



MATHEMATICAL
METHODS
OF STATISTICS
by
Harald Cramér

Professor in the University
of Stockholm

1946

Гаральд Крамер

МАТЕМАТИЧЕСКИЕ МЕТОДЫ СТАТИСТИКИ

Перевод с английского

А. С. МОНИНА и А. А. ПЕТРОВА

под редакцией

академика

А. Н. КОЛМОГОРОВА

Издание второе, стереотипное

ИЗДАТЕЛЬСТВО «МИР»

Москва 1975

Книга выдающегося шведского математика Г. Крамера «Математические методы статистики» — классическое руководство по этой дисциплине. Впервые на русском языке она была издана в 1948 г. и сыграла большую роль в развитии теоретических работ по математической статистике, а также в повышении уровня прикладных работ.

Собственно математической статистике посвящена третья (последняя) часть книги, а ее вторая часть до сих пор является одним из лучших учебных пособий по теории вероятностей.

Книга необходима всем изучающим математическую статистику и ее приложения.

Редакция литературы по математическим наукам

К 20203-011
041(01)-75 11—75 © Перевод на русский язык, «Мир», 1975

ПРЕДИСЛОВИЕ КО ВТОРОМУ РУССКОМУ ИЗДАНИЮ

Монография Г. Крамера является классическим руководством по математической статистике. За четверть века, прошедшие со времени ее появления в русском переводе, эта книга сыграла важную роль в развитии теоретических и прикладных работ по математической статистике и в воспитании квалифицированных специалистов в данной области. Несмотря на обширную литературу, появившуюся позднее, книга Г. Крамера сохраняет большой интерес для современного читателя. В частности, в нашей стране она остается почти обязательной составной частью программ подготовки аспирантов — и математиков, и прикладников,— в круг интересов которых входит математическая статистика. Этим и вызвана необходимость нового издания книги. При этом мы сочли возможным, не изменяя основного текста, дополнить книгу материалом справочного характера, содержащим краткие комментарии к отдельным главам книги, и новой библиографией, что должно помочь читателю ориентироваться в современной математической статистике. Эта работа выполнена А. В. Прохоровым. При составлении дополнения мы воспользовались замечаниями и рекомендациями Л. Н. Большева и Ю. В. Прохорова, которым выражаем искреннюю благодарность.

А. Н. Колмогоров

ИЗ ПРЕДИСЛОВИЯ К ПЕРВОМУ РУССКОМУ ИЗДАНИЮ

Задача книги Г. Крамера в первую очередь чисто научная: создание удовлетворяющей современным требованиям системы изложения математической статистики. Общие контуры этой системы были в последние годы ясны многим исследователям и неоднократно излагались в ряде журнальных статей. Тем не менее, существовавшие до настоящего времени систематические курсы математической статистики строились на теоретическом фундаменте, совершенно не соответствующем современным требованиям, и поэтому в своей практической части в значительной мере приобретали чисто рецензурный характер. Из этих традиционно построенных курсов едва ли не лучшим и самым полным в мировой литературе является русский курс В. И. Романовского («Математическая статистика», 1938). Однако за последнее время широкое развитие получили принципиальные общие исследования по теории статистической проверки гипотез и статистической оценки параметров (или непосредственной «непараметрической» оценки законов распределения), что вызвало необходимость существенно нового изложения самих основ математической статистики. Таким образом, исследования по специальным вопросам математической статистики переросли тот уровень изложения математических и теоретико-вероятностных предпосылок, который все еще по традиции сохранялся во всех руководствах, и постепенно сделались ясной своевременность попытки систематического изложения принципиальных вопросов математической статистики с вполне современных позиций. Ясно было, однако, что окончательную убедительность новая система изложения приобретет лишь в том случае, если будет проделан большой труд по подчинению ей всего богатого материала конкретных фактов, найденного на различных разрозненных путях. Настоящая книга является попыткой осуществления этой программы.

Книга Г. Крамера не свободна от недостатков. Некоторые важные общие направления исследований в математической статистике не нашли в ней должного отражения: «непараметрические задачи», в которых наиболее существенные результаты принадлежат Н. В. Смирнову; временные ряды, серьезная трактовка которых опи-

рается на общую теорию случайных процессов; доверительные границы для случайных величин; только что начавший развиваться «последовательный анализ». Возможно, с другой стороны, что изложение общих основ теории удастся в будущем сделать более элементарным и более доступным (я имею в виду при этом не отказ от пользования общей теорией меры и понятиями измеримости по Борелю и интеграла Лебега, а еще более доступное изложение этих концепций, подкрепленное конкретными примерами, относящимися непосредственно к теории вероятностей и статистике). Наконец, общие концепции Крамера, сводящие задачи научного исследования к «описанию», «анализу» и «предсказанию» (см. гл. 13), имеют заметный оттенок эмпиризма (трудно сказать больше, так как автор очень скрупулезно высказываеться по этим вопросам).

Для чтения книги Крамера необходимо хорошее знание классического математического анализа. Все вспомогательные математические средства, выходящие за эти пределы, изложены на страницах первой части книги. Здесь читатель может познакомиться с языком геометрии n -мерного евклидова пространства, чрезвычайно существенным для понимания действительного замысла многих вполне «классических» выводов в теории вероятностей и математической статистике, с соответствующими этой геометрии фактами из линейной алгебры, с преобразованиями Фурье как в одномерном, так и в n -мерном случае. Наиболее трудными для нематематиков в этой части покажутся, вероятно, главы, посвященные теории меры и теории интегрирования. Читателям, знающим теорию вероятностей, следует, однако, иметь в виду, что определение интеграла Лебега по своей логической сущности просто совпадает с определением математического ожидания, а типичные приемы рассуждения «метрической теории функций» повторяют вывод Чебышевым его знаменного неравенства.

Вторая часть книги посвящена систематическому изложению теории вероятностей. Нам не хотелось бы рекомендовать ее взамен какого-либо обычного курса теории вероятностей. Желательно, чтобы читатель до чтения книги Г. Крамера уже знал элементы теории вероятностей из какого-либо другого руководства. Для читателей-математиков между первой и второй частью книги Г. Крамера можно рекомендовать прочесть маленькую книжку А. Н. Колмогорова «Основные понятия теории вероятностей»¹⁾. Читатель, основательно знакомый с теорией вероятностей, может ограничиться просмотром второй части книги Г. Крамера с целью систематизации и пополнения своих знаний в направлениях, используемых в третьей части.

¹⁾ Второе издание этой книги вышло в издательстве «Наука» в 1974 г., первое — в 1933 г.

Третья часть книги является основной и посвящена собственно математической статистике. Изложение третьей части следует в значительной мере индуктивному пути. После общей предварительной характеристики задач статистической оценки (главы 25—26) рассматривается ряд специальных задач (главы 27—31), и лишь после этого идут уже упомянутые основные, с точки зрения общей теории, главы 32—35, после которых изложение (в главах 36—37) вновь возвращается к более сложным специальным задачам. Эти последние главы — 36 (дисперсионный анализ) и 37 (уравнения регрессии) — написаны несколько конспективно в отличие от глав 27—31, где изложение доведено до числовых примеров, заимствованных из реальной статистической практики. Если учесть объем третьей части (250 страниц), то изложенный в ней конкретный материал следует признать весьма богатым. Впрочем, в качестве сводки специальных статистических приемов и соответствующих формул книга Крамера все же не может заменить трактата В. И. Романовского, охватывающего в более догматическом изложении еще более обширный материал.

Так как книга Крамера, как ясно из сказанного, рассчитана лишь на квалифицированного читателя и не предназначена служить элементарным учебником математической статистики, то мы не сочли целесообразным подвергнуть ее какой-либо переработке, предпочитая, чтобы она дала подготовленному читателю представление об определенном этапе развития науки.

По сравнению с подлинником в русском издании сделаны лишь следующие изменения:

1) Ссылки на иностранную учебную литературу заменены ссылками на русские учебники. Ссылки же автора на научные журнальные статьи оставлены без изменения.

2) Общий указатель терминов разделен на две части. В первой части даны специальные термины из теории вероятностей и математической статистики. При них в скобках указаны страницы книги, где дается их определение, и соответствующие английские термины.

A. N. Колмогоров

ИЗ ПРЕДИСЛОВИЯ АВТОРА

Последние 25 лет отмечены большим прогрессом статистической науки, достигнутым благодаря блестящим работам статистиков английской и американской школы, среди которых на первом месте должно быть упомянуто имя профессора Р. А. Фишера. В течение этого же времени, главным образом благодаря трудам французских и русских математиков, классическое исчисление вероятностей превратилось в чисто математическую теорию, удовлетворяющую современным требованиям строгости.

Целью настоящего труда является объединяющее эти две линии изложение математической теории современных статистических методов, основывающихся на понятии вероятности. Полное понимание теории этих методов требует достаточно обширных познаний из высшей математики. В связи с этим я постарался изложить в книге все необходимое с точки зрения читателя, свободно владеющего элементарными понятиями дифференциального и интегрального исчисления, алгебры и аналитической геометрии.

В первой части книги, являющейся математическим введением, излагаются необходимые для понимания второй и третьей частей математические сведения, предварительное знание которых читателем не предполагается. Особое внимание здесь удалено фундаментальному понятию распределения и интегрированию относительно распределения. В главах 4—5 излагается краткая теория лебеговой меры и интегрирования, а указанные основные понятия вводятся в главах 6—7 в результате непосредственного обобщения этой теории.

Вторая часть книги содержит общую теорию случайных величин и распределений вероятностей, а третья часть посвящена теории выборочных распределений, статистической оценки и критериев значимости. Выбор вопросов, излагаемых в третьей части, по необходимости был несколько произволен, но я постарался изложить в первую очередь вопросы наиболее общего значения. Если они будут изучены полностью, то читатель сможет самостоятельно применять их к частным проблемам. Чтобы ограничить объем книги разумными пределами, я был вынужден оставить в стороне некоторые очень интересные темы, которые я первоначально намеревался

изложить, как, например, теорию случайных процессов статистических временных рядов и периодограмм.

Теория статистических критериев иллюстрирована численными примерами, заимствованными из различных областей приложения этой теории. Количество примеров пришлось значительно урезать из-за недостатка места. Пришлось также отказаться от обсуждения вопросов, связанных с необходимостью последовательного проведения статистических расчетов.

Для изучения второй и третьей частей книги нет необходимости читать первую часть целиком. Читатель, стремящийся как можно скорее проникнуть в суть дела, может ограничиться лишь беглымзнакомством с указанными выше основными понятиями. Для этой цели рекомендуется прочитать главы 1—3 и параграфы 4.1—4.2, 5.1—5.3, 6.1—6.2, 6.4—6.6, 7.1—7.2, 7.4—7.5 и 8.1—8.4. Затем читатель может перейти к главе 13 с тем, чтобы возвращаться к первой части лишь по мере появления соответствующих ссылок.

Книга основана на моих университетских лекциях, читанных приблизительно с 1930 г., и написана в основном в течение 1942—1944 гг. Из-за условий военного времени иностранная научная литература поступала в Швецию неполностью и со значительными задержками, что должно послужить оправданием возможных пробелов в ссылках.

Я обязан профессору Р. А. Фишеру за разрешение воспроизвести таблицы распределений χ^2 и t из книги «Статистические методы для исследователей».

В подготовке книги значительную помощь оказали мне мои друзья. Профессора Гаральд Бор и Эрнст Якобсталь, напечатавшие в Швеции убежище от испытаний военного времени, прочитали часть книги на стадии рукописи и корректуры и оказали мне помощь своей критикой и советами. Профессор Герман Вольд очень внимательно просмотрел корректуру всей книги, и я чрезвычайно благодарен ему за его ценные замечания.

Отделение математической
статистики
Стокгольмского университета
Май 1945

Г. К.

ЧАСТЬ ПЕРВАЯ

**МАТЕМАТИЧЕСКОЕ
ВВЕДЕНИЕ**

ГЛАВЫ 1—3

ТОЧЕЧНЫЕ МНОЖЕСТВА

ГЛАВА I

ОБЩИЕ СВОЙСТВА МНОЖЕСТВ

1.1. Множества. В чистой и прикладной математике часто встречаются положения, когда приходится рассматривать совокупность всевозможных объектов, обладающих некоторыми определенными свойствами. Совокупность объектов, определенная таким образом, называется **множеством**, а каждый объект, принадлежащий такому множеству, называется **элементом множества**.

Элементы множества могут быть объектами любого рода: точками, числами, функциями, вещами, лицами и т. д. Так, например, можно рассматривать: 1) множество всех положительных целых чисел, 2) множество всех точек данной прямой линии, 3) множество всех рациональных функций двух переменных, 4) множество всех людей, родившихся в данной стране и живущих к концу 1940 г.

В первой части книги мы будем иметь дело главным образом с множествами, элементы которых суть точки или числа, но в этой вводной главе мы займемся рассмотрениями, приложимыми в самом общем случае, когда элементами множества могут быть объекты произвольного рода.

В приведенном выше четвертом примере наше множество состоит из конечного, хотя возможно и неизвестного нам, числа элементов, тогда как в трех первых примерах мы, очевидно, имеем дело с множествами, число элементов которых бесконечно. Таким образом, мы должны делать различие между **конечными** и **бесконечными** множествами.

Бесконечное множество называется **счетным**, если его элементы могут быть расположены в **последовательность** $x_1, x_2, \dots, x_n, \dots$ так, чтобы а) каждое x_n было элементом множества и б) каждый элемент множества занимал определенное место в этой последовательности. При таком расположении элементов мы устанавливаем **взаимно-однозначное соответствие** между элементами данного множества и элементами множества всех положительных целых чисел $1, 2, \dots, n, \dots$, которое образует простейший пример счетного множества.

В дальнейшем мы увидим, что существуют также *несчетные бесконечные множества*. Если из такого множества мы выберем какую-либо последовательность элементов $x_1, x_2, \dots, x_n, \dots$, то всегда найдутся элементы этого множества, не принадлежащие нашей последовательности. Можно сказать поэтому, что несчетное множество — бесконечное высшего порядка по сравнению со счетным множеством. Мы покажем в дальнейшем (см. параграф 4.3), что множество всех точек данной прямой линии дает пример несчетного множества.

1.2. Подмножества. Пространство. Если множества S и S_1 таковы, что каждый элемент множества S_1 принадлежит к S , то мы говорим, что S_1 есть *подмножество* множества S , и пишем

$$S_1 \subset S \text{ или } S \supset S_1.$$

Иногда мы будем говорить, что в этом случае S_1 *содержится* в S или S_1 *принадлежит* S . Если S_1 состоит из одного элемента x , то мы будем употреблять то же обозначение $x \subset S$ и говорить, что x *принадлежит* множеству S .

В частном случае, когда выполнены оба соотношения $S_1 \subset S$ и $S \subset S_1$, множества S и S_1 называются *равными*, и мы пишем

$$S = S_1.$$

Иногда удобно рассматривать множество S , которое не содержит ни одного элемента. Мы называем его *пустым множеством* и пишем $S = \emptyset$. Пустое множество является подмножеством любого множества. Если рассматривать пустое множество как частный случай конечного множества, то ясно, что *каждое подмножество конечного множества само конечно*, в то время как *каждое подмножество счетного множества конечно или счетно*. Так, например, множество всех натуральных чисел, заключенных между 20 и 30, есть конечное подмножество множества всех натуральных чисел 1, 2, 3, ..., в то время как множество всех нечетных натуральных чисел 1, 3, 5, ... есть счетное подмножество того же множества.

Во многих исследованиях нас будут интересовать свойства и взаимоотношения различных подмножеств данного множества S . Множество S , содержащее, таким образом, все элементы, которые могут встретиться в нашем исследовании, называется *пространством*. Если, например, мы рассматриваем различные множества точек на данной прямой, то

мы можем выбрать за наше пространство множество S всех ее точек. Подмножество S пространства S называется кратко *множеством в S* .

1.3. Операции над множествами. Пусть теперь дано пространство S и мы рассматриваем различные *множества в S* . Определим сначала операции *сложения, умножения и вычитания* множеств.

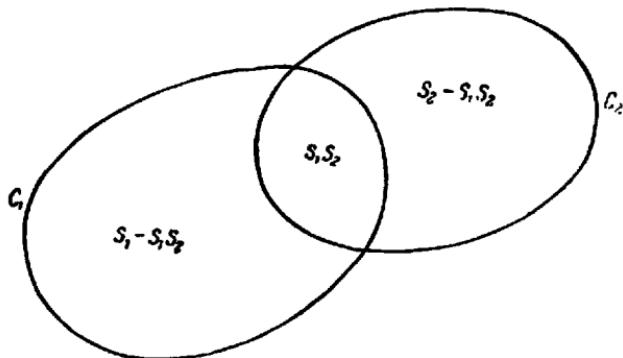
Сумма двух множеств S_1 и S_2 ,

$$S' = S_1 + S_2,$$

есть множество S' всех элементов, принадлежащих по крайней мере одному из множеств S_1 и S_2 .

Произведение

$$S'' = S_1 S_2$$



Фиг. 1. Простейшие операции над множествами.

есть *общая часть* множеств S_1 и S_2 , т. е. множество S'' всех элементов, принадлежащих обоим множествам S_1 и S_2 .

Наконец, *разность*

$$S''' = S_1 - S_2$$

определяется лишь в случае, когда S_2 есть подмножество S_1 , и S''' есть множество всех элементов, принадлежащих S_1 и не принадлежащих S_2 .

Так, если S_1 и S_2 состоят из всех точек, находящихся внутри замкнутых кривых C_1 и C_2 (фиг. 1), то $S_1 + S_2$ есть множество всех точек внутренних по отношению хотя бы к одной из двух кривых, в то время как $S_1 S_2$ есть общая часть обеих областей, ограниченных данными кривыми.

Произведение $S_1 S_2$ есть, очевидно, подмножество каждого из множеств S_1 и S_2 . Разность $S_n - S_1 S_2$ (где $n = 1, 2$) есть множество всех точек S_n , не принадлежащих $S_1 S_2$.

В частном случае, когда S_1 и S_2 не имеют общих элементов, их произведение пусто, т. е. $S_1 S_2 = \emptyset^{\circ}$. С другой стороны, если $S_1 = S_2$, то их разность есть пустое множество, $S_1 - S_2 = \emptyset$.

В частном случае, когда S_2 есть подмножество S_1 , имеем

$$S_1 + S_2 = S_1 \quad \text{и} \quad S_1 S_2 = S_2.$$

Из симметричного характера наших определений суммы и произведения следует, что операции сложения и умножения *коммутативны*, т. е.

$$S_1 + S_2 = S_2 + S_1 \quad \text{и} \quad S_1 S_2 = S_2 S_1.$$

Далее легко видеть, что эти операции *ассоциативны* и *дистрибутивны*, подобно соответствующим операциям арифметики, т. е.

$$(S_1 + S_2) + S_3 = S_1 + (S_2 + S_3),$$

$$(S_1 S_2) S_3 = S_1 (S_2 S_3),$$

$$S_1 (S_2 + S_3) = S_1 S_2 + S_1 S_3.$$

Отсюда следует, что мы можем говорить о сумме и произведении конечного числа множеств

$$S_1 + S_2 + \dots + S_n \quad \text{и} \quad S_1 S_2 \dots S_n,$$

где порядок слагаемых и сомножителей произволен.

Можно распространить определение этих операций на случай счетной последовательности слагаемых или сомножителей. Пусть дана последовательность S_1, S_2, \dots множеств в \mathcal{S} . Определим *сумму*

$$\sum_{v=1}^{\infty} S_v = S_1 + S_2 + \dots$$

как множество всех элементов, принадлежащих по крайней мере одному из множеств S_v , и *произведение*

$$\prod_{v=1}^{\infty} S_v = S_1 S_2 \dots$$

^{°)} В этом случае мы часто будем говорить, что множества S_1 и S_2 не *пересекаются*. (Прим. перев.)

как множество всех элементов, принадлежащих всем S_v . При этом имеем, например, $S(S_1 + S_2 + \dots) = SS_1 + SS_2 + \dots$.

Так, если S_v есть множество всех действительных чисел, таких, что $\frac{1}{v+1} \leq x \leq \frac{1}{v}$, то легко видеть, что $\sum_1^{\infty} S_v$ есть множество всех x , для которых $0 < x \leq 1$, а $\prod_1^{\infty} S_v$ есть пустое множество. Если же S_v есть множество всех x , для которых $0 \leq x \leq \frac{1}{v}$, то сумма $\sum_1^{\infty} S_v$ совпадает с S_1 , а произведение $\prod_1^{\infty} S_v$ состоит из одного элемента, а именно из числа $x=0$.

Важный частный случай операции вычитания имеет место, когда S_1 совпадает со всем пространством S . Разность

$$S^* = S - S$$

есть множество всех элементов нашего пространства, не принадлежащих S , и называется *дополнительным множеством* или просто *дополнением* к S . Очевидно, что $S + S^* = S$, $SS^* = 0$ и $(S^*)^* = S$.

Важно отметить, что дополнение к данному множеству S зависит от того пространства S , в котором S содержится. Если наше пространство есть множество всех точек прямой L , а S есть множество всех точек этой прямой, имеющих положительную координату, то дополнение S^* состоит из начала координат O и всех точек с отрицательной координатой. Если же наше пространство состоит из всех точек некоторой плоскости P , содержащей L , то дополнение S^* к тому же самому множеству S содержит, кроме того, все точки P , не принадлежащие к L . Во всех случаях, когда существует возможность ошибки, будем употреблять выражение: S^* есть дополнение множества S по отношению к S .

Операции сложения и умножения могут быть выражены одна через другую с помощью понятия дополнительного множества. Действительно, для всякой конечной или счетной последовательности S_1, S_2, \dots мы имеем

$$(1.3.1) \quad (S_1 + S_2 + \dots)^* = S_1^* S_2^* \dots, \\ (S_1 S_2 \dots)^* = S_1^* + S_2^* + \dots$$

Первое соотношение показывает, что *дополнение к сумме множеств есть произведение дополнений к слагаемым*. Это непосредственно

следует из данных нами определений. В самом деле, дополнение $(S_1 + S_2 + \dots)^*$ есть множество всех элементов x пространства, не принадлежащих ни одному из множеств S_v , т. е. принадлежащих всем дополнениям S_v^* , т. е. принадлежащих произведению $S_1^* S_2^* \dots$. Второе соотношение вытекает из первого при замене S_v^* на S_v . Аналогично для операции вычитания получаем

$$(1.3.2) \quad S_1 - S_2 = S_1 S_2^*.$$

Читатель легко убедится, что понимание таких соотношений, как (1.3.1) и (1.3.2), сильно облегчается употреблением рисунков типа фиг. 1.

1.4. Последовательности множеств. Когда мы употребляем слово *последовательность* без дальнейшей конкретизации, то мы имеем в виду *конечную* или *счетную* последовательность; последовательность $S_1, S_2, \dots, S_n, \dots$ часто обозначается кратко $\{S_n\}$.

При рассмотрении суммы последовательности множеств

$$S = S_1 + S_2 + \dots$$

часто бывает полезно представить S в виде суммы последовательности множеств, *никакие два из которых не имеют общих элементов*.

Это можно осуществить следующим образом. Положим

$$\begin{aligned} Z_1 &= S_1, \\ Z_2 &= S_1^* S_2, \\ &\dots \dots \dots \dots \\ Z_v &= S_1^* S_2^* \dots S_{v-1}^* S_v, \\ &\dots \dots \dots \dots \end{aligned}$$

Таким образом, Z_v есть множество всех элементов S_v , не принадлежащих ни одному из предшествующих множеств S_1, \dots, S_{v-1} . Легко видеть, что множества Z_μ и Z_v не имеют общих элементов при $\mu \neq v$. Положим, например, $\mu < v$. Тогда Z_μ есть подмножество S_μ , в то время как Z_v есть подмножество S_v^* , т. е. $Z_\mu Z_v = 0$.

Положим теперь $S' = Z_1 + Z_2 + \dots$. Так как $Z_v \subset S_v$ при любом v , то $S' \subset S$. С другой стороны, пусть x есть элемент из S . Согласно определению, x принадлежит по крайней мере одному из множеств S_v . Пусть S_n есть первое множество из последовательности S_1, S_2, \dots , которое содержит x . Тогда, согласно определению Z_n , x принад-

лежит к Z_n , а следовательно, и к S' . Таким образом, имеем $S \subset S'$ и $S' \subset S$, т. е. $S' = S$ и

$$S = Z_1 + Z_2 + \dots$$

Употребим это преобразование, чтобы показать, что *сумма счетной последовательности счетных множеств есть счетное множество*. Если S_v счетно, то Z_v , как подмножество S_v , должно быть конечно или счетно. Пусть элементы Z_v суть x_{v1}, x_{v2}, \dots . Тогда элементы множества $S = \sum S_v = \sum Z_v$, образуют двойную последовательность

$$\begin{aligned} & x_{11}, x_{12}, x_{13}, \dots \\ & x_{21}, x_{22}, x_{23}, \dots \\ & x_{31}, x_{32}, x_{33}, \dots \\ & \cdot \quad \cdot \quad \cdot \quad \cdot \end{aligned}$$

а последняя может быть представлена в виде простой последовательности, получаемой, например, чтением по диагоналям: $x_{11}, x_{12}, x_{21}, x_{13}, x_{22}, x_{31}, \dots$. Легко видеть, что каждый элемент из S займет определенное место в этой последовательности, и, следовательно, множество S счетно.

1.5. Монотонные последовательности. Последовательность S_1, S_2, \dots называется *неубывающей*, если $S_n \subset S_{n+1}$ для всех n . Если $S_n \supset S_{n+1}$ для всех n , то последовательность называется *невозрастающей*. Последовательности этих двух типов называются *монотонными*.

Для неубывающей последовательности имеем

$$S_n = \sum_1^n S_v,$$

и это делает естественным определение *предела* такой последовательности как

$$\lim_{n \rightarrow \infty} S_n = \sum_1^{\infty} S_v.$$

Аналогично, для невозрастающей последовательности имеем

$$S_n = \prod_1^n S_v,$$

и, соответственно, определяем в этом случае

$$\lim_{n \rightarrow \infty} S_n = \prod_1^{\infty} S_v.$$

Пусть S_n есть множество всех точек (x, y, z) , расположенных внутри сферы $x^2 + y^2 + z^2 = 1 - \frac{1}{n}$. Тогда последовательность S_1, S_2, \dots — неубывающая, а $\text{Hm } S_n$ есть множество всех внутренних точек сферы $x^2 + y^2 + z^2 = 1$. С другой стороны, если S_n есть множество всех внутренних точек сферы $x^2 + y^2 + z^2 = 1 + \frac{1}{n}$, то последовательность $\{S_n\}$ невозрастающая, и $\text{Hm } S_n$ есть множество всех внутренних точек и всех точек поверхности сферы $x^2 + y^2 + z^2 = 1$.

Можно распространить определение предела также на некоторые типы последовательностей, которые не являются монотонными. Однако такие случаи не будут встречаться в этой книге.

1.6. Аддитивные классы множеств. Если дано пространство S , то мы можем рассматривать различные *классы множеств* в S . Важное значение будет иметь понятие *аддитивного класса множеств* в S . Класс \mathfrak{E} множеств в S называется *аддитивным* *), если он удовлетворяет следующим трем условиям:

- Все пространство S принадлежит \mathfrak{E} .
- Если все множества последовательности S_1, S_2, \dots принадлежат \mathfrak{E} , то сумма $S_1 + S_2 + \dots$ и произведение $S_1 S_2 \dots$ принадлежат \mathfrak{E} .
- Если S_1 и S_2 принадлежат \mathfrak{E} и $S_2 \subset S_1$, то разность $S_1 - S_2$ принадлежит \mathfrak{E} .

Если \mathfrak{E} есть аддитивный класс, то, производя конечное или счетное число раз операции сложения, умножения и вычитания над элементами класса \mathfrak{E} , мы никогда не выйдем за пределы класса \mathfrak{E} .

Нужно отметить, что приведенные выше три условия не являются независимыми друг от друга. В самом деле, соотношения (1.3.1) и (1.3.2) показывают, что условия а), б), в) эквивалентны следующие условия:

- Все пространство S принадлежит \mathfrak{E} .
- Если все множества последовательности S_1, S_2, \dots принадлежат \mathfrak{E} , то и сумма $S_1 + S_2 + \dots$ принадлежит \mathfrak{E} .

*) В этой книге мы будем всегда употреблять слово „аддитивный“ в том смысле, как в этом параграфе, т. е. по отношению к *конечной* или *счетной* последовательности слагаемых. Нужно отметить, что некоторые авторы употребляют в этом случае термин „полная аддитивность“, в то время как выражения „аддитивность“ или „простая аддитивность“ употребляются для обозначения этого свойства лишь по отношению к *конечному* числу слагаемых.

c_1) Если S принадлежит \mathfrak{E} , то и дополнительное множество S^* принадлежит \mathfrak{E} .

Название „аддитивный класс“ обусловлено тем, что среди приведенных условий важнейшую роль играет условие аддитивности b_1).

Класс всевозможных подмножеств S есть очевидный пример аддитивного класса. В следующей главе мы встретимся с более интересным примером.

ГЛАВА 2

ЛИНЕЙНЫЕ ТОЧЕЧНЫЕ МНОЖЕСТВА

2.1. Интервалы. Пусть наше пространство есть множество R_1 всех точек данной прямой. Любое его подмножество называется *линейным точечным множеством*. Если мы выберем на нашей прямой начало координат, единицу измерения и положительное направление, то, как хорошо известно, можно установить взаимно-однозначное соответствие между всеми действительными числами и всеми точками прямой. Таким образом, мы можем говорить, не делая различия, о точке x прямой или о действительном числе x , соответствующем этой точке. Отметим, что мы рассматриваем лишь точки, соответствующие *конечным* числам.

Простейшим примером линейного точечного множества является *интервал*. Если a и b — некоторые точки, такие, что $a \leq b$, то мы употребляем следующие термины для обозначения множеств всех точек x , удовлетворяющих указанным неравенствам:

$a \leq x \leq b$ — *замкнутый интервал* (a, b) ;

$a < x < b$ — *открытый интервал* (a, b) ;

$a < x \leq b$ — *полуоткрытый интервал* (a, b) , *замкнутый справа*;

$a \leq x < b$ — *полуоткрытый интервал* (a, b) , *замкнутый слева*.

Если мы говорим просто об интервале (a, b) без дальнейших указаний относительно его типа, то подразумевается, что все сказанное справедливо для всех четырех типов интервалов.

В предельном случае, когда $a = b$, мы говорим о *вырожденном* интервале. В этом случае замкнутый интервал стягивается в множество, состоящее из единственной точки $x = a$, а каждый из остальных трех интервалов есть пустое множество.

Если в приведенных выше неравенствах мы позволим b стремиться к $+\infty$, то получим неравенства, определяющие соответственно *замкнутый и открытый бесконечный интервал* $(a, +\infty)$:

$$x \geq a \text{ и } x > a.$$

Аналогично, если a стремится к $-\infty$, то мы получим неравенства

$$x \leq b \quad \text{и} \quad x < b$$

для замкнутого и открытого бесконечного интервала $(-\infty, b)$. Наконец, все пространство R_1 можно рассматривать как бесконечный интервал $(-\infty, +\infty)$.

Ниже будет показано (см. параграф 4.3), что *всякий невырожденный интервал есть несчетное множество*.

Произведение конечной или счетной последовательности интервалов есть интервал, но сумма двух интервалов, вообще говоря, не является интервалом. Чтобы привести пример, когда сумма интервалов есть интервал, рассмотрим $n+1$ точек $a < x_1 < \dots < x_{n-1} < b$. Если все интервалы, встречающиеся в следующем соотношении, полуоткрыты и замкнуты с одной и той же стороны, то мы, очевидно, имеем

$$(a, b) = (a, x_1) + (x_1, x_2) + \dots + (x_{n-1}, b),$$

и никакие два интервала в правой части этого соотношения не имеют общих точек. Это соотношение выполнено и тогда, когда все интервалы замкнуты, но в этом случае каждые два соседних слагаемых имеют одну общую точку. С другой стороны, если все интервалы открыты, то наше соотношение не справедливо.

2.2. Некоторые свойства множеств в R_1 . Рассмотрим некоторое непустое множество S . Если точка a такова, что для любого $\epsilon > 0$ существует по крайней мере одна точка из S , принадлежащая замкнутому интервалу $(a, a+\epsilon)$, и в то же время не существует точки из S , принадлежащей открытому интервалу $(-\infty, a)$, то мы называем a *нижней границей множества S* . Если не существует конечного a , обладающего этим свойством, то мы говорим, что нижняя граница множества S есть $-\infty$. Аналогично мы определим *верхнюю границу β* множества S . Множество называется *ограниченным*, если оно имеет конечные верхнюю и нижнюю границы. Ограниченнное множество есть, таким образом, подмножество замкнутого интервала (α, β) . Самы точки α и β могут как принадлежать, так и не принадлежать множеству S .

Если ϵ есть произвольное положительное число, то открытый интервал $(x-\epsilon, x+\epsilon)$ называется *окрестностью* точки x или, точнее, *ϵ -окрестностью* точки x .

Точка z называется *пределной точкой* для множества S , если каждая окрестность z содержит по крайней мере одну точку из S .

отличную от z . Если это условие выполнено, то легко видеть, что в действительности каждая окрестность z содержит даже бесконечное множество точек из S . При этом сама точка z может и не принадлежать к S . Теорема Больцано — Вейерштрасса утверждает, что *каждое ограниченное бесконечное множество имеет по крайней мере одну предельную точку*. Мы предполагаем это уже известным. Если z есть предельная точка, то множество S содержит последовательность точек x_1, x_2, \dots , такую, что $x_n \rightarrow z$ при $n \rightarrow \infty$.

Точка x из S называется *внутренней точкой* множества S , если существует такое ϵ , при котором ϵ -окрестность x целиком содержится в S . Очевидно, что внутренняя точка всегда является предельной.

Дадим теперь примеры, поясняющие введенные выше понятия. Пусть сперва S есть конечный невырожденный интервал (a, b) . Тогда a есть нижняя, а b — верхняя граница множества S . Каждая точка, принадлежащая *замкнутому* интервалу (a, b) , есть предельная точка для S , а каждая точка, принадлежащая *открытым* интервалу (a, b) , есть внутренняя точка множества S .

Рассмотрим теперь множество R всех рациональных точек $x = \frac{p}{q}$, принадлежащих полуоткрытыму интервалу $0 < x \leq 1$. Если мы выпишем последовательность

$$\begin{aligned} &\frac{1}{1}, \\ &\frac{1}{2}, \frac{2}{2}, \\ &\frac{1}{3}, \frac{2}{3}, \frac{3}{3}, \\ &\frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{4}{4}, \\ &\dots \end{aligned}$$

и затем выбросим все числа $\frac{p}{q}$, в которых p и q имеют общий сомножитель, то каждая точка из R встретится в нашей последовательности ровно один раз, и, следовательно, множество R счетно. Множество R не содержит внутренних точек. Каждая точка замкнутого интервала $(0, 1)$ есть предельная точка для R . Дополнение R^* множества R по отношению к интервалу $0 < x \leq 1$ есть множество всех иррациональных точек этого интервала. R^* несчетно, так как в противном случае интервал $(0, 1)$ был бы суммой двух счетных множеств и, следовательно, сам был бы счетным множеством. Подобно самому R , множество R^* не имеет внутренних точек, и каждая точка замкнутого интервала $(0, 1)$ является предельной для R^* .

Из того, что множество R счетно, непосредственно вытекает, что множество R_n всех рациональных точек x , принадлежащих интервалу $n < x \leq n + 1$,

счетно при любом целом n . Из предложения, доказанного в параграфе 1.4, следует, что *множество всех положительных и отрицательных рациональных чисел счетно*. Действительно, последнее множество есть сумма последовательности $\{R_n\}$, где n принимает все целые значения и, следовательно, в силу параграфа 1.4, есть счетное множество.

2.3. Борелевские множества. Рассмотрим класс всех интервалов в R_1 — замкнутых, открытых, полуоткрытых, вырожденных и невырожденных, конечных и бесконечных, включающий в себя, в частности, само пространство R_1 . Очевидно, что этот класс множеств не является аддитивным, так как сумма двух интервалов не является, вообще говоря, интервалом. *Попытаемся построить аддитивный класс множеств посредством расширения этого класса интервалов.*

В качестве первого обобщения рассмотрим класс \mathfrak{J} всех множеств I , являющихся суммами конечных или счетных последовательностей интервалов. Если I_1, I_2, \dots суть множества, принадлежащие классу \mathfrak{J} , то, согласно 1.4, сумма $I_1 + I_2 + \dots$ есть также сумма конечной или счетной последовательности интервалов и, следовательно, сама принадлежит классу \mathfrak{J} . То же имеет место и для конечного произведения $I_1 I_2 \dots I_n$ вследствие свойства дистрибутивности (см. параграф 1.3). Мы покажем, однако, на примерах, что как бесконечное произведение $I_1 I_2 \dots$, так и разность $I_1 - I_2$ могут и не принадлежать к \mathfrak{J} . Действительно, множество R , рассмотренное в предыдущем параграфе, принадлежит \mathfrak{J} , так как оно является суммой счетной последовательности вырожденных интервалов, каждый из которых содержит единственную точку $\frac{p}{q}$. С другой стороны, разность $(0, 1) - R$ не содержит невырожденного интервала, а если мы попытаемся представить его в виде суммы вырожденных интервалов, то таких интервалов потребуется несчетное множество. Таким образом, эта разность не принадлежит к \mathfrak{J} . Далее, эта разность может быть представлена в виде произведения $I_1 I_2 \dots$, где I_n есть разность между интервалом $(0, 1)$ и множеством, содержащим только n -ю точку множества R . Таким образом, само это произведение множеств из \mathfrak{J} не принадлежит классу \mathfrak{J} .

Хотя в главе 4 класс \mathfrak{J} найдет себе важное применение, ясно, что для нашей теперешней цели этот класс недостаточен. Для того, чтобы построить аддитивный класс, нужно класс \mathfrak{J} расширить, дополнив его множествами более общего характера.

Если мы добавим к \mathfrak{J} все суммы и произведения последовательностей множеств из \mathfrak{J} и все разности между теми множествами из \mathfrak{J} ,

для которых разность определена (некоторые из таких множеств, конечно, могут уже содержаться в \mathfrak{X}), то мы получим расширенный класс множеств. Однако можно показать, что даже этот расширенный класс не будет удовлетворять всем условиям аддитивного класса. Мы, таким образом, должны повторять этот процесс расширения снова и снова, не приходя к концу. Множество, полученное на некотором этапе этого процесса, обладает тем свойством, что оно может быть получено, отправляясь от интервалов, выполнением операций сложения, умножения и вычитания в конечном или счетном числе. Вся совокупность множеств, получаемых таким образом, называется классом \mathfrak{B}_1 борелевских множеств в R_1 и является аддитивным классом. В самом деле, каждое данное борелевское множество может быть образовано, как было описано выше, посредством не более чем счетного числа шагов, и любая сумма, произведение или разность таких множеств будет содержаться в классе всех множеств, получаемых этим путем.

Таким образом, сумма, произведение и разность борелевских множеств сами являются борелевскими множествами. В частности, предел монотонной последовательности (см. параграф 1.5) борелевских множеств всегда есть борелевское множество.

С другой стороны, пусть \mathfrak{E} — произвольный аддитивный класс множеств из R_1 , содержащий все интервалы. Тогда непосредственно из определения аддитивного класса следует, что \mathfrak{E} должно содержать каждое множество, которое может быть получено из интервалов конечным или счетным повторением операций сложения, умножения и вычитания. Таким образом, \mathfrak{E} должно содержать весь класс \mathfrak{B}_1 борелевских множеств, и можно сказать, что *класс \mathfrak{B}_1 есть наименьший аддитивный класс множеств из R_1 , заключающий в себе все интервалы.*

ГЛАВА 3

ТОЧЕЧНЫЕ МНОЖЕСТВА В ПРОСТРАНСТВЕ R_n

3.1. Интервалы. Хорошо известно, что наряду с взаимно-однозначным соответствием между действительными числами x и всеми точками прямой линии можно установить аналогичное соответствие между парами действительных чисел (x_1, x_2) и всеми точками плоскости, а также между тройками действительных чисел (x_1, x_2, x_3) и всеми точками трехмерного пространства.

Обобщая, можно рассматривать систему из n действительных чисел (x_1, x_2, \dots, x_n) как точку или вектор x евклидова n -мерного пространства R_n . Числа x_1, \dots, x_n называются координатами точки x . Как и в одномерном случае, мы рассматриваем лишь точки, соответствующие конечным значениям координат. Расстоянием между точками

$$x = (x_1, \dots, x_n) \text{ и } y = (y_1, \dots, y_n)$$

называется неотрицательная величина

$$|x - y| = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}.$$

Расстояние удовлетворяет так называемому неравенству треугольника:

$$|x - y| \leq |x - z| + |y - z|.$$

Пусть даны $2n$ чисел a_1, \dots, a_n и b_1, \dots, b_n , таких, что $a_\nu \leq b_\nu$, для $\nu = 1, \dots, n$. Множество всех точек x , определенных соотношениями $a_\nu \leq x_\nu \leq b_\nu$, ($\nu = 1, \dots, n$), называется замкнутым n -мерным интервалом. Если все знаки \leq заменить на $<$, то получим открытый интервал, а если в указанных неравенствах встречаются оба вида знаков, то имеем полуоткрытый интервал. В предельном случае, когда $a_\nu = b_\nu$, хотя бы для одного значения ν , интервал называется вырожденным. Если одно или несколько из чисел a_ν стремятся к $-\infty$ или одно или несколько из чисел b_ν стремятся к $+\infty$, то мы получим бесконечный интервал. Как и в параграфе 2.1, все пространство R_n можно рассматривать, как бесконечный интервал.

Ниже мы покажем (см. параграф 4.3), что любой невырожденный интервал есть несчетное множество. Произведение конечной или счетной последовательности интервалов есть также интервал, но сумма двух интервалов не является, вообще говоря, интервалом.

3.2. Некоторые свойства множеств из R_n . Множество S в R_n называется ограниченным, если все точки из S содержатся в конечном интервале.

Если $a = (a_1, \dots, a_n)$ есть данная точка, а ε — положительное число, то множество всех точек x , таких, что $|x - a| < \varepsilon$, называется окрестностью или, точнее, ε -окрестностью точки a .

Определения понятий предельной и внутренней точек и сделанные в параграфе 2.2 в связи с этими понятиями для случая $n = 1$ замечания применяются без изменения в рассматриваемом здесь общем случае.

Мы видели в параграфе 2.2, что множество всех рациональных чисел в R_1 счетно. В силу параграфа 1.4, отсюда вытекает, что множество всех точек плоскости с рациональными координатами счетно, и, далее, по индукции, что *множество всех точек R_n с рациональными координатами счетно*.

3.3. Борелевские множества. Класс интервалов в R_n , так же как и соответствующий класс в R_1 , не является аддитивным классом множеств. Для того чтобы расширением этого класса образовать аддитивный класс множеств, мы поступаем так же, как и в случае интервалов в R_1 .

Так, мы рассматриваем сперва класс \mathfrak{I}_n всех множеств I , которые являются суммами конечных или счетных последовательностей интервалов в R_n . Если I_1, I_2, \dots — множества, принадлежащие к этому классу, то сумма $I_1 + I_2 + \dots$ и конечное произведение $I_1 I_2 \dots I_n$ также принадлежат к \mathfrak{I}_n . Так же как и в случае $n = 1$, бесконечное произведение $I_1 I_2 \dots$ и разность $I_1 - I_2$ могут и не принадлежать к \mathfrak{I}_n .

Расширяя класс \mathfrak{I}_n присоединением всех сумм, произведений и разностей множеств из \mathfrak{I}_n и повторяя этот процесс расширения снова и снова, мы приходим к множествам, которые можно получить из интервалов, производя операции сложения, умножения и вычитания конечное или счетное число раз. *Совокупность множеств, которые могут быть получены этим путем, называется классом \mathfrak{B}_n борелевских множеств в R_n ; \mathfrak{B}_n является аддитивным классом.*

Так же как и в случае $n = 1$, класс \mathfrak{B}_n есть наименьший аддитивный класс множеств из R_n , заключающий в себе все интервалы.

3.4. Линейные множества. В случае $n > 3$ множество всех точек пространства R_n , удовлетворяющих некоторому уравнению $F(x_1, \dots, x_n) = 0$, называется *гиперповерхностью*. Если F есть линейная функция, то соответствующая гиперповерхность называется *гиперплоскостью*. Уравнение гиперплоскости может быть всегда записано в виде

$$a_1(x_1 - m_1) + \dots + a_n(x_n - m_n) = 0,$$

где $m = (m_1, \dots, m_n)$ есть произвольная точка гиперплоскости.

Пусть

$$(3.4.1) \quad H_i = a_{i1}(x_1 - m_1) + \dots + a_{in}(x_n - m_n) = 0, \quad (i = 1, 2, \dots, p)$$

суть уравнения p гиперплоскостей, проходящих через точку m . Уравнения (3.4.1) называются *линейно независимыми*, если не существует

линейной комбинации $k_1H_1 + \dots + k_pH_p$, где k_i — постоянные, не все равные нулю, которая была бы тождественно равна нулю. Соответствующие гиперплоскости также называются линейно независимыми.

Пусть $p < n$. Рассмотрим множество L всех точек пространства \mathbf{R}_n , принадлежащих к p линейно независимым гиперплоскостям (3.4.1). Если рассматривать (3.4.1) как систему линейных уравнений с неизвестными x_1, \dots, x_n , то ее общее решение (см. параграф 11.8) есть

$$x_i = m_i + c_{i1}t_1 + \dots + c_{i,n-p}t_{n-p},$$

где c_{ik} — постоянные, зависящие от коэффициентов a_{ik} , а t_1, \dots, t_{n-p} — произвольные параметры.

Таким образом, координаты точки множества L могут быть представлены линейными функциями от $n-p$ независимых параметров. Поэтому множество L называется *линейным множеством размерности $n-p$* и будет обычно обозначаться через L_{n-p} . Для $p=1$ это гиперплоскость, для $p=n-2$ — обычная плоскость, а для $p=n-1$ — прямая. Обратно, если L_{n-p} есть линейное множество размерности $n-p$ и если $m=(m_1, \dots, m_n)$ есть произвольная точка L_{n-p} , то L_{n-p} может быть представлено как общая часть (или произведение в смысле параграфа 1.3) p линейно независимых гиперплоскостей, проходящих через m .

3.5. Подпространства. Произведение пространств. Рассмотрим пространство \mathbf{R}_n всех точек $x=(x_1, \dots, x_n)$. Выберем группу из $k < n$ координат, скажем x_1, \dots, x_k , и положим все оставшиеся $n-k$ координат равными нулю: $x_{k+1} = \dots = x_n = 0$. Таким образом, мы получим систему из $n-k$ линейно независимых уравнений, которые определяют линейное множество L_k размерности k . Оно называется *k-мерным подпространством, порожденным координатами x_1, \dots, x_k* . Подпространство, порожденное другой группой из k координат, определяется аналогичным образом. Так, в случае $n=3$, $k=2$ двумерное подпространство, порожденное координатами x_1 и x_2 , есть просто плоскость (x_1, x_2) .

Пусть S есть множество в k -мерном подпространстве, порожденном координатами x_1, \dots, x_k . Множество всех точек x из \mathbf{R}_n , таких, что $(x_1, \dots, x_k, 0, \dots, 0) \subset S$, называется *цилиндрическим множеством с основанием S* . В случае $n=3$, $k=2$ это обычный трех-

мерный цилиндр, имеющий своим основанием множество S в плоскости (x_1, x_2) .

Далее, если S_1 и S_2 — множества, расположенные соответственно в подпространствах, порожденных координатами x_1, \dots, x_k и x_{k+1}, \dots, x_n , то множество всех точек x из R_n , таких, что $(x_1, \dots, x_k, 0, \dots, 0) \subset S_1$ и $(0, \dots, 0, x_{k+1}, \dots, x_n) \subset S_2$, называется *прямоугольным множеством со сторонами* S_1 и S_2 . В случае, когда $n=2$, а S_1 и S_2 — одномерные интервалы, это обычный прямоугольник на плоскости (x_1, x_2) .

Наконец, пусть R_m и R_n суть, соответственно, m - и n -мерные пространства. Рассмотрим множество всех пар точек (x, y) , где $x = (x_1, \dots, x_m)$ — точка в R_m , а $y = (y_1, \dots, y_n)$ — точка в R_n . Это множество называется *произведением пространств* R_m и R_n . Оно является $(m+n)$ -мерным пространством, содержащим все точки $(x_1, \dots, x_m, y_1, \dots, y_n)$ в качестве своих элементов. Так, беря $m=n=1$, мы видим, что плоскость (x_1, x_2) можно рассматривать как произведение двух одномерных пространств — прямых x_1 и x_2 . При $m=2, n=1$ мы получим трехмерное пространство (x_1, x_2, x_3) как произведение плоскости (x_1, x_2) на одномерное пространство x_3 и т. д. Распространение данного нами определения произведения пространств на случай больше чем двух пространств-сомножителей не представляет трудностей. Заметим, что введенное здесь произведение пространств есть нечто совершенно отличное от произведения множеств, определенного в параграфе 1.3.

Литература к главам 1—3. Теория точечных множеств была основана Г. Кантором около 1880 г. Она имеет фундаментальное значение для многих областей математики, например для современной теории интегрирования и теории функций. Большинство руководств по этим вопросам содержит главы о точечных множествах. Читатель может обратиться, например, к книгам Бореля [6] и Валле-Пуссена [40]^{1)*}.

¹⁾ Цифры в квадратных скобках относятся к списку цитированной литературы в конце книги. (*Прим. перев.*)

* На русском языке см. П. С. Александров и А. Н. Колмогоров „Введение в теорию функций действительного переменного“ или Ф. Хаусдорф „Теория множеств“. (*Прим. ред.*)

ГЛАВЫ 4—7

ТЕОРИЯ МЕРЫ И ИНТЕГРИРОВАНИЯ в R_1

ГЛАВА 4

МЕРА ЛЕБЕГА ЛИНЕЙНЫХ ТОЧЕЧНЫХ МНОЖЕСТВ

4.1. Длина интервала. *Длиной* конечного интервала (a, b) в R . называется неотрицательная величина $b - a$. Таким образом, длина имеет одно и то же значение как для замкнутого, так и для открытого и полуоткрытого интервалов, имеющих одни и те же концы. Длина вырожденного интервала равна нулю. Длину бесконечного интервала мы, по определению, полагаем равной $+\infty$.

Таким образом, каждому интервалу мы ставим в соответствие определенную неотрицательную длину, которая может быть конечной или бесконечной. Это можно выразить, сказав, что длина $L(i)$ есть *неотрицательная функция интервала* i , и записав

$$L(i) = b - a \quad \text{или} \quad L(i) = +\infty,$$

в зависимости от того, конечен или бесконечен интервал i . Если интервал i есть сумма конечного числа попарно непересекающихся интервалов (см. параграф 2.1):

$$i = i_1 + i_2 + \dots + i_n \quad (i_\mu i_\nu = 0 \text{ при } \mu \neq \nu),$$

то длина всего интервала i равна, очевидно, сумме длин интервалов i_k :

$$L(i) = L(i_1) + L(i_2) + \dots + L(i_n).$$

Покажем теперь, что это соотношение справедливо и в случае счетной последовательности интервалов i_k . Для читателя, который сталкивается с подобными вопросами впервые, это утверждение может показаться тривиальным. Тщательное изучение приведенного ниже доказательства может убедить его в том, что это не так. Для того чтобы дать строгое доказательство нашего утверждения, мы нуждаемся в следующем предложении, известном под названием *леммы Бореля*.

Пусть нам дан конечный замкнутый интервал (a, b) и множество Z интервалов, таких, что каждая точка интервала (a, b) есть

внутренняя точка по крайней мере одного из интервалов, принадлежащих к Z . Тогда существует подмножество Z' множества Z , содержащее только конечное число интервалов, и такое, что каждая точка интервала (a, b) есть внутренняя точка по крайней мере одного из интервалов, принадлежащих Z' .

Разделим интервал (a, b) на n частей равной длины. Лемма будет доказана, если мы сможем показать, что можно подобрать n так, чтобы каждая из n частей (рассматриваемая как замкнутый интервал) целиком содержалась в некотором интервале, принадлежащем Z .

Действительно, предположим, что такого n не существует. Обозначим через i_n первую из n частей (считая от конца a), не содержащуюся целиком ни в каком интервале, принадлежащем к Z . Длина i_n , очевидно, стремится к нулю при $n \rightarrow \infty$. Обозначим середину интервала i_n через x_n и рассмотрим последовательность точек x_1, x_2, \dots . Так как она ограничена, то, по теореме Больцано—Вейерштрасса (см. параграф 2.2), она имеет предельную точку x . Любая окрестность этой точки x содержит некоторый интервал i_n , который не содержит целиком в интервале, принадлежащем к Z . С другой стороны, x есть точка интервала (a, b) и, следовательно, по предположению, есть внутренняя точка некоторого интервала, принадлежащего Z . Полученное противоречие доказывает лемму.

Очевидно, что как сама лемма, так и приведенное выше ее доказательство могут быть непосредственно обобщены на случай пространства произвольной размерности.

Рассмотрим теперь последовательность попарно непересекающихся интервалов $i_\nu = (a_\nu, b_\nu)$, таких, что сумма всех i_ν есть *конечный* интервал $i = (a, b)$:

$$i = \sum_{\nu=1}^{\infty} i_\nu, \quad (i_\mu i_\nu = 0 \text{ при } \mu \neq \nu).$$

Докажем, что

$$(4.1.1) \quad L(i) = \sum_{\nu=1}^{\infty} L(i_\nu).$$

Каково бы ни было n , интервалы i_1, \dots, i_n попарно не пересекаются и содержатся в i ; поэтому

$$\sum_{\nu=1}^n L(i_\nu) \leq L(i).$$

Переходя к пределу при $n \rightarrow \infty$, получим

$$\sum_1^{\infty} L(i_v) \leq L(i).$$

Остается доказать обратное неравенство. Эта часть доказательства не тривиальна.

Рассмотрим множество Z , состоящее из: 1) интервалов i_v , 2) открытых интервалов $(a - \epsilon, a + \epsilon)$ и $(b - \epsilon, b + \epsilon)$, где ϵ положительно и произвольно мало, 3) открытых интервалов $\left(a_v - \frac{\epsilon}{2^v}, a_v + \frac{\epsilon}{2^v}\right)$ и $\left(b_v - \frac{\epsilon}{2^v}, b_v + \frac{\epsilon}{2^v}\right)$, где $v = 1, 2, \dots$. Тогда очевидно, что каждая точка замкнутого интервала (a, b) есть внутренняя точка по крайней мере одного интервала, принадлежащего Z . В силу леммы Бореля, мы можем покрыть i конечным числом интервалов, принадлежащих Z , и сумма длин этих интервалов должна быть больше, чем $L(i) = b - a$. Сумма длин всех интервалов, принадлежащих Z , должна, тем более, превосходить $L(i)$, так что

$$\sum_1^{\infty} L(i_v) + 4\epsilon + 4 \sum_1^{\infty} \frac{\epsilon}{2^v} = \sum_1^{\infty} L(i_v) + 8\epsilon > L(i).$$

Отсюда, так как ϵ произвольно,

$$\sum_1^{\infty} L(i_v) \geq L(i),$$

т. е. соотношение (4.1.1) доказано.

Легко далее показать, что соотношение (4.1.1) имеет место и в случае, когда i — бесконечный интервал. В этом случае имеем $L(i) = +\infty$, и если i_0 есть некоторый конечный интервал, содержащийся в i , то из последней части приведенного выше доказательства вытекает, что

$$\sum_1^{\infty} L(i_v) \geq L(i_0).$$

Так как интервал i бесконечен, то можно выбрать i_0 так, чтобы $L(i_0)$ было больше любой наперед данной величины, и, таким образом, равенство (4.1.1) будет справедливо в том смысле, что обе его части равны $+\infty$.

Таким образом мы доказали, что если интервал разбит на конечное или счетное число попарно непересекающихся интервалов, то длина всего интервала равна сумме длин его частей. Часто выражают это свойство, говоря, что длина $L(i)$ есть *аддитивная функция интервала i* .

4.2. Обобщение. Длина интервала есть *мера* его протяженности. Мы видели в предшествующем параграфе, что основные свойства этой меры суть ее *неотрицательность* и *аддитивность*. Длина интервала $L(i)$ есть *неотрицательная аддитивная функция интервала i* . Значения этой функции могут быть как конечными, так и бесконечными.

Займемся теперь вопросом о возможности определения меры, обладающей теми же основными свойствами для множеств более сложных, чем интервалы. Со всяким множеством S , более или менее общего типа, мы хотим связать конечное или бесконечное*) число $L(S)$ — *меру* множества S , так, чтобы были выполнены следующие три условия:

- $L(S) \geq 0$.
- Если $S = S_1 + S_2 + \dots$ и $S_\mu S_\nu = 0$ (при $\mu \neq \nu$), то $L(S) = L(S_1) + L(S_2) + \dots$
- В частном случае, когда S есть интервал, $L(S)$ совпадает с длиной интервала S .

Таким образом, мы хотим *расширить определение* функции интервала $L(i)$ так, чтобы получить *неотрицательную аддитивную функцию множества $L(S)$* , которая в частном случае, когда S есть интервал i , совпадает с $L(i)$.

Естественно спросить, почему это расширение ограничивается „более или менее общим классом множеств“ и почему нельзя попытаться определить $L(S)$ для *любого* множества S . Можно, однако, показать, что это невозможно. Поэтому мы удовольствуемся тем, что покажем, что *функция множества $L(S)$, удовлетворяющая требованиям а), б), с), может быть определена для класса множеств, заключающего в себе весь класс \mathfrak{B} , борелевских множеств*. Такая функция $L(S)$ называется *лебеговой мерой* множества S . Мы покажем далее, что

*) Для функции множества $L(S)$ и более общих функций множества, рассматриваемых в главе 6, мы будем допускать *бесконечные значения*. С другой стороны, для точечных множеств и обычных функций мы будем иметь дело с бесконечностью только в смысле предела, а не как с определенной точкой или значением (см. параграфы 2.1 и 3.1).

существует только одна функция множества $L(S)$, определенная для всех борелевских множеств и удовлетворяющая условиям а), б), с).

4.3. Мера суммы интервалов. Определим сперва меру $L(I)$ для множеств I , принадлежащих к классу \mathfrak{J} , рассмотренному в параграфе 2,3. Каждое множество из \mathfrak{J} есть сумма конечной или счетной последовательности интервалов. Согласно параграфу 1.4, эти интервалы можно выбрать так, чтобы они попарно не пересекались. (Действительно, если множества S_v , рассматривавшиеся в параграфе 1.4, суть интервалы, то каждое Z , будет суммой конечного числа попарно не пересекающихся интервалов.)

Таким образом, любое множество из \mathfrak{J} может быть представлено в виде

$$(4.3.1) \quad I = i_1 + i_2 + \dots,$$

где i_μ — интервалы, такие, что $i_\mu i_\nu = 0$ при $\mu \neq \nu$. Согласно условиям б) и с) параграфа 4.2, мы должны определить меру $L(I)$ так:

$$(4.3.2) \quad L(I) = L(i_1) + L(i_2) + \dots,$$

где, как и прежде, $L(i_\mu)$ обозначает длину интервала i_μ .

Очевидно, однако, что представление I в форме (4.3.1) не единственно. Пусть

$$(4.3.3) \quad I = j_1 + j_2 + \dots$$

другое представление того же множества I в виде суммы попарно непересекающихся интервалов j_ν . Мы должны показать, что (4.3.1) и (4.3.3) приводят к одному и тому же значению $L(I)$, т. е.

$$(4.3.4) \quad \sum_{\mu=1}^{\infty} L(i_\mu) = \sum_{\nu=1}^{\infty} L(j_\nu).$$

Это может быть доказано следующим образом. Так как $i_\mu \subset I$, то для каждого интервала i_μ имеем

$$i_\mu = i_\mu I = i_\mu \sum_{\nu} j_\nu = \sum_{\nu} i_\mu j_\nu$$

и, таким образом, в силу аддитивности длины,

$$(4.3.5) \quad \begin{aligned} L(i_\mu) &= \sum_{\nu} L(i_\mu j_\nu), \\ \sum_{\mu} L(i_\mu) &= \sum_{\mu} \sum_{\nu} L(i_\mu j_\nu). \end{aligned}$$

Тем же способом получим

$$(4.3.6) \quad \sum_v L(i_v) = \sum_v \sum_{\mu} L(i_{\mu} j_v).$$

Теперь могут представиться три случая: 1) интервалы $i_{\mu} j_v$ все конечны, и двойной ряд $\sum_{\mu, v} L(i_{\mu} j_v)$ с неотрицательными членами сходится; 2) все интервалы $i_{\mu} j_v$ конечны, но указанный двойной ряд расходится; 3) по крайней мере один из интервалов $i_{\mu} j_v$ бесконечен.

В случае 1) выражения, стоящие в правых частях (4.3.5) и (4.3.6), конечны и равны, следовательно, соотношение (4.3.4) выполняется. В случаях 2) и 3) те же выражения оба бесконечны. Итак, в любом случае справедливость (4.3.4) доказана, и, следовательно, определение (4.3.2) приводит однозначно к некоторому (конечному или бесконечному) значению $L(I)$.

Очевидно, что определенная таким образом мера $L(I)$ удовлетворяет условиям а) и с) параграфа 4.2. Остается показать, что условие б) также выполнено.

Пусть I_1, I_2, \dots — последовательность множеств из \mathfrak{J} , таких, что $I_{\mu} I_v = 0$ при $\mu \neq v$, и пусть

$$I_{\mu} = \sum_v i_{\mu v}$$

будет представлением I_{μ} в виде суммы неперекрывающихся интервалов. Тогда

$$I = \sum_{\mu} I_{\mu} = \sum_{\mu} \sum_v i_{\mu v}$$

есть также множество из \mathfrak{J} , и интервалы $i_{\mu v}$ попарно не пересекаются. Если теперь представить двойную последовательность $i_{\mu v}$ в виде простой последовательности i', i'', \dots , (например, нумеруя $i_{\mu v}$ по диагоналям, как указано в параграфе 1.4), то мы будем иметь

$$\begin{aligned} I &= i' + i'' + \dots, \\ L(I) &= L(i') + L(i'') + \dots . \end{aligned}$$

Исследование возможных случаев (аналогичное проведенному выше) покажет, что

$$L(I) = \sum_{\mu} \sum_v L(i_{\mu v}) = \sum_{\mu} L(I_{\mu}).$$

Итак, мы доказали, что соотношение (4.3.2) определяет единственным образом для всех множеств I , принадлежащих к классу \mathfrak{J} , меру $L(I)$, удовлетворяющую условиям а), б), с) параграфа 4.2.

Выполним теперь некоторые свойства меры $L(I)$. Прежде всего рассмотрим последовательность I_1, I_2, \dots множеств из \mathfrak{F} , не предполагая, что множества I_μ и I_ν не пересекаются. Для суммы $I = I_1 + I_2 + \dots$ мы получим, как и выше, представление $I = I' + I'' + \dots$, но интервалы I', I'', \dots могут теперь иметь общие точки. Тогда, применяя преобразование, изложенное в параграфе 1.4, легко видеть, что всегда

$$L(I) \leq L(I') + L(I'') + \dots,$$

что дает

$$(4.3.7) \quad L(I_1 + I_2 + \dots) \leq L(I_1) + L(I_2) + \dots.$$

(В частном случае, когда $I_\mu I_\nu = 0$ при $\mu \neq \nu$, как мы уже видели, это соотношение превращается в равенство.)

Заметим, далее, что любая счетная последовательность точек x_1, x_2, \dots есть множество, входящее в класс \mathfrak{F} , так как каждую точку x_n можно рассматривать как вырожденный интервал с длиной, равной нулю. Тогда из определения (4.3.2) следует, что *мера счетного множества равна нулю*. Отсюда мы получаем простое доказательство свойства, приведенного выше (см. параграфы 1.1 и 2.1) без доказательства: *множество всех точек, принадлежащих невырожденному интервалу, несчетно*. Действительно, мера этого множества равна длине интервала, т. е. есть положительная величина, в то время как мера счетного множества равна нулю. То же свойство имеет место и для невырожденных интервалов в R_n (при $n > 1$).

Наконец, мы докажем следующую теорему, которая будет необходима впоследствии для распространения определения меры на более общие классы множеств:

Если множества I и J из \mathfrak{F} имеют конечную меру, то

$$(4.3.8) \quad L(I + J) = L(I) + L(J) - L(IJ).$$

Рассмотрим сперва случай, когда I и J являются суммами конечного числа интервалов. Из соотношений

$$\begin{aligned} I + J &= I + (J - IJ), \\ J &= IJ + (J - IJ) \end{aligned}$$

мы получим, принимая во внимание, что все множества принадлежат \mathfrak{F} и слагаемые в правых частях наших соотношений не пересекаются,

что

$$L(I+J) = L(I) + L(J-IJ), \\ L(J) = L(IJ) + L(J-IJ),$$

откуда и вытекает (4.3.8).

В общем случае, когда I и J суть суммы конечного или счетного числа интервалов, это рассуждение не подходит, так как мы не можем быть уверены в том, что множество $J - IJ$ входит в \mathfrak{F} (см. параграф 2.3), и, значит, в этом случае мера $L(J - IJ)$ может быть еще не определена. Возьмем представления I и J в виде (4.3.1);

$$I = \sum_{\mu=1}^{\infty} i_{\mu}, \quad J = \sum_{\nu=1}^{\infty} j_{\nu},$$

и положим

$$I_n = \sum_{\mu=1}^n i_{\mu}, \quad J_n = \sum_{\nu=1}^n j_{\nu}.$$

Как было показано выше,

$$L(I_n + J_n) = L(I_n) + L(J_n) - L(I_n J_n).$$

При $n \rightarrow \infty$ каждый член этого соотношения стремится к соответствующему члену соотношения (4.3.8), и тем самым это последнее доказано.

4.4. Внешняя и внутренняя меры ограниченного множества. В предыдущем параграфе мы определили меру $L(I)$ для всех множеств I , принадлежащих классу \mathfrak{F} . Для того чтобы распространить это определение на более общий класс множеств, мы введем теперь две вспомогательные функции — *внутреннюю* и *внешнюю* меры, которые будут определены для любого ограниченного множества из R_1 .

Во всем этом параграфе мы будем рассматривать только *ограниченные* множества. В качестве нашего пространства мы выбираем некоторый фиксированный конечный интервал (a, b) и рассматриваем только такие точки и множества, которые принадлежат (a, b) . Соответственно этому, под дополнением S^* к множеству S мы будем подразумевать дополнение по отношению к (a, b) (см. параграф 1.3).

Для того чтобы определить эти новые функции — внутреннюю и внешнюю меры, рассмотрим множество I , принадлежащее классу \mathfrak{F} , и такое, что $S \subset I \subset (a, b)$. Таким образом, мы *заключаем* множество S

в сумму I интервалов, содержащихся, в свою очередь, в (a, b) . Это всегда можно сделать, так как мы можем, например, выбрать $I = (a, b)$. Множество I имеет меру $L(I)$, определенную в предыдущем параграфе. Рассмотрим множество чисел $L(I)$, которые соответствуют *всем таким множествам I , содержащим S* . Очевидно, что это множество имеет конечную нижнюю границу, так как $L(I) \geq 0$.

Внешней мерой $\bar{L}(S)$ множества S называется нижняя граница чисел $L(I)$. Внутренняя мера $\underline{L}(S)$ множества S определяется соотношением $\underline{L}(S) = (b - a) - \bar{L}(S^)$.*

Так как мы рассматриваем здесь множества S , содержащиеся в интервале (a, b) , который сам является множеством из \mathfrak{X} , то, очевидно,

$$0 \leq \bar{L}(S) \leq b - a, \quad 0 \leq \underline{L}(S) \leq b - a.$$

Далее, непосредственно из определений получаем, что $\bar{L}(S)$ и $\underline{L}(S)$ суть *монотонные* функции от S , т. е. при $S_1 \subset S_2$

$$(4.4.1) \quad \bar{L}(S_1) \leq \bar{L}(S_2), \quad \underline{L}(S_1) \leq \underline{L}(S_2).$$

Действительно, для любого I , такого, что $S_2 \subset I$, имеем $S_1 \subset I$, откуда первое неравенство вытекает непосредственно. Второе неравенство получается из первого переходом к дополнительным множествам.

Далее, если $S \subset I_1$ и $S^* \subset I_2$, то каждая точка интервала (a, b) принадлежит, по крайней мере, одному из множеств I_1 и I_2 . Так как оба множества I_1 и I_2 содержатся в (a, b) , то мы имеем $I_1 + I_2 = (a, b)$ и, следовательно, согласно (4.3.7),

$$L(I_1) + L(I_2) \geq b - a.$$

Выбирая всеми возможными способами множества I_1 и I_2 , содержащие, соответственно, S и S^* , получим, что соответствующее неравенство должно быть выполнено и для нижних границ чисел $L(I_1)$ и $L(I_2)$, так что

$$\bar{L}(S) + \bar{L}(S^*) \geq b - a,$$

т. е.

$$(4.4.2) \quad \underline{L}(S) \leq \bar{L}(S).$$

Пусть S_1, S_2, \dots есть данная последовательность множеств, *пересекающихся или непересекающихся*. Согласно определению внешней

меры, для любого n можно найти такое I_n , что $S_n \subset I_n$ и

$$L(I_n) < \bar{L}(S_n) + \frac{\epsilon}{2^n},$$

где ϵ произвольно мало. Тогда имеем $S_1 + S_2 + \dots \subset I_1 + I_2 + \dots$ и из (4.3.7) получаем

$$\begin{aligned} \bar{L}(S_1 + S_2 + \dots) &\leq L(I_1 + I_2 + \dots) \leq \\ &\leq L(I_1) + L(I_2) + \dots < \\ &< \bar{L}(S_1) + \bar{L}(S_2) + \dots + \epsilon \left(\frac{1}{2} + \frac{1}{4} + \dots \right). \end{aligned}$$

Так как ϵ произвольно, то

$$(4.4.3) \quad \bar{L}(S_1 + S_2 + \dots) \leq \bar{L}(S_1) + \bar{L}(S_2) + \dots.$$

Для того чтобы вывести соответствующее неравенство для внутренней меры $\underline{L}(S)$, рассмотрим два непересекающихся множества S_1 и S_2 . Пусть дополнительные множества S_1^* и S_2^* содержатся, соответственно, в I_1 и I_2 . Тогда

$$(4.4.4) \quad \begin{aligned} (b-a) - \underline{L}(S_1) &= \bar{L}(S_1^*) = \inf L(I_1), \\ (b-a) - \underline{L}(S_2) &= \bar{L}(S_2^*) = \inf L(I_2), \end{aligned}$$

где множества I_1 и I_2 , содержащие S_1^* и S_2^* , выбираются всевозможными способами. Далее, по (1.3.1) имеем

$$(S_1 + S_2)^* = S_1^* S_2^* \subset I_1 I_2,$$

но отсюда мы можем только заключить, что

$$(4.4.5) \quad (b-a) - \underline{L}(S_1 + S_2) = \bar{L}[(S_1 + S_2)^*] \leq \inf L(I_1 I_2),$$

так как, помимо произведения $I_1 I_2$, могут быть и другие множества I , содержащие $(S_1 + S_2)^*$. Из (4.4.4) и (4.4.5) выводим, используя (4.3.8), что

$$\begin{aligned} \underline{L}(S_1 + S_2) - \underline{L}(S_1) - \underline{L}(S_2) &\geq \inf [L(I_1) + L(I_2)] - \inf L(I_1 I_2) - \\ &- (b-a) \geq \\ &\geq \inf [L(I_1) + L(I_2) - L(I_1 I_2)] - \\ &- (b-a) = \\ &= \inf L(I_1 + I_2) - (b-a). \end{aligned}$$

Так как S_1 и S_2 не пересекаются, то $I_1 + I_2 \supset S_1 + S_2 = (S_1 S_2)^* = = (a, b)$. С другой стороны, I_1 и I_2 содержатся в (a, b) , так что $I_1 + I_2 \subset (a, b)$. Таким образом, $I_1 + I_2 = (a, b)$ и

$$\underline{L}(S_1 + S_2) \geq \underline{L}(S_1) + \underline{L}(S_2).$$

Пусть теперь S_1, S_2, \dots есть последовательность попарно непересекающихся множеств. Повторным применением последнего неравенства получим

$$(4.4.6) \quad \underline{L}(S_1 + S_2 + \dots) \geq \underline{L}(S_1) + \underline{L}(S_2) + \dots$$

В частном случае, когда S есть интервал, из определений легко следует, что и $\bar{L}(S)$ и $\underline{L}(S)$ равны длине интервала. Если $I = \sum i_i$ есть множество из \mathfrak{J} , где i_i — непересекающиеся интервалы, то из (4.4.3) и (4.4.6) получим

$$\bar{L}(I) \leq \sum L(i_i), \quad \underline{L}(I) \geq \sum L(i_i)$$

и, следовательно, согласно (4.4.2) и (4.3.2),

$$(4.4.7) \quad \bar{L}(I) = \underline{L}(I) = L(I).$$

Наконец, заметим, что внешняя и внутренняя меры не зависят от интервала (a, b) , в котором, по предположению, содержатся наши множества. Согласно параграфу 2.2, ограниченное множество S всегда содержитя в замкнутом интервале (α, β) , где α и β суть нижняя и верхняя границы S . Если (a, b) есть некоторый интервал, содержащий S , то $a \leq \alpha$ и $b \geq \beta$. Простое рассмотрение покажет, что два интервала (a, b) и (α, β) дадут одни и те же значения внешней и внутренней меры S . Таким образом, величины $\bar{L}(S)$ и $\underline{L}(S)$ зависят только от самого S и не зависят от интервала (a, b) .

4.5. Измеримые множества и лебегова мера. Ограничное множество S называется измеримым, если его внешняя и внутренняя меры равны. Их общее значение обозначается через $L(S)$ и называется лебеговой мерой или просто мерой S :

$$\bar{L}(S) = \underline{L}(S) = L(S).$$

Неограниченное множество S называется измеримым, если произведение $i_x S$, где i_x обозначает замкнутый интервал $(-x, x)$, изме-

римо при любом $x > 0$. Мера $L(S)$ определяется тогда соотношением

$$L(S) = \lim_{x \rightarrow \infty} L(i_x S).$$

Согласно (4.4.1), $L(i_x S)$ есть неубывающая функция от x . Таким образом, этот предел, конечный или бесконечный, всегда существует.

В частном случае, когда S есть множество из \mathfrak{F} , новое определение меры совпадает с определением (4.3.2). Для ограниченного множества I это следует непосредственно из (4.4.7). Для неограниченного множества I мы получим тот же самый результат, рассматривая ограничение множество $i_x I$ и переходя к пределу при $x \rightarrow \infty$.

Согласно (4.4.1), $\bar{L}(S)$ и $\underline{L}(S)$ суть монотонные функции множества S . Из данного нами определения следует, что то же имеет место и для $L(S)$. Для любых двух измеримых множеств S_1 и S_2 , таких, что $S_1 \subset S_2$, мы, таким образом, имеем

$$(4.5.1) \quad L(S_1) \leq L(S_2).$$

Покажем теперь, что мера $L(S)$ удовлетворяет условиям а), б), с) параграфа 4.2. По отношению к условиям а) и с) это следует непосредственно из сказанного выше; остается доказать, что выполнено условие б), что составляет содержание следующей теоремы:

Если попарно непересекающиеся множества S_1, S_2, \dots измеримы, то сумма $S_1 + S_2 + \dots$ также измерима и

$$(4.5.2) \quad L(S_1 + S_2 + \dots) = L(S_1) + L(S_2) + \dots$$

Рассмотрим сперва случай, когда все S_1, S_2, \dots содержатся в некотором конечном интервале (a, b) . Тогда, так как все S_n измеримы то соотношения (4.4.3) и (4.4.6) дают

$$\begin{aligned} \bar{L}(S_1 + S_2 + \dots) &\leq \bar{L}(S_1) + \bar{L}(S_2) + \dots = L(S_1) + L(S_2) + \dots, \\ \underline{L}(S_1 + S_2 + \dots) &\geq \underline{L}(S_1) + \underline{L}(S_2) + \dots = L(S_1) + L(S_2) + \dots. \end{aligned}$$

Согласно (4.4.2), $\underline{L}(S_1 + S_2 + \dots) \leq \bar{L}(S_1 + S_2 + \dots)$ и, следовательно,

$$\bar{L}(S_1 + S_2 + \dots) = \underline{L}(S_1 + S_2 + \dots) = L(S_1 + S_2 + \dots),$$

так что в этом случае наше утверждение доказано.

В общем случае мы рассматриваем произведения $i_x S_1, i_x S_2, \dots$, каждое из которых содержитя в конечном интервале i_x . Проведенное

выше доказательство показывает, что произведение $i_x(S_1 + S_2 + \dots)$ измеримо при любом x и

$$L[i_x(S_1 + S_2 + \dots)] = L(i_x S_1) + L(i_x S_2) + \dots$$

Итак, $S_1 + S_2 + \dots$ измеримо, а так как все члены ряда в правой части последнего равенства суть неубывающие функции от x , то

$$\begin{aligned} L(S_1 + S_2 + \dots) &= \lim_{x \rightarrow \infty} [L(i_x S_1) + L(i_x S_2) + \dots] = \\ &= L(S_1) + L(S_2) + \dots \end{aligned}$$

Таким образом, (4.5.2) доказано, и лебегова мера удовлетворяет всем трем условиям параграфа 4.2.

Множество S , такое, что $L(S) = 0$, называется **множеством меры нуль**. Если внешняя мера $\bar{L}(S) = 0$, то из определения меры следует, что множество S измеримо и имеет меру нуль. Мы уже видели в параграфе 4.3, что, в частности, каждое счетное множество имеет меру нуль. Из сказанного выше легко следуют следующие два предложения:

Любое подмножество множества меры нуль есть множество меры нуль.

Сумма последовательности множеств меры нуль есть множество меры нуль.

Эти предложения являются прямыми следствиями соотношений (4.4.1) и (4.4.3) для внешней меры.

4.6. Класс измеримых множеств. Рассмотрим класс \mathfrak{U} всех измеримых множеств в \mathbb{R}_1 . Мы собираемся доказать, что \mathfrak{U} есть аддитивный класс множеств (см. параграф 1.6). Так как мы уже видели в предыдущем параграфе, что \mathfrak{U} содержит все интервалы, то отсюда следует, что \mathfrak{U} содержит весь класс \mathfrak{B}_1 борелевских множеств, так что *все борелевские множества измеримы*.

Действительно, докажем, что класс \mathfrak{U} удовлетворяет условиям a_1 , b_1 и c_1 параграфа 1.6. Выполнение условия a_1 очевидно.

Рассмотрим условие c_1). Требуется доказать, что *дополнение S^* измеримого множества S само измеримо*. Рассмотрим сперва случай ограниченного множества S . Пусть S^* — дополнение множества S по отношению к некоторому конечному интервалу (a, b) , содержащему S . Так как S измеримо, то, согласно определению внутренней меры, имеем

$$\underline{L}(S^*) = (b - a) - \bar{L}(S) = (b - a) - \underline{L}(S) = \bar{L}(S^*),$$

так что S^* измеримо и имеет меру $(b - a) - L(S)$. В общем случае, когда S измеримо, но не обязательно ограничено, то же рассуждение показывает, что произведение $i_x S^*$, где S^* теперь означает дополнение S по отношению ко всему пространству R_1 , измеримо при любом $x > 0$. Тогда, по определению, S^* измеримо.

Рассмотрим теперь условие b_1). Нужно показать, что *сумма $S_1 + S_2 + \dots$ измеримых множеств S_1, S_2, \dots сама измерима*. В частном случае, когда $S_\mu S_\nu = 0$ (при $\mu \neq \nu$), это уже было доказано в предыдущем параграфе. Остается еще доказать общий случай.

Достаточно рассмотреть случай, когда все S_n содержатся в конечном интервале (a, b) . Действительно, если бы наше утверждение было уже доказано в этом случае, то, рассмотрев множества $i_x S_1, i_x S_2, \dots$ мы нашли бы, что их сумма $i_x (S_1 + S_2 + \dots)$ измерима при любом $x > 0$. Тогда, по определению, $S_1 + S_2 + \dots$ было бы измеримо.

Таким образом, мы должны доказать, что если измеримые множества S_1, S_2, \dots содержатся все в (a, b) , то сумма $S_1 + S_2 + \dots$ измерима.

Сперва докажем это для частного случая, когда мы имеем только два множества S_1 и S_2 . Пусть n — индекс, принимающий значения 1 и 2, и пусть дополнительные множества берутся по отношению к интервалу (a, b) . Так как S_n и S_n^* измеримы, то можно найти такие множества I_n и J_n из \mathfrak{X} , что

$$(4.6.1) \quad S_n \subset I_n \subset (a, b), \quad S_n^* \subset J_n \subset (a, b),$$

а разности $L(I_n) - L(S_n)$ и $L(J_n) - L(S_n^*)$ обе меньше данного $\epsilon > 0$. Теперь, согласно (4.6.1), каждая точка интервала (a, b) должна принадлежать, по крайней мере, одному из множеств I_n и J_n , так что $I_n + J_n = (a, b)$ и, следовательно, согласно (4.3.8),

$$(4.6.2) \quad \begin{aligned} L(I_n J_n) &= L(I_n) + L(J_n) - (b - a) = \\ &= L(I_n) + L(J_n) - L(S_n) - L(S_n^*) < 2\epsilon. \end{aligned}$$

Из (4.6.1) следует далее, что

$$\begin{aligned} S_1 + S_2 &\subset I_1 + I_2, \\ (S_1 + S_2)^* &= S_1^* S_2^* \subset J_1 J_2, \end{aligned}$$

и, следовательно,

$$(4.6.3) \quad \begin{aligned} \bar{L}(S_1 + S_2) &\leq L(I_1 + I_2), \\ \underline{L}(S_1 + S_2) &\geq (b - a) - L(J_1 J_2). \end{aligned}$$

Легко видеть, что $I_1 + I_2 - J_1 J_2 = (a, b)$. Тогда, применяя еще раз (4.3.8), получаем из соотношения (4.6.3)

$$\bar{L}(S_1 + S_2) - \underline{L}(S_1 + S_2) \leq L[(I_1 + I_2)J_1 J_2].$$

Теперь

$$(I_1 + I_2)J_1 J_2 = I_1 J_1 J_2 + I_2 J_1 J_2 \subset I_1 J_1 + I_2 J_2,$$

так что мы получаем с помощью (4.5.1), (4.3.7) и (4.6.2)

$$\bar{L}(S_1 + S_2) - \underline{L}(S_1 + S_2) \leq L(I_1 J_1) + L(I_2 J_2) < 4\varepsilon.$$

Так как ε произвольно, а внутренняя мера никогда не превосходит внешнюю, то отсюда следует, что $\bar{L}(S_1 + S_2) = \underline{L}(S_1 + S_2)$, т. е. $S_1 + S_2$ измеримо.

Отсюда непосредственно следует, что сумма $S_1 + \dots + S_k$ любого *конечного* числа измеримых множеств, содержащихся в (a, b) , измерима. Соотношение $S_1 S_2 \dots S_n = (S_1^* + \dots + S_n^*)^*$ показывает, что то же верно и для произведения.

Рассмотрим, наконец, случай *бесконечной* суммы. С помощью преобразования, изложенного в параграфе 1.4, получим

$$S = S_1 + S_2 + \dots = Z_1 + Z_2 + \dots,$$

где $Z_\nu = S_1^* \dots S_{\nu-1}^* S_\nu$ и $Z_\mu Z_\nu = 0$ (при $\mu \neq \nu$). Так как $S_1^*, \dots, S_{\nu-1}^*$ и S_ν измеримы, то и Z_ν измеримо. Наконец, согласно (4.5.2), сумма $Z_1 + Z_2 + \dots$ измерима.

Итак, мы доказали, что *измеримые множества образуют аддитивный класс \mathfrak{L} . Следовательно, сумма, произведение и разность конечного или счетного числа измеримых множеств измеримы. В частности, борелевские множества измеримы.*

4.7. Измеримые множества и борелевские множества. Класс \mathfrak{L} измеримых множеств в действительности шире, чем класс \mathfrak{B}_1 борелевских множеств. В качестве иллюстрации различия между обоими классами мы упомянем без доказательства следующее предложение: *каждое измеримое множество есть сумма борелевского множества и множества меры нуль*. Все множества, встречающиеся обычно в приложениях математического анализа, суть борелевские множества, и мы будем поэтому, вообще говоря, ограничиваться рассмотрением класса \mathfrak{B}_1 и соответствующего класса \mathfrak{B}_n в n -мерных пространствах.

Докажем теперь утверждение, сформулированное в параграфе 4.2, что лебегова мера есть единственная функция, определенная для всех борелевских множеств и удовлетворяющая условиям а), б), с) параграфа 4.2.

В самом деле, пусть $\Lambda(S)$ есть некоторая функция множества, удовлетворяющая сформулированным условиям. Для каждого I из \mathfrak{X} мы должны, очевидно, иметь $\Lambda(I) = L(I)$, так как наше определение (4.3.2) меры $L(I)$ однозначно продиктовано условиями б) и с) параграфа 4.2. Пусть теперь S есть ограниченное борелевское множество. Заключим S в множество I из класса \mathfrak{X} . Тогда из условий а) и б) следует, что $\Lambda(S) \leq \Lambda(I) = L(I)$. Нижняя граница чисел $L(I)$ для всех I , содержащих S , равна $L(S)$ и, следовательно, $\Lambda(S) \leq L(S)$. Заменяя S его дополнением S^* по отношению к некоторому конечному интервалу, имеем $\Lambda(S^*) \leq L(S^*)$ и, следовательно, $\Lambda(S) \geq L(S)$. Таким образом, $\Lambda(S)$ и $L(S)$ совпадают для всех ограниченных борелевских множеств. Это же совпадение имеет место и для неограниченных борелевских множеств, так как неограниченное борелевское множество может быть, очевидно, представлено как сумма последовательности ограниченных борелевских множеств.

Наконец, докажем теорему относительно меры предела (см. параграф 1.5) монотонной последовательности борелевских множеств. Согласно параграфу 2.3, каждый такой предел есть борелевское множество.

Если S_1, S_2, \dots есть неубывающая последовательность борелевских множеств, то

$$(4.7.1) \quad \lim L(S_n) = L(\lim S_n).$$

То же равенство выполняется и для невозрастающей последовательности борелевских множеств, если только $L(S_1)$ конечно.

Действительно, для неубывающей последовательности мы имеем

$$\lim S_n = S_1 + (S_2 - S_1) + (S_3 - S_2) + \dots$$

и, в силу (4.5.2),

$$\begin{aligned} L(\lim S_n) &= L(S_1) + L(S_2 - S_1) + \dots = \\ &= \lim [L(S_1) + L(S_2 - S_1) + \dots + L(S_n - S_{n-1})] = \lim L(S_n). \end{aligned}$$

Для невозрастающей последовательности, такой, что $L(S_1)$ конечно, требуемое соотношение доказывается переходом к дополнительным множествам S^* по отношению к S_1 . Пример $S_n = (n, +\infty)$ показывает, что требование конечности $L(S_1)$ существенно.

ГЛАВА 5

ИНТЕГРАЛ ЛЕБЕГА ОТ ФУНКЦИЙ ОДНОЙ ПЕРЕМЕННОЙ

5.1. Интеграл от ограниченной функции по множеству конечной меры. Все точечные множества, рассматриваемые далее в этой книге, суть борелевские множества, если специально не оговорено противное*).

Пусть S есть данное множество конечной меры $L(S)$, а $g(x)$ — функция действительного переменного x , определенная для всех значений x , принадлежащих к S . Мы будем предполагать, что $g(x)$ ограничена на S , т. е. что нижняя и верхняя границы значений, принимаемых функцией $g(x)$ на множестве S , конечны. Обозначим эти границы соответственно через m и M , так что $m \leq g(x) \leq M$ для всех x , принадлежащих S . Разобьем S на конечное число попарно не пересекающихся частей S_1, \dots, S_n так, что

$$S = S_1 + S_2 + \dots + S_n \quad (S_\mu S_\nu = 0 \text{ при } \mu \neq \nu).$$

На множестве S_ν функция $g(x)$ имеет нижнюю границу m_ν и верхнюю границу M_ν ; при этом $m \leq m_\nu \leq M_\nu \leq M$.

Теперь определим связанные с этим разбиением *нижнюю и верхнюю суммы Дарбу* соотношениями

$$(5.1.1) \quad z = \sum_1^n m_\nu L(S_\nu), \quad Z = \sum_1^n M_\nu L(S_\nu).$$

Очевидно, что

$$m L(S) \leq z \leq Z \leq M L(S).$$

Непосредственно видно также, что каждое *продолжение* нашего разбиения, т. е. такое разбиение, которое получается при дальнейшем подразбиении частей S_ν , даст нижнюю сумму не меньшую, а верхнюю сумму не большую, чем соответствующие суммы при первоначальном разбиении.

Каждое разбиение множества S на произвольное конечное число попарно не пересекающихся частей дает, согласно (5.1.1), нижнюю сумму z и верхнюю сумму Z . Рассмотрим множество всех *нижних сумм* z и множество всех *верхних сумм* Z , получающихся при всевоз-

*). Для того чтобы дать полное изложение теории интеграла Лебега, необходимо рассматривать *измеримые* множества, а не только борелевские множества. Однако, как указано в параграфе 4.7, ограничение борелевскими множествами вполне достаточно для наших целей.

можных разбиениях множества S . Будем их кратко называть z -множеством и Z -множеством. Оба эти множества ограничены, так как все z и Z заключены между $mL(S)$ и $ML(S)$. Покажем, что *верхняя граница z -множества не превосходит нижней границы Z -множества*. Таким образом, эти два множества имеют, самое большое, одну общую точку, и, не считая этой точки, все элементы z -множества лежат левее всех элементов Z -множества.

Для того чтобы доказать это утверждение, предположим, что z' есть произвольная нижняя сумма, соответствующая разбиению $S = S'_1 + \dots + S'_{n'}$, а Z'' есть произвольная верхняя сумма, соответствующая разбиению $S = S''_1 + \dots + S''_{n''}$. Очевидно, достаточно показать, что $z' \leq Z''$. Это, однако, непосредственно следует из рассмотрения разбиения $S = \sum_{i=1}^{n'} \sum_{k=1}^{n''} S'_i S''_k$, являющегося продолжением каждого из данных разбиений. Если соответствующие этому новому разбиению суммы Дарбу суть z_0 и Z_0 , то мы, согласно сделанному выше замечанию имеем $z' \leq z_0 \leq Z_0 \leq Z''$, и тем самым наше утверждение доказано.

Верхняя граница z -множества (будем обозначать ее $\sup z$) называется *нижним интегралом* функции $g(x)$ на множестве S , а нижняя граница Z -множества (обозначаем ее $\inf Z$) называется *верхним интегралом* функции $g(x)$ на множестве S . Мы пишем

$$(5.1.2) \quad \underline{\int_S} g(x) dx = \sup z, \quad \overline{\int_S} g(x) dx = \inf Z.$$

Из сказанного выше следует, что

$$(5.1.3) \quad mL(S) \leq \underline{\int_S} g(x) dx \leq \overline{\int_S} g(x) dx \leq ML(S).$$

Если верхний и нижний интегралы равны (т. е. если $\sup z = \inf Z$), то говорят, что $g(x)$ *интегрируема в смысле Лебега* или, коротко, *интегрируема* на S . Общее значение обоих интегралов называется *интегралом Лебега* функции $g(x)$ на множестве S , и мы пишем .

$$\int_S g(x) dx = \overline{\int_S} g(x) dx = \underline{\int_S} g(x) dx.$$

Для того чтобы функция $g(x)$ была интегрируема на множестве S , необходимо и достаточно, чтобы для каждого $\epsilon > 0$ существовало

такое разбиение S , что соответствующая разность $Z - z$ меньше ϵ . Действительно, если это условие выполнено, то из наших определений верхнего и нижнего интегралов следует, что разность между ними меньше ϵ , и так как ϵ произвольно, то эти интегралы совпадают. Обратно, если известно, что $g(x)$ интегрируема, то отсюда непосредственно следует, что должны существовать такая нижняя сумма z' и такая верхняя сумма Z'' , что $Z'' - z' < \epsilon$. Разбиение, которое служит продолжением разбиений, соответствующих суммам z' и Z'' , дает нижнюю сумму z_0 и верхнюю сумму Z_0 , такие, что $Z_0 - z_0 < \epsilon$.

Мы видим, что приведенное рассуждение совершенно аналогично обычному определению интеграла Римана. Для интеграла Римана множество S есть интервал, разделенный на конечное число подинтервалов S_ν , а суммы Дарбу образуются согласно (5.1.1), где $L(S_\nu)$ обозначает длину ν -го интервала S_ν . Единственное отличие заключается в том, что в разобранном нами случае мы рассматриваем более общий класс множеств, чем интервалы, так как S и S_ν могут быть любыми борелевскими множествами. В то же время мы заменяем длину интервала S , ее естественным обобщением — мерой множества S_ν .

В частном случае, когда S есть конечный интервал (a, b) , каждое разбиение (a, b) на подинтервалы, рассматриваемое при определении интеграла Римана, есть частный случай разбиений на борелевские множества, фигурирующие в определении интеграла Лебега. Однако мы рассматриваем также разбиения интервала (a, b) на борелевские множества, не являющиеся интервалами. Эти более общие разбиения могут увеличить значение верхней границы z -множества и уменьшить значение нижней границы Z -множества. Мы видим, таким образом, что верхний и нижний интегралы, определенные соотношением (5.1.2), заключены между соответствующими римановыми интегралами. Если $g(x)$ есть функция, интегрируемая в смысле Римана, то последние равны и, следовательно, равны и интегралы (5.1.2), т. е. $g(x)$ интегрируем также и в смысле Лебега, причем значения интегралов Римана и Лебега совпадают.

Если мы рассматриваем функции, интегрируемые в смысле Римана, и их интегралы, распространенные на интервалы, то нет надобности различать оба вида интегралов.

