейшие вероятностные представления	14
§ 3. Частость и вероятность	14
§ 4. Сложение и умножение вероятностей	16
§ 5. Закон больших чисел	18
Глава III. Одномерные совокупности случайных величин	22
§ 6. Понятие о множествах	22
§ 7. Мера Лебега	25
§ 8. Математическое ожидание	28
§ 9. Рассеивание	34
§ 10. Медиана и мода	40
§ 11. Моменты высших порядков	44
§ 12. Производящие функции	47
Глава IV. Распределение дискретной совокупности случайных величин	49
§ 13. Биномиальное распределение	49
§ 14. Закон Пуассона	52
§ 15. Примеры применения закона Пуассона	55
Глава V. Распределение непрерывных случайных величин	58
§ 16. Плотность вероятности	58
§ 17. Понятие о гамма-функции	60
§ 18. Интерполяция биномиального распределения	62
§ 19. Нормальное распределение	68
§ 20. Хи-квадрат-распределение	75
§ 21. Распределение Стьюдента	80
Глава VI. Ошибки измерения	85
§ 22. Основные положения	85
§ 23. Систематические ошибки	94
§ 24. Грубые (выскакивающие) ошибки	95
§ 25. Случайные ошибки	100

Глава VII. Измерение «больших» величин	105
§ 26. Симметричное распределение случайных ошибок	105
§ 27. Оценка среднего и стандарта	109
§ 28. Неравноточные измерения	117
Глава VIII. Измерение малых (существенно положительных) величин	121
§ 29. Асимметрия распределения случайных ошибок	121
§ 30. Применение гамма-распределения	129
§ 31. Оценка медианы	134
§ 32. Оценка прогнозов	146
Глава IX. Связь между измеренными величинами	153
§ 33. Двумерные и многомерные совокупности случайных величин	153
§ 34. Функциональная и стохастическая связи	156
§ 35. Основы теории корреляции	158
§ 36. Нормальная корреляция	167
§ 37. Множественная корреляция	170
§ 38. Среднее и дисперсия функции измеренных величин	172
Глава X. Отыскание формы связи между двумя измеренными величинами	175
§ 39. Постановка задачи	175
§ 40. Применение принципа Лежандра для вычисления поправок при недостаточном числе измерений	180
§ 41. Применение принципа Лежандра при избыточном числе измерений	182
§ 42. Линейная задача	183
§ 43. Вероятнейшая прямая	188
§ 44. Принцип Лапласа и метод Эджуорса	190
§ 45. Нелинейные задачи	195
§ 46. Способ Чебышева	202
Глава XI. Отыскание формы связи между тремя измеренными величинами	210
§ 47. Расширение способа наименьших квадратов на многие переменные	210
§ 48. Семейства эмпирических кривых	212
§ 49. Метод изолиний	215
Глава XII. Метод размерностей	219
§ 50. Основные положения	219
§ 51. П-теорема	220
§ 52. Составление критериев	223
Глава XIII. Логарифмическая анаморфоза	230
§ 53. Общие свойства логарифмических изображений	230
§ 54. Логарифмики сложения и вычитания	240
Глава XIV. Основные формы эмпирических кривых	252
§ 55. Типы монотонных кривых	252
§ 56. Кривые с одним экстремумом	260
Литература	263

Приложение I

Т а б л и ц а 1. Вероятности «выскакивающих» измерений (по Стю- денту)	264
Т а б л и ц а 2. Удвоенная функция Лапласа	
$\delta = 2 \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$	265
Т а б л и ц а 3. Значения t_α в равенстве $\alpha = 2 \int_0^t \delta(t, k) dt$ (по Стю- денту — Фишеру)	266
Т а б л и ц а 4. Значения q , удовлетворяющие равенству $P(s - \varepsilon <$ $< \sigma < s + \varepsilon) = \alpha$	268
Т а б л и ц а 5. Значения χ^2 , удовлетворяющие равенству	
$\delta = \int_0^{\chi^2} p(\chi^2, k) d\chi^2$	269
Т а б л и ц а 6. Основные характеристики гамма-распределения для $\alpha = -\frac{1}{2}, 0, +\frac{1}{2}, 1, 2, \dots, 15$	270

Приложение II

Г. А. Алексеев. Отыскание формы связи между случайными величинами методом квантилей	271
--	-----

Великанов Михаил Андреевич
ОШИБКИ ИЗМЕРЕНИЯ
И
ЭМПИРИЧЕСКИЕ ЗАВИСИМОСТИ

Редактор *В. Н. Дерюгина*
Техн. редактор *А. Г. Алексеев*
Корректоры: *В. В. Мамедова* и *Б. А. Минц*

Сдано в набор 27/II 1962 г. Подписано к печати
25/VI 1962 г. М-06364 Формат 60×90^{1/16}. Бум. л. 9,5
Печ. л. 19. Уч.-изд. л. 19,04 Тираж 1600 экз.
Индекс ГЛ-20. Заказ 102 Цена 1 руб. 10 коп.
Гидрометеорологическое издательство. Ленинград.
В-53, 2-я линия, д. № 23

Типография № 8 УПП Ленсовнархоза.
Ленинград, Прачечный пер., д № 6