Определение интеграла Лебега, конечно, несколько сложнее, чем определение интеграла Римана. Введение этого усложнения оправдывается тем, что свойства интеграла Лебега проще, чем свойства интеграла Римана. Для

того чтобы показать, что интеграл Лебега существует для более общего класса функций, чем класс функций, интегрируемых по Риману, рассмотрим, например, функцию $g(x)$, равную нулю, если x иррационально, и равную 1, если x рационально. В каждом невырожденном интервале эта функция имеет нижнюю границу, равную нулю, и верхнюю границу, равную единице. Поэтому нижняя и верхняя суммы Дарбу в определении риманова интеграла функции $g(x)$ на интервале $(0, 1)$ при любом разбиении $(0, 1)$ на подинтервалы равны соответственно 0 и 1, так что интеграл Римана не существует. Если с другой стороны, мы разобьем интервал $(0, 1)$ на две части S_l и S_r , содержащие соответственно все иррациональные и все рациональные точки интервала, то $g(x)$ равна нулю на множестве S_l и равна единице на множестве S_r . Далее, S_l имеет меру 1, а S_r — меру 0, так что обе суммы Дарбу (5.1.1), соответствующие этому разбиению, равны 0. Тогда верхний и нижний интегралы (5.1.2) также оба равны 0 и, следовательно, интеграл Лебега функции $g(x)$ на интервале $(0, 1)$ существует и равен 0.

Интеграл Лебега на интервале (a, b) обычно обозначается так же, как и интеграл Римана:

$$\int_a^b g(x) dx.$$

Мы увидим ниже (см. параграф 5. 3), что этот интеграл имеет одно и то же значение, независимо от того, рассматриваем ли мы интервал (a, b) как замкнутый, открытый или полуоткрытый. В частном случае, когда $g(x)$ непрерывна для $a \leq x \leq b$, интеграл

$$G(x) = \int_a^x g(t) dt$$

существует (как в смысле Римана, так и в смысле Лебега) и мы имеем

$$(5.1.4) \quad G'(x) = g(x)$$

для всех x из интервала (a, b) .

5.2. B -измеримые функции. Мы говорим, что функция $g(x)$, определенная на множестве S , измерима в смысле Бореля или B -измерима на множестве S , если множество всех точек x из S , для которых $g(x) \leq k$, есть борелевское множество при любом действительном значении k . Мы докажем следующую важную теорему:

Если $g(x)$ ограничена и B -измерима на множестве S конечной меры, то она интегрируема на S .

Предположим, что $m < g(x) \leq M$ для всех x , принадлежащих S . Пусть дано $\varepsilon > 0$. Разобьем интервал (m, M) на подинтервалы точками y_v , такими, что

$$m = y_0 < y_1 < \dots < y_{n-1} < y_n = M,$$

и при этом длина каждого подинтервала меньше ε . Очевидно, что это всегда можно сделать, взяв n достаточно большим. Теперь пусть S_v обозначает множество всех точек, принадлежащих S , и таких, что

$$y_{v-1} < g(x) \leq y_v \quad (v = 1, 2, \dots, n).$$

Тогда $S = S_1 + \dots + S_n$ и $S_\mu S_\nu = 0$ при $\mu \neq \nu$. Далее, S_v есть разность между двумя борелевскими множествами, определенными соответственно неравенствами $g(x) \leq y_v$ и $g(x) \leq y_{v-1}$, и, следовательно, S_v есть борелевское множество. Разность $M_v - m_v$ между верхней и нижней границами функции $g(x)$ на множестве S_v не превосходит $y_v - y_{v-1} < \varepsilon$. Отсюда мы получаем для сумм Дарбу, соответствующих данному разбиению множества S ,

$$Z - z = \sum_1^n (M_v - m_v) L(S_v) < \varepsilon \sum_1^n L(S_v) = \varepsilon L(S).$$

Но ε произвольно мало, и, следовательно, согласно предыдущему параграфу, $g(x)$ интегрируема на S .

Важность доказанной теоремы следует из того, что все функции, обычно встречающиеся в приложениях математического анализа, B -измеримы. Поэтому в дальнейшем мы будем рассматривать только B -измеримые функции. Как и в случае борелевских множеств, предположение B -измеримости рассматриваемых функций специально оговаривается не будет.

Здесь мы остановимся лишь кратко на обосновании этого утверждения, отсылая за дальнейшими деталями к специальным работам, например, Валле-Пуссена [40]. Рассмотрим сперва случай, когда S есть конечный или бесконечный интервал (a, b) , и будем говорить просто о „ E -измеримости“ вместо „ B -измеримости на интервале (a, b) “. Если g_1 и g_2 — B -измеримые функции, то сумма $g_1 + g_2$, разность $g_1 - g_2$ и произведение $g_1 g_2$ также E -измеримы. Докажем это для суммы, другие случаи доказываются аналогично. Пусть дано k и пусть U обозначает множество всех x из (a, b) , для которых $g_1 + g_2 \leq k$, а U'_r и U''_r обозначают соответственно множества, определенные неравенствами $g_1 \leq r$ и $g_2 \leq k - r$. Тогда, по предположению, U'_r и U''_r суть борелевские множества при любых k и r , и можно проверить, что $U = \prod (U'_r + U''_r)$, где r пробегает счетную последовательность положительных и отрицатель-

ных рациональных чисел. Так как, согласно параграфу 2.3, U есть борелевское множество при любом значении k , то, следовательно, $g_1 + g_2$ есть B -измеримая функция. Доказательство непосредственно переносится на случай суммы или произведения любого конечного числа B -измеримых функций.

Рассмотрим теперь бесконечную сумму $g = g_1 + g_2 + \dots$ B -измеримых функций, которую мы предполагаем сходящейся для всех x в интервале (a, b) . Пусть $\varepsilon_1, \varepsilon_2, \dots$ есть убывающая последовательность положительных чисел, стремящаяся к нулю, и пусть Q_{mn} обозначает множество всех точек x из (a, b) , для которых $g_1 + \dots + g_m \leq k + \varepsilon_n$. Тогда Q_{mn} есть борелевское множество и, если мы положим

$$R_{mn} = Q_{mn} Q_{m+1, n}, \dots, U_n = R_{1n} + R_{2n} + \dots, U = U_1 U_2 \dots,$$

то, как в этом нетрудно убедиться, U есть множество точек x из (a, b) , для которых $g(x) \leq k$. Так как при определении U использованы только суммы и произведения борелевских множеств, то U есть борелевское множество, и следовательно, функция $g(x)$ B -измерима. Далее, если g есть предел сходящейся последовательности g_1, g_2, \dots B -измеримых функций, то $g = g_1 + \dots + (g_2 - g_1) + (g_3 - g_2) + \dots$, и, следовательно, функция g B -измерима. Теперь очевидно, что функция $g(x) = cx^n$ B -измерима при любой постоянной c и при любом натуральном n . Следовательно, всякий полином есть B -измеримая функция. Всякая непрерывная функция есть предел некоторой последовательности полиномов и, следовательно, представляет собой B -измеримую функцию. Аналогично, любая функция, полученная посредством предельных переходов из непрерывных функций, B -измерима.

При доказательстве этих предложений предполагалось, что S есть интервал. Пусть функция $g(x)$ B -измерима на (a, b) , а S есть борелевское множество в интервале (a, b) . Тогда функция $e(x)$, равная 1 на S и 0 на S^c , очевидно, B -измерима на (a, b) , и, следовательно, произведение $e(x)g(x)$ есть функция, B -измеримая на (a, b) ; отсюда вытекает, что $g(x)$ B -измерима на S . Если в частности, S есть множество всех x из интервала (a, b) , для которых $g(x) \leq 0$, то $|g(x)| = g(x) - 2e(x)g(x)$. Таким образом, модуль B -измеримой функции есть B -измеримая функция.

Таким образом, когда мы имеем дело с B -измеримыми функциями, то обычные операции над ними и предельные переходы приводят снова к B -измеримым функциям. По теореме, доказанной выше, всякая ограниченная функция, полученная этим путем, интегрируема по Лебегу на любом множестве конечной меры. Для интеграла Римана соответствующее предложение *неверно**); это — одно из свойств, которое делает интеграл Лебега понятием в некотором смысле более простым, чем интеграл Римана.

*). Даже если предел $g(x)$ последовательности функций, интегрируемых по Риману, ограничен на интервале (a, b) , то и тогда нельзя утверждать, что интеграл Римана функции $g(x)$ на интервале (a, b) существует. Рассмотрим, например, последовательность функций g_1, g_2, \dots , где g_n равно 1 для всех ра-

Сделаем, наконец, еще одно замечание, которое будет использовано в дальнейшем (см. параграф 14.5). Пусть $g(x)$ есть B -измеримая функция на множестве S . Уравнение $y = g(x)$ определяет соответствие между переменными x и y . Обозначим через Y некоторое данное множество на оси y , а через X — множество всех x из S , таких, что $y = g(x) \subset Y$. Будем тогда говорить, что множество X соответствует множеству Y . Очевидно, что если Y есть сумма, произведение или разность некоторых множеств Y_1, Y_2, \dots , то X есть сумма, произведение или разность соответствующих множеств X_1, X_2, \dots . Далее, если Y есть замкнутый бесконечный интервал $(-\infty, k)$, то, как мы знаем, X есть борелевское множество. Каждое борелевское множество может быть получено из таких интервалов сложением, умножением и вычитанием. Следовательно, **множество X , соответствующее борелевскому множеству Y , есть борелевское множество.**

5.3. Свойства интеграла. В этом параграфе мы будем рассматривать только ограниченные функции и множества конечной меры. Следующие свойства (5.3.1) — (5.3.4) совершенно аналогичны соответствующим свойствам интеграла Римана и доказываются таким же образом с помощью определений, данных в параграфе 5.1:

$$(5.3.1) \quad \int_S [g_1(x) + g_2(x)] dx = \int_S g_1(x) dx + \int_S g_2(x) dx,$$

$$(5.3.2) \quad \int_S cg(x) dx = c \int_S g(x) dx,$$

$$(5.3.3) \quad mI(S) \leq \int_S g(x) dx \leq M I(S),$$

$$(5.3.4) \quad \int_{S_1 + S_2} g(x) dx = \int_{S_1} g(x) dx + \int_{S_2} g(x) dx,$$

где c — постоянная, m и M обозначают нижнюю и верхнюю границы функции $g(x)$ на S , а множества S_1 и S_2 не имеют общих точек. Свойства (5.3.1) и (5.3.4) непосредственно распространяются на случай

циональных x со знаменателем, меньшим n , и равно 0 для всех остальных точек. Очевидно, что g_n интегрируема по Риману на интервале $(0, 1)$, но предел функции g_n при $n \rightarrow \infty$ есть функция $g(x)$, равная 0 при иррациональном x и 1 при рациональном x , а мы видели уже в предыдущем параграфе, что эта функция не интегрируема в смысле Римана на интервале $(0, 1)$.

произвольного конечного числа слагаемых. Если мы рассмотрим неотрицательную функцию $|g(x)| \pm g(x)$, то из (5.3.3) следует, что

$$(5.3.5) \quad \left| \int_S g(x) dx \right| \leq \int_S |g(x)| dx.$$

В частном случае, когда $g(x)$ тождественно равно 1, неравенства (5.3.3) дают

$$\int_S dx = L(S).$$

Далее, из (5.3.3) следует, что интеграл любой ограниченной функции $g(x)$ на множестве меры 0 равен нулю. В силу (5.3.4) отсюда следует, что, если функции $g_1(x)$ и $g_2(x)$ совпадают для всех x из S , за исключением, может быть, подмножества меры нуль, то

$$\int_S g_1(x) dx = \int_S g_2(x) dx.$$

Следовательно, если произвольным образом изменить значения подинтегральной функции на подмножестве меры нуль, то это не изменит значения интеграла. Мы можем даже допустить, чтобы функция была совсем не определена на подмножестве меры нуль. Мы видим также, что если множества S_1 и S_2 отличаются друг от друга лишь на множестве меры нуль, то интегралы любой ограниченной функции $g(x)$ на множествах S_1 и S_2 совпадают. Отсюда следует, в частности, правильность сделанного в параграфе 5.1 утверждения, что интеграл на интервале имеет одно и то же значение, независимо от того, является ли он замкнутым, открытым или полуоткрытым.

Из сказанного выше следует, что в теории интеграла Лебега часто можно пренебрегать множествами меры нуль. Если некоторое условие выполнено для всех x , принадлежащих рассматриваемому множеству S , за исключением, быть может, некоторых значений x , образующих подмножество меры 0, то мы будем говорить, что это условие выполнено *почти всюду на S или для почти всех значений x из S* .

Докажем теперь важную теорему, принадлежащую Лебегу, относительно интеграла функции, являющейся пределом сходящейся последовательности функций. Будем говорить, что последовательность $g_1(x), g_2(x), \dots$ *равномерно ограничена* на множестве S , если существует постоянная K , такая, что $|g_n(x)| < K$ для всех y и для всех x из S .

Если последовательность $\{g_v(x)\}$ равномерно ограничена на S и если предел $\lim_{v \rightarrow \infty} g_v(x) = g(x)$ существует почти всюду на S , то

$$(5.3.6) \quad \lim_{v \rightarrow \infty} \int_S g_v(x) dx = \int_S g(x) dx.$$

Если $\lim g_v(x)$ существует не для всех x из S , то мы дополняем определение функции $g(x)$ на множестве S , полагая $g(x)=0$ в тех точках x , для которых этот предел не существует. Тогда $|g(x)| \leq K$ для всех x из S и из предыдущего параграфа вытекает, что $g(x)$ B -измерима на S , а следовательно, и интегрируема на S . Пусть теперь дано $\epsilon > 0$. Рассмотрим множество S_n всех точек x из S , в которых $|g_v(x) - g(x)| \leq \epsilon$ при $v = n, n+1, \dots$. Тогда S_n суть борелевские множества, S_1, S_2, \dots образуют неубывающую последовательность, а предельное множество $\lim S_n$ (см. параграф 1.5) содержит каждую точку x из S , для которой $\lim g_v(x)$ существует. Таким образом, согласно предположению, $\lim S_n$ имеет ту же меру, что и S , и, согласно (4.7.1), имеем

$$\lim L(S_n) = L(\lim S_n) = L(S).$$

Итак, мы можем выбрать n такое, что $L(S_n) > L(S) - \epsilon$ или $L(S - S_n) < \epsilon$, и тогда получим для всех $v \geq n$

$$\int_S |g_v(x) - g(x)| dx = \int_{S_n} + \int_{S - S_n} < \epsilon [L(S) + 2K].$$

Так как ϵ произвольно и

$$\left| \int_S g_v(x) dx - \int_S g(x) dx \right| \leq \int_S |g_v(x) - g(x)| dx,$$

то теорема тем самым доказана.

Теорема (5.3.6) может быть сформулирована в другой форме, как теорема о почлененном интегрировании ряда:

Если ряд $\sum_1^{\infty} f_v(x)$ сходится почти всюду на S и если частные суммы $\sum_1^n f_v(x)$ равномерно ограничены на S , то

$$(5.3.7) \quad \int_S \left(\sum_1^{\infty} f_v(x) \right) dx = \sum_1^{\infty} \int_S f_v(x) dx.$$

В этой форме теорема является обобщением равенства (5.3.1) на бесконечное число слагаемых. Покажем теперь, что соответствующее обобщение соотношения (5.3.4) может быть получено как следствие из (5.3.7).

Если $S = S_1 + S_2 + \dots$, где $S_\mu S_\nu = 0$ для $\mu \neq \nu$, то

$$(5.3.8) \quad \int_S g(x) dx = \sum_1^\infty \int_{S_\nu} g(x) dx.$$

Пусть $e_\nu(x)$ есть функция, равная 1 для всех x из S_ν и равная 0 во всех остальных точках. Для каждого x , принадлежащего множеству S , имеем тогда

$$g(x) = \sum_1^\infty e_\nu(x) g(x).$$

Очевидно, что частные суммы этого ряда равномерно ограничены на S . Тогда, в силу (5.3.7),

$$\int_S g(x) dx = \sum_1^\infty \int_S e_\nu(x) g(x) dx = \sum_1^\infty \int_{S_\nu} g(x) dx.$$

В частном случае, когда $g(x) = 1$, соотношение (5.3.8) сводится к аддитивности лебеговой меры (см. (4.5.2)).

5.4. Интеграл от неограниченной функции на множестве конечной меры. В параграфах 5.1 и 5.2 мы видели, что интеграл Лебега

$$\int_S g(x) dx$$

имеет определенное значение при двух предположениях: 1) $g(x)$ ограничена на S и 2) S имеет конечную меру. Попытаемся теперь устранить эти ограничения. В этом параграфе мы будем рассматривать случай, когда множество S все еще имеет конечную меру, но $g(x)$ уже не обязательно ограничена на S .

Пусть a и b — какие-нибудь числа, такие, что $a < b$. Положим

$$g_{a,b}(x) = \begin{cases} a, & \text{если } g(x) < a, \\ g(x), & \text{если } a \leq g(x) \leq b, \\ b, & \text{если } g(x) > b. \end{cases}$$

Очевидно, что $g_{a,b}(x)$ ограничена, B -измерима на S и, следовательно, интегрируема на S . Если предел

$$(5.4.1) \quad \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_S g_{a,b}(x) dx = \int_S g(x) dx$$

существует и имеет конечное значение, то мы будем говорить, что $g(x)$ интегрируема на множестве S , а сам этот предел называется интегралом Лебега функции $g(x)$ на множестве S .

Непосредственно из определения следует, что любая функция интегрируема на множестве меры нуль и что значение этого интеграла равно нулю, как и в случае ограниченной функции.

В определении (5.4.1) мы можем предположить, что $a < 0$, $b > 0$, и тогда

$$g_{a,b}(x) = g_{a,0} + g_{0,b}, \\ |g(x)|_{a,b} = |g|_{a,b} = -g_{-a,0} + g_{0,b}.$$

Для фиксированного x $g_{a,0}(x)$ и $g_{0,b}(x)$ суть неубывающие функции от a и b . Следовательно, обе функции $g(x)$ и $|g(x)|$ интегрируемы тогда и только тогда, когда пределы

$$(5.4.2) \quad \lim_{a \rightarrow -\infty} \int_S g_{a,0}(x) dx \text{ и } \lim_{b \rightarrow +\infty} \int_S g_{0,b}(x) dx$$

оба конечны. Отсюда следует, что интегрируемость функции $g(x)$ эквивалентна интегрируемости функции $|g(x)|$. Далее, отсюда вытекает, что если $g(x)$ интегрируема на S , то она интегрируема и на любом его подмножестве*).

Если для всех x из S имеем $|g(x)| < G(x)$, где $G(x)$ интегрируема на S , то $|g|_{a,b} \leq G_{a,b}$, так что $|g(x)|$, а значит и $g(x)$, интегрируемы на S .

Теперь мы непосредственно обнаружим, что свойства (5.3.2) — (5.3.5) интеграла выполнены для любой интегрируемой $g(x)$. Что касается (5.3.3), то нужно, конечно, заметить, что одна из границ m и M или даже обе могут быть бесконечны.

Перейдем к обобщению (5.3.1), которое несколько сложнее. Предположим, что функции $f(x)$ и $g(x)$ интегрируемы на S . Из того, что

$$|f+g|_{a,0} = 0, \quad |f+g|_{0,b} \leq |f|_{0,b} + |g|_{0,b},$$

*) Напомним, что, согласно сделанному в начале этой главы замечанию рассматриваются только борелевские (и, следовательно, измеримые) подмножества. (Прим. перев.)

следует, что $f(x) + g(x)$ также интегрируема. Мы должны показать, что в этом случае выполняется (5.3.1), т. е.

$$(5.4.3) \quad \int_S (f + g) dx = \int_S f dx + \int_S g dx.$$

Предположим сперва, что f и g неотрицательны на S . Тогда

$$(f + g)_{a,0} = f_{a,0} = g_{a,0} = 0,$$

$$(f + g)_{0,b} \leq f_{0,b} + g_{0,b} \leq (f + g)_{0,2b},$$

и, следовательно,

$$\int_S (f + g)_{a,b} dx \leq \int_S f_{a,b} dx + \int_S g_{a,b} dx \leq \int_S (f + g)_{a,2b} dx.$$

Переходя к пределу при $a \rightarrow -\infty$, $b \rightarrow +\infty$, получим (5.4.3).

В общем случае множество S может быть разбито на шесть попарно не пересекающихся подмножеств, на каждом из которых ни одна из трех функций f , g и $f + g$ не меняет знака. Для каждого из этих подмножеств соотношение (5.4.3) доказывается проведенным выше рассуждением. Складывая результаты и применяя (5.3.4), получим (5.4.3).

Мы, таким образом, показали, что все свойства (5.3.1) — (5.3.5) интеграла выполнены и для нашего более общего случая. Для того чтобы обобщить также свойства, выражаемые соотношениями (5.3.6) — (5.3.8), мы докажем сперва следующую лемму:

Если $g(x)$ интегрируема на множестве S_0 и если задано $\varepsilon > 0$, то всегда существует такое $\delta > 0$, что

$$(5.4.4) \quad \left| \int_S g(x) dx \right| < \varepsilon$$

для каждого подмножества $S \subset S_0$, удовлетворяющего условию $L(S) < \delta$.

Так как нами уже было доказано выполнение условия (5.3.5), то достаточно доказать лемму в предположении, что функция $g(x)$ неотрицательна. В этом случае

$$\int_{S_0} g dx = \lim_{b \rightarrow \infty} \int_{S_0} g_{0,b} dx,$$

и, следовательно, можно найти такое b , что

$$0 \leq \int_{S_0} (g - g_{0,b}) dx < \frac{1}{2} \varepsilon.$$

Так как подинтегральное выражение неотрицательно, то из (5.3.4) и (5.3.3) следует, что для любого подмножества $S \subset S_0$ имеем

$$\int_S (g - g_{0,b}) dx < \frac{1}{2} \varepsilon,$$

или

$$\int_S g dx < \int_S g_{0,b} dx + \frac{1}{2} \varepsilon \leq b L(S) + \frac{1}{2} \varepsilon.$$

Выбирая $\delta = \frac{\varepsilon}{2b}$, получим утверждение леммы.

Следствием этой леммы является следующее утверждение: если $g(x)$ интегрируема на интервале (a, b) , то интеграл $\int_a^x g(t) dt$ есть непрерывная функция от x при $a < x < b$.

Теперь мы можем перейти к обобщению свойства (5.3.6). Предполагая, что $\lim_{v \rightarrow \infty} g_v(x) = g(x)$ почти всюду на S , мы покажем, что соотношение

$$(5.4.5) \quad \lim_{v \rightarrow \infty} \int_S g_v(x) dx = \int_S g(x) dx$$

выполняется, если последовательность $\{g_v(x)\}$ равномерно ограничена некоторой интегрируемой функцией, т. е. если $|g_v(x)| \leq G(x)$ для всех v и для всех x , принадлежащих S , где $G(x)$ интегрируема на множестве S . В частном случае, когда $G(x) = \text{const}$, это сводится к (5.3.6).

Доказательство вполне аналогично доказательству (5.3.6). Из сделанного предположения вытекает, что $|g(x)| \leq G(x)$ почти всюду на S , и, следовательно, g_v и g интегрируемы на S . Пусть задано $\varepsilon > 0$. Обозначим через S_n множество всех x , принадлежащих S , для которых $|g_v(x) - g(x)| \leq \varepsilon$ при всех $v \geq n$. Тогда S_1, S_2, \dots есть неубывающая последовательность и $L(S_n) \rightarrow L(S)$. Используя лемму (5.4.4), определим теперь ε' так, чтобы $\int_{S'} G(x) dx < \varepsilon$ для всех подмножеств $S' \subset S$ с мерой $L(S') < \delta$, и выберем n так, чтобы $L(S_n) > L(S) - \delta$ и, следовательно, $L(S - S_n) < \delta$. Тогда получим для всех $v \geq n$

$$\begin{aligned} \int_S |g_v(x) - g(x)| dx &= \int_{S_n} + \int_{S - S_n} < \varepsilon L(S) + 2 \int_{S - S_n} G(x) dx < \\ &< \varepsilon [L(S) + 2] \end{aligned}$$

и, таким образом, (5.4.5) доказано. Соответствующие обобщения для (5.3.7) и (5.3.8) получаются непосредственно.

5.5. Интеграл на множестве бесконечной меры. Устраним теперь также второе ограничение, упомянутое в начале параграфа 5.4, и рассмотрим интегралы Лебега на множествах бесконечной меры. Пусть S есть борелевское множество бесконечной меры. Обозначим через $S_{a,b}$ произведение (общую часть) множества S и замкнутого интервала (a,b) , где a и b конечны. Разумеется, $S_{a,b}$ есть множество конечной меры.

Если функция $g(x)$ интегрируема на множествах $S_{a,b}$ при всех a и b и если предел

$$\lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_{S_{a,b}} |g(x)| dx = \int_S |g(x)| dx$$

существует и имеет конечное значение, то мы будем говорить, что функция $g(x)$ интегрируема на множестве S *). Легко видеть, что в этом случае предел

$$(5.5.1) \quad \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_{S_{a,b}} g(x) dx = \int_S g(x) dx$$

существует и имеет конечное значение; мы будем поэтому говорить, что интеграл Лебега функции $g(x)$ на множестве S сходится*). Предел (5.5.1) и есть, по определению, значение этого интеграла. Если функция $g(x)$ интегрируема на S , то она также интегрируема и на любом подмножестве множества S .

Если $|g(x)| < G(x)$ для всех x , входящих в S , причем $G(x)$ интегрируема на множестве S , то, как легко видеть, $g(x)$ также интегрируема на S . Так как $|g_1 + g_2| \leq |g_1| + |g_2|$, то сумма двух интегрируемых функций есть интегрируемая функция.

Непосредственно из определения следует, что интегралы функций на множествах бесконечной меры обладают свойствами (5.3.1), (5.3.2) и (5.3.4). Вместо (5.3.3) мы будем иметь только неравенство

$$\int_S g(x) dx \geq 0, \text{ если } g(x) \geq 0$$

*) Строго говоря, мы должны были бы сказать, что функция $g(x)$ абсолютно интегрируема на множестве S и что интеграл функции $g(x)$ на множестве S абсолютно сходится. Так как мы будем пользоваться не абсолютно сходящимися интегралами только в исключительных случаях, то мы можем без недоразумений использовать более простую терминологию, принятую в тексте.

для всех x , принадлежащих S . Однако этого достаточно, чтобы установить (5.3.5) для любой интегрируемой функции $g(x)$.

Теперь приступим к обобщению свойства (5.4.5), которое само является обобщением свойства (5.3.6). Если $\lim_{v \rightarrow \infty} g_v(x) = g(x)$ почти всюду на S и если $|g_v| < G$, где G —функция, интегрируемая на множестве S , то, как и в предыдущем параграфе, отсюда следует, что $|g| \leq G$ почти всюду на S . Следовательно, функция $g(x)$ интегрируема на множестве S , и мы можем выбрать a и b так, чтобы для всех v

$$\int_{S-S_{a,b}} |g_v - g| dx < 2 \int_{S-S_{a,b}} G(x) dx < \frac{1}{2} \epsilon.$$

Множество $S_{a,b}$ имеет конечную меру, и из доказательства (5.4.5) следует, что мы можем выбрать n так, чтобы для всех $v \geq n$ выполнялось неравенство

$$\int_{S_{a,b}} |g_v - g| dx < \frac{1}{2} \epsilon.$$

При $v \geq n$ будем иметь

$$\int_S |g_v - g| dx = \int_{S_{a,b}} + \int_{S-S_{a,b}} < \epsilon.$$

Так как ϵ произвольно, то нами доказана, таким образом, следующая теорема, содержащая (5.3.6) и (5.4.5) как частные случаи:

Если предел $\lim_{v \rightarrow \infty} g_v(x) = g(x)$ существует почти всюду на множестве S конечной или бесконечной меры и если $|g_v(x)| \leq G(x)$ для всех v и для всех x , входящих в S , где $G(x)$ интегрируема на S , то функция $g(x)$ интегрируема на S и

$$(5.5.2) \quad \lim_{v \rightarrow \infty} \int_S g_v(x) dx = \int_S g(x) dx.$$

Теорема (5.5.2) может быть, конечно, также сформулирована подобно (5.3.7) как теорема о почленном интегрировании ряда. Наконец, рассуждение, использование при доказательстве (5.3.8), применимо, очевидно, и в настоящем случае и приводит к следующему обобщению этой теоремы:

Если функция $g(x)$ интегрируема на множестве S и если $S = S_1 + S_2 + \dots$, где $S_\mu S_\nu = 0$ при $\mu \neq \nu$, то

$$(5.5.3) \quad \int_S g(x) dx = \sum_{\nu=1}^{\infty} \int_{S_\nu} g(x) dx.$$

5.6. Интеграл Лебега как аддитивная функция множества.
 Рассмотрим фиксированную *неотрицательную* функцию $f(x)$, интегрируемую на всяком конечном интервале, и положим для каждого борелевского множества S

$$(5.6.1) \quad P(S) = \begin{cases} \int_S f(x) dx, & \text{если } f(x) \text{ интегрируема} \\ & \text{на множестве } S, \\ +\infty & \text{во всех остальных случаях.} \end{cases}$$

Тогда $P(S)$ есть неотрицательная функция множества S , однозначно определенная для всех борелевских множеств S . Пусть теперь $S = S_1 + S_2 + \dots$, где $S_\mu S_\nu = 0$ при $\mu \neq \nu$. Тогда из (5.5.3) следует, что выполняется соотношение аддитивности

$$P(S) = P(S_1) + P(S_2) + \dots,$$

как только $P(S)$ конечно. Однако то же самое соотношение выполняется, даже если $P(S)$ бесконечно. Если бы это было не так, то можно было бы выбрать множества S и S_1, S_2, \dots так, чтобы $P(S) = +\infty$, в то время как сумма $P(S_1) + P(S_2) + \dots$ была бы конечной. Это, однако, повлекло бы за собой соотношение

$$\int_{S_{a,b}} f(x) dx = \sum_{1}^{\infty} \int_{(S_\nu)_{a,b}} f(x) dx \leq \sum_{1}^{\infty} \int_{S_\nu} f(x) dx = \sum_{1}^{\infty} P(S_\nu).$$

Переходя к пределу при $a \rightarrow -\infty, b \rightarrow +\infty$, получим, что функция $f(x)$ интегрируема на множестве S , что противоречит сделанному допущению. Таким образом, функция $P(S)$, определенная соотношением (5.6.1), есть *неотрицательная аддитивная функция множества, определенная для всех борелевских множеств S пространства R_1* .

В частном случае, когда $f(x) = 1$, имеем $P(S) = L(S)$, так что $P(S)$ тождественно равно лебеговой мере множества S . Другой важный частный случай возникает, когда функция $f(x)$ интегрируема по всему пространству R_1 . В этом случае $P(S)$ всегда конечна, и мы имеем для каждого борелевского множества

$$P(S) \leq \int_{-\infty}^{\infty} f(x) dx.$$

ГЛАВА 6

НЕОТРИЦАТЕЛЬНЫЕ АДДИТИВНЫЕ ФУНКЦИИ МНОЖЕСТВА В R_1

6.1. Обобщение меры Лебега и интеграла Лебега. В главе 4 мы определили меру Лебега для борелевского множества S . $L(S)$ есть число, связанное с S , или, как мы выражались, *функция множества* S . Мы видели, что эта функция множества удовлетворяет трем условиям параграфа 4.2, которые требуют, чтобы функция $L(S)$ была а) неотрицательна, б) аддитивна и с) для каждого интервала равна его длине. Наконец, мы видели, что $L(S)$ есть единственная функция множества, удовлетворяющая этим трем условиям.

С другой стороны, если мы опустим условие с), то $L(S)$ не будет уже больше единственной функцией множества, удовлетворяющей остальным двум условиям. Так, например, функция $P(S)$, определенная соотношением (5.6.1), удовлетворяет условиям а) и б), в то время как с) выполняется только в частном случае $f(x) = 1$, когда $P(S) = L(S)$. Другой пример получается следующим образом. Пусть x_1, x_2, \dots — последовательность точек, а p_1, p_2, \dots — последовательность положительных чисел. Теперь положим для каждого борелевского множества S

$$P(S) = \sum_{x_v \in S} p_v,$$

где сумма распространяется на все v , для которых x_v принадлежат S . Очевидно, что функция $P(S)$, определенная таким образом, удовлетворяет условиям а) и б), но не удовлетворяет условию с).

Таким образом, мы приходим к общему понятию *неотрицательной аддитивной функции множества*, как к естественному обобщению меры Лебега $L(S)$. В настоящей главе, в параграфах 6.2—6.4, мы исследуем сначала некоторые общие свойства функций этого типа.

В применениях к теории вероятностей и статистике, которые будут изложены далее в этой книге, фундаментальную роль будет играть один специальный класс неотрицательных аддитивных функций множества. Этот класс будет рассмотрен в параграфах 6.5—6.8.

В главе 7 мы покажем, что вся теория интеграла Лебега может быть обобщена посредством замены в основном определении сумм Дарбу (см. (5.1.1)) меры Лебега $L(S)$ произвольной неотрицательной аддитивной функцией множества $P(S)$.

Обобщенный интеграл, полученный таким образом, называется интегралом Лебега — Стильбеса; он также играет основную роль в применении к теории вероятностей и математической статистике.

6.2. Функции множества и функции точки. Будем рассматривать функцию множества $P(S)$, определенную для всех борелевских множеств S и удовлетворяющую следующим трем условиям:

A) $P(S)$ неотрицательна: $P(S) \geq 0$.

B) $P(S)$ аддитивна: $P(S_1 + S_2 + \dots) = P(S_1) + P(S_2) + \dots$ ($S_\mu S_\nu = 0$ при $\mu \neq \nu$).

C) $P(S)$ конечна для всякого ограниченного множества S .

Мы будем предполагать, что все функции множества, рассматриваемые в дальнейшем, удовлетворяют этим трем условиям.

Из условий A) и B), которые совпадают с соответствующими условиями в частном случае лебеговой меры $L(S)$, мы непосредственно получаем некоторые свойства функции $P(S)$, которые доказываются так же, как и для $L(S)$. Так, если $S_1 \subset S_2$, то

$$(6.2.1) \quad P(S_1) \leq P(S_2).$$

Для пустого множества имеем $P(0) = 0$. Если S_1, S_2, \dots — последовательность множеств (безразлично, пересекающихся или нет), то

$$(6.2.2) \quad P(S_1 + S_2 + \dots) \leq P(S_1) + P(S_2) + \dots$$

(см. соотношение (4.3.7), которое, очевидно, выполняется для любых борелевских множеств). Для неубывающей последовательности S_1, S_2, \dots : имеем (см. (4.7.1))

$$(6.2.3) \quad \lim P(S_n) = P(\lim S_n).$$

Для невозрастающей последовательности то же соотношение справедливо в предположении, что $P(S_1)$ конечно.

Если множество S состоит из всех точек ξ , удовлетворяющих некоторому соотношению, то мы будем часто записывать значение $P(S)$, просто заменяя символ S в скобках соотношением, о котором идет речь. Так, например, если S есть замкнутый интервал (a, b) , то мы будем писать

$$P(S) = P(a \leq \xi \leq b).$$

Если S есть множество, состоящее из одной точки $\xi = a$, то будем писать

$$P(S) = P(\xi = a)$$

и т. д.

Мы называем $P(S)$ функцией множества, так как аргументом этой функции является множество. Для обычной функции $P(x_1, \dots, x_n)$ одной или нескольких переменных аргумент можно рассматривать как точку с координатами x_1, \dots, x_n , и поэтому мы будем часто говорить о такой функции, как о функции точки. Если даны функция множества $P(S)$ и постоянная k , то определим соответствующую функцию точки $F(x; k)$, положив

$$(6.2.4) \quad F(x; k) = \begin{cases} P(k < \xi \leq x) & \text{при } x > k, \\ 0 & \text{при } x = k, \\ -P(x < \xi \leq k) & \text{при } x < k. \end{cases}$$

Каково бы ни было значение параметра k , для любого конечного интервала (a, b)

$$F(b; k) - F(a; k) = P(a < \xi \leq b) \geq 0,$$

откуда видно, что $F(x; k)$ есть неубывающая функция от x . Если в последнем соотношении перейти к пределу при $a \rightarrow -\infty$ или $b \rightarrow +\infty$, то из (6.2.3) следует, что то же соотношение выполняется и для бесконечных интервалов. В частном случае, когда $P(S)$ есть лебегова мера $L(S)$, имеем $F(x; k) = x - k$.

Функции $F(x; k)$, соответствующие двум различным значениям параметра k , разнятся на величину, не зависящую от x . В самом деле, если $k_1 < k_2$, то

$$F(x; k_1) - F(x; k_2) = P(k_1 < \xi \leq k_2).$$

Таким образом, если мы выберем произвольное значение k_0 параметра k и обозначим соответствующую функцию $F(x; k_0)$ просто через $F(x)$, то любая другая функция $F(x; k)$ представляется в виде $F(x) + \text{const}$.

Мы можем, таким образом, сказать, что любой функции $P(S)$, удовлетворяющей условиям А), В), С), соответствует неубывающая функция точки $F(x)$, такая, что

$$(6.2.5) \quad F(b) - F(a) = P(a < \xi \leq b),$$

причем функция $F(x)$ определяется однозначно с точностью до постоянного слагаемого.

Выберем теперь произвольное, но фиксированное значение параметра k и рассмотрим соответствующую функцию $F(x)$. Так как $F(x)$ неубывающая функция, то пределы

$$F(a+0) = \lim_{x \rightarrow a+0} F(x), \quad F(a-0) = \lim_{x \rightarrow a-0} F(x)$$

существуют для всех значений a , и $F(a-0) \leq F(a+0)$. Согласно (6.2.5), имеем для $x > a$

$$F(x) - F(a) = P(a < \xi \leq x).$$

Рассмотрим это соотношение для убывающей последовательности значений x , стремящейся к фиксированному значению a . Соответствующие полуоткрытые интервалы $a < \xi \leq x$ образуют убывающую последовательность множеств. Предел этой последовательности есть пустое множество. Таким образом, согласно (6.2.3), имеем $F(x) - F(a) \rightarrow 0$, т. е.

$$F(a+0) = F(a).$$

С другой стороны, для $x < a$

$$F(a) - F(x) = P(x < \xi \leq a),$$

и аналогичное рассуждение показывает, что

$$F(a-0) = F(a) - P(\xi = a) \leq F(a),$$

Итак, функция $F(x)$ непрерывна справа. Для каждого значения x , такого, что $P(\xi = x) > 0$, функция $F(x)$ имеет разрыв со скачком $P(\xi = x)$. Для каждого значения x , такого, что $P(\xi = x) = 0$, функция $F(x)$ непрерывна.

Каждая точка x , такая, что функция $P(S)$ принимает положительное значение на множестве S , состоящем из единственной точки x , является, таким образом, точкой разрыва функции $F(x)$. Эти точки называются точками разрыва также и для функции $P(S)$, а точки непрерывности функции $F(x)$ называются также точками непрерывности функции $P(S)$.

Точки разрыва функций $P(S)$ и $F(x)$ образуют не более чем счетное множество. Действительно, рассмотрим точку разрыва x в интервале i_n , определенном неравенствами $n < x \leq n+1$, причем $P(\xi = x) > \frac{1}{c}$. Пусть S_v — множество, состоящее из v таких точек, скажем x_1, \dots, x_v . Так как S_v есть подмножество интервала i_n , то

$$P(i_n) \geq P(S_v) = P(\xi = x_1) + \dots + P(\xi = x_v) > \frac{v}{c},$$

или $v < c P(i_n)$. Таким образом, i_n содержит только конечное число

точек x , таких, что $P(\xi=x) > \frac{1}{c}$. Предполагая теперь, что c принимает значение $c=1, 2, \dots$, получим, что множество точек разрыва в интервале I_n не более чем счетно. Суммируя по $n=0, \pm 1, \pm 2, \dots$, получим (см. параграф 1.4) сформулированное предложение.

Пусть теперь X есть множество всех точек разрыва x_1, x_2, \dots функций $P(S)$ и $F(x)$. Положим $P(\xi=x_v)=p_v$. Для любого множества S произведение SX состоит из всех точек x_v , принадлежащих S , в то время как множество $S-SX=SX^*$ содержит все остальные точки множества S . Определим две новые функции P_1 и P_2 соотношениями

$$(6.2.6) \quad P_1(S) = P(SX) = \sum_{x_v \in S} p_v, \quad P_2(S) = P(SX^*).$$

Непосредственно видно, что P_1 и P_2 удовлетворяют условиям А), В), С). Далее, имеем $S=SX+SX^*$ и, следовательно,

$$(6.2.7) \quad P(S) = P_1(S) + P_2(S).$$

Из (6.2.6) следует, что $P_1(S)$ есть сумма скачков p_v во всех точках разрыва x_v , принадлежащих S . Значит $P_1(S)=0$ для всякого множества S , не содержащего точек x_v . С другой стороны, соотношение (6.2.6) показывает, что функция $P_2(S)$ везде непрерывна, так как все точки, принадлежащие X^* , суть точки непрерывности функции $P(S)$. Таким образом, соотношение (6.2.7) дает разложение неотрицательной аддитивной функции множества $P(S)$ на *разрывную часть* $P_1(S)$ и *непрерывную часть* $P_2(S)$.

Если F, F_1 и F_2 — неубывающие функции точки, соответствующие P, P_1 и P_2 , и если во всех трех случаях выбрана одна и та же постоянная k , то из (6.2.4) и (6.2.7) вытекает

$$(6.2.8) \quad F(x) = F_1(x) + F_2(x).$$

В этом соотношении функция $F_2(x)$ непрерывна, а функция $F_1(x)$ есть „ступенчатая функция“, постоянная на каждом интервале, не содержащем точек x_v , и имеющая „скакок“ p_v в каждой точке x_v . Легко видеть, что любая неубывающая функция $F(x)$ может быть представлена в форме (6.2.8), и притом единственным способом, как сумма двух неубывающих функций, одна из которых — ступенчатая, а другая — непрерывная.

6.3. Построение функции множества. Теперь мы докажем следующее обращение теоремы (6.2.5):

Каждой неубывающей функции точки $F(x)$, конечной при всех конечных значениях x и всюду непрерывной справа, соответствует функция множества $P(S)$, однозначно определенная для всех борелевских множеств S , удовлетворяющая условиям А), В), С) параграфа 6.2 и такая, что соотношение

$$F(b) - F(a) = P(a < \xi \leq b)$$

выполнено для любого конечного или бесконечного интервала (a, b) . Очевидно, что две функции $F_1(x)$ и $F_2(x)$ порождают одну и ту же функцию $P(S)$ тогда и только тогда, когда разность $F_1 - F_2$ есть постоянная.

Сравнивая это утверждение с теоремой (6.2.5), находим, что если две функции F_1 и F_2 , отличающиеся друг от друга на постоянную, считать тождественными, то существует взаимооднозначное соответствие между функциями множества $P(S)$ и неубывающими функциями точки $F(x)$.

Прежде всего неубывающая функция точки $F(x)$ определяет неотрицательную функцию интервала $P(i)$, которая может быть определена как приращение функции $F(x)$ на интервале i . Для каждого полуоткрытого интервала $a < x \leq b$ полагаем $P(i) = P(a < x \leq b) = F(b) - F(a)$. Для остальных трех типов интервалов, с теми же конечными точками a и b , определим значение $P(i)$ с помощью простого предельного перехода:

$$(6.3.1) \quad \begin{aligned} P(a \leq x \leq b) &= F(b) - F(a - 0), \\ P(a < x < b) &= F(b - 0) - F(a), \\ P(a < x \leq b) &= F(b) - F(a), \\ P(a \leq x < b) &= F(b - 0) - F(a - 0); \end{aligned}$$

таким образом, функция $P(i)$ определена для всех интервалов i .

Теорема будет доказана, если мы покажем, что существует неотрицательная аддитивная функция множества, определенная для всех борелевских множеств S и совпадающая с $P(i)$ в случае, если множество S есть интервал.

Это есть обобщение проблемы, рассмотренной в главе 4. Там мы имели частный случай $F(x) = x$ с соответствующей функцией интервала i — длиной $L(i)$. Вся теория меры Лебега, как она развивалась в главе 4, состоит в построении неотрицательной аддитивной функции

множества, определенной для всех борелевских множеств S и совпадающей с функцией $L(i)$, когда S есть интервал i . Теперь требуется провести аналогичное построение для случая, когда длина или „ L -мера“ интервала $L(i) = b - a$, заменена более общей „ P -мерой“, определенной соотношением (6.3.1).

Это может быть сделано точно тем же методом, какой мы применяли в частном случае в главе 4. За двумя небольшими исключениями, которые будут оговорены ниже, каждое утверждение и каждая формула главы 4 останутся справедливыми, если 1) слова *мера* и *измеримое* заменить словами *P-мера* и *P-измеримое*, 2) длину $L(i) = b - a$ интервала заменить *P-мерой* $P(i)$ и 3) символы L и \mathfrak{P} заменить символами P и \mathfrak{B} . Этим способом, строго следуя схеме, изложенной в параграфах 4.1 — 4.5, мы устанавливаем существование неотрицательной аддитивной функции множества $P(S)$, однозначно определенной для некоторого класса \mathfrak{B} множеств, называемых *P-измеримыми*, и совпадающей с $P(i)$ в тех случаях, когда множество S есть интервал i . Далее, точно так же, как и в параграфе 4.6, доказывается, что класс \mathfrak{B} всех *P-измеримых* множеств есть аддитивный класс, содержащий все борелевские множества. Наконец, так же как и в параграфе 4.7, доказывается, что $P(S)$ есть единственная неотрицательная аддитивная функция, определенная для всех борелевских множеств, которая сводится к функции интервала $P(i)$ в случае, если множество S есть интервал i . Наша теорема, таким образом, доказана. Более того, из доказательства становится ясно, почему полезно ограничиться рассмотрением борелевских множеств. Действительно, несмотря на то, что класс всех *P-измеримых* множеств зависит от исходной функции $F(x)$, он всегда содержит весь класс \mathfrak{B} , борелевских множеств. Таким образом, каждое борелевское множество всегда *P-измеримо*, и функция множества $P(S)$, соответствующая любой заданной функции $F(x)$, всегда может быть определена для борелевских множеств.

Остается только сделать те оговорки, которые упоминались выше. Содержание первой очень просто и не имеет прямого отношения к доказательству нашей теоремы. В параграфе 4.3 мы доказали, что лебегова мера счетного множества всегда равна нулю. Это следует из того, что счетное множество можно рассматривать как сумму последовательности вырожденных интервалов, каждый из которых имеет длину, равную нулю. Соответствующее предложение для *P*-меры очевидно неверно, если функция $F(x)$ имеет хотя бы одну точку разрыва. Вырожденный

интервал, состоящий из единственной точки a , может иметь положительную P -меру, так как, согласно первому из соотношений (6.3.1),

$$P(x=a) = F(a) - F(a-0).$$

Если счетное множество содержит хотя бы одну точку разрыва функции $F(x)$, то оно имеет положительную P -меру.

Вторая оговорка возникает в связи с обобщением результатов параграфа 4.1, где было доказано, что длина есть аддитивная функция интервала. Для того чтобы установить аналогичное предложение для P -меры, мы должны показать, что

$$(6.3.2) \quad P(i) = P(i_1) + P(i_2) + \dots,$$

где i и i_1, i_2, \dots — интервалы, такие, что $i = i_1 + i_2 + \dots$ и $i_\mu i_\nu = 0$ для $\mu \neq \nu$.

Для непрерывной функции $F(x)$ это доказывается с помощью леммы Бореля, так же, как и в случае соответствующего соотношения (4.1.1), только с заменой длины интервала его P -мерой. Заметим, что в ходе доказательства равенства (4.1.1) мы рассматривали некоторые интервалы, например, интервал $(a-\varepsilon, a+\varepsilon)$, который был выбран так, чтобы его длина равнялась 2ε . При обобщении этого доказательства на случай P -меры мы должны заменить этот интервал интервалом $(a-h, a+h)$, в котором h выбрано так, чтобы P -мера этого интервала $F(a+h) - F(a-h)$ была равна 2ε .

С другой стороны, если $F(x)$ — ступенчатая функция, имеющая на интервале i точки разрыва x_1, x_2, \dots с соответствующими скачками p_1, p_2, \dots , то мы имеем

$$P(i) = \sum_1^{\infty} p_i, \quad P(i_n) = \sum_{x_i \in i_n} p_i.$$

Так как интервалы i_n попарно не пересекаются, то каждая точка x_i принадлежит в точности одному интервалу i_n , и справедливость соотношения (6.3.2) вытекает из свойств сходящихся двойных рядов.

Наконец, согласно замечанию, сделанному в связи с (6.2.8), каждая функция $F(x)$ есть сумма некоторой ступенчатой функции F_1 и некоторой непрерывной функции F_2 , причем обе функции F_1 и F_2 — неубывающие. Для обеих этих функций соотношение (6.3.2) выполняется, следовательно, оно справедливо также и для их суммы $F(x)$. Мы рассмотрели, таким образом, оба отклонения от главы 4 в обоб-

щении результатов главы 4 на случай произвольной P -меры, и доказательство теоремы, сформулированной в начале этого параграфа, этим завершено.

6.4. P -мера. Функция $P(S)$, удовлетворяющая условиям А), В), С) параграфа 6.2, определяет P -меру множества S , которая является обобщением лебеговой меры $L(S)$. Подобно последней P -мера неотрицательна и аддитивна.

Согласно предыдущему параграфу, P -мера однозначно определена для каждого борелевского множества S , если известна соответствующая неубывающая функция точки $F(x)$. Так как, согласно параграфу 6.2, функция $F(x)$ всегда непрерывна справа, то достаточно знать значения функции $F(x)$ во всех ее точках непрерывности.

Если для некоторого множества S мы имеем $P(S)=0$, то будем говорить, что S есть множество P -меры нуль. Согласно (6.2.1), любое подмножество S также имеет тогда P -меру нуль. Сумма последовательности множеств P -меры нуль, согласно (6.2.2), сама имеет P -меру, равную нулю. Если $F(a)=F(b)$, то полуоткрытый интервал $a < x \leq b$ есть множество P -меры нуль.

Если некоторое условие выполнено для всех точек, принадлежащих множеству S , за исключением, может быть, некоторого подмножества точек P -меры нуль, то говорим (см. параграф 5.3), что это условие выполнено *почти всюду* (P) или для *почти всех* (P) точек на S .

6.5. Ограниченные функции множества. Для любого борелевского множества S , согласно (6.2.1), имеем $P(S) \leq P(R_1)$. Если значение $P(R_1)$ конечно, то мы говорим, что функция множества $P(S)$ *ограничена*. Если функция $P(S)$ ограничена, то будем всегда фиксировать постоянное слагаемое в соответствующей неубывающей функции точки $F(x)$, полагая в (6.2.4) $k = -\infty$, так что для любого значения x

$$(6.5.1) \quad F(x) = P(\xi \leq x).$$

Если x стремится в этом соотношении к $-\infty$, то множество всех точек $\xi \leq x$ имеет своим пределом (см. параграф 1.5) пустое множество. Таким образом, согласно (6.2.3), имеем $F(-\infty) = 0$. С другой стороны, если $x \rightarrow +\infty$, то множество точек $\xi \leq x$ имеет своим пределом все пространство R_1 и (6.2.3) дает $F(+\infty) = P(R_1)$. Так как функция $F(x)$ — неубывающая, то для всех x имеем

$$(6.5.2) \quad 0 \leq F(x) \leq P(R_1).$$

6.6. Распределения. Фундаментальную роль в приложениях к теории вероятностей и математической статистике играют такие неотрицательные аддитивные функции множеств, для которых $P(R_1) = 1$. Функция $P(S)$, принадлежащая этому классу, очевидно ограничена, а соответствующая неубывающая функция точки $F(x)$ определяется, согласно (6.5.1), так, что

$$(6.6.1) \quad \begin{aligned} F(x) &= P(\xi \leq x), \\ 0 &\leq F(x) \leq 1, \\ F(-\infty) &= 0, \quad F(+\infty) = 1. \end{aligned}$$

Пара функций $P(S)$ и $F(x)$ этого типа часто будет интерпретироваться посредством *распределения массы по одномерному пространству R_1* . Представим себе единицу массы, распределенную по пространству R_1 так, что для каждого x количество массы, распределенное на бесконечном интервале $\xi \leq x$, равно $F(x)$. Построение функции множества $P(S)$ посредством данной функции точки $F(x)$, описанное в параграфе 6.3, может быть интерпретировано теперь так, что каждое борелевское множество S несет на себе определенное количество массы $P(S)$. Общее количество массы на всей прямой есть $P(R_1) = 1$.

Мы можем задавать это распределение как с помощью функции $P(S)$, так и с помощью соответствующей функции точки $F(x)$. Употребляя терминологию, приспособленную к приложениям этих понятий к теории вероятностей и математической статистике, которые будут делаться в дальнейшем, будем называть функцию $P(S)$ *вероятностной функцией распределения*, а функцию $F(x)$ — *функцией распределения*.

Таким образом, функция распределения есть неубывающая функция точки $F(x)$, которая всюду непрерывна справа и такова, что $F(-\infty) = 0$ и $(F+\infty) = 1$. Обратно, из параграфа 6.3 следует, что любая заданная функция $F(x)$, обладающая этими свойствами, определяет единственное распределение, имеющее $F(x)$ своей функцией распределения.

Если x_0 есть точка разрыва функции $F(x)$ со скачком, равным p_0 , то масса p_0 будет сосредоточена в точке x_0 , которую назовем *точкой сосредоточения массы* в данном распределении. С другой стороны, если x_0 есть точка непрерывности, то количество массы, находящееся в интервале $(x-h, x+h)$ будет стремиться к нулю при $h \rightarrow 0$.

Отношение $\frac{F(x+h) - F(x-h)}{2h}$ есть средняя плотность массы, находящейся в интервале $x-h < \xi \leq x+h$. Если существует производная $F'(x) = f(x)$, то средняя плотность стремится к $f(x)$ при $h \rightarrow 0$, и, таким образом, $f(x)$ представляет собой *плотность массы в точке* x . В приложениях к теории вероятностей $f(x)$ называется *плотностью вероятности* или *функцией плотности* распределения. Всякая плотность вероятности $f(x)$ есть неотрицательная функция, интеграл которой по бесконечному интервалу $(-\infty, \infty)$ равен единице.

Из (6.2.7) и (6.2.8) следует, что каждое распределение может быть разложено на разрывную и непрерывную части

$$(6.6.2) \quad P(S) = c_1 P_1(S) + c_2 P_2(S), \\ F(x) = c_1 F_1(x) + c_2 F_2(x).$$

Здесь c_1 и c_2 — неотрицательные постоянные, такие, что $c_1 + c_2 = 1$; P_1 и F_1 обозначают соответственно вероятностную функцию и функцию распределения массы в точках сосредоточения массы (таким образом, F_1 есть ступенчатая функция). Функции же P_2 и F_2 соответствуют распределению без точек сосредоточения массы (таким образом, функция F_2 всюду непрерывна). Постоянные c_1 и c_2 , так же как и функции P_1 , P_2 , F_1 и F_2 , однозначно определяются заданным распределением.

В предельном случае, когда $c_1 = 1$, $c_2 = 0$, функция распределения $F(x)$ есть ступенчатая функция, и вся масса в этом распределении сосредоточена в точках разрыва функции $F(x)$, каждая из которых заключает в себе массу, равную соответствующему скачку. В противоположном предельном случае, когда $c_1 = 0$, $c_2 = 1$, $F(x)$ всюду непрерывна и не существует ни одной точки, в которой было бы сосредоточено положительное количество массы.

Общая теория распределений в R_1 будет подробно развита в главе 15. В последующих главах 16—19 будут разобраны и проиллюстрированы на примерах некоторые важные специальные распределения. Сейчас же читатель сможет найти графики распределения для случая $c_1 = 1$, $c_2 = 0$ на стр. 191 (фиг. 4—5) и для случая $c_1 = 0$, $c_2 = 1$ на стр. 193 (фиг. 6—7).

6.7. Последовательности распределений. Будем называть интервал (a, b) *интервалом непрерывности* для данной неотрицательной аддитивной функции множества $P(S)$ и для соответствующей функции

точки $F(x)$, если a и b суть точки непрерывности (см. параграф 6.2) функций $P(S)$ и $F(x)$ *. Если две функции множества совпадают для всех интервалов, являющихся интервалами непрерывности для обеих этих функций, то легко видеть, что соответствующие функции точки отличаются друг от друга на постоянное слагаемое, так что сами функции множества совпадают.

Рассмотрим теперь последовательность распределений с вероятностными функциями $P_1(S)$, $P_2(S)$, ... и функциями распределения $F_1(x)$, $F_2(x)$, Будем говорить, что последовательность распределений *сходится*, если существует неотрицательная аддитивная функция множества $P(S)$, такая, что $P_n(S) \rightarrow P(S)$ для любого интервала непрерывности S функции $P(S)$.

Так как всегда $0 \leq P_n(S) \leq 1$, то для сходящейся последовательности имеем $0 \leq P(S) \leq 1$ для любого интервала непрерывности $S = (a, b)$. Если $a \rightarrow -\infty$, $b \rightarrow +\infty$, то из (6.2.3) вытекает, что $P(R_1) \leq 1$. Случай, когда $P(R_1) = 1$, имеет специальный интерес. В этом случае $P(S)$ есть вероятностная функция некоторого распределения, и поэтому мы будем говорить, что наша последовательность *сходится к некоторому распределению*, а именно, к распределению, соответствующему $P(S)$. Обычно только этот случай сходимости интересен в приложениях и нам часто будет нужен критерий, который позволял бы определить, сходится ли наша последовательность распределений к некоторому распределению или нет. Задача отыскания такого критерия решается в дальнейшем (см. параграф 10.4), сейчас же мы только докажем следующее предварительное предложение:

Последовательность распределений с функциями распределения $F_1(x)$, $F_2(x)$, ... сходится к некоторому распределению тогда и только тогда, когда существует функция распределения $F(x)$, такая, что $F_n(x) \rightarrow F(x)$ в каждой точке непрерывности функции $F(x)$. Если такая функция $F(x)$ существует, то она соответствует распределению, являющемуся пределом нашей последовательности распределений, и мы будем коротко говорить, что последовательность $\{F_n(x)\}$ сходится к функции распределения $F(x)$.

Покажем сперва, что сформулированное условие необходимо и что функция распределения $F(x)$ соответствует предельному распределению.

*) Заметим, что *внутренняя* точка интервала может быть и точкой разрыва.

Обозначая, как обычно, через $P_n(S)$ вероятностную функцию, соответствующую $F_n(x)$, мы допустим, что $P_n(S)$ стремится к некоторой вероятностной функции $P(S)$ для любого интервала непрерывности $S = (a, b)$ функции $P(S)$. Обозначим через $F(x)$ функцию распределения, соответствующую $P(S)$. Мы должны показать, что $F_n(x) \rightarrow F(x)$, где x — произвольная точка непрерывности функции $F(x)$. Так как $P(R_1) = 1$, то мы можем выбрать интервал непрерывности $S = (a, b)$, содержащий x , такой, что $P(S) > 1 - \epsilon$, где $\epsilon > 0$ произвольно мало. Тогда $1 - \epsilon < P(S) = F(b) - F(a) \leq 1 - F(a)$, следовательно, $0 \leq F(a) < \epsilon$. Далее, по предположению, имеем $F_n(b) - F_n(a) \rightarrow F(b) - F(a) > 1 - \epsilon$, так что для всех достаточно больших n имеем $F_n(b) - F_n(a) > 1 - 2\epsilon$ или $0 \leq F_n(a) < F_n(b) - 1 + 2\epsilon \leq 2\epsilon$. Так как (a, x) есть интервал непрерывности функции $P(S)$, то, согласно предположению, имеем $F_n(x) - F_n(a) \rightarrow F(x) - F(a)$. Для всех достаточно больших n имеем, таким образом, $|F_n(x) - F(x) - F_n(a) + F(a)| < \epsilon$ и, следовательно, согласно сказанному выше, $|F_n(x) - F(x)| < 3\epsilon$. Так как ϵ произвольно, то $F_n(x) \rightarrow F(x)$.

Обратно, если мы допустим, что $F_n(x)$ стремится к функции распределения $F(x)$ в каждой точке непрерывности функции $F(x)$, и обозначим через $P(S)$ вероятностную функцию, соответствующую $F(x)$, то очевидно, что $F_n(b) - F_n(a) \rightarrow F(b) - F(a)$, т. е. $P_n(S) \rightarrow P(S)$, где S — любой полуоткрытый интервал непрерывности $a < x \leq b$ функции $P(S)$. Далее, так как $F(x)$ есть неубывающая функция, непрерывная при $x = a$ и $x = b$, то $F_n(a - 0) \rightarrow F(a)$ и $F_n(b - 0) \rightarrow F(b)$. Отсюда следует то же соотношение $P_n(S) \rightarrow P(S)$ для любого интервала непрерывности S , замкнутого, открытого или полуоткрытого. Таким образом, наше предложение доказано.

Для того чтобы показать на примере, что последовательность распределений может быть сходящейся, не сходясь в то же время к распределению, мы рассмотрим распределение, при котором вся единица массы сосредоточена в одной точке $x = 0$. Обозначая соответствующую функцию распределения через $\epsilon(x)$, имеем

$$(6.7.1) \quad \epsilon(x) = \begin{cases} 0 & \text{при } x < 0, \\ 1 & \text{при } x \geq 0. \end{cases}$$

Тогда $\epsilon(x - a)$ есть функция распределения, при котором вся масса (равная единице) сосредоточена в точке $x = a$. Рассмотрим теперь последовательность распределений, определяемых функциями распределения $F_n(x) := \epsilon(x - n)$, где $n = 1, 2, \dots$. Очевидно, что, согласно данному выше определению, эта последовательность сходится, так как масса, содержащаяся в лю-

бом интервале, стремится к нулю при $n \rightarrow \infty$. Предельная функция множества, однако, тождественно равна нулю и не является, таким образом, вероятностной функцией. При $n \rightarrow \infty$ масса в наших распределениях исчезает, „уходя“ по направлению к $+\infty$.

Можно было бы спросить, почему в нашем определении сходимости мы не требуем, чтобы $P_n(S) \rightarrow P(S)$ имело место для каждого борелевского множества S . Однако легко показать, что это определение было бы слишком узким. Действительно, рассмотрим последовательность распределений, определяемых функциями распределения $\varepsilon(x - \frac{1}{n})$, где $n = 1, 2, \dots$. При n -м распределении в этой последовательности вся единица массы сосредоточена в точке $x = \frac{1}{n}$. Очевидно, что разумное определение сходимости должно быть таким, чтобы эта последовательность сходилась к распределению, определяемому соотношением (6.7.1), при котором вся единица массы сосредоточена в точке $x = 0$. Легко проверить, что определение сходимости, данное выше, удовлетворяет этому условию. Если, с другой стороны, мы рассмотрим множество S , состоящее из единственной точки $x = 0$, то $P_n(S) = 0$ при любом n , в то время как для предельного распределения $P(S) = 1$, так что $P_n(S)$ не стремится к $P(S)$. Функция распределения $\varepsilon(x - \frac{1}{n})$ стремится к $\varepsilon(x)$ в каждой точке непрерывности функции $\varepsilon(x)$, т. е. при любом $x \neq 0$, но не в точке разрыва $x = 0$.

6.8. Теорема сходимости. Последовательность функций распределения $F_1(x), F_2(x), \dots$ называется *сходящейся*, если существует неубывающая функция $F(x)$, такая, что $F_n(x) \rightarrow F(x)$ в каждой точке непрерывности функции $F(x)$. Всегда имеет место $0 \leq F(x) \leq 1$, но пример $F_n(x) = \varepsilon(x - n)$, рассмотренный в предыдущем параграфе, показывает, что $F(x)$ не всегда является функцией распределения. Таким образом, последовательность $\{F_n(x)\}$ может быть *сходящейся* и в то же время не сходить к функции распределения. Докажем теперь следующее предложение, которое потребуется в дальнейшем. *Каждая последовательность $\{F_n(x)\}$ функций распределения содержит сходящуюся подпоследовательность. Ее предел $F(x)$ всегда может быть определен так, чтобы он был всюду непрерывен справа.*

Пусть r_1, r_2, \dots есть счетное множество всех положительных и отрицательных рациональных чисел, включая нуль. Рассмотрим последовательность $F_1(r_1), F_2(r_1), \dots$. Это — ограниченная бесконечная последовательность действительных чисел, которая по теореме Больцано-Вейерштрасса (см. параграф 2.2) имеет по крайней мере одну предельную точку. Последовательность чисел $\{F_n(r_1)\}$ всегда содержит, таким

образом, сходящуюся подпоследовательность. То же самое можно выразить, говоря, что последовательность функций $\{F_n(x)\}$ всегда содержит подпоследовательность Z_1 , сходящуюся для частного значения $x = r_1$. Таким же образом находим, что Z_1 содержит подпоследовательность, сходящуюся при $x = r_2$. Повторяя этот процесс, мы получаем одну за другой последовательности Z_1, Z_2, \dots , где Z_n есть подпоследовательность последовательности Z_{n-1} , и Z_n сходится для частных значений $x = r_1, r_2, \dots, r_n$. Образуем теперь „диагональную“ последовательность Z , состоящую из первого члена последовательности Z_1 , второго члена последовательности Z_2 и т. д.. Нетрудно видеть, что последовательность Z сходится при любом рациональном значении x .

Пусть члены последовательности Z будут $F_{n_1}(x), F_{n_2}(x), \dots$. Положим

$$\lim_{n \rightarrow \infty} F_{n_i}(r_i) = c_i \quad (i = 1, 2, \dots).$$

Тогда $\{c_i\}$ есть ограниченная последовательность, и так как каждая функция F_{n_i} — неубывающая, то $c_i \leq c_k$ при $r_i \leq r_k$.

Теперь определим функцию $F(x)$ следующим образом:

$$F(x) = \inf c_i \text{ для всех } r_i > x.$$

Непосредственно из определения следует, что $F(x)$ есть ограниченная неубывающая функция от x . Легко доказать также, что $F(x)$ всюду непрерывна справа. Покажем теперь, что в каждой точке непрерывности функции $F(x)$ мы имеем

$$(6.8.1) \quad \lim_{n \rightarrow \infty} F_{n_i}(x) = F(x),$$

т. е. подпоследовательность Z сходится.

Если x есть точка непрерывности функции $F(x)$, то мы можем выбрать $h > 0$ так, чтобы разность $F(x+h) - F(x-h)$ была меньше любого заданного $\epsilon > 0$. Пусть r_i и r_k — рациональные точки, принадлежащие соответственно интервалам $(x-h, x)$ и $(x, x+h)$, так что

$$(6.8.2) \quad F(x-h) \leq c_i \leq F(x) \leq c_k \leq F(x+h).$$

Далее, для каждого y имеем

$$(6.8.3) \quad F_{n_y}(r_i) \leq F_{n_y}(x) \leq F_{n_y}(r_k).$$

$F_{n_i}(r_i)$ и $F_{n_k}(r_k)$ стремятся соответственно к пределам c_i и c_k при $i \rightarrow \infty$. Разность между этими пределами, согласно (6.8.2), меньше ε , и величина $F(x)$ заключена между c_i и c_k . Так как ε произвольно, то $F_{n_\nu}(x)$ стремится к $F(x)$. Таким образом, последовательность Z сходится, и наша теорема доказана.

ГЛАВА 7

ИНТЕГРАЛ ЛЕБЕГА — СТИЛЬТЬЕСА ОТ ФУНКЦИЙ ОДНОГО ПЕРЕМЕННОГО

7.1. Интеграл от ограниченной функции по множеству конечной P -меры. В предыдущей главе мы видели, что теория меры Лебега, развитая в главе 4, может быть обобщена с помощью введения общего понятия неотрицательной аддитивной P -меры. Покажем теперь, что совершенно аналогичное обобщение может быть применено к теории интеграла Лебега, развитой в главе 5.

Предположим, что задана некоторая фиксированная P -мера. Эта мера может быть определена неотрицательной аддитивной функцией множества $P(S)$ или соответствующей неубывающей функцией точки $F(x)$. Мы видели в предыдущей главе, что каждая из этих двух функций может одинаково хорошо служить для определения P -меры.

Пусть, далее, $g(x)$ есть данная функция от x , определенная и ограниченная на данном множестве S конечной P -меры. Так же, как и в параграфе 5.1, мы разбиваем множество S на произвольное конечное число попарно не пересекающихся множеств S_1, S_2, \dots, S_n . Заменяя теперь в основном определении (5.1.1) сумм Дарбу L -меру P -мерой, получим обобщенные суммы Дарбу

$$(7.1.1) \quad z = \sum_1^n m_i P(S_i), \quad Z = \sum_1^n M_i P(S_i),$$

где, как и раньше, m_i и M_i обозначают нижнюю и верхнюю границы функции $g(x)$ на множестве S_i .

Дальнейшее развитие вполне аналогично проведенному в параграфе 5.1. Верхняя граница множества всевозможных значений z называется *нижним интегралом* функции $g(x)$ по множеству S относительно данной P -меры. Нижняя граница множества всевозможных значений Z

есть соответственно *верхний интеграл*. Как и в параграфе 5.1, показывается, что нижний интеграл меньше или равен верхнему интегралу.

Если нижний и верхний интегралы равны, то функция $g(x)$ называется *интегрируемой на множестве S относительно данной P -меры*, а общее значение обоих интегралов называется *интегралом Лебега — Стильтьеса функции $g(x)$ по множеству S относительно данной P -меры* и обозначается одним из следующих двух выражений

$$\int_S g(x) dP(S) = \int_S g(x) dF(x).$$

В тех случаях, когда это не может привести к недоразумению, будем писать вместо $dP(S)$ и $dF(x)$ просто dP и dF . Вместо „интеграл относительно данной P -меры“ будем обычно говорить „интеграл относительно $P(S)$ “ или „относительно $F(x)$ “, в зависимости от того, определена ли P -мера посредством функции $P(S)$ или посредством функции $F(x)$. Пока мы будем иметь дело с функциями одной переменной, мы будем, как правило, употреблять функцию $F(x)$.

В частном случае, когда $F(x) = x$, имеем $P(S) = L(S)$, и очевидно, что данное нами определение интеграла Лебега — Стильтьеса сводится к определению интеграла Лебега, данному в параграфе 5.1. Таким образом, интеграл Лебега — Стильтьеса получается из интеграла Лебега просто заменой в определении интеграла лебеговой меры более общей P -мерой.

Все свойства интеграла Лебега, выведенные в параграфах 5.2 и 5.3, легко обобщаются на случай интеграла Лебега — Стильтьеса. Доказательства этих свойств остаются точно теми же, только везде L -мера заменяется P -мерой. Так, мы находим, что, если $g(x)$ ограничена и B -измерима на множестве S конечной P -меры, то $g(x)$ интегрируема на множестве S относительно $P(S)$. Для ограниченных функций и множеств конечной P -меры, мы получаем следующие обобщения свойств, выведенных в параграфе 5.3:

$$(7.1.2) \quad \int_S (g_1(x) + g_2(x)) dF = \int_S g_1(x) dF + \int_S g_2(x) dF,$$

$$(7.1.3) \quad \int_S cg(x) dF = c \int_S g(x) dF,$$

$$(7.1.4) \quad mP(S) \leq \int_S g(x) dF \leq MP(S),$$

$$(7.1.5) \quad \int_{S_1 + S_2} g(x) dF = \int_{S_1} g(x) dF + \int_{S_2} g(x) dF,$$

$$(7.1.6) \quad \left| \int_S g(x) dF \right| \leq \int_S |g(x)| dF,$$

где c — постоянная, m и M обозначают нижнюю и верхнюю границы функции $g(x)$ на множестве S , а S_1 и S_2 — два непересекающихся множества. Из (7.1.4) следует, что интеграл ограниченной функции по множеству P -меры нуль равен нулю. Таким образом, значение интеграла не изменится от произвольного изменения значений функции $g(x)$ на множестве P -меры нуль.

Имеет место также следующее обобщение предложения (5.3.6). Если последовательность $\{g_n(x)\}$ равномерно ограничена на множестве S и если $\lim_{n \rightarrow \infty} g_n(x) = g(x)$ существует почти всюду (P) на S , то

$$(7.1.7) \quad \lim_{n \rightarrow \infty} \int_S g_n(x) dF = \int_S g(x) dF.$$

Аналогичные обобщения соотношений (5.3.7) и (5.3.8) получаются точно так же, как и в параграфе 5.3.

Если c_1 и c_2 — неотрицательные постоянные, то легко вывести следующее соотношение, не имеющее аналога в случае интеграла Лебега:

$$\int_S g(x) d(c_1 F_1 + c_2 F_2) = c_1 \int_S g(x) dF_1 + c_2 \int_S g(x) dF_2.$$

В частном случае, когда множество S состоит из единственной точки x_0 , непосредственно из определения получаем

$$\int_{\{x=x_0\}} g(x) dF = g(x_0) P(x = x_0).$$

Рассмотрим теперь случай, когда $F(x)$ есть ступенчатая функция (см. параграф 6.2) со скачками p , в точках $x = x_i$, и обозначим множество всех точек x , через X . Используя то, что интеграл по множеству P -меры нуль равен нулю, и упомянутое выше обобщение (5.3.8), получим

$$(7.1.8) \quad \int_S g(x) dF = \int_{SX} g(x) dF = \sum_{x_i \in S} \int_{(x=x_i)} g(x) dF = \sum_{x_i \in S} p_i g(x_i).$$

В частном случае, когда $g(x) = 1$, имеем

$$\int_S dF = \int_S dP = P(S).$$

Мы будем часто рассматривать интегралы от функций $g(x)$, принимающих комплексные значения $g(x) = a(x) + ib(x)$, где $a(x)$ и $b(x)$ — действительные ограниченные функции на множестве S . В этом случае определим интеграл так:

$$\int_S g(x) dF = \int_S a(x) dF + i \int_S b(x) dF.$$

Все свойства, выведенные выше, легко распространяются на интегралы этого типа. Для соотношения (7.1.6) это распространение несколько менее очевидно, чем в других случаях, и мы его здесь проведем. Положим

$$\int_S g(x) dF = re^{iv},$$

где r и v — действительные числа, причем $r \geq 0$. Действительная часть величины $|g(x)| = e^{-iv}g(x)$ всегда неотрицательна. Следовательно, действительный интеграл

$$\begin{aligned} \int_S (|g(x)| - e^{-iv}g(x)) dF &= \int_S |g(x)| dF - r = \\ &= \int_S |g(x)| dF - \left| \int_S g(x) dF \right| \geq 0, \end{aligned}$$

а это эквивалентно соотношению (7.1.6).

7.2. Неограниченные функции и множества бесконечной P -меры. Распространение понятия интеграла Лебега на случаи неограниченных функций и множеств бесконечной меры, рассмотренное в параграфах 5.4 и 5.5, может быть совершенно аналогично проведено и для интеграла Лебега — Стильтьеса. Действительно, любое утверждение и любая формула параграфов 5.4 и 5.5 сохраняются при замене меры Лебега и интеграла Лебега P -мерой и интегралом Лебега — Стильтьеса относительно P -меры.

Таким образом, функция $g(x)$ называется *интегрируемой относительно $P(S)$* (или *относительно $F(x)$*) на множестве S конечной P -меры, если предел (см. (5.4.1)) .

$$\lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_S g_{a,b}(x) dP = \int_S g(x) dP = \int_S g(x) dF$$

существует и имеет конечное значение. В этом случае функция $|g(x)|$ также интегрируема относительно P по множеству S .

Далее, если S есть множество бесконечной P -меры *), то функция $g(x)$ называется интегрируемой относительно P (или относительно F) на множестве S , если (см. параграф 5.5) $g(x)$ интегрируема относительно P по всем множествам $S_{a,b}$ и если предел

$$\lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_{S_{a,b}} |g(x)| dP = \int_S |g(x)| dP = \int_S |g(x)| dF$$

существует и имеет конечное значение. В этом случае предел (см. (5.5.1))

$$(7.2.1) \quad \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_{S_{a,b}} g(x) dP = \int_S g(x) dP = \int_S g(x) dF$$

также существует и конечен, и мы будем говорить, что интеграл Лебега — Стильтьеса функции $g(x)$ относительно P (или относительно F) по множеству S сходится **) и предел (7.2.1) есть, по определению, значение этого интеграла. Если $|g(x)| < G(x)$, где $G(x)$ — интегрируемая функция, то $g(x)$ также интегрируема.

Свойства (7.1.2) — (7.1.6) интеграла Лебега — Стильтьеса справедливы для любых функций, интегрируемых относительно данной P -меры. В случае, когда S есть множество бесконечной P -меры, неравенства (7.1.4) должно быть заменено, однако, неравенством:

$$\int_S g(x) dF \geq 0, \text{ если } g(x) \geq 0 \text{ для любых } x \text{ из } S.$$

Наконец, мы имеем следующее обобщение предложения (7.1.7):

Если $\lim_{\nu} g_\nu(x) = g(x)$ существует почти всюду (P) на множестве S конечной или бесконечной P -меры и если $|g_\nu(x)| < G(x)$ для всех ν и всех x из S , где $G(x)$ есть функция, интегрируемая относительно F на множестве S , то $g(x)$ интегрируема относительно F на S и

$$(7.2.2) \quad \lim_{\nu \rightarrow \infty} \int_S g_\nu(x) dF = \int_S g(x) dF.$$

*) В случае, если функция $P(S)$ ограничена (например, если $P(S)$ есть вероятностная функция, см. параграф 6.6), множество бесконечной P -меры, конечно, не существует.

**) Относительно терминологии может быть сделано такое же замечание, что и в случае (5.5.1).

Обобщение проведенных выше рассмотрений на случай функции $g(x)$, принимающей комплексные значения, очевидно.

В частном случае, когда $F(x) = x$, все наши теоремы сводятся, конечно, к соответствующим теоремам относительно обычного интеграла Лебега.

7.3. Интегралы Лебега — Стильеса с параметром. Мы часто будем иметь дело с интегралами типа

$$u(t) = \int_S g(x, t) dF(x),$$

где t — параметр, а S — данное множество конечной или бесконечной P -меры.

Нам понадобятся некоторые теоремы относительно непрерывности, а также о дифференцировании и интегрировании таких интегралов по параметру t . В частном случае, когда $F(x) = x$, эти теоремы сводятся к теоремам относительно интеграла Лебега.

Предположим, что функция $g(x, t)$ принимает комплексные значения и что для любого фиксированного значения параметра t действительная и мнимая части этой функции суть E -измеримые функции от x , интегрируемые на множестве S относительно $F(x)$. Через $G_1(x)$, $G_2(x)$, ... будем обозначать некоторые функции, интегрируемые на множестве S относительно $F(x)$.

1) **Непрерывность.** Если для почти всех (P) значений x из S функция $g(x, t)$ непрерывна в точке $t = t_0$ и если для всех t в некоторой окрестности точки t_0 имеем $|g(x, t)| < G_1(x)$, то $u(t)$ непрерывна в точке $t = t_0$, так что*)

$$(7.3.1) \quad \lim_{t \rightarrow t_0} \int_S g(x, t) dF(x) = \int_S g(x, t_0) dF(x).$$

Это непосредственно следует из соотношения (7.2.2). В самом деле, для любой последовательности значений t_1, t_2, \dots (из данной окрестности), стремящейся к t_0 , выполняются условия предложения (7.2.2), если взять $g_1(x) = g(x, t_1)$ и $g(x) = g(x, t_0)$. Таким образом, согласно (7.2.2), имеем $u(t_1) \rightarrow u(t_0)$, и, следовательно, то же соотношение выполнено и в том случае, когда t стремится к t_0 непрерывно.

*) Эта теорема справедлива и так же доказывается и в случае $t_0 = \pm \infty$.

Если условия предложения I) выполнены для *всех* t_0 в открытом интервале (a, b) , то $u(t)$ непрерывна во всем интервале.

II) Дифференцирование. Если для почти всех (P) значений x из S выполнены следующие условия:

1) существует частная производная $\frac{\partial g(x, t)}{\partial t}$,

2) $\left| \frac{g(x, t+h) - g(x, t)}{h} \right| < G_3(x)$ для $0 < |h| < h_0$,

где h_0 не зависит от x , то

$$(7.3.2) \quad u'(t) = \frac{d}{dt} \int_S g(x, t) dF(x) = \int_S \frac{\partial g(x, t)}{\partial t} dF(x).$$

Подобно предыдущему предложению, это непосредственно следует из (7.2.2). Для любой последовательности h_1, h_2, \dots , где $|h_v| < h_0$ и $h_v \rightarrow 0$, выполнены условия предложения (7.2.2), если взять

$$g_v(x) = \frac{g(x, t+h_v) - g(x, t)}{h_v} \text{ и } g(x) = \frac{\partial g(x, t)}{\partial t}$$

Таким образом,

$$\frac{u(t+h) - u(t)}{h} = \int_S \frac{g(x, t+h) - g(x, t)}{h} dF(x) \rightarrow \int_S \frac{\partial g(x, t)}{\partial t} dF(x),$$

так что производная $u'(t)$ существует и имеет значение, данное соотношением (7.3.2).

Заметим, что если частная производная $\frac{\partial g}{\partial t}$ существует и удовлетворяет условию $\left| \frac{\partial g(x, t)}{\partial t} \right| < G_3(x)$ для всех t в открытом интервале (a, b) , то из соотношения

$$g(x, t+h) - g(x, t) = h \left(\frac{\partial g}{\partial t} \right)_{t+h}, \quad (0 < h < 1),$$

вытекает, что (7.3.2) справедливо для всех t из (a, b) .

Заметим, что условие 2) предложения II) не выполнено, если мы, например, возьмем $F(x) = x$, $S = (-\infty, \infty)$ и

$$g(x, t) = \begin{cases} e^{t-x} & \text{для } x \geq t, \\ 0 & \text{для } x < t. \end{cases}$$

В этом случае имеем

$$u(t) = \int_{-\infty}^{\infty} g(x, t) dx = \int_t^{\infty} e^{t-x} dx = 1,$$

и (7.3.2) дает

$$u'(t) = \int_{-\infty}^{\infty} \frac{\partial g}{\partial t} dx = \int_t^{\infty} e^{t-x} dx = 1,$$

что, очевидно, неверно. Чтобы правильно вычислить $u'(t)$, надо учесть переменный нижний предел интеграла, тогда

$$u'(t) = \int_t^{\infty} e^{t-x} dx - 1 = 0.$$

III) Интегрирование. Если для почти всех (P) значений x из S функция $g(x, t)$ непрерывна по t на конечном открытом интервале (a, b) и удовлетворяет условию $|g(x, t)| < G_4(x)$ для всех t из (a, b) , то

$$(7.3.3) \quad \int_a^b u(t) dt = \int_a^b \left[\int_S g(x, t) dF(x) \right] dt = \int_S \left[\int_a^b g(x, t) dt \right] dF(x).$$

Далее, если сформулированные здесь условия выполняются для любого конечного интервала (a, b) и если, кроме того,

$$\int_{-\infty}^{\infty} |g(x, t)| dt < G_5(x), \text{ то } *)$$

$$(7.3.4) \quad \int_{-\infty}^{\infty} u(t) dt = \int_S \left[\int_{-\infty}^{\infty} g(x, t) dt \right] dF(x).$$

Рассмотрим сперва случай конечного интервала (a, b) . Для почти всех (P) значений x из S интеграл

$$h(x, t) = \int_a^t g(x, \tau) d\tau$$

имеет, согласно (5.1.4), частную производную $\frac{\partial h(x, t)}{\partial t} = g(x, t)$ для всех t из (a, b) , так что $\left| \frac{\partial h(x, t)}{\partial t} \right| < G_4(x)$. Далее, $|h(x, t)| < (b-a) G_4(x)$, следовательно, $h(x, t)$ интегрируема на S относительно $F(x)$. Записывая

$$v(t) = \int_S h(x, t) dF(x),$$

*) Необходимые изменения этих условий, в случае, когда мы хотим интегрировать $u(t)$ на (a, ∞) или $(-\infty, b)$, очевидны.

мы можем применить к $v(t)$ замечание к теореме II) и найти

$$v'(t) = \int_S g(x, t) dF(x) = u(t).$$

Согласно I), функция $u(t)$ непрерывна на (a, b) , так что разность

$$\Delta(t) = \int_a^t u(\tau) d\tau - v(t)$$

имеет производную $\Delta'(t) = u(t) - v'(t) = 0$. Для $t = a$ имеем $h(x, a) = 0$, $v(a) = 0$ и, значит, $\Delta(a) = 0$. Следовательно, $\Delta(t) = 0$ для $a \leq t \leq b$ и, в частности, $\Delta(b) = 0$, что тождественно (7.3.3).

Если выполняются условия второй части теоремы, то (7.3.3) имеет место для любого конечного интервала (a, b) , и мы получим

$$\begin{aligned} \int_a^b |u(t)| dt &\leq \int_a^b \left[\int_S |g(x, t)| dF(x) \right] dt = \int_S \left[\int_a^b |g(x, t)| dt \right] dF(x) \leq \\ &\leq \int_S G_5(x) dF(x). \end{aligned}$$

Таким образом, интеграл $\int_{-\infty}^{\infty} |u(t)| dt$ сходится. Если в соотношении (7.3.3) заставить a и b стремиться соответственно к $-\infty$ и $+\infty$, то левая часть этого соотношения будет стремиться к левой части соотношения (7.3.4). Применяя (7.2.2), находим, что при этом правая часть (7.3.3) стремится к правой части (7.3.4). Таким образом (7.3.4) доказано.

Теоремы, доказанные в этом параграфе, показывают, что при некоторых условиях предельный переход, дифференцирование и интегрирование можно производить под знаком интеграла.

7.4. Интегралы Лебега — Стильеса относительно распределения. Если $P(S)$ есть вероятностная функция распределения (см. параграф 6.6), то интеграл

$$(7.4.1) \quad \int_{R_1} g(x) dP = \int_{-\infty}^{\infty} g(x) dP = \int_{-\infty}^{\infty} g(x) dF$$

может быть конкретно, хотя и недостаточно точно, интерпретирован как взвешенное среднее значение функции $g(x)$ для всех значений x ,

причем веса составляются из масс dP или dF , сосредоточенных в окрестности каждой точки x . Сумма всех весов равна единице, так как

$$\int_{-\infty}^{\infty} dP = \int_{-\infty}^{\infty} dF = P(R_1) = 1.$$

Каждая ограниченная B -измеримая функция $g(x)$ интегрируема относительно P (или F) на интервале $(-\infty, +\infty)$.

Если распределение масс представляется, согласно (6.6.2), как сумма двух компонент, то интеграл (7.4.1) равен

$$\int_{-\infty}^{\infty} g(x) dF = c_1 \int_{-\infty}^{\infty} g(x) dF_1 + c_2 \int_{-\infty}^{\infty} g(x) dF_2,$$

где первое слагаемое правой части сводится, как показывает соотношение (7.1.8), к сумме по точкам сосредоточения массы в данном распределении.

Если для некоторого целого положительного v функция x^v интегрируема относительно $F(x)$ по $(-\infty, +\infty)$, то интеграл

$$\alpha_v = \int_{-\infty}^{\infty} x^v dF(x)$$

называется *моментом v-го порядка* или просто *v-м моментом* распределения, и мы говорим, что v -й момент *существует*. Тогда, как легко видеть, любой момент порядка $v' < v$ также существует.

Из элементарной механики известно, что первый момент α_1 есть абсцисса *центра тяжести* массы распределения, а второй момент α_2 представляет собой *момент инерии* массы относительно перпендикулярной оси, проходящей через точку $x=0$. Моменты распределения будут играть большую роль в приложениях, разбираемых далее в этой книге.

Если для некоторого $k > 0$ функция распределения $F(x)$ удовлетворяет условиям (относительно принятых обозначений см. параграф 12.1)

$$F(x) = O(|x|^{-k}), \text{ если } x \rightarrow -\infty, \\ 1 - F(x) = O(x^{-k}), \quad \text{если } x \rightarrow +\infty,$$

то любой момент порядка $v < k$ существует. Чтобы это доказать, достаточно, согласно параграфу 7.2, установить, что интеграл функции $|x|^v$ относительно $F(x)$ по интервалу (a, b) меньше некоторой постоянной, не зависящей от

a и *b*. По предположению,

$$\int_{2^r-1}^{2^r} |x|^v dF(x) \leq 2^r (F(2^r) - F(2^{r-1})) \leq 2^r (1 - F(2^{r-1})) < \frac{C}{2^{r(k-v)}},$$

где *C* не зависит от *r*. Аналогичное соотношение имеет место и для интеграла по $(-2^r, -2^{r-1})$. Суммируя по $r = 1, 2, \dots$ и присоединяя интеграл по $(-1, 1)$, который не превосходит 1, получим для любого интервала (a, b)

$$\int_a^b |x|^v dF(x) < 1 + \frac{2C}{2^{k-v} - 1},$$

и, таким образом, *v*-й момент существует.

7.5. Интеграл Римана — Стильтьеса. Рассмотрим интеграл Лебега — Стильтьеса

$$(7.5.1) \quad \int_I g(x) dF(x)$$

в частном случае, когда *I* есть конечный полуоткрытый интервал

$$I = (a < x \leq b),$$

а *g(x)* непрерывна на *I* и стремится к конечному пределу при $x \rightarrow a+$.

Разобьем интервал *I* на *n* подинтервалов $i_v = (x_{v-1} < x \leq x_v)$ точками

$$a = x_0 < x_1 < \dots < x_n = b$$

и рассмотрим суммы Дарбу (7.1.1), соответствующие разбиению $I = i_1 + \dots + i_n$. Получаем

$$(7.5.2) \quad Z = \sum_1^n M_v [F(x_v) - F(x_{v-1})],$$
$$z = \sum_1^n m_v [F(x_v) - F(x_{v-1})],$$

где *m_v* и *M_v* суть нижняя и верхняя границы функции *g(x)* на *i_v*. Пусть теперь $\epsilon > 0$ задано. По предположению, существует δ , такое, что $M_v - m_v < \epsilon$ при $x_v - x_{v-1} < \delta$ для всех *v*, и мы имеем

$$Z - z < \epsilon [F(b) - F(a)].$$

Таким образом, если n стремится к бесконечности и в то же время максимум длин подинтервалов i_v стремится к нулю, то Z и z стремятся к общему пределу, который должен быть равен интегралу (7.5.1):

$$(7.5.3) \quad \lim_{n \rightarrow \infty} Z = \lim_{n \rightarrow \infty} z = \int_a^b g(x) dF(x).$$

Таким образом, в рассматриваемом здесь частном случае простые выражения (7.5.2) сумм Дарбу достаточны для определения значения интеграла Лебега—Стильтьеса. Если положить $F(x) = x$, то эти выражения эквивалентны суммам Дарбу, рассматриваемым в теории обычного интеграла Римана. Соответственно интеграл, определяемый соотношением (7.5.3), называется *интегралом Римана — Стильтьеса*. Из изложенного выше следует, что если этот интеграл существует, то он имеет всегда то же значение, что и соответствующий интеграл Лебега — Стильтьеса.

Если в каждом подинтервале i_v выбрать произвольную точку ξ_v , то, очевидно, имеем

$$(7.5.4) \quad \lim_{n \rightarrow \infty} \sum_1^n g(\xi_v) [F(x_v) - F(x_{v-1})] = \int_a^b g(x) dF(x),$$

так как сумма в левой части заключена между z и Z .

Интеграл Римана—Стильтьеса (7.5.3) существует даже в более общем случае, когда $g(x)$ ограничена на (a, b) и имеет не больше конечного числа точек разрыва r_v , при условии, что функция $F(x)$ непрерывна во всех точках r_v . Действительно, в этом случае мы можем заключить точки r_v в подинтервалы i_v , которые составят сколь угодно малую часть сумм z и Z .

В частном случае, когда $F(x)$ непрерывна на (a, b) и имеет непрерывную производную $F'(x)$ всюду, за исключением, может быть, конечного числа точек, имеем для каждого интервала i_v , не содержащего этих точек,

$$F(x_v) - F(x_{v-1}) = (x_v - x_{v-1}) F'(\xi_v),$$

где ξ_v — некоторая точка, принадлежащая i_v . Из (7.5.4) следует, что в этом случае интеграл (7.5.3) сводится к обычному интегралу Римана

$$(7.5.5) \quad \int_a^b g(x) dF(x) = \int_a^b g(x) F'(x) dx.$$

Все эти свойства непосредственно распространяются на случай функции $g(x)$, принимающей комплексные значения, а также на случай бесконечных интервалов (a, b) при условии, что функция $g(x)$ интегрируема на (a, b) относительно $F(x)$. Если это условие выполнено, то имеет место, например, следующее обобщение (7.5.4):

$$(7.5.6) \quad \lim_{n \rightarrow \infty} \sum_{v=1}^n g(\xi_v) [F(x_v) - F(x_{v-1})] = \int_{-\infty}^{\infty} g(x) dF(x),$$

где, как и раньше, максимум длин подинтервалов (x_{v-1}, x_v) стремится к нулю при $n \rightarrow \infty$ и в то же время $x_0 \rightarrow -\infty$ и $x_n \rightarrow +\infty$.

Предположим, что даны две неубывающие функции $F(x)$ и $G(x)$, непрерывные на замкнутом интервале (a, b) , за исключением, может быть, конечного числа точек разрыва, являющихся внутренними точками интервала (a, b) . Предположим, далее, что ни одна точка интервала (a, b) не является точкой разрыва *обеих* функций F и G . Выбирая подинтервалы так, чтобы ни одно x_v не являлось точкой разрыва, имеем

$$F(b)G(b) - F(a)G(a) = \sum_{v=1}^n [F(x_v)G(x_v) - F(x_{v-1})G(x_{v-1})] = \\ = \sum_{v=1}^n F(x_v)[G(x_v) - G(x_{v-1})] + \sum_{v=1}^n G(x_{v-1})[F(x_v) - F(x_{v-1})].$$

Два слагаемых в правой части заключены соответственно между нижними и верхними суммами Дарбу интегралов $\int FdG$ и $\int GdF$. Переходя к пределу, мы получим, таким образом, *формулу интегрирования по частям*:

$$(7.5.7) \quad \int_a^b d(FG) = \int_a^b FaG + \int_a^b GdF.$$

Рассмотрим, наконец, последовательность распределения $F_1(x), F_2(x), \dots$, сходящуюся к неубывающей функции $F(x)$ в каждой точке непрерывности последней. (Согласно параграфу 6.7, предел $F(x)$ неизбежно является функцией распределения.) Пусть функция $g(x)$ всюду непрерывна. При любом конечном интервале (a, b) , таком, что a и b суть точки непрерывности функций $F(x)$, рассмотрение сумм Дарбу, определяющих соответствующие интегралы, показывает, что

$$(7.5.8) \quad \lim_{n \rightarrow \infty} \int_a^b g(x) dF_n(x) = \int_a^b g(x) dF(x).$$

Предположим, далее, что для каждого $\epsilon > 0$ можно найти такое A , что

$$\int_{-\infty}^{-A} |g(x)| dF_n(x) + \int_A^{\infty} |g(x)| dF_n(x) < \epsilon$$

для $n = 1, 2, \dots$. Всегда можно выбрать A так, чтобы функция $F(x)$ была непрерывна в точке $x = A$; тогда, согласно (7.5.8),

$$\int_A^B |g(x)| dF_n(x) \rightarrow \int_A^B |g(x)| dF(x),$$

где $B \geq A$ — другая точка непрерывности функции $F(x)$. Таким образом, последний интеграл не превосходит ϵ для любого $B > A$, и для интеграла по $(-B, -A)$ имеет место такое же соотношение. Следовательно, $g(x)$ интегрируема на $(-B, -A)$ относительно $F(x)$. Если в (7.5.8) взять $a = -A$ и $b = +A$, то каждый интеграл отличается от соответствующего ему интеграла на $(-\infty, +\infty)$ не больше чем на 2ϵ . Так как ϵ произвольно, то

$$(7.5.9) \quad \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} g(x) dF_n(x) = \int_{-\infty}^{\infty} g(x) dF(x).$$

Это соотношение непосредственно распространяется и на функции $g(x)$, принимающие комплексные значения.

Литература к главам 4—7. Классическая теория интегрирования получила свое законченное развитие в знаменитой работе Римана (1854). Около 1900 г. Борелем и Лебегом была основана теория меры точечных множеств и последним было введено понятие интеграла, носящего его имя. Интеграл относительно неубывающей функции $F(x)$ был рассмотрен уже в 1894 г. Стильтьесом и в 1913 г. Радоном [205], исследовавшим свойства аддитивных функций множества и теорию интегрирования относительно таких функций.

Существует большое число работ, посвященных современной теории интегрирования. Читатель, в частности, отсылается к книгам самого Лебега [23], Валле-Пуссена [40] и Сакса [33]. Книга Валле-Пуссена дает прекрасное введение в теорию интеграла Лебега и содержит также несколько глав относительно аддитивных функций множества, в то время как другие две книги углубляются в более трудные части теории.

ТЕОРИЯ МЕРЫ И ИНТЕГРИРОВАНИЯ В R_n

ГЛАВА 8

МЕРА ЛЕБЕГА И ДРУГИЕ АДДИТИВНЫЕ ФУНКЦИИ МНОЖЕСТВА в R_n

8.1. Мера Лебега в R_n . Элементарная мера протяженности одномерного интервала есть его *длина*. Соответствующая мера для двумерного интервала (см. параграф 3.1) есть его *площадь*, а для трехмерного интервала — *объем*.

Вообще, если i есть n -мерный интервал, определенный неравенствами

$$a_v \leqslant x_v \leqslant b_v, \quad (v = 1, 2, \dots, n),$$

то его n -мерным объемом называется неотрицательная величина

$$L(i) = \prod_1^n (b_v - a_v).$$

Для открытого, полуоткрытого и замкнутого интервалов с общими границами a_v и b_v объем имеет одно и то же значение. Объем вырожденного интервала всегда равен нулю. Для бесконечного невырожденного интервала положим $L(i) = +\infty$.

Лемма Бореля (см. параграф 4.1) непосредственно распространяется на случай n измерений. Слегка обобщая доказательство (4.1.1), находим, что $L(i)$ есть *аддитивная функция интервала*.

Так же, как и в параграфе 4.2, можно поставить вопрос, может ли мера с теми же основными свойствами, что и $L(i)$, быть определена для более общего класса множеств, чем интервалы. Мы, таким образом, хотим найти неотрицательную аддитивную функцию множества $L(S)$, определенную для всех борелевских множеств в R_n и принимающую значение $L(i)$ в случае, когда S есть интервал i . В параграфах 4.3—4.7 мы детально исследовали эту проблему в случае $n = 1$ и показали, что она имеет единственное решение: решением служит мера Лебега в R_1 . Случай произвольного n не требует никаких изменений. Каждое утверждение и каждая формула параграфов 4.3—4.7 останутся

справедливыми, если заменить везде линейные множества n -мерными, а длину интервала — n -мерным объемом.

Итак, существует неотрицательная аддитивная функция множества $L(S)$, однозначно определенная для всех борелевских множеств в R_n и равная в частном случае, когда S есть интервал, его n -мерному объему. $L(S)$ называется n -мерной мерой Лебега *) множества S .

8.2. Неотрицательные аддитивные функции множества в R_n

Точно так же, как и в одномерном случае, мы можем и для $n > 1$ рассматривать неотрицательные аддитивные функции множества $P(S)$ более общего вида, чем n -мерная мера Лебега $L(S)$.

Будем рассматривать функции множества $P(S)$, определенные для всех борелевских множеств S в R_n и удовлетворяющие условиям А), В), С) параграфа 6.2. Непосредственно видно, что эти условия не содержат ссылок на число измерений. Соотношения (6.2.1) — (6.2.3) выполняются, очевидно, при любом числе измерений.

С каждой функцией множества $P(S)$ этого типа можно связать функцию точки $F(x) = F(x_1, \dots, x_n)$ подобно тому, как это делается для одномерного случая посредством равенства (6.2.4). Однако непосредственное обобщение (6.2.4) на случай произвольного n несколько затруднительно. Поэтому мы рассмотрим только случай ограниченной $P(S)$, когда определение соответствующей функции точки может быть упрощено подобно тому, как в одномерном случае это достигается посредством (6.5.1).

Этим мы будем заниматься в следующем параграфе.

Как и в случае $n = 1$, любая неотрицательная аддитивная функция множества $P(S)$ в R_n определяет некоторую n -мерную P -меру множества S , которая служит обобщением n -мерной меры Лебега $L(S)$. Замечания параграфа 6.4 относительно множеств P -меры нуль применимы в случае любого числа измерений.

*) Чтобы быть вполне точным, нужно было бы принять обозначение, явно указывающее на число измерений, например $L_n(S)$ вместо $L(S)$. Однако это не может привести к недоразумению, если все время помнить, что мера данного точечного множества относится к пространству, в котором оно содержится. Так, если мы рассмотрим, например, интервал $(0, 1)$ на прямой линии как множество точек в R_1 , то его (одномерная) мера равна 1. Если, с другой стороны, мы возьмем ось x на плоскости и рассмотрим тот же интервал как множество точек в R_2 , то мы имеем дело с вырожденным интервалом, (двумерная) мера которого равна нулю.

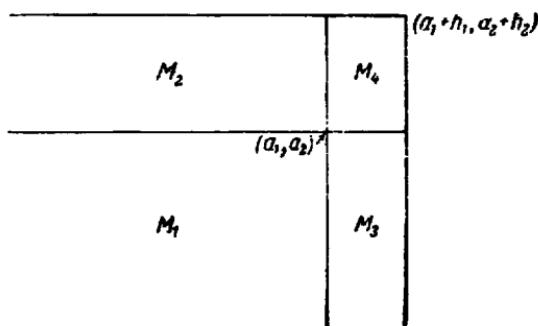
8.3. Ограниченные функции множества. Если значение $P(R_n)$ конечно, то мы говорим (см. параграф 6.5), что функция $P(S)$ *ограничена*. При этом всегда $P(S) \leq P(R_n)$. Обобщая (6.5.1), определим для ограниченной функции $P(S)$ функцию

$$(8.3.1) \quad F(x) = F(x_1, \dots, x_n) = P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n).$$

Очевидно, что по каждому переменному x_i функция $F(x)$ не убывает, везде непрерывна справа, и для всех x (см. (6.5.2))

$$0 \leq F(x) \leq P(R_n).$$

В одномерном случае значение $P(S)$ для полуоткрытого интервала I_1 ,



Фиг. 2. Функции множества и функции точки в R_2 .

определенного неравенством $a < x \leq a + h$, задается, согласно (6.2.5), разностью первого порядка функции $F(x)$:

$$P(I_1) = \Delta F(a) = F(a + h) - F(a).$$

Эта формула может быть обобщена на случай произвольного n . Рассмотрим сперва функцию множества $P(S)$ в R_2 и двумерный интервал I_2 , определенный неравенствами $a_1 < x_1 \leq a_1 + h_1$, $a_2 < x_2 \leq a_2 + h_2$. Имеем

$$(8.3.2) \quad P(I_2) = \Delta_2 F(a_1, a_2) = \\ = F(a_1 + h_1, a_2 + h_2) - F(a_1, a_2 + h_2) - F(a_1 + h_1, a_2) + F(a_1, a_2).$$

Это ясно из фиг. 2. Если M_1, M_2, M_3, M_4 суть значения, принимаемые функцией $P(S)$ в каждой из прямоугольных областей, указанных на чертеже, то из аддитивности $P(S)$ получаем

$$M_4 = (M_1 + M_2 + M_3 + M_4) - (M_1 + M_2) - (M_1 + M_3) + M_1,$$

что, согласно определению (8.3.1) функции $F(x)$, эквивалентно (8.3.2). Обобщение на случай произвольного n очевидно. Если $P(S)$ есть

функция множества в R_n , а i_n — полуоткрытый интервал, заданный неравенствами $a_v < x_v \leq a_v + h_v$, ($v = 1, 2, \dots, n$), то

$$(8.3.3) \quad P(i_n) = \Delta_n F(a_1, \dots, a_n) = F(a_1 + h_1, \dots, a_n + h_n) - F(a_1, a_2 + h_2, \dots, a_n + h_n) - \dots - F(a_1 + h_1, \dots, a_{n-1} + h_{n-1}, a_n) + \dots + (-1)^n F(a_1, \dots, a_n).$$

Любой ограниченной $P(S)$ в R_n соответствует, таким образом, функция точки $F(x_1, \dots, x_n)$, которая не убывает и непрерывна справа по любому переменному x_v и обладает свойством, что n -я разность, определяемая соотношением (8.3.3), всегда неотрицательна. Обратно, обобщая рассуждения параграфа 6.3, мы видим, что для каждой данной функции F , обладающей этими свойствами, однозначно определяется функция множества $P(S)$, удовлетворяющая условиям А), В), С) параграфа 6.2, которая для любого интервала i_n принимает значение, данное равенством (8.3.3).

Если одна из переменных, например x_v , стремится к $-\infty$, в то время как остальные остаются фиксированными, то, как и в параграфе 6.5, оказывается, что F стремится к нулю. Аналогично, если все x_v одновременно стремятся к $+\infty$, то F стремится к $P(R_n)$.

Если все аргументы функции F , за исключением одного, скажем x_v , стремятся к $+\infty$, то F стремится к пределу — к некоторой неубывающей ограниченной функции $F_v(x_v)$ переменной x_v . Согласно параграфу 6.2, функция $F_v(x_v)$ имеет не более чем счетное множество точек разрыва $z'_v; z''_v, \dots$. Будем рассматривать эти точки как *исключенные значения* переменной x_v , тем самым допуская для x_v лишь значения, отличные от z'_v, z''_v, \dots . Точно так же каждая из переменных x_1, \dots, x_n имеет свое собственное конечное или счетное множество исключенных значений. В *каждой неисключенной точке* $x = (x_1, \dots, x_n)$ (т. е. такой, в которой ни одно из значений x_v не является исключенным) функция F непрерывна. Это следует из неравенства

$$|F(x + h) - F(x)| \leq F(x + |h|) - F(x - |h|) \leq \sum_1^n [F_v(x_v + |h_v|) - F_v(x_v - |h_v|)],$$

где $h = (h_1, \dots, h_n)$ есть произвольная точка, $|h|$ обозначает точку $(|h_1|, \dots, |h_n|)$, а сумма $x + h$, разность $x - h$ и т. п. образуются

по правилам векторного сложения (см. параграфы 11.1—11.2). Фиг. 2 поможет уяснить это неравенство.

Всякий n -мерный интервал, такой, что ни одна из его границ a_i и b_i не является исключенным значением соответствующей переменной x_i , называется *интервалом непрерывности* функции $P(S)$. Значения, принимаемые функцией $P(S)$, когда S есть интервал непрерывности, будут меняться, очевидно, непрерывным образом при небольших изменениях a_i и b_i . Если две ограниченные функции множества в \mathbb{R}_n совпадают для всех интервалов, являющихся интервалами непрерывности для них обеих, то (см. параграф 6.7) эти функции совпадают.

8.4. Распределения. Неотрицательные аддитивные функции $P(S)$, такие, что $P(\mathbb{R}_n) = 1$, играют, подобно соответствующим одномерным функциям (см. параграф 6.6), фундаментальную роль в приложениях. Согласно предыдущему параграфу, функция точки $F(x)$, связанная с функцией множества $P(S)$ этого класса, удовлетворяет соотношениям

$$(8.4.1) \quad \begin{aligned} F(x) &= F(x_1, \dots, x_n) = P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n), \\ 0 &\leq F(x) \leq 1, \quad \Delta_n F \geq 0, \\ F(-\infty, x_2, \dots, x_n) &= \dots = F(x_1, \dots, x_{n-1}, -\infty) = 0, \\ F(+\infty, \dots, +\infty) &= 1. \end{aligned}$$

Как и в одномерном случае, функции $P(S)$ и $F(x)$ будут интерпретироваться посредством *распределения единицы массы* по пространству \mathbb{R}_n , при котором каждое борелевское множество S несет массу $P(S)$. Как в параграфе 6.6, мы можем задавать распределение как функцией множества $P(S)$, так и соответствующей функцией точки $F(x)$, представляющей количество массы, заключенной в бесконечном интервале $\xi_1 \leq x_1, \dots, \xi_n \leq x_n$. Разница между этими двумя способами определения, конечно, только формальная, и в каждом данном случае мы употребляем тот или другой из них в зависимости от удобства. Как в параграфе 6.6, $P(S)$ будем называть *вероятностной функцией распределения*, а $F(x)$ — *функцией распределения*. Таким образом, функция распределения $F(x) = F(x_1, \dots, x_n)$ не убывает и всюду непрерывна справа по каждой переменной x_i и обладает свойством, что n -я разность, определенная соотношением (8.3.3), всегда неотрицательна. Обратно, из предыдущего параграфа следует, что любая заданная функция F , обладающая этими свойствами, есть функция распределения некоторого однозначно определенного распределения в \mathbb{R}_n .

Если множество, состоящее из единственной точки $x = a$ несет положительное количество массы, то a называется *точкой сосредоточения массы* в данном распределении. Множество всех точек сосредоточения массы не более чем счетно, как это вытекает из простого обобщения соответствующего доказательства в параграфе 6.2. Очевидно, что каждая точка сосредоточения массы a есть точка разрыва функции распределения F . В случае $n = 1$ мы видели (в параграфе 6.6), что, обратно, функция F непрерывна во всех точках x , за исключением точек сосредоточения массы. *Это, вообще говоря, неверно при $n > 1$.* Действительно, в многомерном пространстве масса может быть сосредоточена на линиях, поверхностях и гиперповерхностях так, что точек, несущих положительное количество массы, может не существовать, в то время как функция F может быть разрывна в некоторых точках. В предыдущем параграфе мы видели, однако, что возможно исключить некоторые значения каждой переменной x_i , так что функция F будет непрерывной во всех „исключенных“ точках.

Рассмотрим, например, распределение единицы массы с равномерной плотностью по интервалу $(0, 1)$ оси x_2 в плоскости (x_1, x_2) . Очевидно, что это распределение не имеет точек сосредоточения массы, и тем не менее соответствующая функция распределения $F(x_1, x_2)$ разрывна в каждой точке $(0, x_2)$, где $x_2 > 0$. Соответственно, функция $F_1(x_1) = \lim_{x_2 \rightarrow +\infty} F(x_1, x_2)$, рассмотренная в предыдущем параграфе, разрывна при $x_1 = 0$, которое является единственным „исключенным“ значением x_1 . Для x_2 исключенных значений нет, и, таким образом, $F(x_1, x_2)$ непрерывна в любой точке (x_1, x_2) , где $x_1 \neq 0$.

Мы увидим далее, что каждое распределение в R_n может быть единственным образом представлено в виде (6.6.2) как сумма двух компонент, первая из которых соответствует распределению, при котором вся масса заключена в точках сосредоточения массы, в то время как вторая компонента соответствует распределению без точек сосредоточения массы. Из сказанного выше следует, что при $n > 1$ мы не можем утверждать, что функция распределения второй компоненты непрерывна.

Пусть I обозначает n -мерный интервал

$$x_y - h_y < \xi_y \leq x_y + h_y \quad (y = 1, 2, \dots).$$

Отношение

$$\frac{P(I)}{L(I)} = \frac{\Delta_n F}{2^n h_1 h_2 \dots h_n},$$

где разность $\Delta_n F$ определяется соотношением (8.3.3), представляет собой среднюю плотность массы на интервале I . Если существует частная производная

$$f(x_1, \dots, x_n) = \frac{\partial^n F}{\partial x_1 \partial x_2 \dots \partial x_n},$$

то средняя плотность стремится к этому значению, если все h_i стремятся к нулю, и соответственно $f(x_1, \dots, x_n)$ представляет *плотность массы в точке x* . Как и в одномерном случае, эта функция будет называться *плотностью вероятности или функцией плотности*.

Пусть $F(x_1, \dots, x_n)$ — функция распределения для некоторого данного распределения. Если все переменные, за исключением фиксированного x_v , стремятся к $+\infty$, то F будет стремиться (см. параграф 8.3) к пределу $F_v(x_v)$, который является некоторой функцией распределения переменной x_v ; так, например, $F_1(x_1) = F(x_1, +\infty, \dots, +\infty)$. Функция $F_v(x_v)$ определяет некоторое одномерное распределение, которое мы будем называть *частным распределением* переменной x_v . Мы можем конкретно представить себе это частное распределение так: предположим, что каждая частица массы в первоначальном n -мерном распределении движется к оси x_v по прямой перпендикулярной оси x_v , до тех пор, пока не осядет на эту ось. Когда, наконец, вся масса спроектируется таким образом на ось x_v , то распределение, порожденное на этой оси, и есть частное распределение x_v . Каждая переменная x_v имеет, конечно, свое собственное частное распределение, которое может быть отлично от частных распределений других переменных.

Выделим теперь какую-нибудь группу из $k < n$ переменных, скажем x_1, \dots, x_k , и предположим, что остальные $n - k$ переменных стремятся к $+\infty$. Тогда F будет стремиться к некоторой функции распределения переменных x_1, \dots, x_k , которая определяет *k -мерное частное распределение* этой группы переменных. Это распределение можно конкретно представить с помощью проектирования массы первоначального распределения в k -мерное пространство переменных x_1, \dots, x_k (см. параграф 3.5). Пусть P — вероятностная функция n -мерного распределения, а $P_{1, \dots, k}$ — вероятностная функция частного распределения переменных x_1, \dots, x_k . Пусть, далее, S' — произвольное множество в k -мерном пространстве переменных x_1, \dots, x_k , а S есть цилиндрическое множество (см. параграф 3.5) всех точек x в R_n , которые проектируются в точки подпространства, принадлежащие S' . Тогда имеем,

очевидно,

$$(8.4.2) \quad P_1, \dots, k(S') = P(S),$$

что является аналитическим выражением проекции массы первоначального n -мерного распределения в k -мерное подпространство переменных x_1, \dots, x_k .

Теория распределений в R_n будет далее развита в главах 21—24.

8.5. Последовательности распределений. Как и в одномерном случае (см. параграф 6.7), мы будем говорить, что последовательность распределений в R_n *сходится*, если соответствующие вероятностные функции сходятся к неотрицательной аддитивной функции множества $P(S)$ на каждом интервале непрерывности этой последней. Если, кроме того, предел $P(S)$ есть вероятностная функция, т. е. если $P(R_n) = 1$, то мы говорим, что *последовательность сходится к распределению*. С точки зрения приложений важен, вообще говоря, только этот последний случай сходимости.

Для последовательности сходящейся, но не сходящейся к распределению, имеем $P(R_n) < 1$, что можно интерпретировать, говоря, что некоторая часть массы в наших распределениях „исчезает в бесконечности“ при переходе к пределу.

Простое обобщение результатов параграфа 6.7 показывает, что последовательность распределений сходится к распределению тогда и только тогда, когда соответствующие функции распределения F_1, F_2, \dots стремятся к функции распределения F во всех „не исключенных“ точках последней (см. параграф 8.3). Дальнейший критерий для решения вопроса, сходится ли данная последовательность распределений к некоторому распределению или нет, будет дан в параграфе 10.7.

Как и в параграфе 6.8, будем говорить, что последовательность функций распределения F_1, F_2, \dots сходится, если существует такая функция F , неубывающая по каждому переменному x , что $F_n \rightarrow F$ в каждой „не исключенной“ точке функции F . Всегда имеем $0 \leq P \leq 1$, но, соответственно сказанному выше, F не обязательно представляет собой функцию распределения. Далее, имеет место следующее обобщение предложения, доказанного в параграфе 6.8 для одномерного случая. *Каждая последовательность функций распределения содержит сходящуюся подпоследовательность*. Это может быть доказано непосредственным обобщением доказательства параграфа 6.8, которое мы здесь проводить не будем.

8.6. Распределения в произведении пространств. Рассмотрим два пространства R_m и R_n с переменными точками $x = (x_1, \dots, x_m)$ и $y = (y_1, \dots, y_n)$. Предположим, что в каждом из этих пространств задано некоторое распределение и пусть F_1 и P_1 суть функция распределения и вероятностная функция распределения в R_m , а F_2 и P_2 имеют аналогичный смысл для распределений в R_n .

В *произведении пространств* $R_m \cdot R_n$ (см. параграф 3.5) размерности $m+n$ обозначим переменную точку через $z = (x, y) = (x_1, \dots, x_m, y_1, \dots, y_n)$. Если S_1 и S_2 суть соответственно множества в пространствах R_m и R_n , то обозначим через S прямоугольное множество (см. параграф 3.5) всех таких точек $z = (x, y)$ в произведении пространств, для которых $x \in S_1$ и $y \in S_2$.

Почти очевидно, что мы можем всегда найти бесконечное число распределений в произведении пространств, таких, что для каждого из них частные распределения (см. параграф 8.4), соответствующие подпространствам R_m и R_n , совпадают с двумя заданными распределениями в этих пространствах. Среди этих распределений отметим, в частности, одно, имеющее специальный интерес для приложений. Это распределение задается следующей теоремой:

Существует одно и только одно распределение в произведении пространств $R_m \cdot R_n$, такое, что

$$(8.6.1) \quad P(S) = P_1(S_1) P_2(S_2)$$

для всех прямоугольных множеств S , определенных соотношениями $x \in S_1$ и $y \in S_2$. Это распределение задается функцией распределения

$$(8.6.2) \quad F(z) = F_1(x) F_2(y)$$

для всех точек $z = (x, y)$.

Заметим сперва, что функция $F(z)$, определенная равенством (8.6.2), является, конечно, функцией распределения в $R_m \cdot R_n$, так как она обладает характеристическими свойствами функций распределения, сформулированными в параграфе 8.4. Рассмотрим теперь распределение, заданное функцией $F(z)$. Из (8.3.3) следует, что

$$P(I) = P_1(I_1) P_2(I_2)$$

для любого полуоткрытого интервала $I = (I_1, I_2)$, определенного неравенствами типа $a_i < x_i \leq b_i$, $c_j < y_j \leq d_j$. Каждое борелевское множество S , в R_m может быть получено из интервалов I_1 применением

операций сложения и вычитания. (Согласно (1.3.1), операция умножения может быть сведена к сложению и вычитанию.) Из аддитивности функции P_1 следует, что для любого прямоугольного множества вида $S = (S_1, I_2)$ имеем

$$P(S) = P_1(S_1)P_2(I_2).$$

Поступая таким же образом с интервалами I_2 , мы, наконец, получим (8.6.1). С другой стороны, каждое распределение, удовлетворяющее условию (8.6.1), удовлетворяет также условию (8.6.2), потому что последнее является просто частным случаем предыдущего. Так как распределение однозначно определяется своей функцией распределения, то, таким образом, может существовать только одно распределение, удовлетворяющее условию (8.6.1).

Если в (8.6.1) положить $S_2 = R_n$, то из (8.4.2) будет следовать, что частное распределение, соответствующее подпространству R_m , совпадает с заданным в этом пространстве распределением с вероятностной функцией P_1 . Аналогично, полагая $S_1 = R_m$, находим, что частное распределение в R_n совпадает с заданным в этом пространстве распределением.

Заметим, наконец, что теорема может быть обобщена на распределения в произведении любого числа пространств. Доказательство вполне аналогично проведенному выше, а соотношения (8.6.1) и (8.6.2) заменяются очевидными обобщениями:

$$P = P_1 P_2 \dots P_k \text{ и } F = F_1 F_2 \dots F_k.$$

ГЛАВА 9

ИНТЕГРАЛ ЛЕБЕГА — СТИЛЬСЕА ОТ ФУНКЦИЙ n ПЕРЕМЕННЫХ

9.1. Интеграл Лебега — Стильсеса. Теория интеграла Лебега — Стильсеса от функций одного переменного, развитая в главе 7, может быть непосредственно обобщена на функции n переменных. Если в выражениях (7.1.1) для сумм Дарбу допустить, что $P(S)$ обозначает неотрицательную аддитивную функцию множества в R_n , а m , и M , суть нижняя и верхняя границы данной функции $g(x) = g(x_1, \dots, x_n)$ на n -мерном множестве S , то *интеграл Лебега — Стильсеса*

$$(9.1.1) \quad \int_S g(x) dP = \int_S g(x_1, \dots, x_n) dP$$

оказывается определенным точно так же, как и в одномерном случае.

Функция $g(x)$ называется *B*-измеримой на множестве S , если при любом действительном k подмножество всех точек $x \in S$ таких, что $g(x) \leq k$, есть борелевское множество. Все замечания относительно *B*-измеримых функций, сделанные в параграфе 5.2, без труда распространяются на функции n переменных.

Если $g(x)$ ограничена и *B*-измерима на множестве S конечной P -меры, то она интегрируема на множестве S относительно P . Определения интеграла и интегрируемости в случае неограниченной функции $g(x)$ и в случае множества S бесконечной P -меры получаются непосредственным обобщением определений параграфа 7.2. Все свойства интеграла, упомянутые в параграфах 7.1—7.3, легко распространяются на случай n переменных, все доказательства полностью аналогичны доказательствам для случая $n=1$.

В частном случае, когда $P(S)$ есть n -мерная лебегова мера, получим *интеграл Лебега* функции $g(x)$, который часто записывается в форме обычного кратного интеграла

$$\int_S g(x) dL = \int_S g(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Если S есть интервал, а $g(x)$ интегрируема в смысле Римана на этом интервале, то интеграл Лебега совпадает с обычным кратным интегралом Римана, как мы это уже отмечали для одномерного случая (см. параграф 5.1).

9.2. Интегралы Лебега — Стильеса относительно распределения. Замечания, сделанные на эту тему в параграфе 7.4, очевидно, применимы также и в случае $n > 1$.

Моменты распределения в R_n суть интегралы

$$a_{v_1, \dots, v_n} = \int_{R_n} x_1^{v_1} \dots x_n^{v_n} dP,$$

где v_i — неотрицательные целые числа. Как и в одномерном случае, мы будем говорить, что соответствующий момент *существует*, если функция $x_1^{v_1} \dots x_n^{v_n}$ интегрируема на R_n относительно P .

Рассмотрим теперь интеграл

$$(9.2.1) \quad \int_{R_n} g(x_1, \dots, x_n) dP$$

в случае, когда функция g зависит только от некоторого числа переменных, скажем x_1, \dots, x_k , где $k < n$. Обозначим через R_k k -мерное подпространство этих переменных. Предположим сперва, что функция g ограничена, и рассмотрим разбиения

$$R_k = S'_1 + \dots + S'_q,$$

$$R_n = S_1 + \dots + S_q,$$

где S'_v суть борелевские множества в R_k , такие, что $S'_\mu S'_v = 0$ при $\mu \neq v$, а S_v обозначает цилиндрическое множество (см. параграф 3.5) в R_n с основанием S'_v .

Верхняя сумма Дарбу

$$Z = M_1 P(S_1) + \dots + M_q P(S_q),$$

соответствующая интегралу (9.2.1), совпадает, согласно (8.4.2), с суммой

$$Z = M_1 P_{1, \dots, k}(S'_1) + \dots + M_q P_{1, \dots, k}(S'_q)$$

(где $P_{1, \dots, k}$ обозначает вероятностную функцию частного распределения переменных x_1, \dots, x_k), которая является верхней суммой Дарбу для k -мерного интеграла

$$\int_{R_k} g dP_{1, \dots, k}.$$

Так как то же соотношение справедливо и для нижних сумм Дарбу, то, следовательно, для любой ограниченной функции $g(x_1, \dots, x_k)$ имеем

$$(9.2.2) \quad \int_{R_n} g(x_1, \dots, x_k) dP = \int_{R_k} g(x_1, \dots, x_k) dP_{1, \dots, k},$$

так что в этом случае n -мерный интеграл сводится к k -мерному.

Легко видеть, что это соотношение выполняется, когда g интегрируема на R_k относительно $P_{1, \dots, k}$, даже если g не ограничена. Можно предполагать также, что функция g принимает комплексные значения.

9.3. Теорема о повторном интегрировании. Если функция $g(x, y)$ непрерывна в прямоугольнике $a \leq x \leq b, c \leq y \leq d$, то

известно, что справедливы равенства

$$\int_a^b \left(\int_c^d g(x, y) dy \right) dx = \int_c^d \left(\int_a^b g(x, y) dx \right) dy,$$

так что двойной интеграл может быть выражен двумя способами в виде повторного интеграла. Соответствующая теорема справедлива и для случая интеграла Лебега — Стильтьеса любого числа измерений, мы сейчас докажем ее для одного специального случая.

Используя те же обозначения, что и в параграфе 8.6, рассмотрим две вероятностные функции P_1 и P_2 соответственно в пространствах R_m и R_n и однозначно определенную вероятностную функцию P в произведении пространств $R_m \cdot R_n$, удовлетворяющую условию (8.6.1). Пусть S_1 и S_2 обозначают множества, заданные соответственно в R_m и R_n , а $S = (S_1, S_2)$ — прямоугольное множество со „сторонами” S_1 и S_2 . Пусть, далее, $g(x)$ и $h(y)$ суть заданные функции точки соответственно в R_m и R_n , такие, что $g(x)$ интегрируема на S_1 относительно P_1 , а $h(y)$ интегрируема на S_2 относительно P_2 .

Тогда функция $g(x)h(y)$ интегрируема на $S = (S_1, S_2)$ относительно P и

$$(9.3.1) \quad \int_S g(x)h(y) dP = \int_{S_1} g(x) dP_1 \int_{S_2} h(y) dP_2.$$

Предположим сперва, что функции $g(x)$ и $h(y)$ ограничены и неотрицательны. Рассмотрим суммы Дарбу для интегралов в равенстве (9.3.1), соответствующие разбиениям $S_1 = S_1^{(1)} + \dots + S_1^{(n)}$, $S_2 = S_2^{(1)} + \dots + S_2^{(m)}$, $S = \sum_{i,j} S^{(i,j)}$, где $S^{(i,j)}$ обозначает прямоугольное множество $(S_1^{(i)}, S_2^{(j)})$. Если обозначить эти суммы через z и Z для интеграла в левой части и через z_1, Z_1 и z_2, Z_2 для двух интегралов в правой части, то имеем

$$z_1 z_2 \leq z \leq Z \leq Z_1 Z_2.$$

Теперь (9.3.1) вытекает непосредственно из определения интеграла. Заменив, далее, g и h на $g' - g''$ и $h' - h''$, где g' , g'' , h' , h'' суть ограниченные неотрицательные функции, мы установим (9.3.1) для любых действительных ограниченных функций g и h . Распространение на произвольные интегрируемые функции и на функции, принимающие комплексные значения, следует непосредственно из определения интегралов для таких функций.

9.4. Интеграл Римана — Стильтьеса. Содержание параграфа 7.5 может быть также обобщено на случай n переменных, причем мы должны употреблять функцию точки $F(x_1, \dots, x_n)$ и разность $\Delta_n F$ вместо функции точки $F(x)$ и разности $F(x_v) - F(x_{v-1})$.

В частности, если для всех точек интервала $I(a_v \leq x_v \leq b_v; v = 1, \dots, n)$ существует непрерывная производная $\frac{\partial^n F}{\partial x_1 \dots \partial x_n}$, а функция $g(x)$ непрерывна на I , то интеграл (9.1.1) при $S = I$ может быть представлен как кратный интеграл Римана:

$$\int_I g(x) dP = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} g(x_1, \dots, x_n) \frac{\partial^n F}{\partial x_1 \dots \partial x_n} dx_1 \dots dx_n.$$

Это свойство непосредственно распространяется на случай функции $g(x)$, принимающей комплексные значения, и на случай бесконечных интервалов, при условии, что $g(x)$ интегрируема на I относительно P .

9.5. Неравенство Шварца. Рассмотрим две действительные функции $g(x)$ и $h(x)$, такие, что их квадраты g^2 и h^2 интегрируемы относительно P на множестве S в R_n . Квадратичная форма

$$\int_S [ug(x) + vh(x)]^2 dP = u^2 \int_S g^2 dP + 2uv \int_S gh dP + v^2 \int_S h^2 dP$$

неотрицательна для всех действительных значений переменных u и v . Следовательно (см. параграф 11.10), детерминант этой формы неотрицателен, откуда вытекает

$$(9.5.1) \quad \left(\int_S gh dP \right)^2 \leq \int_S g^2 dP \cdot \int_S h^2 dP.$$

РАЗЛИЧНЫЕ ВОПРОСЫ

ГЛАВА 10

ИНТЕГРАЛЫ ФУРЬЕ

Для применений к теории вероятностей и статистике нам понадобится ряд теорем относительно некоторых специальных классов интегралов Фурье; мы выведем их в этой главе. Относительно общей теории этих вопросов см. Бахнер [4], Тичмарш [38] и Винер [41].

10.1. Характеристическая функция распределения в R_1 . Пусть $F(x)$ обозначает некоторую одномерную функцию распределения (см. параграф 6.6), а t — действительное число. Функция $g(x) = e^{itx} = \cos tx + i \sin tx$, согласно параграфу 7.4, интегрируема на $(-\infty, \infty)$ относительно $F(x)$, так как $|e^{itx}| = 1$. Функция действительного переменного t

$$(10.1.1) \quad \varphi(t) = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

называется *характеристической функцией* распределения, соответствующего $F(x)$.

Вообще говоря, $\varphi(t)$ есть функция переменного t , принимающая комплексные значения. Очевидно, всегда $\varphi(0) = 1$ и для всех значений t

$$\begin{aligned} |\varphi(t)| &\leq \int_{-\infty}^{\infty} dF(x) = 1, \\ \varphi(-t) &= \overline{\varphi(t)}, \end{aligned}$$

(через \bar{a} обозначаем комплексную величину, сопряженную с a). Далее, из параграфа 7.3 следует, что $\varphi(t)$ непрерывна для всех действительных t .

Если момент k -го порядка (см. параграф 7.4) данного распределения существует, то из параграфа 7.3 вытекает, что мы можем дифференцировать равенство (10.1.1) k раз по t , так что для $0 \leq v \leq k$ имеем

$$(10.1.2) \quad \varphi^{(v)}(t) = i^v \int_{-\infty}^{\infty} x^v e^{itx} dF(x).$$

Отсюда, согласно параграфу 7.3, следует, что функция $\varphi^{(v)}(t)$ непрерывна для всех действительных t , и мы имеем

$$\varphi^{(v)}(0) = i^v \int_{-\infty}^{\infty} x^v dF(x) = i^v a_v.$$

В окрестности точки $t=0$ имеем, таким образом, разложение в ряд Маклорена:

$$(10.1.3) \quad \varphi(t) = 1 + \sum_1^k \frac{a_v}{v!} (it)^v + o(t^k),$$

где остаточный член, деленный на t^k , стремится к нулю при $t \rightarrow 0$ (см. параграф 12.1).

Обратно, если известно, что характеристическая функция имеет для частного значения $t=0$ конечную производную порядка $2k$, то эта производная равна пределу

$$\varphi^{(2k)}(0) = \lim_{t \rightarrow 0} \int_{-\infty}^{\infty} \left(\frac{e^{itx} - e^{-itx}}{2t} \right)^{2k} dF(x) = (-1)^k \lim_{t \rightarrow 0} \int_{-\infty}^{\infty} \left(\frac{\sin tx}{t} \right)^{2k} dF(x).$$

Для любого конечного интервала (a, b) имеем, согласно (7.1.7),

$$\int_a^b x^{2k} dF(x) = \lim_{t \rightarrow 0} \int_a^b \left(\frac{\sin tx}{x} \right)^{2k} dF(x) \leq |\varphi^{(2k)}(0)|.$$

Следовательно, момент a_{2k} существует и соотношение (10.1.2) выполняется для $0 \leq v \leq 2k$ и всех значений t .

Мы видим, таким образом, что свойства дифференцируемости $\varphi(t)$ связаны с поведением функции $F(x)$ при больших значениях x , так как от этого поведения зависит, существуют или нет моменты a_v . Можно также показать, что, обратно, поведение функции $\varphi(t)$ в бесконечности связано с непрерывностью и свойствами дифференцируемости функции $F(x)$. Предположим, например, что функция $F(x) = f(x)$ существует для всех x , за исключением, быть может, конечного числа точек. Согласно (7.5.5), имеем

$$(10.1.4) \quad \varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx,$$

и можно показать, что $\varphi(t)$ стремится к нулю при $t \rightarrow \pm\infty$. Если, кроме того, n -я производная $f^{(n)}(x)$ существует для всех x , причем

$|f^{(n)}(x)|$ интегрируема на $(-\infty, \infty)$, то повторное интегрирование по частям показывает, что

$$|\varphi(t)| < \frac{K}{|t|^n}$$

для всех t , где K — некоторая постоянная. Здесь мы, однако, не будем проводить подробного доказательства этого утверждения.

Предположим, с другой стороны, что $F(x)$ есть ступенчатая функция со скачками высоты p_v в точках $x = x_v$. Тогда, согласно (7.1.8), имеем

$$(10.1.5) \quad \varphi(t) = \sum_v p_v e^{itx_v},$$

причем этот ряд сходится абсолютно и равномерно для всех t , так как $\sum_v p_v = 1$. Каждый член ряда есть периодическая функция от t и, следовательно, не стремится к нулю при $t \rightarrow \pm\infty$. Так, например, характеристическая функция функции распределения $\varepsilon(x)$, определенной соотношением (6.7.1), тождественно равна 1.

Не каждая функция $\varphi(t)$ может быть характеристической функцией некоторого распределения. Для этого необходимо, чтобы функция $\varphi(t)$ была непрерывной и удовлетворяла условиям: $|\varphi(t)| \leq 1$, $\varphi(0) = 1$ и $\varphi(-t) = \overline{\varphi(t)}$. Однако эти условия недостаточны для того, чтобы функция была характеристической функцией распределения. Если, например, $\varphi(t)$ вблизи $t = 0$ имеет вид $\varphi(t) = 1 + O(t^2 + \delta)$, где $\delta > 0$, то из (10.1.3) следует, что распределение, соответствующее функции $\varphi(t)$, должно иметь $a_1 = a_2 = 0$. Это означает (см. параграф 16.1), что вся масса данного распределения сосредоточена в точке $t = 0$, следовательно, соответствующая функция распределения есть $\varepsilon(x)$, а характеристическая функция $\varphi(t) = 1$. Итак, в этом случае $\varphi(t)$ не может быть характеристической функцией, если она не равна тождественно единице. Так

например, функции e^{-t^4} и $\frac{1}{1+t^4}$ не являются характеристическими функциями, хотя обе они удовлетворяют приведенным выше необходимым условиям.

Известны различные *необходимые и достаточные* условия того, чтобы данная функция была характеристической. Наиболее простым таким условием представляется следующее (Крамер [71]): *Для того чтобы данная ограниченная непрерывная функция $\varphi(t)$ была характеристической функцией некоторого распределения, необходимо и достаточно, чтобы $\varphi(0) = 1$ и чтобы функция*

$$\psi(x, A) = \int_0^A \int_0^A \varphi(t-u) e^{ix(t-u)} dt du$$

была действительна и неотрицательна для всех действительных x и всех $A > 0$.

Легко видеть, что эти условия необходимы. Действительно, если $\varphi(t)$ есть характеристическая функция, соответствующая функции распределения $F(x)$, то

$$\varphi(x, A) = 2 \int_{-\infty}^{\infty} \frac{1 - \cos A(x+y)}{(x+y)^2} dF(y),$$

а последнее выражение, очевидно, действительно и неотрицательно. Доказательство достаточности этих условий опирается на свойства некоторых интегралов, аналогичных тем, которые используются в двух следующих параграфах. Однако оно несколько сложнее, и мы его здесь приводить не будем.

10.2. Некоторые вспомогательные функции. Рассмотрим функции

$$s(h, T) = \frac{2}{\pi} \int_0^T \frac{\sin ht}{t} dt,$$

$$c(h, T) = \frac{2}{\pi} \int_0^T \frac{1 - \cos ht}{t^2} dt,$$

где h действительно, а $T > 0$. Очевидно, что $s(h, T) \geq 0$ и

$$s(-h, T) = -s(h, T), \quad c(-h, T) = c(h, T).$$

С помощью простых преобразований получим для $h > 0$

$$s(h, T) = \frac{2}{\pi} \int_0^{hT} \frac{\sin t}{t} dt,$$

$$c(h, T) = \frac{2h}{\pi} \int_0^{hT} \frac{\sin t}{t^2} dt - \frac{2}{\pi} \cdot \frac{1 - \cos hT}{T}.$$

В учебниках интегрального исчисления доказывается, что интеграл

$$\int_0^x \frac{\sin t}{t} dt$$

ограничен для всех $x > 0$ и имеет предел $\frac{\pi}{2}$ при $x \rightarrow \infty$.

Отсюда следует, что функция $s(h, T)$ ограничена при всех действительных h и всех $T > 0$ и что, равномерно для $|h| > \delta > 0$,

$$(10.2.1) \quad \lim_{T \rightarrow \infty} s(h, T) = \begin{cases} 1 & \text{для } h > 0, \\ 0 & \text{для } h = 0, \\ -1 & \text{для } h < 0. \end{cases}$$

Далее, для всех действительных h получаем

$$(10.2.2) \quad \lim_{T \rightarrow \infty} c(h, T) = \frac{2}{\pi} \int_0^\infty \frac{1 - \cos ht}{t^2} dt = |h|.$$

10.3. Теоремы единственности для характеристических функций в R_1 . Если $(a - h, a + h)$ есть интервал непрерывности (см. параграф 6.7) функции распределения $F(x)$, то

$$(10.3.1) \quad F(a + h) - F(a - h) = \lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \frac{\sin ht}{t} e^{-ita} \varphi(t) dt.$$

Эта важная теорема (Леви [24]) показывает, что распределение однозначно определяется своей характеристической функцией. Действительно, если два распределения имеют одну и ту же характеристическую функцию, то, согласно доказанной теореме, эти два распределения совпадают для любого интервала, являющегося для них обоих интервалом непрерывности. Следовательно, согласно параграфу 6.7, эти распределения тождественны.

Для того чтобы доказать эту теорему, пишем

$$J = \frac{1}{\pi} \int_{-T}^T \frac{\sin ht}{t} e^{-ita} \varphi(t) dt = \frac{1}{\pi} \int_{-T}^T \frac{\sin ht}{t} e^{-ita} dt \int_{-\infty}^{\infty} e^{itx} dF(x).$$

Модуль функции $\frac{\sin ht}{t} e^{ita}$ не больше, чем h , так что условия, установленные в параграфе 7.3 для изменения порядка интегрирования, выполнены. Следовательно,

$$\begin{aligned} J &= \frac{1}{\pi} \int_{-\infty}^{\infty} dF(x) \int_{-T}^T \frac{\sin ht}{t} e^{ita} dt = \\ &= \frac{2}{\pi} \cdot \int_{-\infty}^{\infty} dF(x) \int_0^T \frac{\sin ht}{t} \cos(x - a)t dt = \int_{-\infty}^{\infty} g(x, T) dF(x), \end{aligned}$$

где

$$\begin{aligned} g(x, T) &= \frac{2}{\pi} \int_0^T \frac{\sin ht}{t} \cos(x - a)t dt = \frac{1}{\pi} \int_0^T \frac{\sin(x - a + h)t}{t} dt - \\ &- \frac{1}{\pi} \int_0^T \frac{\sin(x - a - h)t}{t} dt = \frac{1}{2} s(x - a + h, T) - \frac{1}{2} s(x - a - h, T). \end{aligned}$$

Таким образом, согласно предшествующему параграфу, $|g(x, T)|$ меньше некоторой абсолютной постоянной, и мы имеем

$$\lim_{T \rightarrow \infty} g(x, T) = \begin{cases} 0 & \text{при } x < a - h, \\ \frac{1}{2} & \text{при } x = a - h, \\ 1 & \text{при } a - h < x < a + h, \\ \frac{1}{2} & \text{при } x = a + h, \\ 0 & \text{при } x > a + h. \end{cases}$$

Следовательно, можно применить теорему (7.2.2). Так как $F(x)$ непрерывна при $x = a \pm h$, то получим

$$\lim_{T \rightarrow \infty} J = \int_{a-h}^{a+h} dF(x) = F(a+h) - F(a-h),$$

так что (10.3.1) доказано.

В частном случае, когда $|\varphi(t)|$ интегрируема на $(-\infty, \infty)$, из (10.3.1) следует, что

$$\frac{F(x+h) - F(x-h)}{2h} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\sin ht}{ht} e^{-itx} \varphi(t) dt,$$

если F непрерывна в точках $x \pm h$. Если h стремится к нулю, то подинтегральная функция стремится к $e^{-itx} \varphi(t)$, в то время как ее модуль не превосходит интегрируемой функции $|\varphi(t)|$. Таким образом, применяя (7.3.1), мы убедимся, что производная $F'(x) = f(x)$ существует для всех x и что

$$(10.3.2) \quad f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt.$$

Тогда $f(x)$ есть плотность вероятности (см. параграф 6.6) данного распределения, и из параграфа 7.3 следует, что $f(x)$ непрерывна для всех значений x . Обращаем внимание на взаимную обратимость соотношений (10.3.2) и (10.1.4).

Для того чтобы определить $F(x)$ посредством соотношения (10.3.1), мы должны знать функцию $\varphi(t)$ на всем бесконечном интервале $(-\infty, \infty)$. Знание же функции $\varphi(t)$ на некотором конечном интервале недостаточно для определения функции $F(x)$.

В самом деле, Гнеденко [117] дал пример двух характеристических функций, совпадающих на конечном интервале, но не тождественных при всех t . Мы приведем несколько более простой пример, принадлежащий Хинчину. Функции

$$\varphi_1(t) = \begin{cases} 1 - |t| & \text{при } |t| \leq 1, \\ 0 & \text{при } |t| > 1, \end{cases}$$

$$\varphi_2(t) = \frac{1}{2} + \frac{4}{\pi^2} \left(\frac{\cos \pi t}{1^2} + \frac{\cos 3\pi t}{3^2} + \frac{\cos 5\pi t}{5^2} + \dots \right)$$

являются характеристическими. $\varphi_1(t)$ есть характеристическая функция распределения, определенного плотностью вероятности

$$f_1(x) = \frac{1 - \cos x}{\pi x^2},$$

то можно видеть, полагая в (10.3.3) $h = 1$, $F(x) = e(x)$ и $\varphi(t) = 1$. Функция $\varphi_2(t)$ соответствует распределению, при котором масса $\frac{1}{2}$ сосредоточена в точке $x = 0$ и массы $\frac{2}{n^2\pi^2}$ — в точках $x = n\pi$ ($n = \pm 1, \pm 3, \dots$). Суммируя тригонометрический ряд для $\varphi_2(t)$, убеждаемся, что $\varphi_1(t) = \varphi_2(t)$ при $|t| \leq 1$. С другой стороны, для $|t| > 1$ функция $\varphi_1(t)$ равна нулю, а функция $\varphi_2(t)$ — периодическая с периодом 2.

Теперь перейдем к доказательству формулы, которая тесно связана с (10.3.1), но отличается от нее тем, что она содержит абсолютно сходящийся интеграл. В следующем параграфе эта формула найдет важное применение.

Для любых действительных a и $h > 0$ имеем

$$(10.3.3) \quad \int_0^h [F(a+z) - F(a-z)] dz = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \cos ht}{t^2} e^{-itz} \varphi(t) dt.$$

Преобразуем интеграл в правой части так же, как и в доказательстве (10.3.1); при этом изменение порядка интегрирования законно в силу 7.3. Обозначая правую часть (10.3.3) через J_1 , получим

$$J_1 = \frac{1}{\pi} \int_{-\infty}^{\infty} dF(x) \int_{-\infty}^{\infty} \frac{1 - \cos ht}{t^2} e^{it(x-a)} dt =$$

$$= \frac{2}{\pi} \int_{-\infty}^{\infty} dF(x) \int_0^{\infty} \frac{1 - \cos ht}{t^2} \cos(x-a)t dt.$$

Из (10.2.2), как и выше, следует

$$\begin{aligned} J_1 &= \int_{-\infty}^{\infty} \frac{|x-a+h| + |x-a-h| - 2|x-a|}{2} dF(x) = \\ &= \int_{a-h}^{a+h} (h - |x-a|) dF(x). \end{aligned}$$

Применяя формулу интегрирования по частям (7.5.7) к последнему интегралу, взятому по каждому из интервалов $(a-h, a)$ и $(a, a+h)$ в отдельности, убеждаемся, наконец, что J_1 совпадает с выражением в левой части (10.3.3), так что соотношение (10.3.3) доказано.

10.4. Теорема непрерывности для характеристических функций в R_1 . В предыдущем параграфе мы видели, что существует взаимно-однозначное соответствие между распределениями и их характеристическими функциями $\varphi(t)$: функция $F(x)$ всегда однозначно определяется соответствующей характеристической функцией $\varphi(t)$, и преобразование, с помощью которого мы переходим от $F(x)$ к $\varphi(t)$ или обратно, всегда единственно. Докажем теперь теорему, которая показывает, что при некоторых условиях это преобразование также и *непрерывно*, так что соотношения $F_n(x) \rightarrow F(x)$ и $\varphi_n(t) \rightarrow \varphi(t)$ равносильны.

Эта теорема очень важна для приложений, так как она дает критерий, который часто позволяет решить, сходится ли данная последовательность распределений к распределению или нет. В параграфе 6.7 мы видели, что последовательность распределений сходится к распределению тогда и только тогда, когда соответствующая последовательность функций распределения сходится к функции распределения. Однако в приложениях иногда бывает очень трудно прямо обнаружить сходимость последовательности функций распределения, в то время как вопрос о сходимости соответствующей последовательности характеристических функций может быть решен сравнительно легко. В таких случаях мы будем часто пользоваться следующей теоремой, принадлежащей Леви [24, 25] и Крамеру [11]:

Дана последовательность распределений с функциями распределения $F_1(x), F_2(x), \dots$ и характеристическими функциями $\varphi_1(t), \varphi_2(t), \dots$. Для того чтобы последовательность $\{F_n(x)\}$ сходилась к функции распределения $F(x)$, необходимо и достаточно, чтобы при любом t последовательность $\{\varphi_n(t)\}$ сходилась к пределу $\varphi(t)$, непрерывному при $t = 0$.

Если это условие выполнено, то предел $\varphi(t)$ совпадает с характеристической функцией предельной функции распределения $F(x)$.

Покажем сперва, что это условие необходимо и что предел $\varphi(t)$ служит характеристической функцией функции распределения $F(x)$. Действительно, это непосредственно следует из (7.5.9), так как условия этого соотношения, очевидно, выполнены, если взять $g(x) = e^{itx}$.

Основная трудность заключается в доказательстве достаточности этого условия. Предположим, что $\varphi_n(t)$ стремится при каждом t к пределу $\varphi(t)$, который непрерывен при $t = 0$, и докажем, что при этом предположении $F_n(x)$ стремится к некоторой функции распределения $F(x)$. Если это будет доказано, то из первой части теоремы будет следовать, что предел $\varphi(t)$ совпадает с характеристической функцией функции распределения $F(x)$.

Согласно параграфу 6.8, последовательность $\{F_n(x)\}$ содержит подпоследовательность $\{F_{n_y}(x)\}$, сходящуюся к некоторой неубывающей функции $F(x)$, где $F(x)$ можно определить так, чтобы она была всюду непрерывна справа. Докажем сперва, что $F(x)$ есть функция распределения. Так как, очевидно, $0 \leq F(x) \leq 1$, то достаточно доказать, что $F(+\infty) - F(-\infty) = 1$. Из (10.3.3), полагая $a = 0$, получим

$$\int_0^h F_{n_y}(z) dz - \int_{-h}^0 F_{n_y}(z) dz = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \cos ht}{t^2} \varphi_{n_y}(t) dt.$$

В обеих частях этого равенства мы можем заставить h стремиться к бесконечности под знаками интегралов. Действительно, интегралы слева берутся по конечным интервалам, в которых F_{n_y} равномерно ограничены и стремятся почти всюду к F , так что мы можем применить (5.3.6). Справа модуль подинтегральной функции не превосходит функции

$$\frac{1 - \cos ht}{t^2},$$

интегрируемой на $(-\infty, \infty)$, так что мы можем применить более общую теорему (5.5.2). Таким образом, разделив на h , получим

$$\begin{aligned} \frac{1}{h} \int_0^h F(z) dz - \frac{1}{h} \int_{-h}^0 F(z) dz &= \frac{1}{\pi h} \int_{-\infty}^{\infty} \frac{1 - \cos ht}{t^2} \varphi(t) dt = \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \cos t}{t^2} \varphi\left(\frac{t}{h}\right) dt. \end{aligned}$$

В этом соотношении мы возьмем последовательность значений h , стремящуюся к бесконечности. Тогда первый член будет стремиться, очевидно, к $F(+\infty) - F(-\infty)$. С другой стороны, $\varphi(t)$ непрерывна при $t = 0$, так что $\varphi\left(\frac{t}{h}\right)$ стремится при любом t к пределу $\varphi(0)$. Имеем $\varphi(0) = \lim_{n \rightarrow \infty} \varphi_n(0)$, но $\varphi_n(0) = 1$ при любом n , так как $\varphi_n(t)$ есть характеристическая функция. Следовательно, $\varphi(0) = 1$.

Применяя еще раз (5.5.2), мы получим для последнего интеграла, используя (10.2.2),

$$F(+\infty) - F(-\infty) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \cos t}{t^2} dt = 1.$$

Таким образом, $F(+\infty) = 1$, $F(-\infty) = 0$, и предел $F(x)$ последовательности $\{F_n(x)\}$ есть функция распределения. Согласно же первой части доказательства, предел $\varphi(t)$ последовательности $\{\varphi_n(t)\}$ совпадает с характеристической функцией распределения $F(x)$.

Рассмотрим теперь какую-нибудь другую сходящуюся подпоследовательность последовательности $\{F_n(x)\}$ и обозначим предел этой новой подпоследовательности через $F^*(x)$; при этом предположим, что $F^*(x)$ всюду непрерывна справа. Так же, как и раньше, показываем, что $F^*(x)$ есть функция распределения. По предположению, характеристические функции новой подпоследовательности имеют для всех значений t тот же предел $\varphi(t)$, что и раньше; следовательно, $\varphi(t)$ есть характеристическая функция обеих функций распределения $F(x)$ и $F^*(x)$. Согласно теореме единственности (10.3.1), $F(x) = F^*(x)$ для всех x .

Итак, каждая сходящаяся подпоследовательность последовательности $\{F_n(x)\}$ имеет один и тот же предел $F(x)$. Это, однако, эквивалентно утверждению, что последовательность $\{F_n(x)\}$ сходится к $F(x)$, и так как мы показали, что $F(x)$ есть функция распределения, то наша теорема доказана.

В параграфе 10.1 мы видели, что характеристическая функция непрерывна при любом t . Таким образом, из доказанной выше теоремы следует, что предел $\varphi(t)$ последовательности характеристических функций непрерывен при любом t , если только он непрерывен при $t = 0$. Однако условие, что предел должен быть непрерывен при $t = 0$, существенно.

В самом деле, покажем на примере, что утверждение теоремы может быть неверно, если пренебречь этим условием. Пусть $F_n(x)$ — функция рас-

пределения, определенная равенствами

$$F_n(x) = \begin{cases} 0 & \text{при } x \leq -n, \\ \frac{x+n}{2n} & \text{при } -n < x < n, \\ 1 & \text{при } x \geq n. \end{cases}$$

Соответствующая плотность вероятности равна $\frac{1}{2n}$ на интервале $(-n, n)$ и исчезает вне этого интервала. Соответствующая характеристическая функция, согласно (10.1.4), есть

$$\varphi_n(t) = \frac{1}{2n} \int_{-n}^n e^{itx} dx = \frac{\sin nt}{nt}.$$

Если n стремится к бесконечности, то $\varphi_n(t)$ сходится при любом t к пределу $\varphi(t)$, определенному равенствами

$$\varphi(t) = \begin{cases} 1 & \text{при } t = 0, \\ 0 & \text{при } t \neq 0. \end{cases}$$

Таким образом, этот предел не непрерывен при $t = 0$. Соответственно этому при любом фиксированном x имеем $F_n(x) \rightarrow \frac{1}{2}$, так что предел последовательности $F_n(x)$ не является функцией распределения.

В случае $F_n(x) = \varepsilon(x - n)$, рассмотренном в параграфе 6.7, имеем $\varphi_n(t) = e^{int}$, так что последовательность характеристических функций никогда не сходится, за исключением t , кратных 2π . Соответственно этому, при любом фиксированном x имеем $F_n(x) \rightarrow 0$, так что предел последовательности $F_n(x)$ не является функцией распределения, как мы уже убедились в параграфе 6.7.

10.5. Некоторые интегралы. Выведем теперь некоторые формулы, которыми будем пользоваться в дальнейшем. В учебниках интегрального исчисления приводится интеграл

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

Подставляя $x \sqrt{\frac{h}{2}}$ вместо x , получим для любого $h > 0$

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}hx^2} dx = \sqrt{\frac{2\pi}{h}}.$$

На основании параграфа 7.3 легко видеть, что это равенство можно дифференцировать по h любое число раз, так что

$$(10.5.1) \quad \int_{-\infty}^{\infty} x^{2v} e^{-\frac{1}{2}hx^2} dx = \frac{(2v)!}{2^v v!} \sqrt{2\pi} h^{-v-1/2} \quad (v=0, 1, 2, \dots).$$

Рассмотрим теперь интеграл

$$\int_{-\infty}^{\infty} e^{itx - \frac{1}{2}hx^2} dx = \int_{-\infty}^{\infty} \sum_0^{\infty} \frac{(itx)^v}{v!} e^{-\frac{1}{2}hx^2} dx.$$

Частные суммы ряда под интегралом в правой части не превосходят функции $e^{|tx| - \frac{1}{2}hx^2}$, интегрируемой на $(-\infty, \infty)$. Таким образом, согласно (5.5.2), этот ряд можно почленно интегрировать. Так как члены нечетного порядка (относительно t), очевидно, исчезают, то получим

$$(10.5.2) \quad \int_{-\infty}^{\infty} e^{itx - \frac{1}{2}hx^2} dx = \sum_0^{\infty} \frac{(it)^v}{v!} \int_{-\infty}^{\infty} x^v e^{-\frac{1}{2}hx^2} dx = \\ = \sum_0^{\infty} \frac{(it)^{2v}}{(2v)!} \cdot \frac{(2v)!}{2^v v!} \sqrt{2\pi} h^{-v-1/2} = \sqrt{\frac{2\pi}{h}} e^{-\frac{t^2}{2h}}.$$

Полагая $h=1$ и вводя функцию

$$(10.5.3) \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt,$$

имеем, следовательно,

$$(10.5.4) \quad \int_{-\infty}^{\infty} e^{itx} d\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx - \frac{x^2}{2}} dx = e^{-\frac{t^2}{2}}.$$

Соотношение (10.5.3) показывает, что $\Phi(x)$ есть неубывающая и всюду непрерывная справа функция, такая, что $\Phi(-\infty)=0$ и $\Phi(+\infty)=1$. Таким образом, $\Phi(x)$ есть функция распределения, и из (10.5.4) видно, что соответствующая характеристическая функция есть $e^{-\frac{t^2}{2}}$. Распределение, определяемое функцией $\Phi(x)$, имеет важное значение и называется *нормальным распределением*. Оно будет рассмотрено

в главе 17. Повторным интегрированием по частям получим из (10.5.4)

$$(10.5.5) \quad \int_{-\infty}^{\infty} e^{itx} d\Phi^{(n)}(x) = (-it)^n e^{-\frac{t^2}{2}}.$$

Далее рассмотрим интеграл

$$(10.5.6) \quad \frac{1}{2} \int_{-\infty}^{\infty} e^{itx - |x|} dx = \int_0^{\infty} \cos tx e^{-x} dx = \\ = \left[\frac{t \sin tx - \cos tx}{1 + t^2} e^{-x} \right]_0^{\infty} = \frac{1}{1 + t^2}.$$

Это выражение можно рассматривать как характеристическую функцию, соответствующую плотности вероятности $f(x) = \frac{1}{2} e^{-|x|}$. Так как эта характеристическая функция интегрируема на $(-\infty, \infty)$, то из (10.3.2) получим

$$(10.5.7) \quad \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{-itx}}{1 + t^2} dt = e^{-|x|}.$$

10.6. Характеристическая функция распределения в R_n . Если рассматривать $t = (t_1, \dots, t_n)$ и $x = (x_1, \dots, x_n)$ как векторы-столбцы (см. параграф 11.2), соответствующие точкам в R_n , то через $t'x$ мы обозначим произведение, образованное по правилу (11.2.1) умножения векторов:

$$t'x = t_1 x_1 + \dots + t_n x_n.$$

Определение (10.1.1) характеристической функции одномерного распределения обобщается тогда следующим образом:

$$(10.6.1) \quad \varphi(t) = \varphi(t_1, \dots, t_n) = \int_{R_n} e^{it'x} dP,$$

где $P = P(S)$ есть вероятностная функция распределения в R_n . Характеристическая функция $\varphi(t)$ распределения является, таким образом функцией n действительных переменных t_1, \dots, t_n . Очевидно, всегда $\varphi(0, \dots, 0) = 1$, и для всех значений переменных t_1, \dots, t_n

$$|\varphi(t)| \leq 1, \quad \varphi(-t) = \overline{\varphi(t)}.$$

Далее, функция $\varphi(t)$ всюду непрерывна. Если существуют все моменты распределения (см. параграф 9.2) до некоторого порядка, то в окрестности точки $t=0$ имеем разложение функции $\varphi(t)$, аналогичное (10.1.3).

Следующая теорема, которая является непосредственным обобщением теоремы единственности (10.3.1), показывает, что распределение в R_n однозначно определяется своей характеристической функцией.

Если интервал I , определенный неравенствами $a_v - h_v < x_v < a_v + h_v$ ($v = 1, \dots, n$), есть интервал непрерывности (см. параграф 8.3) функции $P(S)$, то

$$(10.6.2) \quad P(I) = \lim_{T \rightarrow \infty} \frac{1}{\pi^n} \int_{-T}^T \dots \int_{-T}^T \prod_{v=1}^n \frac{\sin h_v t_v}{t_v} e^{-it_v u_v} \varphi(t) dt_1 \dots dt_n.$$

Доказательство этой теоремы является простым обобщением доказательства (10.3.1). В частном случае, когда функция $|\varphi(t)|$ интегрируема на R_n , находим, как и в (10.3.2), что плотность вероятности (см. параграф 8.4) $\frac{\partial^n F}{\partial x_1 \dots \partial x_n} = f(x_1, \dots, x_n) = f(x)$ существует и непрерывна при всех x и, кроме того,

$$(10.6.3) \quad f(x) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-it_v x_v} \varphi(t) dt_1 \dots dt_n.$$

Обратная формула, соответствующая (10.1.4),

$$(10.6.4) \quad \varphi(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{it_v x_v} f(x) dx_1 \dots dx_n.$$

получается из (10.6.1) и справедлива, если плотность вероятности $f(x)$ существует и непрерывна, за исключением, может быть, некоторых точек, принадлежащих конечному числу гиперповерхностей в R_n .

Нам понадобится также следующее обобщение теоремы (10.3.3), которое доказывается так же, как и в одномерном случае:

Пусть I_{z_1, \dots, z_n} есть интервал, определенный неравенствами

$$a_v - z_v < x_v < a_v + z_v \quad (v = 1, \dots, n).$$

Для любых действительных a_v и положительных h_v , имеем

$$(10.6.5) \quad \int\limits_0^{h_1} \dots \int\limits_0^{h_n} P(I_{z_1, \dots, z_n}) dz_1 \dots dz_n = \\ = \frac{1}{\pi^n} \int\limits_{-\infty}^{\infty} \dots \int\limits_{-\infty}^{\infty} \prod_{v=1}^n \frac{1 - \cos h_v t_v}{t_v^2} e^{-it_v a_v} \varphi(t) dt_1 \dots dt_n.$$

10.7. Теорема непрерывности для характеристических функций в R_n . Теорема непрерывности, доказанная в параграфе 10.4, может быть непосредственно обобщена на многомерные распределения. Согласно параграфу 8.5, последовательность распределений в R_n сходится тогда и только тогда, когда соответствующие функции распределения сходятся к некоторой функции распределения. Как и в одномерном случае, часто в приложениях бывает проще решать вопрос о сходимости соответствующей последовательности характеристических функций. В этом случае будет полезна следующая теорема:

Дана последовательность распределений в R_n с функциями распределения $F_1(x), F_2(x), \dots$ и характеристическими функциями $\varphi_1(t), \varphi_2(t), \dots$. Для того чтобы последовательность $\{F_n(x)\}$ сходилась к некоторой функции распределения $F(x)$, необходимо и достаточно, чтобы при любом t последовательность $\{\varphi_n(t)\}$ сходилась к пределу $\varphi(t)$, непрерывному при $t = 0$.

Если это условие выполнено, то предел $\varphi(t)$ совпадает с характеристической функцией предельной функции распределения $F(x)$.

Доказательство необходимости этого условия вполне аналогично соответствующей части доказательства в параграфе 10.4 и использует обобщение соотношения (7.5.9) на интегралы в R_n (см. параграф 9.4). Отсюда также следует, что предел $\varphi(t)$ есть характеристическая функция функции распределения $F(x)$. Для того чтобы доказать, что наше условие достаточно, рассмотрим подпоследовательность $\{F_{m_p}(x)\}$, сходящуюся (см. параграф 8.5) к пределу $F(x) = F(x_1, \dots, x_n)$, который является функцией неубывающей и непрерывной справа по каждой переменной x_v . Мы хотим показать, что $F(x)$ есть функция распределения, т. е. что соответствующая неотрицательная аддитивная функция множества $P(S)$ есть вероятностная функция. Для этой цели достаточно показать, что $P(R_n) = 1$. Затем применяем (10.6.5) к каждой $\varphi_{m_p}(t)$, полагая все $a_v = 0$. Если μ стремится к бесконечности, то тем же

рассуждением, что и в параграфе 10.4, получим

$$\begin{aligned} & \frac{1}{h_1 \dots h_n} \int_0^{h_1} \dots \int_0^{h_n} P(I_{z_1, \dots, z_n}) dz_1 \dots dz_n = \\ & = \frac{1}{\pi^n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{v=1}^n \frac{1 - \cos t_v}{t_v^2} \varphi\left(\frac{t_1}{h_1}, \dots, \frac{t_n}{h_n}\right) dt_1 \dots dt_n. \end{aligned}$$

Выбирая h_v , стремящимися к бесконечности, получим, совершенно аналогично с одномерным случаем,

$$P(R_n) = \prod_{v=1}^n \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \cos t_v}{t_v^2} dt_v = 1,$$

так что предел $P(S)$ последовательности $\{P_{m_k}(S)\}$ есть некоторая вероятностная функция. Завершается доказательство так же, как и в параграфе 10.4.

ГЛАВА II

МАТРИЦЫ, ДЕТЕРМИНАНТЫ И КВАДРАТИЧНЫЕ ФОРМЫ

Содержание настоящей главы излагается во многих учебниках в элементарной форме, вполне пригодной для нашей цели. Мы отсылаем читателя, в частности, к книгам Эйткена [1], Бокера [3], Бора и Моллерупа [5] *). Мы ограничимся здесь тем, что дадим для удобства читателя краткий обзор — во многих случаях без законченных доказательств — основных определений и свойств, которые будут употребляться в дальнейшем, проводя полностью доказательства лишь некоторых теорем, не содержащихся в учебниках.

11.1. Матрицы. *Матрица A порядка $m \cdot n$ есть прямоугольная таблица чисел или элементов a_{ik} , расположенных в m строках и n столбцах:*

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Кратко пишем $A = \{a_{ik}\}$ или, если мы хотим выделить порядок матрицы, пишем A_{mn} вместо A . Будем всегда предполагать, что элементы a_{ik} — действительные числа.

* См. также А. Г. Курош, „Курс высшей алгебры“, 1946, или И. М. Гельфанд, „Лекции по линейной алгебре“, 1948. (Прим. ред.)

В частном случае, когда $m = n = 1$, матрица A состоит из единственного элемента a_{11} ; мы будем идентифицировать такую матрицу с числом a_{11} .

Две матрицы A и B мы будем называть *равными* и писать $A = B$ тогда и только тогда, когда они имеют одинаковый порядок и все соответствующие элементы их равны: $a_{ik} = b_{ik}$ для всех i и k . Определим теперь три *операции над матрицами*:

1. *Произведение* матрицы A на число c определяется как матрица, получаемая умножением всех элементов A на c . Таким образом, $cA = Ac = B$, где элементы B суть $b_{ik} = ca_{ik}$. Если $c = -1$, пишем $-A$ вместо $(-1)A$.

2. *Сумма* двух матриц A и B определяется только в случае, когда обе матрицы имеют один и тот же порядок. Тогда сумма $C = A + B$ определяется как матрица того же порядка с элементами $c_{ik} = a_{ik} + b_{ik}$.

3. *Произведение* двух матриц A и B определяется только в случае, когда *первый сомножитель* A имеет порядок $m \cdot r$, а *второй сомножитель* B — порядок $r \cdot n$, так что число столбцов первого сомножителя совпадает с числом строк второго сомножителя. Тогда произведение $C = AB$ или $C_{mn} = A_{mr}B_{rn}$ определяется как матрица порядка $m \cdot n$ с элементами c_{ik} , заданными выражением

$$c_{ik} = \sum_{j=1}^r a_{ij} b_{jk}.$$

Элемент в i -й строке и k -м столбце произведения есть, таким образом, сумма всех произведений соответствующих элементов из i -й строки первого сомножителя и k -го столбца второго сомножителя.

Определенные таким образом три операции над матрицами *ассоциативны* и *дистрибутивны*. Кроме того, две первые операции *коммутативны*, в то время как третья операция, вообще говоря, не коммутативна. Таким образом имеем, например,

$$(A + B) + C = A + (B + C), \quad (AB)C = A(BC), \\ C(A + B) = CA + CB, \quad (A + B)C = AC + BC, \\ A + B = B + A, \quad c(A + B) = cA + cB,$$

но, вообще говоря, $AB = BA$ не справедливо. Даже если оба произведения AB и BA определены, они могут не совпадать. Таким образом, необходимо различать *умножение справа* и *умножение слева*.

Из этих свойств следует, например, что линейная комбинация $c_1A_1 + \dots + c_pA_p$ однозначно определена, если только все матрицы A_i одного и того же порядка, и что порядок слагаемых может быть изменен произвольным образом. Аналогично, произведение $D_{mn} = A_{mr}B_{rs}C_{st}$ однозначно определено, но здесь нельзя менять порядок сомножителей.

Элементы d_{hk} матрицы D даются выражением

$$d_{hk} = \sum_{i=1}^r \sum_{j=1}^s a_{hi}b_{ij}c_{jk}.$$

Если задана матрица $A = \{a_{ik}\}$ порядка $m \cdot n$, то *транспонированной* матрицей $A' = \{a'_{ik}\}$ называется матрица порядка $n \cdot m$, такая, что $a'_{ik} = a_{ki}$. Таким образом, строки матрицы A' являются столбцами матрицы A , а столбцы матрицы A' являются строками матрицы A . Имеем, очевидно,

$$(A')' = A, \quad (A + B)' = A' + B', \quad (AB)' = B'A'.$$

Любая матрица, получаемая вычеркиванием некоторого числа строк и столбцов из матрицы A , называется *подматрицей* матрицы A . В частности, каждый элемент матрицы A есть подматрица порядка $1 \cdot 1$, а строки и столбцы суть подматрицы соответственно порядков $1 \cdot n$ и $m \cdot 1$.

Если $m = n$, то матрица A называется *квадратной*. Вследствие ассоциативности умножения матриц ясно, что следует понимать под степенями A^2, A^3, \dots квадратной матрицы A . Элементы $a_{11}, a_{22}, \dots, a_{nn}$ квадратной матрицы образуют *главную диагональ* матрицы и называются *диагональными элементами*.

Квадратная матрица, симметричная относительно своей главной диагонали, называется *симметричной*. Симметричная матрица совпадает со своей транспонированной, так что $A' = A$ или $a_{ki} = a_{ik}$. Для произвольной матрицы $A = A_{mn}$ произведения AA' и $A'A$ суть симметричные матрицы соответственно порядков $m \cdot m$ и $n \cdot n$.

Симметричная матрица, все недиагональные элементы которой равны нулю, называется *диагональной*. Если A_{mn} — произвольная матрица, а D_{mm} и D_{nn} — диагональные матрицы, то произведение $D_{mm}A_{mn}$ получается умножением строк матрицы A на соответствующие диагональные элементы матрицы D , а произведение $A_{mn}D_{nn}$ получается

умножением столбцов матрицы A на соответствующие диагональные элементы матрицы D .

Единичной матрицей I называется диагональная матрица, все диагональные элементы которой равны 1. Для любой матрицы $A = A_{mn}$ имеем

$$IA = AI = A,$$

где I обозначает единичную матрицу порядка $m \cdot m$ в первом произведении и порядка $n \cdot n$ — во втором произведении.

Матрица (не обязательно квадратная), все элементы которой равны нулю, называется *нулевой* и обозначается через 0.

11.2. Векторы. *Вектор* есть матрица, состоящая из одной строки или из одного столбца; она называется *вектором-строкой* в первом случае и *вектором-столбцом* — во втором. Таким образом, вектор-строка $x = \{x_1, \dots, x_n\}$ есть матрица порядка $1 \cdot n$, а вектор-столбец

$$x = \begin{Bmatrix} x_1 \\ \vdots \\ x_n \end{Bmatrix}$$

есть матрица порядка $n \cdot 1$. Для упрощения записи будем, однако, обычно записывать последний вектор в форме $x = (x_1, \dots, x_n)$; круглые скобки вместо фигурных указывают на то, что вектор нужно представлять себе как вектор-столбец. Большинство векторов, встречающихся в приложениях, будут именно этого рода.

Транспонированием вектора-столбца $x = (x_1, \dots, x_n)$ получаем вектор-строку $x' = \{x_1, \dots, x_n\}$, и обратно.

Если $x = (x_1, \dots, x_n)$ и $y = (y_1, \dots, y_n)$ суть два вектора-столбца, то произведение $x'y$ есть матрица порядка $1 \cdot 1$, т. е. обычное число

$$(11.2.1) \quad x'y = x_1y_1 + \dots + x_ny_n.$$

В частном случае при $x = y$ имеем

$$x'x = x_1^2 + \dots + x_n^2.$$

Произведения xy' и xx' , с другой стороны, являются не числами, но матрицами порядка $n \cdot n$.

Векторы x_1, \dots, x_p называются *линейно зависимыми*, если существуют такие числа c_i , не все равные нулю, что

$$c_1x_1 + \dots + c_px_p = 0.$$

В противном случае векторы x_1, \dots, x_p называются *линейно независимыми*. Аналогично, p функций f_1, \dots, f_p одной или более переменных называются линейно зависимыми, если соотношение $c_1f_1 + \dots + c_pf_p = 0$, где c_i — некоторые постоянные, не все равные нулю, выполнено при всех значениях переменных. Если существуют несколько линейных соотношений такого вида, то они называются *независимыми*, если соответствующие векторы $\mathbf{c} = (c_1, \dots, c_p)$ линейно независимы.

11.3. Матричное обозначение линейных преобразований. Линейное преобразование

$$(11.3.1) \quad \begin{aligned} x_1 &= a_{11}y_1 + a_{12}y_2 + \dots + a_{1n}y_n, \\ x_2 &= a_{21}y_1 + a_{22}y_2 + \dots + a_{2n}y_n, \\ &\dots \dots \dots \dots \dots \dots \\ x_m &= a_{m1}y_1 + a_{m2}y_2 + \dots + a_{mn}y_n \end{aligned}$$

устанавливает некоторое соотношение между двумя системами переменных x_1, \dots, x_m и y_1, \dots, y_n , где m не обязательно равно n . Матрица $A = A_{mn} = \{a_{ik}\}$ называется матрицей этого преобразования.

Если теперь $\mathbf{x} = (x_1, \dots, x_m)$ и $\mathbf{y} = (y_1, \dots, y_n)$ рассматривать как векторы-столбцы, то правые стороны равенств (11.3.1) суть элементы матрицы-произведения $A\mathbf{y}$ порядка $m \cdot 1$, т. е. некоторого вектора-столбца. Таким образом, (11.3.1) выражает, что соответствующие элементы векторов-столбцов \mathbf{x} и $A\mathbf{y}$ равны, так что в матричном обозначении преобразование (11.3.1) принимает простую форму $\mathbf{x} = A\mathbf{y}$.

11.4. Матричное обозначение для билинейных и квадратичных форм. В векторах-столбцах \mathbf{x} и \mathbf{y} предыдущего параграфа будем рассматривать теперь x_i и y_k как два множества независимых переменных и образуем матрицу-произведение $\mathbf{x}'A\mathbf{y}$, где $A = A_{mn} = \{a_{ik}\}$. Эта матрица имеет порядок $1 \cdot 1$, т. е. это — обычное число

$$(11.4.1) \quad \mathbf{x}'A\mathbf{y} = \sum_{i,k} a_{ik}x_i y_k,$$

где $i = 1, 2, \dots, m$ и $k = 1, 2, \dots, n$. Таким образом, *билинейная форма* переменных x_i и y_k в правой части последнего соотношения имеет простое выражение в матричном обозначении.

В важном частном случае, когда $m = n$, $\mathbf{x} = \mathbf{y}$ и матрица A симметрична, билинейная форма (11.4.1) принимает вид

$$(11.4.2) \quad \mathbf{x}'A\mathbf{x} = \sum_{i,k=1}^n a_{ik}x_i x_k,$$

где $a_M = a_{ik}$. Это выражение называется *квадратичной формой* переменных x_1, \dots, x_n и часто обозначается $Q(\mathbf{x})$ или $Q(x_1, \dots, x_n)$. В матричном обозначении имеем, таким образом, $Q(\mathbf{x}) = \mathbf{x}' A \mathbf{x}$. Симметричная матрица A называется *матрицей квадратичной формы* Q . Если, в частности, $A = I$, то имеем $Q = \mathbf{x}' I \mathbf{x} = \mathbf{x}' \mathbf{x} = x_1^2 + \dots + x_n^2$.

Матричные выражения (11.4.1) и (11.4.2) особенно удобны для изучения *линейных преобразований* билинейных и квадратичных форм.

Так, если в квадратичной форме $Q(x_1, \dots, x_n) = \sum_{i,k=1}^n a_{ik} x_i x_k$ вводятся новые переменные y_1, \dots, y_m с помощью линейного преобразования $\mathbf{x} = \mathbf{C}\mathbf{y}$, где $\mathbf{C} = \mathbf{C}_{nm}$, то результат есть квадратичная форма $Q_1(y_1, \dots, y_m)$ относительно новых переменных:

$$Q(x_1, \dots, x_n) = Q_1(y_1, \dots, y_m) = \sum_{i,k=1}^m b_{ik} y_i y_k,$$

и матричное выражение (11.4.2) непосредственно дает

$$\mathbf{Q} = \mathbf{x}' A \mathbf{x} = \mathbf{y}' \mathbf{C}' A \mathbf{C} \mathbf{y} = \mathbf{y}' B \mathbf{y},$$

где $B = \mathbf{C}' A \mathbf{C}$. Транспонируя, убеждаемся, что это симметричная матрица, и, таким образом, матрица преобразованной формы есть $\mathbf{C}' A \mathbf{C}$. Ее порядок, конечно, есть $m \cdot m$.

11.5. Детерминанты. Каждой квадратной матрице $A = A_{nn} = \{a_{ik}\}$ соответствует некоторое число A , называемое *детерминантом* матрицы; оно обозначается

$$A = |A| = |a_{ik}| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

и определяется как сумма

$$A = \sum \pm a_{1r_1} a_{2r_2} \dots a_{nr_n},$$

где вторые индексы r_1, \dots, r_n пробегают все $n!$ возможных перестановок чисел $1, \dots, n$, а знак каждого слагаемого берется $+$ или $-$ в зависимости от того, является ли соответствующая перестановка четной или нечетной. Число n называется *порядком* детерминанта.

Детерминанты квадратной матрицы A и транспонированной матрицы A' равны: $A = A'$. Если переставить две строки или два столбца в матрице A , то детерминант меняет знак. Следовательно, если две строки или два столбца матрицы A одинаковы, то ее детерминант равен нулю. Если A, B и C — квадратные матрицы, такие, что $AB = C$, то соответствующие детерминанты удовлетворяют соотношению $AB = C$.

Если A — произвольная матрица (не обязательно квадратная), то детерминант любой ее квадратной подматрицы называется *минором* матрицы A . Если матрица A квадратная, то *главными минорами* называются миноры, диагональные элементы которых суть диагональные элементы матрицы A .

В квадратной матрице $A = \{a_{ik}\}$ алгебраическим дополнением элемента a_{ik} называется минор A_{ik} , получаемый вычеркиванием i -й строки и k -го столбца, умноженный на $(-1)^{i+k}$. Имеют место следующие важные тождества:

$$(11.5.1) \quad \sum_{j=1}^n a_{ij} A_{kj} = \begin{cases} A & \text{при } i = k, \\ 0 & \text{при } i \neq k, \end{cases}$$

$$(11.5.2) \quad \sum_{i=1}^n a_{ji} A_{jk} = \begin{cases} A & \text{при } j = k, \\ 0 & \text{при } j \neq k, \end{cases}$$

и далее

$$(11.5.3) \quad A = a_{11} A_{11} - \sum_{i,k=2}^n a_{1i} a_{ik} A_{11..ik},$$

где $A_{11..ik}$ есть алгебраическое дополнение элемента a_{ik} в A_{11} .

11.6. Ранг матрицы A (не обязательно квадратной) есть максимальное число r , такое, что матрица A содержит по крайней мере один минор порядка r , не равный нулю. Если все миноры матрицы A равны нулю, то A есть нулевая матрица и мы полагаем $r = 0$. Если $A = A_{mn}$, то ранг r не превосходит меньшего из чисел m и n .

Будем рассматривать строки и столбцы матрицы A как векторы. Если A — матрица ранга r , то в ней можно найти r линейно независимых строк, в то время как любые $r+1$ строк линейно зависимы. То же справедливо и для столбцов.

Если A_1, A_2, \dots, A_p — матрицы рангов r_1, r_2, \dots, r_p , то ранг суммы $A_1 + \dots + A_p$ не больше суммы $r_1 + \dots + r_p$, а ранг произведения не превосходит наименьшего из чисел r_1, \dots, r_p .

Если квадратная матрица $A = A_{nn}$ такова, что $A \neq 0$, то она имеет ранг n . Такие матрицы называются *невырожденными*, а квадратные матрицы A , для которых $A = 0$, называются *вырожденными*. Если произвольную матрицу B умножить (слева или справа) на невырожденную матрицу A , то произведение будет иметь тот же ранг, что и B . Линейное преобразование называется вырожденным или невырожденным в зависимости от того, вырожденная или нет матрица этого преобразования.

Если матрица A симметрична и имеет ранг r , то существует по меньшей мере один главный минор матрицы A порядка r , не равный нулю. Следовательно, в частности, ранг диагональной матрицы равен числу диагональных элементов, отличных от нуля.

Ранг квадратичной формы $Q = x'Ax = \sum_{i,k=1}^n a_{ik}x_i x_k$ равен, по определению, рангу матрицы A . Соответственно, форма Q называется вырожденной или невырожденной в зависимости от того, вырождена матрица A или нет. Невырожденное линейное преобразование не изменяет ранга формы. Если при таком преобразовании Q переходит в $\sum_1^r x_i y_i^2$, где $x_i \neq 0$ при $i = 1, 2, \dots, r$, то ранг формы Q равен r . Ранг есть наименьшее число независимых переменных, к функции которых Q может быть приведена всевозможными невырожденными линейными преобразованиями.

Часто бывает полезно следующее предложение. Если Q может быть записано в виде $Q = L_1^2 + \dots + L_p^2$, где L_i суть линейные функции переменных x_1, \dots, x_n , и если существуют ровно h независимых линейных соотношений (см. параграф 11.2) между L_i , то ранг формы Q равен $p - h$. Следовательно, если мы знаем, что существует *не менее* h таких линейных соотношений, то ранг формы Q не больше $p - h$.

11.7. Присоединенная и обратная матрицы. Пусть $A = \{a_{ik}\}$ — квадратная матрица и пусть, как и раньше, A_{ik} обозначает алгебраическое дополнение элемента a_{ik} . Если мы образуем матрицу $\{A_{ik}\}$, а затем транспонируем ее, то получим новую матрицу $A^* = \{a_{ik}^*\}$, где

$a_{ik}^* = A_{ki}$. Будем называть матрицу A^* *присоединенной* к A . Из тождеств (11.5.1) и (11.5.2) получим

$$(11.7.1) \quad AA^* = A^*A = AI = \begin{pmatrix} A & 0 & \dots & 0 \\ 0 & A & \dots & 0 \\ \vdots & \ddots & \ddots & \\ 0 & 0 & \dots & A \end{pmatrix}$$

Для алгебраического дополнения A_{ik}^* элемента $a_{ik}^* = A_{ik}$ в матрице A^* получим соотношение

$$(11.7.2) \quad A_{ik}^* = A^{n-2} a_{ik}.$$

Это только частный случай общего соотношения, которое выражает любой минор матрицы A^* через A и его миноры. Здесь мы только сошлемся на дальнейший частный случай

$$(11.7.3) \quad \begin{vmatrix} A_{11} & A_{11} \\ A_{1k} & A_{1k} \end{vmatrix} = A_{11} A_{1k} - A_{11} A_{1k} = AA_{11 \cdot 1k}.$$

Если матрица A — невырожденная, то матрица $A^{-1} = \frac{1}{A} A^* = \left\{ \frac{A_{ij}}{A} \right\}$ называется *обратной* к A . Из (11.7.1) получим

$$(11.7.4) \quad AA^{-1} = A^{-1}A = I.$$

Матричные уравнения $AX = I$ и $XA = I$ оба имеют одно решение, а именно: $X = A^{-1}$. Следовательно, детерминант матрицы A^{-1} равен A^{-1} . Далее, имеют место следующие соотношения:

$$\begin{aligned} (A^{-1})^{-1} &= A, \\ (A^{-1})' &= (A')^{-1}, \\ (AB)^{-1} &= B^{-1} A^{-1} \end{aligned}$$

Если матрица A симметрична, то $A_{ik} = A_{ki}$, так что матрицы A^* и A^{-1} также симметричны. Матрица, обратная к диагональной матрице D с диагональными элементами d_1, \dots, d_n , есть также диагональная матрица D^{-1} с диагональными элементами $d_1^{-1}, \dots, d_n^{-1}$.

Если $Q = \mathbf{x}' A \mathbf{x}$ есть невырожденная квадратичная форма, то форма $Q^{-1} = \mathbf{x}' A^{-1} \mathbf{x}$ называется *обратной* к Q . Очевидно, что $(Q^{-1})^{-1} = Q$.

Пусть $\mathbf{x} = (x_1, \dots, x_n)$ и $\mathbf{t} = (t_1, \dots, t_n)$ — переменные векторы-столбцы. Если новые переменные $\mathbf{y} = (y_1, \dots, y_m)$ и $\mathbf{u} = (u_1, \dots, u_m)$

вводятся посредством преобразований

$$(11.7.5) \quad \mathbf{y} = \mathbf{C}x, \quad t = \mathbf{C}'u,$$

где $\mathbf{C} = \mathbf{C}_{mn}$, то

$$(11.7.6) \quad t'x = u'Cx = u'y.$$

Билинейная форма $t'x = t_1x_1 + \dots + t_nx_n$ преобразовывается, таким образом, в билинейную форму $u'y = u_1y_1 + \dots + u_my_m$ от новых переменных. Два множества переменных x_i и t_i , которые преобразуются согласно (11.7.5), называются *контрагredientными* множествами переменных. В частном случае, когда $m = n$ и матрица \mathbf{C} невырожденная, соотношение (11.7.5) может быть записано в виде

$$(11.7.7) \quad \mathbf{y} = \mathbf{C}x, \quad u = (\mathbf{C}')^{-1}t.$$

11.8. Линейные уравнения. Мы будем рассматривать здесь только некоторые частные случаи. *Неоднородная* система

$$(11.8.1) \quad \begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= h_1, \\ \dots & \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= h_n, \end{aligned}$$

эквивалентна матричному соотношению $Ax = h$, где

$$A = \{a_{ik}\}, \quad x = (x_1, \dots, x_n) \text{ и } h = (h_1, \dots, h_n).$$

Если матрица A невырожденная, то можно обе стороны этого соотношения умножить слева на A^{-1} и, таким образом, получить единственное решение $x = A^{-1}h$ или, в развернутой форме,

$$(11.8.2) \quad x_k = \frac{1}{A} \sum_{i=1}^n h_i A_{ik} \quad (k = 1, 2, \dots, n).$$

Таким образом, x_k выражается дробью с знаменателем A и числителем, равным детерминанту, получаемому из A заменой элементов k -го столбца элементами h_1, \dots, h_n . Этот классический результат принадлежит Крамеру (1750 г.).

Рассмотрим теперь *однородную* систему

$$(11.8.3) \quad \begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= 0, \\ \dots & \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= 0, \end{aligned}$$

или, в матричном обозначении, $Ax = \mathbf{0}$, где m не обязательно равно n . Согласно 11.6, матрица A имеет ранг $r \leq n$. Если $r = n$, то система (11.8.3) имеет только тривиальное решение $x = \mathbf{0}$. С другой стороны, если $r < n$, то можно найти $n - r$ линейно независимых векторов c_1, \dots, c_{n-r} , таких, что общее решение системы (11.8.3) может быть записано в виде $x = t_1 c_1 + \dots + t_{n-r} c_{n-r}$, где t_i — произвольные постоянные.

11.9. Ортогональные матрицы. Характеристические числа. Квадратная матрица $C = \{c_{ik}\}$ называется *ортогональной*, если $CC' = I$. Следовательно, $C^2 = 1$, так что ее детерминант $C = |C| = \pm 1$. Очевидно, что, транспонируя ортогональную матрицу C , получим снова ортогональную матрицу C' . Далее, $C^{-1} = C'$, и, таким образом, согласно определению обратной матрицы, $C_{ik} = C c_{ik}$ для всех i и k и следовательно, в силу тождеств (11.5.1) и (11.5.2),

$$(11.9.1) \quad \sum_{j=1}^n c_{ij} c_{kj} = \begin{cases} 1 & \text{при } i = k, \\ 0 & \text{при } i \neq k. \end{cases}$$

$$(11.9.2) \quad \sum_{j=1}^n c_{ji} c_{jk} = \begin{cases} 1 & \text{при } i = k, \\ 0 & \text{при } i \neq k. \end{cases}$$

Произведение $C_1 C_2$ двух ортогональных матриц одного порядка, есть ортогональная матрица. Если дано произвольное число $p < n$ строк $c_{i1}, c_{i2}, \dots, c_{ip}$ ($i = 1, 2, \dots, p$), для которых выполняются соотношения (11.9.1), то можно всегда найти еще $n - p$ строк, таких, что, добавляя их, мы получим ортогональную матрицу порядка $n \times n$. То же справедливо, конечно, и для столбцов.

Линейное преобразование $x = Cy$, где C есть ортогональная матрица, называется *ортогональным преобразованием*. Квадратичная форма $x'x = x_1^2 + \dots + x_n^2$ инвариантна относительно такого преобразования, т. е. она преобразуется в форму $y'C'Cy = y'y = y_1^2 + \dots + y_n^2$ с той же матрицей I . Обратное преобразование $y = C^{-1}x$ также ортогонально, так как матрица $C^{-1} = C'$ ортогональна.

Ортогональные преобразования имеют важное геометрическое значение. Действительно, ортогональное преобразование можно рассматривать как аналитическое выражение преобразования координат в n -мерном евклидовом пространстве, которое осуществляется вращением осей прямоугольной системы координат около фиксированного начала коор-

динат. Расстояние $(x_1^2 + \dots + x_n^2)^{1/2}$ от начала координат до точки (x_1, \dots, x_n) инвариантно относительно всякого такого вращения.

Если A — произвольная симметрическая матрица, то всегда можно найти такую ортогональную матрицу C , что произведение $C'AC$ есть диагональная матрица:

$$(11.9.3) \quad C'AC = K = \begin{pmatrix} x_1 & 0 & \dots & 0 \\ 0 & x_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & x_n \end{pmatrix}.$$

Любая другая ортогональная матрица, удовлетворяющая этому условию, дает те же самые диагональные элементы x_1, \dots, x_n , хотя, возможно, расположенные в другом порядке. Числа x_1, \dots, x_n , которые зависят, таким образом, только от матрицы A , называются *характеристическими числами* матрицы A . Эти числа суть n корней *векового уравнения*

$$(11.9.4) \quad |A - xI| = \begin{vmatrix} a_{11} - x & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - x & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} - x \end{vmatrix} = 0,$$

и все они действительны. Так как матрица C невырожденная, то матрицы A и K имеют один и тот же ранг (см. параграф 11.6). Следовательно, ранг матрицы A равен числу корней x_i , отличных от нуля. Из (11.9.3) получаем, рассматривая детерминанты обеих частей этого соотношения и учитывая, что $C^2 = 1$,

$$(11.9.5) \quad A = x_1 x_2 \dots x_n.$$

Если матрица A невырожденная, то тождество

$$(11.9.6) \quad |A^{-1} - xI| = (-x)^n A^{-1} \left| A - \frac{1}{x} I \right|$$

показывает, что характеристические числа матрицы A^{-1} суть числа обратные к характеристическим числам матрицы A .

Наконец, пусть B есть матрица порядка $m \cdot n$, где $m \leq n$. Если матрица B имеет ранг m , то все характеристические числа симметрической матрицы BB' порядка $m \cdot m$ положительны. Следовательно, в частности, матрица BB' — невырожденная. Это доказывается без труда, если взять в (11.9.3) $A = BB'$ и выразить произвольное характеристическое число x_i посредством правила умножения.

11.10. Неотрицательные квадратичные формы. Если для всех действительных значений переменных x_1, \dots, x_n мы имеем

$$Q(x_1, \dots, x_n) = \sum_{i,k=1}^n a_{ik}x_i x_k \geq 0,$$

где $a_{ik} = a_{ki}$, то форма Q называется *неотрицательной* квадратичной формой. Если, кроме того, знак равенства в последнем соотношении имеет место только тогда, когда все x_i равны нулю, то форма Q называется *положительно определенной*. Форма Q называется *полуопределенной положительной*, если она неотрицательна, но не является положительно определенной. Свойство формы быть неотрицательной, положительно определенной или полуопределенной положительной, очевидно, инвариантно относительно невырожденных линейных преобразований.

Симметричная матрица $A = \{a_{ik}\}$ называется неотрицательной, положительно определенной или полуопределенной положительной в зависимости от того, каким из этих свойств обладает форма $Q = \mathbf{x}'A\mathbf{x}$.

Ортогональное преобразование $\mathbf{x} = \mathbf{C}\mathbf{y}$, где \mathbf{C} — ортогональная матрица, фигурирующая в преобразовании (11.9.3), превращает Q в форму, содержащую только квадраты:

$$(11.10.1) \quad Q(x_1, \dots, x_n) = x_1 y_1^2 + x_2 y_2^2 + \dots + x_n y_n^2$$

или в матричном обозначении $\mathbf{x}'A\mathbf{x} = \mathbf{y}'K\mathbf{y}$, где x_i — характеристические числа матрицы A , а K — соответствующая диагональная матрица из (11.9.3). Это же самое ортогональное преобразование преобразует форму $Q = \mathbf{x}(x_1^2 + \dots + x_n^2)$ в $(x_1 - x)y_1^2 + \dots + (x_n - x)y_n^2$. Если x не превосходит наименьшего характеристического числа матрицы A , то последняя форма, очевидно, неотрицательна и, следовательно, форма $Q = \mathbf{x}(x_1^2 + \dots + x_n^2)$ с матрицей $A - xI$ обладает тем же свойством.

Если Q — положительно определенная форма, то форма, стоящая в правой части соотношения (11.10.1), обладает тем же свойством, и, следовательно, при этом все характеристические числа x_i положительны. Следовательно, согласно (11.9.5), $A > 0$, так что матрица A невырожденная.

Если, с другой стороны, Q есть полуопределенная положительная форма, то аналогичное рассуждение показывает, что по крайней мере одно из характеристических чисел равно нулю, так что $A = 0$. Если Q имеет ранг r , то существует ровно r положительных характери-

стических чисел, а остальные $n-r$ равны нулю. В этом случае существует ровно $n-r$ линейно независимых векторов $x_p = (x_1^{(p)}, \dots, x_n^{(p)})$, таких, что $Q(x_p) = 0$.

Геометрический смысл рассмотренного выше ортогонального преобразования заключается в том, что соответствующим вращением системы координат гиперповерхность второго порядка, изображаемая уравнением $Q(x_1, \dots, x_n) = \text{const}$, приводится к своим главным осям. Если Q — положительно определенная форма, то уравнение $Q = \text{const}$ представляет n -мерный эллипсоид с полуосами $x_i^{-\frac{1}{2}}$. Для полуопределенных форм Q получаются различные классы эллиптических цилиндров.

Если форма Q положительно определенная, то, приравнивая одну или несколько x_i нулю, мы получим положительно определенную форму. Следовательно, любой главный минор Q положителен. Для полуопределенной положительной формы Q то же рассуждение показывает, что любой главный минор неотрицателен. Следовательно, в частности, если в неотрицательной форме Q не встречается квадрат x_i^2 , то Q не зависит от x_i . Действительно, в противном случае главный минор $a_{ii}a_{kk} - a_{ik}^2$ был бы отрицателен при некотором k . Обратно, если величины $A, A_{11}, A_{11.22}, \dots, A_{11..n-1, n-1}$ все положительны, то форма Q положительно определенная.

Подстановка $x = A^{-1}y$ преобразует форму $Q = x'Ax$ в обратную форму $Q^{-1} = y'A^{-1}y$. Таким образом, если Q — положительно определенная форма, то Q^{-1} также положительно определенная, и обратно. В этом можно убедиться и непосредственно из (11.9.6). Рассмотрим теперь соотношение (11.5.3) для положительно определенной симметричной матрицы A . Так как любая главная подматрица матрицы A также положительно определенная, то последнее слагаемое в правой части соотношения (11.5.3) есть положительно определенная квадратичная форма переменных a_{12}, \dots, a_{1n} , так что мы имеем $0 < A \leq a_{11}A_{11}$ и, вообще,

$$(11.10.2) \quad 0 < A \leq a_{ii}A_{ii} \quad (i = 1, 2, \dots, n).$$

Повторяя то же рассуждение, получим

$$(11.10.3) \quad 0 < A \leq a_{11}a_{22} \dots a_{nn}.$$

Знак равенства здесь имеет место только в случае, когда A — диагональная матрица. Вообще же для неотрицательных матриц соотношение

(11.10.3) будет выполняться, конечно, если заменить знак $<$ знаком \leqslant .

11.11. Разложение формы $\sum_1^n x_i^2$. В некоторых приложениях к статистике встречаются различные соотношения вида

$$(11.11.1) \quad \sum_1^n x_i^2 = Q_1 + \dots + Q_k,$$

где Q_i ($i = 1, 2, \dots, k$) — неотрицательная квадратичная форма переменных x_1, \dots, x_n ранга r_i .

Рассмотрим сперва частный случай $k=2$ и предположим, что существует ортогональное преобразование, переводящее Q_1 в сумму r_1 квадратов: $Q = \sum_1^{r_1} y_i^2$. Применим это преобразование к обеим частям соотношения (11.11.1). Левая часть перейдет в $\sum_1^n y_i^2$ и, следовательно, Q_2 перейдет в $\sum_{r_1+1}^n y_i^2$. Таким образом, ранг формы Q_2 есть $r_2 = n - r_1$, и все ее характеристические числа равны 0 или 1. В качестве примера рассмотрим тождество

$$(11.11.2) \quad \sum_1^n x_i^2 = n\bar{x}^2 + \sum_1^n (x_i - \bar{x})^2,$$

где $\bar{x} = \frac{1}{n} \sum_1^n x_i$. Ортогональное преобразование $y = Cx$, где первая строка матрицы C есть $\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}$, переводит форму $n\bar{x}^2 = \left(\frac{x_1}{\sqrt{n}} + \frac{x_2}{\sqrt{n}} + \dots + \frac{x_n}{\sqrt{n}}\right)^2$ в y_1^2 . Таким образом, это преобразование переводит $\sum_1^n (x_i - \bar{x})^2$ в $\sum_2^n y_i^2$. В разложении $\sum_1^n x_i^2$, согласно (11.11.2), два члена в правой части имеют соответственно ранг 1 и $n-1$.

Рассмотрим теперь соотношение (11.11.1) при произвольном $k > 1$. Докажем следующее предложение, принадлежащее Кохрену [66] (см. также [154]):

Если $\sum_1^k r_i = n$, то существует ортогональное преобразование $x = Cy$, переводящее все Q_i в суммы квадратов такого вида:

$$Q_1 = \sum_1^{r_1} y_i^2, Q_2 = \sum_{r_1+1}^{r_1+r_2} y_i^2, \dots, Q_k = \sum_{n-r_k+1}^n y_i^2,$$

где никакие две различные формы Q_i не содержат общей переменной y_i .

Докажем эту теорему с помощью индукции. Для $k=1$ справедливость теоремы очевидна. Таким образом, нужно показать, что если теорема справедлива при разложении на $k-1$ членов, то она справедлива также и при разложении на k членов. Чтобы это показать применим к (11.11.1) ортогональное преобразование $x = C_1 z$, переводящее Q_1 в $\sum_1^{r_1} z_i^2$. Это дает

$$\sum_1^{r_1} (1 - x_i) z_i^2 + \sum_{r_1+1}^n z_i^2 = Q'_2 + \dots + Q'_k,$$

где Q'_2, \dots, Q'_k обозначают формы, в которые перешли при этом Q_2, \dots, Q_k . Утверждаем теперь, что все x_i равны 1. Предположим, действительно, что p чисел x_i отличны от 1, а остальные равны 1. Обе части последнего соотношения суть квадратичные формы переменных z_1, \dots, z_n . Ранг левой части равен $n - r_1 + p$, а ранг правой части, согласно параграфу 11.6, не больше чем $r_2 + \dots + r_k = n - r_1$. Таким образом, $p = 0$ и все $x_i = 1$, так что

$$(11.11.3) \quad \sum_{r_1+1}^n z_i^2 = Q'_2 + \dots + Q'_k.$$

Здесь переменные z_1, \dots, z_{r_1} не входят в левую часть; мы теперь покажем, что эти переменные не входят также ни в одно слагаемое правой части. Если бы, например, Q'_2 зависело от z_1 , то, согласно предыдущему параграфу, Q'_2 содержало бы член cz_1^2 , где $c > 0$. Так как коэффициенты при z_1^2 в Q'_3, \dots, Q'_k неотрицательны, то это предположение приводит к противоречию с (11.11.3).

Таким образом, (11.11.3) дает представление $\sum_{r_1+1}^n z_i^2$ в виде суммы $k-1$ неотрицательных форм от переменных z_{r_1+1}, \dots, z_n . По предпо-

ложению, теорема Кохрена для этого разложения справедлива. Таким образом, существует ортогональное преобразование $n - r_1$ переменных, переводящее z_{r_1+1}, \dots, z_n в новые переменные y_{r_1+1}, \dots, y_n , такие, что

$$(11.11.4) \quad Q'_2 = \sum_{r_1+1}^{r_1+r_2} y_i^2, \dots, Q'_k = \sum_{n-r_k+1}^n y_i^2.$$

Если дополнить это преобразование r_1 уравнениями $z_1 = y_1, \dots, z_{r_1} = y_{r_1}$, то получим ортогональное преобразование n переменных $\mathbf{z} = \mathbf{C}_2 \mathbf{y}$, для которого соотношение (11.11.4) выполнено.

Результат последовательного выполнения преобразований $\mathbf{x} = \mathbf{C}_1 \mathbf{z}$ и $\mathbf{z} = \mathbf{C}_2 \mathbf{y}$ будет ортогональным преобразованием $\mathbf{x} = \mathbf{C}_1 \mathbf{C}_2 \mathbf{y}$, так как произведение двух ортогональных матриц есть ортогональная матрица. Это преобразование обладает всеми требуемыми свойствами, и, таким образом, теорема доказана.

Заметим, что если относительно соотношения (11.11.1) известно только, что каждая форма Q_i неотрицательна и ранг ее не превосходит r_i , где $\sum_i^k r_i = n$, то мы можем сразу заключить, что Q_i действительно имеет ранг r_i , так что условия теоремы Кохрена выполнены. Действительно, так как ранг суммы квадратичных форм не больше суммы рангов слагаемых, то, обозначая через r'_i ранг формы Q_i , имеем

$$n \leq \sum_1^k r'_i \leq \sum_1^k r_i = n.$$

Таким образом, $\sum r'_i = \sum r_i$ и $r'_i \leq r_i$. Отсюда, очевидно, следует, что $r'_i = r_i$ для всех i .

Заметим, наконец, что теорема Кохрена, очевидно, остается в силе, если заменить левую часть соотношения (11.11.1) квадратичной формой Q некоторого числа переменных, которая ортогональным преобразованием может быть переведена в $\sum_1^n x_i^2$.

11.12. Некоторые интегральные формулы. Докажем сперва важную формулу

$$(11.12.1a) \quad \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{\frac{1}{2} x' A x} dx_1 \dots dx_n = \frac{(2\pi)^{n/2}}{\sqrt{A}} e^{-\frac{1}{2} \operatorname{tr} A},$$

или в обычных обозначениях

$$(11.12.1b) \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\sum_{j=1}^n t_j x_j - \frac{1}{2} Q(x_1, \dots, x_n)} dx_1 \dots dx_n = \\ = \frac{(2\pi)^{n/2}}{\sqrt{A}} e^{-\frac{1}{2} t^T Q^{-1} t},$$

где Q — положительно определенная квадратичная форма с матрицей A , а $t = (t_1, \dots, t_n)$ — некоторый действительный вектор. Как и в предыдущем параграфе, A есть детерминант $|A|$, а Q^{-1} есть обратная форма, определенная в параграфе 11.7. Для $n=1$ формулы (11.12.1a — b) сводятся к (10.5.2).

Для того чтобы доказать (11.12.1), вводим новые переменные $y = (y_1, \dots, y_n)$ подстановкой $x = Cy$ с ортогональной матрицей C (см. (11.9.3)), такой, что $C'AC = K$, где K — диагональная матрица, образованная характеристическими числами x_j матрицы A . В то же время мы заменяем вектор t новым вектором $u = (u_1, \dots, u_n)$ посредством контрагredientной подстановки (см. (11.7.7)) $t = (C')^{-1}u$, которая в этом случае сводится к $t = Cu$, так как матрица C ортогональна. Согласно (11.7.6), имеем $t^T x = u^T y$. Так как $C = \pm 1$, то, обозначая интеграл в левой части соотношения (11.12.1) через J , получим

$$J = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{iu^T y - \frac{1}{2} y^T Ky} dy_1 \dots dy_n = \prod_{j=1}^n \int_{-\infty}^{\infty} e^{iu_j y_j - \frac{1}{2} x_j y_j^2} dy_j.$$

Применяя к каждому множителю правой части формулу (10.5.2), получим

$$J = \frac{(2\pi)^{n/2}}{\sqrt{x_1 x_2 \dots x_n}} e^{-\frac{1}{2} \sum_{j=1}^n \frac{u_j^2}{x_j}} = \frac{(2\pi)^{n/2}}{\sqrt{A}} e^{-\frac{1}{2} u^T K^{-1} u},$$

так как, согласно 11.7, диагональная матрица с диагональными элементами $\frac{1}{x_j}$ совпадает с K^{-1} , а в силу (11.9.5), мы имеем $A = x_1 x_2 \dots x_n$. Так как матрица C ортогональна, то мы имеем $K^{-1} = (C'AC)^{-1} = C^{-1}A^{-1}(C')^{-1} = C'A^{-1}C$. Следовательно, $u^T K^{-1} u = u^T C'A^{-1}Cu = t^T A^{-1}t$ и, таким образом,

$$J = \frac{(2\pi)^{n/2}}{\sqrt{A}} e^{-\frac{1}{2} t^T A^{-1} t},$$

т. е. формула (11.12.1а) доказана. Полагая, в частности, $t=0$, получим формулу

$$(11.12.2) \quad \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2}Q(x_1, \dots, x_n)} dx_1 \dots dx_n = \frac{\pi^{n/2}}{\sqrt{A}}.$$

Это имеет место даже для матрицы A с комплексными элементами при условии, что матрица, составленная из действительных частей этих элементов, является положительно определенной.

Рассмотрим далее интеграл

$$V = \int_{Q(x_1, \dots, x_n) < c^2} \dots \int dx_1 \dots dx_n,$$

представляющий n -мерный „объем“ области, ограниченной эллипсондом $Q=c^2$. Использованное выше ортогональное преобразование, вытекающее из простой замены $y_i = \frac{c}{\sqrt{x_i}} z_i$, показывает, что

$$V = \frac{c^n}{\sqrt{A}} \int \dots \int dz_1 \dots dz_n \cdot \sum_1^n z_i^2 < 1$$

Последний интеграл представляет объем n -мерной „единичной сферы“,

и мы покажем ниже, что он равен $\frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)}$, так что

$$(11.12.3) \quad V = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} \cdot \frac{c^n}{\sqrt{A}}.$$

Наконец нам понадобится значение интеграла,

$$B_{ik} = \int_{Q < c^2} \dots \int x_i x_k dx_1 \dots dx_n,$$

распространенного на ту же область, что и интеграл V . Делая такую же замену переменных, что и в случае V , находим теми же вычислениями, что матрица B с элементами B_{ik} есть

$$B = g_n C K^{-1} C' = g_n A^{-1},$$

где

$$g_n = \frac{c^{n+2}}{\sqrt{A}} \int \dots \int_{\sum z_i^2 < 1} z_1^2 dz_1 \dots dz_n.$$

Ниже мы покажем, что

$$g_n = \frac{c^{n+2}}{\sqrt{A}} \cdot \frac{\pi^{n/2}}{2 \Gamma\left(\frac{n}{2} + 2\right)} = \frac{c^n V}{n+2},$$

так что

$$(11.12.4) \quad B_{lk} = \frac{c^2 V}{n+2} \cdot \frac{A_{kl}}{A}.$$

Использованные выше интегралы Дирихле

$$J_1 = \int \dots \int dz_1 \dots dz_n \text{ и } J_2 = \int \dots \int z_1^2 dz_1 \dots dz_n,$$

взятые по n -мерной единичной сфере $\sum_1^n z_i^2 < 1$, могут быть вычислены по-средством преобразования

$$\begin{aligned} z_1 &= \cos \varphi_1, \\ z_2 &= \sin \varphi_1 \cos \varphi_2, \\ z_3 &= \sin \varphi_1 \sin \varphi_2 \cos \varphi_3, \\ &\dots \\ z_n &= \sin \varphi_1 \dots \sin \varphi_{n-1} \cos \varphi_n, \end{aligned}$$

которое устанавливает взаимнооднозначное соответствие между областями $\sum z_i^2 < 1$ и $0 < \varphi_l < \pi$ ($l = 1, 2, \dots, n$). Якобиан этого преобразования равен $(-1)^n (\sin \varphi_1)^n (\sin \varphi_2)^{n-1} \dots \sin \varphi_n$. С помощью соотношения

$$\int_0^\pi (\sin \varphi)^n d\varphi = 2 \int_0^{\frac{\pi}{2}} (\sin \varphi)^n d\varphi = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n+2}{2}\right)} \sqrt{\pi},$$

которое доказывается подстановкой $x = \sin^2 \varphi$ и применением (12.4.2), получим

$$J_1 = \int_0^\pi (\sin \varphi_1)^n d\varphi_1 \dots \int_0^\pi \sin \varphi_n d\varphi_n = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)},$$

$$\begin{aligned} J_2 &= \int_0^\pi (\sin \varphi_1)^n \cos^2 \varphi_1 d\varphi_1 \int_0^\pi (\sin \varphi_2)^{n-1} d\varphi_2 \dots \int_0^\pi \sin \varphi_n d\varphi_n = \\ &= \frac{\pi^{n/2}}{2 \Gamma\left(\frac{n}{2} + 2\right)}. \end{aligned}$$

РАЗЛИЧНЫЕ ДОПОЛНЕНИЯ

12.1. Символы O , o и \sim . Когда мы исследуем поведение функции $f(x)$ при $x \rightarrow 0$, при $x \rightarrow \infty$ или при других предельных переходах, часто бывает желательно сравнить порядок величины $f(x)$ с порядком величины некоторой известной простой функции $g(x)$. В этом случае мы будем часто употреблять следующие обозначения.

1) Если $\frac{f(x)}{g(x)}$ остается ограниченным, когда x стремится к своему пределу, то мы пишем $f(x) = O(g(x))$ и говорим, что „ $f(x)$ имеет порядок не больший, чем $g(x)$ “.

2) Если $\frac{f(x)}{g(x)}$ стремится к нулю, то пишем $f(x) = o(g(x))$ и говорим, что „ $f(x)$ высшего порядка малости сравнительно с $g(x)$ “.

3) Если $\frac{f(x)}{g(x)}$ стремится к единице, то пишем $f(x) \sim g(x)$ и говорим, что „ $f(x)$ асимптотически равна $g(x)$ “.

Таким образом, при $x \rightarrow \infty$ имеем, например,

$$ax + b = O(x), \quad x^n = o(e^x), \quad \frac{x^8}{x + \log x} \sim x.$$

Символы $O(x)$, $o(1)$ и т. п. употребляются часто без отношения к определенной функции $f(x)$. Так, например, $O(x)$ будет обозначать любую функцию порядка не большего, чем x , $O(1)$ — любую ограниченную функцию, а $o(1)$ — любую функцию, стремящуюся к нулю.

В качестве примера рассмотрим еще функцию $f(x)$, имеющую в некоторой окрестности $x=0$ непрерывные производные до n -го порядка включительно. Тогда имеет место разложение Маклорена

$$f(x) = \sum_0^n \frac{f^{(v)}(0)}{v!} x^v + R_n(x),$$

где

$$R_n(x) = \frac{f^{(n)}(0)x - f^{(n)}(0)}{n!} x^n \quad (0 < \theta < 1).$$

По предположению, $f^{(n)}(0x) - f^{(n)}(0)$ стремится к нулю при $x \rightarrow 0$. Согласно сказанному выше, при $x \rightarrow 0$ можно писать

$$f(x) = \sum_0^n \frac{f^{(v)}(0)}{v!} x^v + o(x^n).$$

Это соотношение, которое справедливо и для функций $f(x)$, принимающих комплексные значения, уже было использовано в (10.1.3).

12.2. Формула Эйлера — Маклорена. Определим последовательность вспомогательных функций $P_1(x), P_2(x), \dots$ с помощью тригонометрических выражений

$$P_{2k}(x) = \sum_{v=1}^{\infty} \frac{\cos 2v\pi x}{2^{2k-1} (v\pi)^{2k}},$$

(12.2.1)

$$P_{2k+1}(x) = \sum_{v=1}^{\infty} \frac{\sin 2v\pi x}{2^{2k} (v\pi)^{2k+1}}.$$

Все эти функции периодические с периодом 1, так что

$$P_n(x+1) = P_n(x).$$

Для $n > 1$ ряд, представляющий $P_n(x)$, абсолютно и равномерно сходится для всех действительных x , так что $P_n(x)$ ограничена и непрерывна на всем интервале $(-\infty, \infty)$.

С другой стороны, ряд для $P_1(x)$ сходится лишь условно; хорошо известно, что $P_1(x) = -x + \frac{1}{2}$ при $0 < x < 1$. Обозначая через $[x]$ наибольшее целое число, не превосходящее x , и используя периодичность функции $P_1(x)$, получим для всех нецелых значений x

$$P_1(x) = [x] - x + \frac{1}{2}.$$

Таким образом, каждое целое число является точкой разрыва функции $P_1(x)$, и мы имеем $|P_1(x)| < \frac{1}{2}$ для всех x .

Для целых значений x имеем

$$P_{2k}(m) = \frac{1}{2^{2k-1} \pi^{2k}} \sum_{v=1}^{\infty} \frac{1}{v^{2k}} = (-1)^{k-1} \frac{B_{2k}}{(2k)!},$$

$$P_{2k+1}(m) = 0.$$

Числа B_v , появляющиеся здесь, суть числа *Бернулли*, определяемые соотношением

$$(12.2.2) \quad \frac{x}{ex-1} = \sum_{v=0}^{\infty} \frac{B_v}{v!} x^v.$$

Имеем

$$B_0 = 1, \quad B_1 = -\frac{1}{2}, \quad B_2 = \frac{1}{6}, \quad B_4 = -\frac{1}{30}, \quad B_6 = \frac{1}{42}, \dots$$

При нечетных $\nu \geq 3$ все B_ν равны нулю. Для $n > 1$ имеем

$$\frac{d}{dx} P_n(x) = (-1)^{n-1} P_{n-1}(x).$$

При $n > 2$ это соотношение справедливо для всех x , а при $n = 2$ — только для нецелых значений x .

Рассмотрим теперь функцию $g(x)$, непрерывную и имеющую непрерывную производную $g'(x)$ для всех x , заключенных в замкнутом интервале $(a + n_1 h, a + n_2 h)$, где a и $h > 0$ — постоянные, а n_1 и n_2 — положительные или отрицательные целые числа. Для любого целого ν , такого, что $n_1 \leq \nu < n_2$, интегрированием по частям находим

$$h \int_{\nu}^{n_2+1} P_1(x) g'(a + hx) dx = -\frac{1}{2} g(a + \nu h) - \frac{1}{2} g(a + (\nu + 1) h) + \\ + \int_{\nu}^{n_2+1} g(a + hx) dx.$$

Следовательно, суммируя по $\nu = n_1, \dots, n_2 - 1$, получим

$$(12.2.3) \quad \sum_{n_1}^{n_2} g(a + \nu h) = \int_{n_1}^{n_2} g(a + hx) dx + \frac{1}{2} g(a + n_1 h) + \\ + \frac{1}{2} g(a + n_2 h) - h \int_{n_1}^{n_2} P_1(x) g'(a + hx) dx.$$

Это простейший случай *формулы Эйлера — Маклорена*, которая часто бывает полезна при суммировании рядов. Если $g(x)$ имеет непрерывные производные высших порядков, то последний член может быть преобразован повторным интегрированием по частям, и мы получим общую формулу

$$(12.2.4) \quad \sum_{n_1}^{n_2} g(a + \nu h) = \int_{n_1}^{n_2} g(a + hx) dx + \frac{1}{2} g(a + n_1 h) + \\ + \frac{1}{2} g(a + n_2 h) - \sum_1^s \frac{B_{2\nu}}{(2\nu)!} h^{2\nu-1} [g^{(2\nu-1)}(a + n_1 h) - g^{(2\nu-1)}(a + n_2 h)] + \\ + (-1)^{s+1} h^{2s+1} \int_{n_1}^{n_2} P_{2s+1}(x) g^{(2s+1)}(a + hx) dx,$$

где s может быть любым неотрицательным целым числом; при этом предполагается, что все встречающиеся в формуле производные существуют и непрерывны.

Если ряд $\sum_{-\infty}^{\infty} g(a+hy)$ и интеграл $\int_{-\infty}^{\infty} g(a+hy) dx$ оба сходятся, то из формулы (12.2.3) получим

$$(12.2.5) \quad \sum_{-\infty}^{\infty} g(a+hy) = \int_{-\infty}^{\infty} g(a+hx) dx - h \int_{-\infty}^{\infty} P_1(x) g'(a+hx) dx,$$

где последний интеграл также должен сходиться. Если, кроме того, для $y=1, 2, \dots, s$, $g^{(2y-1)}(x) \rightarrow 0$ при $x \rightarrow \pm\infty$, то из (12.2.4) получим

$$(12.2.6) \quad \sum_{-\infty}^{\infty} g(a+hy) = \int_{-\infty}^{\infty} g(a+hx) dx + \\ + (-1)^{s+1} h^{2s+1} \int_{-\infty}^{\infty} P_{2s+1}(x) g^{(2s+1)}(a+hx) dx.$$

Если в (12.2.3) взять $g(x) = \frac{1}{x}$, $a=0$, $h=1$, $n_1=1$, $n_2=n$, то получим

$$\sum_1^n \frac{1}{y} = \log n + \frac{1}{2} + \frac{1}{2n} + \int_1^n \frac{P_1(x)}{x^2} dx.$$

Из определения $P_1(x)$ легко видеть, что

$$0 < \int_n^{\infty} \frac{P_1(x)}{x^2} dx < \frac{1}{8n^2},$$

так что имеем

$$(12.2.7) \quad \sum_1^n \frac{1}{y} = \log n + C + \frac{1}{2n} + O\left(\frac{1}{n^2}\right),$$

где

$$C = \frac{1}{2} + \int_1^{\infty} \frac{P_1(x)}{x^2} dx = 0,5772\dots$$

есть так называемая *постоянная Эйлера*.

12.3. Гамма-функция. Гамма-функция $\Gamma(p)$ определяется для всех действительных $p > 0$ соотношением

$$(12.3.1) \quad \Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx.$$

Согласно параграфу 7.3, эта функция непрерывна и имеет непрерывные производные всех порядков:

$$\Gamma^{(r)}(p) = \int_0^\infty x^{p-1} (\log x)^r e^{-x} dx$$

для любого $p > 0$. Если p стремится к 0 или к $+\infty$, то $\Gamma(p)$ стремится к $+\infty$. Так как вторая производная всегда положительна, то $\Gamma(p)$ имеет единственный минимум на $(0, \infty)$. Приближенное вычисление показывает, что минимум находится в точке $p_0 = 1,4616$, где функция принимает значение $\Gamma(p_0) = 0,8856$.

Интегрированием по частям получаем из (12.3.1) для любого $p > 0$

$$\Gamma(p+1) = p \Gamma(p).$$

Если p есть положительное целое число n , то повторное применение последнего равенства дает

$$\Gamma(n+1) = n!,$$

так как $\Gamma(1) = 1$.

Далее, из (12.3.1) получим

$$(12.3.2) \quad \int_0^\infty x^{\lambda-1} e^{-ax} dx = \frac{\Gamma(\lambda)}{a^\lambda},$$

где $a > 0$, $\lambda > 0$. Если заменить здесь a через $a+it$ и разложить множитель e^{-itx} в ряд, то можно показать, что последнее соотношение справедливо и для комплексных значений a в предположении, что действительная часть a положительна *).

Согласно (12.3.2), функция

$$(12.3.3) \quad f(x; a, \lambda) = \begin{cases} \frac{a^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-ax} & \text{при } x > 0 \\ 0 & \text{при } x \leq 0 \end{cases}$$

обладает относительно переменной x основными свойствами функции плотности (см. параграф 6.6): эта функция всюду неотрицательна, и ее интеграл по $(-\infty, \infty)$ равен 1. Соответствующее распределение играет важную роль в приложениях (см. например, параграфы 18.1 и 19.4). Его харак-

*) Читатель, знакомый с теоремой Коши об интегрировании функции комплексного переменного, может вывести (12.3.2) для комплексного a с помощью простого применения этой теоремы.

теристическая функция —

$$(12.3.4) \quad \int_{-\infty}^{\infty} e^{itx} f(x; a, \lambda) dx = \frac{a^\lambda}{\Gamma(\lambda)} \int_0^{\infty} x^{\lambda-1} e^{-(a-it)x} dx = \\ = \frac{a^\lambda}{\Gamma(\lambda)} \cdot \frac{\Gamma(\lambda)}{(a-it)^\lambda} = \frac{1}{\left(1 - \frac{it}{a}\right)^\lambda}.$$

12.4. Бэта-функция. Бэта-функция $B(p, q)$ определена для всех действительных $p > 0, q > 0$ соотношением

$$(12.4.1) \quad B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx.$$

Докажем важное соотношение

$$(12.4.2) \quad B(p, q) = \frac{\Gamma(p) \Gamma(q)}{\Gamma(p+q)}.$$

Интеграл

$$\int_0^{\infty} t^{p+q-1} x^{p-1} e^{-t(1+x)} dx = \Gamma(p) t^{q-1} e^{-t},$$

рассматриваемый как функция параметра t , удовлетворяет условиям теоремы III параграфа 7.3 на любом интервале (ε, ∞) при $\varepsilon > 0$, так что

$$\Gamma(p) \int_{\varepsilon}^{\infty} t^{q-1} e^{-t} dt = \int_0^{\infty} dx \int_{\varepsilon}^{\infty} t^{p+q-1} x^{p-1} e^{-t(1+x)} dt.$$

Когда ε стремится к нулю, левая часть стремится к $\Gamma(p) \Gamma(q)$. В правой части интеграл относительно t стремится, возрастаая, к пределу $\Gamma(p+q) \frac{x^{p-1}}{(1+x)^{p+q}}$, интегрируемому относительно x по $(0, \infty)$. Согласно (5.5.2), получим

$$\Gamma(p) \Gamma(q) = \Gamma(p+q) \int_0^x \frac{x^{p-1}}{(1+x)^{p+q}} dx.$$

Вводя новое переменное $y = \frac{x}{1+x}$, получим (12.4.2).

Полагая, в частности, в (12.4.2) $p = q$, получим, вводя новое переменное $y = 2x - 1$,

$$(12.4.3) \quad \frac{\Gamma^2(p)}{\Gamma(2p)} = \int_0^1 x^{p-1} (1-x)^{p-1} dx = 2^{2-2p} \int_0^1 (1-y^2)^{p-1} dy.$$

При $p = \frac{1}{2}$ это дает

$$\Gamma^2\left(\frac{1}{2}\right) = 2 \int_0^1 \frac{dy}{\sqrt{1-y^2}} = \pi, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

С другой стороны, полагая в (12.4.3) $y^2 = z$, получим

$$(12.4.4) \quad \frac{\Gamma^2(p)}{\Gamma(2p)} = 2^{1-2p} \int_0^1 (1-z)^{p-1} z^{-1/2} dz = 2^{1-2p} \frac{\Gamma(p)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(p + \frac{1}{2}\right)},$$

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

Если определить функцию $\beta(x; p, q)$ соотношением

$$(12.4.5) \quad \beta(x; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1} (1-x)^{q-1}$$

для $0 < x < 1$ и положить $\beta(x; p, q) = 0$ вне этого интервала, то из (12.4.1) и (12.4.2) будет следовать, что эта функция обладает основными свойствами функции плотности. Соответствующее распределение, при котором вся масса заключена в интервале $(0, 1)$, будет далее рассмотрено в параграфе 18.4.

12.5. Формула Стирлинга. Выведем теперь известную формулу Стирлинга, дающую *асимптотическое выражение для $\Gamma(p)$* при больших значениях p . Докажем сперва соотношение

$$(12.5.1) \quad \Gamma(p) = \lim_{n \rightarrow \infty} \frac{n! n^p}{p(p+1)\dots(p+n)}$$

при любом $p > 0$.

Повторным интегрированием по частям находим

$$\int_0^n x^{p-1} \left(1 - \frac{x}{n}\right)^n dx = \frac{n! n^p}{p(p+1)\dots(p+n)}.$$

Левая часть этого соотношения может быть записана в виде

$$\int_0^\infty g(x, n) dx, \quad \text{где } g(x, n) = x^{p-1} \left(1 - \frac{x}{n}\right)^n \text{ при } 0 < x < n \text{ и } g(x, n) = 0$$

при $x \geq n$. Когда n стремится к бесконечности, $g(x, n)$ стремится к $x^{p-1} e^{-x}$ для любого $x > 0$, и легко видеть, что всегда $0 \leq g(x, n) < x^{p-1} e^{-x}$. Отсюда, в силу (5.5.2), получим (12.5.1).

Из (12.5.1) следует, что $\log \Gamma(p) = \lim_{n \rightarrow \infty} S_n$, где

$$S_n = p \log n + \sum_1^n \log \nu - \sum_0^n \log(p + \nu).$$

Применяя формулу Эйлера — Маклорена (12.2.3) к обеим суммам в этом выражении, получим после некоторых упрощений

$$\begin{aligned} S_n &= \left(p - \frac{1}{2}\right) \log p - \left(p + n + \frac{1}{2}\right) \log \left(1 + \frac{p}{n}\right) + \\ &\quad + 1 - \int_1^n \frac{P_1(x)}{x} dx + \int_0^n \frac{P_1(x)}{p+x} dx. \end{aligned}$$

Когда n стремится к бесконечности, второе слагаемое в правой части стремится к $-p$, а оба интеграла сходятся (хотя и не абсолютно), так как $P_1(x)$ последовательно меняет знак. Таким образом, получим

$$(12.5.2) \quad \log \Gamma(p) = \left(p - \frac{1}{2}\right) \log p - p + k + R(p),$$

где k — постоянная, а остаточный член $R(p)$ выражается так:

$$R(p) = \int_0^\infty \frac{P_1(x)}{p+x} dx.$$

Этот интеграл может быть преобразован повторным интегрированием по частям, как показано в (12.2.4), и мы, таким образом, получим

$$R(p) = \sum_1^s \frac{B_{2s}}{2s(2s-1)p^{2s-1}} + (-1)^s (2s)! \int_0^\infty \frac{P_{2s+1}(x)}{(p+x)^{2s+1}} dx$$

для $s = 0, 1, 2, \dots$. Интеграл в этом равенстве абсолютно сходится при любом $s > 0$, и его модуль меньше, чем

$$A \int_p^\infty \frac{dx}{x^{2s+1}} = \frac{A}{2sp^{2s}},$$

где A — некоторая постоянная. Отсюда, в частности, следует, что $R(p) \rightarrow 0$ при $p \rightarrow \infty$.

Для того чтобы найти значение постоянной k в равенстве (12.5.2), заметим, что, согласно (12.4.4),

$$\log \Gamma(2p) = \log \Gamma(p) + \log \Gamma\left(p + \frac{1}{2}\right) + (2p - 1) \log 2 - \frac{1}{2} \log \pi.$$

Заменив здесь гамма-функции их выражениями, полученными из (12.5.2), и заставляя p стремиться к бесконечности, мы находим после некоторых упрощений

$$k = \frac{1}{2} \log 2\pi.$$

Таким образом, нами доказана *формула Стирлинга*:

$$(12.5.3) \quad \log \Gamma(p) = \left(p - \frac{1}{2}\right) \log p - p + \frac{1}{2} \log 2\pi + R(p),$$

где

$$R(p) = \int_0^\infty \frac{P_1(x)}{p+x} dx = \frac{1}{12p} + O\left(\frac{1}{p^3}\right) = \frac{1}{12p} - \frac{1}{360p^3} + O\left(\frac{1}{p^5}\right) = \dots$$

Из формулы Стирлинга мы выводим, в частности, асимптотическое выражение

$$n! = \Gamma(n+1) \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$$

и, далее, если $p \rightarrow \infty$, а h фиксировано,

$$\frac{\Gamma(p+h)}{\Gamma(p)} \sim p^h.$$

Дифференцируя, получим из формулы Стирлинга

$$(12.5.4) \quad \begin{aligned} \frac{\Gamma'(p)}{\Gamma(p)} &= \log p - \frac{1}{2p} - \int_0^\infty \frac{P_1(x)}{(p+x)^2} dx, \\ \frac{\Gamma''(p)}{\Gamma(p)} - \left(\frac{\Gamma'(p)}{\Gamma(p)}\right)^2 &= \frac{1}{p} + \frac{1}{2p^2} + 2 \int_0^\infty \frac{P_1(x)}{(p+x)^3} dx. \end{aligned}$$

При $p = 1$ первое равенство дает

$$(12.5.5) \quad \Gamma'(1) = -\frac{1}{2} - \int_0^\infty \frac{P_1(x)}{(1+x)^2} dx = -\frac{1}{2} - \int_1^\infty \frac{P_1(x)}{x^2} dx = -C,$$

где C — постоянная Эйлера, определенная равенством (12.2.7). Дифференцируя равенство $\Gamma(p+1) = p \Gamma(p)$, получим, далее,

$$\frac{\Gamma'(p+1)}{\Gamma(p+1)} = \frac{1}{p} + \frac{\Gamma'(p)}{\Gamma(p)},$$

и, следовательно, для целых значений p

$$(12.5.6) \quad \frac{\Gamma'(n)}{\Gamma(n)} = 1 + \frac{1}{2} + \dots + \frac{1}{n-1} - C.$$

Применение формулы Эйлера — Маклорена (12.2.3) дает.

$$\sum_n \frac{1}{n^3} = \frac{1}{n} + \frac{1}{2n^2} + 2 \int_n^\infty \frac{P_1(x)}{x^3} dx.$$

Полагая во втором равенстве (12.5.4) $p = n$, мы, таким образом, получим

$$(12.5.7) \quad \frac{\Gamma''(n)}{\Gamma(n)} - \left(\frac{\Gamma'(n)}{\Gamma(n)} \right)^2 = \sum_n \frac{1}{n^3} = \frac{\pi^2}{6} - \sum_1^{n-1} \frac{1}{n^3}.$$

12.6. Ортогональные полиномы. Пусть $F(x)$ — некоторая функция распределения с конечными моментами (см. параграф 7.4) a , всех порядков. Скажем, что x_0 есть *точка возрастания* функции $F(x)$, если $F(x_0 + h) > F(x_0 - h)$ при любом $h > 0$.

Предположим сперва, что множество точек возрастания функции F бесконечно. Покажем, что тогда существует последовательность полиномов $p_0(x), p_1(x), \dots$, однозначно определенных следующими условиями:

а) Полином $p_n(x)$ имеет степень n , и коэффициент при x^n положителен.

б) Полиномы $p_n(x)$ удовлетворяют условиям ортогональности

$$\int_{-\infty}^{\infty} p_m(x) p_n(x) dF(x) = \begin{cases} 1 & \text{при } m = n, \\ 0 & \text{при } m \neq n. \end{cases}$$

Полиномы $p_n(x)$ называются *ортогональными полиномами*, связанными с распределением, соответствующим функции $F(x)$.

Заметим, что при любом $n \geq 0$ квадратичная форма $n+1$ переменных u_0, u_1, \dots, u_n

$$\int_{-\infty}^{\infty} (u_0 + u_1 x + \dots + u_n x^n)^2 dF(x) = \sum_{i,k=0}^n a_{i+k} u_i u_k$$

является положительно определенной. Так как, согласно предположению, $F(x)$ имеет не меньше $n+1$ точек возрастания и по крайней мере одна из этих точек отлична от всех n нулей полинома $u_0 + u_1 x + \dots + u_n x^n$, то этот интеграл всегда положителен, если только

не все a_i равны нулю. Отсюда следует (см. параграф 11.10), что детерминант рассматриваемой формы положителен:

$$D_n = \begin{vmatrix} a_0 & a_1 & \dots & a_n \\ a_1 & a_2 & \dots & a_{n+1} \\ \dots & \dots & \dots & \dots \\ a_n & a_{n+1} & \dots & a_{2n} \end{vmatrix} > 0.$$

Очевидно, $p_0(x) = 1$. Теперь положим

$$p_n(x) = a_0 + a_1 x + \dots + a_n x^n,$$

где $n > 0$, и попытаемся определить коэффициенты a_i из условий a) и b). Так как полином $p_i(x)$ имеет степень i , то x^i может быть представлено как линейная комбинация полиномов $p_0(x), \dots, p_i(x)$. Следовательно,

$$\int_{-\infty}^{\infty} x^i p_n(x) dF(x) = 0$$

при $i = 0, 1, \dots, n-1$. Выполняя интегрирование, мы получим, таким образом, n линейных однородных уравнений с $n+1$ неизвестными a_0, \dots, a_n ; следовательно, любой полином $p_n(x)$, удовлетворяющий нашим условиям, должен иметь вид

$$(12.6.1) \quad p_n(x) = K \begin{vmatrix} a_0 & a_1 & \dots & a_n \\ \cdot & \cdot & \cdot & \cdot \\ a_{n-1} & a_n & \dots & a_{2n-1} \\ 1 & x & \dots & x^n \end{vmatrix},$$

где K — некоторая постоянная. При $K \neq 0$ этот полином имеет степень n , так как коэффициент при x^n равен $D_{n-1} > 0$. Таким образом, полином $p_n(x)$ однозначно определен условием $\int p_n^2 dF = 1$ и тем, что коэффициент при x^n должен быть положительным*). Таким образом, мы установили существование однозначно определенной последовательности ортогональных полиномов, соответствующих распределению с бесконечным числом точек возрастания.

Если $F(x)$ имеет только N точек возрастания, то из приведенного выше доказательства легко вытекает, что полиномы $p_n(x)$ существуют

*) Можно показать, что $K = (D_{n-1} D_n)^{-\frac{1}{2}}$. См., например, Сеге [36].

и однозначно определены для $n=0, 1, \dots, N-1$. Детерминанты D_n в этом случае положительны при $n=0, 1, \dots, N-1$, но при $n \geq N$ мы имеем $D_n = 0$.

Рассмотрим, в частности, распределение с непрерывной плотностью вероятности $f(x) = F'(x)$, и пусть $p_0(x), p_1(x), \dots$ — соответствующие ортогональные полиномы. Если $g(x)$ — какая-нибудь другая функция плотности, то мы можем попытаться разложить $g(x)$ в ряд

$$(12.6.2) \quad g(x) = b_0 p_0(x) f(x) + b_1 p_1(x) f(x) + \dots$$

Умножим это соотношение на $p_n(x)$ и предположим, что можно интегрировать (12.6.2) почленно. Тогда соотношения ортогональности дают

$$(12.6.3) \quad b_n = \int_{-\infty}^{\infty} p_n(x) g(x) dx.$$

Отсюда, в частности, $b_0 = 1$. Разложения такого вида иногда бывают полезны для аналитического представления распределений.

Дадим теперь некоторые примеры ортогональных полиномов.

1. Полиномы Эрмита $H_n(x)$ определяются соотношениями

$$(12.6.4) \quad \left(\frac{d}{dx}\right)^n e^{-\frac{x^2}{2}} = (-1)^n H_n(x) e^{-\frac{x^2}{2}} \quad (n=0, 1, 2, \dots).$$

$H_n(x)$ есть полином степени n ; имеем

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1,$$

$$(12.6.5) \quad \begin{aligned} H_3(x) &= x^3 - 3x, \quad H_4(x) = x^4 - 6x^2 + 3, \\ H_5(x) &= x^5 - 10x^3 + 15x, \quad H_6(x) = x^6 - 15x^4 + 45x^2 - 15, \\ &\dots \end{aligned}$$

Повторным интегрированием по частям получим

$$(12.6.6) \quad \int_{-\infty}^{\infty} H_m(x) H_n(x) d\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H_m(x) H_n(x) e^{-\frac{x^2}{2}} dx = \\ = \begin{cases} n! & \text{при } m=n, \\ 0 & \text{при } m \neq n, \end{cases}$$

откуда видно, что $\left\{ \frac{1}{\sqrt{n!}} H_n(x) \right\}$ есть последовательность ортогональных полиномов, связанных с нормальным распределением (см. (10.5.3)). Отметим также разложения

$$(12.6.7) \quad \sum_0^{\infty} \frac{H_n(x)}{n!} t^n = e^{-\frac{t^2}{2} + tx}$$

$$(12.6.8) \quad \sum_0^{\infty} \frac{H_n(x) H_n(y)}{n!} t^n = \frac{1}{\sqrt{1-t^2}} e^{\frac{tx^2+ty^2-2t\bar{x}\bar{y}}{2(1-t^2)}}, \quad (|t| < 1).$$

Первое из этих разложений легко вытекает из определения (12.6.4). Доказательство второго, данное Крамером, можно найти в книге Шарлье [9а], стр. 50—53.

2. Полиномы Лагерра $L_n^{(\lambda)}(x)$ определяются соотношениями

$$\left(\frac{d}{dx}\right)^n (x^n + \lambda - 1)e^{-x} = (-1)^n n! L_n^{(\lambda)}(x) x^{\lambda-1} e^{-x},$$

которые дают

$$L_0(x) = 1, \quad L_1(x) = x - \lambda, \quad L_2(x) = \frac{x^2 - 2(\lambda + 1)x + \lambda(\lambda + 1)}{2}, \dots$$

Повторным интегрированием по частям находим

$$\frac{1}{\Gamma(\lambda)} \int_0^{\infty} L_m^{(\lambda)}(x) L_n^{(\lambda)}(x) x^{\lambda-1} e^{-x} dx = \begin{cases} \binom{n+\lambda-1}{n} & \text{при } m=n, \\ 0 & \text{при } m \neq n, \end{cases}$$

так что $\left\{ \frac{L_n^{(\lambda)}(x)}{\sqrt{\binom{n+\lambda-1}{n}}} \right\}$ есть последовательность ортогональных полиномов, связанных с распределением, определяемым функцией плотности $f(x; a, \lambda)$ (см. (12.3.3)) при $a=1$.

3. Рассмотрим распределение, которое получим, поместив массы $\frac{1}{N}$ в каждую из N точек x_1, x_2, \dots, x_N . Соответствующая функция распределения есть ступенчатая функция со скачками высоты $\frac{1}{N}$ в каждой точке x_i . Пусть $p_0(x), \dots, p_{N-1}(x)$ — ортогональные полиномы, связанные с этим распределением. Согласно сказанному выше, они определены однозначно. Соотношения ортогональности сводятся к

$$\frac{1}{N} \sum_{i=1}^N p_m(x_i) p_n(x_i) = \begin{cases} 1 & \text{при } m=n, \\ 0 & \text{при } m \neq n. \end{cases}$$

Этими полиномами удобно воспользоваться, например, при решении следующей проблемы. Предположим, что мы имеем N наблюдаемых точек $(x_1, y_1), \dots, (x_N, y_N)$ и хотим найти параболу $y = q(x)$ степени $n < N$, дающую наилучшее приближение к наблюденным ординатам в смысле принципа наименьших квадратов, т. е. такую, что

$$U = \frac{1}{N} \sum_{i=1}^N (y_i - q(x_i))^2$$

имеет наименьшее значение. Записываем $q(x)$ в виде

$$q(x) = c_0 p_0(x) + \dots + c_n p_n(x)$$

Обычное правило нахождения минимума непосредственно дает

$$c_r = \frac{1}{N} \sum_{i=1}^N y_i p_r(x_i) .$$

для $r = 0, 1, \dots, n$, а соответствующее значение $\min U$ есть

$$\min U = \frac{1}{N} \sum_{i=1}^N y_i^2 - c_0^2 - c_1^2 - \dots - c_n^2.$$

Случай, когда точки x_i — равноотстоящие, особенно важен в приложениях. В этом случае численные значения $q(x)$ и $\min U$ могут быть найдены сравнительно легко. См., например, Эшер [82] и Эйткен [50]. См. ниже теорию параболической регрессии в параграфе 21.6.

ЧАСТЬ ВТОРАЯ

**СЛУЧАЙНЫЕ ВЕЛИЧИНЫ
И РАСПРЕДЕЛЕНИЯ
ВЕРОЯТНОСТЕЙ**

ГЛАВА 13

СТАТИСТИКА И ВЕРОЯТНОСТЬ

13.1. Случайные эксперименты. В самых разнообразных областях практической и научной деятельности встречаются случаи, когда некоторые эксперименты или наблюдения могут быть повторены большое число раз при одинаковых условиях. В каждом таком случае наше внимание бывает сосредоточено на *результате наблюдения*, выражаемом некоторым числом характеристических признаков.

Во многих случаях эти характеристики имеют непосредственно количественный характер: при каждом наблюдении нечто вычисляется или измеряется. В других случаях характеристики являются качественными: наблюдался, например, цвет какого-либо объекта, наступление или ненаступление некоторого события в связи с каждым экспериментом, и т. д. В последнем случае характеристики всегда можно выразить в количественной форме в некоторой условной системе обозначений. Всегда, когда это удобно, мы можем предполагать, что результат каждого наблюдения выражается некоторым количеством величин.

1. Если совершается ряд бросаний обычной игральной кости, то каждое бросание дает в результате одно из чисел 1, 2, ..., 6.

2. Если измеряются длина и вес тела каждого представителя из некоторой группы животных одного и того же вида, то в результате наблюдения над каждым индивидом получаются два числа.

3. Если производятся выборки из ежедневной продукции сталелитейного завода и измеряются твердость стали, сопротивление на разрыв и содержание углерода, серы и фосфора, то результат каждого наблюдения выражается пятью числами.

4. Если наблюдаются цены k различных товаров в определенные промежутки времени, то результат каждого наблюдения выражается k числами.

5. Если регистрируется пол каждого ребенка, родившегося в данном районе, то результат каждого наблюдения непосредственно

числами не выражается. Однако, можно условиться обозначать рождение мальчика единицей, а рождение девочки — нулем, тогда наши результаты будут выражены в числах.

В некоторых случаях наблюдаемое явление изучено настолько хорошо, что можно считать оправданным точное предсказание результата каждого отдельного наблюдения. Так, например, если наши эксперименты заключаются в ежегодном определении числа затмений солнца, наблюдавшихся в данной обсерватории, то мы можем на основании астрономических вычислений без колебаний предсказать точное значение этого числа на ближайший год.

Подобное положение имеет место в каждом случае, когда предполагается, что законы, управляющие явлением, известны и при этом достаточно просты, чтобы их можно было использовать для практических вычислений.

Однако в большинстве случаев наши знания недостаточно точны для того, чтобы можно было предсказывать результаты отдельных наблюдений. Так обстоит дело в примерах 1—5, приведенных выше. Даже если приняты все меры к тому, чтобы контролировать условия, влияющие на эксперимент, результат эксперимента может при этом меняться от одного наблюдения к другому самым неправильным образом, сводя на нет все наши попытки предсказания результата. В таких случаях мы будем говорить, что имеем дело с последовательностью *случайных экспериментов*.

Каждая систематическая сводка результатов подобной последовательности экспериментов будет образовывать *множество статистических данных*, относящихся к изучаемому явлению. Основным объектом статистической теории является исследование возможности *получить надежные выводы из статистических данных* и выработка методов, с помощью которых эти выводы могут быть получены.

Прежде чем изучать эти вопросы, в ближайших двух параграфах мы предварительно рассмотрим некоторые общие свойства случайных экспериментов.

13.2. Примеры. Повидимому, невозможно дать точное определение того, что подразумевается под словом „случайный“. Смысл этого слова лучше всего разъяснить на примерах.

Если многократно бросать обыкновенную монету, стараясь сохранять условия эксперимента по возможности во всех отношениях не-

изменными, то обнаружится, что мы не в состоянии предсказать, упадет ли монета при данном бросании вверх гербом или решеткой. Если в первом бросании выпал герб, и в следующем бросании мы старались придать монете то же самое начальное положение и движение, то окажется, тем не менее, что невозможно обеспечить последующее выпадение герба.

Даже если мы постараемся построить машину, бросающую монету каждый раз совершенно одинаково, то все же невероятно, чтобы нам удалось предсказывать результаты отдельных бросаний. Наоборот, результат эксперимента всегда будет меняться неконтролируемым образом от одного бросания к другому.

Сначала может показаться, что такое явление трудно объяснить. Если мы станем на точку зрения детерминизма, то мы должны будем утверждать, что результат каждого бросания однозначно определяется начальным положением и движением монеты (при фиксированных внешних условиях, таких, как сопротивление воздуха и физические свойства стола). Поэтому может показаться, что теоретически возможно дать точное предсказание результата отдельного бросания монеты, коль скоро известно ее начальное состояние, и что возможно получить любой желаемый результат, задавшись соответствующим начальным состоянием. Минутное размышление покажет, однако, что даже ничтожно малое изменение начального состояния может оказать решающее влияние на результат. Практически начальное состояние никогда не известно точно, а лишь с некоторым приближением. Таким образом, если мы и попытаемся поддерживать полную неизменность начальных состояний в течение ряда бросаний монеты, мы все же никогда не сможем исключить небольшие отклонения, величина которых зависит от точности механизма, осуществляющего бросания. В пределах точности приближения всегда будет находиться множество различных начальных состояний, приводящих к обоим возможным результатам — выпадению герба или решетки; поэтому точное предсказание результата отдельного бросания практически невозможно.

Такие же замечания можно высказать и относительно ряда других аналогичных случаев: бросаний кости в примере 1 предыдущего параграфа и вообще относительно всех обычных азартных игр в кости или карты.

Предположим, далее, что мы выделили некоторое количество людей одного и того же возраста и по прошествии некоторого времени,

например, одного года, регистрируем, живо ли данное лицо или умерло за истекшее время. Предположим, что мы можем получить подробные сведения о состоянии здоровья каждого наблюдаемого лица, а также о его занятиях, привычках и т. д. Тем не менее очевидно, что точно предсказать смерть отдельного человека невозможно, так как причины, определяющие тот или иной исход, слишком многочисленны и сложны, чтобы их можно было точно учесть. Даже наблюдатель, располагающий гораздо более совершенными биологическими знаниями, чем те, которые доступны нам в настоящее время, придет к такому же практическому выводу, вследствие многочисленности и сложности причин в рассматриваемом явлении.

Примеры 2 и 4 предыдущего параграфа обнаруживают большую аналогию с рассмотренным сейчас примером. Управляющие явления законы не известны достаточно хорошо ни в том, ни в другом случае, и даже если бы они были известны значительно лучше, чем в настоящее время, природа каждого явления столь сложна, что индивидуальное предсказание все равно было бы практически невозможно. В этих и во многих других аналогичных случаях результаты наблюдений обнаруживают такую же нерегулярность от случая к случаю, как и в рассмотренных выше примерах.

Важно отметить, что такое положение может истретиться даже в тех случаях, когда мы можем считать законы явления полностью известными, если только эти законы достаточно сложны. Возьмем, например, явление солнечного затмения, упомянутое в предыдущем параграфе. Мы допускаем, что предсказание ежегодного числа затмений возможно, и, при наличии необходимых таблиц, подобное предсказание может осуществить каждый. Однако если таблиц нет, то производство необходимых вычислений было бы чрезвычайно трудной задачей. Если эту трудность считать непреодолимой, то предсказание окажется практически невозможным, и колебания числа затмений за год будут казаться похожими на колебания результатов в серии азартных игр.

Предположим, наконец, что наш эксперимент состоит в ряде повторных измерений некоторой физической константы, причем метод измерения и влияющие на эксперимент внешние условия поддерживаются неизменными, насколько это возможно, в течение всего ряда измерений. Хорошо известно, что, несмотря на все предосторожности, принятые наблюдателем, последовательные измерения дают, вообще говоря, различные результаты. Это явление обычно приписывается действию боль-

шого количества малых возмущающих факторов, общий эффект которых создает некоторую суммарную „ошибку“ в каждом отдельном измерении. Величина этой ошибки колеблется от одного наблюдения к другому неправильным образом, что делает невозможным предсказание результата в каждом отдельном измерении.

Подобные же замечания применимы также к колебаниям качества продукции, как в примере 3 предыдущего параграфа. Малые и неконтролируемые изменения производственного процесса и колебания качества сырья в совокупности создают неправильные колебания свойств окончательной продукции.

Рассмотренные выше примеры являются представителями большой и важной группы случайных экспериментов. Малые изменения начального состояния наблюдаемых объектов, которые не удается обнаружить нашими инструментами, могут создать значительные изменения окончательного результата. Сложный характер законов наблюдаемого явления также может сделать точное вычисление практически, если не теоретически, невозможным. Наконец, не поддающееся контролю воздействие малых возмущающих факторов может привести к неправильным отклонениям от гипотетического „истинного значения“.

Ясно, конечно, что между этими видами случайности нет резкого различия. По большей части, дело вкуса — относить ли колебания, например, в результатах обстрела цели, главным образом за счет малых изменений начального состояния снаряда, сложной природы баллистических законов или за счет малых возмущающих факторов. Сущность дела состоит в том, что во всех случаях, когда мы находимся в тех или иных условиях такого рода, точное предсказание результатов индивидуальных экспериментов оказывается невозможным, и обнаруживаются неправильные колебания, характерные для случайного эксперимента.

Сейчас мы увидим, что в такого рода случаях среди всех неправильностей в колебаниях результатов экспериментов обнаруживается некоторая характерная закономерность, которая и будет служить базисом для математической теории статистики.

13.3. Статистическая устойчивость. Мы видели, что в последовательности случайных экспериментов невозможно предсказывать индивидуальные результаты, так как в этих результатах обнаруживаются неправильные случайные колебания, не поддающиеся точному учету.

Однако, как только мы перенесем свое внимание с индивидуальных экспериментов на *последовательность экспериментов* в целом, положение коренным образом изменится, и обнаружится чрезвычайно важное явление: *Несмотря на неправильное поведение индивидуальных результатов, средние результаты достаточно длинной последовательности случайных экспериментов обнаруживают поразительную устойчивость.*

Чтобы пояснить эту важную закономерность, рассмотрим определенный случайный эксперимент \mathfrak{E} , который может быть повторен большое количество раз при одинаковых условиях. Пусть S —множество всех возможных *a priori* различных результатов индивидуального эксперимента, а S —некоторое его фиксированное подмножество *). Если в некотором отдельном эксперименте получен результат ξ , принадлежащий подмножеству S , то будем говорить, что произошло *событие*, определяемое соотношением $\xi \subset S$, или, короче, *событие* $\xi \subset S$. Мы будем часто обозначать событие одной буквой E и записывать $E = E (\xi \subset S)$; поэтому не будем делать различия между „событием E “ и „событием $\xi \subset S$ “.

Если наш эксперимент \mathfrak{E} состоит в бросании игральной кости, то множество S содержит шесть чисел 1, 2, ..., 6.

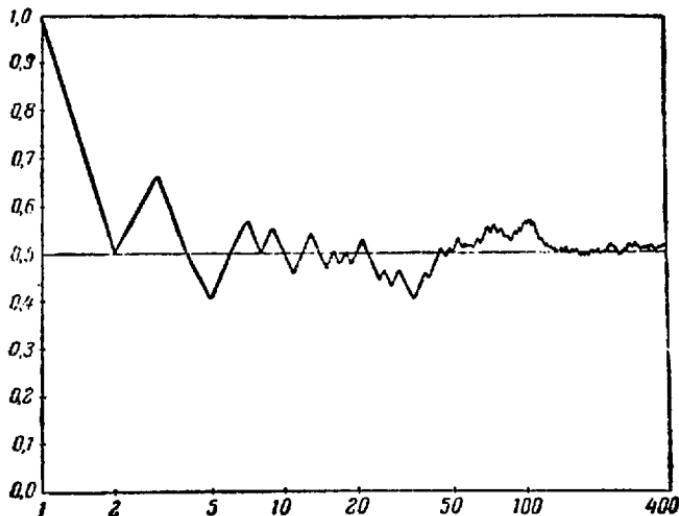
Пусть S обозначает, например, подмножество, состоящее из трех чисел 2, 4, 6. Тогда событие $\xi \subset S$ произойдет в каждом бросании кости, в котором выпадает четное число.

Если мы занимаемся измерением некоторой физической константы x , значение которой *a priori* совершенно неизвестно, то, по крайней мере теоретически, результатом измерения может оказаться любое действительное число. Тогда S будет одномерным пространством R_1 . Пусть S обозначает, например, замкнутый интервал (a, b) . Тогда событие $\xi \subset S$ будет происходить всякий раз, когда измерение дает значение ξ , такое, что $a \leq \xi \leq b$.

Будем повторять наш эксперимент \mathfrak{E} большое число раз, всякий раз наблюдая, происходит ли событие $E = E (\xi \subset S)$. Если мы обнаружим, что в первых n экспериментах событие E произошло ровно y раз, то отношение $\frac{y}{n}$ будет называться *частотой* события E в последовательности, образованной первыми n экспериментами.

*) Здесь мы предполагаем, что S —некоторое множество простой структуры, так что можно непосредственно наблюдать, принадлежит ξ к S или нет. В следующей главе этот вопрос будет рассмотрен с более общей точки зрения.

При наблюдении частоты $\frac{v}{n}$ фиксированного события E для возрастающих значений n обнаруживается, что эта частота, вообще говоря, имеет тенденцию принимать при больших значениях n более или менее постоянное значение.



Фиг. 3. Частота выпадения „герба“ в последовательности бросаний монеты. (Логарифмическая шкала по оси абсцисс).

Иллюстрацией к этому явлению служит фиг. 3, на которой изображены изменения частоты $\frac{v}{n}$ события „выпадение герба“ в последовательности бросаний монеты. Как видно из графика, частота сильно колеблется при малых значениях n , но постепенно амплитуда колебаний становится меньше, и из графика может получиться впечатление, что, если бы ряд экспериментов был неограниченно продолжен при неизменных условиях, то частота приближалась бы к некоторому определенному идеальному или предельному значению, очень близкому к $\frac{1}{2}$.

Из многочисленных опытов известно, что описанная *устойчивость частоты* обычно наблюдается в длинном ряду повторяющихся при неизменных условиях случайных экспериментов. Для события типа $\xi \subset S$, наблюдающегося в связи с таким рядом экспериментов, как правило, будет получаться график, вообще говоря, такого же характера, как

и в частном случае, иллюстрированном на фиг. 3. Более того, в тех случаях, когда это утверждение не оправдывается, тщательное исследование обычно обнаруживает, что неизменность условий эксперимента так или иначе нарушена. Мы вправе поэтому выдвинуть предположение, что, вообще говоря, частота рассматриваемого здесь типа приближалась бы к спределенному идеальному значению, если бы соответствующий ряд экспериментов можно было неограниченно продолжить.

Предположение такого рода, конечно, не может быть ни доказано, ни опровергнуто фактическими опытами, так как мы никогда не можем осуществить бесконечную последовательность экспериментов. Однако опыты подтверждают менее точное предположение, что для каждого события E , связанного со случайным экспериментом \mathcal{E} , можно указать такое число P , что в длинном ряду повторений эксперимента \mathcal{E} частота события E окажется приблизительно равной P .

Такова типичная форма *статистической устойчивости*, служащей эмпирической основой статистики. Теперь мы должны попытаться придать точный смысл тем несколько неопределенным выражениям, которые встречались в приведенном выше утверждении, исследовать законы, управляющие подобной статистической устойчивостью, и показать, как эти законы могут применяться для извлечения выводов из эмпирического материала. Для решения этой задачи мы должны прежде всего постараться построить *математическую теорию явлений*, в которых обнаруживается статистическая устойчивость. Предварительно полезно будет привести в следующем параграфе некоторые общие замечания о природе и предмете любой математической теории, описывающей некоторую группу эмпирических явлений.

Описанное выше замечательное свойство частот впервые было замечено в области азартных игр, простейшим частным случаем которых служит наш пример с бросанием монеты. Уже давно было подмечено, что во всех распространенных играх в карты, кости и т. п. частота определенного исхода данной игры группируется вблизи некоторого числа, если игра повторяется много раз. Попытки дать математическое толкование некоторым обнаруженным фактам такого рода послужили непосредственной причиной возникновения (около 1650 г.) и развития математической теории вероятностей в трудах Паскаля, Ферма, Гюйгенса и Якова Бернулли. Позднее такого же рода устойчивость была обнаружена для частот, связанных с различными демографическими данными, и на этом была основана статистическая теория народонаселения. Область применения статистических методов посте-

ленно расширялась, и в настоящее время можно считать эмпирически установленным, что „устойчивость длинного ряда“ частот есть общая характеристика случайных экспериментов, производящихся при неизменных условиях.

В некоторых случаях, особенно тогда, когда наблюдению подвергаются совокупности людей (или других биологических особей), статистическую устойчивость часто интерпретируют, рассматривая такие совокупности как *выборки* из некоторой очень обширной или даже бесконечной *генеральной совокупности*.

Рассмотрим сначала случай конечной генеральной совокупности, состоящей из N индивидуумов. Для каждого наблюдаемого индивидуума отметим некоторую характеристику ξ и обозначим через E некоторое специальное событие вида $\xi \in S$. Частота события E в выборке из n наблюденных индивидуумов стремится при возрастании размеров выборки к частоте события E в полной совокупности и фактически достигает этого значения, если положить $n = N$, т. е. если наблюдаются все индивидуумы в полной совокупности.

Понятие *бесконечной генеральной совокупности* есть математическая абстракция, как и представление о том, что данный случайный эксперимент можно повторить бесконечное число раз. Бесконечную совокупность можно рассматривать, как предельный случай конечной совокупности, когда число N всех индивидуумов неограниченно возрастает. Частота события E в выборке из n индивидуумов из бесконечной совокупности всегда подвержена случайным колебаниям, пока n конечно, но кажется естественным предположение, что при неограниченном возрастании числа n эта частота в конце концов достигнет „истинного“ значения, соответствующего частоте события E в полной бесконечной совокупности.

Такая интерпретация, посредством понятия выборки, может быть распространена на любые случайные эксперименты. Действительно, каждую конечную последовательность повторений случайного эксперимента можно представлять себе, как выборку из гипотетической бесконечной совокупности всех осуществимых при данных условиях экспериментов. Мы вернемся к этому в главе 25, где понятие о выборке будет подвергнуто дальнейшему исследованию.

13.4. Объект математической теории. Если в некоторой группе доступных наблюдению явлений обнаруживается определенная закономерность, то можно попытаться построить математическую теорию этих явлений. Такую теорию можно считать *математической моделью* запаса эмпирических фактов — данных, полученных наблюдениями.

В качестве отправной точки мы выбираем некоторые из наиболее существенных и наиболее простых черт той закономерности, которая обнаружена в полученных данных. Эти признаки мы выражаем в упрощении.

щенной и идеализированной форме как математические предложения,— аксиомы, которые мы кладем в основу нашей теории. Из аксиом получаются различные теоремы с помощью чистой логической дедукции, без каких-либо ссылок на опыт. Логически непротиворечивая система утверждений, построенная таким способом на аксиоматической основе, образует нашу математическую теорию.

Двумя классическими примерами такого построения являются геометрия и теоретическая механика. Геометрия является системой чисто математических предложений, служащей математической моделью большой группы эмпирических фактов, связанных с положением и конфигурацией различных тел в пространстве. Геометрия ограничивается сравнительно небольшим количеством аксиом, введенных без доказательства. Как только аксиомы выбраны, полная система геометрических теорем получается из них чистой логической дедукцией. При выборе аксиом мы руководствуемся закономерностями, обнаруженными в эмпирических данных, которыми мы располагаем. Однако, аксиомы могут быть выбраны различным образом, и, соответственно, существуют различные геометрии: Эвклида, Лобачевского и т. д. Каждая из этих геометрий есть логически непротиворечивая система математических предложений, опирающаяся на свою систему аксиом. Аналогичным образом, теоретическая механика есть система математических теорем, служащая математической моделью наблюдаемых фактов, связанных с равновесием и движением тел.

Каждое предложение в такого рода системе *истинно*, в математическом смысле этого слова, если оно правильно выведено из соответствующих аксиом. С другой стороны, важно подчеркнуть, что никакое предложение любой математической теории *не доказывает* ничего относительно явлений, имеющих место в действительности. Точки, линии, плоскости и т. д., рассматриваемые в абстрактной геометрии, не являются предметами, воспринимаемыми нами из непосредственного опыта. Чистая теория принадлежит целиком к области понятий и имеет дело с абстрактными объектами, которые полностью определяются своими свойствами, выраженными посредством аксиом. Для этих объектов предложения теории точны и истинны. Но никакое предложение относительно этих идеальных объектов никогда не содержит логического доказательства свойств предметов, воспринимаемых посредством опыта. Математические рассуждения принципиально непригодны для доказательства физических фактов.

Так, например, теорема эвклидовой геометрии о том, что сумма углов треугольника равна π , полностью верна для идеального треугольника, определяемого в абстрактной геометрии. Но отсюда не следует, что сумма углов, измеренных в конкретном треугольнике, необходимо должна равняться π , так же как из теорем классической механики не следует, что солнце и планеты необходимо должны двигаться в соответствии с ньютоновым законом тяготения. Эти вопросы могут быть решены только непосредственным наблюдением фактов.

Однако, некоторые предложения математической теории могут быть практически *испытаны*. Так, теорему эвклидовой геометрии о сумме углов треугольника можно сопоставить с измерениями конкретных треугольников. Если из таких непосредственных проверок обнаружится, что различные следствия теории действительно согласуются достаточно точно с имеющимися в нашем распоряжении эмпирическими данными, то можно быть более или менее уверенным в том, что существует некоторое сходство между математической теорией и структурой наблюдаемого реального мира. Мы ожидаем также, что согласие теории с опытом будет иметь место в дальнейшем и распространяться на те следствия теории, которые еще не подвергались непосредственной проверке; такого рода ожиданием мы руководствуемся в нашей деятельности.

Так обстоит дело, например, с эвклидовой геометрией. Какое бы предложение этой теории ни сопоставлялось с эмпирическими наблюдениями, всегда обнаруживалось совпадение, достаточное для всех обычных практических целей (отсюда необходимо исключить некоторые приложения, связанные с новейшими выводами физики). Таким образом, хотя нельзя *логически доказать*, что сумма углов конкретного треугольника должна равняться π , мы считаем *практически несомненным*, т. е. пригодным для практических действий, что наши измерения дадут сумму углов, *приблизительно равную* этому значению. Более того, мы уверены, что такая же согласованность будет обнаруживаться для каждого предложения, выведенного из аксиом Эвклида, которое мы стали бы испытывать на практике.

Естественно, что наша уверенность в дальнейшей согласованности между теорией и опытом будет расти по мере того, как будут накапливаться результаты, свидетельствующие о такой согласованности. Заметим, что „практическая несомненность“ утверждений эвклидовой геометрии отличается от практической несомненности утверждений,

связанных, например, со вторым законом термодинамики. Далее, *точность* ожидаемого совпадения результатов теории и опыта не всегда будет одной и той же. Тогда как в некоторых случаях даже наиболее чувствительные инструменты не могут обнаружить ни малейшего расхождения, в других случаях, где научный „закон“ выражает лишь главные черты наблюдаемых явлений, появляются отклонения, интерпретируемые как „ошибки“ или „возмущения“.

В тех случаях, когда установлено более или менее точное и постоянное согласие теории с опытом, математическая теория приобретает *практическое значение*, совершенно независимое от ее чисто математического интереса. Тогда теория может быть использована для различных практических целей. Большинство обычных применений математической теории можно грубым образом разбить на три рубрики: *описание, анализ и предсказание*.

Прежде всего теория может употребляться для чисто *описательных* целей. Большое количество эмпирических данных с помощью теории можно свести к сравнительно небольшому числу характеристик, представляющих в сжатом виде существенную информацию, полученную из опыта. Так, например, сложная совокупность астрономических наблюдений, относящихся к движению планет, в сжатом виде резюмируется системой Коперника.

Далее, результаты теории могут применяться как инструмент для научного *анализа* наблюдаемых явлений. Этой цели служит почти всякое научное исследование. Основной принцип, лежащий в основе таких применений теории, может быть выражен следующим образом: *каждая теория, не приспособленная к фактам, должна быть изменена*. Предположим, например, что мы пытаемся определить, влияют ли на интересующее нас явление изменения какого-нибудь фактора. Мы пытаемся тогда построить теорию, по которой такое влияние не будет иметь места, и сравниваем следствия из этой теории с нашими наблюдениями. Если в каком-нибудь месте мы обнаружим явное несогласие, то это будет означать, что мы должны приступить к исправлению нашей теории, чтобы учесть то влияние, которым мы пренебрегли.

Наконец, можно использовать теорию для *предсказания* событий, которые произойдут при данных условиях. Так, например, с помощью геометрической и механической теории астроном может предсказать дату затмения. Это является непосредственным применением указанного выше принципа, согласно которому ожидается, что согласие теории

с фактами будет иметь место также и для будущих событий. Этот же принцип применяется тогда, когда теоретические сведения используются для осуществления некоторого определенного события, например, когда специалист по теории стрельбы дает указание, как следует направить орудие, чтобы поразить цель.

13.5. Математическое понятие вероятности. Теперь мы приступим к построению теории, которая будет служить математической моделью явлений, обнаруживающих статистическую устойчивость. Нам нужна теория, которая базировалась бы на основных фактах, характеризующих устойчивость такого рода, и которую можно было бы использовать в различных направлениях, как это описано в предыдущем параграфе.

Закладывая фундамент этой теории, мы будем стараться возможно точнее придерживаться классического построения, описанного в предыдущем параграфе. В случае геометрии, например, мы знаем, что с помощью определенных действий с линейкой и куском мела мы можем создавать некоторые объекты, известные в повседневной речи как точки, прямые линии и т. п. Эмпирическое изучение свойств этих объектов обнаруживает наличие некоторых закономерностей. Тогда мы постулируем существование абстрактных копий этих объектов: точек, прямых линий и т. д. в чистой геометрии. Далее, основные черты наблюденных закономерностей мы фиксируем, в идеализированной форме, в качестве геометрических аксиом.

Подобным же образом в нашем случае мы знаем, что посредством некоторых действий, а именно, осуществляя ряд известных экспериментов, мы получаем совокупность наблюдаемых чисел, называемых частотами. Эмпирическое изучение поведения частот обнаруживает наличие некоторой типичной формы устойчивости, описанной в параграфе 13.3. Рассмотрим некоторое событие E , связанное со случайным экспериментом \mathfrak{E} . Согласно параграфу 13.3, частота события E в исследовательности из n повторений эксперимента \mathfrak{E} обнаруживает тенденцию стать постоянной, когда n возрастает; мы вынуждены высказать предположение, что при больших n частота с практической несомненностью становится приблизительно равной некоторому определенному числу P .

В нашей математической теории мы должны, таким образом, ввести определенное число P , которое будем называть вероятностью события E в случайному эксперименте \mathfrak{E} .

Всякий раз, когда мы говорим, что вероятность события E в эксперименте \mathfrak{E} равна P , точный смысл этого утверждения заключается просто в следующем: практически несомненно, что частота события E в длинном ряду повторений эксперимента \mathfrak{E} будет приблизительно равной P^*). Это утверждение будет называться также частотной интерпретацией вероятности P .

Вероятность P , введенная таким путем, будет являться абстрактной копией эмпирической частоты. Будет обнаружено, что для определения вероятности P окажется необходимым точно задать как тип случайного эксперимента \mathfrak{E} , так и событие E . Однако обычно мы будем считать эксперимент \mathfrak{E} фиксированным, и тогда можно без какой бы то ни было неопределенности говорить о вероятности события E .

Для дальнейшего развития теории мы должны будем рассмотреть основные свойства частот и выразить эти свойства в идеализированном виде как утверждения относительно свойств соответствующих вероятностей. Эти утверждения, вместе с постулатом существования вероятностей, будут служить аксиомами в нашей теории. В настоящем параграфе мы должны добавить лишь несколько предварительных замечаний, сами аксиомы будут сформулированы в следующей главе.

Для каждой частоты $\frac{v}{n}$ мы, очевидно, имеем $0 \leq \frac{v}{n} \leq 1$. Так как, согласно определению, каждая вероятность P приблизительно равна некоторой частоте, то естественно предположить, что P удовлетворяет соответствующему неравенству

$$0 \leq P \leq 1.$$

Это действительно будет одним из свойств, выраженных нашими аксиомами.

Если E — невозможное событие, т. е. событие, которое при осуществлении эксперимента \mathfrak{E} никогда не может произойти, то каждая частота события E должна равняться нулю, и соответственно мы положим $P=0$. С другой стороны, если известно, что для некоторого события E мы имеем $P=0$, то E не необходимо будет невозможным событием. Действительно, частотная интерпретация вероятности P показывает лишь то, что частота $\frac{v}{n}$ события E при больших n приблизительно равна нулю, так что в длинной серии экспериментов событие E произойдет

* На дальнейших стадиях исследования (в параграфе 16.3) мы сможем придать этому утверждению более точную форму.

лишь в очень малой доле всех случаев. Такой же вывод справедлив не только в случае $P = 0$, но даже при более общем предположении, что $0 \leq P < \epsilon$, где ϵ — некоторое очень маленькое число. Если E — событие такого рода, и если эксперимент \mathcal{E} осуществляется всего один раз, то можно считать практически несомненным, что событие E не произойдет. Этот пример частотной интерпретации вероятности будет часто применяться в дальнейшем.

Аналогично, если E — достоверное событие, т. е. событие, которое в эксперименте \mathcal{E} происходит всегда, то мы положим $P = 1$. С другой стороны, если известно, что $P = 1$, то мы не можем считать событие E достоверным и можем утверждать лишь, что в длинной серии экспериментов событие E будет происходить всегда, за исключением случаев, доля которых от общего числа очень мала. Такой же вывод справедлив и при более общем предположении, что $1 - \epsilon < P \leq 1$, где ϵ — очень малое число. Если E — событие такого рода, и если эксперимент \mathcal{E} осуществляется всего один раз, то можно считать практически несомненным, что событие E произойдет.

В литературе встречается множество различных мнений относительно обоснования теории вероятностей. Никакое из этих мнений не может считаться общепризнанным. Мы закончим настоящий параграф очень кратким обзором основных точек зрения.

Теория вероятностей возникла при изучении проблем, связанных с обычными азартными играми (см. параграф 13.3). Во всех этих играх результаты *a priori* распадаются на конечное число случаев, которые можно считать совершенно симметричными, как, например, случаи шести сторон игральной кости, 52 карт в обычной карточной колоде и т. д. Этот факт, казалось, мог служить основой для рационального объяснения устойчивости частот. Математики XVIII столетия воспользовались этим фактом и ввели знаменитый *принцип равновозможных случаев*, который был точно сформулирован Лапласом в классическом труде [22], как основной принцип всей теории. Ранее этот принцип принимался авторами в иезивном виде. В соответствии с этим принципом, разбиение на „равновозможные“ случаи осуществимо в любых наблюдениях, и вероятность события есть отношение числа случаев, благоприятных явлению, к общему числу возможных случаев.

Слабость этого определения очевидна. Прежде всего, неизвестно, каким образом решать, можно ли считать два случая равновозможными. Более того, трудно, а то и невозможно, точно указать разбиение на равновозможные случаи результатов опытов, не принадлежащих к азартным играм. Много труда было положено на попытки преодолеть эти трудности и улучшить классическое определение вероятности.

С другой стороны, многие авторы старались заменить классическое определение вероятности каким-нибудь существенно отличным определением.

На современных работах этого направления сказалась общая тенденция к построению всякой математической теории на аксиоматической основе. Так, некоторые авторы пытаются ввести систему аксиом, непосредственно основывающуюся на свойствах частот. Основным представителем этой школы является Мизес (см. [27], [28], [159]), который определяет вероятность события как предел частоты $\frac{y}{n}$ этого события, когда n стремится к бесконечности.

Существование этого предела, в строгом математическом смысле, постулируется как первая аксиома теории. Хотя такое определение на первый взгляд, бесспорно, кажется очень заманчивым, оно ведет к некоторым математическим трудностям, которые лишают это определение большей части его кажущейся простоты. Кроме того, предлагаемое определение вероятности приводит к смешению эмпирических и теоретических элементов, а современные аксиоматические теории обычно избегают этого смешения. Указанное определение вероятности можно сравнить, например, с определением геометрической точки, как предела пятен мела неограниченно убывающих размеров, а подобного определения современная аксиоматическая геометрия не вводит.

Другая школа выбирает в качестве отправной точки те же данные опыта как и частотная школа, но не постулирует существования определенных пределов для частот, а вводит вероятность события просто как число, связанное с этим событием. Аксиомы теории, выражающие правила, по которым можно оперировать с этими числами, суть идеализированные утверждения относительно наблюдаемых свойств частот. Теория этой школы с чисто математической точки зрения изложена Колмогоровым [21]. Более или менее похожих точек зрения придерживаются Дуб [75], Феллер [84] и Нейман [30]. К тому же кругу идей принадлежит работа автора [11], а настоящая книга представляет собой попытку построить на тех же принципах теорию статистики.

До сих пор мы занимались исключительно теорией вероятностей, рассматривая последнюю как математическую теорию явлений, в которых обнаруживается статистическая устойчивость. Согласно этой точке зрения, вероятности имеют своих двойников в наблюдаемых частотах, и вероятность, приписываемая фиксированному событию, в принципе должна быть доступна эмпирической проверке. Различие между описанными выше школами заключается главным образом в основаниях и в математическом изложении предмета, в то время как с точки зрения приложений все эти школы в значительной степени эквивалентны.

В коренной оппозиции к описанным выше направлениям стоит более общая концепция теории вероятностей — теория *степеней правдоподобия*, предлагаемая, например, Кейпсом [20] и Джессфрисом [18]. Согласно этой теории в ее наиболее развитой форме, данной Джессфрисом, каждое предложение имеет вероятность, выражаемую числом. Так, например, можно было бы выразить численно степень „практической несомненности“ будущего согласия некоторой математической теорией с опытными данными (см. параграф 13.4). Подобным же образом должна была бы существовать определенная численная вероятность истиности всякого утверждения, вроде «Железная

маска" был братом Людовика XIV», „нынешняя европейская война закончится в течение года“*) или „на планете Марс существует органическая жизнь“ Вероятности такого рода не имеют непосредственного отношения к случайным экспериментам и не имеют, очевидно, частотной интерпретации. В настоящей книге мы не будем заниматься вопросами, можно ли выразить численно такие вероятности и, в случае утвердительного ответа, могут ли такие численные выражения принести какую-нибудь пользу.

ГЛАВА II

ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ И АКСИОМЫ

14.1. Случайные величины. (Аксиомы 1—2.) Рассмотрим некоторый случайный эксперимент \mathfrak{E} , который может быть повторен большое число раз при неизменных условиях. Будем предполагать, что результат каждого отдельного эксперимента дается некоторым числом действительных величин $\xi_1, \xi_2, \dots, \xi_k$, где $k \geq 1$.

Введем соответствующую переменную точку или вектор $\xi = (\xi_1, \dots, \xi_k)$ в k -мерном пространстве R_k . Мы будем называть ξ *k -мерной случайной величиной* **).

Каждое осуществление эксперимента \mathfrak{E} дает в результате *наблюденное значение* величины ξ , координаты которого суть значения величин ξ_1, \dots, ξ_k , наблюденные в этом конкретном случае.

Пусть S обозначает некоторое достаточно простое точечное множество в R_k , например k -мерный интервал (см. параграф 3.1). Рассмотрим событие $\xi \subset S$, которое может произойти или не произойти при каждом отдельном осуществлении эксперимента \mathfrak{E} . Мы будем предполагать, что это событие имеет определенную вероятность P , в смысле, установленном в параграфе 13.5. Число P будет, очевидно, зависеть от множества S и соответственно будет обозначаться одним из следующих выражений:

$$P = P(S) = P(\xi \subset S).$$

Таким образом, мы видим, что вероятность можно рассматривать как *функцию множества*. Разумно требовать, чтобы эта функция множеств была однозначно определена, по крайней мере, для всех

*) Книга писалась в 1942—1944 гг. (Прим. ред.)

**) Во всем изложении общей теории случайные величины будут предпочтительно обозначаться буквами ξ и η . Жирный шрифт употребляется для многомерных величин ($k > 1$), обычный шрифт — для одномерных.

k -мерных интервалов. Однако ясно, что неудобно ограничиваться рассмотрением только интервалов. Может понадобиться рассмотреть вероятности событий, которым соответствуют, например, множества, получающиеся из интервалов с помощью операций сложения, вычитания и умножения (см. параграф 1.3). Мы видели в параграфах 2.3 и 3.3, что с помощью таких операций мы получим класс борелевских множеств в R_k в качестве естественного расширения класса всех интервалов. Поэтому представляется разумным непосредственно распространить наши рассмотрения на этот класс и предположить, что $P(S)$ определена для каждого борелевского множества. Если S — некоторое борелевское множество сложной структуры, то событие $\xi \subset S$ может и не поддаваться непосредственному наблюдению, и введение вероятностей для таких событий следует считать теоретической идеализацией. Однако некоторые следствия из теории будут всегда поддаваться непосредственному наблюдению, и практическое значение теории будет оцениваться согласованностью между поддающимися наблюдению следствиями теории и эмпирическими фактами. Итак, мы сформулируем нашу первую аксиому.

Аксиома 1. Каждой случайной величине ξ в R_k соответствует некоторая функция множества $P(S)$, однозначно определенная для всех борелевских множеств в R_k и такая, что значение $P(S)$ представляет собой вероятность события (или соотношения) $\xi \subset S$.

Как мы уже видели в параграфе 13.5, естественно предполагать, что всякая вероятность P удовлетворяет неравенству $0 \leq P \leq 1$. Далее, при каждом осуществлении эксперимента \mathcal{E} наблюденное значение величины ξ должно лежать где-нибудь в R_k , так что событие $\xi \subset R_k$ есть достоверное событие, и поэтому, в соответствии с параграфом 13.5, мы положим $P(R_k) = 1$.

Пусть теперь S_1 и S_2 — два множества в R_k без общих точек *).

Рассмотрим последовательность n повторений эксперимента \mathcal{E} , и пусть

$$\begin{aligned} v_1 &—\text{число наступлений события } \xi \subset S_1, \\ v_2 &= \quad " \quad " \quad " \quad \xi \subset S_2, \\ v &= \quad " \quad " \quad " \quad \xi \subset S_1 + S_2. \end{aligned}$$

*.) Как уже было установлено в параграфе 5.1, мы рассматриваем только борелевские множества.

Тогда, очевидно, $v = v_1 + v_2$, и поэтому соответствующие частоты удовлетворяют соотношению

$$\frac{v}{n} = \frac{v_1}{n} + \frac{v_2}{n}.$$

При больших значениях n , согласно предположению, практически несомненно, что частоты $\frac{v}{n}$, $\frac{v_1}{n}$, $\frac{v_2}{n}$ приблизительно равны соответственно вероятностям $P(S_1 + S_2)$, $P(S_1)$ и $P(S_2)$. Поэтому разумно потребовать, чтобы вероятность P обладала свойством аддитивности:

$$P(S_1 + S_2) = P(S_1) + P(S_2).$$

Наши рассуждения немедленно распространяются на любое конечное число множеств. Однако, чтобы получить простую и последовательную математическую теорию, мы должны ввести дальнейшую идеализацию. Действительно, мы будем предполагать, что свойство аддитивности функции $P(S)$ может быть распространено даже на счетные последовательности попарно не перекрывающихся множеств S_1, S_2, \dots , так что $P(S_1 + S_2 + \dots) = P(S_1) + P(S_2) + \dots$ (как и в случае аксиомы 1, здесь приходится, конечно, вводить соотношения, которые нельзя наблюдать непосредственно). В терминах, введенных в параграфах 6.2 и 8.2, мы можем теперь сформулировать нашу вторую аксиому.

Аксиома 2. Функция $P(S)$ есть неотрицательная аддитивная функция множества в R_k , такая, что $P(R_k) = 1$.

Согласно параграфам 6.6 и 8.4, каждая функция множества $P(S)$ со свойствами, указанными в аксиоме 2, задает некоторое *распределение* в R_k , которое можно конкретно интерпретировать как распределение единичной массы по пространству R_k , при котором множество S содержит массу $P(S)$.

Это распределение будет называться *распределением вероятностей случайной величины* ξ , а функция множества $P(S)$ будет называться *вероятностной функцией для* ξ . Функция точки $F(x) = F(x_1, \dots, x_k)$, которая определена формулой (6.6.1) в случае $k = 1$ и формулой (8.4.1) в общем случае и соответствует функции $P(S)$, будет называться *функцией распределения для* ξ . Как показано в параграфах 6.6 и 8.4, распределение однозначно определяется как функцией множеств $P(S)$, так и функцией точки $F(x)$.

В заключение заметим, что аксиомы 1 и 2 могут быть вкратце сформулированы следующим образом: *Каждая случайная величина имеет единственное распределение вероятностей.*

Если, например, эксперимент \mathcal{E} заключается в бросании игральной кости, то соответствующая случайная величина ξ — число выпавших очков — может принимать значения 1, 2, ..., 6, и только такие значения. Тогда наши аксиомы утверждают, что в R_1 существует распределение масс p_1, p_2, \dots, p_6 сосредоточенных в точках 1, 2, ..., 6, такое, что p_r представляет собой вероятность события $\xi = r$, и $\sum_1^6 p_r = 1$. С другой стороны, — и это важно отметить, — из аксиом 1 и 2 не следует равенства $p_r = \frac{1}{6}$ для всех r . Действительно, числа p_r можно считать физическими постоянными той игральной кости, которой мы пользуемся, и вопрос об их численной величине не может быть решен с помощью аксиом теории вероятностей, точно так же как размеры и вес кости не определяются геометрическими и механическими аксиомами. Однако, опыт показывает, что для тщательно сделанной кости частота каждого события $\xi = r$ в длинной серии бросаний обычно близка к $\frac{1}{6}$; в соответствии с этим, прибегая в дальнейшем для иллюстрации к примеру с костью, мы часто будем предполагать, что все p_r равны $\frac{1}{6}$. Это, однако, лишь предположение, а не логическое следствие из аксиом.

Если, далее, эксперимент \mathcal{E} заключается в измерении роста ξ людей из некоторой заданной группы, то для ξ можно предполагать любое значение лежащее в некоторых пределах, и наши аксиомы утверждают, что в R_1 существует непрерывная аддитивная функция множеств $P(S)$, представляющая вероятность того, что ξ имеет значение, принадлежащее к множеству S .

Аксиомы 1 и 2 для рассмотренного здесь класса случайных величин эквивалентны аксиомам Колмогорова [21]. Однако, аксиомы Колмогорова применимы к случайным величинам, определенным в пространствах более общего характера, чем рассматриваемые здесь. Те же аксиомы, что и выше, использованы в работе [11] автора настоящей книги.

14.2. Составные величины. (Аксиома 3.) Сначала мы рассмотрим один частный случай. Пусть случайные эксперименты \mathcal{E} и \mathcal{F} связаны соответственно с одномерными случайными величинами ξ и η . Таким образом, результат эксперимента \mathcal{E} представляется величиной ξ , а результат эксперимента \mathcal{F} — другой величиной η . Часто случается, что нам приходится рассматривать *составной эксперимент* (\mathcal{E}, \mathcal{F}), который заключается в осуществлении обоих экспериментов \mathcal{E} и \mathcal{F} при соблюдении известных правил и в совместном наблюдении обоих результатов.

Это означает, что мы наблюдаем переменную точку (ξ, η) , координатами которой являются результаты ξ и η экспериментов \mathfrak{E} и \mathfrak{F} . Поэтому мы можем рассматривать точку (ξ, η) как значение некоторой двумерной величины, которая называется *составной величиной*, определяемой величинами ξ и η . Пространство значений этой составной величины есть двумерное пространство — произведение (см. параграф 3.5) одномерных пространств величин ξ и η .

Пусть эксперимент \mathfrak{E} заключается в бросании одной игральной кости, а эксперимент \mathfrak{F} — в бросании другой кости; тогда составной эксперимент $(\mathfrak{E}, \mathfrak{F})$ заключается в бросании обеих костей. Результатом эксперимента \mathfrak{E} является число ξ , которое может принимать значения 1, 2, ..., 6, то же самое справедливо и для результата η эксперимента \mathfrak{F} . Тогда составная величина (ξ, η) выражает совместный результат бросания обеих костей, и ее возможными „значениями“ являются 36 пар чисел $(1, 1), \dots, (6, 6)$.

Если, с другой стороны, эксперимент \mathfrak{E} заключается в измерении роста ξ женатых мужчин, а эксперимент \mathfrak{F} — в измерении роста η замужних женщин, то составной эксперимент $(\mathfrak{E}, \mathfrak{F})$ может состоять, например, в измерении роста (ξ, η) обоих супругов. Точка (ξ, η) в этом случае может занимать любое место в некоторой определенной части плоскости.

Принцип образования составных величин можно применять и в более общих случаях. Пусть случайные эксперименты $\mathfrak{E}_1, \dots, \mathfrak{E}_n$ связаны соответственно со случайными величинами ξ_1, \dots, ξ_n размерностей k_1, \dots, k_n ; рассмотрим составной эксперимент $(\mathfrak{E}_1, \dots, \mathfrak{E}_n)$, заключающийся в осуществлении каждого из экспериментов \mathfrak{E}_i и совместном наблюдении всех результатов. Мы получим составную величину (ξ_1, \dots, ξ_n) , представляемую точкой $(k_1 + \dots + k_n)$ -мерного пространства — произведения пространств всех ξ_i (см. параграф 3.5).

Эмпирическое изучение частот, связанных с составными экспериментами, обнаруживает такую же статистическую устойчивость, как и для составляющих экспериментов. Каждый эксперимент, составленный из случайных экспериментов, действительно имеет характер случайного эксперимента. В соответствии с этим мы можем сформулировать нашу третью аксиому:

Аксиома 3. *Если ξ_1, \dots, ξ_n — случайные величины, то каждая составная величина (ξ_1, \dots, ξ_n) также является случайной величиной.*

Тогда из предыдущих аксиом следует, что каждая составная величина имеет единственное распределение вероятностей в $(k_1 + \dots + k_n)$ -мерном пространстве. Это распределение часто будет называться *совместным распределением величин ξ_1, \dots, ξ_n* .

Рассмотрим теперь две случайные величины ξ и η размерностей соответственно k_1 и k_2 . Пусть P_1 и P_2 обозначают вероятностные функции величин ξ и η , а P — вероятностную функцию составной величины (ξ, η) . Если S — некоторое множество в пространстве величины ξ , то выражение $P(\xi \subset S)$ представляет вероятность того, что составная величина (ξ, η) принимает значение, принадлежащее цилиндрическому множеству (см. параграф 3.5), определяемому соотношением $\xi \subset S$ или, другими словами, вероятность того, что ξ имеет значение, принадлежащее множеству S , безотносительно к значению величины η . Подобным же образом, если T — некоторое множество в пространстве величины η , то $P(\eta \subset T)$ есть вероятность того, что η принимает значение из T , безотносительно к значению ξ .

Таким образом, мы имеем

$$(14.2.1) \quad P(\xi \subset S) = P_1(S), \quad P(\eta \subset T) = P_2(T),$$

и, согласно (8.4.2), это означает, что *частные распределения* ($k_1 + k_2$)-мерного совместного распределения относительно подпространств величин ξ и η совпадают соответственно с распределениями величин ξ и η . Очевидно, это можно обобщить на случай любого числа составляющих величин. Если массу в совместном распределении „спроектировать“ на подпространство какой-нибудь из составляющих величин, то получающееся при этом частное распределение всегда совпадает с распределением соответствующей составляющей величины.

Важный случай образования составной величины возникает при рассмотрении последовательности повторений случайного эксперимента \mathcal{E} . Пусть составной эксперимент заключается в n -кратном осуществлении одного и того же эксперимента \mathcal{E} и в наблюдении результатов ξ_1, \dots, ξ_n всех n повторных экспериментов. Тогда результатом этого составного эксперимента будет значение составной величины (ξ_1, \dots, ξ_n) , выражающее совместный результат всех n повторений эксперимента \mathcal{E} .

Если, например, \mathcal{E} заключается в бросании кости, то соответствующая одномерная случайная величина ξ имеет шесть возможных значений $1, 2, \dots, 6$. Составная величина (ξ_1, \dots, ξ_n) выражает совместный результат n последовательных бросаний кости, и ее „значениями“ служат 6^n систем из n чисел: $(1, \dots, 1), \dots, (6, \dots, 6)$. По аксиоме 3, в R_n существует распределение вероятностей для (ξ_1, \dots, ξ_n) с определенными вероятностями $P_1, \dots, 1 + \dots + P_6, \dots, 6$, соответствующими всевозможным значениям этой составной величины.

В тех вопросах, где одновременно рассматриваются различные случайные величины, мы всегда будем предполагать, что для всех отно-

связанных к изучаемому вопросу случайных величин задано правило их комбинирования, так что составная величина определена. Как правило, мы будем при этом употреблять $P(S)$ для обозначения вероятностной функции составной величины.

14.3. Условные распределения. Пусть ξ и η — случайные величины размерностей k_1 и k_2 , связанные со случайными экспериментами \mathcal{E} и \mathcal{F} . Пусть P обозначает вероятностную функцию составной величины (ξ, η) , S и T — множества соответственно в пространствах величин ξ и η . Тогда выражение $P(\xi \subset S, \eta \subset T)$ представляет собой вероятность события, определяемого совокупностью соотношений $\xi \subset S, \eta \subset T$, или, другими словами, вероятность того, что составная величина (ξ, η) имеет значение, принадлежащее прямоугольному множеству со сторонами S и T (см. параграф 3.5).

Допустим, что $P(\xi \subset S) > 0$. Введем при этом условии новую величину $P(\eta \subset T | \xi \subset S)$, определяемую соотношением

$$(14.3.1) \quad P(\eta \subset T | \xi \subset S) = \frac{P(\xi \subset S, \eta \subset T)}{P(\xi \subset S)}.$$

Подобным же образом, допустив, что $P(\eta \subset T) > 0$, мы введем другую величину $P(\xi \subset S | \eta \subset T)$, положив

$$(14.3.2) \quad P(\xi \subset S | \eta \subset T) = \frac{P(\xi \subset S, \eta \subset T)}{P(\eta \subset T)}.$$

Чтобы оправдать названия, которые будут даны этим величинам, мы рассмотрим теперь некоторые важные свойства последних.

Предположим сначала, что в (14.3.2) множество T фиксировано, а S есть переменное множество в пространстве R_{k_1} величины ξ . Тогда правая часть равенства (14.3.2) будет неотрицательной аддитивной функцией множества S . Если $S = R_{k_1}$, то прямоугольное множество $\xi \subset R_{k_1}$, $\eta \subset T$ совпадает с цилиндрическим множеством $\eta \subset T$ (см. параграф 3.5), так что правая часть в (14.3.2) будет иметь значение 1. Таким образом, при фиксированном T $P(\xi \subset S | \eta \subset T)$ есть неотрицательная аддитивная функция множества S , которая при $S = R_{k_1}$ принимает значение 1. Другими словами, $P(\xi \subset S | \eta \subset T)$ при фиксированном T есть вероятностная функция для некоторого распределения в R_{k_1} . Точно так же можно показать, что $P(\eta \subset T | \xi \subset S)$ при фиксированном S есть вероятностная функция для некоторого распределения в пространстве R_{k_2} величины η . Теперь мы покажем, что в некотором

обобщенном смысле эти величины можно действительно считать вероятностями, допускающими определенную частотную интерпретацию.

Рассмотрим последовательность Z из n повторений составного эксперимента (ξ, η). Каждый из n экспериментов, являющихся элементами последовательности Z , дает в результате „значение“ составной величины (ξ, η). Пусть в последовательности Z

v_1 обозначает число наступлений события $\xi \subset S$,

v_2 " " " " $\eta \subset T$,

v " " " " $\xi \subset S, \eta \subset T$,

а Z_1 , Z_2 и Z — соответствующие подпоследовательности из Z тек экспериментов, в результате которых имели место соответственно первое, второе и третье события. Очевидно, третье событие происходит тогда и только тогда, когда происходят оба первых события, так что Z состоит в точности из общих элементов подпоследовательностей Z_1 и Z_2 .

В соответствии с частотной интерпретацией вероятностей (см. параграф 13.5) практически несомненно, что при больших n приближение удовлетворяются соотношения

$$P(\xi \subset S) = \frac{v_1}{n}, \quad P(\eta \subset T) = \frac{v_2}{n}, \quad P(\xi \subset S, \eta \subset T) = \frac{v}{n}.$$

Поэтому, согласно формулам (14.3.1) и (14.3.2), мы получаем приближенные равенства

$$(14.3.3) \quad P(\eta \subset T | \xi \subset S) = \frac{v}{v_1}, \quad P(\xi \subset S | \eta \subset T) = \frac{v}{v_2}.$$

Рассмотрим теперь v_1 элементов подпоследовательности Z_1 . Эти элементы являются теми случаями из наших n повторных экспериментов, в которых произошло событие $\xi \subset S$. Среди этих случаев существует точно v таких, в которых, кроме того, произошло событие $\eta \subset T$; последние v случаев образуют подпоследовательность Z . Таким образом, отношение $\frac{v}{v_1}$ есть частота события $\eta \subset T$ в подпоследовательности Z_1 , или можно сказать еще так: $\frac{v}{v_1}$ есть условная частота события $\eta \subset T$ при условии $\xi \subset S$. Простой перестановкой получается соответствующее свойство отношения $\frac{v}{v_2}$. Приближенные соотношения (14.3.3) дают частотную интерпретацию для выражений $P(\eta \subset T | \xi \subset S)$ и $P(\xi \subset S | \eta \subset T)$, чем спрощдается введение следующих определений:

Число $P(\eta \subset T | \xi \subset S)$, определяемое формулой (14.3.1), называется *условной вероятностью события* $\eta \subset T$ при условии $\xi \subset S$. В соответствии с этим распределение в R_{k_2} , определяемое формулой (14.3.1) при фиксированном S , называется *условным распределением величины* η при условии $\xi \subset S$. Аналогичные определения вводятся для величины $P(\xi \subset S | \eta \subset T)$, определенной формулой (14.3.2).

Следует заметить, что *условная вероятность определена только в том случае, когда вероятность соответствующего условия отлична от нуля.*

Если $P(\xi \subset S)$ и $P(\eta \subset T)$ обе отличны от нуля, то из формул (14.3.1) и (14.3.2) получается соотношение

$$(14.3.4) \quad P(\xi \subset S, \eta \subset T) = P(\xi \subset S)P(\eta \subset T | \xi \subset S) = \\ = P(\eta \subset T)P(\xi \subset S | \eta \subset T).$$

В примере, рассмотренном в предыдущем параграфе, где ξ — рост женатого мужчины, а η — рост его жены, данные, соответствующие *всем наблюденным значениям* ξ , определяют распределение величины ξ . Так, например, вероятность соотношения $a < \xi \leq b$ будет приближенно определяться частотой соответствующего события в совокупности наших данных.

Предположим теперь, что из наших данных выбираются подмножества всех случаев, в которых η превышает некоторую заданную константу c . Данные для ξ в *случаях, принадлежащих к этому подмножеству*, определяют условное распределение величины ξ при условии $\eta > c$. Так, например, частота события $a < \xi \leq b$ внутри указанного подмножества является определенной выше условной частотой и при большом числе наблюдений с практической несомненностью становится приблизительно равной условной вероятности соотношения $a < \xi \leq b$ при условии $\eta > c$. Множеством S здесь служит интервал $a < \xi \leq b$, а множеством T — интервал $\eta > c$.

Очевидно, что в этом случае есть основания предполагать, что эта условная вероятность отлична от вероятности в полной совокупности данных, так как в среднем рослые женщины, соответствующие условию $\eta > c$, становятся женами рослых мужчин чаще, чем малорослых.

С другой стороны, пусть ξ попрежнему обозначает рост женатого мужчины, а η обозначает на этот раз рост замужней женщины в *непосредственно следующей супружеской паре* в том списке, из которого мы черпаем наши данные. В этом случае нет причин ожидать, что условная вероятность соотношения $a < \xi \leq b$ при условии $\eta > c$ будет отличаться от безусловной вероятности $P(a < \xi \leq b)$. Наоборот, следует ожидать, что условное распределение величины ξ не зависит от каких-либо гипотез относительно величины η , и, обратно, условное распределение η не должно зависеть от гипотез относительно ξ . Если это условие выполнено, то мы имеем дело

со случаем независимых случайных величин, который будет исследован в следующем параграфе.

14.4. Независимые величины. Рассмотрим важный частный случай введенных в предыдущем параграфе понятий, когда для любых множеств S и T удовлетворяется условие мультипликативности

$$(14.4.1) \quad P(\xi \subset S, \eta \subset T) = P(\xi \subset S)P(\eta \subset T).$$

Соотношения (14.3.1) и (14.3.2) показывают, что в этом случае

$$(14.4.2) \quad P(\xi \subset S | \eta \subset T) = P(\xi \subset S), \text{ если } P(\eta \subset T) > 0,$$

$$(14.4.3) \quad P(\eta \subset T | \xi \subset S) = P(\eta \subset T), \text{ если } P(\xi \subset S) > 0,$$

так что *условное распределение величины ξ не зависит от каких-либо гипотез относительно величины η , и наоборот.*

В этом случае мы будем говорить, что ξ и η *независимые случайные величины* и что события $\xi \subset S$ и $\eta \subset T$ суть *независимые события*.

Обратно, предположим, что одно из двух последних соотношений, например (14.4.2), выполняется при любых множествах S и T , таких, что условная вероятность в левой части равенства определена, т. е. что $P(\eta \subset T) > 0$. Тогда из формулы (14.3.2) следует, что при этом выполняется условие мультипликативности (14.4.1). Однако при $P(\eta \subset T) = 0$ равенство (14.4.1) тривиально, так как при этом обе части равенства обращаются в нуль. Таким образом, (14.4.1) выполняется для всех S и T , и отсюда мы получаем (14.4.3). Итак, *каждое из соотношений (14.4.2) или (14.4.3) является необходимым и достаточным условием независимости.*

Теперь мы дадим другое необходимое и достаточное условие. Пусть P_1 и P_2 — вероятностные функции для ξ и η , а функциями распределения для ξ , η и (ξ, η) являются соответственно

$$F_1(x) = F_1(x_1, \dots, x_{k_1}) = P_1(\xi_1 \leq x_1, \dots, \xi_{k_1} \leq x_{k_1}),$$

$$F_2(y) = F_2(y_1, \dots, y_{k_2}) = P_2(\eta_1 \leq y_1, \dots, \eta_{k_2} \leq y_{k_2}),$$

$$F(x, y) = F(x_1, \dots, x_{k_1}, y_1, \dots, y_{k_2}) = P(\xi_i \leq x_i, \eta_j \leq y_j),$$

где $i = 1, 2, \dots, k_1$ и $j = 1, 2, \dots, k_2$. Условие мультипликативности (14.4.1) может быть записано в виде

$$(14.4.4) \quad P(\xi \subset S, \eta \subset T) = P_1(S)P_2(T).$$

В параграфе 8.6 было показано, что если P_1 и P_2 — заданные вероятностные функции в пространствах величин ξ и η , то в произведении этих пространств существует одно и только одно распределение, удовлетворяющее условию (14.4.4). Оно определяется функцией распределения

$$(14.4.5) \quad F(x, y) = F_1(x)F_2(y).$$

Итак, (14.4.5) есть необходимое и достаточное условие независимости величин ξ и η .

Рассмотрим теперь n случайных величин ξ_1, \dots, ξ_n с вероятностными функциями P_1, \dots, P_n и функциями распределения F_1, \dots, F_n . Пусть P и F обозначают соответственно вероятностную функцию и функцию распределения составной величины (ξ_1, \dots, ξ_n) . Непосредственно обобщая сказанное выше, мы будем говорить, что ξ_1, \dots, ξ_n — независимые случайные величины, если для любых множеств S_1, \dots, S_n выполняется условие мультипликативности

$$(14.4.6) \quad P(\xi_1 \subset S_1, \dots, \xi_n \subset S_n) = \prod_{r=1}^n P(\xi_r \subset S_r) = \prod_{r=1}^n P_r(S_r).$$

Используя заключительное замечание параграфа 8.6, мы обнаружим, что условие (14.4.5) можно непосредственно обобщить и показать, что равенство $F = F_1F_2 \dots F_n$ есть необходимое и достаточное условие независимости величин ξ_r . Если ξ_r и составная величина $(\xi_1, \dots, \xi_{r-1})$ при $r=2, 3, \dots, n$ независимы, то ξ_1, \dots, ξ_n независимы. Это непосредственно следует из определения независимости (14.4.6).

Если в последовательности ξ_1, ξ_2, \dots случайных величин группа ξ_1, \dots, ξ_n при любом n состоит из независимых величин, то мы говорим, что ξ_1, ξ_2, \dots есть последовательность независимых величин. Важный случай такой последовательности встречается при рассмотрении последовательности повторений случайного эксперимента \mathcal{E} . Если последовательные эксперименты производятся при совершенно одинаковых условиях, то нельзя допустить, что результаты первых $n-1$ экспериментов влияли на вероятность P определенного результата n -го эксперимента. Это означает, что распределение случайной величины ξ_n , связанной с n -м экспериментом, не зависит от каких-либо гипотез относительно значений составной величины $(\xi_1, \dots, \xi_{n-1})$, т. е. ξ_n и $(\xi_1, \dots, \xi_{n-1})$ независимы. Согласно сказанному

выше, отсюда следует, что ξ_1, ξ_2, \dots образуют последовательность независимых величин. Такая последовательность повторений случайного эксперимента \mathcal{E} будет коротко называться последовательностью независимых повторений эксперимента \mathcal{E} . Мы всегда будем считать, если только не оговорено противное, что рассматриваются независимые последовательности повторений.

Рассмотрим составной эксперимент, заключающийся в двух бросаниях некоторой игральной кости. Будем повторять этот составной эксперимент большое число раз, сохраняя условия каждого отдельного бросания неизменными, насколько это возможно. Тогда мы можем изучить поведение условной частоты какого-нибудь данного результата второго бросания при какой-либо гипотезе относительно результата первого бросания. Многочисленные опыты не обнаружили никакого влияния таких гипотез на поведение условной частоты, так что мы вправе предполагать, что случайные величины, связанные с этими двумя бросаниями, независимы. Такое же положение обнаруживается при рассмотрении составного эксперимента, образованного n бросаниями, где n может иметь любое значение; в соответствии с этим мы предполагаем, что последовательность бросаний, осуществляемых при одинаковых условиях, образует последовательность независимых повторений в установленном выше смысле.

Предположим теперь, что в каждом бросании все шесть возможных результатов имеют вероятность $\frac{1}{6}$. Тогда по (14.4.6) каждый из 6^n возможных результатов n последовательных бросаний будет иметь вероятность $\left(\frac{1}{6}\right)^n$.

Рассмотрим, наконец, n независимых величин ξ_1, \dots, ξ_n . Если в условии мультипликативности (14.4.6) некоторое число множеств S , будет совпадать с пространствами соответствующих величин, то из условия (14.4.6) будет следовать, что *любые $n_1 < n$ из этих величин независимы*.

Обратное предложение неверно. Это показывает построенный С. Н. Бернштейном пример трех одномерных величин ξ, η, ζ , любые две из которых независимы, но все три величины ξ, η, ζ не независимы. Пусть трехмерное распределение составной величины (ξ, η, ζ) таково, что в каждой из четырех точек $(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)$ сосредоточена масса $\frac{1}{4}$. Тогда легко убедиться, что каждое одномерное частное распределение содержит массы $\frac{1}{2}$, сосредоточенные в двух точках 0 и 1, а каждое двумерное частное распределение содержит массы $\frac{1}{4}$, сосредоточенные в четырех точках $(0, 0), (1, 0), (0, 1)$ и $(1, 1)$. Отсюда следует, что любые две величины независимы. Например, мы

имеем

$$P(\xi = 1, \eta = 1) = P(\xi = 1)P(\eta = 1) = \left(\frac{1}{2}\right)^2 = \frac{1}{4},$$

и нетрудно видеть, что аналогичные соотношения справедливы для любых событий $\xi \subset S$, $\eta \subset T$, так что (14.4.1) выполняется. Но *три величины ξ , η , ζ не независимы*, так как мы имеем

$$P(\xi = 1, \eta = 1, \zeta = 1) = \frac{1}{4},$$

$$P(\xi = 1)P(\eta = 1)P(\zeta = 1) = \left(\frac{1}{2}\right)^3 \neq \frac{1}{4}.$$

14.5. Функции случайных величин. Рассмотрим сначала случай одномерной случайной величины ξ с вероятностной функцией P . Предположим, что при каждом осуществлении случайного эксперимента, с которым связана величина ξ , мы наблюдаем не самое величину ξ , а некоторую действительную функцию $g(\xi)$, конечную и однозначно определенную для всех действительных ξ . Как обычно, мы предполагаем, что $g(\xi)$ *B*-измерима (см. параграф 5.2).

Уравнение $\eta = g(\xi)$ определяет соответствие между величинами ξ и η . Обозначим через Y некоторое данное множество на оси η и через X — соответствующее множество (см. параграф 5.2) всех тех ξ , для которых $\eta = g(\xi) \subset Y$. В параграфе 5.2 было показано, что множество X , соответствующее любому борелевскому множеству Y , само является борелевским множеством. Если X и Y — соответствующие множества, то $\eta \subset Y$ тогда и только тогда, если $\xi \subset X$, так что два события $\eta \subset Y$ и $\xi \subset X$ полностью эквивалентны. Последнее событие, согласно аксиоме 1, имеет определенную вероятность $P(X)$, следовательно, событие $\eta \subset Y$ должно иметь такую же вероятность.

Итак, мы видим, что каждая функция $\eta = g(\xi)$ случайной величины ξ сама является случайной величиной с распределением вероятностей, определяемым распределением величины ξ . Действительно, если Q обозначает вероятностную функцию для η , то из сказанного выше следует, что для каждого борелевского множества Y мы имеем

$$(14.5.1) \quad Q(Y) = P(X),$$

где X есть множество, соответствующее множеству Y . Если, в частности, в качестве Y выбрать замкнутый интервал $(-\infty, y]$ и обозначить через S_y множество тех ξ , для которых $\eta = g(\xi) \leq y$, то функцией распределения величины η будет

$$(14.5.2) \quad G(y) = Q(\eta \leq y) = P(S_y).$$

Пусть распределение величины ξ интерпретируется обычным образом, как распределение массы на оси ξ . Представим себе, что каждая частица массы в этом распределении движется от своего исходного места на оси ξ сначала в вертикальном направлении, пока она не достигнет кривой $\eta = g(\xi)$, и затем горизонтально по направлению к оси η . Полученное таким путем распределение массы на оси η будет распределением, заданным формулой (14.5.1).

Наши рассмотрения немедленно обобщаются на случай произвольного числа измерений. Пусть $\xi = (\xi_1, \dots, \xi_r)$ — случайная величина в j -мерном пространстве R_j , P — ее вероятностная функция. Рассмотрим k -мерную векторную функцию $\eta = g(\xi) = (\eta_1, \dots, \eta_k)$, конечную и однозначно определенную для всех ξ в R_j . Сам вектор η представляется точкой k -мерного пространства R_k . Предположим, что каждая компонента η_i , величины η есть B -измеримая функция (см. параграф 9.1) величин ξ_1, \dots, ξ_r . Отсюда, как и в одномерном случае, будет следовать, что η есть случайная величина в R_k с вероятностной функцией Q , определяемой соотношением (14.5.1), в котором Y обозначает теперь произвольное множество в R_k , а X — соответствующее множество тех ξ из R_j , для которых $\eta = g(\xi) \subset Y$.

Если для множества Y соответствующее множество X пусто, то, конечно, $Q(Y) = 0$. Условие, что функция $g(\xi)$ конечна и однозначно определена для всех ξ из R_j , можно, очевидно, заменить более общим условием, что точки ξ , в которых $g(\xi)$ не является конечной или однозначно определенной, образуют множество S , такое, что $P(S) = 0$.

Возьмем, например, $g(\xi) = (\xi_1, \dots, \xi_r)$, где $r < j$, так что $g(\xi)$ будет просто проекцией точки ξ на некоторое r -мерное подпространство (см. параграф 3.5). Тогда вероятностной функцией для $g(\xi)$ будет $Q(Y) = P(X)$, где Y — множество из этого подпространства, а X — цилиндрическое множество (см. параграф 3.5) в R_j , определяемое соотношением $(\xi_1, \dots, \xi_r, 0, \dots, 0) \subset Y$. Соответствующим распределением будет частное распределение (см. параграф 8.4) для (ξ_1, \dots, ξ_r) , которое получается проектированием исходного распределения на выделенное r -мерное подпространство. Если положить, в частности, $r = 1$, то мы видим, что каждая компонента ξ_r случайной величины ξ сама является случайной величиной с частным распределением, получающимся проектированием исходного распределения на ось ξ_r .

Функцию $\eta = g(\xi_1, \dots, \xi_n)$ от n случайных величин можно рассматривать как функцию от составной величины (ξ_1, \dots, ξ_n) . Таким образом, согласно сказанному выше, η всегда является случайной величиной

с распределением вероятностей, однозначно определяемым совместным распределением величин ξ_1, \dots, ξ_n .

Если ξ_1, \dots, ξ_n — независимые величины, то непосредственно видно, что величины $g_1(\xi_1), \dots, g_n(\xi_n)$ также независимы.

14.6. Заключение. Содержание настоящей главы можно вкратце резюмировать следующим образом. Из эмпирических данных, связанных со случайными экспериментами, мы извлекли основной факт статистической устойчивости, т. е. устойчивости длинного ряда частот. В нашей математической теории мы идеализировали этот факт, постулируя существование абстрактных дубликатов частот — *математических вероятностей*. Затем процесс идеализации был продвинут еще на один шаг нашим предположением, что свойство аддитивности вероятностей можно распространить с конечных последовательностей событий на счетные. Таким путем мы пришли к понятиям *случайной величины* и ее *распределения вероятностей*.

Далее мы высказали допущение, что любое количество случайных экспериментов можно объединить в составной случайный эксперимент, обнаруживающий тот же вид статистической устойчивости, как и составляющие эксперименты. Таким способом мы ввели представление о *совместном распределении вероятностей* некоторого числа случайных величин.

Изучение некоторых условных частот заставило нас ввести их абстрактные дубликаты под названием *условных вероятностей*. Последние связаны с некоторым *условным распределением* случайной величины, которое, в частном случае, дает основание для введения важного понятия *независимых случайных величин*. Наконец, было показано, что B -измеримая функция любого количества случайных величин сама является случайной величиной с распределением вероятностей, однозначно определяемым совместным распределением аргументов.

Таким образом, мы заложили фундамент чисто математической теории случайных величин и распределений вероятностей. Теперь нашей ближайшей задачей будет детальная разработка этой теории; этой цели будут посвящены дальнейшие главы части II. В главах 15—20 мы в основном будем заниматься одномерными случайными величинами и распределениями, а в главах 21—24 будет изучен многомерный случай.

В части III мы обратимся к вопросам проверки математической теории опытом и использования теоретических результатов для целей получения статистических выводов из эмпирических данных.

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И РАСПРЕДЕЛЕНИЯ В R_1

ГЛАВА 15

ОБЩИЕ СВОЙСТВА

15.1. Функция распределения и плотность вероятности. Рассмотрим некоторую одномерную случайную величину ξ . Согласно аксиомам 1 и 2 параграфа 14.1, ξ обладает определенным *распределением вероятности* в R_1 . Это распределение может быть конкретно интерпретировано как такое распределение единицы массы по R_1 , при котором количество массы $P(S)$, приходящееся на произвольное борелевское множество S , равно вероятности того, что величина ξ принимает значение, принадлежащее S .

Как мы видели в параграфе 6.6, такое распределение может быть задано как посредством неотрицательной аддитивной функции множества $P(S)$, называемой *вероятностной функцией* величины ξ , так и посредством соответствующей функции точки $F(x)$, определенной соотношением

$$P(\xi \leq x) = F(x)$$

и называемой *функцией распределения* величины ξ . В рассматриваемом случае одномерного распределения мы будем обычно пользоваться функцией $F(x)$.

Общие свойства функций распределения уже были исследованы нами в параграфе 6.6. В частности, там было доказано, что любая функция распределения $F(x)$ есть неубывающая функция от x , всюду непрерывная справа, и такая, что $F(-\infty) = 0$ и $F(+\infty) = 1$. Разность $F(b) - F(a)$ представляет вероятность того, что величина ξ принимает значение, принадлежащее интервалу $a < \xi \leq b$:

$$P(a < \xi \leq b) = F(b) - F(a).$$

Если x_0 есть точка разрыва функции $F(x)$ со скачком, равным p_0 , то из параграфа 6.6 следует, что в точке x_0 сосредоточена масса p_0 ; это означает, что вероятность для величины ξ принять значение x_0 равна p_0 :

$$P(\xi = x_0) = p_0.$$

С другой стороны, если в некоторой точке x существует производная $F'(x) = f(x)$, то $f(x)$ представляет плотность массы в этой точке и называется *плотностью вероятности* или *функцией плотности* данной случайной величины. Вероятность того, что величина ξ принимает значение, принадлежащее интервалу $x < \xi < x + \Delta x$, будет тогда для малых Δx асимптотически равна $f(x)\Delta x$, что записывается в обычных дифференциальных обозначениях

$$P(x < \xi < x + dx) = f(x)dx.$$

Этот дифференциал будем называть *элементом вероятности* данного распределения.

Любая функция $\eta = g(\xi)$ случайной величины ξ , согласно параграфу 14.5, сама есть случайная величина с функцией распределения, заданной соотношением (14.5.2). Рассмотрим два простых примера, которые будут часто встречаться в дальнейшем.

В случае линейной функции $\eta = a\xi + b$ соотношение $\eta \leq y$ эквивалентно соотношению $\xi \leq \frac{y-b}{a}$ или $\xi \geq \frac{y-b}{a}$, в зависимости от того, положительно или отрицательно число a . Из (14.5.2) следует, что η имеет функцию распределения

$$(15.1.1) \quad G(y) = \begin{cases} F\left(\frac{y-b}{a}\right) & \text{при } a > 0, \\ 1 - F\left(\frac{y-b}{a}\right) & \text{при } a < 0, \end{cases}$$

где $F(x)$ есть функция распределения величины ξ . Такое выражение $G(y)$ при $a < 0$ справедливо только для тех y , для которых $\frac{y-b}{a}$ есть точка непрерывности функции F . В точках разрыва, по нашему обычному соглашению, функция $G(y)$ должна быть определена так, чтобы она была непрерывна справа. Если плотность вероятности $f(x) = F'(x)$ существует для всех значений x , то η имеет плотность вероятности

$$(15.1.2) \quad g(y) = G'(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right).$$

Рассмотрим затем функцию $\eta = \xi^2$. Величина η всегда неотрицательна и при $y > 0$ соотношение $\eta \leq y$ эквивалентно соотношению

$-V_y \leq \xi \leq V_y$. Поэтому η имеет функцию распределения

$$(15.1.3) \quad G(y) = \begin{cases} 0 & \text{при } y < 0, \\ F(V_y) - F(-V_y) & \text{при } y \geq 0. \end{cases}$$

На этот раз последнее выражение справедливо только тогда, когда $-V_y$ есть точка непрерывности функции F . Если плотность вероятности $f(x) = F'(x)$ существует для всех x , то η имеет плотность вероятности

$$(15.1.4) \quad g(y) = G'(y) = \begin{cases} 0 & \text{при } y < 0, \\ \frac{1}{2V_y}(f(V_y) + f(-V_y)) & \text{при } y > 0. \end{cases}$$

15.2. Два простых типа распределений. В большинстве вопросов, встречающихся в приложениях к статистике, мы имеем дело с распределениями одного из двух простых типов — *дискретного и непрерывного*.

1. **Дискретный тип.** Случайная величина ξ называется величиной дискретного типа или обладающей распределением этого типа, если вся распределенная масса содержится в точках сосредоточения массы *) и если, кроме того, всякий конечный интервал содержит не более чем конечное число точек сосредоточения массы. (Согласно параграфу 6.2, множество всех точек сосредоточения массы конечно или счетно.) Обозначим точки сосредоточения массы через x_1, x_2, \dots , а соответствующие массы через p_1, p_2, \dots . Распределение случайной величины ξ будет полностью описано, если указать, что для любого y вероятность того, что ξ принимает значение x_v , равна p_v :

$$P(\xi = x_v) = p_v.$$

Для любого множества S , не содержащего ни одной из точек x_v , имеем, с другой стороны,

$$P(\xi \subset S) = 0.$$

Так как вся распределенная масса равна единице, то

$$\sum_v p_v = 1.$$

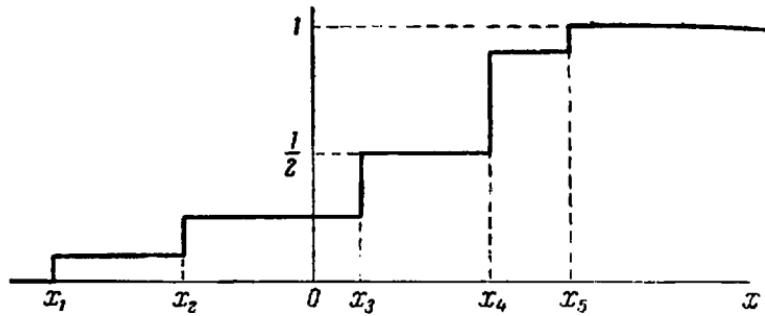
*) Это соответствует случаю $c_1 = 1, c_2 = 0$ в (6.6.2).

Функция распределения $F(x)$ задается при этом соотношением

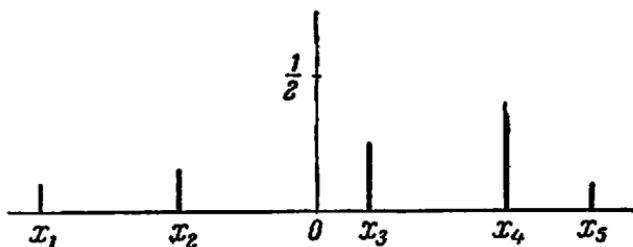
$$(15.2.1) \quad F(x) = P(\xi \leq x) = \sum_{x_v \leq x} p_v,$$

где суммирование распространено на все значения индекса v , для которых $x_v \leq x$. Таким образом, $F(x)$ есть ступенчатая функция (см. параграфы 6.2 и 6.6), равная постоянной на любом интервале, не содержащем точек x_v , и имеющая в каждой точке x_v скачок p_v .

Распределение дискретного типа может быть графически представлено посредством графика функции $F(x)$ или графика, на котором каждой точке x_v отнесена ордината высоты p_v , как показано на фиг. 4 и 5.



Фиг. 4. Функция распределения дискретного типа. (Заметьте, что медиана неопределенна; см. стр. 201)



Фиг. 5. Вероятности, соответствующие распределению на фиг. 4.

В приложениях к статистике величины дискретного типа встречаются, например, тогда, когда исследуемая случайная величина представляет число каких-нибудь единиц, например: число поросят в помете, число телефонных вызовов на данной станции в течение одного часа, число банкротств в течение года. В этих случаях точки сосредоточения массы x , суть просто целые неотрицательные числа $0, 1, 2, \dots$.

2. Непрерывный тип. Величина ξ называется величиной непрерывного типа или обладающей распределением этого типа, если функция распределения $F(x)$ всюду непрерывна *), и если, кроме того, плотность вероятности $f(x) = F'(x)$ существует и непрерывна для всех значений x , за исключением, быть может, некоторого множества, не обладающего ни одной конечной предельной точкой. В этом случае функция распределения $F(x)$ есть

$$F(x) = P(\xi \leq x) = \int_{-\infty}^x f(t) dt.$$

Это распределение не имеет точек сосредоточения массы и, следовательно, вероятность того, что величина ξ принимает частное значение x_0 , равна нулю при любом x_0 :

$$P(\xi = x_0) = 0.$$

Вероятность того, что величина ξ принимает значение, принадлежащее конечному или бесконечному интервалу (a, b) , имеет, таким образом, одно и то же значение, независимо от того, рассматриваем ли мы интервал (a, b) как замкнутый, открытый или полуоткрытый; она равна

$$P(a < \xi < b) = F(b) - F(a) = \int_a^b f(t) dt.$$

Так как вся распределенная масса равна единице, то

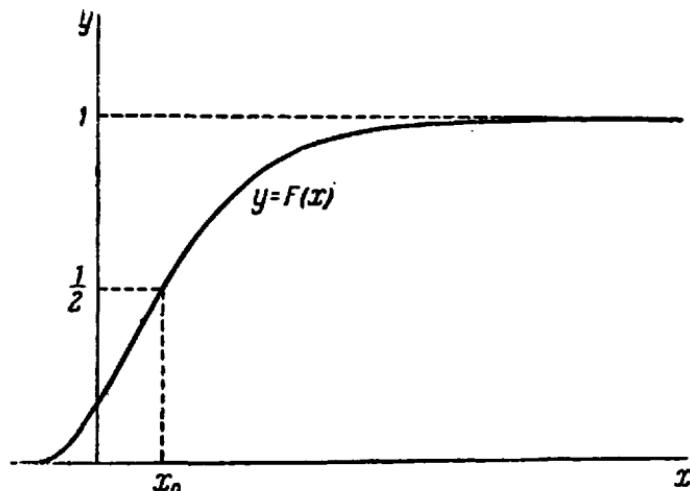
$$\int_{-\infty}^{\infty} f(t) dt = 1.$$

Распределение непрерывного типа может быть графически представлено графиком функции распределения $F(x)$ или плотности вероятности $f(x)$, как показано на фиг. 6 и 7. Кривая $y = f(x)$ называется *кривой плотности* данного распределения.

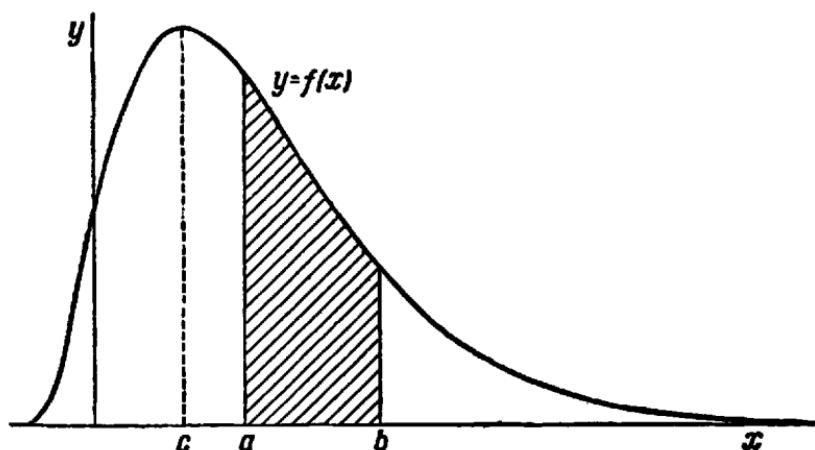
В приложениях к статистике величины непрерывного типа встречаются тогда, когда мы имеем дело с измерением величин, могущих принимать любое значение в некоторых пределах, например, цена товара, рост человека, размер урожая. В этих случаях величины рассматриваются как непрерывные, хотя, строго говоря, фактические данные всегда разрывны, так как каждое

*) Это соответствует случаю $c_1 = 0$, $c_2 = 1$ в (6.6.2).

измерение выражается целым числом, кратным наименьшей единице измерения данной величины. Так, цена выражается в денежных единицах, длина может быть выражена в сантиметрах, вес в килограммах и т. д. Таким образом, всегда, когда для теоретических целей величины этого рода рассматриваются как непрерывные, совершаются некоторая математическая идеализация действительно наблюдаемых фактов.



Фиг. 6. Функция распределения непрерывного типа. (Заметьте, что это распределение имеет единственную медиану в точке x_0 ; см. стр. 201)



Фиг. 7. Функция плотности распределения, показанного на фиг. 6. Заштрихованная площадь соответствует вероятности $P(a < \xi \leq b)$. Распределение имеет единственную моду (см. стр. 202) в точке c . Асимметрия (см. стр. 207) положительна.

15.3. Средние значения. Рассмотрим случайную величину ξ с функцией распределения $F(x)$. Пусть $g(\xi)$ есть функция, интегрируемая на $(-\infty, \infty)$ относительно F (см. параграф 7.2). Интеграл

$$\int_{-\infty}^{\infty} g(x) dF(x)$$

интерпретировался в параграфе 7.4 как взвешенное среднее значение функции $g(x)$ для всех значений x , причем весами служили массы dF , сосредоточенные в окрестности точек x .

В соответствии с этим будем называть такой интеграл *средним значением* или *математическим ожиданием* случайной величины $g(\xi)$ и писать

$$(15.3.1) \quad E(g(\xi)) = \int_{-\infty}^{\infty} g(x) dF(x).$$

Вообще, если ξ есть j -мерная случайная величина с вероятностной функцией $P(S)$ и если $g(\xi)$ есть одномерная функция (см. параграф 14.5; частный случай при $k=1$) от ξ , интегрируемая по R_j относительно $P(S)$, то мы определяем среднее значение $g(\xi)$ соотношением

$$(15.3.2) \quad E(g(\xi)) = \int g(x) dP(S).$$

Для функций $g(\xi) = a(\xi) + ib(\xi)$, принимающих комплексные значения, мы пользуемся той же формулой для определения среднего значения и получаем

$$E(g(\xi)) = E(a(\xi)) + i E(b(\xi)).$$

В тех случаях, когда это не может привести к недоразумению, мы будем писать просто $E g(\xi)$ или $E(g)$ вместо $E(g(\xi))$.

В случае одномерного распределения *дискретного* типа среднее значение сводится, согласно (7.1.8), к конечной или бесконечной сумме:

$$E(g(\xi)) = \sum_v p_v g(x_v),$$

а для непрерывного типа в предположении, что функция $g(x)$ непрерывна всюду кроме, может быть, конечного числа точек, мы получаем, согласно (7.5.5), обычный риманов интеграл:

$$E(g(\xi)) = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

Условие, что функция $g(x)$ должна быть интегрируема на $(-\infty, \infty)$ относительно F , в последних двух частных случаях эквивалентно *абсолютной сходимости* ряда или интеграла, представляющего среднее значение. Таким образом, содержание этого условия заключается в том, что среднее значение существует. Это условие всегда выполнено в случае ограниченной функции $g(\xi)$, как указано в параграфе 7.4.

Рассмотрим теперь две случайные величины ξ и η , определенные в пространствах R' и R'' любого числа измерений, с вероятностными функциями P_1 и P_2 . Пусть $g(\xi)$ и $h(\eta)$ —две действительные или комплексные функции, такие, что средние значения $Eg(\xi)$ и $Eh(\eta)$ существуют. Рассмотрим сумму $g(\xi) + h(\eta)$. Согласно параграфу 14.5, эта сумма есть случайная величина, которую можно рассматривать как функцию составной величины (ξ, η) . Если R обозначает пространство составной величины, а P —соответствующую вероятностную функцию, то среднее значение суммы имеет вид

$$E(g(\xi) + h(\eta)) = \int_R (g(x) + h(y)) dP = \int_{R'} g(x) dP_1 + \int_{R''} h(y) dP_2.$$

Согласно (9.2.2), последние два интеграла сводятся к

$$\int_{R'} g(x) dP_1 = Eg(\xi) \text{ и } \int_{R''} h(y) dP_2 = Eh(\eta),$$

так что

$$(15.3.3) \quad E(g(\xi) + h(\eta)) = Eg(\xi) + Eh(\eta).$$

Это соотношение непосредственно распространяется на произвольное конечное число слагаемых, и мы получаем, таким образом, следующую важную теорему:

Среднее значение суммы случайных величин равно сумме средних значений слагаемых в предположении, что эти последние существуют.

Нужно заметить, что эта теорема была доказана без каких бы то ни было предположений относительно зависимости между слагаемыми. В случае же среднего значения произведения невозможно получить столь общий результат. Употребляя те же значения, что и выше, имеем

$$E(g(\xi)h(\eta)) = \int_R g(x)h(y) dP.$$

Для того чтобы привести этот интеграл к более простому виду, предположим, что величины ξ и η независимы, так что вероятностная функция P удовлетворяет соотношению мультипликативности (14.4.4). Согласно замечанию в конце параграфа 14.5, величины $g(\xi)$ и $h(\eta)$, при этом также независимы. При этом предположении формула для среднего значения сводится, согласно (9.3.1), к

$$(15.3.4) \quad E(g(\xi)h(\eta)) = \int_{\mathbb{R}'} g(x) dP_1 \int_{\mathbb{R}''} h(y) dP_2 = Eg(\xi) Eh(\eta).$$

Распространение на произвольное конечное число сомножителей производится непосредственно, так что имеет место следующая теорема:

Среднее значение произведения независимых случайных величин равно произведению средних значений сомножителей в предположении, что эти средние значения существуют.

Рассмотрим, наконец, некоторые простые частные случаи соотношений (15.3.3) и (15.3.4). Если ξ — одномерная случайная величина, такая, что среднее значение $E(\xi)$, которое получается, если положить в (15.3.1) $g(\xi) = \xi$, существует, то для любых постоянных a и b

$$(15.3.5) \quad E(a\xi + b) = aE(\xi) + b.$$

Полагая $E(\xi) = m$, имеем, в частности,

$$(15.3.6) \quad E(\xi - m) = m - m = 0.$$

Полагая в теореме сложения (15.3.3) $g(\xi) = \xi$, $h(\eta) = \eta$, получим

$$(15.3.7) \quad E(\xi + \eta) = E(\xi) + E(\eta).$$

Если ξ и η независимы, то теорема умножения (15.3.4) дает

$$(15.3.8) \quad E(\xi\eta) = E(\xi) E(\eta).$$

15.4. Моменты. Моменты одномерного распределения были введены в параграфе 7.4. Если для некоторого целого положительного числа v функция x^v интегрируема на $(-\infty, \infty)$ относительно $F(x)$, то среднее значение

$$(15.4.1) \quad a_v = E(\xi^v) = \int_{-\infty}^{\infty} x^v dF(x)$$

называется моментом порядка v , или просто v -м моментом этого распределения или соответствующей случайной величины, и мы говорим,

что момент конечен или существует. Очевидно, что a_0 всегда существует и равен единице.

Если a_v существует, то функция $|x|^v$ также интегрируема, так что существует v -й абсолютный момент

$$(15.4.2) \quad \beta_v = E(|\xi|^v) = \int_{-\infty}^{\infty} |x|^v dF(x).$$

Следовательно, если a_k существует, то существуют a_v и β_v , для $0 \leq v \leq k$.

Для распределения дискретного типа моменты, согласно параграфу 15.3, выражаются рядами

$$a_v = \sum_i p_i x_i^v,$$

а для распределения непрерывного типа — римановыми интегралами

$$a_v = \int_{-\infty}^{\infty} x^v f(x) dx.$$

Только когда ряд или интеграл, представляющий момент, абсолютно сходится, мы говорим, что этот момент существует.

Первый момент a_1 равен среднему значению или, короче, среднему самой случайной величины и будет часто обозначаться через m :

$$a_1 = E(\xi) = m.$$

Если c — постоянная, то величины

$$E[(\xi - c)^v] = \int_{-\infty}^{\infty} (x - c)^v dF(x)$$

называются моментами относительно точки c . При $c = 0$ получаем обычные моменты. Аналогично определяются абсолютные моменты относительно c . Моменты относительно среднего m называются часто центральными моментами. Они особенно важны и обозначаются

$$(15.4.3) \quad \mu_v = E[(\xi - m)^v] = \int_{-\infty}^{\infty} (x - m)^v dF(x).$$

Раскрывая $(x - m)^\nu$, находим

$$(15.4.4) \quad \begin{aligned} \mu_0 &= 1, \\ \mu_1 &= 0, \\ \mu_2 &= a_2 - m^2, \\ \mu_3 &= a_3 - 3ma_2 + 2m^3, \\ \mu_4 &= a_4 - 4ma_3 + 6m^2a_2 - 3m^4, \\ &\dots \end{aligned}$$

Для второго момента относительно точки c имеем

$$E[(\xi - c)^2] = E[(\xi - m + m - c)^2] = \mu_2 + (c - m)^2 \geq \mu^2,$$

так что второй момент имеет минимальное значение при $c = m$.

Моменты любой функции $g(\xi)$ суть средние значения степеней $g(\xi)$. В частном случае линейной функции $g(\xi) = a\xi + b$ момент a' , задается выражением

$$a' = E[(a\xi + b)^\nu] = a^\nu a_\nu + \binom{\nu}{1} a^{\nu-1} b a_{\nu-1} + \dots + b^\nu.$$

В параграфе 7.4 было дано простое достаточное условие того, что момент данного порядка k существует. Заметим далее, что если случайная величина ξ ограничена, т. е. если можно найти конечные a и b , такие, что $P(a < \xi < b) = 1$, то все моменты конечны и $|a_\nu| \leq |a|^\nu + |b|^\nu$.

Докажем теперь одно важное неравенство для абсолютных моментов β_ν , определенных соотношением (15.4.2). Квадратичная форма переменных u и v

$$\int_{-\infty}^{\infty} (u|x|^{\frac{\nu-1}{2}} + v|x|^{\frac{\nu+1}{2}})^2 dF(x) = \beta_{\nu-1} u^2 + 2\beta_\nu uv + \beta_{\nu+1} v^2,$$

очевидно, неотрицательна. Следовательно, согласно параграфу 11.10 детерминант этой формы неотрицателен, т. е. $\beta_{\nu-1}\beta_{\nu+1} - \beta_\nu^2 \geq 0$ или

$$(15.4.5) \quad \beta_\nu^{2\nu} \leq \beta_{\nu-1}^\nu \beta_{\nu+1}^\nu.$$

Полагая последовательно $\nu = 1, 2, \dots, \nu$ и перемножая все полученные таким образом неравенства, получим $\beta_{\nu+1}^{\nu+1} \leq \beta_{\nu+1}^\nu$ или, окончательно,

$$(15.4.6) \quad \beta_\nu^{\frac{1}{\nu}} \leq \beta_{\nu+1}^{\frac{1}{\nu+1}} \quad (\nu = 1, 2, \dots).$$

Часто важно знать, однозначно ли задается распределение последовательностью своих моментов. Мы не будем углубляться в пол-

ное исследование этой трудной проблемы и удовольствуемся доказательством следующего критерия, который часто бывает полезен.

Пусть $a_0 = 1, a_1, a_2, \dots$ суть моменты некоторой функции распределения $F(x)$, каждый из которых предполагается конечным.

Предположим, что ряд $\sum_0^{\infty} \frac{a_n}{n!} r^n$ абсолютно сходится при некотором $r > 0$. Тогда $F(x)$ есть единственная функция распределения, обладающая моментами a_0, a_1, a_2, \dots .

Покажем сперва, что $\frac{\beta_n}{n!} r^n \rightarrow 0$ при $n \rightarrow \infty$. Для четных n это следует прямо из нашего предположения. Для нечетных n это также верно, так как, в силу (15.4.5), имеем неравенство

$$\frac{\beta_n}{n!} r^n \leq \left(\frac{\beta_{n-1}}{(n-1)!} r^{n-1} \right)^{\frac{1}{2}} \left(\frac{\beta_{n+1}}{(n+1)!} r^{n+1} \right)^{\frac{1}{2}} \sqrt{\frac{n+1}{n}}.$$

Для любого целого $n > 0$ и любого действительного z имеем разложение Маклорена

$$e^{iz} = \sum_0^{n-1} \frac{(iz)^n}{n!} + \theta \frac{z^n}{n!},$$

где θ — действительное или комплексное число, модуль которого не превосходит единицы. Отсюда, посредством (10.1.2), мы получим для характеристической функции $\varphi(t)$ функции $F(x)$ такое разложение:

$$\begin{aligned} \varphi(t+h) &= \int_{-\infty}^{\infty} e^{ithx} e^{itx} dF(x) = \\ &= \sum_0^{n-1} \frac{(ih)^n}{n!} \int_{-\infty}^{\infty} x^n e^{ithx} dF(x) + \theta \frac{h^n}{n!} \int_{-\infty}^{\infty} |x|^n dF(x) = \\ &= \sum_0^{n-1} \frac{h^n}{n!} \varphi^{(n)}(t) + \theta \frac{\beta_n h^n}{n!}. \end{aligned}$$

При $|h| < r$ остаток стремится к нулю, так что при любом t характеристическая функция $\varphi(t+h)$ может быть разложена в ряд Тейлора, сходящийся при $|h| < r$.

Полагая сперва $t = 0$, находим, что ряд (пишем теперь t вместо h)

$$(15.4.7) \quad \varphi(t) = \sum_0^{\infty} \frac{a_n}{n!} (it)^n$$

представляет функцию $\varphi(t)$ по крайней мере на интервале $-r < t < r$. В этом интервале функция $\varphi(t)$ определена, таким образом, своими моментами a_i , однозначно. В точках $t = \pm \frac{1}{2}r$ ряд, получаемый дифференцированием равенства (15.4.7) любое число раз, сходится, так что все производные $\varphi^{(n)}\left(\pm \frac{1}{2}r\right)$ могут быть вычислены из соотношения (15.4.7), т. е. по моментам a_n . Эти производные служат коэффициентами разложения в ряд Тейлора для $\varphi\left(\pm \frac{1}{2}r + h\right)$, которое сходится и представляет $\varphi(t)$ при $|h| < r$. Тем самым область определения $\varphi(t)$ расширяется теперь до интервала $-\frac{3}{2}r < t < \frac{3}{2}r$. Из последнего разложения мы можем теперь вычислить производные $\varphi^{(n)}(t)$ в точках $t = \pm r$ и использовать их в качестве коэффициентов ряда Тейлора для $\varphi(\pm r + h)$ и т. д. Этот процесс, который можно продолжать неограниченно, показывает, что моменты a_i однозначно определяют характеристическую функцию $\varphi(t)$ для всех значений t^*). Из теоремы единственности (10.3.1) следует, что функция распределения также однозначно определена моментами a_i , и наша теорема доказана.

В частном случае, когда $F(x)$ есть функция распределения ограниченной случайной величины, из замечания, сделанного выше, следует, что условия этой теоремы всегда выполнены.

15.5. Характеристики расположения. В приложениях бывает важно суметь описать распределение хотя бы в общих чертах, посредством немногих простых параметров. Прежде всего, часто мы хотим характеризовать расположение значений случайной величины в данном распределении с помощью некоторого *типичного значения* этой величины, которое можно представлять себе как *центральную точку* распределения. Существуют различные способы вычисления таких типичных параметров. Рассмотрим здесь три наиболее важных характеристики такого рода, а именно *среднее, медиану и моду*.

Среднее $E(\xi) = m$ — первый момент распределения — было уже определено в предыдущем параграфе. В терминах нашей механической интерпретации распределения вероятностей как распределения массы среднее имеет важный конкретный смысл: это — абсцисса

*) Этот процесс в теории аналитических функций носит название *аналитического продолжения*.

центра тяжести в данном распределении массы (см. параграф 7.4). Это свойство придает среднему роль некоторого типичного параметра.

Если точка x_0 разделяет всю распределенную массу на две равные части, каждая из которых содержит массу $\frac{1}{2}$, то x_0 называется *медианой* распределения. Таким образом, медиана есть любой корень уравнения $F(x) = \frac{1}{2}$. Чтобы исследовать возможные случаи, рассмотрим кривую $F(x)$, причем каждый вертикальный отрезок, соответствующий скачку функции F , будем рассматривать как часть нашей кривой. Мы получим некоторую связную неубывающую кривую (см. фиг. 4 и 6). Эта кривая имеет по крайней мере одну точку пересечения с прямой $y = \frac{1}{2}$. Если существует только одна точка пересечения, то абсцисса этой точки есть единственная медиана распределения (см. фиг. 6). Может случиться, однако, что кривая $F(x)$ и прямая $y = \frac{1}{2}$ имеют общий замкнутый интервал (см. фиг. 4). В этом случае абсцисса любой точки интервала, удовлетворяющего уравнению $F(x) = \frac{1}{2}$, может быть принята за медиану распределения.

Мы видим, таким образом, что *каждое распределение имеет по крайней мере одну медиану*. Мы будем различать *определенный случай*, когда медиана единственна, и *неопределенный случай*, когда медианы заполняют некоторый замкнутый интервал.

Среднее значение может вообще не существовать. Однако даже в тех случаях, когда оно существует, иногда предпочитают брать в качестве типичного параметра медиану, так как на среднее могут оказывать значительное влияние массы, расположенные на большом расстоянии от основного сгустка распределения.

Как показано в предыдущем параграфе, среднее m характеризуется тем, что второй момент относительно c имеет минимальное значение при $c = m$. Аналогичным свойством обладает и медиана: *первый абсолютный момент $E(|\xi - c|)$ принимает минимальное значение, когда c есть медиана*. Это справедливо даже в неопределенном случае, и момент $E(|\xi - c|)$ имеет одинаковое значение при любом возможном значении медианы c . Действительно, обозначая медиану (или, в неопределенном случае, любое значение медианы)

через μ , будем иметь

$$E(|\xi - c|) = \begin{cases} E(|\xi - \mu|) + 2 \int_{\mu}^c (c - x) dF(x) \text{ при } c > \mu, \\ E(|\xi - \mu|) + 2 \int_c^{\mu} (x - c) dF(x) \text{ при } c < \mu. \end{cases}$$

Вторые слагаемые в правых частях, очевидно, положительны, если только c не совпадает с μ или (в неопределенном случае) с каким-нибудь другим значением медианы, когда эти слагаемые обращаются в нуль.*). Доказательство приведенных здесь соотношений представляется читателю в качестве упражнений.

Мода распределения определяется только для двух простых типов распределений, рассмотренных в параграфе 15.2. Для *непрерывного* распределения *модой* называется любая точка x_0 максимума плотности вероятности $f(x)$. Единственная мода существует, таким образом, только в случае, когда кривая плотности $y = f(x)$ имеет один максимум (см. фиг. 7). Такие *унимодальные* распределения часто встречаются в приложениях к статистике. Если кривая плотности имеет два или более максимумов, то распределение называется соответственно *бимодальным* или *мультимодальным*. Для дискретного распределения предположим, что точки сосредоточения массы x , расположены в порядке возрастания. Тогда точка x_0 называется *модой* распределения, если $p_{x_0} > p_{x_0-1}$ и $p_{x_0} > p_{x_0+1}$. Унимодальное, бимодальное и мультимодальное распределения определяются в этом случае так же, как и для непрерывных распределений.

В частном случае, когда распределение симметрично относительно некоторой точки a , имеем $F(a+x) + F(a-x) = 1$, если только $a \pm x$ быть точками непрерывности функции F . Отсюда видно, что среднее (если оно существует) и медиана равны a . Если, кроме того, распределение унимодально, то мода также равна a .

15.6. Характеристики рассеяния. Если известно типичное значение случайной величины, то часто бывает необходимо вычислить некоторый параметр, показывающий, насколько широко разбросаны

*) В частном случае, когда μ есть точка разрыва функции F , обычное определение интегралов в правых частях должно быть несколько изменено, так как эти интегралы должны тогда включать *половину* приращения, вызванного разрывом.

значения этой величины по каждую сторону от типичного значения. Параметр такого рода называется характеристикой *рассеяния* или, иногда, характеристикой *сосредоточения*. Рассеяние и сосредоточение находятся, конечно, в обратной зависимости: чем больше рассеяние, тем меньше сосредоточение, и наоборот.

Если наше типичное значение есть среднее m распределений, то в качестве меры рассеяния естественно рассматривать второй момент μ_2 относительно среднего. Он называется *дисперсией* величины и представляет *момент инерции* распределенной массы относительно перпендикулярной оси, проходящей через центр тяжести (см. параграф 7.4). Всегда, конечно, $\mu_2 \geq 0$. Если $\mu_2 = 0$, то из определения μ_2 следует, что вся распределенная масса содержится в точке m (см. параграф 16.1).

Чтобы иметь характеристику рассеяния той же размерности, что и сама случайная величина ξ , часто предпочитают рассматривать неотрицательное значение квадратного корня из μ_2 , которое называется *стандартным* (или *средним квадратическим*) *отклонением* величины и обозначается через $D(\xi)$ или иногда через σ . Для любой величины, для которой второй момент существует, имеем

$$D^2(\xi) = \sigma^2 = \mu_2 = E[(\xi - E\xi)^2] = E(\xi^2) - E^2(\xi).$$

Из (15.3.5) следует, что для любых постоянных a и b

$$D(a\xi + b) = |a| D(\xi).$$

Если величина ξ имеет среднее значение m и стандартное отклонение σ , то часто рассматривают соответствующую *нормированную величину* $\frac{\xi - m}{\sigma}$, представляющую отклонение ξ от среднего m , выражённое в единицах стандартного отклонения σ . Из последнего соотношения и из (15.3.5) следует, что нормированная величина имеет среднее значение, равное нулю, и стандартное отклонение, равное единице:

$$E\left(\frac{\xi - m}{\sigma}\right) = 0, \quad D\left(\frac{\xi - m}{\sigma}\right) = 1.$$

Если ξ и η — независимые случайные величины, то из (15.3.8) следует, что

$$(15.6.1) \quad D^2(\xi + \eta) = D^2(\xi) + D^2(\eta).$$

Это соотношение непосредственно распространяется на любое конечное число слагаемых. Если случайные величины ξ_1, \dots, ξ_n

независимы, то

$$(15.6.2) \quad D^2(\xi_1 + \dots + \xi_n) = D^2(\xi_1) + \dots + D^2(\xi_n).$$

Мы уже видели, что второй момент минимален, когда он берется относительно среднего, а первый абсолютный момент минимален, когда он берется относительно медианы (см. параграфы 15.4 и 15.5). Если в качестве типичного значения взять медиану μ , то за характеристику рассеяния естественно принять первый абсолютный момент

$$E(|\xi - \mu|),$$

который называется средним отклонением величины. Иногда средним отклонением называется первый абсолютный момент относительно среднего значения, однако употреблять этот термин в таком смысле не рекомендуется.

Точно так же, как с помощью уравнения $F(x) = \frac{1}{2}$ была определена медиана, можно с помощью уравнения $F(\zeta_p) = p$, где p есть некоторое данное число, такое, что $0 < p < 1$, определить величину ζ_p , которую будем называть *квантилью* распределения порядка p . Подобно медиане квантиль ζ_p может быть неопределенной. Квантиль $\zeta_{1/4}$ есть, конечно, сама медиана. Знание квантилей ζ_p для некоторого разумно выбранного множества значений p , как, например, $p = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$ или $p = 0,1; 0,2; \dots; 0,9$ дает, очевидно, хорошее представление о расположении и рассеянии данного распределения. Квантили $\zeta_{1/4}$ и $\zeta_{3/4}$ называются соответственно *нижней* и *верхней квартилью*, а величины $\zeta_{0,1}; \zeta_{0,2}; \dots$ называются *декилиями*. Величина $\frac{1}{2}(\zeta_{3/4} - \zeta_{1/4})$ иногда используется как характеристика рассеяния и называется *семи-интерквартильной широтой* *).

Если вся распределенная масса заключена в конечном интервале, то существует верхняя граница g всех точек x , для которых $F(x) = 0$, и нижняя граница G всех точек x , для которых $F(x) = 1$. Интервал (g, G) содержит всю распределенную массу. Его длина $G - g$

*) В книге В. И. Романовского „Математическая статистика“ эта величина носит название „вероятного отклонения“. (Прим. перев.)

называется *широтой* распределения и может быть использована как характеристика рассеяния.

Если интервал (g, G) известен, то мы уже имеем некоторое представление о расположении и рассеянии распределения. Для распределения, широта которого конечна, такие интервалы, как $(m - \sigma, m + \sigma)$ или $(\zeta_{1/2}, \zeta_{3/2})$, хотя они и не содержат всей распределенной массы, могут быть также использованы в качестве некоторых геометрических характеристик расположения и рассеяния распределения (см. параграф 21.10).

Все характеристики расположения, рассеяния и других аналогичных свойств в большой степени произвольны. Это вполне естественно, так как свойства, описываемые такими параметрами, определены слишком расплывчато, чтобы каждое из них можно было точно характеризовать с помощью одного числа. Каждая характеристика имеет свои достоинства и недостатки, и характеристика, вполне пригодная в одном случае, может оказаться более или менее бесполезной в другом.

Если, в частности, за характеристику рассеяния выбрать дисперсию σ^2 или стандартное отклонение σ , то мерой разброса массы в распределении со средним $m = 0$ будет служить величина

$$E(\xi^2) = \int_{-\infty}^{\infty} x^2 dF(x).$$

Сосредоточение случайной величины около точки $m = 0$ будет изменяться той же величиной: чем меньше $E(\xi^2)$, тем больше это сосредоточение, и наоборот. Итак, $E(\xi^2)$ рассматривается как мера отклонения ξ от нуля. Это — одна из форм выражения известного *принципа наименьших квадратов*, с которым мы будем часто сталкиваться в дальнейшем. Из сказанного выше следует, что нет логической необходимости принимать этот принцип. Напротив, принятие или непринятие этого принципа есть в значительной степени вопрос соглашения. Основной довод в пользу этого принципа заключается в сравнительной простоте приемов вычислений, к которым он приводит. Так, например, имеет место простое правило сложения (15.6.2), в то время как другие характеристики рассеяния, рассмотренные выше, подобному правилу не подчиняются.

15.7. Теорема Чебышева. Докажем теперь следующее обобщение теоремы, принадлежащей Чебышеву:

Пусть $g(\xi)$ есть неотрицательная функция случайной величины ξ . Тогда для любого $K > 0$ имеем

$$(15.7.1) \quad P[g(\xi) \geq K] \leq \frac{Eg(\xi)}{K},$$

где P обозначает, как обычно, вероятностную функцию величины ξ .

Если обозначить через S множество всех ξ , удовлетворяющих неравенству $g(\xi) \geq K$, то справедливость теоремы вытекает непосредственно из соотношения

$$Eg(\xi) = \int_{-\infty}^{\infty} g(x) dF \geq K \int_S dF = KP(S).$$

Очевидно, что теорема остается справедливой и сохраняет то же доказательство, если заменить ξ случайной величиной ξ любого числа измерений.

Полагая, в частности, $g(\xi) = (\xi - m)^2$, $K = k^2\sigma^2$, где m и σ суть среднее значение и стандартное отклонение величины ξ , получим для любого $k > 0$ неравенство Биенэме — Чебышева:

$$(15.7.2) \quad P(|\xi - m| \geq k\sigma) \leq \frac{1}{k^2}.$$

Это неравенство показывает, что количество распределенной массы, лежащей вне интервала $m - k\sigma < \xi < m + k\sigma$, превосходит $\frac{1}{k^2}$ и, таким образом, дает хорошее представление о том, в каком смысле σ может быть использована как мера рассеяния.

Для распределения со средним значением m и стандартным отклонением σ , при котором в каждой точке $x = m \pm k\sigma$ заключена масса $\frac{1}{2k^2}$, а в точке $x = m$ масса $1 - \frac{1}{k^2}$, имеем $P(|\xi - m| \geq k\sigma) = \frac{1}{k^2}$ и, таким образом, видно, что верхний предел вероятности (15.7.2) не может, вообще говоря, быть улучшен.

С другой стороны, если ограничиться некоторыми классами распределений, то неравенство (15.7.2) иногда удается улучшить. Так, Гауссом было показано (в 1821 г.), что для **униодальных** распределений (см. параграф 15.5)

непрерывного типа при любом $k > 0$ имеем

$$(15.7.3) \quad P(|\xi - x_0| \geq k\tau) \leq \frac{4}{9k^2},$$

где x_0 — мода, а $\tau^2 = \sigma^2 + (x_0 - m)^2$ — второй момент относительно моды. Простое доказательство этого соотношения указано в упражнении 4 на стр. 283. Отсюда получается следующее неравенство для отклонения от среднего:

$$(15.7.4) \quad P(|\xi - m| \geq k\sigma) \leq \frac{4}{9} \cdot \frac{1 + s^2}{(k - |s|)^2}$$

при любом $k > |s|$, где s обозначает пирсоновскую меру асимметрии (см. выражение (15.8.3)). Для небольших значений $|s|$ это неравенство часто дает меньшее значение предела левой части, чем (15.7.2). Так, если $|s| < 0,25$, то вероятность отклонения, превышающего 3σ , согласно (15.7.4), меньше, чем 0,0624, в то время как (15.7.2) дает менее точный предел 0,1111. Для вероятности отклонения, превышающего 4σ , соотношение (15.7.4) дает 0,0336 а (15.7.2) дает 0,0625.

15.8. Характеристики асимметрии и эксцесса. В симметричном распределении каждый момент нечетного порядка относительно среднего, очевидно, равен нулю (если он вообще существует). Любой момент, не равный нулю, можно, таким образом, рассматривать как характеристику *асимметрии* данного распределения. Простейшая из этих характеристик μ_3 выражается в единицах измерения самой случайной величины в третьей степени. Чтобы свести размерность этой величины к нулю и получить, таким образом, *абсолютную* характеристику, рассматривают в качестве меры асимметрии отношение

$$(15.8.1) \quad \gamma_1 = \frac{\mu_3}{\sigma^3}.$$

Мы будем называть γ_1 *коэффициентом асимметрии*.

В приложениях к статистике часто встречаются унимодальные непрерывные распределения типа, изображенного на фиг. 7 (см. стр. 193), когда кривая плотности такова, что по одну сторону от моды расположена ее «длинная часть», а по другую — «короткая часть». На фиг. 7 длинная часть кривой плотности расположена справа от моды и в выражении μ_3 кубы положительных отклонений перевесят отрицательные кубы, так что коэффициент γ_1 будет положителен. В таких случаях будем говорить, что распределение обладает *положительной асимметрией*. Аналогично мы будем говорить об *отрицательной асимметрии*, если коэффициент γ_1 отрицателен; в этом случае длинная часть кривой плотности расположена слева от моды.

Выражая четвертый момент μ_4 в абсолютных единицах, как это было сделано выше для μ_3 , получим так называемый *коэффициент эксцесса*

$$(15.8.2) \quad \gamma_2 = \frac{\mu_4}{\sigma^4} - 3,$$

который иногда используется в качестве характеристики сглаженности кривой плотности около ее центра. Для нормального распределения (см. параграф 17.2) коэффициент эксцесса γ_2 равен нулю. Положительное значение γ_2 указывает на то, что кривая плотности в окрестности моды имеет более высокую и более острую вершину, чем нормальная кривая. Обратно, отрицательное значение γ_2 указывает на более низкий и более плоский характер вершины сравнительно с нормальной кривой. В первом случае обычно говорят о *положительном эксцессе* по сравнению с нормальной кривой, во втором — об *отрицательном эксцессе*. Это словоупотребление, однако, имеет свои недостатки см. параграф 17.6).

Часто в литературе вместо величин γ_1 и γ_2 используют величины $\beta_1 = \gamma_1^2$ и $\beta_2 = \gamma_2 + 3$.

Предлагались и многие другие характеристики асимметрии и эксцесса. Так, К. Пирсон ввел в качестве меры асимметрии разность между средним и модой, поделенную на стандартное отклонение:

$$(15.8.3) \quad s = \frac{m - x_0}{\sigma}.$$

Для класса распределений, входящих в систему Пирсона (см. параграф 19.4) можно показать, что

$$s = \frac{\gamma_1(\gamma_2 + 6)}{2(5\gamma_2^2 - 6\gamma_1^2 + 6)}.$$

Если γ_1 и γ_2 малы, то это приближенно дает

$$s = \frac{1}{2} \gamma_1 \text{ или } x_0 = m - \frac{1}{2} \gamma_1 \sigma.$$

Последнее соотношение также приближенно выполняется для распределений, задаваемых разложениями Эджвортса и Шарлье (см. параграфы 17.6—17.7). Шарлье употребляет коэффициент $S = -\frac{1}{2} \gamma_1$ в качестве меры асимметрии и $E = \frac{1}{8} \gamma_2$ как меру эксцесса.

15.9. Характеристические функции. Среднее значение функции $e^{it\xi}$

$$(15.9.1) \quad \varphi(t) = E(e^{it\xi}) = \int_{-\infty}^{\infty} e^{itx} dF(x),$$

есть функция действительной переменной t и называется *характеристической функцией* случайной величины ξ или *характеристической функцией* соответствующего распределения. Математическую теорию характеристических функций читатель найдет в главе 10. Там, в частности, было установлено, что существует взаимнооднозначное соответствие между распределениями и характеристическими функциями. Если два распределения совпадают, то их характеристические функции также совпадают, и обратно. Из этого свойства вытекают важные следствия. Во многих вопросах, когда бывает нужно найти распределение некоторой данной случайной величины, сравнительно просто найти ее характеристическую функцию. Если окажется, что она совпадает с характеристикой функцией некоторого уже известного распределения, то наше искомое распределение должно совпадать с этим последним.

Характеристическая функция функции $g(\xi)$ есть среднее значение $e^{itg(\xi)}$. В частном случае линейной функции $g(\xi) = a\xi + b$ характеристическая функция имеет вид

$$(15.9.2) \quad E(e^{it(a\xi+b)}) = e^{bit}\varphi(at).$$

Так, например, функция $-\xi$ имеет своей характеристической функцией $\varphi(-t) = \overline{\varphi(t)}$. Далее, нормированная случайная величина $\frac{\xi - m}{\sigma}$ имеет характеристическую функцию

$$E(e^{it\frac{\xi-m}{\sigma}}) = e^{-\frac{mit}{\sigma}} \varphi\left(\frac{t}{\sigma}\right).$$

15.10. Семи-инварианты. Если k -й момент распределения существует, то, согласно (10.1.3), характеристическая функция может быть разложена в ряд Маклорена при малых значениях t :

$$(15.10.1) \quad \varphi(t) = 1 + \sum_1^k \frac{a_v}{v!} (it)^v + o(t^k).$$

Для функции $\log(1+z)$ соответствующее разложение имеет вид

$$\log(1+z) = \frac{z}{1} - \frac{z^2}{2} + \dots + \frac{z^k}{k} + o(z^k).$$

Заменяя $1+z$ через $\varphi(t)$, получим после перестановки слагаемых такое разложение:

$$(15.10.2) \quad \log \varphi(t) = \sum_1^k \frac{x_n}{n!} (it)^n + o(t^k).$$

Коэффициенты x_n , введенные Тиле [37], называются *семи-инвариантами* или *кумюлянтами* распределения.

Для того чтобы вывести соотношения между моментами a_n и семи-инвариантами x_n , воспользуемся формально тождествами

$$\begin{aligned} \log \varphi(t) &= \log \left(1 + \sum_1^{\infty} \frac{a_n}{n!} (it)^n \right) = \sum_1^{\infty} \frac{x_n}{n!} (it)^n, \\ \varphi(t) &= 1 + \sum_1^{\infty} \frac{a_n}{n!} (it)^n = e^{\sum_1^{\infty} \frac{x_n}{n!} (it)^n}, \end{aligned}$$

не исследуя вопроса о существовании моментов и о сходимости ряда. Очевидно, что x_n есть полином от переменных a_1, \dots, a_n и, обратно a_n есть полином от x_1, \dots, x_n . В частности, имеем

$$\begin{aligned} (15.10.3) \quad x_1 &= a_1 = m, \\ x_2 &= a_2 - a_1^2 = a^2, \\ x_3 &= a_3 - 3a_1a_2 + 2a_1^3, \\ x_4 &= a_4 - 3a_2^2 - 4a_1a_3 + 12a_1^2a_2 - 6a_1^4, \\ &\dots \end{aligned}$$

и обратно

$$\begin{aligned} (15.10.4) \quad a_1 &= x_1, \\ a_2 &= x_2 + x_1^2, \\ a_3 &= x_3 + 3x_1x_2 + x_1^3, \\ a_4 &= x_4 + 3x_2^2 + 4x_1x_3 + 6x_1^2x_2 + x_1^4, \\ &\dots \end{aligned}$$

Можно выразить x_n через центральные моменты μ_n :

$$\begin{aligned} (15.10.5) \quad x_1 &= m, \\ x_2 &= \mu_2 = \sigma^2, \\ x_3 &= \mu_3, \\ x_4 &= \mu_4 - 3\mu_2^2, \\ x_5 &= \mu_5 - 10\mu_2\mu_3, \\ x_6 &= \mu_6 - 15\mu_2\mu_4 - 10\mu_3^2 + 30\mu_2^3, \\ &\dots \end{aligned}$$

Отсюда коэффициенты асимметрии и эксцесса, введенные в параграфе 15.8, равны соответственно

$$\gamma_1 = \frac{x_3}{x_2^{3/2}} \text{ и } \gamma_2 = \frac{x_4}{x_2^2}.$$

Семи-инвариантны x'_v линейной функции $g(\xi) = a\xi + b$ находятся согласно (15.9.2), из разложения

$$\log [e^{b/t} \varphi(at)] = \sum_1^k \frac{x'_v}{v!} (it)^v + o(t^k).$$

Сравнивая с (15.10.2), получим

$$x'_v = ax_v + b \text{ и } x'_v = a^v x_v \text{ при } v > 1.$$

15.11. Независимые величины. Пусть ξ и η — случайные величины с функциями распределения F_1 и F_2 и с вероятностной функцией P их совместного распределения. Согласно (14.4.5), для независимости ξ и η необходимо и достаточно, чтобы для всех x и y совместная функция распределения этих величин удовлетворяла соотношению *)

$$(15.11.1) \quad F(x, y) = P(\xi \leq x, \eta \leq y) = F_1(x) \cdot F_2(y).$$

Если обе величины обладают распределениями одного и того же простого типа, то условие независимости можно выразить в более удобной форме.

Рассмотрим сперва случай, когда обе рассматриваемые величины — *дискретного* типа с распределениями, заданными равенствами

$$P(\xi = x_v) = p_v, \quad P(\eta = y_v) = q_v,$$

где $v = 1, 2, \dots$. Тогда легко видеть, что условие независимости (15.11.1) эквивалентно соотношениям

$$(15.11.2) \quad P(\xi = x_\mu, \eta = y_v) = p_\mu q_v,$$

для всех значений μ и v .

*) Другое необходимое и достаточное условие будет дано в параграфе 21.3.

В случае, когда обе величины — непрерывного типа, дифференцируем равенство (15.11.1) по x и по y и получим

$$(15.11.3) \quad f(x, y) = \frac{\partial^2 F}{\partial x \partial y} = f_1(x) f_2(y),$$

где f_1 и f_2 — плотности вероятности величин ξ и η , а f есть, согласно параграфу 8.4, плотность вероятности совместного распределения обеих величин или совместная функция плотности величин ξ и η . Обратно, из (15.11.3), интегрируя, получаем (15.11.1).

Таким образом, (15.11.2) есть необходимое и достаточное условие независимости двух величин дискретного типа, а (15.11.3) необходимое и достаточное условие независимости двух величин непрерывного типа. Оба условия непосредственно распространяются на любое конечное число случайных величин.

15.12. Сложение независимых случайных величин. Пусть ξ и η — независимые случайные величины с известными распределениями. Согласно параграфу 14.5, сумма $\xi + \eta$ имеет распределение, однозначно определенное распределениями величин ξ и η . Во многих вопросах бывает нужно выразить функцию распределения, характеристическую функцию, моменты и т. д. этого распределения через соответствующие функции и величины данных распределений ξ и η . Эта задача может быть, конечно, обобщена на случай суммы произвольного конечного числа независимых величин.

Рассмотрим сперва характеристические функции. Пусть $\varphi_1(t)$, $\varphi_2(t)$, $\varphi(t)$ обозначают соответственно характеристические функции величин ξ , η , $\xi + \eta$. Тогда, согласно теореме (15.3.4) о среднем значении произведения независимых сомножителей, имеем

$$\varphi(t) = E(e^{it(\xi+\eta)}) = E(e^{it\xi} e^{it\eta}) = E(e^{it\xi}) E(e^{it\eta}) = \varphi_1(t) \varphi_2(t).$$

Это соотношение непосредственно распространяется на случай произвольного конечного числа случайных величин: если ξ_1, \dots, ξ_n — независимые величины с характеристическими функциями $\varphi_1(t), \dots, \varphi_n(t)$ то характеристическая функция $\varphi(t)$ суммы $\xi_1 + \dots + \xi_n$ есть

$$(15.12.1) \quad \varphi(t) = \varphi_1(t) \varphi_2(t) \dots \varphi_n(t).$$

Итак, мы имеем следующую важную теорему, выражющую основное свойство характеристических функций:

Характеристическая функция суммы независимых случайных величин равна произведению характеристических функций сомножителей.

Выразим теперь функцию распределения суммы $\xi + \eta$ через функции распределения F_1 и F_2 слагаемых. Применим к этой задаче общий метод нахождения функции распределения по ее характеристической функции (см. параграфы 10.3 и 15.9). Рассмотрим интеграл

$$F(x) = \int_{-\infty}^{\infty} F_1(x-z) dF_2(z).$$

Так как функция F_1 ограничена, то этот интеграл, согласно параграфу 7.1, определен и имеет конечное значение при любом x . $F_1(x-z)$ при любом фиксированном z есть неубывающая функция от x , всюду непрерывная справа и стремящаяся к 1 при $x \rightarrow +\infty$ и к 0 при $x \rightarrow -\infty$. Рассмотрим разность $F(x+h) - F(x)$, где $h > 0$. Из (7.1.4) следует, что эта разность неотрицательна, а из (7.3.1) — что она стремится к нулю при $h \rightarrow 0$. Далее, из (7.3.1) следует, что $F(x)$ стремится к 1 при $x \rightarrow +\infty$ и к 0 при $x \rightarrow -\infty$.

Итак, $F(x)$ есть функция распределения. Соответствующая характеристическая функция

$$\int_{-\infty}^{\infty} e^{itx} dF(x)$$

есть, согласно (7.5.6), предел при $n \rightarrow \infty$ сумм s_n вида

$$s_n = \sum_1^n e^{itx_v} [F(x_v) - F(x_{v-1})]$$

в предположении, что максимум длины подинтервалов (x_{v-1}, x_v) стремится к нулю и при этом $x_0 \rightarrow -\infty$ и $x_n \rightarrow +\infty$. Вводя в это соотношение интегральное выражение функции $F(x)$, получим

$$s_n = \int_{-\infty}^{\infty} s'_n e^{itz} dF_2(z),$$

где

$$s'_n = \sum_1^n e^{itz'} [F_1(x'_v) - F_1(x'_{v-1})],$$

$$x'_v = x_v - z.$$

Для каждого фиксированного z суммы s'_n при $n \rightarrow \infty$ стремятся к пределу

$$\lim s'_n = \int_{-\infty}^{\infty} e^{itz} dF_1(x) = \varphi_1(t).$$

Далее, величины s'_n равномерно ограничены, так как

$$|s'_n| \leq \sum_1^n [F_1(x'_v) - F_1(x'_{v-1})] \leq 1.$$

Согласно (7.1.7), отсюда следует, что

$$\lim s_n = \varphi_1(t) \int_{-\infty}^{\infty} e^{itz} dF_2(z) = \varphi_1(t) \varphi_2(t).$$

Таким образом, характеристическая функция, соответствующая функции распределения $F(x)$, совпадает с характеристической функцией $\varphi(t) = \varphi_1(t) \varphi_2(t)$ суммы $\xi + \eta$, следовательно, $F(x)$ есть искомая функция распределения. Так как, очевидно, в нашем выводе функции F_1 и F_2 можно всюду поменять местами, то нами установлена следующая теорема:

Функция распределения $F(x)$ суммы двух независимых случайных величин есть

$$(15.12.2) \quad F(x) = \int_{-\infty}^{\infty} F_1(x-z) dF_2(z) = \int_{-\infty}^{\infty} F_2(x-z) dF_1(z),$$

где F_1 и F_2 суть функции распределения слагаемых^{**}.

Если три функции распределения F , F_1 и F_2 удовлетворяют соотношению (15.12.2), то будем говорить, что F есть *композиция функций* F_1 и F_2 , и употреблять следующее сокращенное обозначение:

$$(15.12.2a) \quad F(x) = F_1(x) * F_2(x) = F_2(x) * F_1(x).$$

Функции же $F_1(x)$ и $F_2(x)$ будем называть *компонентами* этой операции.

^{**}) Читателю следует попробовать дать прямое доказательство этой теоремы, не пользуясь характеристическими функциями. Для этого надо показать, что в двумерном распределении независимых случайных величин ξ и η количество массы $F(x)$, заключенной в полуплоскости $\xi + \eta \leq x$, задается равенством (15.12.2). См. Крамер [11].

Согласно (15.12.1), такое символическое умножение функций распределения соответствует обычному умножению характеристических функций.

Если даны три независимые случайные величины ξ_1, ξ_2 и ξ_3 , то, повторяя доказательство (15.12.2) с некоторыми очевидными изменениями, получим, что сумма $\xi_1 + \xi_2 + \xi_3$ имеет функцию распределения $(F_1 * F_2) * F_3 = F_1 * (F_2 * F_3)$. Ясно, что это можно обобщить на любое конечное число компонент и что операция композиции коммутативна и ассоциативна. Для суммы n независимых случайных величин $\xi_1 + \dots + \xi_n$ функция распределения есть

$$(15.12.3) \quad F = F_1 * F_2 * \dots * F_n.$$

Рассмотрим теперь следующие два частных случая композиции функции распределения из двух компонент по правилу (15.12.2):

a) Обе компоненты — дискретного типа (см. параграф 15.2).

b) Обе компоненты — непрерывного типа, и по крайней мере одна из функций плотности, скажем $f_1 = F'_1$, ограничена для всех x .

В случае a) пусть x_1, x_2, \dots и y_1, y_2, \dots обозначают соответственно точки разрыва функций F_1 и F_2 . Тогда очевидно, что вся масса распределения, задаваемого композицией F , содержится в точках $x_r + y_s$, где r и s принимают значения 1, 2, Если множество всех таких точек не имеет конечных предельных точек, то композиция также принадлежит к дискретному типу. Это условие соблюдается, например, в случае, когда все x_r и y_s неотрицательны или когда хотя бы одна из последовательностей $\{x_r\}$ и $\{y_s\}$ конечна.

В случае b) первый интеграл в соотношении (15.12.2) удовлетворяет условиям теоремы (7.3.2). Далее, согласно (7.3.1) и (7.5.5), производная $F'(x) = f(x)$ непрерывна при всех x и может быть выражена римановым интегралом:

$$(15.12.4) \quad f(x) = \int_{-\infty}^{\infty} f_1(x-z) f_2(z) dz = \int_{-\infty}^{\infty} f_2(x-z) f_1(z) dz.$$

Таким образом, композиция принадлежит к непрерывному типу, и плотность вероятности $f(x)$ всюду непрерывна.

Возвращаясь к общему случаю, обозначим соответственно через m_1, m_2 и m средние значения и через σ_1, σ_2 и σ стандартные отклонения величин ξ , η и $\xi + \eta$. Так как ξ и η независимы, то, согласно

(15.3.7) и (15.6.1), имеем

$$(15.12.5) \quad m = m_1 + m_2, \quad \sigma^2 = \sigma_1^2 + \sigma_2^2.$$

Для моментов высшего порядка относительно среднего общее выражение выводится из соотношения

$$\mu_v = E[(\xi + \eta - m)^v] = E[(\xi - m_1 + \eta - m_2)^v].$$

Так как любой первый момент относительно среднего равен нулю, то имеем, в частности, употребляя очевидные обозначения,

$$(15.12.6) \quad \begin{aligned} \mu_3 &= \mu_3^{(1)} + \mu_3^{(2)}, \\ \mu_4 &= \mu_4^{(1)} + 6\mu_2^{(1)}\mu_2^{(2)} + \mu_4^{(2)}, \\ &\dots \end{aligned}$$

Выражения для моментов композиции непосредственно распространяются на случай любого конечного числа случайных величин. Так для суммы n независимых случайных величин имеем следующие простые выражения для моментов трех первых порядков:

$$(15.12.7) \quad \begin{aligned} m &= m_1 + m_2 + \dots + m_n, \\ \sigma^2 &= \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2, \\ \mu_3 &= \mu_3^{(1)} + \mu_3^{(2)} + \dots + \mu_3^{(n)}. \end{aligned}$$

Для моментов высших порядков ($v > 3$) формулы становятся более сложными.

Наконец, рассмотрим semi-инварианты композиции. Теорема умножения характеристических функций дает

$$\log \varphi(t) = \log \varphi_1(t) + \log \varphi_2(t).$$

Следовательно, из (15.10.2) получаем $x_v = x_v^{(1)} + x_v^{(2)}$. Это простое правило композиции есть главное основание для введения semi-инвариантов. Оно непосредственно распространяется на случай n независимых случайных величин:

$$(15.12.8) \quad x_v = x_v^{(1)} + x_v^{(2)} + \dots + x_v^{(n)}.$$

РАЗЛИЧНЫЕ ДИСКРЕТНЫЕ РАСПРЕДЕЛЕНИЯ

16.1. Функция $\epsilon(x)$. Простейшее дискретное распределение то, при котором вся масса сосредоточена в единственной точке, скажем, в точке $x = 0$. Таково распределение случайной величины ξ , которая „почти всегда“ равна нулю, т. е. такова, что $P(\xi = 0) = 1$. Соответствующая функция распределения $\epsilon(x)$ задана посредством соотношения

$$(16.1.1) \quad \epsilon(x) = \begin{cases} 0 & \text{при } x < 0, \\ 1 & \text{при } x \geq 0. \end{cases}$$

Как мы уже заметили в параграфе 10.1, соответствующая характеристическая функция тождественно равна единице.

Вообще, случайная величина, почти всегда равная x_0 , имеет функцию распределения $\epsilon(x - x_0)$ и характеристическую функцию e^{itx_0} . Среднее значение этой случайной величины есть x_0 , а ее стандартное отклонение равно нулю. Обратно, если известно, что стандартное отклонение некоторой случайной величины равно нулю, то (см. параграф 15.6) вся распределенная масса содержится в единственной точке, и, следовательно, функция распределения имеет вид $\epsilon(x - x_0)$.

Общая функция распределения дискретного типа, заданная соотношением (15.2.1), может быть записана в виде:

$$(16.1.2) \quad F(x) = \sum_{v} p_v \epsilon(x - x_v).$$

Рассмотрим частный случай дискретной случайной величины ξ с распределением

$$(16.1.3) \quad \xi = \begin{cases} 1 & \text{с вероятностью } p, \\ 0 & \text{с вероятностью } q = 1 - p. \end{cases}$$

В следующем параграфе случайные величины, обладающие распределением такого рода, найдут себе важное применение. Из (16.1.2) мы получим функцию распределения такой величины ξ :

$$F(x) = p \epsilon(x - 1) + q \epsilon(x);$$

характеристическая функция, следовательно, имеет вид

$$(16.1.4) \quad \varphi(t) = pe^{it} + q = 1 + p(e^{it} - 1).$$

Среднее значение и дисперсия величины ξ есть

$$(16.1.5) \quad E(\xi) = p \cdot 1 + q \cdot 0 = p,$$

$$D^2(\xi) = E((\xi - p)^2) = p(1 - p)^2 + q(0 - p)^2 = pq.$$

16.2. Биномиальное распределение. Пусть \mathfrak{E} есть данный случайный эксперимент. Обозначим через E событие, вероятность наступления которого при каждом осуществлении эксперимента \mathfrak{E} равна p . Рассмотрим ряд из n независимых повторений эксперимента \mathfrak{E} (см. параграф 14.4) и определим случайную величину ξ_r , связанную с r -м повторением эксперимента, положив

$$\xi_r = \begin{cases} 1, & \text{если событие } E \text{ происходит при } r\text{-м повторении эксперимента } \mathfrak{E} \text{ (соответствующая вероятность} = p), \\ 0, & \text{если событие } E \text{ не происходит при } r\text{-м повторении эксперимента } \mathfrak{E} \text{ (соответствующая вероятность} = q = 1 - p). \end{cases}$$

При этом случайные величины ξ_1, \dots, ξ_n независимы и каждая величина ξ_i имеет распределение вероятностей (16.1.3), рассмотренное в предыдущем параграфе.

Очевидно что ξ , обозначает *число наступлений* события E в r -м эксперименте, следовательно, сумма

$$v = \xi_1 + \xi_2 + \dots + \xi_n$$

есть *число наступлений события E в ряде из n повторений эксперимента \mathfrak{E}* .

Так как v есть сумма n независимых случайных величин, то v есть случайная величина*), распределение которой может быть найдено методами параграфа 15.12. Таким образом, согласно (15.12.7) и (16.1.5), получим следующие выражения для среднего, дисперсии и стандартного отклонения величины v :

$$(16.2.1) \quad E(v) = np, \quad D^2(v) = npq, \quad D(v) = \sqrt{npq}.$$

Отношение $\frac{v}{n}$ выражает *частоту* события E в нашем ряде из n повторений эксперимента \mathfrak{E} . Для среднего значения и стандартного отклонения величины $\frac{v}{n}$ имеем

$$(16.2.2) \quad E\left(\frac{v}{n}\right) = p, \quad D\left(\frac{v}{n}\right) = \sqrt{\frac{pq}{n}}.$$

* В общей теории, развитой в предыдущей главе, мы систематически употребляли для обозначения случайных величин буквы ξ и η . Однако в дальнейшем было бы неудобно строго следовать этому правилу. Поэтому мы часто будем употреблять для обозначения случайных величин и другие греческие и латинские буквы. Необходимо поэтому всегда с большим вниманием следить за значением различных букв, входящих в формулы.

Характеристическая функция величины ν , согласно (15.12.1), есть произведение характеристических функций всех ξ_r . Таким образом, из (16.1.4) получаем

$$(16.2.3) \quad E(e^{it\nu}) = (pe^{it} + q)^n = (1 + p(e^{it} - 1))^n.$$

Раскрывая это выражение, находим

$$E(e^{it\nu}) = \sum_{r=0}^n \binom{n}{r} p^r q^{n-r} e^{itr}.$$

Согласно (10.1.5), это есть характеристическая функция случайной величины, принимающей значения $r = 0, 1, \dots, n$ с вероятностями $P_r = \binom{n}{r} p^r q^{n-r}$. Вследствие взаимнооднозначного соответствия между распределениями и характеристическими функциями можно заключить (см. параграф 15.9), что распределение вероятностей для ν задается соотношением

$$(16.2.4) \quad P(\nu = r) = P_r = \binom{n}{r} p^r q^{n-r} \quad (r = 0, 1, \dots, n).$$

Это — *биномиальное распределение*, простейшие свойства которого предполагаются уже известными. Это распределение дискретного типа, зависящее от двух параметров n и p , где n есть положительное целое число, а $0 < p < 1$. (Случай $p = 0$ и $p = 1$ тривиальны и исключаются из дальнейших рассмотрений.) Соответствующая функция распределения

$$(16.2.5) \quad B_n(x; p) = P(\nu \leq x) = \sum_{r \leq x} \binom{n}{r} p^r q^{n-r}$$

есть ступенчатая функция со скачками P_r в $n+1$ точках сосредоточения массы $r = 0, 1, \dots, n$.

Для того чтобы найти моменты μ_r относительно среднего значения биномиального распределения, рассмотрим характеристическую функцию отклонения $u - np$:

$$\begin{aligned} E(e^{it(u-np)}) &= e^{-npit} (pe^{it} + q)^n = (pe^{qit} + qe^{-pit})^n = \\ &= \left[\sum_{r=0}^{\infty} (pq^r + q(-p)^r) \frac{(it)^r}{r!} \right]^n. \end{aligned}$$

Отсюда мы видим, что все моменты μ_r , конечны и могут быть найдены сравнением коэффициентов в соотношении

$$\sum_0^{\infty} \mu_r \frac{t^r}{r!} = \left[\sum_0^{\infty} (pq^r + q(-p)^r) \frac{t^r}{r!} \right]^n.$$

В частности, находим

$$(16.2.6) \quad \begin{aligned} \mu_2 &= \sigma^2 = pq, \\ \mu_3 &= npq(q-p), \\ \mu_4 &= 3n^2 p^2 q^2 + npq(1-6pq), \\ &\dots \end{aligned}$$

Для коэффициентов асимметрии и эксцесса имеем, следовательно, выражения

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{q-p}{\sqrt{npq}} = \frac{1-2p}{\sqrt{npq}}, \quad \gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{1-6pq}{npq}.$$

Асимметрия положительна при $p < \frac{1}{2}$, отрицательна при $p > \frac{1}{2}$ и равна нулю при $p = \frac{1}{2}$. Оба коэффициента γ_1 и γ_2 стремятся к нулю при $n \rightarrow \infty$.

Пусть ν_1 и ν_2 — две независимые случайные величины, имеющие обе биномиальные распределения с одним и тем же значением параметра p и со значениями n_1 и n_2 параметра n . Можно, например, взять ν_1 и ν_2 равными числу наступлений события E в двух независимых рядах из n_1 и n_2 повторений эксперимента \mathcal{E} .

Сумма $\nu_1 + \nu_2$ будет равна тогда числу наступлений события E в ряде из $n_1 + n_2$ повторений эксперимента \mathcal{E} . Характеристическая функция величины $\nu_1 + \nu_2$ есть (см. параграф 15.12)

$$E(e^{it(\nu_1 + \nu_2)}) = E(e^{it\nu_1}) E(e^{it\nu_2}) = \\ = (pe^{it} + q)^{n_1} (pe^{it} + q)^{n_2} = (pe^{it} + q)^{n_1 + n_2}.$$

Мы пришли к характеристической функции биномиального распределения с параметрами p и $n_1 + n_2$. Таким образом, сложение двух независимых случайных величин с функциями распределения $B_{n_1}(x; p)$ и $B_{n_2}(x; p)$ дает (как в этом можно убедиться, конечно, и непосредственно) случайную величину с функцией распределения $B_{n_1+n_2}(x; p)$. В сокращенных обозначениях (см. (15.12.2а)) это может быть записано в виде

$$B_{n_1}(x; p) * B_{n_2}(x; p) = B_{n_1+n_2}(x; p).$$

Таким образом, биномиальное распределение *воспроизводит само себя* при сложении независимых случайных величин. Этот факт мы будем называть *теоремой сложения* для биномиальных распределений. Позднее мы увидим, что подобные (хотя и менее очевидные) теоремы

сложения имеют место также для некоторых других важных распределений.

16.3. Теорема Бернулли. Для частоты $\frac{v}{n}$, рассмотренной в предыдущем параграфе, имеем, согласно (16.2.2).

$$E\left(\frac{v}{n}\right) = p, \quad D\left(\frac{v}{n}\right) = \sqrt{\frac{pq}{n}}.$$

Применим теперь неравенство Бъенэме — Чебышева (15.7.2), полагая $k = \epsilon \sqrt{\frac{n}{pq}}$, где ϵ — некоторое данное положительное число. Обозначая через P вероятностную функцию случайной величины v , получим следующий результат:

$$(16.3.1) \quad P\left(\left|\frac{v}{n} - p\right| \geq \epsilon\right) \leq \frac{pq}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}.$$

Если δ обозначает другое данное положительное число, то при $n > \frac{1}{4\delta\epsilon^2}$ вероятность в левой части неравенства (16.3.1) меньше δ . Так как δ произвольно мало, то нами доказана следующая теорема:

Вероятность того, что частота $\frac{v}{n}$ отличается по модулю от своего среднего значения p не меньше чем на ϵ , стремится к нулю при $n \rightarrow \infty$, как бы мало ни было ϵ .

Эта классическая теорема была доказана первоначально Яковом Бернулли в его посмертной работе „Ars Conjectandi“ (1713) совершенно иным способом. Бернулли рассматривал две дополнительные вероятности

$$\tilde{\omega} = P\left(\left|\frac{v}{n} - p\right| \geq \epsilon\right) = \sum_{|r-np| \geq n\epsilon} \binom{n}{r} p^r q^{n-r},$$

$$1 - \tilde{\omega} = P\left(\left|\frac{v}{n} - p\right| < \epsilon\right) = \sum_{|r-np| < n\epsilon} \binom{n}{r} p^r q^{n-r}.$$

и доказывал прямой оценкой слагаемых биномального разложения, что для любого данного $\epsilon > 0$ отношение $\frac{1 - \tilde{\omega}}{\tilde{\omega}}$ превзойдет заданную величину, если выбрать достаточно большое n .

Согласно предыдущему параграфу, величина v связана с составным экспериментом, заключающимся в n повторениях исходного эксперимента (§). Таким образом, согласно параграфу 13.5, любое вероятностное утверждение относительно v является утверждением относительно приближенного значения

частоты некоторого события в ряде повторений составного эксперимента. Таким образом, *частотная интерпретация* (см. параграф 13.5) любого такого вероятностного утверждения связана с рядом повторений *составного эксперимента*.

Рассмотрим, например, частотную интерпретацию (см. параграф 13.5) определенной выше вероятности $\bar{\omega}$. Мы осуществляем n повторений эксперимента E и регистрируем число v наступлений события E . Это — первое осуществление нашего составного эксперимента. Если наблюдение число v удовлетворяет неравенству $\left| \frac{v}{n} - p \right| \geq \varepsilon$, то мы условимся говорить, что при первом осуществлении составного эксперимента имеет место *событие E'* . Вероятность события E' есть $\bar{\omega}$.

Затем мы осуществляем всю серию из n повторений основного эксперимента E большое число n' раз, получая в конечном счете ряд из n' повторений составного эксперимента. Общее число осуществлений эксперимента E будет, конечно, равно $n'p$. Пусть v' есть число наступлений события E' в полном ряде из n' повторений составного эксперимента. Тогда частотная интерпретация вероятности $\bar{\omega}$ заключается в следующем утверждении: для больших значений n' практически несомненно, что частота $\frac{v'}{n'}$ приблизительно равна $\bar{\omega}$.

Теорема Бернулли, выраженная соотношением (16.3.1), показывает, что при $n > \frac{1}{4\delta^2}$ имеем $\bar{\omega} < \delta$, где δ есть заданное произвольно малое число. В длинной серии повторений составного эксперимента (т. е. при больших n') мы должны ожидать, что событие $\left| \frac{v}{n} - p \right| \geq \varepsilon$ будет наступать с частотой, меньшей δ . Выбирая δ очень малым, мы можем считать практически несомненным (см. параграф 13.5), что при *единственном* осуществлении составного эксперимента, т. е. в *одном ряде из n повторений эксперимента E* , событие $\left| \frac{v}{n} - p \right| \geq \varepsilon$ не наступит.

Значение δ , которое нужно выбрать для того, чтобы достичь достаточной степени „практической несомненности“, зависит от допускаемого риска ошибки наших предсказаний. Предположим, что мы согласились рассматривать некоторое значение δ_0 как достаточно малое для нашей цели. Возвращаясь к исходному событию E , имеющему вероятность p , можно дать следующую, более точную, чем в параграфе 13.5, формулировку частотной интерпретации:

Пусть задано $\varepsilon > 0$. Если выбрать $n > \frac{1}{4\delta_0\varepsilon^2}$, то практически несомненно, что в единственном ряде из n повторений эксперимента E мы будем иметь $\left| \frac{v}{n} - p \right| < \varepsilon$.

Это утверждение может быть названо *частотной интерпретацией теоремы Бернулли*. Подобно всем частотным интерпретациям, это не математическая теорема, а утверждение относительно некоторых наблюденных фактов, которое должно быть справедливым, если наша математическая теория имеет какое-либо практическое значение.

16.4. Теорема Муавра. Случайная величина

$$(16.4.1) \quad v = \xi_1 + \xi_2 + \dots + \xi_n,$$

рассмотренная в двух предыдущих параграфах, имеет, согласно (16.2.1), среднее значение np и стандартное отклонение \sqrt{npq} . Нормированная величина (см. параграф 15.6)

$$(16.4.2) \quad \lambda = \frac{v - np}{\sqrt{npq}}$$

имеет, таким образом, среднее 0 и стандартное отклонение 1. Преобразование, с помощью которого мы переходим от v к λ , состоит, конечно, только в смещении начала координат и изменении масштаба на оси абсцисс. Ординаты графика, изображающего распределение вероятностей, будут одинаковыми для обеих случайных величин.

Действительно, используя обозначения предыдущего параграфа, получим для $r = 0, 1, \dots, n$

$$P\left(\lambda = \frac{r - np}{\sqrt{npq}}\right) = P(v = r) = \binom{n}{r} p^r q^{n-r}.$$

Функция распределения и характеристическая функция величины v задаются соотношениями (16.2.5) и (16.2.3). Обозначая через $F_n(x)$ и $\varphi_n(t)$ соответствующие функции нормированной величины λ , получим (см. параграф 15.9)

$$(16.4.3) \quad F_n(x) = B_n(np + x\sqrt{npq}; p),$$

$$\varphi_n(t) = \left(pe^{\frac{qt}{\sqrt{npq}}} + qe^{-\frac{pt}{\sqrt{npq}}}\right)^n.$$

Рассмотрим теперь *поведение распределения вероятностей величины λ при возрастании n и фиксированном p .* Преобразуем сперва выражение характеристической функции $\varphi_n(t)$.

Для любого целого $k > 0$ и любого действительного z имеем разложение Маклорена

$$(16.4.4) \quad e^{iz} = \sum_0^{k-1} \frac{(iz)^r}{r!} + o \frac{z^k}{k!},$$

где ϑ означает действительное или комплексное число, по модулю не превышающее единицу. Используя это разложение для $k=3$, получим

$$pe^{\frac{qit}{\sqrt{n}pq}} = p + \frac{pqit}{\sqrt{n}pq} - \frac{pq^2t^2}{2npq} + \vartheta \frac{pq^3t^3}{3!(npq)^{3/2}},$$

$$qe^{-\frac{p|t|}{\sqrt{n}pq}} = q - \frac{pqit}{\sqrt{n}pq} - \frac{p^2qt^2}{2npq} + \vartheta \frac{p^3qt^3}{3!(npq)^{3/2}},$$

и, следовательно,

$$\varphi_n(t) = \left(1 - \frac{t^2}{2n} + \vartheta \frac{t^3}{(npq)^{3/2}}\right)^n.$$

Полагая

$$y = -\frac{t^2}{2} + \vartheta \frac{t^3}{(pq)^{3/2}\sqrt{n}},$$

получаем

$$\log \varphi_n(t) = y \cdot \frac{n}{y} \log \left(1 + \frac{y}{n}\right)$$

При $n \rightarrow \infty$ и фиксированном t величина y стремится, очевидно, к $-\frac{t^2}{2}$. Следовательно, $\frac{y}{n}$ стремится к нулю, а $\frac{n}{y} \log \left(1 + \frac{y}{n}\right)$ стремится к 1. Отсюда следует, что $\log \varphi_n(t)$ стремится к $-\frac{t^2}{2}$ и, наконец, что

$$\varphi_n(t) \rightarrow e^{-\frac{t^2}{2}}$$

при любом t

Теперь мы можем применить теорему непрерывности для характеристических функций (см. параграф 10.4). Мы только что доказали, что последовательность характеристических функций $\{\varphi_n(t)\}$, определенных соотношением (16.4.3), сходится при любом t к непрерывной

функции $e^{-\frac{t^2}{2}}$. Согласно теореме непрерывности, заключаем, что

1) предел $e^{-\frac{t^2}{2}}$ есть характеристическая функция некоторой функции распределения и 2) последовательность функций распределения $\{F_n(x)\}$, определенная соотношением (16.4.3), сходится к функции распределения, соответствующей характеристической функции $e^{-\frac{t^2}{2}}$.

Согласно (10.5.3) и (10.5.4), имеем

$$e^{-\frac{t^2}{2}} = \int_{-\infty}^{\infty} e^{itx} d\Phi(x),$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt,$$

так что $e^{-\frac{t^2}{2}}$ есть характеристическая функция функции распределения $\Phi(x)$, заданной последним соотношением. Эта весьма важная *нормальная функция распределения* будет предметом исследования следующей главы. Сейчас заметим только, что $\Phi(x)$ непрерывна при любом x . Нами доказана, таким образом, следующая *пределная теорема для биномиального распределения*, полученная впервые Муавром в 1733 г.

При любых фиксированных x и p имеем

$$(16.4.5) \quad \lim_{n \rightarrow \infty} B_n(np + x\sqrt{npq}; p) = \Phi(x).$$

Таким образом, биномиальное распределение величины $y = \xi_1 + \dots + \xi_n$, нормированной согласно (16.4.2), стремится к нормальному распределению при $n \rightarrow \infty$. Мы увидим позднее (см. параграф 17.4), что это только частный случай более общей и важной теоремы относительно распределения суммы большого числа независимых случайных величин. Использованный выше метод доказательства был выбран с целью подготовить читателя к доказательству этой общей теоремы. Однако в рассматриваемом сейчас частном случае биномиального распределения можно достигнуть того же результата более прямым методом без использования характеристических функций, как это обычно и делается в учебниках. Ограничимся здесь лишь несколькими краткими указаниями; детали вывода читатель сможет найти в обычных руководствах по теории вероятностей.

Соотношение (16.4.5) эквивалентно соотношению

$$(16.4.6) \quad \sum_{np + \lambda_1 \sqrt{npq} < v \leq np + \lambda_2 \sqrt{npq}} \binom{n}{v} p^v q^{n-v} \rightarrow \Phi(\lambda_2) - \Phi(\lambda_1) = \\ = \frac{1}{\sqrt{2\pi}} \int_{\lambda_1}^{\lambda_2} e^{-\frac{t^2}{2}} dt$$

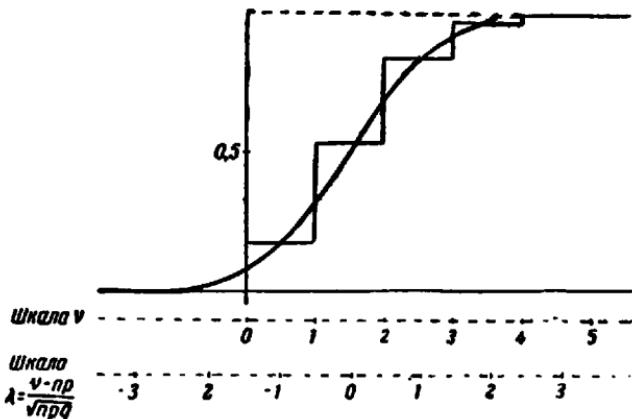
для любого фиксированного интервала (λ_1, λ_2) . Соотношение (16.4.6) может быть доказано простой оценкой слагаемых в биномиальном разложении. Для этой цели выражаем факториалы в биномиальных

коэффициентах (16.4.6) с помощью формулы Стирлинга (12.5.3). После некоторых вычислений получим выражение

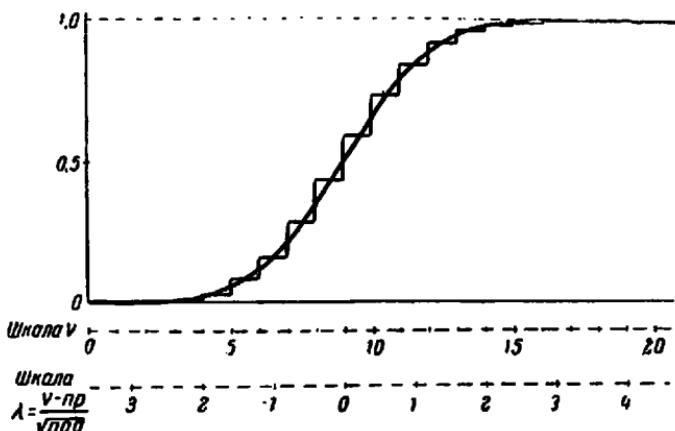
$$(16.4.7) \quad \binom{n}{v} p^v q^{n-v} = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{1}{2}\left(\frac{v-np}{\sqrt{npq}}\right)^2} + \theta \frac{C}{n},$$

где C есть величина, зависящая от p , но не зависящая от v и n , а θ имеет тот же смысл, что и раньше. Таким образом, левая часть соотношения (16.4.6) равна

$$\frac{1}{\sqrt{2\pi npq}} \sum e^{-\frac{1}{2}\left(\frac{v-np}{\sqrt{npq}}\right)^2} + \theta \frac{(l_2 - l_1) C}{\sqrt{n}},$$

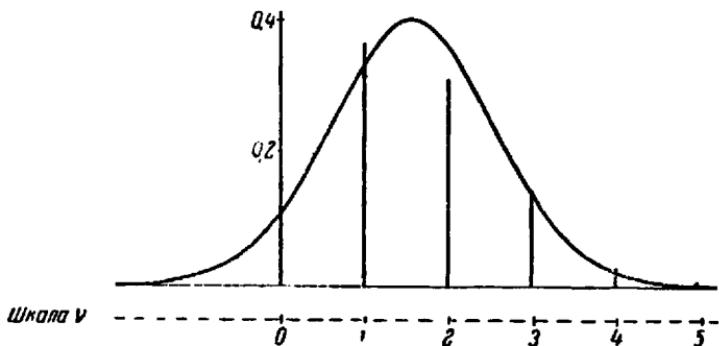


Фиг. 8. Функция распределения v (или λ) и нормальная функция распределения. $p = 0,3$; $n = 5$.



Фиг. 9. Функция распределения v (или λ) и нормальная функция распределения. $p = 0,3$; $n = 30$.

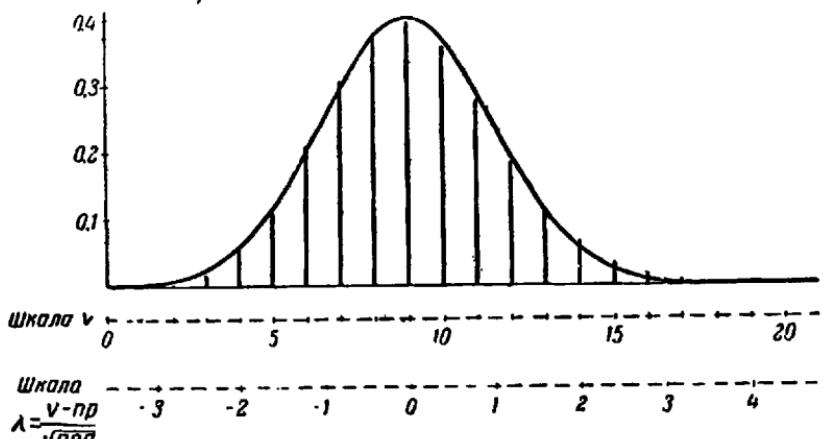
где сумма распространяется на те же значения u , что и в (16.4.6). При $n \rightarrow \infty$ второе слагаемое в этом выражении стремится к нулю; первое же слагаемое представляет собою сумму Дарбу для интеграла в правой части соотношения (16.4.6) и, следовательно, стремится к этому интегралу при $n \rightarrow \infty$. Таким образом, соотношение (16.4.6) доказано.



$$\text{Шкала } u \quad \dots -3 -2 -1 0 1 2 3 \dots$$

$$\lambda = \frac{v - np}{\sqrt{npq}}$$

Фиг. 10. $\sqrt{npq} \binom{n}{v} p^v q^{n-v}$ и нормальная функция плотности. $p = 0,3$; $n = 5$



$$\text{Шкала } u \quad \dots -3 -2 -1 0 1 2 3 4 \dots$$

$$\lambda = \frac{v - np}{\sqrt{npq}}$$

Фиг. 11. $\sqrt{npq} \binom{n}{v} p^v q^{n-v}$ и нормальная функция плотности. $p = 0,3$; $n = 30$.

Для графической иллюстрации предельной теоремы (16.4.5) можно прибегнуть к простому сравнению графиков функций B_n и Φ на фиг. 8 и 9. Можно, однако, также использовать соотношение (16.4.7). Предполагая, что u стремится к бесконечности при $n \rightarrow \infty$ таким

образом, что $\frac{v-np}{\sqrt{npq}}$ стремится к конечному пределу x , получим

$$\sqrt{npq} \left(\frac{v-np}{\sqrt{npq}} \right) p^v q^{n-v} \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Если на шкале v принять за начало координат среднее np , за единицу масштаба стандартное отклонение \sqrt{npq} , и если в то же время каждую вероятность P_v умножить на \sqrt{npq} , то концы соответствующих ординат при $n \rightarrow \infty$ будут приближаться к кривой плотности $y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ нормального распределения, как это показано на фиг. 10—11.

16.5. Распределение Пуассона. В предыдущем параграфе мы видели, что дискретное биномиальное распределение может быть преобразовано с помощью предельного перехода в новое распределение непрерывного типа, а именно, в нормальное распределение.

Соответствующим изменением предельного перехода можно также получить предельное распределение дискретного типа. Предположим, что в биномиальном распределении вероятность p зависит от n так, что p стремится к нулю при $n \rightarrow \infty$. Точнее, предположим, что

$$(16.5.1) \quad p = \frac{\lambda}{n},$$

где λ — некоторая положительная постоянная. Для вероятности P_r , заданной соотношением (16.2.4), получим при $n \rightarrow \infty$

$$\begin{aligned} P_r &= \frac{n(n-1)\dots(n-r+1)}{r!} \left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^{n-r} = \\ &= \frac{\lambda^r}{r!} \left(1 - \frac{\lambda}{n}\right)^n \frac{\left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{r-1}{n}\right)}{\left(1 - \frac{\lambda}{n}\right)^r} \rightarrow \frac{\lambda^r}{r!} e^{-\lambda} \end{aligned}$$

для любого фиксированного $r = 0, 1, 2, \dots$. Сумма всех предельных значений равна единице, так как

$$\sum_{r=0}^{\infty} \frac{\lambda^r}{r!} e^{-\lambda} = e^{\lambda} \cdot e^{-\lambda} = 1.$$

Если случайная величина ξ имеет распределение вероятностей

$$(16.5.2) \quad P(\xi=r) = \frac{\lambda^r}{r!} e^{-\lambda} \text{ при } r=0, 1, 2, \dots,$$

то говорят, что она обладает *распределением Пуассона*. Это — дискретное распределение с одним параметром λ , который всегда положителен. Все точки $r = 0, 1, 2, \dots$ суть точки сосредоточения массы. Два случая этого распределения изображены на фиг. 12—13.

Характеристическая функция распределения Пуассона есть

$$(16.5.3) \quad E(e^{it\xi}) = \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} e^{-\lambda} e^{itr} = e^{\lambda(e^{it}-1)}.$$

Согласно (15.10.2), это показывает, что семи-инварианты этого распределения все конечны и равны λ . Для вычисления первых двух семи-инвариантов находим среднее значение и стандартное отклонение

$$E(\xi) = \lambda, \quad D(\xi) = \sqrt{\lambda}.$$

Полагая $p = \frac{\lambda}{n}$ во втором выражении (16.2.3) характеристической функции биномиального распределения, можно без труда убедиться в том, что эта функция стремится при $n \rightarrow \infty$ к характеристической функции (16.5.3) распределения Пуассона.

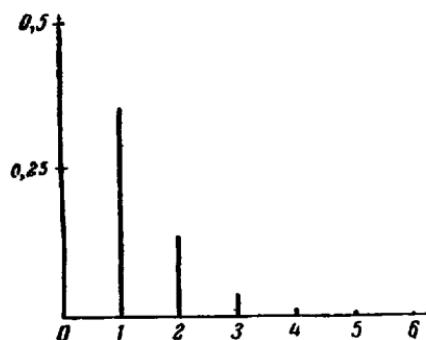
Тогда из теоремы непрерывности (см. параграф 10.4) следует, что биномиальное распределение стремится к распределению Пуассона, что подтверждает результат, уже полученный при непосредственном изучении вероятности P_r .

Легко также показать, что, не нарушая результата, можно заменить условие (16.5.1) более общим условием: $np \rightarrow \lambda$.

Наконец, если независимые случайные величины ξ_1 и ξ_2 обладают распределениями Пуассона с параметрами λ_1 и λ_2 , то сумма $\xi_1 + \xi_2$ имеет характеристическую функцию

$$e^{\lambda_1(e^{it}-1)} e^{\lambda_2(e^{it}-1)} = e^{(\lambda_1 + \lambda_2)(e^{it}-1)},$$

т. е. характеристическую функцию распределения Пуассона с параметром $\lambda_1 + \lambda_2$. Таким образом, сумма $\xi_1 + \xi_2$ обладает распределением Пуассона с параметром $\lambda_1 + \lambda_2$. Мы видим, что распределение Пуассона, как и биномиальное, воспроизводит само себя при сложении

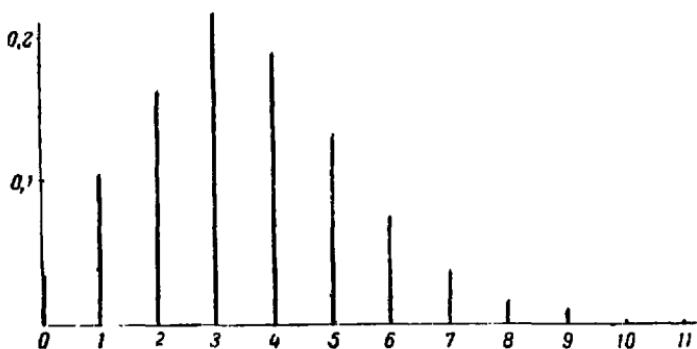


Фиг. 12. Распределение Пуассона, $\lambda = 0,8$

независимых величин. Обозначая через $F(x; \lambda)$ его функцию распределения, получим теорему сложения для распределения Пуассона

$$(16.5.4) \quad F(x; \lambda_1) * F(x; \lambda_2) = F(x; \lambda_1 + \lambda_2).$$

В приложениях к статистике распределение Пуассона встречается часто, когда мы имеем дело с числом наступлений некоторого события при большом числе наблюдений, и при малой вероятности наступления этого события в каждом отдельном наблюдении, например: число самоубийств за год, число взвешенных частиц в пробе малого объема из большого количества суспензии и т. д. (См., например, Борткевич [63а].)



Фиг. 13. Распределение Пуассона, $\lambda = 3,5$.

В одном важном классе приложений основной случайный эксперимент состоит в наблюдении числа наступлений некоторого события за промежуток времени t , причем выбор t произволен. Такое положение встречается, например, при рассмотрении числа телефонных вызовов за различные промежутки времени. Предположим в этом случае, что числа наступлений события за непересекающиеся промежутки времени независимы. Предположим далее, что вероятность того, что за промежуток времени Δt событие произойдет ровно один раз, при малом Δt равна

$$\lambda \Delta t + o(\Delta t),$$

где λ — постоянная, а вероятность того, что это событие за то же время Δt наступит более одного раза, есть $o(\Delta t)$. Разделив промежуток времени t на n равных частей, можно рассматривать эти n частей как n повторений случайного эксперимента, где вероятность наступления события в каждом отдельном случае есть

$$\frac{\lambda t}{n} + o\left(\frac{1}{n}\right).$$

Заставляя n стремиться к бесконечности, находим, что общее число наступлений события за время t обладает распределением Пуассона с параметром λt .

Величинами такого типа, кроме уже упомянутого числа телефонных вызовов являются: число расщепившихся радиоактивных атомов, число претензий к страховой компании и т. д.

16.6. Обобщенное биномиальное распределение Пуассона. Предположим, что n случайных экспериментов E_1, \dots, E_n таковы, что случайные величины, связанные с этими экспериментами, независимы. С каждым экспериментом мы связываем событие E_r , вероятность наступления которого при осуществлении эксперимента E_r равна $p_r = 1 - q_r$.

Допустим, что каждый из экспериментов E_1, \dots, E_n осуществляется ровно один раз и при этом в каждом случае регистрируется, наступает ли связанное с данным экспериментом событие или нет. В этом случае мы говорим о ряде *независимых испытаний*. Если при осуществлении эксперимента E_r , связанного с ним событие E_r , наступает, то говорим, что r -е испытание имеет *благоприятный исход*, в противном случае говорим, что оно имеет *неблагоприятный исход*. Пусть u есть общее число благоприятных исходов во всех n испытаниях. Каково распределение вероятностей величины u ?

В частном случае, когда все эксперименты E_r и все события E_r тождественны, величина u сводится к величине, рассмотренной в параграфе 16.2, и искомое распределение есть биномиальное распределение. Описанный общий случай был рассмотрен Пуассоном [32].

Так же, как и в параграфе 16.2, определяем величину ξ_r , связанную с r -м испытанием и принимающую значение 1 при благоприятном исходе (с вероятностью p_r) и значение 0 при неблагоприятном исходе (с вероятностью $q_r = 1 - p_r$). Величины ξ_1, \dots, ξ_n независимы и каждая из них имеет распределение вида (16.1.3). Как и в предыдущем случае, общее число благоприятных исходов равно $u = \xi_1 + \dots + \xi_n$.

Характеристическая функция случайной величины u есть произведение характеристических функций всех ξ_r :

$$E(e^{itu}) = \prod_{r=1}^n (p_r e^{itr} + q_r).$$

Возможные значения u суть $u = 0, 1, \dots, n$, и вероятность того, что u принимает значение r , равна коэффициенту при e^{itr} в разложении написанного выше произведения.

Для среднего значения и дисперсии величины v имеем следующие выражения:

$$(16.6.1) \quad E(v) = \sum_1^n E(\xi_r) = \sum_1^n p_r,$$

$$D^2(v) = \sum_1^n D^2(\xi_r) = \sum_1^n p_r q_r.$$

Обозначая через P вероятностную функцию величины v и через p среднее арифметическое $\frac{1}{n} \sum_1^n p_r$, получаем с помощью неравенства Бьенэме — Чебышева (15.7.2) результат, аналогичный (16.3.1):

$$(16.6.2) \quad P\left(\left|\frac{v}{n} - p\right| \geq \epsilon\right) \leq \frac{\sum p_r q_r}{n^2 \epsilon^2} \leq \frac{1}{4n\epsilon^2}.$$

Таким образом, имеем следующее обобщение теоремы Бернулли, найденное Пуассоном:

Вероятность того, что частота благоприятных исходов $\frac{v}{n}$ отличается по модулю от среднего арифметического вероятностей p_r не меньше чем на ϵ , стремится к нулю при $n \rightarrow \infty$, как бы мало ни было ϵ .

Частотная интерпретация этой обобщенной теоремы вполне аналогична частотной интерпретации теоремы Бернулли, данной в параграфе 16.3. Рассмотрим, в частности, случай, когда все вероятности p_r равны p . Мы тогда видим, что в длинном ряде независимых испытаний, при которых вероятность благоприятного исхода постоянно равна p , хотя испытания могут относиться к различным экспериментам, практически несомненно, что частота благоприятных исходов приблизительно равна p .

Существует также обобщение теоремы Муавра (16.4.5) на наш случай. Оно здесь доказываться не будет, но ниже мы выведем его как частный случай из более общей теоремы, доказанной в параграфе 17.4.

Для дисперсии величины v получено значение $D^2(v) = \sum p_r q_r$. В ряде из n испытаний с постоянной вероятностью $p = \frac{1}{n} \sum p_r$ соответствующая дисперсия равна $p p q$, где $q = 1 - p = \frac{1}{n} \sum q_r$. Для того чтобы сравнить эти две дисперсии, пишем:

$$\begin{aligned} \sum p_r q_r &= \sum (p + p_r - p)(q + q_r - q) := \\ &= \sum (p + p_r - p)(q + p - p_r) = p p q - \sum (p_r - p)^2. \end{aligned}$$

Таким образом, „пуассоновская дисперсия“ $\sum p_r q_r$ всегда меньше соответствующей „бернуллиевой дисперсии“ $p p q$. На первый взгляд этот результат

может показаться несколько неожиданным. Он становится более естественным, если рассмотреть предельный случай, когда все вероятности p_r равны 0 или 1 (причем фактически встречаются как те, так и другие). Тогда пуассоновская дисперсия равна 0, а бернуlliева дисперсия положительна.

ГЛАВА 17

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

17.1. Нормальные функции. *Нормальная функция распределения*, которая уже встречалась в параграфах 10.5 и 16.4, определяется равенством

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Соответствующая *нормальная плотность вероятности* есть

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

Графики этих функций даны на фиг. 14—15, а их численные значения — в таблице I в конце книги.

Среднее значение этого распределения равно 0, а его стандартное отклонение равно 1, как это видно из (10.5.1):

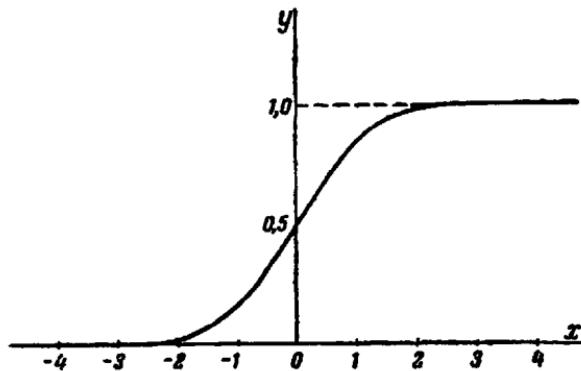
$$\int_{-\infty}^{\infty} x d\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} dx = 0,$$

(17.1.1)

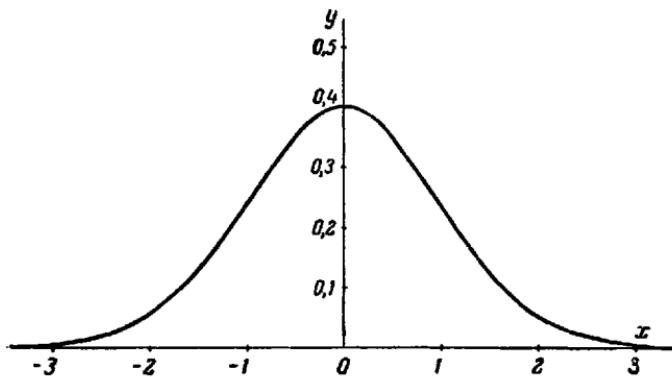
$$\int_{-\infty}^{\infty} x^2 d\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = 1.$$

Вообще, все моменты нечетного порядка равны нулю, а моменты четного порядка равны, согласно (10.5.1),

$$(17.1.2) \quad \int_{-\infty}^{\infty} x^{2y} d\Phi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2y} e^{-\frac{x^2}{2}} dx = 1 \cdot 3 \dots (2y - 1).$$



Фиг. 14. Нормальная функция распределения.



Фиг. 15. Нормальная плотность вероятности.

Наконец, характеристическая функция есть, согласно (10.5.4),

$$(17.1.3) \quad \int_{-\infty}^{\infty} e^{itx} d\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx - \frac{x^2}{2}} dx = e^{-\frac{t^2}{2}}.$$

17.2. Нормальное распределение. Случайная величина ξ называется *нормально распределенной с параметрами m и σ* или, короче, *нормальной (m, σ)*, если функция распределения величины ξ есть

$\Phi\left(\frac{x-m}{\sigma}\right)$, где $\sigma > 0$ и m — постоянные. Тогда плотность вероятности есть

$$\frac{1}{\sigma} \Phi'\left(\frac{x-m}{\sigma}\right) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

и из (17.1.1) получаем

$$E(\xi) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (m + \sigma x) e^{-\frac{x^2}{2}} dx = m,$$

$$D^2(\xi) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} (x - m)^2 e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \sigma^2,$$

т. е. m и σ обозначают, как обычно, среднее значение и стандартное отклонение нашей величины.

Кривая плотности

$$y = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

симметрична и унимодальна (см. параграф 15.5) и достигает своего максимума в точке $x = m$, так что m есть одновременно среднее, медиана и мода нашего распределения. При $x = m \pm \sigma$ кривая имеет две точки перегиба. Изменение значения m вызывает только смещение кривой без изменения ее формы, тогда как изменение величины σ вызывает изменение масштаба на обеих координатных осях. Площадь, заключенная между кривой и осью абсцисс, равна, конечно, единице. Кривые, соответствующие некоторым различным значениям σ , изображены на фиг. 16.

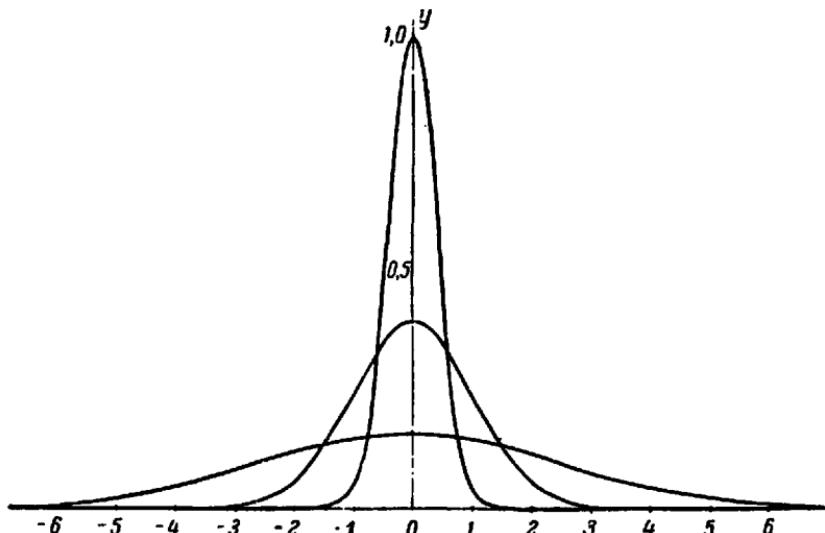
Чем меньше σ , тем большая часть массы сосредоточена в окрестности точки $x = m$. В предельном случае $\sigma = 0$ вся масса сосредоточена в точке $x = m$, и, следовательно, (см. параграф 16.1) функция распределения сводится к $\epsilon(x - m)$. Этот случай будет рассматриваться как вырожденный предельный случай и будет называться *несобственным* нормальным распределением. Соответствующая функция распределения $\Phi\left(\frac{x-m}{0}\right)$ будет всегда интерпретироваться как $\epsilon(x - m)$.

Часто бывает важно уметь найти вероятность P того, что нормально распределенная величина отличается по модулю от своего среднего больше, чем на некоторое кратное своего стандартного отклонения — $\lambda\sigma$. Эта вероятность равна сумме площадей двух „остатков“, отрезан-

ных от кривой плотности ординатами, проходящими через точки $x = m \pm \lambda\sigma$. Вследствие симметричности распределения

$$P = P(|\xi - m| > \lambda\sigma) = 2(1 - \Phi(\lambda)) = \frac{2}{\sqrt{2\pi}} \int_{\lambda}^{\infty} e^{-\frac{x^2}{2}} dx.$$

Обратно, можно рассматривать λ как функцию от P , определенную последним соотношением. Тогда λ выражает (в единицах стандартного



Фиг. 16. Нормальные кривые плотности. $m = 0$; $\sigma = 0,4; 1,0; 2,5$

отклонения σ) то отклонение от среднего значения m , которое превышается с данной вероятностью P . Если P выражается в процентах, скажем $P = \frac{p}{100}$, то соответствующее $\lambda = \lambda_p$ называется *p-процентным значением нормального отклонения* $\frac{\xi - m}{\sigma}$. Некоторые численные значения p как функции от λ_p и λ_p как функции от p приведены в таблице II в конце книги. Для значения λ_p при $p = 50$ квартили (см. параграф 15.6) нормального распределения равны $m \pm 0,6745 \sigma$. Далее, 5-процентное значение $\frac{\xi - m}{\sigma}$ равно приблизительно 2,0; 1-процентное значение близко к 2,6; 0,1-процентное значение близко к 3,3. Отклонения, превышающие учетверенное стандартное отклонение, имеют крайне малые вероятности.

Нормированная величина $\frac{\xi - m}{\sigma}$ имеет функцию распределения $\Phi(x)$ и, следовательно, согласно (17.1.3), ее характеристическая функция есть $e^{-\frac{t^2}{2}}$. Из (15.9.2) следует, что величина ξ имеет характеристическую функцию

$$(17.2.1) \quad E(e^{it\xi}) = e^{mit - \frac{1}{2}\sigma^2 t^2}.$$

Из этого выражения, согласно (15.10.2), находятся семи-инварианты

$$(17.2.2) \quad x_1 = m, \quad x_2 = \sigma^2, \quad x_3 = x_4 = \dots = 0.$$

Моменты относительно среднего значения величины ξ равны

$$(17.2.3) \quad \mu_{2v+1} = 0, \quad \mu_{2v} = 1 \cdot 3 \dots (2v-1) \sigma^{2v}.$$

В частности, коэффициенты асимметрии и эксцесса (см. параграф 15.8) равны

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = 0, \quad \gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = 0.$$

Заметим, наконец, что если величина ξ нормальна (m, σ), то из (15.1.1) следует, что линейная функция $a\xi + b$ нормальна ($am + b, |a|\sigma$).

17.3. Сложение независимых нормальных величин. Пусть ξ_1, \dots, ξ_n — независимые случайные величины, из которых каждая ξ_v распределена нормально с параметрами m_v и σ_v . Рассмотрим сумму

$$\xi = \xi_1 + \xi_2 + \dots + \xi_n.$$

Обозначая через m и σ среднее значение и стандартное отклонение величины ξ , имеем, согласно (15.12.7),

$$(17.3.1) \quad \begin{aligned} m &= m_1 + m_2 + \dots + m_n, \\ \sigma^2 &= \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2. \end{aligned}$$

По правилу умножения (15.12.1) характеристическая функция величины ξ есть произведение характеристических функций всех ξ_v . Из выражения (17.2.1) для характеристической функции нормального распределения получим

$$E(e^{it\xi}) = \prod_{v=1}^n e^{m_v it - \frac{1}{2}\sigma_v^2 t^2} = e^{mit - \frac{1}{2}\sigma^2 t^2}.$$

Это — характеристическая функция нормального распределения с параметрами m и σ . Таким образом, нами доказана следующая важная теорема сложения для нормального распределения:

Сумма любого числа независимых нормально распределенных величин сама распределена нормально:

$$(17.3.2) \quad \Phi\left(\frac{x-m_1}{\sigma_1}\right) * \Phi\left(\frac{x-m_2}{\sigma_2}\right) * \dots * \Phi\left(\frac{x-m_n}{\sigma_n}\right) = \Phi\left(\frac{x-m}{\sigma}\right),$$

где m и σ заданы соотношениями (17.3.1).

Упомянем без доказательства следующее обращение этой теоремы (Крамер [11]): *Если сумма n независимых величин $\xi = \xi_1 + \dots + \xi_n$ распределена нормально, то каждая из величин ξ_i сама распределена нормально.* Таким образом, верно не только, что нормальное распределение при композиции воспроизводит само себя, но и, кроме того, то, что нормальное распределение не может быть точно представлено как композиция не нормально распределенных компонент. С другой стороны, мы видели в предыдущем параграфе, что при очень общих условиях композиция большого числа не нормальных компонент представляет *приближенно* нормальное распределение.

Так как, согласно предыдущему параграфу, линейная функция нормальной величины сама нормальна, то из (17.3.2) следует, что линейная функция $a_1\xi_1 + a_2\xi_2 + \dots + a_n\xi_n + b$ независимых нормальных величин сама нормальна с параметрами $m = a_1m_1 + \dots + a_nm_n + b$ и $\sigma^2 = a_1^2\sigma_1^2 + \dots + a_n^2\sigma_n^2$. В частности, имеет место следующая важная теорема:

Если ξ_1, \dots, ξ_n независимы и все нормальны (m_i, σ_i), то среднее арифметическое $\bar{\xi} = \frac{1}{n} \sum_1^n \xi_i$ само нормально $\left(m, \frac{\sigma}{\sqrt{n}}\right)$.

17.4. Центральная предельная теорема. Рассмотрим сумму

$$(17.4.1) \quad \xi = \xi_1 + \dots + \xi_n$$

n независимых величин, где ξ_i имеет среднее m_i и стандартное отклонение σ_i . Среднее значение m и стандартное отклонение σ суммы ξ задаются тогда обычными выражениями (17.3.1).

В предыдущем параграфе мы видели, что если величины ξ_i распределены нормально, то и сумма ξ распределена нормально. С другой стороны, теорема Муавра (см. параграф 16.4) показывает, что в частном случае, когда величины ξ_i имеют простое распределение (16.1.3), распределение суммы *приблизительно* нормально для больших значений n . Действительно, теорема Муавра утверждает, что в этом частном случае функция распределения нормированной величины $\frac{\xi - m}{\sigma}$ стремится к нормальной функции $\Phi(x)$ при $n \rightarrow \infty$. Весьма замечательно, что

утверждение теоремы Муавра справедливо и при гораздо более общих условиях.

Удобно ввести следующую терминологию. Если распределение случайной величины X зависит от параметра n и если две величины m_0 и σ_0 (зависящие или не зависящие от n) могут быть выбраны так, что функция распределения величины $\frac{X - m_0}{\sigma_0}$ стремится к $\Phi(x)$ при $n \rightarrow \infty$, то говорим, что X асимптотически нормальна (m_0, σ_0). Отсюда не следует что среднее и стандартное отклонения величины X стремятся к m_0 и σ_0 , — эти моменты могут даже не существовать, — это просто эквивалентно тому, что для любого интервала (a, b) , не зависящего от n ,

$$\lim_{n \rightarrow \infty} P(m_0 + a\sigma_0 < X < m_0 + b\sigma_0) = \Phi(b) - \Phi(a).$$

Таким образом, например, величина γ , рассматривавшаяся в теореме Муавра, асимптотически нормальна (p, \sqrt{pq}).

Так называемая *центральная предельная теорема* в математической теории вероятностей может быть выражена теперь следующим образом: *Каковы бы ни были распределения величин ξ_r , при некоторых весьма общих условиях, сумма $\xi = \xi_1 + \dots + \xi_n$ асимптотически нормальна (m, σ), где m и σ заданы соотношениями (17.3.1).*

Эта фундаментальная теорема была впервые высказана Лапласом [22] в 1812 г. Строгое доказательство при довольно общих условиях было дано Ляпуновым [146], [147] в 1901 г. Проблема нахождения более общих условий, при которых справедлива эта теорема, была решена Феллером, Хинчином и Леви [85], [86], [140], [145]. Докажем здесь эту теорему для двух частных случаев, достаточных для большинства приложений к статистике.

Рассмотрим сперва *случай равных компонент*, т. е. случай, когда все ξ_r в (17.4.1) имеют одинаковые распределения. В этом случае $m = nm_1$, $\sigma = \sigma_1\sqrt{n}$ и нормированная величина может быть записана в виде

$$\frac{\xi - m}{\sigma} = \frac{\xi - nm_1}{\sigma_1\sqrt{n}} = \frac{1}{\sigma_1\sqrt{n}} \sum_1^n (\xi_r - m_1),$$

где все отклонения $\xi_r - m_1$ имеют одинаковые распределения. Обозначим через $\varphi_1(t)$ характеристическую функцию любого такого отклонения, а через $F(x)$ и $\varphi(t)$ — функцию распределения и характеристи-

ческую функцию нормированной величины $\frac{\xi - m}{\sigma}$. Тогда из (15.9.2) и (15.12.1) будет следовать, что

$$(17.4.2) \quad \varphi(t) = \left[\varphi_1 \left(\frac{t}{\sigma_1 \sqrt{n}} \right) \right]^n.$$

Первые два момента величины $\xi_i - m_1$ равны 0 и σ_1^2 , так что, согласно (10.1.3), для соответствующей характеристической функции будем иметь выражение

$$\varphi_1(t) = 1 - \frac{1}{2} \sigma_1^2 t^2 + o(t^2).$$

Заменяя t на $\frac{t}{\sigma_1 \sqrt{n}}$, получим из (17.4.2)

$$\varphi(t) = \left(1 - \frac{t^2}{2n} + \frac{\zeta(n, t)}{n} \right)^n,$$

где для каждого фиксированного t величина $\zeta(n, t)$ стремится к нулю при $n \rightarrow \infty$. Отсюда следует, что $\varphi(t) \rightarrow e^{-\frac{t^2}{2}}$ при любом t , и, следовательно, мы можем, так же как и в параграфе 16.4, заключить, что соответствующая функция распределения $F(x)$ стремится к $\Phi(x)$ при любом x . Таким образом, мы получаем следующий случай центральной предельной теоремы, доказанный впервые Линнебергом и Леви [24], [148]:

Если независимые случайные величины ξ_1, ξ_2, \dots имеют все одно и то же распределение вероятностей и если каждое ξ_i имеет среднее значение m_1 и стандартное отклонение σ_1 , то сумма $\xi = \sum_1^n \xi_i$ асимптотически нормальна $(nm_1, \sigma_1 \sqrt{n})$. Следовательно, среднее арифметическое $\bar{\xi} = \frac{1}{n} \sum_1^n \xi_i$ асимптотически нормально $(m_1, \frac{\sigma_1}{\sqrt{n}})$.

В случае равных компонент для выполнения центральной предельной теоремы достаточно, таким образом, предположить, что общее распределение всех ξ_i имеет конечный момент второго порядка. При переходе к общему случаю, когда не предполагается, что переменные ξ_i имеют одинаковое распределение, существование всех вторых моментов величин ξ_i уже недостаточно, и мы должны наложить на ξ_i некоторые дополнительные условия. Эти дополнительные условия сводятся, в сущности говоря, к предположению, что вероятность для отдель-

ного слагаемого ξ_v , составить значительную часть суммы ξ мала. Интересное достаточное условие такого рода было найдено Линдебергом. Однако мы приведем здесь только следующую несколько менее общую теорему, принадлежащую Ляпунову:

Пусть ξ_1, ξ_2, \dots суть независимые случайные величины. Обозначим через m_v среднее значение и через σ_v стандартное отклонение величины ξ_v . Предположим, что третий абсолютный момент величины ξ_v относительно ее среднего

$$\rho_v^3 = E(|\xi_v - m_v|^3)$$

конечен при любом v и положим

$$\rho^3 = \rho_1^3 + \rho_2^3 + \dots + \rho_n^3.$$

Если выполняется условие

$$\lim_{n \rightarrow \infty} \frac{\rho}{\sigma} = 0;$$

то сумма $\xi = \sum_1^n \xi_v$ асимптотически нормальна (m, σ), где m и σ заданы соотношениями (17.3.1).

В частном случае, когда все ξ_v имеют одно и то же распределение, имеем $\rho^3 = n\rho_1^3$, $\sigma^2 = n\sigma_1^2$ и, таким образом, $\frac{\rho}{\sigma} = \frac{\rho_1}{\sigma_1 \sqrt{n}}$, так что условие $\lim_{n \rightarrow \infty} \frac{\rho}{\sigma} = 0$ выполнено. Нельзя, однако, заключить, что доказанная выше теорема Линде берга — Леви есть частный случай теоремы Ляпунова, так как первая не предполагает существования третьих моментов.

Чтобы доказать теорему Ляпунова, обозначим через $\varphi_v(t)$ характеристическую функцию v -го отклонения $\xi_v - m_v$, а через $\varphi(t)$ — характеристическую функцию нормированной суммы $\frac{\xi - m}{\sigma} = \frac{1}{\sigma} \sum_1^n (\xi_v - m_v)$.

Тогда из (15.9.2) и (15.12.1) будет следовать, что

$$(17.4.4) \quad \varphi(t) = \prod_1^n \varphi_v\left(\frac{t}{\sigma}\right).$$

Как и раньше, достаточно доказать, что для любого фиксированного t имеем $\varphi(t) \rightarrow e^{-\frac{t^2}{2}}$ при $n \rightarrow \infty$, так как тогда наша теорема будет

непосредственно следовать из теоремы непрерывности (см. параграф 10.4). Используя (16.4.4) при $k=3$, получим

$$\varphi_v(t) = E(e^{it(\xi_v - \rho_v)}) = 1 - \frac{1}{2} \sigma_v^2 t^2 + \theta \frac{\rho_v^3 t^3}{6},$$

где, как и в параграфе 16.4, θ обозначает некоторое число, не превосходящее единицу по модулю. Далее получим

$$\log \varphi_v\left(\frac{t}{c}\right) = \log \left(1 - \frac{\sigma_v^2 t^2}{2c^2} + \theta \frac{\rho_v^3 t^3}{6c^3}\right) = \log(1+z),$$

где

$$z = -\frac{\sigma_v^2 t^2}{2c^2} + \theta \frac{\rho_v^3 t^3}{6c^3}.$$

Вследствие условия (17.4.3) для всех достаточно больших значений n будем иметь

$$\frac{\rho_v}{c} \leq \frac{\rho}{c} < 1,$$

и, следовательно, замечая, что, в силу (15.4.6), $\sigma_v \leq \rho_v$ при любом v ,

$$z = \theta \frac{\rho_v^2 t^2}{2c^2} + \theta \frac{\rho_v^3 t^3}{6c^3} = \theta \frac{\rho_v^2}{c^3} \left(\frac{t^2}{2} + \frac{|t|^3}{6}\right).$$

Условие (17.4.3) показывает, что $z \rightarrow 0$ при $n \rightarrow \infty$ для любого фиксированного t . Таким образом, $|z| < \frac{1}{2}$ для всех достаточно больших n .

При $|z| < \frac{1}{2}$ имеем, однако,

$$\begin{aligned} \log(1+z) &= \frac{z}{1} - \frac{z^2}{2} \left(1 - \frac{2}{3} z + \frac{2}{4} z^2 - \dots\right) = \\ &= z + \frac{1}{2} \theta z^2 \left(1 + \frac{1}{2} + \frac{1}{2^2} + \dots\right) = z + \theta z^2 \end{aligned}$$

и, следовательно,

$$\begin{aligned} \log \varphi_v\left(\frac{t}{c}\right) &= -\frac{\sigma_v^2}{c^2} \cdot \frac{t^2}{2} + \theta \frac{\rho_v^3}{c^3} \cdot \frac{t^3}{6} + \theta \frac{\rho_v^4}{c^4} \left(\frac{t^2}{2} + \frac{|t|^3}{6}\right)^2 = \\ &= -\frac{\sigma_v^2}{c^2} \cdot \frac{t^2}{2} + \theta \frac{\rho_v^3}{c^3} \left(\frac{1}{6} |t|^3 + \left(\frac{1}{2} t^2 + \frac{1}{6} |t|^3\right)^2\right). \end{aligned}$$

Суммируя по $v = 1, 2, \dots, n$, получаем, согласно (17.4.4),

$$\log \varphi(t) = -\frac{t^2}{2} + \theta \frac{\rho^3}{c^3} \left(\frac{1}{6} |t|^3 + \left(\frac{1}{2} t^2 + \frac{1}{6} |t|^3\right)^2\right).$$

Из условия (17.4.3) следует, что при $n \rightarrow \infty$ $\log \varphi(t)$ стремится к $-\frac{t^2}{2}$ для каждого фиксированного t , и, таким образом, теорема Ляпунова доказана.

В случае величины $\chi := \sum_1^n \xi_r$ (см. параграф 16.6), выражающей число благоприятных исходов в ряде из n независимых испытаний с вероятностями p_1, \dots, p_n , имеем

$$\rho_r^3 = E(|\xi_r - p_r|^3) = p_r q_r (p_r^2 + q_r^2) \leq p_r q_r,$$

$$\rho^3 \leq \sum_1^n p_r q_r, \quad \sigma^2 = \sum_1^n p_r q_r,$$

и, таким образом,

$$\frac{\rho}{\sigma} \leq \left(\sum_1^n p_r q_r \right)^{-\frac{1}{6}}.$$

Если ряд $\sum_1^\infty p_r q_r$ расходится, то условие Ляпунова (17.4.3) выполнено и, следовательно, величина χ асимптотически нормальна $\left(\sum_1^n p_r, \sqrt{\sum_1^n p_r q_r} \right)$. Для расходимости ряда $\sum p_r q_r$ достаточно, чтобы существовало число $c > 0$, такое, что $c < p_r < 1 - c$ при любом r . Если же ряд $\sum p_r q_r$ сходится, то можно доказать (см. Крамер [11]), что величина χ не является асимптотически нормальной.

17.5. Дополнительные замечания к центральной предельной теореме. Центральная предельная теорема была развита впоследствии в различных направлениях. В этом параграфе мы сделаем несколько кратких замечаний по этому поводу, в то время как следующие параграфы будут посвящены некоторым частным вопросам, принадлежащим к тому же кругу идей.

1. Теоремы предыдущего параграфа имели дело исключительно с *функциями распределения* случайных величин. Было показано, что функция распределения нормированной суммы $\frac{\xi - m}{\sigma}$ стремится к нормальной функции распределения $\Phi(x)$. Если все компоненты ξ_r — непрерывного типа, то возникает вопрос, стремится ли функция плотности величины $\frac{\xi - m}{\sigma}$ к нормальной функции плотности $\Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Действительно, можно показать (Крамер [11], [70]), что

это имеет место, если наложить на компоненты некоторые общие условия (см. (17.7.4)).

2. В проблемах теоретической статистики часто приходится рассматривать функцию $g(\xi_1, \dots, \xi_n)$ от n независимых случайных величин, где n может рассматриваться как большое число. Если функция g имеет непрерывные производные первого и второго порядка в окрестности точки $\mathbf{m} = (m_1, \dots, m_n)$, где m_v обозначает среднее значение величины ξ_v , то можно взять разложение Тейлора

$$(17.5.1) \quad g(\xi_1, \dots, \xi_n) = g(m_1, \dots, m_n) + \sum_1^n c_v(\xi_v - m_v) + R;$$

здесь c_v есть значение $\frac{\partial g}{\partial \xi_v}$ в точке \mathbf{m} , а остаточный член R содержит производные второго порядка. Первое слагаемое в правой части постоянно, а второе слагаемое есть сумма n независимых случайных величин, каждая из которых имеет среднее значение, равное нулю. Согласно центральной предельной теореме, при некоторых общих условиях сумма первых двух слагаемых асимптотически нормальна со средним значением, равным первому слагаемому. Во многих важных случаях можно показать, что в пределе при $n \rightarrow \infty$ присутствие слагаемого R не оказывает влияния на распределение, так что функция g при больших значениях n распределена асимптотически нормально (см. Мизес [157], [158]). Мы вернемся к этому вопросу в главе 28.

3. Центральная предельная теорема может быть распространена на различные случаи, когда величины ξ_v не являются независимыми. Укажем здесь только на один случай (см. Крамер [10]) важный для приложений, особенно к биологии. За дальнейшими сведениями отсылаем читателя к книге Леви [25] и статьям Бернштейна [63], Кэптейна [135] и Виксела [230]. Условимся употреблять здесь терминологию, непосредственно связанную с некоторыми биологическими приложениями. Если наша случайная величина есть размер некоторого определенного органа, который мы наблюдаем, то действительный размер этого органа в конкретном частном случае можно рассматривать часто как совместный эффект большого числа взаимно независимых причин, действующих в определенном порядке в течение роста организма. Если эти причины просто складывают свои эффекты, которые предполагаются случайными величинами, то, согласно центральной предельной теореме, их сумма распределена асимптотически нормально.

Однако, вообще говоря, предположение, что взаимодействие причин выражается простым сложением, не кажется правдоподобным. Более естественно предположение, что каждая причина даст некоторый импульс, эффект которого зависит от силы импульса и от размера органа, уже достигнутого к моменту действия импульса.

Предположим, что мы имеем n импульсов ξ_1, \dots, ξ_n , действующих в порядке возрастания их индексов. Будем рассматривать их как независимые случайные величины. Обозначим через x , размер органа, достигнутый благодаря действию импульсов ξ_1, \dots, ξ_n . Можно предположить, например, что прирост, вызванный импульсом ξ_{n+1} , пропорционален ξ_{n+1} и некоторой функции $g(x_n)$ от размера органа в данный момент времени:

$$(17.5.2) \quad x_{n+1} = x_n + \xi_{n+1} g(x_n).$$

Мы будем иметь, следовательно,

$$\xi_1 + \xi_2 + \dots + \xi_n = \sum_0^{n-1} \frac{x_{n+1} - x_n}{g(x_n)}.$$

Если каждый импульс вызывает только незначительный прирост органа, то приближенно

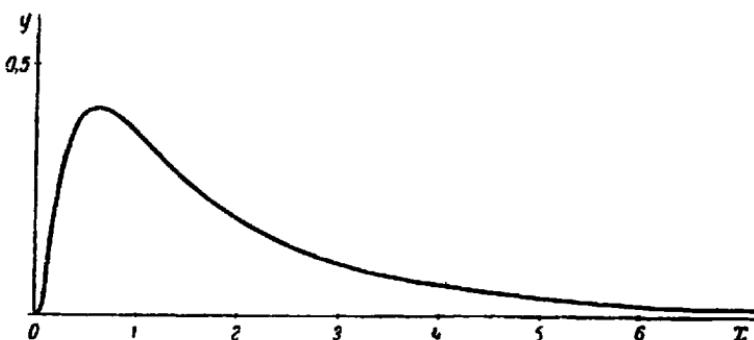
$$\xi_1 + \xi_2 + \dots + \xi_n = \int_{x_0}^x \frac{dt}{g(t)},$$

где $x = x_n$ обозначает окончательный размер органа. По предположению, величины ξ_1, \dots, ξ_n независимы, и n можно рассматривать как большое число. В условиях центральной предельной теоремы имеем, таким образом, что в пределе функция случайной величины x , появляющаяся в правой части последнего соотношения, распределена нормально.

Рассмотрим, например, случай $g(t) = t$. Эффект каждого импульса прямо пропорционален тогда размеру органа в соответствующий момент времени. В этом случае мы, таким образом, находим, что $\log x$ распределен нормально. Вообще, если $\log(x - a)$ нормален (m, σ), то легко видеть, что величина x имеет при $x > a$ функцию плотности

$$(17.5.3) \quad \frac{1}{\sqrt{(x-a)V2\pi}} e^{-\frac{(\log(x-a)-m)^2}{2\sigma^2}},$$

а при $x \leq a$ ее функция плотности равна нулю. Соответствующая кривая плотности унимодальна и положительно асимметрична. Она изображена на фиг. 17. Эта функция плотности *логарифмически-нормального распределения* может быть использована при разложении в ряд, подобно тому как функция плотности нормального распределения используется в следующем параграфе.



Фиг. 17. Кривая плотности логарифмически-нормального распределения
 $a = 0; m = 0,46; \sigma = 1$

Аналогичные рассуждения могут быть применены также в других случаях, например, в некоторых областях экономической статистики. Рассмотрим распределение дохода или имущества в некоторой популяции. Каждое конкретное значение на шкале имущества можно рассматривать как эффект большого числа импульсов, каждый из которых вызывает некоторое приращение богатства. Конечно, допущение, что эффект такого импульса будет пропорционален уже достигнутому богатству, несколько спорно. Если все-таки принять это допущение, то нужно ожидать, что распределение дохода или имущества приблизительно логарифмически-нормальное. Для низких значений дохода логарифмически-нормальная кривая довольно хорошо согласуется с действительными кривыми дохода (Квензель [201], [202]). Для умеренных же и высоких доходов более подходящим, вообще говоря, представляется распределение Парето (см. параграф 19.3).

17.6. Ортогональное разложение, основанное на нормальном распределении. Рассмотрим случайную величину

$$(17.6.1) \quad \xi = \xi_1 + \xi_2 + \dots + \xi_n,$$

являющуюся суммой n независимых случайных величин. В условиях центральной предельной теоремы функция распределения нормированной величины $\frac{\xi - m}{\sigma}$ при больших n приближенно равна $\Phi(x)$. Далее,

если все компоненты ξ_i имеют распределения непрерывного типа, то плотность вероятности $f(x) = F'(x)$ будет (см. параграф 17.5) при некоторых общих условиях равномерности приближенно равна нормальной функции плотности *) $\varphi(x) = \Phi'(x)$. Если положить

$$(17.6.2) \quad \begin{aligned} F(x) &= \Phi(x) + R(x), \\ f(x) &= \varphi(x) + r(x), \end{aligned}$$

то $R(x)$ и $r(x) = R'(x)$ будут малы при больших значениях n , так что $\Phi(x)$ и $\varphi(x)$ можно рассматривать как первые приближения к функциям $F(x)$ и $f(x)$. Тогда естественно поставить вопрос, можем ли мы при дальнейшем анализе остаточных членов $R(x)$ и $r(x)$ найти более точные приближения, например в форме некоторого разложения $R(x)$ и $r(x)$ в ряды.

Этот же вопрос можно рассматривать с более общей точки зрения. В приложениях мы часто сталкиваемся с функциями плотности и функциями распределения, которые приблизительно нормальны даже в случаях, когда нет основания предполагать, что соответствующая случайная величина образуется в виде (17.6.1) как сумма независимых величин. Тогда естественно записывать эти функции в виде (17.6.2) и пытаться найти подходящее разложение для остаточных членов.

Рассмотрим здесь два различных типа таких разложений. В настоящем параграфе мы будем иметь дело с разложением по ортогональным полиномам, известным под названием ряда Грама — Шарлье типа А (см. [9], [65], [118]), а следующий параграф будет посвящен асимптотическому разложению, введенному Эджвортом. В обоих случаях, поскольку полные выводы очень сложны, мы ограничимся формальными разложениями и краткой формулировкой основных результатов.

Рассмотрим сперва случайную величину ξ с распределением непрерывного типа, не предполагая, что ξ представляется в виде (17.6.1). Как обычно, обозначаем через m среднее значение и через σ стандартное отклонение величины ξ , а через μ_n — центральный момент n -го порядка (см. параграф 15.4) величины ξ , который предполагается конечным

*) Как правило, мы используем букву φ для обозначения характеристической функции. Однако в параграфах 17.6 и 17.7 $\varphi(x)$ будет обозначать нормальную функцию плотности $\varphi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, а для обозначения характеристических функций будет служить буква ψ .

при любом y . Рассмотрим нормированную величину $\frac{\xi - m}{\sigma}$ и обозначим через $F(x)$ ее функцию распределения, а через $f(x) = F'(x)$ — ее функцию плотности.

Для любой функции плотности рассмотрим разложение в виде

$$(17.6.3) \quad f(x) = c_0 \varphi(x) + \frac{c_1}{1!} \varphi'(x) + \frac{c_2}{2!} \varphi''(x) + \dots,$$

где c_v — постоянные коэффициенты. Согласно (12.6.4), $\varphi^{(v)}(x) = (-1)^v H_v(x) \varphi(x)$, где $H_v(x)$ есть полином Эрмита степени v , и, таким образом, (17.6.3) есть фактически разложение вида (12.6.2) по ортогональным полиномам. Определим теперь, так же как и в параграфе 12.6, коэффициенты этого разложения, предполагая возможность почлененного интегрирования ряда (17.6.3). Умножая этот ряд на $H_v(x)$ и интегрируя, непосредственно из соотношений ортогональности (12.6.6) получим

$$(17.6.4) \quad c_v = (-1)^v \int_{-\infty}^{\infty} H_v(x) f(x) dx.$$

Функция плотности $f(x)$ нормированной величины $\frac{\xi - m}{\sigma}$ имеет среднее значение, равное нулю, стандартное отклонение, равное единице, и y -й момент, равный $\frac{\mu_y}{\sigma^y}$. Соответственно находим $c_0 = 1$, $c_1 = c_2 = 0$, так что разложение (17.6.3) и разложение, полученное формальным интегрированием, могут быть записаны в виде

$$(17.6.5) \quad \begin{aligned} F(x) &= \Phi(x) + \frac{c_3}{3!} \Phi^{(3)}(x) + \frac{c_4}{4!} \Phi^{(4)}(x) + \dots, \\ f(x) &= \varphi(x) + \frac{c_3}{3!} \varphi^{(3)}(x) + \frac{c_4}{4!} \varphi^{(4)}(x) + \dots, \end{aligned}$$

где c_v заданы равенствами (17.6.4). Из выражений (12.6.5) первых полиномов Эрмита получим, в частности, обозначая через γ_1 и γ_2 коэффициенты асимметрии и эксцесса (см. параграф 15.8) величины ξ .

$$(17.6.6) \quad \begin{aligned} c_3 &= -\frac{\mu_3}{\sigma^3} = -\gamma_1, \\ c_4 &= \frac{\mu_4}{\sigma^4} - 3 = \gamma_2, \\ c_5 &= -\frac{\mu_5}{\sigma^5} + 10 \frac{\mu_3}{\sigma^3}, \\ c_6 &= \frac{\mu_6}{\sigma^6} - 15 \frac{\mu_4}{\sigma^4} + 30. \end{aligned}$$

Со всякой нормированной величиной $\frac{\xi - m}{\sigma}$, имеющей конечные моменты всех порядков, можно, таким образом, формально связать разложения (17.6.5), коэффициенты которых заданы равенством (17.6.4). Будут ли, однако, эти разложения действительно сходиться и представлять $f(x)$ и $F(x)$?

Можно показать (см., например, Крамер [69], [70]), что если только интеграл

$$(17.6.6a) \quad \int_{-\infty}^{\infty} e^{\frac{x^2}{4}} dF(x)$$

сходится, то первый ряд (17.6.5) будет сходиться при любом x к сумме $F(x)$. Если, кроме того, функция плотности $f(x)$ имеет ограниченную вариацию на $(-\infty, \infty)$, то второй ряд (17.6.5) сходится к $f(x)$ в каждой точке непрерывности функции $f(x)$. С другой стороны, можно показать (см. упражнение 18, стр. 286), что, если эти условия не выполнены, то разложения могут расходиться. Таким образом, законность разложений (17.6.5) можно утверждать только для сравнительно небольшого класса распределений. Действительно, большинство важных распределений, встречающихся в следующих двух главах, не входит в этот класс.

Однако в практических приложениях в большинстве случаев знание свойств сходимости наших разложений не имеет большого значения. Что в действительности интересно знать, это — *дает ли небольшое число слагаемых* (обычно не более двух или трех) *достаточно хорошее приближение к функциям $f(x)$ и $F(x)$* . Если это имеет место, то нас не интересует больше вопрос, сходятся или расходятся наши бесконечные ряды. Обратно, если мы знаем, что один из рядов (17.6.5) сходится, то это не принесет практической пользы, если для того, чтобы получить частную сумму ряда, дающую достаточно хорошее приближение, необходимо вычислить большое число коэффициентов c_n .

Указанный вопрос приобретает особенное значение, когда рассматривается величина ξ , образованная в виде (17.6.1). Как указано выше, при некоторых общих условиях функции $F(x)$ и $f(x)$ при большом n приближенно равны $\Phi(x)$ и $\varphi(x)$. Улучшится ли приближение, если ввести в формулу (17.6.2) член, содержащий третью производную? Даст ли рассмотрение дальнейших членов разложения еще лучшее

приближение? Мы увидим ниже, что здесь мы сталкиваемся с вопросом, относящимся к *асимптотическим свойствам* наших разложений при больших значениях n .

Для упрощения алгебраических выкладок рассмотрим *случай равных компонент* (см. параграф 17.4), когда ξ_1, \dots, ξ_n в (17.6.1) имеют одно и то же распределение со средним значением m_1 и стандартным отклонением σ_1 , так что $m = nm_1$, $\sigma = \sigma_1 \sqrt{n}$. В этом случае мы предполагаем изучать поведение коэффициентов c , *A-рядов* (17.6.5) при больших значениях n .

Пусть $\Phi(t)$ — характеристическая функция нормированной величины $\frac{\xi - m}{\sigma}$, а $\psi_1(t)$ — характеристическая функция отклонения $\xi_1 - m_1$. Тогда, согласно (17.4.2),

$$\Phi(t) = \left[\psi_1 \left(\frac{t}{\sigma_1 \sqrt{n}} \right) \right]^n.$$

Обозначим через x_v семи-инварианты величины $\xi - m = \sum_1^n (\xi_i - m_1)$, а через x'_v — сечи-инварианты величины $\xi_1 - m_1$ ($v = 1, 2, \dots$) и положим

$$(17.6.7) \quad \lambda_v = \frac{x_v}{c_v}, \quad \lambda'_v = \frac{x'_v}{c'_1}.$$

Тогда, согласно (15.12.8),

$$(17.6.8) \quad x_v = nx'_v, \quad \lambda_v = \frac{\lambda'_v}{\frac{v^2}{n^2} - 1}.$$

По определению характеристической функции $\Phi(t)$, имеем

$$e^{\frac{t^2}{2}} \Phi(t) = \int_{-\infty}^{\infty} e^{\frac{t^2}{2} + itx} f(x) dx,$$

и, следовательно, согласно (12.6.7), получаем разложение

$$(17.6.9) \quad e^{\frac{t^2}{2}} \Phi(t) = \sum_0^{\infty} \frac{c_v}{v!} (-it)^v,$$

или

$$(17.6.10) \quad \Phi(t) = e^{-\frac{t^2}{2}} + \frac{c_3}{3!} (-it)^3 e^{-\frac{t^2}{2}} + \frac{c_4}{4!} (-it)^4 e^{-\frac{t^2}{2}} + \dots,$$

где c_v задано равенством (17.6.4).

Вообще говоря, нельзя утверждать, что степенной ряд, стоящий в правой части (17.6.10), сходится; мы можем только сказать, что это соотношение выполняется асимптотически при малых значениях t в том же смысле, как и соотношение (10.1.3).

Сравнивая (17.6.10) с разложением

$$(17.6.11) \quad f(x) = \varphi(x) + \frac{c_3}{3!} \varphi^{(3)}(x) + \frac{c_4}{4!} \varphi^{(4)}(x) + \dots,$$

видим, что члены этих двух разложений соответствуют друг другу посредством следующего соотношения, получаемого из (10.5.5):

$$(17.6.12) \quad \int_{-\infty}^{\infty} e^{itx} \varphi^{(v)}(x) dx = (-it)^v e^{-\frac{t^2}{2}}. \quad (v = 0, 1, 2, \dots).$$

Как указано по аналогичному поводу в параграфе 15.10, со степенным рядом вида (17.6.9) можно обращаться чисто формально, не заботясь о сходимости, до тех пор, пока нас интересуют алгебраические соотношения между различными параметрами, такими, как c_v и λ'_v . Следовательно в соответствии с параграфом 15.10, используя (17.6.7), мы можем написать

$$\begin{aligned} \psi_1(t) &= e^{\sum_1^{\infty} \frac{\lambda'_v}{v!} (it)^v}, \\ \psi(t) &= \left[\psi_1\left(\frac{t}{\sigma_1 \sqrt{n}}\right) \right]^n = e^{n \sum_1^{\infty} \frac{\lambda'_v}{v!} \left(\frac{it}{\sigma_1 \sqrt{n}}\right)^v}. \end{aligned}$$

Величина $\xi_1 - m_1$ имеет среднее значение, равное нулю, и стандартное отклонение σ_1 . Следовательно, $\lambda'_1 = 0$ и $\lambda'_2 = \sigma_1^2$, так что $\lambda'_1 = 0$ и $\lambda'_2 = 1$. Отсюда последнее соотношение может быть записано в виде

$$(17.6.13) \quad e^{\frac{n}{2}} \psi(t) = e^{n \sum_3^{\infty} \frac{\lambda'_v}{v!} \left(\frac{it}{\sigma_1 \sqrt{n}}\right)^v}.$$

Чтобы получить явное выражение для c_v через λ'_v , остается только расположить это выражение по степеням t и отождествить получающийся

ряд с (17.6.9). При этом получаем

$$c_3 = -\frac{\lambda'_3}{n^{1/2}}, \quad c_4 = \frac{\lambda'_4}{n},$$

(17.6.14)

$$c_5 = -\frac{\lambda'_5}{n^{3/2}}, \quad c_6 = \frac{\lambda'_6}{n^2} + \frac{10\lambda'^2_3}{n},$$

и вообще

$$\sum_{v=0}^{\infty} \frac{c_v}{v!} (it)^v = \sum_{k=0}^{\infty} \frac{n^k}{k!} \left[\sum_{v=0}^{\infty} \frac{\lambda'_v}{v!} \left(\frac{it}{\sqrt{n}} \right)^v \right]^k,$$

откуда

$$(17.6.15) \quad c_v = \frac{a_{v1}n + a_{v2}n^2 + \dots + a_{v[\gamma/3]}n^{[\gamma/3]}}{n^{v/2}},$$

где $\left[\frac{\gamma}{3}\right]$ обозначает наибольшее целое число $\leq \frac{\gamma}{3}$, а a_{vk} — полиномы относительно переменных λ'_v , не зависящие от n . Таким образом,

$$c_v = O(n^{[\gamma/3]-v/2})$$

при $n \rightarrow \infty$. Следующая таблица показывает порядок величины c_v для первых значений v .

v	порядок c_v
3	$n^{-1/2}$
4, 6	n^{-1}
5, 7, 9	$n^{-3/2}$
8, 10, 12	n^{-2}
11, 13, 15	$n^{-5/2}$

Мы видим, что порядок величины члена А-ряда с возрастанием v не убывает монотонно. Преложим, например, что мы хотим вычислить частную сумму ряда (17.6.11) с учетом всех членов, вызывающих поправки к $\varphi(x)$ порядка $n^{-1/2}$ или n^{-1} . Тогда из нашей таблицы следует, что мы должны рассматривать члены до $v=6$ включительно. Для того чтобы вычислить коэффициенты c_v этих членов по формулам (17.6.6) или (17.6.14), нужно знать моменты μ , или семиинварианты λ' до шестого порядка. Соотношения (17.6.14) показывают,

однако, что поправки порядка $n^{-\frac{1}{2}}$ и n^{-1} не содержат семи-инвариантов порядка выше четвертого, так что в действительности нет необходимости вычислять эти семи-инварианты. Если мы захотим пойти далее и учесть члены, содержащие множители $n^{-\frac{3}{2}}$, n^{-2} и т. д., то легко видеть, что мы столкнемся с подобным же явлением.

Таким образом, A -ряд Грама — Шарлье не может рассматриваться как удовлетворительное решение проблемы разложения функций $F(x)$ и $f(x)$. Действительно, нам нужен ряд, который давал бы разложение по степеням $n^{-\frac{1}{2}}$, и такой, что для определения членов до некоторого порядка не приходилось бы вычислять лишние моменты или семи-инварианты. Этим условиям удовлетворяет ряд Эджвортса, который мы рассмотрим в следующем параграфе.

17.7. Асимптотическое разложение, основанное на нормальном распределении.

В предыдущем параграфе разложение функции

$$(17.7.1) \quad e^{\frac{t}{2}} \phi(t) = e^{\sum_{v=1}^{\infty} \frac{\lambda'_v}{v!} \left(\frac{it}{\sqrt{n}} \right)^v}$$

по степеням t давало выражение для коэффициентов c , A -ряда. Однако эта же функция может быть также разложена и другими способами, например, по степеням $n^{-\frac{1}{2}}$. Записывая

$$e^{\frac{t}{2}} \phi(t) = e^{(it)^2 \sum_1^{\infty} \frac{\lambda'_v+2}{(v+2)!} \left(\frac{it}{\sqrt{n}} \right)^v} = \sum_{h=0}^{\infty} \frac{(it)^{2h}}{h!} \left[\sum_{v=1}^{\infty} \frac{\lambda'_v+2}{(v+2)!} \left(\frac{it}{\sqrt{n}} \right)^v \right]^h,$$

получим

$$\phi(t) = e^{-\frac{t^2}{2}} + \sum_1^{\infty} \frac{b_{v,v+2} (it)^{v+2} + b_{v,v+4} (it)^{v+4} + \dots + b_{v,3v} (it)^{3v}}{n^{v/2}} e^{-\frac{t^2}{2}},$$

где $b_{v,v+2h}$ — полиномы относительно переменных $\lambda'_3, \dots, \lambda'_{v-h+3}$, не зависящие от n . Согласно (17.6.12), имеем соответствующее разложение по степеням $n^{-\frac{1}{2}}$:

$$(17.7.2) \quad f(x) = \varphi(x) + \sum_1^{\infty} (-1)^v \frac{b_{v,v+2} \varphi^{(v+2)}(x) + \dots + b_{v,3v} \varphi^{(3v)}(x)}{n^{v/2}}.$$

Записывая в одну строку члены одного порядка относительно n и выписывая первые члены разложения, имеем

$$\begin{aligned}
 f(x) &= \varphi(x) - \\
 &- \frac{1}{3!} \cdot \frac{\lambda_3'}{n^{1/2}} \varphi^{(3)}(x) + \\
 &- \frac{1}{4!} \cdot \frac{\lambda_4'}{n} \varphi^{(4)}(x) + \frac{10}{6!} \cdot \frac{\lambda_3'^2}{n} \varphi^{(6)}(x) - \\
 &- \frac{1}{5!} \cdot \frac{\lambda_5'}{n^{3/2}} \varphi^{(5)}(x) - \frac{35}{7!} \cdot \frac{\lambda_3' \lambda_4'}{n^{5/2}} \varphi^{(7)}(x) - \frac{280}{9!} \cdot \frac{\lambda_3'^3}{n^3} \varphi^{(9)}(x) + \\
 &\dots
 \end{aligned}$$

Согласно (17.6.7) и (17.6.8), коэффициенты этого разложения могут быть выражены через семи-инварианты x_i , которые в свою очередь могут быть заменены центральными моментами μ_i с помощью соотношений (15.10.5). Этим путем мы получаем ряд, введенный Эджвортом [80]:

$$\begin{aligned}
 f(x) &= \varphi(x) - \\
 (17.7.3) \quad &- \frac{1}{3!} \cdot \frac{\mu_3}{\sigma^3} \varphi^{(3)}(x) + \\
 &+ \frac{1}{4!} \cdot \left(\frac{\mu_4}{\sigma^4} - 3 \right) \varphi^{(4)}(x) + \frac{10}{6!} \left(\frac{\mu_3}{\sigma^3} \right)^2 \varphi^{(6)}(x) - \\
 &- \frac{1}{5!} \left(\frac{\mu_5}{\sigma^5} - 10 \frac{\mu_3}{\sigma^3} \right) \varphi^{(5)}(x) - \frac{35}{7!} \frac{\mu_3}{\sigma^3} \left(\frac{\mu_4}{\sigma^4} - 3 \right) \varphi^{(7)}(x) - \frac{280}{9!} \left(\frac{\mu_3}{\sigma^3} \right)^3 \varphi^{(9)}(x) + \\
 &\dots
 \end{aligned}$$

где в каждой строке выписаны члены одного и того же порядка. Для того чтобы получить соответствующее разложение для функции распределения $F(x)$, нужно только заменить $\varphi(x)$ на $\Phi(x)$.

Асимптотические свойства этих рядов были исследованы Крамером ([11], [70]); им было показано, что при довольно общих условиях ряд (17.7.2) действительно дает асимптотическое разложение $f(x)$ по степеням $n^{-1/2}$, с остаточным членом порядка первого отброшенного члена. Аналогичный результат справедлив и для $F(x)$. Если рассматривать только первые члены соответствующих рядов, то, в частности, получим

$$(17.7.4) \quad |F(x) - \Phi(x)| < \frac{A}{\sqrt{n}}, \quad |f(x) - \varphi(x)| < \frac{B}{\sqrt{n}},$$

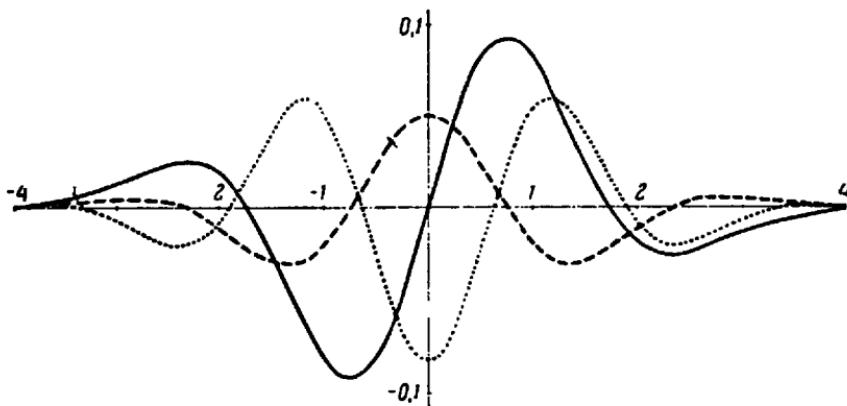
где A и B — постоянные *).

*). Эссином [83] и Бергстрёмом [62] было показано, что это неравенство для $|F - \Phi|$ выполнено при единственном условии, что семи-инвариант x_3' конечен.

Члены порядка $n^{-1/2}$ в ряде Эджворта содержат моменты μ_3, \dots, μ_{n+2} , которые как раз и необходимы для приближения этого порядка. На практике обычно не рекомендуется идти дальше третьего и четвертого момента. Члены, содержащие эти моменты, часто дают достаточно хорошее приближение. Для численных подсчетов необходима таблица производных $\varphi^{(v)}(x)$ — они даны в таблице I в конце книги.

Вводя коэффициенты асимметрии и эксцесса γ_1 и γ_2 (см. параграф 15.8), можно записать выражение для $f(x)$ до членов порядка n^{-1} в виде

$$(17.7.5) \quad f(x) = \varphi(x) - \frac{\gamma_1}{3!} \varphi^{(1)}(x) + \frac{\gamma_2}{4!} \varphi^{(4)}(x) + \frac{10\gamma_1^2}{6!} \varphi^{(6)}(x).$$



Фиг. 18. Производные нормальной функции плотности $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
 $\frac{1}{3!} \varphi^{(3)}(x) \text{ — сплошная линия}, \frac{1}{4!} \varphi^{(4)}(x) \text{ — пунктирная линия}, \frac{10}{6!} \varphi^{(6)}(x) \text{ — точечная линия}.$

Графики производных $\varphi^{(3)}$, $\varphi^{(4)}$ и $\varphi^{(6)}$ с коэффициентами, встречающимися в (17.7.5), изображены на фиг. 18. Кривые $\varphi^{(4)}$ и $\varphi^{(6)}$ симметричны относительно $x = 0$, а третья производная $\varphi^{(3)}$ вводит в выражение асимметричный элемент.

При больших x выражение (17.7.5) будет давать иногда небольшие отрицательные значения для $f(x)$. Это, конечно, вполне согласуется с тем, что (17.7.5) дает *приближенное*, но не *точное* выражение для функции плотности.

Для моды x_0 функции плотности получим из (17.7.5) приближенное выражение $x_0 = -\frac{1}{2} \gamma_1$, т. е. характеристику асимметрии Шарлье.

Далее, имеем

$$\frac{f(0) - \varphi(0)}{\varphi(0)} = \frac{1}{8} \gamma_2 - \frac{5}{24} \gamma_1^2.$$

Первое слагаемое справа характеризует эксцесс кривой плотности $y=f(x)$ относительно нормальной кривой $y=\varphi(x)$ в точке $x=0$ *). Именно эту величину, т. е. $\frac{1}{8} \gamma_2$, Шарлье ввел в качестве характеристики эксцесса. Однако, чтобы получить выражение эксцесса с точностью до членов порядка n^{-1} , надо еще ввести и второе слагаемое с γ_1^2 (см. параграф 15.8).

17.8. Роль нормального распределения в статистике. Нормальное распределение было найдено впервые Муавром [29] в 1733 г. в связи с его исследованием предела биномиального распределения, разобранного в параграфе 16.4.

Открытие Муавра, повидимому, прошло незамеченным и только некоторое время спустя нормальное распределение было снова открыто Гауссом [16] в 1809 г. и Лапласом [22] в 1812 г. Последний затрагивал эту тему уже в некоторых работах, написанных около 1780 г., хотя и не углублялся в нее до своей знаменитой работы 1812 года. Гаусс и Лаплас пришли к нормальной функции в связи со своей работой по теории ошибок наблюдений. Лаплас, кроме того, дал первую (несовершенную) формулировку рассмотренной нами выше центральной предельной теоремы и дал большое число важных приложений нормального распределения к различным вопросам теории вероятностей.

Под влиянием великих работ Гаусса и Лапласа долгое время считалось едва ли не аксиомой, что практически все статистические распределения должны приближаться к нормальному распределению как к идеальной предельной форме, если только мы можем располагать достаточно большим числом достаточно точных наблюдений. Отклонение случайной величины от ее среднего рассматривалось как „ошибка“, подчиненная „закону ошибок“, выраженному нормальным распределением.

*). Если вместо сравнения ординат в среднем значении $x=0$ сравнить ординаты в модах этих двух кривых, то получим в первом приближении

$$\frac{f(x_0) - \varphi(0)}{\varphi(0)} = \frac{1}{8} \gamma_2 - \frac{1}{12} \gamma_1^2.$$

Хотя такая точка зрения определенно преувеличивает роль нормального распределения и должна быть значительно изменена, тем не менее несомненно, что в большом числе важных приложений встречаются распределения по крайней мере приблизительно нормальные. Это имеет место, например, в случае распределений ошибок физических и астрономических измерений, большого числа демографических и биологических распределений, и т. д.

Центральная предельная теорема дает теоретическое объяснение этих эмпирических фактов. Согласно „гипотезе элементарных ошибок“, введенной Хагеном и Бесселем, суммарная ошибка физического или астрономического измерения рассматривается как сумма большого числа взаимно независимых элементарных ошибок. Тогда, согласно центральной предельной теореме, суммарная ошибка должна быть распределена приблизительно нормально. Часто, например в биологических исследованиях, естественно видеть в случайной величине суммарный эффект большого числа независимых причин. Такая же точка зрения может быть применена к случайным величинам, встречающимся во многих вопросах техники и экономики. Так, общее потребление электрической энергии, вырабатываемой некоторой станцией, есть сумма величин энергии, приходящихся на долю отдельных потребителей, прибыль или убыток страховой компании есть сумма прибылей и убытков по каждому отдельному полису и т. д.

В случаях этого рода можно ожидать по крайней мере приблизительно нормальных распределений. Если же число компонент недостаточно велико, или если различные компоненты нельзя рассматривать строго как слагаемые, или их нельзя считать независимыми, то видоизменения центральной предельной теоремы, приведенные в параграфах 17.5—17.7, могут все же показать, что распределение приближенно нормально; в других случаях они могут привести к распределениям, тесно связанным с нормальным, например, к распределению, задаваемому соотношением вида (17.7.3) или к логарифмически-нормальному распределению (17.5.3).

В условиях центральной предельной теоремы среднее арифметическое большого числа независимых величин распределено приблизительно нормально. Замечания, сделанные в связи с выражением (17.5.1), показывают, что то же имеет место и для некоторых функций более общего характера, чем среднее. Эти свойства имеют фундаментальное значение для многих методов, используемых в статистической практике,

где мы в значительной мере имеем дело со средними и другими аналогичными функциями наблюденных значений случайных величин (см. гл. 28).

Существует известное замечание Липмана (цитируемое Пуанкаре [31]), гласящее, что „каждый уверен в справедливости закона ошибок, экспериментаторы — потому, что они думают, что это математическая теорема, математики — потому, что они думают, что это экспериментальный факт“. Стоит отметить, что обе стороны совершенно правы, если только это их убеждение не слишком безусловно: математическое доказательство говорит нам, что *при некоторых ограничительных условиях* мы вправе ожидать нормального распределения, а статистический опыт показывает, что в действительности распределения являются часто *приближенно нормальными*.

ГЛАВА 18

РАЗЛИЧНЫЕ РАСПРЕДЕЛЕНИЯ, СВЯЗАННЫЕ С НОРМАЛЬНЫМ РАСПРЕДЕЛЕНИЕМ

В этой главе мы будем рассматривать распределения некоторых простых функций нормально распределенных величин. Все эти распределения имеют важные статистические приложения; они еще появятся по различным поводам в части III.

18.1. χ^2 -распределение. Пусть ξ — случайная величина, нормальная $(0,1)$. Функция плотности ее квадрата ξ^2 есть, согласно (15.1.4),

$$\frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}$$

при $x > 0$. При $x \leq 0$ функция плотности равна нулю. Соответствующую этой функции плотности характеристическую функцию

$$\int_0^\infty e^{itx} \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} dx = (1 - 2it)^{-\frac{1}{2}}$$

получим, положив в (12.3.4) $a = \lambda = \frac{1}{2}$.

Пусть теперь каждая из n независимых случайных величин ξ_1, \dots, ξ_n нормальна $(0,1)$. Рассмотрим величину

$$(18.1.1) \quad \chi^2 = \sum_1^n \xi_v^2.$$

Каждая величина ξ^2 имеет характеристическую функцию $(1-2it)^{-n/2}$; следовательно, по теореме умножения (15.12.1), сумма χ^2 имеет характеристическую функцию

$$(18.1.2) \quad E(e^{it\chi^2}) = (1-2it)^{-n/2}.$$

Мы пришли к характеристической функции, которая может быть получена, если в (12.3.4) положить $a = \frac{1}{2}$, $\lambda = \frac{1}{2}n$; соответствующее распределение определяется, таким образом, функцией плотности $f(x; \frac{1}{2}, \frac{1}{2}n)$ (см. формулу (12.3.3)). Введем обозначение для этой функции плотности, полагая при любом $n = 1, 2 \dots$

$$(18.1.3) \quad k_n(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{при } x > 0, \\ 0 & \text{при } x \leq 0. \end{cases}$$

Итак, $k_n(x)$ есть плотность вероятности величины χ^2 , т. е.

$$k_n(x) dx = P(x < \chi^2 < x + dx).$$

Соответствующая функция распределения равна при $x \leq 0$ нулю, а при $x > 0$

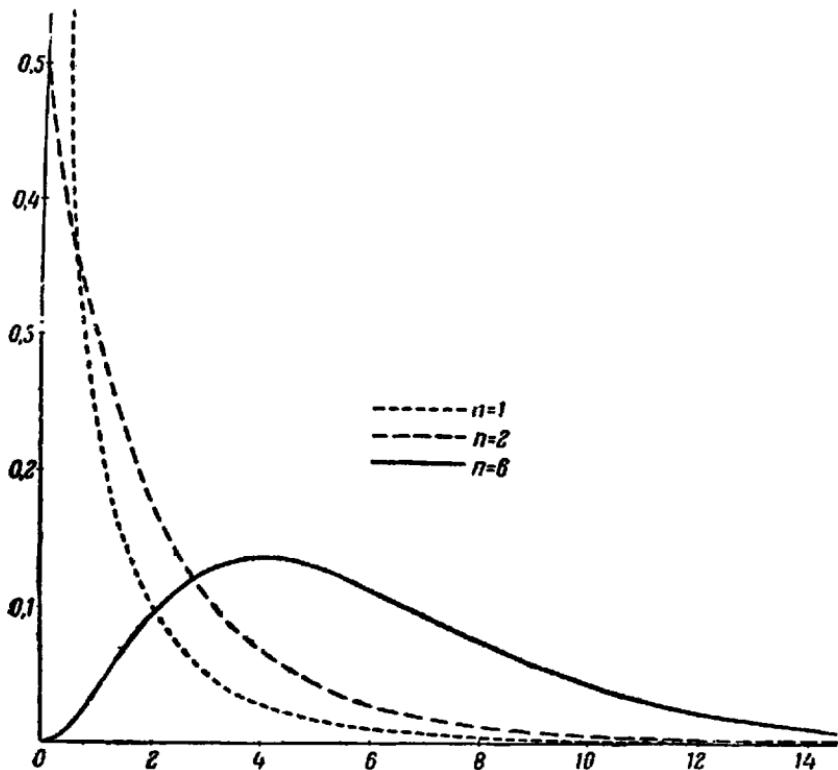
$$(18.1.4) \quad K_n(x) = P(\chi^2 \leq x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \int_0^x t^{\frac{n}{2}-1} e^{-\frac{t}{2}} dt.$$

Распределение, заданное функцией плотности $k_n(x)$ или функцией распределения $K_n(x)$, известно под названием χ^2 -распределения. Приложения этого важного распределения к статистике излагаются в главе 30. χ^2 -распределение содержит параметр n , который часто называется числом степеней свободы этого распределения. Смысл этого термина будет объяснен в главе 29. χ^2 -распределение было впервые рассмотрено Хельмертом [125] и К. Пирсоном [183].

При $n \leq 2$ функция плотности $k_n(x)$ постоянно убывает для $x > 0$, а при $n > 2$ имеет единственный максимум в точке $x = n - 2$. Графики функций $k_n(x)$ для некоторых значений n изображены на фиг. 19.

Моменты a_v и семи-инварианты x_v , χ^2 -распределения конечны при всех v , и их общие выражения могут быть получены, например, из характеристической функции (18.1.2), с помощью формул параграфов 10.1 и 15.10:

$$(18.1.5) \quad a_v = n(n+2)\dots(n+2v-2), \\ x_v = 2^{v-1}(v-1)!n.$$



Фиг. 19. χ^2 -распределение. Кривые плотности для $n = 1, 2, 6$.

Отсюда, в частности,

$$(18.1.6) \quad E(\chi^2) = a_1 = n, \quad D^2(\chi^2) = a_2 - a_1^2 = 2n.$$

Пусть χ_1^2 и χ_2^2 — две независимые величины, распределенные согласно (18.1.4) со значениями параметра n_1 и n_2 . Тогда выражение (18.1.2) показывает, что характеристическая функция суммы $\chi_1^2 + \chi_2^2$ есть

$$(1 - 2it)^{-\frac{n_1}{2}}(1 - 2it)^{-\frac{n_2}{2}} = (1 - 2it)^{-\frac{n_1+n_2}{2}}.$$

Таким образом, χ^2 -распределение подобно биномиальному распределению, распределению Пуассона и нормальному распределению, воспроизводит само себя при композиции, и мы имеем *теорему сложения*:

$$(18.1.7) \quad K_{n_1}(x) * K_{n_2}(x) = K_{n_1+n_2}(x).$$

Это соотношение можно рассматривать как очевидное следствие определения (18.1.1) величины χ^2 , так как сумма $\chi_1^2 + \chi_2^2$ есть сумма $n_1 + n_2$ независимых квадратов.

Подробные таблицы, относящиеся к χ^2 -распределению, можно найти в различных местах (см., например, [262], [264], [265]).

Во многих приложениях бывает важно найти вероятность P того, что величина χ^2 принимает значение, превышающее данную величину χ_0^2 . Эта вероятность равна площади, ограниченной ветвью кривой плотности, расположенной справа от ординаты, проходящей через точку $x = \chi_0^2$. Таким образом,

$$P = P(\chi^2 > \chi_0^2) = \int_{\chi_0^2}^{\infty} k_n(x) dx = 1 - K_n(\chi_0^2).$$

Обычно более удобно табулировать χ_0^2 как функцию вероятности P . Если P выражается в процентах, скажем, $P = \frac{p}{100}$, то соответствующее $\chi_0^2 = \chi_p^2$ называется *p-процентным значением* χ^2 для n степеней свободы. (См. таблицу III в конце книги). Дадим теперь некоторые простые преобразования χ^2 -распределения, которые часто бывают необходимы в приложениях.

Если каждая из независимых величин x_1, \dots, x_n нормальна $(0, \sigma)$, где $\sigma > 0$ — какая угодно постоянная, то величины $\frac{x_1}{\sigma}, \dots, \frac{x_n}{\sigma}$ независимы и нормальны $(0, 1)$. Таким образом, согласно сказанному выше, плотность вероятности величины $\sum_1^n \left(\frac{x_i}{\sigma}\right)^2$ равна $k_n(x)$. Тогда, по

формуле (15.1.2), получаем функцию плотности величины $\sum_1^n x_i^2$:

$$(18.1.8) \quad \frac{1}{\sigma^2} k_n\left(\frac{x}{\sigma^2}\right) = \frac{1}{2^{n/2} \sigma^n \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2\sigma^2}}, \quad (x > 0).$$

Подобными же простыми преобразованиями находим функцию плотности среднего арифметического $\frac{1}{n} \sum_{v=1}^n x_v^2$, неотрицательного квадратного корня $\sqrt{\sum_{v=1}^n x_v^2}$ и квадратного корня из среднего арифметического

$\sqrt{\frac{1}{n} \sum_{v=1}^n x_v^2}$. Результаты даны в следующей ниже таблице.

Величины x_1, \dots, x_n повсюду предполагаются независимыми и нормальными $(0, \sigma)$. При $x < 0$ все функции плотности равны нулю.

Величина	Плотность вероятности ($x > 0$)
$\sum_{v=1}^n x_v^2$	$\frac{1}{\sigma^2} k_n \left(\frac{x}{\sigma^2} \right) = \frac{1}{2^{n/2} \sigma^n \Gamma \left(\frac{n}{2} \right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2\sigma^2}}$
$\frac{1}{n} \sum_{v=1}^n x_v^2$	$\frac{n}{\sigma^2} k_n \left(\frac{nx}{\sigma^2} \right) = \frac{\left(\frac{n}{2} \right)^{n/2}}{\sigma^n \Gamma \left(\frac{n}{2} \right)} x^{\frac{n}{2}-1} e^{-\frac{nx}{2\sigma^2}}$
$\sqrt{\sum_{v=1}^n x_v^2}$	$\frac{2x}{\sigma^2} k_n \left(\frac{x^2}{\sigma^2} \right) = \frac{2}{2^{n/2} \sigma^n \Gamma \left(\frac{n}{2} \right)} x^{n-1} e^{-\frac{x^2}{2\sigma^2}}$
$\sqrt{\frac{1}{n} \sum_{v=1}^n x_v^2}$	$\frac{2n}{\sigma^2} k_n \left(\frac{nx^2}{\sigma^2} \right) = \frac{2 \left(\frac{n}{2} \right)^{n/2}}{\sigma^n \Gamma \left(\frac{n}{2} \right)} x^{n-1} e^{-\frac{nx^2}{2\sigma^2}}$

Если горизонтальное и вертикальное отклонения u и v снаряда от центра цели независимы и нормальны $(0, \sigma)$, то расстояние $r = \sqrt{u^2 + v^2}$ имеет плотность вероятности

$$\frac{2x}{\sigma^2} k_2 \left(\frac{x^2}{\sigma^2} \right) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}.$$

Если компоненты u , v и w скорости молекулы в прямоугольной системе координат независимы и нормальны $(0, \sigma)$, то скорость $r = \sqrt{u^2 + v^2 + w^2}$ имеет плотность вероятности

$$\frac{2x}{\sigma^2} k_3 \left(\frac{x^2}{\sigma^2} \right) = \sqrt{\frac{2}{\pi}} \cdot \frac{x^2}{\sigma^3} e^{-\frac{x^2}{2\sigma^2}}.$$

18.2. Распределение Стьюдента. Предположим, что $n+1$ случайных величин ξ и ξ_1, \dots, ξ_n независимы и нормальны $(0, \sigma)$. Положим $\eta = \sqrt{\frac{1}{n} \sum_1^n \xi_i^2}$ (берется положительное значение квадратного корня) и рассмотрим величину

$$(18.2.1) \quad t = \frac{\xi}{\eta} = \frac{\xi}{\sqrt{\frac{1}{n} \sum_1^n \xi_i^2}}.$$

Пусть $S_n(x)$ обозначает функцию распределения величины t , так что

$$S_n(x) = P(t \leq x) = P\left(\frac{\xi}{\eta} \leq x\right).$$

По предположению, величины ξ и η независимы; следовательно, согласно (15.11.3), их совместная функция плотности есть произведение функций плотности ξ и η . Величина ξ нормальна $(0, \sigma)$ и η имеет функцию плотности, данную в последней строке таблицы предыдущего параграфа, так что совместная функция плотности есть *)

$$\frac{1}{\sigma} \Phi'\left(\frac{\xi}{\sigma}\right) \cdot \frac{2n\eta}{\sigma^2} k_n\left(\frac{n\eta^2}{\sigma^2}\right) = c_n \eta^{n-1} e^{-\frac{\xi^2 + n\eta^2}{2\sigma^2}},$$

где $\eta > 0$ и

$$c_n = \sqrt{\frac{2}{\pi}} \frac{\left(\frac{n}{2}\right)^{n/2}}{\sigma^{n+1} \Gamma\left(\frac{n}{2}\right)}.$$

Вероятность соотношения $\frac{\xi}{\eta} \leq x$ выражается интегралом от совместной функции плотности по области, определенной неравенствами

*) К правило, мы до сих пор пользовались для обозначения случайной величины и аргумента ее функции распределения или функции плотности соответствующими друг другу буквами греческого и латинского алфавитов и употребляли, например, выражение: „Случайная величина ξ имеет функцию плотности $f(x)$ “. Однако, если рассматриваются одновременно различные величины, то практически иногда удобно отходить от этого правила и использовать в обоих случаях одну и ту же букву. Мы будем, таким образом, по временам употреблять такие выражения, как „случайная величина ξ имеет функцию плотности $f(\xi)$ “ или „случайные величины ξ и η имеют совместную функцию плотности $f(\xi, \eta)$ “.

$\eta > 0$, $\xi < x\eta$:

$$S_n(x) = c_n \iint_{\substack{\xi > 0 \\ \eta > 0 \\ \xi < x\eta}} \eta^{n-1} e^{-\frac{\xi^2 + n\eta^2}{2\sigma^2}} d\xi d\eta.$$

Вводя новые переменные u , v с помощью подстановки

$$(18.2.2) \quad \xi = uv, \quad \eta = v,$$

якобиан которой $\frac{\partial(\xi, \eta)}{\partial(u, v)} = v$, получим

$$\begin{aligned} S_n(x) &= c_n \int_{-\infty}^x du \int_0^\infty v^n e^{-\frac{n+u^2}{2\sigma^2} v^2} dv = \\ (18.2.3) \quad &= 2^{\frac{n-1}{2}} \sigma^{n+1} \Gamma\left(\frac{n+1}{2}\right) c_n \int_{-\infty}^x \frac{du}{(n+u^2)^{\frac{n+1}{2}}} = \\ &= \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^x \frac{du}{\left(1+\frac{u^2}{n}\right)^{\frac{n+1}{2}}}. \end{aligned}$$

Соответствующая функция плотности $s_n(x) = S'_n(x)$ существует для всех значений x и задается выражением

$$(18.2.4) \quad s_n(x) = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

Распределение, задаваемое функцией плотности $s_n(x)$ или функцией распределения $S_n(x)$, известно под названием *распределения Стьюдента* или *t-распределения*. Оно было впервые использовано в одной важной статистической проблеме В. Госсетом, писавшим под псевдонимом „Стьюдент“ (Student) [221]. Как и в случае χ^2 -распределения, параметр n часто называется *числом степеней свободы* нашего распределения (см. параграф 29.2).

Из выражения функции плотности $s_n(x)$ видно, что это распределение не зависит от стандартного отклонения σ основных величин ξ и ξ . Этого, конечно, и следовало ожидать, так как величина t есть однородная функция нулевой степени от основных переменных. Далее мы увидим, что это распределение унимодально и симметрично отно-

сительно $x = 0$. При $v < n$ момент v -го порядка конечен. В частности, среднее значение конечно при $n > 1$, а стандартное отклонение — при $n > 2$. Вследствие симметричности распределения все существующие моменты нечетного порядка равны нулю; несложное вычисление дает

$$D^2(t) = \int_{-\infty}^{\infty} x^2 s_n(x) dx = \frac{n}{n-2},$$

и вообще при $2v < n$

$$\mu_{2v} = a_{2v} = \frac{1 \cdot 3 \dots (2v-1) n^v}{(n-2)(n-4)\dots(n-2v)}.$$

Вероятность того, что величина t отличается по модулю больше чем на заданную величину t_0 от своего среднего значения (равного нулю), так же как и в случае нормального распределения, равна сумме площадей отрезанных от кривой плотности ординатами, проходящими через точки $\pm t_0$. В силу симметрии t -распределения, она равна

$$(18.2.5) \quad P = P(|t| > t_0) = 2 \int_{t_0}^{\infty} s_n(x) dx = 2(1 - S_n(t_0)).$$

Исходя из этого соотношения, можно табулировать отклонение t_0 как функцию вероятности P . Если $P = \frac{p}{100}$, то соответствующее $t_0 = t_p$ называется *p-процентным значением t* для n степеней свободы. Численные значения этой функции даны в таблице IV в конце книги.

Для больших значений n величина t асимптотически нормальна (0,1) в соответствии с соотношениями

$$\lim_{n \rightarrow \infty} S_n(x) = \Phi(x), \quad \lim_{n \rightarrow \infty} s_n(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

которые будут доказаны в параграфе 20.2. Для небольших значений n , однако, t -распределение заметно отличается от предельного нормального распределения, как видно из таблицы IV, где числа, соответствующие предельному случаю, даны в графе $n = \infty$. График распределения Стьюдента для $n = 3$ вместе с приведенной для сравнения нормальной кривой дан на фиг. 20. Из этого графика очевидно, что вероятности больших отклонений от среднего значительно больше при t -распределении, чем при нормальном распределении.

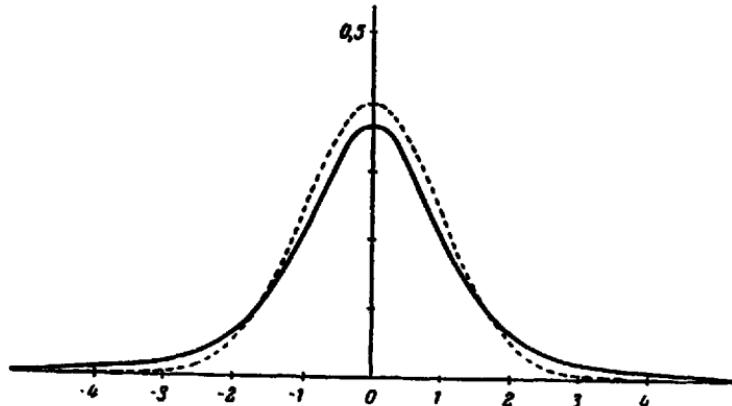
Если вместо величины t , определенной соотношением (18.2.1), рассмотреть величину

$$\tau = \frac{\xi_1}{\eta} = \frac{\xi_1}{\sqrt{\frac{1}{n} \sum_{v=1}^n \xi_v^2}} \quad (n > 1),$$

то числитель и знаменатель не будут независимыми, и распределение такой величины нельзя вывести таким путем, как раньше. Очевидно, что всегда $\tau^2 \leq n$, так что плотность вероятности τ , конечно, равна нулю вне интервала $(-\sqrt{n}, \sqrt{n})$.

Полагая

$$t' = \sqrt{\frac{n-1}{n}} \cdot \frac{\tau}{\sqrt{1 - \frac{\tau^2}{n}}} = \frac{\xi_1}{\sqrt{\frac{1}{n-1} \sum_{v=2}^n \xi_v^2}},$$



Фиг. 20. Кривая плотности распределения Стьюдента при $n = 3$. Пунктиром показана кривая нормальной плотности, $m = 0$, $\sigma = 1$

видим, что t' задается соотношением типа (18.2.1), в котором n заменено на $n-1$. Таким образом, величина t' распределена по закону Стьюдента с функцией распределения $S_{n-1}(x)$. Если τ возрастает от $-\sqrt{n}$ до \sqrt{n} , то, как мы увидим ниже, t' постоянно возрастает от $-\infty$ до $+\infty$. Следовательно, неравенство $\tau < x$ эквивалентно неравенству

$$t' < \sqrt{\frac{n-1}{n}} \cdot \frac{x}{\sqrt{1 - \frac{x^2}{n}}},$$

и мы имеем

$$\begin{aligned} P(\tau < x) &= P\left(t' < \sqrt{\frac{n-1}{n}} \cdot \frac{x}{\sqrt{1 - \frac{x^2}{n}}}\right) = \\ &= s_{n-1}\left(\sqrt{\frac{n-1}{n}} \cdot \frac{x}{\sqrt{1 - \frac{x^2}{n}}}\right). \end{aligned}$$

Таким образом, нами найдена функция распределения величины τ . Дифференцируя по x , получим выражение для плотности вероятности τ

$$(18.2.7) \quad \sqrt{\frac{n-1}{n}} \left(1 - \frac{x^2}{n}\right)^{-\frac{3}{2}} s_{n-1}\left(\sqrt{\frac{n-1}{n}} \cdot \frac{x}{\sqrt{1 - \frac{x^2}{n}}}\right) = \\ = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \left(1 - \frac{x^2}{n}\right)^{\frac{n-3}{2}},$$

где $|x| \leq \sqrt{n}$. При $n=2$ кривая плотности „U-образна“, т. е. имеет минимум в точке среднего значения $x=0$. При $n=3$ функция плотности постоянна, и мы получаем *прямоугольное распределение* (см. параграф 19.1). При $n > 3$ это распределение унимодально и симметрично относительно $x=0$. При всех n среднее значение этого распределения равно нулю, а стандартное отклонение — единице.

18.3. z -распределение Фишера. Предположим, что $m+n$ случайных величин $\xi_1, \dots, \xi_m, \eta_1, \dots, \eta_n$ независимы и нормальны $(0, \sigma)$. Положим

$$\xi = \sum_1^m \xi_v^2, \quad \eta = \sum_1^n \eta_v^2,$$

и рассмотрим величину

$$(18.3.1) \quad x = \frac{\xi}{\eta} = \frac{\sum_1^m \xi_v^2}{\sum_1^n \eta_v^2}.$$

Пусть F_{mn} обозначает функцию распределения величины x . Так как величины ξ и η обе неотрицательны, то $x \geq 0$, и $F_{mn}(x)$ равна нулю при $x < 0$. При $x > 0$ для нахождения $F_{mn}(x)$ можно воспользоваться тем же приемом, что и в предыдущем параграфе. Так как, по предположению, величины ξ и η независимы, то $F_{mn}(x)$ равна ин-

тегралу от произведения функций плотности величин ξ и η по области, определенной неравенствами $\eta > 0$, $0 < \xi < x\eta$. Функции плотности величин ξ и η могут быть взяты из таблицы в параграфе 18.1. Получаем

$$F_{mn}(x) = a_{mn} \iint_{\substack{\eta > 0 \\ 0 < \xi < x\eta}} \xi^{\frac{m}{2}-1} \eta^{\frac{n}{2}-1} e^{-\frac{\xi+\eta}{2\sigma^2}} d\xi d\eta,$$

где

$$a_{mn} = \frac{1}{2^{\frac{m+n}{2}} \sigma^{m+n} \Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)}.$$

Вводя новые переменные u , v с помощью подстановки (18.2.2), находим

$$\begin{aligned} F_{mn}(x) &= a_{mn} \int_0^x u^{\frac{m}{2}-1} du \int_0^\infty v^{\frac{m+n}{2}-1} e^{-\frac{u+v}{2\sigma^2}} dv = \\ &= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \int_0^x \frac{u^{\frac{m}{2}-1}}{(u+1)^{\frac{m+n}{2}}} du. \end{aligned}$$

Дифференцируя, получим функцию плотности $f_{mn}(x) = F'_{mn}(x)$ величины x :

$$(18.3.2) \quad f_{mn}(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \cdot \frac{x^{\frac{m}{2}-1}}{(x+1)^{\frac{m+n}{2}}}, \quad (x > 0).$$

Подобно t -распределению, это распределение не зависит от σ . В частном случае $m = 1$ величина nx имеет то же выражение, что и квадрат величины t , определенной равенством (18.2.1).

В дисперсионном анализе Фишера (см. гл. 36) рассматривается величина z , определенная соотношением

$$(18.3.3) \quad e^{2z} = \frac{n}{m} x = \frac{\frac{1}{m} \sum_{i=1}^m \xi_i^2}{\frac{1}{n} \sum_{i=1}^n \eta_i^2}.$$

Среднее значение и дисперсия величины e^{2x} легко находятся из распределения величины x :

$$(18.3.4) \quad E(e^{2x}) = \frac{n}{m} E(x) = \frac{n}{n-2}, \quad (n > 2),$$

$$D^2(e^{2x}) = \left(\frac{n}{m}\right)^2 D^2(x) = \frac{2n^2(m+n-2)}{m(n-2)^3(n-4)}, \quad (n > 4).$$

При $m > 2$ распределение величины e^{2x} имеет единственную моду в точке $x = \frac{m-2}{m} \cdot \frac{n}{n+2}$.

Чтобы найти распределение самой величины z , заметим, что когда x возрастает от 0 до ∞ , то, как видно из (18.3.3), z монотонно возрастает от $-\infty$ до $+\infty$. Таким образом, неравенство $z < x$ эквивалентно неравенству $x < \frac{m}{n} e^{2x}$ и функция распределения величины z есть

$$P(z < x) = P\left(z < \frac{m}{n} e^{2x}\right) = F_{mn}\left(\frac{m}{n} e^{2x}\right).$$

Дифференцируя по x , получим для плотности вероятности величины z выражение, данное Фишером ([13], [94]):

$$(18.3.5) \quad 2 \frac{m}{n} e^{2x} f_{mn}\left(\frac{m}{n} e^{2x}\right) = 2m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \cdot \frac{e^{mx}}{(me^{2x} + n)^{\frac{m+n}{2}}}.$$

18.4. Бэта-распределение. Употребляя те же обозначения, что и в предыдущем параграфе, рассмотрим величину *)

$$(18.4.1) \quad \lambda = \frac{x}{1+x} = \frac{1}{\sum_1^m \xi_v^2 + \sum_1^n \eta_v^2}.$$

Имеем, очевидно, $0 \leq \lambda \leq 1$, так что плотность вероятности λ равна нулю вне интервала $(0, 1)$. Когда x возрастает от 0 до ∞ , λ монотонно возрастает от 0 до 1. Неравенство $\lambda < x$ эквивалентно, таким образом, неравенству $x < \frac{x}{1-x}$, и функция распределения

*) В частном случае $m = 1$, величина $(n+1)\lambda$ выражается так же, как квадрат величины τ , определенной равенством (18.2.6).

величины λ есть

$$P(\lambda < x) = P\left(x < \frac{x}{1-x}\right) = F_{mn}\left(\frac{x}{1-x}\right).$$

Отсюда мы получаем плотность вероятности λ :

$$(18.4.2) \quad \frac{1}{(1-x)^2} f_{mn}\left(\frac{x}{1-x}\right) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} x^{\frac{m}{2}-1} (1-x)^{\frac{n}{2}-1}$$

Это — частный случай функции плотности $\beta(x; p, q)$, заданной соотношением (12.4.5) при $p = \frac{m}{2}$, $q = \frac{n}{2}$. В общем случае распределение, заданное функцией плотности

$$(18.4.3) \quad \beta(x; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1} (1-x)^{q-1}, \quad (0 < x < 1, p > 0, q > 0),$$

называется *бета-распределением*. Его v -й момент есть

$$(18.4.4) \quad \int_0^1 x^v \beta(x; p, q) dx = \frac{\Gamma(p+v)}{\Gamma(p)} \cdot \frac{\Gamma(p+q)}{\Gamma(p+q+v)}.$$

Следовательно, в частности, среднее значение равно $\frac{p}{p+q}$, а дисперсия равна

$$\frac{pq}{(p+q)^2(p+q+1)}.$$

При $p > 1$, $q > 1$ существует единственная мода в точке $x = \frac{p-1}{p+q-2}$.

ГЛАВА 19

ДРУГИЕ НЕПРЕРЫВНЫЕ РАСПРЕДЕЛЕНИЯ

19.1. Прямоугольное распределение. Будем говорить, что случайная величина ξ имеет *прямоугольное распределение*, если ее функция плотности равна $\frac{1}{2h}$ на некотором конечном интервале $(a-h, a+h)$ и равна нулю вне этого интервала. Тогда кривая плотности изображается прямоугольником с основанием $(a-h, a+h)$ и высотой $\frac{1}{2h}$. Будем также говорить, что в этом случае величина ξ равноз-

мерно распределена по $(a-h, a+h)$. Среднее значение этого распределения равно a , а дисперсия равна $\frac{h^2}{3}$.

Ошибка, происходящая от округления числа, может часто рассматриваться как равномерно распределенная по интервалу $(-\frac{1}{2}, \frac{1}{2})$ в единицах последнего знака.

С помощью линейного преобразования интервал, на котором наша величина равномерно распределена, может быть переведен в любой заданный интервал. Так, например, величина $\eta = \frac{\xi - a + h}{2h}$ равномерно распределена по интервалу $(0, 1)$. Соответствующая функция плотности есть

$$f_1(x) = \begin{cases} 1 & \text{на интервале } (0, 1), \\ 0 & \text{вне интервала } (0, 1). \end{cases}$$

Если независимые величины η_1, η_2, \dots равномерно распределены по $(0, 1)$, то очевидно, что сумма $\eta_1 + \dots + \eta_n$ ограничена интервалом $(0, n)$. Если $f_n(x)$ обозначает плотность вероятности для $\eta_1 + \dots + \eta_n$, то отсюда следует, что $f_n(x)$ равна нулю вне $(0, n)$. Далее, из (15.12.4) следует, что

$$f_{n+1}(x) = \int_{-\infty}^{\infty} f_1(x-t) f_n(t) dt = \int_{x-1}^x f_n(t) dt.$$

С помощью простых вычислений получаем из этих соотношений:

$$f_2(x) = \begin{cases} x & \text{при } 0 < x < 1, \\ x - 2(x-1) & \text{при } 1 < x < 2, \end{cases}$$

$$f_3(x) = \begin{cases} \frac{1}{2}x^2 & \text{при } 0 < x < 1, \\ \frac{1}{2}(x^2 - 3(x-1)^2) & \text{при } 1 < x < 2, \\ \frac{1}{2}(x^2 - 3(x-1)^2 + 3(x-2)^2) & \text{при } 2 < x < 3. \end{cases}$$

С помощью индукции можно проверить следующее общее выражение

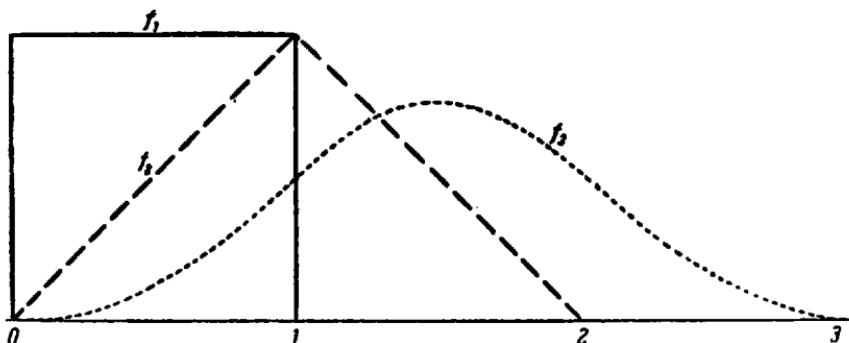
$$f_n(x) = \frac{1}{(n-1)!} \left[x^{n-1} - \binom{n}{1} (x-1)^{n-1} + \binom{n}{2} (x-2)^{n-1} - \dots \right],$$

где $0 < x < n$, а суммирование продолжается до тех пор, пока аргументы $x, x-1, x-2, \dots$ остаются положительными.

Функция f_1 есть разрывная функция плотности, f_2 непрерывна, но имеет разрывную производную, f_3 имеет непрерывную первую производную, но разрывную вторую производную, и т. д. Графики f_1 , f_2 и f_3 показаны на фиг. 21. Среднее значение и стандартное отклонение суммы $\eta_1 + \dots + \eta_n$ равны $\frac{n}{2}$ и $\sqrt{\frac{n}{12}}$, так что плотность вероятности нормированной суммы есть

$$\sqrt{\frac{n}{12}} f_2\left(\frac{n}{2} + x \sqrt{\frac{n}{12}}\right).$$

При возрастании n это выражение быстро приближается к нормальной функции плотности $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.



Фиг. 21. Прямоугольное распределение и родственные ему распределения

Выражение для $f_2(x)$, данное выше, может быть записано в виде

$$f_2(x) = 1 - |1 - x|, \quad (0 < x < 2).$$

Эта функция плотности и любая функция плотности, полученная из нее линейным преобразованием, задает распределение, называемое иногда *треугольным*.

19.2. Распределения Коши и Лапласа. В частном случае $n=1$ распределение Стьюдента (18.2.4) имеет функцию плотности

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

характеристическая функция которой, согласно (10.5.7), есть $e^{-|t|}$. С помощью линейного преобразования, получим функцию плотности

$$(19.2.1) \quad c(x; \lambda, \mu) = \frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + (x - \mu)^2}$$

с характеристической функцией

$$(19.2.2) \quad e^{\mu it - \lambda |t|},$$

где $\lambda > 0$. Распределение с функцией плотности $c(x; \lambda, \mu)$ или с соответствующей функцией распределения $C(x; \lambda, \mu)$ называется *распределением Коши*. Это распределение унимодально и симметрично относительно точки $x = \mu$, являющейся модой и медианой этого распределения. Ни один из его моментов положительного порядка (в том числе даже среднее значение) не существует. Квартили (см. параграф 15.6) равны $\mu \pm \lambda$, так что семи-нинтерквартильная широта равна λ .

Если величина ξ распределена согласно (19.2.1), то любая линейная функция $a\xi + b$ имеет распределение того же типа с параметрами $\lambda' = |a|\lambda$ и $\mu' = a\mu + b$.

Вид характеристической функции (19.2.2) непосредственно показывает, что это распределение воспроизводит само себя при композиции, так что имеет место следующая *теорема сложения*:

$$(19.2.3) \quad C(x; \lambda_1, \mu_1) * C(x; \lambda_2, \mu_2) = C(x; \lambda_1 + \lambda_2, \mu_1 + \mu_2).$$

Отсюда мы выводим следующее интересное свойство распределения Коши: *Если величины ξ_1, \dots, ξ_n независимы и имеют все одно и то же распределение Коши, то среднее арифметическое $\bar{\xi} = \frac{1}{n} \sum_1^n \xi_i$ имеет то же распределение, что и каждое ξ_i .*

Два взаимные интеграла Фурье (10.5.6) и (10.5.7), связывают распределение Коши с *распределением Лапласа*, которое имеет функцию плотности $\frac{1}{2} e^{-|x|}$. Последняя имеет конечные моменты любого порядка, а ее производная разрывна при $x = 0$. С помощью линейного преобразования получим функцию плотности

$$(19.2.4) \quad \frac{1}{2\lambda} e^{-\frac{|x-\mu|}{\lambda}}$$

с характеристической функцией

$$\frac{e^{\mu it}}{1 + \lambda^2 t^2}.$$

19.3. Усеченные распределения. Предположим, что мы имеем дело со случайной величиной ξ , связанной со случайным экспериментом \mathfrak{E} . Пусть, как обычно, P и F обозначают соответственно вероят-

ностную функцию и функцию распределения величины ξ . Из последовательности повторений эксперимента \mathfrak{E} выбираем подпоследовательность, для которой наблюденное значение ξ принадлежит некоторому фиксированному множеству S_0 . Распределение ξ в группе выбранных случаев есть *условное распределение величины ξ при условии $x \in S_0$* . Согласно (14.3.1) или (14.3.2), условная вероятность события $\xi \in S$, где S есть подмножество множества S_0 , может быть записана в виде

$$P(\xi \in S | \xi \in S_0) = \frac{P(\xi \in S)}{P(\xi \in S_0)}.$$

Случай, когда S_0 есть интервал $a < \xi \leq b$, часто встречается в приложениях; при этом мы отбрасываем все наблюдения, при которых были зарегистрированы значения величины $\xi \leq a$ или $\geq b$. Остающиеся наблюдения дают *усеченное распределение* с функцией распределения

$$F(x | a < \xi \leq b) = \begin{cases} 0 & \text{при } x \leq a, \\ \frac{F(x) - F(a)}{F(b) - F(a)} & \text{при } a < x \leq b, \\ 1 & \text{при } x > b. \end{cases}$$

Если существует плотность вероятности $f(x) = F'(x)$, то усеченное распределение имеет функцию плотности, равную

$$f(x | a < \xi \leq b) = \frac{f(x)}{\int_a^b f(t) dt}$$

для всех x из интервала (a, b) и равную нулю вне этого интервала. Как a , так и b могут быть, разумеется, бесконечны.

1. Усеченное нормальное распределение. Предположим, что рост лиц, являющихся на призывную комиссию, можно рассматривать как случайную величину, нормальную (m, σ). Если голыми к военной службе признаются лишь те лица, рост которых превосходит некоторый фиксированный предел x_0 , то рост отобранных призывников дает *усеченное нормальное распределение* с функцией распределения

$$\frac{\Phi\left(\frac{x-m}{\sigma}\right) - \Phi\left(\frac{x_0-m}{\sigma}\right)}{1 - \Phi\left(\frac{x_0-m}{\sigma}\right)}, \quad (x > x_0).$$

Если положить

$$= \frac{\Phi'\left(\frac{x_0-m}{\sigma}\right)}{1 - \Phi\left(\frac{x_0-m}{\sigma}\right)}$$

то первые два момента усеченного распределения равны

$$a_1 = m + \lambda\sigma, \quad a_2 = m^2 + \lambda\sigma(x_0 + m) + \sigma^2.$$

Если x_0 , a_1 и a_2 даны, а m и σ неизвестны, то для определения m и σ имеем два уравнения. Таблицы, дающие численное решение этих уравнений были даны К. Пирсоном [264].

2. Распределение Парето. Усеченные распределения часто встречаются в различных задачах экономической статистики. Так, например, налоговые органы обычно интересуются распределением годовых доходов тех лиц, годовой доход которых превосходит некоторый предел x_0 , установленный законами о налогообложении. Это и некоторые аналогичные распределения иногда считаются приближенно совпадающими с *распределением Парето*, заданным соотношением

$$P(\xi > x) = \left(\frac{x_0}{x}\right)^{\alpha}, \quad (x > x_0, \quad \alpha > 0).$$

Функция плотности этого распределения равна $\frac{a}{x_0} \left(\frac{x_0}{x}\right)^{\alpha+1}$ при $x > x_0$ и нулю при $x \leq x_0$. При $\alpha > 1$ среднее значение конечно и равно $\frac{a}{\alpha-1} x_0$. Медиана распределения есть $2^{\frac{1}{\alpha}} x_0$. По вопросам, связанным с распределением Парето мы отсылаем читателя к работам Хагстрема [121], [122].

19.4. Система Пирсона. Для большинства непрерывных распределений, изложенных в главах 17—19, функция плотности $y = f(x)$ удовлетворяет дифференциальному уравнению вида

$$(19.4.1) \quad y' = \frac{x+a}{b_0+b_1x+b_2x^2} y,$$

где a и b_i суть постоянные. Легко проверить, что это выполняется, например, для нормального распределения, χ^2 -распределения, бета-распределения, распределений Стьюдента, Фишера и Парето. Любое распределение, получаемое из одного из перечисленных посредством линейного преобразования случайной величины, будет, конечно, удовлетворять уравнению того же вида.

Дифференциальное уравнение (19.4.1) определяет систему кривых плотности, введенную К. Пирсоном (см. [180], [181], [184] и т. д.). Можно показать, что постоянные уравнения (19.4.1) могут быть выражены через первые четыре момента распределения, если только эти моменты конечны. Решения классифицируются по характеру корней уравнения $b_0 + b_1x + b_2x^2 = 0$. Этим способом получается большое разнообразие возможных типов кривых плотности. Знание первых четырех

моментов для любой функции плотности, принадлежащей этой системе, достаточно для полного определения функции. Полное описание пирсоновских типов было дано Элдертоном [12], к книге которого мы и отсылаем читателя. Здесь мы только упомянем некоторые из наиболее интересных типов. Постоянный множитель A , встречающийся во всех приведенных ниже уравнениях, может быть в каждом случае определен так, чтобы интеграл правой части относительно x по указанной области был равен единице.

Тип I. $y = A(x - a)^{p-1}(b - x)^{q-1}; a < x < b; p > 0, q > 0.$

При $a = 0, b = 1$, получим бета-распределение (18.4.3) как частный случай распределения типа I. Полагая $p = q = \frac{1}{2}b^2, a = -b$ и заставляя b стремиться к бесконечности, получим нормальное распределение в качестве предельного случая. Другой предельный случай получаем, полагая $q = ba$; при $b \rightarrow \infty$ получим после изменения обозначений следующий

Тип III. $y = A(x - \mu)^{\lambda-1}e^{-\alpha(x-\mu)}; x > \mu; \alpha > 0, \lambda > 0.$

Это — обобщение функции плотности $f(x; \alpha, \lambda)$, определенной равенством (12.3.3), и α *fortiori*, обобщение χ^2 -распределения (18.1.3).

Тип VI. $y = A(x - a)^{p-1}(x - b)^{q-1}; x > b; a < b, q > 0, p + q < 1.$

Этот тип содержит в качестве частного случая (при $a = -1, b = 0$) распределение (18.3.2).

Тип VII. $y = \frac{A}{((x - a)^2 + \beta^2)^m}; -\infty < x < \infty; m > \frac{1}{2}.$

Этот тип содержит как частный случай распределение Стьюдента.

ГЛАВА 20

НЕКОТОРЫЕ ТЕОРЕМЫ О СХОДИМОСТИ

20.1. Сходимость распределений и случайных величин. Если дана последовательность случайных величин ξ_1, ξ_2, \dots с функциями распределения $F_1(x), F_2(x), \dots$, то часто бывает важно знать, сходится ли (см. параграф 6.7) эта последовательность функций распределения к некоторой предельной функции распределения $F(x)$. Так,

например, центральная предельная теорема утверждает, что некоторая последовательность функций распределения сходится к нормальной функции распределения $\Phi(x)$. В следующем параграфе мы дадим другие важные примеры сходимости к нормальному распределению.

Важно заметить, что всякое утверждение относительно сходимости *последовательности функций распределения* $\{F_n(x)\}$ необходимо отличать от утверждения относительно сходимости *последовательности случайных величин* $\{\xi_n\}$. В этой книге мы не будем проводить полное исследование вопроса о сходимости последовательностей случайных величин. По этим вопросам мы можем отослать читателя к книгам Фреше [15] и Леви [25]. Здесь мы будем пользоваться только понятием *сходимости по вероятности*, которое будет рассмотрено в параграфах 20.3—20.6.

20.2. Сходимость некоторых распределений к нормальному.

1. Распределение Пуассона. Согласно параграфу 16.5, величина ξ , имеющая распределение Пуассона, имеет среднее значение λ , стандартное отклонение $\sqrt{\lambda}$ и характеристическую функцию $e^{\lambda(e^{it}-1)}$. Нормированная величина $\frac{\xi - \lambda}{\sqrt{\lambda}}$ имеет, таким образом, характеристическую функцию

$$e^{-it\sqrt{\lambda} + \lambda(e^{\frac{it}{\sqrt{\lambda}}} - 1)} = e^{-\frac{t^2}{2} + \frac{(it)^2}{3! \sqrt{\lambda}} + \dots}$$

При $\lambda \rightarrow \infty$ это выражение стремится к $e^{-\frac{t^2}{2}}$ и согласно теореме непрерывности (см. параграф 10.4), соответствующая функция распределения стремится к $\Phi(x)$. Таким образом, величина ξ асимптотически нормальна ($\lambda, \sqrt{\lambda}$).

2. χ^2 -распределение. При n степенях свободы величина χ^2 имеет, согласно (18.1.6) и (18.1.2), среднее значение n , стандартное отклонение $\sqrt{2n}$ и характеристическую функцию $(1 - 2it)^{-n/2}$. Таким образом, нормированная величина $\frac{\chi^2 - n}{\sqrt{2n}}$ имеет характеристическую функцию

$$e^{-it\sqrt{n/2}} \left(1 - it\sqrt{\frac{2}{n}}\right)^{-n/2}.$$

При любом фиксированном t можно выбрать n настолько большим, чтобы это выражение могло быть записано в виде

$$\left(1 + \frac{t^2}{n} + o\left(\frac{2}{n}\right)^{\frac{n}{2}} |t|^3\right)^{-n/2},$$

где $|o| \leq 1$.

При $n \rightarrow \infty$ это выражение, очевидно, стремится к $e^{-\frac{x^2}{2}}$; следовательно, функция распределения величины $\frac{\chi^2 - n}{\sqrt{2n}}$ стремится к $\Phi(x)$, так что величина χ^2 асимптотически нормальна $(n, \sqrt{2n})$.

Рассмотрим теперь вероятность неравенства $\sqrt{2\chi^2} < \sqrt{2n} + x$, которое может быть записано также в виде

$$\chi^2 < n + \left(x + \frac{x^2}{2\sqrt{2n}}\right) \sqrt{2n}.$$

Когда x фиксировано, а n стремится к ∞ , $\frac{x^2}{2\sqrt{2n}}$ стремится к нулю, так что вероятность последнего неравенства стремится к тому же пределу, что и вероятность неравенства $\chi^2 < n + x\sqrt{2n}$, т. е. к $\Phi(x)$. Таким образом, величина $\sqrt{2\chi^2}$ асимптотически нормальна $(\sqrt{2n}, 1)$. Согласно Р. Фишеру [13], эта аппроксимация будет улучшена, если заменить $2n$ на $2n-1$ и рассматривать $\sqrt{2\chi^2}$ как нормально распределенную величину со средним значением $\sqrt{2n}-1$ и стандартным отклонением, равным единице. При $n \geq 30$ это дает аппроксимацию, часто достаточную для практических целей.

3. Распределение Стьюдента. Плотность вероятности (18.2.4) распределения Стьюдента может быть записана в виде

$$(20.2.1) \quad s_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \cdot \frac{1}{\sqrt{2n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Согласно формуле Стирлинга (12.5.3), первый сомножитель стремится к единице при $n \rightarrow \infty$, и при любом фиксированном x имеем

$$-\frac{n+1}{2} \log\left(1 + \frac{x^2}{n}\right) \rightarrow -\frac{x^2}{2}.$$

откуда

$$(20.2.2) \quad s_n(x) \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Далее, пусть r обозначает наибольшее целое число, содержащееся в $\frac{n+1}{2}$. Тогда $r \geq \frac{n}{2}$ и при всех $n \geq 1$ мы имеем для всех действительных x

$$\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}} \geq \left(1 + \frac{x^2}{n}\right)^r \geq 1 + r \frac{x^2}{n} \geq 1 + \frac{x^2}{2}.$$

Таким образом, последовательность $\{s_n(x)\}$ равномерно ограничена функцией вида $A \left(1 + \frac{1}{2}x^2\right)^{-1}$, так что (5.5.2) дает

$$(20.2.3) \quad S_n(x) = \int_{-\infty}^x s_n(t) dt \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \Phi(x).$$

4. Бэта-распределение. Пусть ξ — случайная величина, имеющая бэта-распределение (18.4.3) со значениями параметров pq и pq . Среднее значение и дисперсия равны тогда, согласно 18.4,

$\frac{pq}{p+q}$ и $\frac{pq}{(p+q)^2(pq+pq+1)}$. Пусть теперь p и q фиксированы, а n стремится к бесконечности. Тогда с помощью вычислений, подобных проделанных выше, можно доказать, что плотность вероятности нормированной величины стремится к нормальной функции плотности $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ и что соответствующая функция распределения стремится к нормальной функции распределения $\Phi(x)$.

20.3. Сходимость по вероятности. Пусть ξ_1, ξ_2, \dots — последовательность случайных величин и пусть $F_n(x)$ и $\varphi_n(t)$ обозначают функцию распределения и характеристическую функцию величины ξ_n . Будем говорить (см. Кантелли [64], Слуцкий [214], Фреше [112]), что ξ_n сходится по вероятности к постоянной c , если при любом $\varepsilon > 0$ вероятность соотношения $|\xi_n - c| > 0$ стремится к нулю при $n \rightarrow \infty$.

Так, если ξ обозначает частоту $\frac{y}{n}$ события E при n повторениях случайного эксперимента E , то теорема Бернулли (см. параграф 16.3) утверждает, что $\frac{y}{n}$ сходится по вероятности к p .

Очевидно, для того чтобы последовательность ξ_n сходилась по вероятности к c , необходимо и достаточно, чтобы функция распределения $F_n(x)$ стремилась при каждом фиксированном $x \neq c$ к функции распределения $\varepsilon(x - c)$, определенной в параграфе 16.1.

Согласно теореме непрерывности (см. параграф 10.4) это условие эквивалентно тому, что характеристическая функция $\varphi_n(t)$ стремится при любом фиксированном t к пределу e^{cit} .

20.4. Теорема Чебышева. Докажем следующую теорему, принадлежащую в основном Чебышеву.

Пусть ξ_1, ξ_2, \dots — случайные величины и пусть m_n и σ_n обозначают среднее значение и стандартное отклонение величины ξ_n . Если $\sigma_n \rightarrow 0$ при $n \rightarrow \infty$, то $\xi_n - m_n$ сходится по вероятности к нулю.

Чтобы доказать эту теорему, достаточно применить неравенство Бьенэме — Чебышева (15.7.2) к величине $\xi_n - m_n$. Тогда видно, что вероятность соотношения $|\xi_n - m_n| > \varepsilon$ не превосходит $\frac{\sigma_n^2}{\varepsilon^2}$ и, следовательно, по предположению, стремится к нулю при $n \rightarrow \infty$.

Предположим теперь, что величины ξ_1, ξ_2, \dots независимы, и положим

$$\bar{\xi} = \frac{1}{n} \sum_{v=1}^n \xi_v, \quad \bar{m} = \frac{1}{n} \sum_{v=1}^n m_v.$$

Тогда имеем следующее следствие нашей теоремы:

Если

$$(20.4.1) \quad \sum_{v=1}^n \sigma_v^2 = o(n^2),$$

то $\bar{\xi} - \bar{m}$ сходится по вероятности к нулю.

Действительно, величина $\bar{\xi}$ имеет среднее значение \bar{m} и стандартное отклонение $\frac{1}{n} \sqrt{\sum_{v=1}^n \sigma_v^2}$. По предположению, последнее выражение стремится к нулю при $n \rightarrow \infty$, и, следовательно, наше утверждение следует из теоремы Чебышева.

В частном случае, когда ξ_n суть переменные, рассмотренные в параграфе 16.6 в связи с рядом независимых испытаний, то σ_n ограничены, и, значит, условие (20.4.1) выполнено. Тогда наше следствие сводится к обобщению Пуассона теоремы Бернулли.

20.5. Теорема Хинчина. Если даже не предполагать существования конечных стандартных отклонений у величин ξ_n , рассмотренных в предыдущем параграфе, то все же можно получить некоторый результат, аналогичный следствию из теоремы Чебышева. Рассмотрим только случай, когда все ξ_n имеют одинаковое распределение вероятностей, и докажем следующую теорему, принадлежащую Хинчину [139]:

Пусть независимые случайные величины ξ_1, ξ_2, \dots имеют одну и ту же функцию распределения $F(x)$ и пусть $F(x)$ имеет конечное среднее значение m . Тогда величина $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$ сходится по вероятности к m .

Если $\varphi(t)$ — характеристическая функция распределения величины ξ_n , то характеристической функцией величины $\bar{\xi}$ служит $(\varphi(\frac{t}{n}))^n$. Согласно (10.1.3), при $t \rightarrow 0$ имеем

$$\varphi(t) = 1 + mit + o(t),$$

и, следовательно, для любого фиксированного t при $n \rightarrow \infty$

$$(\varphi(\frac{t}{n}))^n = (1 + \frac{mit}{n} + o(\frac{1}{n}))^n \rightarrow e^{mit}.$$

В силу параграфа 20.3, теорема этим доказана.

20.6. Теорема о сходимости. В различных приложениях бывает полезна следующая теорема:

Пусть ξ_1, ξ_2, \dots — последовательность случайных величин с функциями распределения F_1, F_2, \dots . Предположим, что $F_n(x)$ стремится к функции распределения $F(x)$ при $n \rightarrow \infty$.

Пусть, далее, η_1, η_2, \dots — другая последовательность случайных величин. предположим, что η_n сходится по вероятности к некоторой постоянной c . Положим

$$(20.6.1) \quad X_n = \xi_n + \eta_n, \quad Y_n = \xi_n \eta_n, \quad Z_n = \frac{\xi_n}{\eta_n}.$$

Тогда функция распределения величины X_n стремится к $F(x-c)$. Далее, если $c > 0$, то функция распределения величины Y_n стремится к $F(\frac{x}{c})$, а функция распределения величины Z_n к $F(cx)$. (Изменения в формулировке, необходимые в случае $c < 0$, очевидны.)

Важно отметить, что в этой теореме не требуется независимости входящих в нее случайных величин.

Достаточно доказать какое-нибудь одно из утверждений теоремы, другие доказательства проводятся аналогично. Возьмем, например, величины Z_n . Пусть x — точка непрерывности функции $F(x)$. Обозначим через P_n совместную вероятностную функцию величин ξ_n и η_n . Нужно доказать, что

$$P_n \left(\frac{\xi_n}{\eta_n} \leqslant x \right) \rightarrow F(cx)$$

при $n \rightarrow \infty$. Множество S всех точек плоскости (ξ_n, η_n) таких, что $\frac{\xi_n}{\eta_n} \leqslant x$, есть сумма непересекающихся множеств S_1 и S_2 , определенных неравенствами:

$$\begin{aligned} S_1 : \quad & \frac{\xi_n}{\eta_n} \leqslant x, \quad |\eta_n - c| \leqslant \varepsilon, \\ S_2 : \quad & \frac{\xi_n}{\eta_n} \leqslant x, \quad |\eta_n - c| > \varepsilon. \end{aligned}$$

Таким образом, $P_n(S) = P_n(S_1) + P_n(S_2)$. S_2 есть подмножество множества $|\eta_n - c| > \varepsilon$, и, по предположению, $P_n(S_2) \rightarrow 0$ при любом $\varepsilon > 0$.

Далее, $P_n(S_1)$ заключена между пределами

$$P_n(\xi_n \leqslant (c - \varepsilon)x, |\eta_n - c| \leqslant \varepsilon) \text{ и } P_n(\xi_n \leqslant (c + \varepsilon)x, |\eta_n - c| \leqslant \varepsilon).$$

Каждый из этих пределов отличается от соответствующей величины

$$P_n(\xi_n \leqslant (c \pm \varepsilon)x) = F_n((c \pm \varepsilon)x)$$

меньше, чем на $P_n(|\eta_n - c| > \varepsilon)$. При $n \rightarrow \infty$ последняя величина стремится к нулю, и мы, таким образом, видим, что $P_n(S)$ заключена между двумя пределами, которые выбором достаточно малого ε можно сделать сколь угодно близкими к $F(x)$. Тем самым наша теорема доказана.

Отсюда мы выводим следующее утверждение, принадлежащее Слуцкому [214]:

Если случайные величины $\xi_n, \eta_n, \dots, \rho_n$ сходятся по вероятности соответственно к постоянным x, y, \dots, r , то любая рациональная функция $R(\xi_n, \eta_n, \dots, \rho_n)$ сходится по вероятности к постоянной $R(x, y, \dots, r)$, если только $R(x, y, \dots, r)$ конечно. Отсюда, в частности, следует, что любая степень $R^k(\xi_n, \eta_n, \dots, \rho_n)$ при $k > 0$ сходится по вероятности к $R^k(x, y, \dots, r)$.

Упражнения к главам 15—20

1. Случайная величина ξ имеет плотность вероятности $f(x)$. Найти плотность вероятностей величин $\eta = \frac{1}{\xi}$ и $\zeta = \cos \xi$. Дать условия существования моментов величин η и ζ .

2. При любом $k > 1$ функция $f(x) = \frac{k}{2(1+|x|)^{k+1}}$ есть плотность вероятности с широтой $(-\infty, \infty)$. Показать, что n -й момент существует тогда и только тогда, когда $n < k$.

3. Неравенство (15.4.6) для абсолютных моментов β_n есть частный случай следующего неравенства, принадлежащего Ляпунову [147]: для любых неотрицательных n, p, q (не обязательно целых)

$$\log \beta_{n+p} \leq \frac{q}{p+q} \log \beta_n + \frac{p}{p+q} \log \beta_{n+p+q}.$$

При $n=0, q=1$ это сводится к (15.4.6), так как $\beta_0=1$. Общее неравенство выражает, что хорда, соединяющая две точки кривой $y = \log \beta_x (x > 0)$, лежит целиком выше кривой, так что $\log \beta_x$ есть выпуклая функция x . (Подробное доказательство см., например, в книге Успенского [39], стр. 265.)

4. Если $g(x)$ не возрастает при $x > 0$, то при любом $k > 0$ имеем

$$k^2 \int_k^\infty g(x) dx \leq \frac{4}{9} \int_0^\infty x^2 g(x) dx.$$

Сперва докажите, что это неравенство выполнено в частном случае, когда $g(x)$ постоянна при $0 < x < c$ и равна нулю при $x > c$. Затем определите функцию $h(x)$, которая равна $g(k)$ при $0 < x < k+a$ и равна нулю при $x > k+a$, где a определяется соотношением $ag(k) = \int_k^\infty g(x) dx$, и покажите, что

$$k^2 \int_k^\infty g(x) dx = k^2 \int_k^\infty h(x) dx \leq \frac{4}{9} \int_0^\infty x^2 h(x) dx \leq \frac{1}{9} \int_0^\infty x^3 g(x) dx.$$

Используйте этот результат для доказательства неравенств (15.7.3) и (15.7.4).

5. Если $F(x)$ есть функция распределения со средним значением 0 и стандартным отклонением σ , то $F(x) \leq \frac{\sigma^2}{\sigma^2 + x^2}$ при $x < 0$ и $F(x) \geq \frac{x^2}{\sigma^2 + x^2}$

при $x > 0$. Для $x < 0$ это следует из неравенств

$$-x = \int_{-\infty}^{\infty} (y - x) dF \leq \int_x^{\infty} (y - x) dF,$$

$$x^2 \leq \left(\int_x^{\infty} (y - x) dF \right)^2 \leq \int_x^{\infty} dF \cdot \int_x^{\infty} (y - x)^2 dF \leq (1 - F(x)) (x^2 + x^2).$$

Для $x > 0$ доказательство аналогично. Показать на примере, что эти неравенства не могут быть улучшены.

6. Неравенство Бененэме — Чебышева (15.7.2) может быть улучшено, если известен какой-нибудь центральный момент μ_{2n} ($n > 1$). Имеем, например, для $k > 1$

$$P(|z - m| \geq k) \leq \frac{\mu_4 - \sigma^4}{\mu_4 + k^4 \sigma^4 - 2k^2 \sigma^4} = \frac{\gamma_2 + 2}{(k^2 - 1)^2 + \gamma_2 + 2}.$$

Применить (15.7.1) с $K = 1$ и $g(z) = 1 + \frac{\sigma^2(k^2 - 1)((z - m)^2 - k^2 \sigma^2)}{\mu_4 + k^4 \sigma^4 - 2k^2 \sigma^4}$.

7. Покажите, используя (15.4.6), что семи-инвариант x_n произвольного распределения удовлетворяет неравенству $|x_n| \leq n \sigma^2$. (Крамер [11], стр. 39).

8. Докажите неравенство $|a + b|^n \leq 2^{n-1}(|a|^n + |b|^n)$. Выведите отсюда, что если n -е моменты x и y существуют, то существует и n -й момент суммы $x + y$.

9. Пусть $G(p, q) = \sum_{r > np} \binom{n}{r} p^r q^{n-r}$. Покажите, что первый абсолютный момент этого биномиального распределения относительно среднего есть

$$E(|v - np|) = 2pq \left(\frac{\partial G}{\partial p} - \frac{\partial G}{\partial q} \right) = 2p \binom{n}{v} p^v q^{n-v+1},$$

где v есть наименьшее целое число $> np$. При больших n отсюда следует, что

$$E(|v - np|) \approx \sqrt{\frac{2npq}{\pi}}.$$

10. Покажите, что если $1 - F(x) = O(e^{-cx})$ при $x \rightarrow +\infty$ и $F(x) = O(e^{-c|x|})$ при $x \rightarrow -\infty$ ($c > 0$), то распределение однозначно определяется своими моментами.

11. Факториальными моментами (Стефенсен [217]) какого-нибудь дискретного распределения называются величины $a_{[v]} = \sum_r p_r x_r^{[v]}$, где $x^{[v]}$ обозначает выражение $x(x-1)\dots(x-v+1)$. Аналогично центральные факториальные моменты суть $\mu_{[v]} = \sum_r p_r (x_r - m)^{[v]}$. Выразите $a_{[v]}$ и $\mu_{[v]}$ через обычные моменты. Покажите, что $(x+y)^{[v]} = x^{[v]} + \binom{v}{1} x^{[v-1]} y^{[1]} + \dots + y^{[v]}$ и выведите отсюда соотношения между $a_{[v]}$ и $\mu_{[v]}$.

12. Характеристическая функция распределения в предыдущем упражнении есть $\varphi(t) = \sum_r p_r e^{itx_r}$. Заменяя здесь e^{it} на t , получим производящую функцию $\phi(t) = \sum_r p_r t x_r$. Покажите, что $\phi^{(v)}(1) = a_{[v]}$ и, в частности, $E(x) = \psi'(1)$, $D^2(x) = \psi''(1) + \psi'(1) - (\psi'(1))^2$. Используя этот результат, выведите выражения $a_{[v]} = n^{[v]} p^v$ для биномиального распределения и $a_{[v]} = \lambda^v$ для распределения Пуассона.

13. а) Производим ряд независимых испытаний, при каждом из которых вероятность благоприятного исхода равна $p = 1 - q$ до тех пор, пока не получим подряд v благоприятных исходов, где $v > 0$ задано. Пусть p_{nv} обозначает вероятность того, что для этой цели необходимо ровно n испытаний. Найдите производящую функцию

$$\Phi(t) = \sum_{n=1}^{\infty} p_{nv} t^n = \frac{p^v t^v (1 - pt)}{1 - t + p^v q t^{v+1}}$$

и покажите, что

$$E(n) = \Phi'(1) = \frac{1 - p^v}{p^v q}.$$

б) С другой стороны, предположим, что мы производим n испытаний, где n задано и наблюдаем наибольшее число μ произошедших подряд благоприятных исходов при этих n испытаниях. Обозначая через P_{nv} вероятность того, что $\mu \leq v$, покажите, что

$$P_{nv} = 1 - p_{1v} - \dots - p_{nv},$$

и, следовательно,

$$\Psi(t) = \sum_{n=1}^{\infty} P_{nv} t^n = \frac{1 - \Phi(t)}{1 - t} = \frac{1 - p^v t^v}{1 - t + p^v q t^{v+1}}.$$

Отсюда можно показать (Крамер [68]), что $P_{nv} = e^{-\mu p^v q}$ равномерно стремится к нулю при $n \rightarrow \infty$ для $1 \leq v \leq n$. Отсюда следует, что для больших n

$$E(\mu) = \frac{\log n}{\log \frac{1}{p}} + O(1), \quad D^2(\mu) = O(1).$$

14. Величина ξ нормальна (m, σ). Покажите, что среднее отклонение есть

$$E(|\xi - m|) = \sigma \sqrt{\frac{2}{\pi}} = 0,79788 \sigma.$$

15. В обоих случаях центральной предельной теоремы, доказанной в параграфе 17.4, имеем $E\left|\frac{\xi - m}{\sigma}\right| \rightarrow \sqrt{\frac{2}{\pi}}$ при $n \rightarrow \infty$. Использовать (7.5.9) и (9.5.1). (См. упражнение 9.)

16. Пусть независимые величины ξ_1, ξ_2, \dots таковы, что ξ_v может принимать значения 0 и $\pm u$ соответственно с вероятностями $1 - v^{-2x}$, $\frac{1}{2} v^{-2x}$.

Таким образом, ξ_v имеет среднее значение, равное 0 , и стандартное отклонение, равное 1 . Покажите, что условие Ляпунова (17.4.3) выполнено при $x < \frac{1}{2}$ и не выполнено при $x \geq \frac{1}{2}$. Таким образом, при $x < \frac{1}{2}$ сумма $\xi =$

$= \sum_1^n \xi_v$ асимптотически нормальна $(0, \sqrt{n})$. При $x > \frac{1}{2}$ вероятность того,

что $\xi_1 = \dots = \xi_n = 0$ не стремится к нулю при $n \rightarrow \infty$, так что в этом случае распределение ξ не стремится к нормальному. Последний результат имеет место и при $a = \frac{1}{2}$ (см. Крамер [11], стр. 79).

17. Если a_1 и a_2 — первые два момента логарифмически-нормального распределения (17.5.3) и если η — действительный корень уравнения $\eta^3 + 3\eta - \gamma_1 = 0$, где γ_1 есть коэффициент асимметрии, то параметры a , m и s этого распределения задаются соотношениями

$$a = a_1 - \frac{\sqrt{a_2 - a_1^2}}{\eta}, \quad s^2 = \log(1 + \eta^2),$$

$$m = \log(a_1 - a) - \frac{1}{2}s^2.$$

18. Рассмотрим разложение (17.6.3) функции плотности $f(x)$ в ряд Грама — Шарлье и возьмем $f(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{x^2}{2s^2}}$. При $x = 0$ имеем $f(0) = \frac{1}{s\sqrt{2\pi}}$, и разложение принимает вид

$$\frac{1}{s\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}} \sum_{v=0}^{\infty} \frac{(2v)!}{2^{2v} (v!)^2} (1 - s^2)^v.$$

Однако это разложение верно только при $s^2 \leq 2$. При $s^2 > 2$ этот ряд расходится. Найдите a и β , такие, чтобы $af(x) + \beta f(s^2x)$ было плотностью вероятности некоторой нормированной случайной величины, и покажите на этом примере, что коэффициент $\frac{1}{4}$ в условии сходимости (17.6.6а) не может быть заменен меньшим числом.

19. Вычислите коэффициенты γ_1 и γ_2 распределений, рассмотренных в главе 18.

20. Если величина η равномерно распределена на интервале $(a - h, a + h)$, то характеристическая функция величины η есть $\frac{\sin ht}{ht} e^{alt}$. Если ξ — произвольная случайная величина, независимая от η , с характеристической функцией $\varphi(t)$, то сумма $\xi + \eta$ имеет характеристическую функцию $\frac{\sin ht}{ht} e^{alt} \varphi(t)$. Покажите, что с помощью этого результата формула (10.3.3) может быть выведена непосредственно из (10.3.1).

21. Пусть n есть случайная величина, имеющая распределение Пуассона с вероятностями $\frac{x^n}{n!} e^{-x}$, где $n = 0, 1, \dots$. Если рассматривать параметр x как случайную величину с плотностью вероятности $\frac{a^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-ax}$, ($x > 0$),

то вероятность того, что λ принимает данное значение v , равна

$$\int_0^{\infty} \frac{x^v}{v!} e^{-x} \frac{a^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-ax} dx = \left(\frac{a}{1+a} \right)^\lambda \left(\frac{-1}{v} \right) \frac{(-1)^v}{(1+a)^v}.$$

Найдите характеристическую функцию, среднее значение и стандартное отклонение этого распределения, которое называется *отрицательно-биномиальным распределением*.

22. Независимые величины x_1, x_2, \dots имеют одинаковое распределение со средним значением 0 и стандартным отклонением 1. Используя теоремы параграфов 20.5 и 20.6, покажите, что величины

$$y = \sqrt{n} \frac{x_1 + \dots + x_n}{x_1^2 + \dots + x_n^2} \text{ и } z = \frac{x_1 + \dots + x_n}{\sqrt{x_1^2 + \dots + x_n^2}}$$

асимптотически нормальны $(0,1)$.

23. Если x_n и y_n асимптотически нормальны, соответственно $\left(a, \frac{h}{\sqrt{n}} \right)$ и $\left(b, \frac{k}{\sqrt{n}} \right)$, где $b \neq 0$, то величина $z_n = \sqrt{n} \frac{z_n - a}{y_n}$ асимптотически нормальна $\left(0, \frac{h}{b} \right)$. Заметьте, что в этом случае не фигурирует условие независимости.

ГЛАВА 21

СЛУЧАЙ ДВУХ ИЗМЕРЕНИЙ

21.1. Два простых типа распределений. Рассмотрим две одномерные случайные величины ξ и η . Совместное распределение вероятностей (см. параграф 14.2) величин ξ и η есть распределение в R_2 , или двумерное распределение. Этот случай будет разобран в настоящем параграфе, прежде чем мы перейдем к общему случаю случайных величин и распределений в n -мерном пространстве.

Согласно параграфу 8.4, мы можем определить совместное распределение величин ξ и η как *вероятностной функцией* $P(S)$, представляющей вероятность соотношения $(\xi, \eta) \subset S$, так и *функцией распределения* $F(x, y)$, задаваемой соотношением

$$F(x, y) = P(\xi \leq x, \eta \leq y).$$

Мы будем часто интерпретировать распределение вероятностей посредством распределения единицы массы по плоскости (ξ, η) . Проектированием массы двумерного распределения на одну из координатных осей получим (см. параграф 8.4) *частное распределение* соответствующей величины. Обозначая через $F_1(x)$ функцию распределения частного распределения величины ξ и через $F_2(x)$ соответствующую функцию для η , имеем

$$F_1(x) = P(\xi \leq x) = F(x, \infty),$$

$$F_2(x) = P(\eta \leq y) = F(\infty, y).$$

Как и в одномерном случае (см. параграф 15.2) удобно ввести в рассмотрение два простых типа распределений: *дискретный* и *непрерывный*.

1. Дискретный тип. Говорят, что двумерное распределение принадлежит к дискретному типу, если соответствующие частные распределения принадлежат оба к дискретному типу, определенному в параграфе 15.2. Тогда в каждом частном распределении вся масса заключена в некоторых точках сосредоточения массы, причем любой

конечный интервал содержит не более конечного числа этих точек. Обозначим через x_1, x_2, \dots и y_1, y_2, \dots соответственно точки сосредоточения массы в частных распределениях величин ξ и η . Тогда вся масса двумерного распределения будет сосредоточена в точках пересечения прямых $\xi = x_i$ и $\eta = y_k$, т. е. в точках (x_i, y_k) , где i и k независимо друг от друга принимают значения 1, 2, 3, Если обозначить массу, заключенную в точке (x_i, y_k) , через p_{ik} , то

$$(21.1.1) \quad P(\xi = x_i, \eta = y_k) = p_{ik},$$

а для каждого множества S , не содержащего точек (x_i, y_k) , имеет место равенство $P(S) = 0$. Так как вся масса распределения равна единице, то всегда

$$\sum_{i,k} p_{ik} = 1.$$

Для некоторых комбинаций индексов может оказаться, конечно, что $p_{ik} = 0$. Точки (x_i, y_k) , для которых $p_{ik} > 0$, являются точками сосредоточения массы распределения.

Рассмотрим теперь частное распределение величины ξ , с точками сосредоточения массы x_1, x_2, \dots . Если p_i есть масса, заключенная в точке x_i , то очевидно, что

$$(21.1.2) \quad p_i = P(\xi = x_i) = \sum_k p_{ik}.$$

Аналогично, в частном распределении величины η , в точках y_k заключена масса

$$(21.1.3) \quad p_k = P(\eta = y_k) = \sum_i p_{ik}.$$

Согласно (15.11.2), для независимости величин ξ и η необходимо и достаточно, чтобы для всех i и k

$$(21.1.4) \quad p_{ik} = p_i p_k.$$

2. Непрерывный тип. Говорят, что двумерное распределение принадлежит к непрерывному типу, если функция распределения $F(x, y)$ непрерывна и если плотность вероятности (см. параграф 8.4)

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y}$$

существует и непрерывна всюду, за исключением может быть некоторого множества точек, принадлежащих конечному числу кривых.

Для любого множества S

$$P(S) = \int_S f(x, y) dx dy,$$

и, в частности, при $S = R_2$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1.$$

Частное распределение величины ξ имеет функцию распределения

$$P(\xi \leq x) = \int_{-\infty}^x \int_{-\infty}^{\infty} f(t, u) dt du = \int_{-\infty}^x f_1(t) dt,$$

где

$$(21.1.5) \quad f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

Если в некоторой точке $x = x_0$ функция $f(x, y)$ непрерывна по x для почти всех (см. параграф 5.3) значений y и если в некоторой окрестности x_0 имеет место неравенство $f(x, y) < G(y)$, в котором функция $G(y)$ интегрируема на $(-\infty, \infty)$, то из (7.3.1) следует, что $f_1(x)$ непрерывна в точке $x = x_0$. Во всех случаях, которые будут встречаться в приложениях, эти условия выполнены для всех $x = x_0$ за исключением может быть только конечного числа точек. В таких случаях $f_1(x)$ имеет не больше конечного числа точек разрыва, так что частное распределение величины ξ принадлежит к непрерывному типу и имеет плотность вероятности $f_1(x)$. Аналогично находим, что частное распределение η имеет плотность вероятности

$$(21.1.6) \quad f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Согласно (15.11.3), для независимости величин ξ и η необходимо и достаточно, чтобы для всех x и y выполнялось равенство

$$(21.1.7) \quad f(x, y) = f_1(x)f_2(y).$$

21.2 Средние значения, моменты. Среднее значение функции $g(\xi, \eta)$, интегрируемой на R_2 относительно двумерной вероятностной функции $P(S)$ было определено формулой (15.3.2) как интеграл

$$(21.2.1) \quad E(g(\xi, \eta)) = \int_{R_2} g(x, y) dP(S).$$

Для распределения, принадлежащего к одному из двух простых типов, этот интеграл сводится к сумме или обычному интегралу Римана, как отмечено в параграфе 15.3 для одномерного случая. Основные правила вычисления средних значений уже были выведены в параграфе 15.3 для любого числа измерений.

Моменты распределения (см. 9.2) суть средние значения

$$(21.2.2) \quad a_{ik} = E(\xi^i \eta^k) = \int_{K_2} x^i y^k dP(S),$$

где i и k — неотрицательные целые числа. Сумма $i+k$ индексов есть *порядок момента* a_{ik} .

Моменты $a_{i0} = E(\xi^i)$ и $a_{0k} = E(\eta^k)$ тождественны с моментами одномерных частных распределений величин ξ и η , как это показывает равенство (9.2.2). В частности, полагаем

$$a_{10} = E(\xi) = m_1, \quad a_{01} = E(\eta) = m_2.$$

Точка с координатами $\xi = m_1$, $\eta = m_2$ есть *центр тяжести* распределенной в плоскости массы. Для моментов относительно центра тяжести будем употреблять обозначения

$$(21.2.3) \quad \mu_{ik} = E((\xi - m_1)^i (\eta - m_2)^k).$$

В частности, имеем $\mu_{10} = \mu_{01} = 0$ и $\mu_{20} = \sigma_1^2$, $\mu_{02} = \sigma_2^2$, где σ_1 и σ_2 — стандартные отклонения величин ξ и η .

Между моментами a_{ik} и *центральными моментами* μ_{ik} имеют место соотношения, аналогичные данным в параграфе 15.4 для одномерного случая. Для моментов второго порядка имеем

$$(21.2.4) \quad \mu_{20} = a_{20} - m_1^2, \quad \mu_{11} = a_{11} - m_1 m_2, \quad \mu_{02} = a_{02} - m_2^2.$$

Момент μ_{11} часто называется *смешанным моментом* второго порядка.

В частном случае, когда величины ξ и η независимы, по теореме умножения (15.3.4) имеем $a_{ik} = a_{i0} a_{0k}$ и $\mu_{ik} = \mu_{i0} \mu_{0k}$. Таким образом, в частности, в этом случае

$$\mu_{11} = \mu_{10} \mu_{01} = 0.$$

Для любых действительных t и u имеем

$$(21.2.5) \quad E[(t(\xi - m_1) + u(\eta - m_2))^2] = \mu_{20} t^2 + 2\mu_{11} tu + \mu_{02} u^2.$$

Левая часть этого тождества есть среднее значение квадрата и, следовательно, неотрицательна. Отсюда следует, что правая часть (21.2.5) есть неотрицательная квадратичная форма (см. параграф 11.10)

от t и u ; следовательно, матрица вторых моментов $M = \begin{pmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{pmatrix}$ неотрицательна и

$$(21.2.6) \quad \mu_{20}\mu_{02} - \mu_{11}^2 \geq 0.$$

Ранг r матрицы M (см. параграф 11.6) может принимать одно из значений 0, 1 и 2. При $r=2$ в (21.2.6) имеет место знак $>$, а при $r=1$ и $r=0$ — знак $=$. Покажем теперь, что некоторые простые свойства распределения непосредственно связаны со значением r .

$r=0$ тогда и только тогда, когда вся распределенная масса содержитя в единственной точке.

$r=1$ тогда и только тогда, когда вся масса распределена на некоторой прямой линии, но не в одной ее точке.

$r=2$ тогда и только тогда, когда не существует прямой линии, содержащей всю распределенную массу.

Очевидно, доказательство достаточно провести для случаев $r=0$ и $r=1$, а случай $r=2$ будет являться их следствием. Если $r=0$, то $\mu_{20}=\mu_{02}=0$, так что масса при каждом из частных распределений сосредоточена в одной точке (см. параграф 16.1). Тогда в двумерном распределении вся масса должна быть сосредоточена в центре тяжести m_1, m_2 . Обратно, если известно, что вся распределенная масса содержится в одной точке, то отсюда непосредственно следует, что $\mu_{20}=\mu_{02}=0$ и, следовательно, согласно (21.2.6), $\mu_{11}=0$, так что ранг матрицы M равен нулю.

Далее, если $r=1$, то форма (21.2.5) является полуопределенной (см. параграф 11.10) и, следовательно, принимает значение, равное нулю, при некоторых $t=t_0$ и $u=u_0$, не равных нулю одновременно. Это возможно только, если вся распределенная масса лежит на прямой

$$(21.2.7) \quad t_0(\xi - m_1) + u_0(\eta - m_2) = 0.$$

Обратно, если известно, что вся распределенная масса лежит на прямой линии, но не заключена вся в одной точке, то очевидно, что эта прямая должна проходить через центр тяжести, и, следовательно, должно удовлетворяться соотношение вида (21.2.7). Тогда среднее значение в левой части соотношения (21.2.5) равно нулю при $t=t_0$, $u=u_0$, так что квадратичная форма в правой части этого соотношения является полуопределенной, и, следовательно, матрица M имеет ранг, равный единице. Таким образом, наша теорема доказана.

Пусть теперь мы имеем распределение, при котором оба момента μ_{20} и μ_{02} положительны. (Это означает, что ранг матрицы M равен 1 или 2.) Определим тогда величину ρ , положив

$$(21.2.8) \quad \rho = \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}} = \frac{\mu_{11}}{\sigma_1\sigma_2}.$$

Согласно (21.2.6), $\rho^2 \leq 1$ или $-1 \leq \rho \leq 1$. Далее, $\rho^2 = 1$ тогда и только тогда, когда матрица M имеет ранг 1, т. е. когда вся распределенная масса лежит на некоторой прямой линии. В частном случае, когда величины ξ и η независимы, имеем $\mu_{11} = 0$ и, следовательно, $\rho = 0$.

Величина ρ называется *коэффициентом корреляции* величин ξ и η ; мы будем иметь с ним дело далее, в параграфе 21.7.

Предположим, что заданы величины m_1 , m_2 и μ_{20} , μ_{11} , μ_{02} , такие, что квадратичная форма $\mu_{20}t^2 + 2\mu_{11}tu + \mu_{02}u^2$ неотрицательна. Тогда всегда можно найти распределение, имеющее моменты первого порядка m_1 , m_2 и центральные моменты второго порядка μ_{20} , μ_{11} , μ_{02} . Поставленным условиям удовлетворяет, например, дискретное распределение, при котором масса $\frac{1+\rho}{4}$ помещена в каждой из двух точек $(m_1 + \sigma_1, m_2 + \sigma_2)$ и $(m_1 - \sigma_1, m_2 - \sigma_2)$ а масса $\frac{1-\rho}{4}$ — в каждой из двух точек $(m_1 + \sigma_1, m_2 - \sigma_2)$ и $(m_1 - \sigma_1, m_2 + \sigma_2)$. Величины σ_1 , σ_2 и ρ определяются, конечно, приведенными выше соотношениями.

21.3. Характеристические функции. Среднее значение

$$(21.3.1) \quad \varphi(t, u) = E(e^{i(t\xi+u\eta)}) = \int_{R^2} e^{i(t\xi+u\eta)} dP$$

есть *характеристическая функция* двумерной случайной величины (ξ, η) или соответствующего распределения. Мы будем также часто называть $\varphi(t, u)$ *совместной* характеристической функцией двух одномерных величин ξ и η .

Согласно теории характеристических функций, развитой в главе 10, взаимнооднозначное соответствие между одномерными распределениями и их характеристическими функциями (см. параграф 15.9) распространяется на распределения произвольного числа измерений. Если два распределения совпадают, то их характеристические функции также совпадают, и обратно.

Если моменты второго порядка совместного распределения величин ξ и η конечны, то в окрестности точки $t = u = 0$ имеет место

разложение, аналогичное (10.1.3):

$$(21.3.2) \quad \varphi(t, u) = 1 + \frac{i}{1!} (\alpha_{10}t + \alpha_{01}u) + \\ + \frac{i^2}{2!} (\alpha_{20}t^2 + 2\alpha_{11}tu + \alpha_{02}u^2) + o(t^2 + u^2) = \\ = e^{i(m_1 t + m_2 u)} \left[1 + \frac{i^2}{2!} (\mu_{20}t^2 + 2\mu_{11}tu + \mu_{02}u^2) + o(t^2 + u^2) \right].$$

В важном частном случае, когда средние значения m_1 и m_2 оба равны нулю, будем иметь:

$$(21.3.3) \quad \varphi(t, u) = 1 - \frac{1}{2} (\mu_{20}t^2 + 2\mu_{11}tu + \mu_{02}u^2) + o(t^2 + u^2).$$

Характеристические функции частных распределений величин ξ и η суть

$$(21.3.4) \quad E(e^{it\xi}) = \varphi(t, 0) \text{ и } E(e^{iu\eta}) = \varphi(0, u).$$

Если величины ξ и η независимы, то

$$\varphi(t, u) = E(e^{it\xi} \cdot e^{iu\eta}) = E(e^{it\xi}) E(e^{iu\eta}),$$

так что совместная характеристическая функция $\varphi(t, u)$ равна произведению характеристических функций частных распределений величин ξ и η .

Обратно, допустим, что совместная характеристическая функция величин ξ и η имеет вид $\varphi_1(t) \cdot \varphi_2(u)$. Вводя в это произведение, если нужно, постоянный множитель, мы можем, очевидно, добиться того, чтобы $\varphi_1(0) = \varphi_2(0) = 1$; тогда из (21.3.4) будет следовать, что функции $\varphi_1(t)$ и $\varphi_2(u)$ являются соответственно характеристическими функциями величин ξ и η . Если двумерный интервал $a_1 < \xi < b_1$, $a_2 < \eta < b_2$ есть интервал непрерывности (см. параграф 8.3) совместного распределения величин ξ и η , то из (10.3.1) и (10.6.2) следует, что

$$P(a_1 < \xi < b_1, a_2 < \eta < b_2) = P(a_1 < \xi < b_1) P(a_2 < \eta < b_2).$$

Заставляя a_1 и a_2 стремиться к $-\infty$, получим, в частности, пользуясь теми же обозначениями, что и в параграфе 21.1, $F(x, y) = F_1(x) \cdot F_2(y)$ для всех x и y , являющихся соответственно точками непрерывности функций F_1 и F_2 . Согласно общим свойствам непрерывности функций распределения, полученное соотношение непосредственно распространяется на все значения x и y . Из (14.4.5) следует

тогда, что величины ξ и η независимы. Таким образом, нами доказана следующая теорема:

Для независимости двух одномерных случайных величин необходимо и достаточно, чтобы их совместная характеристическая функция имела вид

$$(21.3.5) \quad \varphi(t, u) = \varphi_1(t) \varphi_2(u).$$

21.4. Условные распределения. Условное распределение случайной величины η при условии, что другая величина ξ принадлежит некоторому данному множеству S , было определено в параграфе 14.3. В настоящем параграфе мы рассмотрим этот вопрос несколько более подробно для распределений, принадлежащих к одному из двух простых типов, определенных в параграфе 21.1.

1. Дискретный тип. Рассмотрим дискретное распределение, заданное соотношением (21.1.1). Пусть x_i — некоторое значение, для которого частная вероятность $P(\xi = x_i) = \sum_k p_{ik} = p_i$ положительна.

Тогда *условная вероятность* события $\eta = y_k$ при условии $\xi = x_i$, согласно (14.3.1), равна

$$(21.4.1) \quad P(\eta = y_k | \xi = x_i) = \frac{P(\xi = x_i, \eta = y_k)}{P(\xi = x_i)} = \frac{p_{ik}}{p_i}.$$

При фиксированном x_i условные вероятности различных возможных значений y_k определяют *условное распределение* величины η при условии $\xi = x_i$. Сумма всех этих условных вероятностей равна, конечно, единице.

Если (ξ, η) -распределение интерпретировать как распределение единицы массы в точках (x_i, y_k) , то это условное распределение получаем, фиксируя x_i и умножая каждую массу, расположенную на прямой $\xi = x_i$, на $\frac{1}{p_i}$, так чтобы сумма всех масс стала равной единице.

Условное среднее значение функции $g(\xi, \eta)$ при условии $\xi = x_i$ определяется как среднее значение функции $g(x_i, \eta)$ относительно условного распределения величины η , определенного соотношением (21.4.1):

$$(21.4.2) \quad E(g(\xi, \eta) | \xi = x_i) = \frac{\sum_k p_{ik} g(x_i, y_k)}{\sum_k p_{ik}}.$$

При $g(\xi, \eta) = \eta$ получим *условное среднее* величины η , которое является ординатой центра тяжести массы, лежащей на прямой $\xi = x_i$:

$$(21.4.3) \quad E(\eta | \xi = x_i) = m_2^{(i)} = \frac{\sum_k p_{ik} y_k}{\sum_k p_{ik}}.$$

С другой стороны, положив $g(\xi, \eta) = (\eta - m_2^{(i)})^2$, получим *условную дисперсию* величины η .

Условное распределение величины ξ при условии $\eta = y_k$ и соответствующие условные средние значения определяются с помощью перестановки случайных величин в данных выше выражениях.

В частном случае, когда величины ξ и η независимы, соотношение 21.1.4) показывает, что $p_{ik} = p_i p_{k|}$ и, таким образом,

$$(21.4.4) \quad P(\xi = x_i | \eta = y_k) = p_i = P(\xi = x_i), \\ P(\eta = y_k | \xi = x_i) = p_{k|} = P(\eta = y_k),$$

в соответствии с общими соотношениями (14.4.2) и (14.4.3).

2. Непрерывный тип. Пусть $f(x, y)$ — совместная плотность вероятности величин ξ и η . Рассмотрим интервал $(x, x+h)$, такой, что масса, заключенная в вертикальной полосе $x < \xi < x+h$, представляющая вероятность

$$P(x < \xi < x+h) = \int_x^{x+h} \int_{-\infty}^{\infty} f(x, y) dx dy,$$

положительна. Тогда условная вероятность события $\eta \leq y$ при условии $x < \xi < x+h$, согласно (14.3.1), равна

$$P(\eta \leq y | x < \xi < x+h) = \frac{P(x < \xi < x+h, \eta \leq y)}{P(x < \xi < x+h)} = \\ = \frac{\int_x^{x+h} \int_{-\infty}^y f(x, y) dx dy}{\int_x^{x+h} \int_{-\infty}^{\infty} f(x, y) dx dy}$$

Это — функция распределения, соответствующая *условному распределению величины η при условии $x < \xi < x+h$* . Она выражает количество массы, заключенной в полосе $x < \xi < x+h$ и принадлежащей прямой $\eta = y$, деленное на количество всей массы, заключенной в указанной полосе. Пусть теперь $h \rightarrow 0$. Если условия непрерывности,

связанные с (21.1.5), выполнены в точке x и если частная плотность вероятности $f_1(x)$ принимает в точке x положительное значение, то из (5.1.4) следует, что условная функция распределения стремится к пределу

$$(21.4.5) \quad \lim_{h \rightarrow 0} P(\eta \leq y | x < \xi < x + h) = \frac{\int_{-\infty}^y f(x, \eta) d\eta}{\int_{-\infty}^{\infty} f(x, \eta) d\eta} = \\ = \frac{\int_{-\infty}^y f(x, \eta) d\eta}{f_1(x)}.$$

При фиксированном x этот предел как функция от y представляет собой, очевидно, функцию распределения и называется *условной функцией распределения величины η при условии $\xi = x$* .

Если функция $f(x, y)$ непрерывна по y , то полученную условную функцию распределения можно дифференцировать по y , и мы получим соответствующую *условную плотность вероятности* величины η :

$$(21.4.6) \quad f(y|x) = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, \eta) d\eta} = \frac{f(x, y)}{f_1(x)}.$$

Условное среднее значение функции $g(\xi, \eta)$ при условии $\xi = x$ в этом случае равно

$$E[g(\xi, \eta) | \xi = x] = \int_{-\infty}^{\infty} g(x, y) f(y|x) dy = \frac{\int_{-\infty}^{\infty} g(x, y) f(x, y) dy}{f_1(x)}.$$

Умножая на $f_1(x)$ и интегрируя по x , получим

$$(21.4.7) \quad E g(\xi, \eta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy = \\ = \int_{-\infty}^{\infty} E[g(\xi, \eta) | \xi = x] f_1(x) dx.$$

Условное среднее и условная дисперсия величины η равны соответственно:

$$(21.4.8) \quad E(\eta | \xi = x) = m_2(x) = \frac{\int_{-\infty}^{\infty} y f(x, y) dy}{\int_{-\infty}^{\infty} f(x, y) dy},$$

$$(21.4.9) \quad D^2(\eta | \xi = x) = \frac{\int_{-\infty}^{\infty} (y - m_2(x))^2 f(x, y) dy}{\int_{-\infty}^{\infty} f(x, y) dy}.$$

Точка с координатами $\xi = x$, $\eta = m_2(x)$ есть предел центра тяжести массы, заключенной в полосе $x < \xi < x + h$, при $h \rightarrow 0$.

Условное распределение величины ξ при данном значении η и соответствующие условные средние значения определяются аналогично. Так, например, условная плотность вероятности величины ξ при условии $\eta = y$ есть

$$(21.4.10) \quad f(x | y) = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(\xi, y) d\xi} = \frac{f(x, y)}{f_2(y)} = \frac{f_1(x) f(y | x)}{f_2(y)},$$

а условное среднее $E(\xi | \eta = y) = m_1(y)$ есть среднее величины ξ соответствующее плотности вероятности $f(x | y)$.

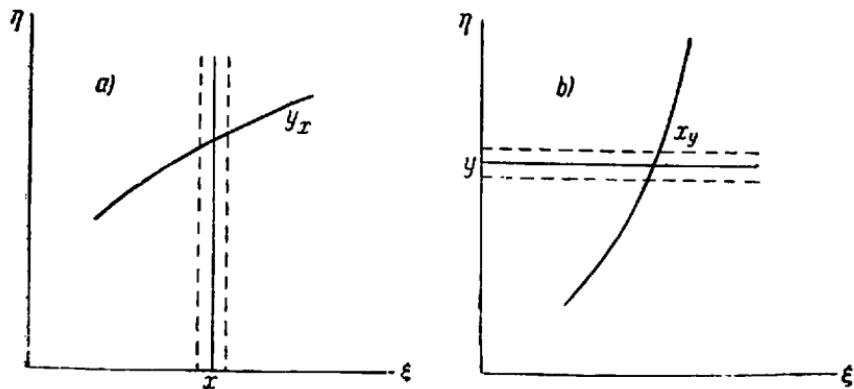
Если величины ξ и η независимы, то $f(x, y) = f_1(x)f_2(y)$. Отсюда следует, что при этом условная плотность вероятности одной величины не зависит от предположений, сделанных относительно значения другой величины, и равна плотности вероятности соответствующего частного распределения. Соответственно, условные средние значения для обеих величин совпадают со средними значениями частных распределений.

$$(21.4.11) \quad m_1(y) = m_1, \quad m_2(x) = m_2.$$

21.5. Регрессия, I. Пусть случайные величины ξ и η имеют совместное распределение непрерывного типа. Предположим, что соответствующая плотность вероятности $f(x, y)$ удовлетворяет условиям непрерывности, связанным с (21.1.5), при любом x , в котором частная плотность вероятности $f_1(x)$ положительна.

Тогда, согласно предыдущему параграфу, условная плотность вероятности $f(y|x)$, заданная соотношением (21.4.6), представляет распределение массы в бесконечной узкой вертикальной полосе, проходящей через точку $\xi=x$. Мы можем здесь рассматривать ξ как *независимое переменное*; тогда фиксированному значению $\xi=x$ соответствует распределение вероятностей *зависимой* переменной η с плотностью вероятности $f(y|x)$.

Рассмотрим теперь некоторое типичное значение этого условного η -распределения, такое, как среднее, мода, медиана и т. п. Вообще говоря, это значение будет зависеть от x и может быть обозначено



Фиг. 22. а) Регрессия η на ξ . б) Регрессия ξ на η .

через y_x . Если x меняется, то точка (x, y_x) описывает некоторую кривую. По виду этой кривой мы можем судить о расположении условного η -распределения для различных значений ξ (см. фиг. 22а).

Кривая этого типа называется *кривой регрессии*, и говорят, что она изображает *регрессию η на ξ* . В дальнейшем, если не оговорено противное, мы будем всегда выбирать за y_x условное среднее $m_2(x)$ величины η , заданное соотношением (21.4.8), и получать, таким образом, *кривую регрессии для среднего величины η как траекторию точки $(x, m_2(x))$ при переменном x* :

$$(21.5.1) \quad y = m_2(x) = E(\eta | \xi = x).$$

Если вместо ξ рассматривать η как независимое переменное, то условная плотность вероятности зависимого переменного ξ при фиксированном значении $\eta=y$ будет задана соотношением (21.4.10). Любое типичное значение x_y условного распределения величины ξ приводит

к некоторой кривой регрессии, изображающей *регрессию* ξ на η (см. фиг. 22, б). Таким образом, кривая регрессии для среднего величины ξ есть траектория точки $(m_1(y), y)$ при переменном y ; она имеет уравнение

$$(21.5.2) \quad x = m_1(y) = E(\xi | \eta = y).$$

Две кривые регрессии (21.5.1) и (21.5.2) не будут, вообще говоря, совпадать. Во многих важных для приложений случаях обе кривые регрессии будут прямыми, или, по крайней мере, приблизительно прямыми. Так, например, в частном случае, когда ξ и η независимы, из (21.4.11) следует, что кривые регрессии суть прямые линии, параллельные координатным осям и проходящие через центр тяжести (m_1, m_2) . Если кривая регрессии есть прямая, то будем говорить, что мы имеем дело со случаем *линейной регрессии*.

Кривые регрессии (21.5.1) и (21.5.2) обладают одним важным свойством минимальности. Попробуем найти среди всех функций $g(\xi)$ такую, которая дает *возможно лучшее представление* другой величины η . Понимая выражение „возможно лучшее“ в смысле принципа наименьших квадратов (см. параграф 15.6), мы должны определить $g(\xi)$ так, чтобы сделать выражение

$$(21.5.3) \quad \begin{aligned} E[\eta - g(\xi)]^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [y - g(x)]^2 f(x, y) dx dy = \\ &= \int_{-\infty}^{\infty} f_1(x) dx \int_{-\infty}^{\infty} [y - g(x)]^2 f(y|x) dy \end{aligned}$$

возможно более малым (см. (21.4.7)). Однако, согласно параграфу 15.4, интеграл по y в последнем выражении при любом значении x принимает наименьшее значение тогда, когда функция $g(x)$ равна условному среднему $m_2(x)$. Итак, среди всех возможных функций $g(\xi)$ *минимум выражения* $E[\eta - g(\xi)]^2$ достигается при функции $g(\xi) = m_2(\xi)$, изображаемой графически *кривой регрессии* (21.5.1). Аналогично, выражение $E[\xi - h(\eta)]^2$ достигает минимума при функции $h(\eta) = m_1(y)$, которая соответствует *кривой регрессии* (21.5.2).

Аналогичные определения могут быть введены в случае распределения дискретного типа, заданного соотношением (21.1.1). Для любого значения x_i величины ξ , для которого частная вероятность p_i положительна, условное распределение величины η задано соотношением (21.4.1). Рассмотрим некоторое типичное значение этого распределения, например, условное сред-

нее $m_2^{(i)}$, заданное соотношением (21.4.3). Когда ξ принимает все возможные значения x_i , мы получаем последовательность точек $(x_i, m_2^{(i)})$, представляющую регрессию η на ξ . Регрессия ξ на η представляется последовательностью точек $(m_1^{(k)}, y_k)$, где $m_1^{(k)}$ — условное среднее величины ξ при условии $\eta = y_k$. В обоих случаях можно соединить точки, соответствующие последовательным значениям i или k , прямолинейными отрезками и рассматривать образованные таким образом ломаные как кривые регрессии дискретного распределения.

21.6. Регрессия, II. В литературе часто называют кривыми регрессии другой тип кривых, отличный от рассмотренного нами в предыдущем параграфе. Сейчас мы перейдем к рассмотрению этого другого типа кривых.

В экстремальной задаче, рассмотренной в связи с (21.5.3), мы пытались найти *среди всех возможных функций* $g(\xi)$ такую, для которой среднее значение величины $(\eta - g(\xi))^2$ имело бы возможно меньшее значение, и обнаружили, что решение этой задачи дается кривой регрессии (21.5.1). Вместо того чтобы рассматривать всевозможные функции $g(\xi)$, можно, однако, ограничиться функциями, принадлежащими некоторому данному классу, например, всевозможными линейными функциями, всевозможными полиномами данной степени n и т. д. Итак, будем искать среди всех функций $g(\xi)$, принадлежащих данному классу, такую, которая дает возможно лучшее представление величины η в смысле принципа наименьших квадратов. В этом случае наша экстремальная задача может иметь попрежнему определенное решение, но оно будет, вообще говоря, соответствовать кривой, отличной от кривой регрессии (21.5.1). Кривые, полученные этим путем, будут называться *кривыми средней квадратической регрессии*.

Простейший случай есть случай *линейной средней квадратической регрессии*. В этом случае мы ищем наилучшую *линейную аппроксимацию величины* η посредством величины ξ , т. е. такую линейную функцию $g(\xi) = a + \beta\xi$, для которой среднее значение величины $(\eta - g(\xi))^2$ принимает наименьшее возможное значение. Употребляя обозначения, введенные в параграфе 21.2, и предполагая $\mu_{20} > 0$, $\mu_{02} > 0$, мы можем написать

$$(21.6.1) \quad \begin{aligned} E(\eta - a - \beta\xi)^2 &= E(\eta - m_2 - \beta(\xi - m_1) + m_2 - a - \beta m_1)^2 = \\ &= \mu_{20}\beta^2 - 2\mu_{11}\beta + \mu_{02} + (m_2 - a - \beta m_1)^2. \end{aligned}$$

Простое вычисление показывает, что поставленная экстремальная задача имеет единственное решение

$$(21.6.2) \quad \beta = \beta_{21} = \frac{\mu_{11}}{\mu_{20}} = \frac{\rho \sigma_2}{\sigma_1}, \quad a = m_2 - \beta_{21} m_1,$$

где ρ — коэффициент корреляции, определенный равенством (21.2.8). Таким образом, прямая средней квадратической регрессии имеет уравнение

$$(21.6.3) \quad y = m_2 + \frac{\rho \sigma_2}{\sigma_1} (x - m_1).$$

Эта линия проходит через точку (m_1, m_2) и ее уравнение может быть также записано в виде

$$(21.6.4) \quad \frac{y - m_2}{\sigma_2} = \rho \frac{x - m_1}{\sigma_1}.$$

Заметим, что эта прямая может быть определена для любого распределения, обе дисперсии которого конечны и положительны, а не только для непрерывных и дискретных распределений, как кривые регрессии предыдущего параграфа.

Величина β_{21} , определенная первым равенством (21.6.2), есть *коэффициент регрессии η на ξ* . Если ввести в (21.6.1) значения a и β , данные равенствами (21.6.2), то (21.6.1) достигнет своего минимального значения

$$(21.6.5) \quad E_{\min} (\eta - a - \beta \xi)^2 = \frac{\mu_{20}\mu_{02} - \mu_{11}^2}{\mu_{20}} = \sigma_2^2 (1 - \rho^2).$$

Выражение $E(\eta - a - \beta \xi)^2 = \int_R (y - a - \beta x)^2 dP$ можно рассматривать как взвешенное среднее квадрата расстояния по вертикали $y - a - \beta x$ между частицей массы dP с координатами (x, y) и прямой $y = a + \beta x$. Так как это среднее становится минимальным для прямой регрессии (21.6.4), то эта прямая дает среди всех остальных прямых *наилучшее* в смысле принципа наименьших квадратов *представление нашего распределения*, если расстояния до этой прямой измеряются вдоль оси y .

В случае распределения, при котором кривая регрессии $y = m_2(x)$, определенная соотношением (21.5.1), существует, выражение $E(\eta - a - \beta \xi)^2$ может быть записано в виде

$$E(\eta - m_2(\xi))^2 + 2E[(\eta - m_2(\xi))(m_2(\xi) - a - \beta \xi)] + \\ + E(m_2(\xi) - a - \beta \xi)^2.$$

Согласно (21.4.7) и (21.4.8), второе слагаемое в этом выражении равно нулю. Таким образом, для любых α и β

$$(21.6.6) \quad E(\eta - \alpha - \beta\xi)^2 = E(\eta - m_2(\xi))^2 + E(m_2(\xi) - \alpha - \beta\xi)^2.$$

В этом соотношении первое слагаемое правой части не зависит от α и β , так что второе слагаемое достигает своего минимума при тех же значениях α и β , что и левая часть соотношения, т. е. для значений (21.6.2). Так как $m_2(x) - \alpha - \beta x$ есть расстояние по вертикали между кривой регрессии $y = m_2(x)$ и прямой $y = \alpha + \beta x$, то прямая средней квадратической регрессии (21.6.4) может рассматриваться как *прямая, дающая наилучшее приближение к кривой регрессии $y = m_2(x)$* , если измерять расстояния вдоль оси y . Отсюда непосредственно вытекает, что *в случае, когда кривая регрессии $y = m_2(x)$ есть прямая линия, она совпадает с прямой средней квадратической регрессии* (21.6.4).

Мы рассмотрели линейную среднюю квадратическую регрессию η на ξ . В случае регрессии ξ на η нужно найти значения α и β , при которых выражение

$$(21.6.7) \quad E(\xi - \alpha - \beta\eta)^2 = \int_{R_s} (x - \alpha - \beta y)^2 dP$$

принимает возможно меньшес значение. Так же, как и выше, находим, что эта задача имеет единственное решение и что прямая, дающая минимум, может рассматриваться как прямая, дающая среди всех остальных прямых наилучшее представление массы нашего распределения (а также и кривой регрессии $x = m_1(y)$) при условии, что расстояния измеряются по горизонтали, т. е. вдоль оси x . Уравнение этой линии — *прямой средней квадратической регрессии величины ξ* — может быть записано в виде

$$(21.6.8) \quad \frac{y - m_2}{\sigma_2} = \frac{1}{\rho} \cdot \frac{x - m_1}{\sigma_1},$$

а коэффициент регрессии имеет выражение

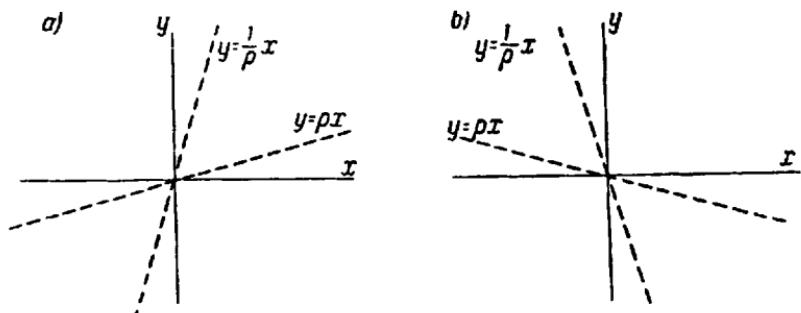
$$(21.6.9) \quad \beta = \beta_{12} = \frac{\mu_{11}}{\mu_{02}} = \frac{\rho\sigma_1}{\sigma_2},$$

в то время как соответствующее значение минимума выражения (21.6.7) есть

$$(21.6.10) \quad E_{\min} (\xi - \alpha - \beta\eta)^2 = \sigma_1^2(1 - \rho^2).$$

Обе прямые средней квадратической регрессии (21.6.4) и (21.6.8) проходят через центр тяжести (m_1 , m_2). Эти две прямые никогда не совпадают, за исключением предельных случаев $\rho = \pm 1$, когда вся распределенная масса лежит на одной прямой (см. параграф 21.2). В этом случае обе прямые регрессии совпадают с этой прямой.

При $\rho = 0$ уравнения прямых средней квадратической регрессии сводятся к $y = m_2$ и $x = m_1$, так что эти линии параллельны соответствующим координатным осям. Этот случай имеет место, например, тогда, когда величины ξ и η независимы (см. параграфы 21.2 и 21.7).



Фиг. 23. Прямые средней квадратической регрессии.

$m_1 = m_2 = 0$, $\sigma_1 = \sigma_2 = 1$. а) $\rho > 0$, б) $\rho < 0$

Если нормировать наши случайные величины, переместив начало координат в центр тяжести и выбрав σ_1 и σ_2 соответственно за единицы измерения величин ξ и η , то уравнения прямых средней квадратической регрессии принимают простой вид $y = \rho x$ и $y = -\frac{x}{\rho}$. Если ρ не равно ни нулю, ни ± 1 , то эти прямые расположены, как показано на фиг. 23а и 23б.

Если вместо измерения расстояния между точкой и прямой линией вдоль одной из координатных осей рассматривать *кратчайшее* расстояние, то мы получим новый тип прямых регрессии. Пусть d есть кратчайшее расстояние между точкой (ξ, η) и прямой L . Если прямая L определена так, что $E(d^2)$ принимает наименьшее возможное значение, то она называется *прямой ортогональной средней квадратической регрессии*. Эта прямая наилучшим образом представляет (ξ, η) -распределение, если расстояния измеряются по перпендикулярам к прямой.

$E(d^2)$ можно рассматривать как *момент инерции* распределенной массы относительно прямой L . Для данного направления прямой L он всегда достигает минимума, когда L проходит через центр тяжести. Таким образом, уравнение L может быть записано в виде $(\xi - m_1) \sin \varphi - (\eta - m_2) \cos \varphi = 0$

где φ есть угол между L и положительным направлением оси ξ . Тогда момент инерции есть

$$E(d^2) = E((\xi - m_1) \sin \varphi - (\eta - m_2) \cos \varphi)^2 = \\ = \mu_{20} \sin^2 \varphi - 2\mu_{11} \sin \varphi \cos \varphi + \mu_{02} \cos^2 \varphi.$$

Если на прямой L по обе стороны от центра тяжести отложить отрезок, длина которого обратно пропорциональна $\sqrt{E(d^2)}$, то геометрическое место концов таких отрезков при различных значениях φ есть *эллипс инерции* нашего распределения. Легко найти уравнение этого эллипса:

$$\frac{(\xi - m_1)^2}{c_1^2} - \frac{2\rho(\xi - m_1)(\eta - m_2)}{c_1 c_2} + \frac{(\eta - m_2)^2}{c_2^2} = c^2.$$

При различных значениях c мы получим семейство гомотетичных эллипсов с общим центром (m_1, m_2) . Направления главных осей этого семейства эллипсов получаются из уравнения

$$\operatorname{tg} 2\varphi = \frac{2\mu_{11}}{\mu_{20} - \mu_{02}},$$

а уравнения этих осей суть

$$(21.6.11) \quad \eta - m_2 = \frac{2\mu_{11}}{\mu_{20} - \mu_{02} \pm \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}} (\xi - m_1).$$

В последнем соотношении нижний знак соответствует большой оси эллипса, следовательно, — минимуму $E(d^2)$, т. е. прямой ортогональной средней квадратической регрессии. В случае

$$\mu_{11} = \mu_{20} = \mu_{02} = 0$$

задача становится неопределенной. Во всех остальных случаях существует единственное решение.

Параболическая средняя квадратическая регрессия порядка $n > 1$ есть обобщение линейной средней квадратической регрессии. В этом случае мы хотим определить полином $g(\xi) = \beta_0 + \dots + \beta_n \xi^n$, при котором среднее значение $M = E(\eta - g(\xi))^2$ имеет возможно меньшее значение. Тогда кривая $y = g(x)$ есть парабола n -го порядка, дающая наилучшую аппроксимацию распределения массы, или кривой регрессии $y = m_2(x)$.

Предполагая, что все моменты, встречающиеся в наших формулах, конечны, получим следующие условия минимума:

$$\frac{1}{2} \frac{\partial M}{\partial \beta_v} = E[\xi^v (g(\xi) - \eta)] = \beta_0 a_{v,0} + \dots + \beta_n a_{v,n,0} - a_{v,1} = 0,$$

где $v = 0, 1, \dots, n$. Если моменты $a_{v,k}$ известны, то для определения $n+1$ неизвестных β_0, \dots, β_n имеем, таким образом, $n+1$ уравнений.

Вычисления, необходимые для определения неизвестных коэффициентов, могут быть значительно упрощены, если полином регрессии $g(x)$ рассматривается как линейная комбинация ортогональных полиномов $p_v(x)$, связанных с частным распределением величины ξ (см. параграф 12.6). Для всех порядков, для которых эти полиномы определены однозначно, имеем

$$(21.6.12) \quad E(p_m(\xi)p_n(\xi)) = \\ = \int_{-\infty}^{\infty} p_m(x)p_n(x)dF_1(x) = \begin{cases} 1 & \text{при } m=n, \\ 0 & \text{при } m \neq n, \end{cases}$$

где полином $p_n(x)$ имеет степень n , а $F_1(x)$ обозначает частную функцию распределения величины ξ . Любой полином $g(x)$ степени n может быть тогда записан в виде

$$g(x) = c_0 p_0(x) + \dots + c_n p_n(x)$$

с постоянными коэффициентами c_0, \dots, c_n . Условия минимума приобретают тогда такой вид:

$$(21.6.13) \quad \frac{1}{2} \frac{\partial M}{\partial c_v} = E[p_v(\xi)(g(\xi) - \eta)] = c_v - E(\eta p_v(\xi)) = 0.$$

Отсюда коэффициенты c_v получаются непосредственно: $c_v = E(\eta p_v(\xi))$. Далее, выражение для c_v не зависит от степени n . Таким образом, если известен, например, полином регрессии степени n и ищется соответствующий полином степени $n+1$, то нужно вычислить только слагаемое $c_{n+1} p_{n+1}(x)$. Вводя коэффициенты c_v в выражение для среднего значения M , находим минимальное значение M :

$$(21.6.14) \quad E_{\min}(\eta - g(\xi))^2 = E(\eta^2) - c_0^2 - \dots - c_n^2.$$

Заметим, наконец, что для выполнения приведенных выше соотношений иссущественно, что $p_v(x)$ суть полиномы. Любая последовательность функций, удовлетворяющая условиям ортогональности (21.6.12), может быть использована для образования кривой средней квадратической регрессии $y = g(x) = \sum c_v p_v(x)$, и соотношения (21.6.13) и (21.6.14) будут выполняться, каков бы ни был вид функций $p_v(x)$.

21.7. Коэффициент корреляции. Согласно (21.2.8), коэффициент корреляции ρ величин ξ и η определяется выражением

$$\rho = \frac{\mu_{11}}{c_1 c_2} = \frac{E[(\xi - m_1)(\eta - m_2)]}{\sqrt{E(\xi - m_1)^2 E(\eta - m_2)^2}},$$

и мы видели в параграфе 21.2, что $-1 \leq \rho \leq 1$. Коэффициент корреляции является важной характеристикой (ξ, η) -распределения. Его основные свойства тесно связаны с прямыми средней квадратической регрессии

$$(21.7.1) \quad \begin{aligned} \frac{y - m_2}{\sigma_2} &= \rho \cdot \frac{x - m_1}{\sigma_1}, \\ \frac{y - m_2}{\sigma_2} &= \frac{1}{\rho} \cdot \frac{x - m_1}{\sigma_1}, \end{aligned}$$

каждая из которых дает наилучшую аппроксимацию (ξ, η) -распределения в смысле, определенном в предыдущем параграфе. Точность аппроксимации, даваемой каждой из этих прямых, характеризуется соответственно выражениями

$$(21.7.2) \quad \begin{aligned} E_{\min}(\eta - \alpha - \beta\xi)^2 &= \sigma_2^2(1 - \rho^2), \\ E_{\min}(\xi - \alpha - \beta\eta)^2 &= \sigma_1^2(1 - \rho^2). \end{aligned}$$

Таким образом, если из одной величины вычесть ее наилучшее приближение посредством линейной функции другой величины, то дисперсия „остатка“ уменьшится сравнительно с дисперсией самой величины в отношении $(1 - \rho^2):1$. Выражения (21.7.2) называются иногда *остаточными дисперсиями* величин η и ξ .

Если $\rho = 0$, то дисперсию величины η нельзя уменьшить, вычитая из нее какую бы то ни было линейную функцию от ξ , и, наоборот, дисперсию величины ξ нельзя уменьшить, вычитая из нее линейную функцию от η . В этом случае будем говорить, что наши величины *некоррелированы*.

Если $\rho \neq 0$, то дисперсию величины η можно уменьшить, вычитая из нее некоторую линейную функцию от ξ , и, обратно, дисперсию величины ξ можно уменьшить, вычитая из нее некоторую линейную функцию от η . Максимальная величина, на которую можно таким способом уменьшить дисперсию, тем больше, чем больше абсолютная величина коэффициента корреляции ρ . При $\rho \neq 0$ мы будем говорить, что наши величины *коррелированы* и что корреляция *положительна* или *отрицательна* соответственно при $\rho > 0$ и $\rho < 0$.

Если ρ достигает одного из своих экстремальных значений ± 1 , то, как показывает (21.7.2), остаточные дисперсии равны нулю. Мы уже показали в 21.2, что этот случай имеет место тогда и только тогда, когда вся масса (ξ, η) -распределения лежит на некоторой прямой, совпадающей тогда с обеими прямыми регрессии (21.7.1). В этом

экстремальном случае существует точная функциональная зависимость между нашими величинами: каждому значению ξ соответствует единственное значение η , и обратно. Каждая величина является линейной функцией другой, и обе величины меняются в одном направлении, если $\rho = +1$, и в разных направлениях, если $\rho = -1$.

Вследствие этих свойств коэффициент корреляции ρ можно рассматривать как меру прямолинейности (ξ, η) -распределения. Эта прямолинейность достигает своего максимума при $\rho = \pm 1$, когда вся распределенная масса лежит на некоторой прямой линии. Противоположный крайний случай $\rho = 0$, когда дисперсия одной из величин не может быть уменьшена вычитанием из нее какой-либо линейной функции другой величины.

В параграфе 21.2 уже было показано, что в частном случае, когда величины ξ и η независимы, $\rho = 0$. Таким образом, *две независимые величины всегда некоррелированы*. Важно отметить, что обратное не имеет: *две некоррелированные величины не обязательно независимы*.

Действительно, рассмотрим одномерную плотность вероятности $g(x)$, отличную от нуля только при $x > 0$ и имеющую конечный второй момент. Тогда

$$f(x, y) = \frac{g(\sqrt{x^2 + y^2})}{2\pi\sqrt{x^2 + y^2}}$$

есть функция плотности двумерного распределения, при котором плотность массы на любой окружности $x^2 + y^2 = c^2$ постоянна. Центр тяжести есть $m_1 = m_2 = 0$ и, в силу симметрии распределения, имеем $\mu_{11} = 0$, следовательно, $\rho = 0$. Таким образом, величины с таким совместным распределением некоррелированы. Однако, согласно (15.11.3), для *независимости* этих величин необходимо и достаточно, чтобы функция $f(x, y)$ представлялась в виде $f_1(x)f_2(y)$, а это условие выполняется не всегда, как это видно на примере $g(x) = e^{-x}$.

Если ρ есть коэффициент корреляции величин ξ и η , то непосредственно из определения следует, что величины $\xi' = a\xi + b$ и $\eta' = c\eta + d$ имеют коэффициент корреляции $\rho' = \rho \operatorname{sgn}(ac)$, где $\operatorname{sgn} x = +1$ при $x > 0$ и $\operatorname{sgn} x = -1$ при $x < 0$.

В частном случае дискретного распределения, при котором каждая величина способна принимать только два возможных значения (соответственно x_1, x_2 и y_1, y_2), после некоторых упрощений, используя

обозначения параграфа 21.1, находим, что

$$(21.7.3) \quad \rho = \frac{p_{11}p_{22} - p_{12}p_{21}}{\sqrt{p_{11}p_{22}p_{12}p_{21}}} \operatorname{sgn}[(x_1 - x_2)(y_1 - y_2)].$$

21.8. Линейное преобразование случайных величин. Рассмотрим линейное преобразование случайных величин ξ и η , соответствующее повороту координатных осей около центра тяжести. Вводим новые переменные X и Y , определенные соотношениями

$$(21.8.1) \quad \begin{aligned} X &= (\xi - m_1) \cos \varphi + (\eta - m_2) \sin \varphi, \\ Y &= -(\xi - m_1) \sin \varphi + (\eta - m_2) \cos \varphi. \end{aligned}$$

Обратное преобразование записывается так:

$$(21.8.2) \quad \begin{aligned} \xi &= m_1 + X \cos \varphi - Y \sin \varphi, \\ \eta &= m_2 + X \sin \varphi + Y \cos \varphi. \end{aligned}$$

Если угол поворота φ определяется уравнением $\operatorname{tg} 2\varphi = \frac{2\mu_{11}}{\mu_{20} - \mu_{02}}$, то

$$E(XY) = \mu_{11} \cos 2\varphi - \frac{1}{2}(\mu_{20} - \mu_{02}) \sin 2\varphi = 0,$$

так что величины X и Y некоррелированы. В частном случае $\mu_{11} = \mu_{20} - \mu_{02} = 0$, когда уравнение для φ неопределенно, $E(XY) = 0$ при любом φ . Таким образом, величины ξ и η всегда можно представить в виде линейных функций двух некоррелированных величин.

Рассмотрим частный случай, когда матрица вторых моментов $M = \begin{pmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{pmatrix}$ имеет ранг 1 (см. параграф 21.2). Тогда $\rho = \pm 1$ и вся распределенная масса лежит на прямой $\eta - m_2 = \frac{\rho\sigma_2}{\sigma_1}(\xi - m_1)$. Определим теперь угол поворота φ из уравнения $\operatorname{tg} \varphi = \frac{\rho\sigma_2}{\sigma_1}$. Тогда из (21.8.1) находим

$$\begin{aligned} E(Y^2) &= \sigma_1^2 \sin^2 \varphi - 2\rho\sigma_1\sigma_2 \sin \varphi \cos \varphi + \sigma_2^2 \cos^2 \varphi = \\ &= (\sigma_1 \sin \varphi - \rho\sigma_2 \cos \varphi)^2 = 0. \end{aligned}$$

Таким образом, дисперсия величины Y равна нулю, так что величина почти всюду равна нулю (см. параграф 16.1). Если в (21.8.2) положить $Y = 0$, то получающиеся соотношения между ξ , η и X будут выполняться с вероятностью, равной единице. Итак, две величины ξ и η , такие, что матрица M их вторых моментов имеет ранг 1, могут быть

с вероятностью, равной единице, представлены в виде линейных функций одной случайной величины.

21.9. Корреляционное отношение и средняя квадратическая связанность. Рассмотрим две величины ξ и η с распределением непрерывного типа, при котором условное среднее $m_2(x)$ есть непрерывная функция от x . Полагая в (21.6.6) $\alpha = m_2$, $\beta = 0$, получим

$$(21.9.1) \quad \sigma_2^2 = E(\eta - m_2)^2 = E(\eta - m_2(\xi))^2 + E(m_2(\xi) - m_2)^2.$$

Мы видим, что дисперсия величины η может быть представлена как сумма двух компонент, а именно, среднего квадратического отклонения η от ее условного среднего $m_2(\xi)$ и среднего квадратического отклонения $m_2(\xi)$ от ее среднего m_2 .

Определим теперь величину $\theta_{\eta\xi}$, положив

$$(21.9.2) \quad \theta_{\eta\xi}^2 = \frac{1}{\sigma_2^2} E(m_2(\xi) - m_2)^2 = \frac{1}{\sigma_2^2} \int_{-\infty}^{\infty} (m_2(x) - m_2)^2 f_1(x) dx.$$

Эта величина $\theta_{\eta\xi}$ есть *корреляционное отношение*^{*)} τ_η на ξ , введенное К. Пирсоном. В приложениях обычно рассматривают квадрат этой величины, и поэтому ее знак можно оставить неопределенным. Из (21.9.1) получим

$$(21.9.3) \quad 1 - \theta_{\eta\xi}^2 = \frac{1}{\sigma_2^2} E(\eta - m_2(\xi))^2,$$

и, следовательно,

$$(21.9.4) \quad 0 \leq \theta_{\eta\xi}^2 \leq 1.$$

Далее, пишем уравнение первой прямой средней квадратической регрессии (21.7.1) в виде $y = \alpha + \beta x$ и подставляем эти значения α и β в (21.6.6). Тогда, в силу (21.7.2) и (21.9.3), получим после некоторых упрощений

$$(21.9.5) \quad \theta_{\eta\xi}^2 = \rho^2 + \frac{1}{\sigma_2^2} E(m_2(\xi) - x - \hat{\beta}\xi)^2.$$

Отсюда следует, что $\theta_{\eta\xi}^2 = 0$ тогда и только тогда, когда $m_2(x)$ не зависит от x . Действительно, если $m_2(x)$ постоянно, то кривая регрессии $y = m_2(x)$ есть горизонтальная прямая; следовательно,

^{*)} В литературе обычно корреляционное отношение обозначается буквой τ . Этим обозначением мы не можем здесь пользоваться, так как через τ нас обозначена случайная величина.

$\rho = \beta = 0$ и $\theta_{\eta\xi}^2 = 0$. Обратное доказывается аналогично. Далее, из (21.9.3) видно, что $\theta_{\eta\xi}^2 = 1$ тогда и только тогда, когда вся распределенная масса лежит на кривой регрессии $y = m_2(x)$, так что существует точная функциональная зависимость между нашими величинами. Для промежуточных значений $\theta_{\eta\xi}^2$ (21.9.3) показывает, что корреляционное отношение можно рассматривать как меру стремления массы накапливаться около кривой регрессии.

Если регрессия η на ξ линейна, так что $y = m_2(x)$ есть прямая линия, то (21.9.5) показывает, что $\theta_{\eta\xi}^2 = \rho^2$, и (21.9.3) сводится к первому из соотношений (21.7.2). В этом случае вычисление корреляционного отношения не дает ничего нового, если уже известен коэффициент корреляции ρ .

С другой стороны, в случае нелинейной регрессии $\theta_{\eta\xi}^2$ всегда пре-
восходит ρ^2 на величину, характеризующую отклонение кривой $y = m_2(x)$ от прямой, дающей наилучшую аппроксимацию нашего рас-
пределения в смысле, указанном ранее.

Корреляционное отношение $\theta_{\eta\xi}$ на η определяется, конечно, ана-
логичным образом.

Для распределения *дискретного* типа корреляционное отношение может быть определено аналогично, причем соотношения (21.9.2) и (21.9.3) заменяются соотношениями

$$(21.9.2a) \quad \theta_{\eta\xi}^2 = \frac{1}{\sigma_2^2} E(m_2^{(i)} - m_2)^2 = \frac{1}{\sigma_2^2} \sum_i p_i (m_2^{(i)} - m_2)^2,$$

$$(21.9.3a) \quad 1 - \theta_{\eta\xi}^2 = \frac{1}{\sigma_2^2} E(\eta_i - m_2^{(i)})^2,$$

где p_i и $m_2^{(i)}$ определяются соответственно соотношениями (21.1.2) и (21.4.3). Соотношения (21.9.4), (21.9.5) и сделанные выше заключе-
ния относительно свойств корреляционного отношения справедливы с очевидными видоизменениями и в этом случае.

Как коэффициент корреляции, так и корреляционное отношение служат для характеристики „степени зависимости“ между двумя величи-
нами в смысле, истолкованном выше. Для этой же цели предлагались и многие другие характеристики. Из них мы упомянем здесь только среднюю квадратическую связьность, введенную К. Пирсоном. Рас-
смотрим две величины ξ , η с совместным распределением дискретного типа, определенным равенствами (21.1.1), и предположим, что число

возможных значений конечно для обеих величин. Тогда вероятности p_{ik} образуют матрицу, имеющую, скажем, m строк и n столбцов. Так как любой столбец или любая строка, состоящие исключительно из нулей, могут быть отброшены, то можно предположить, что каждая строка и каждый столбец содержат по крайней мере один положительный элемент, так что все суммы p_{il} и p_{ik} положительны. Тогда *средняя квадратическая связанность* распределения определяется так:

$$(21.9.6) \quad \varphi^2 = \sum_{i, k} \frac{(p_{ik} - p_{il} p_{ik})^2}{p_{il} p_{ik}} = \sum_{i, k} \frac{p_{ik}^2}{p_{il} p_{ik}} - 1.$$

Согласно (21.1.4), $\varphi^2 = 0$ тогда и только тогда, когда наши величины независимы. С другой стороны, из последнего выражения, в силу неравенств $p_{ik} \leq p_{il}$ и $p_{ik} \leq p_{ik}$, следует, что $\varphi^2 \leq q - 1$, где $q = \min(m, n)$ обозначает наименьшее из чисел m и n или их общее значение, если $m = n$. Далее, знак равенства в последнем соотношении имеет место тогда и только тогда, когда одна из величин есть однозначно определенная функция другой. Таким образом, $0 \leq \frac{\varphi^2}{q-1} \leq 1$

и величина $\frac{\varphi^2}{q-1}$ может быть использована (на нормированной шкале) как мера степени зависимости между величинами ξ и η .

В частном случае $m = n = 2$ после некоторых преобразований получаем

$$(21.9.7) \quad \varphi^2 = \frac{(p_{11}p_{22} - p_{12}p_{21})^2}{p_{11}p_{22}p_{12}p_{21}}.$$

Таким образом, при этом φ^2 есть квадрат коэффициента корреляции ρ , заданного соотношением (21.7.3). В этом случае $q = 2$ и $\frac{\varphi^2}{q-1}$ совпадает с φ^2 . Далее, φ^2 принимает свое максимальное значение 1 только в двух случаях: $p_{12} = p_{21} = 0$ или $p_{11} = p_{22} = 0$.

21.10. Эллипс рассеяния. Рассмотрим некоторую одномерную случайную величину ξ' со средним значением m и стандартным отклонением σ . Если ξ' — другая случайная величина, равномерно распределенная (см. параграф 19.1) по интервалу $(m - \sigma\sqrt{3}, m + \sigma\sqrt{3})$, то легко видеть, что ξ' имеет то же среднее значение и то же стандартное отклонение, что и ξ . Таким образом, интервал $(m - \sigma\sqrt{3}, m + \sigma\sqrt{3})$ можно считать геометрической характеристикой концентрации ξ -распределения около его центра тяжести m (см. также параграф 15.6).

Будем теперь искать аналогичную геометрическую характеристику концентрации *двумерного* распределения около его центра тяжести (m_1, m_2). Для этой цели ищем кривую, окружающую точку (m_1, m_2), и такую, что если единица массы равномерно распределена по площади, ограниченной этой кривой, то такое распределение имеет те же моменты первого и второго порядка, что и данное распределение. (Под „равномерным распределением“ мы подразумеваем, конечно, распределение с постоянной плотностью вероятности.)

В этой общей форме задача, конечно, является неопределенной, и мы ограничимся отысканием *эллипса*, обладающего указанными свойствами. Для упрощения записи предположим, что $m_1 = m_2 = 0$. Пусть μ_{20} , μ_{11} и μ_{02} — центральные моменты второго порядка данного распределения. Предположим, что $\rho^2 < 1$, так что наше распределение не принадлежит предельному типу, при котором вся масса лежит на некоторой прямой.

Рассмотрим неотрицательную квадратичную форму

$$q(\xi, \eta) = a_{11}\xi^2 + 2a_{12}\xi\eta + a_{22}\eta^2.$$

Согласно (11.12.3), площадь, ограниченная эллипсом $q = c^2$, равна $\frac{\pi c^2}{\sqrt{A}}$, где $A = a_{11}a_{22} - a_{12}^2$. Если единица массы равномерно распределена по этой площади, то моменты первого порядка этого распределения, очевидно, равны нулю, а моменты второго порядка, согласно (11.12.4) равны

$$\frac{c^2}{4} \cdot \frac{a_{20}}{A}, \quad -\frac{c^2}{4} \cdot \frac{a_{12}}{A} \quad \text{и} \quad \frac{c^2}{4} \cdot \frac{a_{11}}{A}.$$

Нужно определить c и a_{ik} так, чтобы эти моменты совпадали соответственно с μ_{20} , μ_{11} и μ_{02} . Легко видеть, что это выполняется при $c^2 = 4$ и

$$a_{11} = \frac{\mu_{02}}{M}, \quad a_{12} = -\frac{\mu_{11}}{M}, \quad a_{22} = \frac{\mu_{20}}{M},$$

где $M = \mu_{20}\mu_{02} - \mu_{11}^2$. Получаемая таким образом форма $q(\xi, \eta)$ является обратной (см. параграф 11.7) к форме

$$Q(\xi, \eta) = \mu_{20}\xi^2 + 2\mu_{11}\xi\eta + \mu_{02}\eta^2.$$

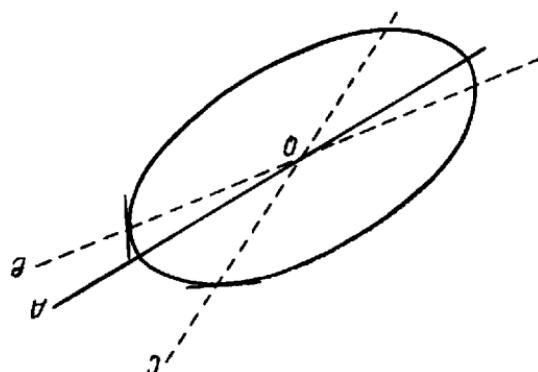
Возвращаясь к общему случаю произвольного центра тяжести (m_1, m_2) и заменяя μ_{ik} их выражениями через σ_1, σ_2 и ρ , получим, что

равномерное распределение единицы массы по площади, ограниченной эллипсом

$$(21.10.1) \quad \frac{1}{1 - \rho^2} \left(\frac{(\xi - m_1)^2}{\sigma_1^2} - \frac{2\rho(\xi - m_1)(\eta - m_2)}{\sigma_1 \sigma_2} + \frac{(\eta - m_2)^2}{\sigma_2^2} \right) = 4,$$

имеет те же моменты первого и второго порядков, что и данное распределение. Этот эллипс называется эллипсом рассеяния, соответствующим данному распределению.

Область, ограниченную эллипсом (21.10.1), можно, таким образом, рассматривать как двумерный аналог интервала $(m - 3\sqrt{3}, m + 3\sqrt{3})$.



Фиг. 24. Эллипс рассеяния и прямые регрессии, $\rho > 0$. Q — центр тяжести, QA — прямая ортогональной средней квадратической регрессии, QB — прямая средней квадратической регрессии η на ξ , QC — прямая средней квадратической регрессии ξ на η .

постоянной c^2 , то получим для различных значений c^2 семейство подобных эллипсов с общим центром (m_1, m_2) , которое совпадает с семейством эллипсов инерции, рассмотренным в параграфе 21.6. Общая большая ось этих эллипсов совпадает с прямой ортогональной средней квадратической регрессии нашего распределения (см. параграф 21.6). Обычные прямые средней квадратической регрессии являются диаметрами этих эллипсов, сопряженными с одной из координатных осей (фиг. 24).

21.11. Сложение независимых случайных величин. Рассмотрим двумерные случайные величины $x_1 = (\xi_1, \eta_1)$ и $x_2 = (\xi_2, \eta_2)$. Определим сумму $x = x_1 + x_2$ по правилу векторного сложения

$$x = (\xi, \eta) = (\xi_1 + \xi_2, \eta_1 + \eta_2).$$

Если два распределения в R_2 с общим центром тяжести таковы, что эллипс рассеяния одного из них лежит целиком в другом, то говорят, что первое распределение имеет меньшее рассеяние по сравнению со вторым. Это понятие найдет важное применение в теории оценок (см. параграф 32.7).

Если заменить постоянную 4 в уравнении (21.10.1) произвольной

Согласно параграфу 14.5, x есть двумерная случайная величина с распределением, однозначно определенным совместным распределением величин x_1 и x_2 .

Предположим теперь, что величины x_1 и x_2 *независимы* (см. параграф 14.4), и обозначим через $\varphi(t, u)$, $\varphi_1(t, u)$ и $\varphi_2(t, u)$ характеристические функции величин x , x_1 и x_2 . Согласно теореме (15.3.4), имеем

$$(21.11.1) \quad \begin{aligned} \varphi(t, u) &= E(e^{i(\xi+u\eta)}) = E(e^{i(\xi_1+u\eta_1)} \cdot e^{i(\xi_2+u\eta_2)}) = \\ &= \varphi_1(t, u) \varphi_2(t, u). \end{aligned}$$

Обобщение на случай произвольного числа слагаемых очевидно. Таким образом, мы получаем ту же теорему, что и для одномерных величин (см. параграф 15.12):

Характеристическая функция суммы независимых слагаемых есть произведение характеристических функций слагаемых.

Рассмотрим теперь сумму $x = x_1 + x_2 + \dots + x_n$, где величины $x_v = (\xi_v, \eta_v)$ независимы и все имеют одно и то же двумерное распределение. Предположим, что это последнее распределение имеет конечные моменты второго порядка $\mu_{20}, \mu_{11}, \mu_{02}$ и что моменты первого порядка равны нулю: $m_1 = m_2 = 0$. Если $\varphi(t, u)$ есть характеристическая функция этого распределения величины x_v , то, согласно (21.3.3), имеем

$$(21.11.2) \quad \varphi(t, u) = 1 - \frac{1}{2} (\mu_{20}t^2 + 2\mu_{11}tu + \mu_{02}u^2) + o(t^2 + u^2).$$

С другой стороны, имеем $x = (\xi_1 + \dots + \xi_n, \eta_1 + \dots + \eta_n)$ и

$$\frac{x}{\sqrt{n}} = \left(\frac{\xi_1 + \dots + \xi_n}{\sqrt{n}}, \frac{\eta_1 + \dots + \eta_n}{\sqrt{n}} \right).$$

Если $\varphi_n(t, u)$ есть характеристическая функция величины $\frac{x}{\sqrt{n}}$, то из сказанного выше следует, что

$$\varphi_n(t, u) = \left[\varphi \left(\frac{t}{\sqrt{n}}, \frac{u}{\sqrt{n}} \right) \right]^n.$$

Заменяя в (21.11.2) t и u на $\frac{t}{\sqrt{n}}$ и $\frac{u}{\sqrt{n}}$, получим

$$\varphi_n(t, u) = \left[1 - \frac{\mu_{20}t^2 + 2\mu_{11}tu + \mu_{02}u^2}{2n} + \frac{\delta(n, t, u)}{n} \right]^n,$$

где при любых фиксированных t и u величина $\delta(n, t, u)$ стремится к нулю при $n \rightarrow \infty$. Следовательно, таким же образом, как и при

доказательстве теоремы Линдеберга — Леви (см. параграф 17.4), получим

$$(21.11.3) \quad \lim_{n \rightarrow \infty} \varphi_n(t, u) = e^{-\frac{1}{2}(\mu_2 t^2 + 2\mu_{11}tu + \mu_{02}u^2)}.$$

Следовательно, $\varphi_n(t, u)$ при любых t и u стремится к пределу, который, очевидно, является функцией непрерывной в точке $(t, u) = (0, 0)$. Согласно теореме непрерывности для характеристических функций, доказанной в параграфе 10.7, мы можем утверждать, что этот предел есть характеристическая функция некоторого распределения, которое в свою очередь служит пределом распределения величины $\frac{x}{\sqrt{n}}$ при $n \rightarrow \infty$.

Итак, если двумерные величины x_1, x_2, \dots независимы и все имеют одно и то же распределение с конечными моментами второго порядка и моментами первого порядка, равными нулю, то распределение величины $\frac{x_1 + \dots + x_n}{\sqrt{n}}$ стремится при $n \rightarrow \infty$ к некоторому предельному распределению; характеристическая функция предельного распределения задается правой частью соотношения (21.11.3). Исключая тривиальное ограничение $m_1 = m_2 = 0$, это есть обобщение теоремы Линдеберга — Леви (см. параграф 17.4) на двумерный случай.

Нужно заметить, что относительно моментов второго порядка здесь предположено только, что они конечны. Предположим теперь, что заданы величины μ_{20}, μ_{11} и μ_{02} , такие, что квадратичная форма

$$\mu_{20}t^2 + 2\mu_{11}tu + \mu_{02}u^2$$

неотрицательна; тогда можно найти (см. параграф 21.2) распределение, при котором $m_1 = m_2 = 0$ и вторые моменты которого равны данным $\mu_{20}, \mu_{11}, \mu_{02}$. Если принять это распределение за распределение каждой из величин x_i в доказанной выше теореме, то выражение в правой части (21.11.3) будет характеристической функцией распределения некоторой величины x . Вводя постоянный вектор $m = (m_1, m_2)$, получим, что величина $m + x$ имеет характеристическую функцию

$$(21.11.4) \quad e^{t(m_1 + m_2u) - \frac{1}{2}(\mu_{20}u^2 + 2\mu_{11}(u + m_2u^2))}$$

Распределение, соответствующее этой характеристической функции, есть *двумерное нормальное распределение*, которое будет рассмотрено в следующем параграфе.

21.12. Нормальное распределение. Переходим теперь к изучению распределения, соответствующего характеристической функции (21.11.4). Нужно различать два случая, соответственно тому, является ли неотрицательная квадратичная форма

$$Q(t, u) = \mu_{20}t^2 + 2\mu_{11}tu + \mu_{02}u^2$$

положительно-определенной или только полуопределенной (см. параграф 11.10). В первом случае будем говорить, что мы имеем *собственное нормальное распределение*, во втором случае — *несобственное нормальное распределение*. Говоря просто о *нормальном распределении*, мы будем иметь в виду оба случая.

Рассмотрим сперва случай *положительно-определенной* формы $Q(t, u)$. При этом обратная форма $Q^{-1}(x, y)$ существует и имеет выражение (см. параграф 21.10):

$$Q^{-1}(x, y) = \frac{\mu_{20}x^2 - 2\mu_{11}xy + \mu_{02}y^2}{M} = \frac{1}{1 - \rho^2} \left(\frac{x^2}{\sigma_1^2} - \frac{2\rho xy}{\sigma_1 \sigma_2} + \frac{y^2}{\sigma_2^2} \right),$$

где $M = \mu_{20}\mu_{02} - \mu_{11}^2 = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$. Из (11.12.1b) получим

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t(tx + uy) - \frac{1}{2} Q^{-1}(x, y)} dx dy = 2\pi \sqrt{M} e^{-\frac{1}{2} Q(t, u)},$$

или, заменяя x на $x - m_1$ и y на $y - m_2$,

$$\begin{aligned} & \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t(tx + uy) - \frac{1}{2} Q^{-1}(x - m_1, y - m_2)} dx dy = \\ & = e^{t(m_1 t + m_2 u) - \frac{1}{2} Q(t, u)} \end{aligned}$$

Последнее соотношение показывает, что *функция*

$$(21.12.1) \quad f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2} Q^{-1}(x - m_1, y - m_2)}$$

есть двумерная функция плотности с характеристической функцией

$$(21.12.2) \quad \varphi(t, u) = e^{t(m_1 t + m_2 u) - \frac{1}{2} Q(t, u)}.$$

Разложение (21.3.2) для характеристической функции показывает, что величины m_i и μ_{ik} имеют для этого распределения свой обычный смысл средних значений и центральных моментов второго порядка. Функция $f(x, y)$, определенная равенством (21.12.1), есть *нормальная функция плотности* от двух переменных. Она имеет максимальное значение в центре тяжести (m_1, m_2) . Подобные эллипсы

$$(21.12.3) \quad \frac{1}{2(1-\rho^2)} \left(\frac{(x-m_1)^2}{\sigma_1^2} + \frac{2\rho(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2} \right) = c^2,$$

уже встречавшиеся раньше в параграфах 21.6 и 21.10, играют в этом случае роль *кривых равных вероятностей*. Действительно, для любой точки, принадлежащей к эллипсу (21.12.3), имеем

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-c^2}.$$

Так как, согласно (11.12.3), площадь кольца между эллипсами, соответствующими значениям c и $c+dc$, равна

$$4\pi\sigma_1\sigma_2\sqrt{1-\rho^2}cdc,$$

то масса, лежащая в этом кольце, равна $2ce^{-c^2}dc$, и, следовательно, масса, лежащая вне эллипса (21.12.3), равна

$$\int_c^{\infty} 2ce^{-c^2}dc = e^{-c^2}.$$

Форма эллипса равных вероятностей (21.12.3) дает хорошее представление о виде нормальной поверхности плотности $z=f(x, y)$. При $\rho=0$, $\sigma_1=\sigma_2$ эллипсы превращаются в окружности. Когда ρ приближается к $+1$ или -1 , эллипсы становятся более тонкими и вытянутыми, что является показателем стремления массы сосредоточиваться около общей большой оси этих эллипсов, являющейся прямой ортогональной средней квадратической регрессии (см. параграф 21.6) нашего распределения.

О величине (ξ, η) с плотностью вероятности (21.12.1) говорят, что она обладает *собственным нормальным распределением*. Согласно (21.3.4), характеристическая функция частного распределения величины ξ есть

$$\varphi(t, 0) = e^{im_1t - \frac{1}{2}\sigma_1^2t^2}.$$

Согласно параграфу 17.2, величина ξ нормальна (m_1, σ_1) с частной плотностью вероятности

$$f_1(x) = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}}$$

Заменой индексов получаем соответствующее выражение для частной плотности вероятности $f_2(y)$ величины η .

В частном случае, когда $\rho = 0$, видно, что $f(x, y) = f_1(x)f_2(y)$, откуда следует, что величины ξ и η независимы. Таким образом, для нормального распределения справедливо утверждение, что две некоррелированные величины независимы, хотя мы видели в параграфе 21.7, что в общем случае это неверно.

Условная плотность вероятности величины η при условии $\xi = x$, согласно (21.4.6), равна

$$(21.12.4) \quad f(y|x) = \frac{f(x,y)}{f_1(x)} = \\ = \frac{1}{\sigma_2 \sqrt{2\pi(1-\rho^2)}} \cdot e^{-\frac{1}{2\sigma_2^2(1-\rho^2)} \left(y - m_2 - \frac{\rho\sigma_2}{\sigma_1}(x - m_1) \right)^2}$$

Это — нормальная функция плотности относительно y со средним значением

$$m_2(x) = m_2 + \frac{\rho\sigma_2}{\sigma_1}(x - m_1)$$

и стандартным отклонением $\sigma_2 \sqrt{1 - \rho^2}$. Таким образом, регрессия η на ξ линейна, и условная дисперсия величины η не зависит от значения, принимаемого ξ . Аналогичные свойства условного распределения величины ξ при данном значении η выводятся точно так же.

Если неотрицательная форма $Q(t, u)$ является полуопределенной, то детерминант M равен нулю, и обратной формы не существует (см. параграфы 11.7 и 11.10). Однако из предыдущего параграфа следует, что выражение (21.12.2) попрежнему является характеристической функцией некоторого распределения, которое будет называться *несобственным нормальным распределением*. Согласно параграфу 21.2, при этом распределении вся масса сосредоточена в одной точке или на одной прямой линии, соответственно тому, равен ли ранг матрицы M нулю или единице.

В этом случае очевидно, что не существует конечной двумерной функции плотности. Однако несобственное нормальное распределение

можно всегда рассматривать как предел последовательности собственных нормальных распределений. Чтобы убедиться в этом, рассмотрим последовательность собственных нормальных распределений, соответствующих данным значениям m_1 и m_2 и последовательности положительно определенных форм $Q_v(t, u) = Q(t, u) + \epsilon_v^2(t^2 + u^2)$, где $\epsilon_v \rightarrow 0$. Соответствующие характеристические функции стремятся, конечно, к пределу (21.12.2), и, согласно теореме непрерывности (см. параграф 10.7), эти собственные распределения стремятся тогда к данному несобственному распределению.

Рассмотрим какое-нибудь несобственное нормальное распределение с матрицей вторых моментов M ранга 1. Согласно параграфу 21.8, соответствующие величины ξ и η с вероятностью, равной единице, могут быть представлены как линейные функции одной величины X . Обратно, X есть линейная функция от ξ и η , и характеристическая функция величины X имеет вид $e^{mt - \frac{1}{2} \sigma^2 t^2}$, так что величина X распределена нормально. Случай, когда матрица M имеет ранг 0, можно рассматривать как предельный случай $\sigma = 0$. Таким образом, имеет место следующий результат:

Двумерное несобственное нормальное распределение можно рассматривать как обычное одномерное нормальное распределение на некоторой прямой линии.

При $m_1 = m_2 = 0$ из (12.6.8) получаем следующее разложение нормальной плотности вероятности по степеням ρ :

$$(21.12.5) \quad f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_1^2} - \frac{2\rho xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)} = \\ = \frac{1}{\sigma_1\sigma_2} \sum_0^\infty \frac{\Phi^{(v+1)}\left(\frac{x}{\sigma_1}\right)\Phi^{(v+1)}\left(\frac{y}{\sigma_2}\right)}{v!} \rho^v.$$

Этот ряд можно проинтегрировать почленно и получить соответствующее разложение нормальной функции распределения

$$(21.12.6) \quad \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv = \sum_0^\infty \frac{\Phi^{(v)}\left(\frac{x}{\sigma_1}\right)\Phi^{(v)}\left(\frac{y}{\sigma_2}\right)}{v!} \rho^v$$

При $x = y = 0$ из (21.12.5) получим

$$\sum_0^\infty \frac{[\Phi^{(v+1)}(0)]^2}{v!} \rho^v = \frac{1}{2\pi\sqrt{1-\rho^2}},$$

и, следовательно, после интегрирования по ρ

$$\sum_{v=1}^{\infty} \frac{[\Phi^{(v)}(0)]^2}{v!} \rho^v = \frac{1}{2\pi} \int_0^\rho \frac{dr}{\sqrt{1-r^2}} = \frac{1}{2\pi} \arcsin \rho.$$

Теперь соотношение (21.12.6) дает

$$\int_{-\infty}^0 \int_{-\infty}^0 f(u, v) du dv = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho.$$

Из симметрии функции плотности $f(x, y)$ следует, что в первой и третьей четвертях плоскости (x, y) заключено по $\frac{1}{4} + \frac{1}{2\pi} \arcsin \rho$ массы в каждой, а во второй и четвертых заключены массы $\frac{1}{4} - \frac{1}{2\pi} \arcsin \rho$. Эти соотношения были найдены Стильттьесом [220] и Шеппартом [211].

ГЛАВА 22

ОБЩИЕ СВОЙСТВА РАСПРЕДЕЛЕНИЙ В R_n

22.1. Два простых типа распределений. Условные распределения. Совместное распределение вероятностей n одномерных случайных величин ξ_1, \dots, ξ_n (см. параграф 14.2) есть распределение в n -мерном пространстве R_n величины $x = (\xi_1, \dots, \xi_n)$.

Вероятностная функция (см. параграф 8.4) этого распределения есть функция множества $P(S) = P(x \subset S)$, которая для любого множества S из R_n представляет вероятность соотношения $x \subset S$. *Функция распределения* есть функция от n действительных переменных, определенная соотношением (8.3.1):

$$F(x_1, \dots, x_n) = P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n).$$

Распределение однозначно определяется как функцией P , так и функцией F .

Как и раньше, мы будем часто использовать механическую интерпретацию распределения вероятностей в виде распределения единицы массы по пространству R_n . Если выделить группу из k величин $\xi_{v_1}, \dots, \xi_{v_k}$ и спроектировать массу исходного n -мерного распределения на k -мерное пространство выделенных величин, то получим (см. параграф 8.4) k -мерное частное распределение величин $\xi_{v_1}, \dots, \xi_{v_k}$. Соответствующая частная функция распределения получается, как и в двумерном случае, если

остальные $n - k$ переменных в F положить равными $+\infty$. Так, например, частная функция распределения величины ξ_1 есть $F_1(x) = F(x, \infty, \dots, \infty)$. Аналогичным образом получается частная функция распределения любой величины ξ_i .

Как и в случаях $n = 1$ и 2 (см. параграфы 15.2 и 21.1) рассмотрим два простых типа распределений: *дискретный* и *непрерывный*. Их определения и свойства совершенно аналогичны приведенным в параграфе 21.1, мы ограничимся здесь лишь несколькими краткими замечаниями.

Для распределения *дискретного* типа на каждой оси ξ_i имеем конечное или счетное множество точек x_{i_1}, x_{i_2}, \dots сосредоточения массы частного распределения величины ξ_i . Тогда вся масса n -мерного распределения величины $x = (\xi_1, \dots, \xi_n)$ заключена в точках сосредоточения массы $(x_{i_1}, \dots, x_{i_n})$, каждая из которых содержит массу $p_{i_1 \dots i_n} \geq 0$, так что

$$P(\xi_1 = x_{i_1}, \dots, \xi_n = x_{i_n}) = p_{i_1 \dots i_n},$$

$$\sum_{i_1, \dots, i_n} p_{i_1 \dots i_n} = 1.$$

Частное распределение любой группы из k величин также принадлежит дискретному типу и соответствующие значения p получаются, так же как и в (21.1.2) и (21.1.3), суммированием $p_{i_1 \dots i_n}$ по всем значениям остальных $n - k$ величин.

Для распределения *непрерывного* типа, функция распределения всюду непрерывна и *плотность вероятности* или *функция плотности* (см. параграф 8.4)

$$f(x_1, \dots, x_n) = \frac{\partial^n F}{\partial x_1 \dots \partial x_n}$$

существует и непрерывна всюду, за исключением, может быть, некоторых точек, принадлежащих конечному числу гиперповерхностей в R_n . Дифференциал $f(x_1, \dots, x_n) dx_1 \dots dx_n$ называется *элементом вероятности* (см. параграф 15.1) нашего распределения. Плотность вероятности частного распределения любой группы из k величин получается интегрированием $f(x_1, \dots, x_n)$ по остальным $n - k$ величинам, как это было показано для двумерного случая соотношениями (21.1.5) и (21.1.6).

Если величины ξ_1, \dots, ξ_n имеют распределение непрерывного типа, то *условная плотность вероятности* величин ξ_1, \dots, ξ_k при

условиях $\xi_{k+1} = x_{k+1}, \dots, \xi_n = x_n$ задается выражением, обобщающим (21.4.10):

$$(22.1.1) \quad f(x_1, \dots, x_k | x_{k+1}, \dots, x_n) = \frac{f(x_1, \dots, x_n)}{\int\limits_{-\infty}^{\infty} \dots \int\limits_{-\infty}^{\infty} f(\xi_1, \dots, \xi_k, x_{k+1}, \dots, x_n) d\xi_1 \dots d\xi_k}.$$

Наконец, рассмотрим две величины $x = (\xi_1, \dots, \xi_m)$ и $y = (\eta_1, \dots, \eta_n)$, такие, что $(m+n)$ -мерная составная величина (x, y) имеет распределение непрерывного типа. Тогда, обобщая (21.1.7), находим, что для независимости величин x и y необходимо и достаточно, чтобы

$$(22.1.2) \quad f(x_1, \dots, x_m, y_1, \dots, y_n) = f_1(x_1, \dots, x_m) f_2(y_1, \dots, y_n),$$

где f, f_1 и f_2 суть соответственно плотности вероятностей величин (x, y) , x и y . Это условие непосредственно обобщается на случай произвольного числа величин x, y, \dots

22.2. Замена переменных в непрерывном распределении. Пусть $x = (\xi_1, \dots, \xi_n)$ — случайная величина в R_n . Рассмотрим m функций

$$(22.2.1) \quad \eta_i = g_i(\xi_1, \dots, \xi_n), \quad (i = 1, 2, \dots, m),$$

где m не обязательно равно n . Тогда, согласно параграфу 14.5, вектор $y = (\eta_1, \dots, \eta_m)$ представляет собой случайную величину в m -мерном пространстве R_m с распределением вероятностей, которое однозначно задается распределением величины x .

Рассмотрим здесь только частный случай, когда $m = n$ и распределение величины x принадлежит к непрерывному типу. Если функции g_i удовлетворяют некоторым условиям, то, как мы сейчас покажем, распределение величины y может быть точно определено.

Предположим, что для всех x , таких, что плотность вероятности $f(x_1, \dots, x_n)$ отлична от нуля, выполнены следующие условия:

А) Функции g_i всюду однозначны и непрерывны и, кроме того, имеют непрерывные частные производные $\frac{\partial \eta_i}{\partial \xi_k}$ во всех точках x , за исключением, может быть, некоторых точек, принадлежащих к конечному числу гиперповерхностей.

Б) Соотношения (22.2.1), в которых мы полагаем $m = n$, определяют взаимооднозначное соответствие между точками $x = (\xi_1, \dots, \xi_n)$, и $y = (\eta_1, \dots, \eta_n)$, так что, обратно, $\xi_i = h_i(\eta_1, \dots, \eta_n)$ при $i = 1, \dots, n$, где функции h_i однозначны.

Рассмотрим точку x , не принадлежащую ни к какой из „особых“ гиперповерхностей, в которой якобиан $\frac{\partial(\eta_1, \dots, \eta_n)}{\partial(\xi_1, \dots, \xi_n)} = \left| \frac{\partial \eta_i}{\partial \xi_k} \right|$ отличен от нуля. Тогда якобиан обратного преобразования $J = \frac{\partial(\xi_1, \dots, \xi_n)}{\partial(\eta_1, \dots, \eta_n)} = \left| \frac{\partial \xi_i}{\partial \eta_k} \right|$ конечен в точке y , соответствующей точке x , так как

$$\frac{\partial(\eta_1, \dots, \eta_n)}{\partial(\xi_1, \dots, \xi_n)} \cdot \frac{\partial(\xi_1, \dots, \xi_n)}{\partial(\eta_1, \dots, \eta_n)} = 1.$$

Если S — достаточно малая окрестность точки x , а T — соответствующее ей множество в y -пространстве, то J конечен во всех точках T , и мы имеем

$$(22.2.2) \quad P(S) = \int_S f(x_1, \dots, x_n) dx_1 \dots dx_n = \\ = \int_T f(x_1, \dots, x_n) |J| dy_1 \dots dy_n,$$

где в последнем интеграле x_i должны быть заменены их выражениями $x_i = h_i(y_1, \dots, y_n)$ через y_j .

Элемент вероятности распределения величины x преобразовывается, таким образом, согласно равенству

$$(22.2.3) \quad f(x_1, \dots, x_n) dx_1 \dots dx_n = f(x_1, \dots, x_n) |J| dy_1 \dots dy_n,$$

где в правой части $x_i = h_i(y_1, \dots, y_n)$. Плотность вероятности новой величины $y = (\eta_1, \dots, \eta_n)$ равна, следовательно, $f(x_1, \dots, x_n) |J|$.

Если $n=1$ и преобразование $\eta = g(\xi)$ или $\xi = h(\eta)$ однозначно в обоих направлениях, то (22.2.3) сводится к

$$f(x) dx = f[h(y)] |h'(y)| dy,$$

где коэффициент при dy есть плотность вероятности величины η . Примером может служить выражение (15.1.2), относящееся к линейному преобразованию $\eta = a\xi + b$ или $\xi = \frac{\eta - b}{a}$.

Предположим теперь, что условие В) не выполнено. Каждой точке x попрежнему соответствует одна и только одна точка y , но обратное преобразование не однозначно: некоторой точке y может соответствовать несколько точек x . В этом случае нужно разделить x -пространство на несколько частей, таких, что в каждой части соответствие однозначно в обоих направлениях. Тогда масса, лежащая на множестве T в y -пространстве, равна сумме масс, лежащих в множествах, соответствующих множеству T в каждой из частей x -пространства. Каждая из этих составляющих масс представляется кратным интегралом, который может быть преобразован по формуле (22.2.2). Отсюда сле-

дует, что плотность вероятности величины y может быть выражена в виде $\sum f_y | J_y |$, где сумма распространяется на всевозможные точки x , соответствующие данной точке y , а f_y и J_y — соответствующие значения $f(x_1, \dots, x_n)$ и J .

В случае $n=1$ примером такого рода может служить преобразование $\xi = \xi^y$, рассмотренное в параграфе 15.1. Выражение (15.1.4) для плотности вероятности является, очевидно, частным случаем общего выражения $\sum f_y | J_y |$. Более сложные примеры встречаются в параграфе 29.3.

22.3. Средние значения, моменты. Среднее значение функции $g(\xi_1, \dots, \xi_n)$, интегрируемой на R_n относительно n -мерной вероятностной функции $P(S)$, было определено (см. равенство (15.3.2)) интегралом

$$Eg(\xi_1, \dots, \xi_n) = \int_{R_n} g(x_1, \dots, x_n) dP.$$

Моменты этого распределения (см. параграфы 9.2 и 21.2) определяются так:

$$(22.3.1) \quad \alpha_{v_1 \dots v_n} = E(\xi_1^{v_1} \dots \xi_n^{v_n}) = \int_{R_n} x_1^{v_1} \dots x_n^{v_n} dP,$$

причем $v_1 + \dots + v_n$ есть *порядок* момента. Для моментов первого порядка будем употреблять обозначение

$$m_i = E(\xi_i) = \int_{R_n} x_i dP.$$

Точка $m = (m_1, \dots, m_n)$ есть *центр тяжести* массы n -мерного распределения.

Центральные моменты $\mu_{v_1 \dots v_n}$ или моменты относительно точки m получаются заменой в (22.3.1) степеней $\xi_i^{v_i}$ степенями $(\xi_i - m_i)^{v_i}$. *Центральные моменты второго порядка* играют важную роль в дальнейшем и, если не будет оговорено обратное, мы будем всегда предполагать, что они конечны. При $n > 2$ обычные обозначения для этих моментов неудобны из-за большого числа необходимых индексов. Чтобы упростить запись, введем следующие обозначения, положив

$$(22.3.2) \quad \begin{aligned} \lambda_{ii} &= \sigma_i^2 = E(\xi_i - m_i)^2, \\ \lambda_{ik} &= \rho_{ik} \sigma_i \sigma_k = E((\xi_i - m_i)(\xi_k - m_k)). \end{aligned}$$

Таким образом, λ_{ii} обозначает дисперсию, σ_i — стандартное отклонение величины ξ_i , а λ_{ik} обозначает смешанный второй момент величин ξ_i и ξ_k . Коэффициент корреляции $\rho_{ik} = \frac{\lambda_{ik}}{\sigma_i \sigma_k}$ определен, конечно, только когда σ_i и σ_k положительны.

Очевидно, имеем $\lambda_{ki} = \lambda_{ik}$, $\rho_{ki} = \rho_{ik}$ и $\rho_{ii} = 1$. В частном случае $n = 2$ имеем $\lambda_{11} = \mu_{20}$, $\lambda_{12} = \mu_{11}$, $\lambda_{22} = \mu_{02}$.

Обобщая (21.2.5), находим, что среднее значение

$$(22.3.3) \quad E \left(\sum_1^n t_i (\xi_i - m_i) \right)^2 = \sum_{i,k=1}^n \lambda_{ik} t_i t_k$$

неотрицательно, так что правая часть соотношения (22.3.3) есть неотрицательная квадратичная форма от переменных t_1, \dots, t_n . Матрица этой формы есть *матрица вторых моментов*

$$\Lambda = \begin{Bmatrix} \lambda_{11} \dots \lambda_{1n} \\ \dots \dots \dots \\ \lambda_{n1} \dots \lambda_{nn} \end{Bmatrix},$$

а форма, получаемая заменой $t_i = \frac{u_i}{\sigma_i}$, соответствует *матрице коэффициентов корреляции*

$$P = \begin{Bmatrix} \rho_{11} \dots \rho_{1n} \\ \dots \dots \dots \\ \rho_{n1} \dots \rho_{nn} \end{Bmatrix},$$

которая определена, если только все σ_i положительны.

Таким образом, симметричные матрицы Λ и P обе неотрицательны (см. параграф 11.10). Между Λ и P существует соотношение

$$\Lambda = P \Sigma,$$

где Σ обозначает диагональную матрицу с диагональными элементами $\sigma_1, \dots, \sigma_n$. Согласно параграфу 11.6, матрицы Λ и P имеют одинаковый ранг. Для соответствующих детерминантов $|\Lambda| = |\lambda_{ik}|$ и $P = |\rho_{ik}|$ имеем $|\Lambda| = \sigma_1^2 \dots \sigma_n^2 P$. Из (11.10.3) получим

$$(22.3.4) \quad 0 \leq |\Lambda| \leq \lambda_{11} \dots \lambda_{nn}, \quad 0 \leq P \leq \rho_{11} \dots \rho_{nn} = 1.$$

В частном случае, когда $\lambda_{ik} = 0$ при $i \neq k$, будем говорить, что величины ξ_1, \dots, ξ_n некоррелированы. Матрица вторых моментов Λ в этом случае является диагональной матрицей, и $|\Lambda| = \lambda_{11} \dots \lambda_{nn}$. Если, кроме того, все σ_i положительны, то матрица коэффициентов корреляции P существует и совпадает с единичной матрицей I , так что $P = 1$. Более того, равенства $|\Lambda| = \lambda_{11} \dots \lambda_{nn}$ и $P = 1$ имеют место *только* в случае некоррелированных величин.

22.4. Характеристические функции. Характеристическая функция n -мерной случайной величины $\mathbf{x} = (\xi_1, \dots, \xi_n)$ есть функция вектора $t = (t_1, \dots, t_n)$, определенная как среднее значение функции $e^{it'x}$

$$\varphi(t) = E(e^{it'x}) = \int_{R_n} e^{it'x} dP,$$

где, в соответствии с (11.2.1), $t'x = t_1\xi_1 + \dots + t_n\xi_n$. Свойства характеристической функции двумерной величины (см. параграф 21.3) непосредственно распространяются на случай произвольного n . В частности, в окрестности точки $t = 0$ имеем разложение, обобщающее (21.3.2),

$$(22.4.1) \quad \varphi(t) = e^{itm} \left(1 + \frac{t^2}{2!} \sum_{j,k} \lambda_{jk} t_j t_k + o\left(\sum_j t_j^2\right) \right).$$

При $m = 0$ это сводится к

$$(22.4.2) \quad \varphi(t) = 1 - \frac{1}{2} \sum_{j,k} \lambda_{jk} t_j t_k + o\left(\sum_j t_j^2\right).$$

Семи-инварианты n -мерного распределения определяются посредством разложения $\log \varphi$, так же как и в параграфе 15.10 для случая $n = 1$.

Так же как и в параграфе 21.3, можно показать, что для независимости величин \mathbf{x} и \mathbf{y} необходимо и достаточно, чтобы их совместная характеристическая функция имела вид

$$\varphi(t, u) = \varphi_1(t) \varphi_2(u).$$

Характеристическая функция частного распределения любой группы из k величин, выбранных из ξ_1, \dots, ξ_n , получается из $\varphi(t)$, если положить $t_i = 0$ для всех $n - k$ остальных величин. Таким образом, совместная характеристическая функция величин ξ_1, \dots, ξ_k есть

$$(22.4.3) \quad E(e^{it_1\xi_1 + \dots + t_k\xi_k}) = \varphi(t_1, \dots, t_k, 0, \dots, 0).$$

22.5. Ранг распределения. Ранг распределения в R_n (Фриш [113]; см. также Лукомский [151]) определяется как ранг r матрицы вторых моментов Λ (или как равный ему ранг матрицы коэффициентов корреляции P). Распределение называется *собственным* или *несобственным* в зависимости от того, имеет ли место $r < n$ или $r = n$.

В частном случае $n = 2$, Λ совпадает с матрицей M , рассмотренной в параграфе 21.2. Там было показано, что ранг матрицы M

непосредственно связан с некоторыми свойствами вырождения распределения. Докажем теперь, что аналогичная связь существует и в случае произвольного n .

Распределение в R_n является собственным тогда и только тогда, когда не существует гиперплоскости в R_n , которая содержит всю распределенную массу.

Для того чтобы распределение в R_n имело ранг $r < n$, необходимо и достаточно, чтобы вся распределенная масса принадлежала некоторому линейному множеству L_r размерности r , но не принадлежала бы ни одному линейному множеству размерности, меньшей r .

Очевидно, достаточно доказать вторую часть нашей теоремы, так как первая часть является ее непосредственным следствием. Напомним, что, согласно параграфу 3.4, линейное множество размерности r в R_n определяется $n-r$ независимыми линейными соотношениями между координатами.

Предположим сперва, что задано распределение ранга $r < n$. Квадратичная форма матрицы A

$$(22.5.1) \quad Q(\mathbf{t}) = \sum_{i,k} t_i t_k = E \left(\sum_i t_i (\xi_i - m_i) \right)^2$$

имеет тогда ранг r , и поэтому (см. параграф 11.10) существует ровно $n-r$ линейно независимых векторов $\mathbf{t}_p = (t_1^{(p)}, \dots, t_n^{(p)})$, таких, что $Q(\mathbf{t}_p) = 0$.

Для каждого вектора \mathbf{t}_p соотношение (22.5.1) показывает, что равенство

$$(22.5.2) \quad \sum_i t_i^{(p)} (\xi_i - m_i) = 0$$

должно выполняться с вероятностью $= 1$. Тогда $n-r$ соотношений, соответствующих $n-r$ векторам \mathbf{t}_p , определяют некоторое линейное множество L_r , которое содержит всю распределенную массу; так как любой вектор \mathbf{t} , такой, что $Q(\mathbf{t}) = 0$, должен быть линейной комбинацией векторов \mathbf{t}_p , то не может существовать линейного множества меньшей размерности, обладающего тем же свойством.

Обратно, если известно, что вся распределенная масса принадлежит некоторому линейному множеству L_r , но не существует линейного множества меньшей размерности, обладающего этим свойством, то, прежде всего, очевидно, что L_r проходит через центр тяжести \mathbf{m} , так что каждое из $n-r$ независимых соотношений, определяющих L_r , должно иметь вид (22.5.2). Тогда соответствующее множество коэффициентов $t_i^{(p)}$ определяет, согласно (22.5.1), вектор \mathbf{t}_p , такой, что $Q(\mathbf{t}_p) = 0$.

и так как существует ровно $n-r$ независимых соотношений такого рода, то $Q(t)$ имеет, согласно параграфу 11.10, ранг r , и наша теорема доказана.

Итак, для распределения ранга $r < n$ существует ровно $n-r$ независимых линейных соотношений между величинами ξ_i , которые выполнены с вероятностью, равной единице. Как пример мы можем рассмотреть случай $n=3$. Несобственное распределение в R_3 имеет ранг 2, 1 или 0, соответственно тому, лежит ли вся масса в некоторой плоскости, на некоторой прямой, или в некоторой точке, или соответственно тому, существует ли одно, два или три независимых линейных соотношений между величинами ξ_1, ξ_2, ξ_3 , которые выполняются с вероятностью, равной единице.

22.6. Линейное преобразование величин. Пусть случайные величины ξ_1, \dots, ξ_n обладают данным распределением в R_n , таким, что $m=0$. Рассмотрим линейное преобразование

$$(22.6.1) \quad \tau_i = \sum_{k=1}^n c_{ik} \xi_k \quad (i=1, 2, \dots, m),$$

с матрицей $C = C_{mn} = \{c_{ik}\}$, где m не обязательно равно n . В матричном обозначении (см. параграф 11.3) преобразование (22.6.1) имеет вид $y = Cx$. Это преобразование определяет новую случайную величину $y = (\tau_1, \dots, \tau_m)$ с m -мерным распределением, однозначно определенным данным n -мерным распределением величины x (см. параграфы 14.5 и 22.2).

Очевидно, каждая величина τ_i имеет среднее значение, равное нулю. Далее, полагая $\lambda_{ik} = E(\xi_i \xi_k)$, $\mu_{ik} = E(\tau_i \tau_k)$, получим из (22.6.1)

$$\mu_{ik} = \sum_{r,s=1}^n c_{ir} \lambda_{rs} c_{ks}.$$

Это справедливо даже при $m \neq n$ и показывает, что матрицы вторых моментов $\Lambda = \Lambda_{nn} = \{\lambda_{ik}\}$ и $M = M_{mm} = \{\mu_{ik}\}$ связаны с соотношением (22.6.2)

$$M = C \Lambda C'.$$

Если в характеристической функции $\varphi(t)$ величины x заменить t_1, \dots, t_n новыми переменными u_1, \dots, u_m посредством контрагredientного преобразования (см. (11.7.5)) $t = Cu$, то, согласно (11.7.6), $t'x = u'y$ и, следовательно,

$$(22.6.3) \quad \varphi(t) = E(e^{it'x}) = E(e^{iu'y}) = \psi(u),$$

где $\psi(u) = \psi(u_1, \dots, u_m)$ — характеристическая функция новой случайной величины y .

Из (22.6.2) мы заключаем, основываясь на свойствах ранга матрицы произведения (см. параграф 11.6), что *ранг распределения величины у никогда не превосходит ранга распределения величины x*.

Рассмотрим теперь частный случай $m = n$ и предположим, что матрица $C = C_{mn}$ не вырождена. Тогда, согласно параграфу 11.6, матрицы A и M имеют одинаковый ранг, так что в этом случае преобразование (22.6.1) *не влияет на ранг распределения*. Выберем, в частности, в качестве матрицы C ортогональную матрицу, такую, чтобы преобразованная матрица M была диагональной (см. параграф 11.9). Отсюда будет следовать, что $\mu_{ik} = 0$ при $i \neq k$, так что величины η_1, \dots, η_n некоррелированы (ср. случай $n = 2$, параграф 21.8). В этом случае обратная матрица C^{-1} существует (см. параграф 11.7), и обратное преобразование $x = C^{-1}y$ показывает, что величины ξ_i могут быть представлены как линейные функции величин η_i . Если распределение величины x имеет ранг r , то диагональная матрица M содержит ровно r положительных диагональных элементов, а все другие ее элементы равны нулю. Если $r < n$, то всегда можно занумеровать величины η_i так, чтобы $\mu_{11}, \dots, \mu_{rr}$ были положительны. Тогда при $i = r+1, \dots, n$ будем иметь $\mu_{ii} = E(\eta_i^2) = 0$; отсюда видно, что величина η_i почти всюду равна нулю. Таким образом, получаем следующее обобщение результатов параграфа 21.8:

Если распределение n величин ξ_1, \dots, ξ_n имеет ранг r , то величины ξ_i могут быть с вероятностью, равной единице, представлены в виде линейных функций r некоррелированных величин η_1, \dots, η_r .

Понятие *сходимости по вероятности* (см. параграф 20.3) непосредственно распространяется на случай многомерных величин. Величина $x = (\xi_1, \dots, \xi_n)$ называется сходящейся по вероятности к постоянному вектору $a = (a_1, \dots, a_n)$, если ξ_i сходится по вероятности к a_i при $i = 1, \dots, n$. Нам понадобится следующий аналог теоремы сходимости из параграфа 20.6, который можно установить, непосредственно обобщив доказательство, данное для одномерного случая:

Предположим, что для всех $v = 1, 2, \dots$

$$y_v = Ax_v + z_v,$$

где x_v , y_v и z_v суть n -мерные случайные величины, а A — матрица порядка $n \times n$ с постоянными элементами. Предположим далее, что при $v \rightarrow \infty$ n -мерное распределение величины x_v стремится к неко-

торому предельному распределению, а z , сходится по вероятности к нулю. Тогда предельное распределение величин y , задается линейным преобразованием $y = Ax$, где величина x имеет распределение, к которому стремится распределение величины x .

22.7. Эллипсоид рассеяния. Определение эллипса рассеяния, данное в параграфе 21.10, может быть обобщено на любое число измерений. Пусть величины ξ_1, \dots, ξ_n имеют собственное распределение в R_n с $m=0$ и с центральными моментами второго порядка λ_{ik} . Рассмотрим неотрицательную квадратичную форму

$$q(\xi_1, \dots, \xi_n) = \sum_{i,k} a_{ik} \xi_i \xi_k.$$

Если единица массы распределена равномерно (т. е. так, что плотность вероятности постоянна) по области, ограниченной n -мерным эллипсоидом $q=c^2$, то моменты первого порядка этого распределения, очевидно, будут равны нулю, а моменты второго порядка будут, согласно (11:12.4), равны

$$\frac{c^2}{n+2} \cdot \frac{\Lambda_{ik}}{\Lambda} \quad (i, k = 1, 2, \dots, n).$$

Теперь нужно определить c и a_{ik} так, чтобы эти моменты совпадали с данными моментами λ_{ik} . Легко видеть, что это имеет место при $c^2=n+2$ и

$$a_{ik} = \frac{\Lambda_{ki}}{\Lambda} = \frac{\Lambda_{ik}}{\Lambda}.$$

Таким образом, эллипсоид

$$(22.7.1) \quad q(\xi_1, \dots, \xi_n) = \sum_{i,k} \frac{\Lambda_{ik}}{\Lambda} \xi_i \xi_k = n+2$$

обладает требуемым свойством. Он называется *эллипсоидом рассеяния*, соответствующим данному распределению, и служит геометрическим описанием сопредоточения распределенной массы около начала координат. Изменения в определении, которые нужно сделать в случае произвольного m , очевидны. Если два распределения с общим центром тяжести таковы, что один из эллипсоидов рассеяния лежит целиком в другом, то говорят, что первое распределение является более сопредоточенным, чем последнее.

Квадратичная форма (22.7.1) является обратной к форме

$$Q(\xi_1, \dots, \xi_n) = \sum_{i,k} \lambda_{ik} \xi_i \xi_k.$$

(Так как матрица Λ симметрична, то в обратной матрице можно заменить Λ_{kl} на Λ_{lk} — см. параграф 11.7.) n -мерный объем эллипсоида (22.7.1) равен, согласно (11.12.3),

$$\frac{(n+2)^{n/2} \pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)} V \bar{\Lambda} = \frac{(n+2)^{n/2} \pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)} \sigma_1 \dots \sigma_n V P,$$

где детерминанты $\Lambda = |\lambda_{lk}|$ и $P = |\rho_{lk}|$ положительны, так как распределение собственное. Если $\sigma_1, \dots, \sigma_n$ заданы, то из (22.3.4) следует, что этот объем достигает своего максимума, когда величины некоррелированы ($P = 1$). С другой стороны, объем этот стремится к нулю, если ρ_{lk} стремятся к коэффициентам корреляции некоторого несобственного распределения. Отношение между объемом и его максимальным значением равно $V \bar{P}$; эта величина называется *коэффициентом разброса* данного распределения (Фриш [113]). Его можно рассматривать как меру „невырожденности“ распределения. При $n=2$ имеем $V \bar{P} = V \sqrt{1 - \rho^2}$.

С другой стороны, квадрат объема эллипса пропорционален детерминанту $\Lambda = \sigma_1^2 \dots \sigma_n^2 P$; такое выражение называется *обобщенной дисперсией* распределения (Уилкс [232]). При $n=1$ Λ сводится к обычной дисперсии σ^2 , а при $n=2$ имеем $\Lambda = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$.

Наконец, заметим, что тождественность семейств подобных эллипсов рассеивания и эллипсов инерции, указанная в параграфе 21.10 для двумерного случая, не имеет места при $n > 2$.

ГЛАВА 23

РЕГРЕССИЯ И КОРРЕЛЯЦИЯ В R_n

23.1. Поверхности регрессии. Кривые регрессии, введенные в параграфе 21.5, могут быть обобщены на любое число измерений, если распределение принадлежит к одному из наших двух простых типов. Рассмотрим, например, n величин ξ_1, \dots, ξ_n с распределением непрерывного типа. *Условное среднее значение* величины ξ_1 при условиях $\xi_i = x_i$ ($i = 2, \dots, n$) равно

$$E(\xi_1 | \xi_2 = x_2, \dots, \xi_n = x_n) = m_1(x_2, \dots, x_n) = \frac{\int_{-\infty}^{\infty} x_1 f(x_1, \dots, x_n) dx_1}{\int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1}$$

Геометрическое место точек (m_1, x_2, \dots, x_n) при всевозможных значениях x_2, \dots, x_n есть *поверхность регрессии* для среднего значения величины ξ_1 и имеет уравнение

$$x_1 = m_1(x_2, \dots, x_n),$$

что является непосредственным обобщением (21.5.2).

23.2. Линейная средняя квадратическая регрессия. Рассмотрим теперь n величин ξ_1, \dots, ξ_n с совершенно произвольным распределением, имеющим конечные моменты второго порядка. Для упрощения записи будем предполагать в этой главе, что $m = 0$. Формулы для произвольного центра тяжести будут получаться просто заменой в данных ниже соотношениях ξ_i на $\xi_i - m_i$.

Плоскость средней квадратической регрессии величины ξ_1 относительно ξ_2, \dots, ξ_n определяется как такая гиперплоскость

$$(23.2.1) \quad \xi_1 = \beta_{12 \cdot 34 \dots n} \xi_2 + \beta_{13 \cdot 24 \dots n} \xi_3 + \dots + \beta_{1n \cdot 23 \dots n-1} \xi_n,$$

которая дает наилучшую аппроксимацию n -мерного распределения в том смысле, что среднее значение

$$(23.2.2) \quad E(\xi_1 - \beta_{12 \cdot 34 \dots n} \xi_2 - \dots - \beta_{1n \cdot 23 \dots n-1} \xi_n)^2$$

имеет наименьшее возможное значение. Таким образом, выражение в правой части соотношения (23.2.1) является *наилучшей линейной оценкой* величины ξ_1 величинами ξ_2, \dots, ξ_n в смысле обращения в минимум выражения (23.2.2). Мы можем рассматривать здесь ξ_2, \dots, ξ_n как независимые переменные, а ξ_1 как зависимую переменную, которая приближенно представляется или оценивается линейной комбинацией независимых переменных.

Аналогично определяем плоскость средней квадратической регрессии для любой другой величины ξ_i . В этом случае, конечно, ξ_i занимает место зависимой переменной, а остальные переменные $\xi_1, \dots, \xi_{i-1}, \xi_{i+1}, \dots, \xi_n$ рассматриваются как независимые.

Для *коэффициентов регрессии** мы используем здесь обозначения, введенные Юлом [251]. Каждый коэффициент β имеет n индексов, причем первые два индекса отделены точкой от остальных $n - 2$

*) Они часто называются также *частными коэффициентами регрессии*.

индексов. В первой группе из двух индексов, которые называются *первичными* индексами, первый индекс относится к зависимому переменному, а второй к той из независимых переменных, при которой стоит данный коэффициент. Таким образом, порядок первичных индексов существен. Вторая группа индексов, которые мы будем называть *вторичными*, указывает на номера оставшихся переменных, взятых в произвольном порядке. Иногда, когда это не сможет привести к недоразумению, мы будем опускать эту вторую группу индексов.

Чтобы определить коэффициенты регрессии, дифференцируем выражение (23.2.2) по каждому из $n - 1$ неизвестных коэффициентов β и получаем таким образом $n - 1$ уравнений

$$\begin{aligned}\lambda_{22}\beta_{12} + \lambda_{23}\beta_{13} + \dots + \lambda_{2n}\beta_{1n} &= \lambda_{21}, \\ \lambda_{32}\beta_{12} + \lambda_{33}\beta_{13} + \dots + \lambda_{3n}\beta_{1n} &= \lambda_{31}, \\ \vdots &\quad \vdots \\ \lambda_{n2}\beta_{12} + \lambda_{n3}\beta_{13} + \dots + \lambda_{nn}\beta_{1n} &= \lambda_{n1}.\end{aligned}$$

Вторичные индексы здесь опущены, так что $\beta_{1k}, \beta_{2k}, \dots, \beta_{(k+1)n}$ обозначены просто через β_{ik} . Детерминант этой системы уравнений равен Λ_{11} — алгебраическому дополнению элемента λ_{11} в детерминанте $|\Lambda| = |\lambda_{ik}|$.

Предположим сперва, что распределение величины x — собственное (см. параграф 22.5). Матрица вторых моментов Λ и матрица коэффициентов корреляции P являются тогда положительно-определенными, так что $\Lambda_{11} > 0$, и, согласно (11.8.2), наши уравнения имеют единственное решение

$$(23.2.3) \quad \beta_{ik} = -\frac{\Lambda_{1k}}{\Lambda_{11}} = -\frac{\sigma_i}{\sigma_k} \cdot \frac{P_{ik}}{P_{11}}.$$

Простой перестановкой индексов получим соответствующее выражение

$$(23.2.4) \quad \beta_{ik} = -\frac{\Lambda_{ik}}{\Lambda_{ii}} = -\frac{\sigma_i}{\sigma_k} \cdot \frac{P_{ik}}{P_{ii}}$$

для коэффициента β_{ik} в плоскости регрессии величины ξ_i .

Итак, в собственном распределении плоскость регрессии каждой величины относительно остальных величин определена однозначно, и коэффициенты корреляции задаются соотношением (23.2.4). В частном случае n некоррелированных величин все коэффициенты регрессии равны нулю, так как $\Lambda_{ik} = 0$ при $i \neq k$.

Предположим теперь, что распределение величины x — несобственное с рангом $r < n$. Тогда может иметь место равенство $\Lambda_{ii} = 0$ и соответственно некоторые коэффициенты регрессии могут быть бесконечны или неопределены. В качестве примера рассмотрим случай $n = 3$. Для распределения ранга 2 вся масса лежит в некоторой плоскости. Если эта плоскость не параллельна ни одной из координатных осей, то очевидно, что все три плоскости регрессии совпадают с этой плоскостью, так что все коэффициенты регрессии конечны и однозначно определены. Если же эта плоскость параллельна одной из осей, скажем, оси ξ_1 , то в двумерном частном распределении величин ξ_2 и ξ_3 вся распределенная масса лежит на некоторой прямой. Матрица вторых моментов этого частного распределения имеет детерминант Λ_{11} , и, таким образом, $\Lambda_{11} = 0$. В этом случае можно сказать, что плоскость регрессии величины ξ_1 параллельна оси ξ_1 , так что по крайней мере один из коэффициентов регрессии $\beta_{12.3}$ и $\beta_{13.2}$ бесконечен. Для распределения ранга 1 или 0 вся распределенная масса сосредоточена на некоторой прямой или в некоторой точке. Тогда каждая плоскость регрессии должна проходить через эту линию или эту точку, в остальном оставаясь совершенно неопределенной.

Как и в параграфе 21.6, мы можем показать, что плоскость средней квадратической регрессии (23.2.1) есть также плоскость, дающая наилучшую аппроксимацию поверхности регрессии $x_1 = m_1(x_2, \dots, x_n)$, для всех распределений, для которых последняя существует. Если известно, что поверхность регрессии есть плоскость, то она совпадает с плоскостью средней квадратической регрессии.

Рассмотрим, далее, группу из $h < n$ величин ξ , скажем $\xi_i, \xi_j, \dots, \xi_q$, h -мерное частное распределение этих величин имеет в качестве своей матрицы вторых моментов некоторую подматрицу Λ^* матрицы Λ . Тогда можно образовать плоскость регрессии величины ξ_i относительно ξ_j, \dots, ξ_q , и коэффициенты регрессии будут заданы выражениями, аналогичными (23.2.4), в которых Λ_{ii} и Λ_{ik} заменяются соответствующими алгебраическими дополнениями в детерминанте $|\Lambda^*|$. Если, в частности, рассмотреть группу из $n - 1$ величин $\xi_1, \dots, \xi_{j-1}, \xi_{j+1}, \dots, \xi_n$, то получим

$$(23.2.5) \quad \beta_{ik} = -\frac{\Lambda_{jj.ik}}{\Lambda_{jj.ii}},$$

где $\Lambda_{jj.ik}$ — алгебраическое дополнение элемента λ_{ik} в Λ_{jj} (см. (11.5.3)).

23.3. Остатки, остаточная дисперсия.

Предположим, что $\Lambda_{11} \neq 0$.

Разность

$$(23.3.1) \quad \eta_{1.23\dots n} = \xi_1 - \beta_{12}\xi_2 - \dots - \beta_{1n}\xi_n,$$

где коэффициенты регрессии β_{1k} заданы соотношением (23.2.3), можно рассматривать как часть величины ξ_1 , остающуюся после вычитания из нее ее наилучшей линейной оценки величинами ξ_2, \dots, ξ_n . Она называется *остатком* величины ξ_1 относительно ξ_2, \dots, ξ_n .

Остаток некоррелирован с любой из "вычитаемых" величин. Действительно, вводя выражение коэффициентов β , имеем

$$(23.3.2) \quad \eta_{1.23\dots n} = \frac{1}{\Lambda_{11}} \sum_1^n \Lambda_{1k} \xi_k.$$

Следовательно, $E(\eta_{1.23\dots n}) = 0$ и

$$(23.3.3) \quad E(\xi_i \eta_{1.23\dots n}) = \frac{1}{\Lambda_{11}} \sum_{k=1}^n \lambda_{ik} \Lambda_{1k} = \begin{cases} \frac{\Lambda}{\Lambda_{11}} & \text{при } i = 1, \\ 0 & \text{при } i = 2, 3, \dots, n. \end{cases}$$

Отсюда следует, что *остаточная дисперсия* $\sigma_{1.23\dots n}^2 = E(\eta_{1.23\dots n}^2)$ задается выражением

$$(23.3.4) \quad \sigma_{1.23\dots n}^2 = E(\xi_1 \eta_{1.23\dots n}) = \frac{\Lambda}{\Lambda_{11}} = \sigma_1^2 \frac{P}{P_{11}}$$

и что остатки $\eta_{1.23\dots n}$ и $\eta_{i,j,k\dots q}$ некоррелированы в предположении, что все индексы i, j, \dots, q встречаются среди вторичных индексов первого остатка.

Остаточную дисперсию $\sigma_{1.23\dots n}^2$ можно рассматривать как меру приближения при представлении ξ_1 посредством линейной комбинации величин ξ_2, \dots, ξ_n . В случае $n = 2$, выражение (23.3.4) сводится, в соответствии с (21.7.2), к $\sigma_{1.2}^2 = \sigma_1^2(1 - \rho^2)$.

23.4. Частная корреляция. Корреляция между величинами ξ_1 и ξ_2 характеризуется коэффициентом корреляции ρ_{12} , который иногда называется также *полным коэффициентом корреляции* ξ_1 и ξ_2 . Если рассматривать ξ_1 и ξ_2 совместно с остальными $n - 2$ величинами ξ_3, \dots, ξ_n , то можно считать, что изменение величин ξ_1 и ξ_2 вызывается в какой-то мере изменением остальных величин. Остатки $\eta_{1.34\dots n}$ и $\eta_{2.34\dots n}$ представляют, согласно предыдущему параграфу, те части величин ξ_1 и ξ_2 , которые остаются после вычитания из них их наилучших линейных оценок величинами ξ_3, \dots, ξ_n . Таким образом, коэф-

ффициент корреляции между этими двумя остатками можно рассматривать как характеристику корреляции между величинами ξ_1 и ξ_2 после устранения изменений, вызванных влиянием ξ_3, \dots, ξ_n . Он называется *частным коэффициентом корреляции* ξ_1 и ξ_2 относительно ξ_3, \dots, ξ_n и обозначается через $\rho_{12.34\dots n}$. (Порядок индексов в каждой из групп, разделенных точкой, конечно не существен.) Таким образом, имеем

$$(23.4.1) \quad \rho_{12.34\dots n} = \frac{E(\eta_{1.34\dots n} \eta_{2.34\dots n})}{\sqrt{E(\eta_{1.34\dots n}^2) E(\eta_{2.34\dots n}^2)}}.$$

Это выражение, будучи обычным коэффициентом корреляции между двумя случайными величинами, должно удовлетворять соотношению $-1 \leq \rho_{12.34\dots n} \leq 1$.

Остатки $\eta_{1.34\dots n}$ и $\eta_{2.34\dots n}$ могут быть выражены в виде, аналогичном (23.3.2), если использовать выражение (23.2.5) для коэффициентов регрессии в группе из $n - 1$ величин. Тогда получим два следующих выражения, аналогичных (23.3.4):

$$E(\eta_{1.34\dots n}^2) = E(\xi_1 \eta_{1.34\dots n}) = \frac{\Lambda_{22}}{\Lambda_{22.11}} = \frac{\Lambda_{22}}{\Lambda_{11.22}},$$

$$E(\eta_{2.34\dots n}^2) = E(\xi_2 \eta_{2.34\dots n}) = \frac{\Lambda_{11}}{\Lambda_{11.22}},$$

и далее

$$\begin{aligned} E(\eta_{1.34\dots n} \eta_{2.34\dots n}) &= E(\xi_1 \xi_2 \eta_{2.34\dots n}) = \\ &= \frac{1}{\Lambda_{11.22}} \sum_k \lambda_{1k} \Lambda_{11.2k} = -\frac{\Lambda_{12}}{\Lambda_{11.22}}. \end{aligned}$$

Подставляя эти выражения в (23.4.1), получим простую формулу

$$(23.4.2) \quad \rho_{12.34\dots n} = -\frac{\Lambda_{12}}{\sqrt{\Lambda_{11} \Lambda_{22}}} = -\frac{P_{12}}{\sqrt{P_{11} P_{22}}}.$$

Перестановкой индексов получим аналогичное выражение для частного коэффициента корреляции любых двух величин ξ_i и ξ_k относительно оставшихся $n - 2$ величин.

Таким образом, любой частный коэффициент корреляции может быть выражен через центральные моменты λ_{ik} или через полные коэф-

фициенты корреляции ρ_{ik} рассматриваемых величин. Так, например, в случае $n=3$

$$(23.4.3) \quad \rho_{12 \cdot 3} = \frac{\rho_{12}\rho_{23} - \rho_{13}\rho_{23}}{\sqrt{(1 - \rho_{13}^2)(1 - \rho_{23}^2)}}.$$

В частном случае некоррелированных величин из (23.4.2) следует, что все частные коэффициенты корреляции, подобно полным коэффициентам корреляции, равны нулю. Так, например, $\rho_{12 \cdot 34 \dots n} = \rho_{12} = 0$. Если же корреляция между величинами существует, то $\rho_{12 \cdot 34 \dots n}$, вообще говоря, отличен от ρ_{12} . Легко видеть, например, из (23.4.3), что ρ_{12} и $\rho_{12 \cdot 3}$ могут иметь различные знаки и что любой из этих коэффициентов может быть равен нулю, в то время как другой отличен от нуля.

Если полные коэффициенты корреляции ρ_{ik} известны, то частные коэффициенты корреляции могут быть вычислены непосредственно из (23.4.2) и аналогичных выражений при других индексах. Численные расчеты могут быть упрощены использованием некоторых рекурентных соотношений, таких, как

$$(23.4.4) \quad \rho_{12 \cdot 34 \dots n} = \frac{\rho_{12 \cdot 34 \dots n-1} - \rho_{1n \cdot 34 \dots n-1}, \rho_{2n \cdot 34 \dots n-1}}{\sqrt{(1 - \rho_{1n \cdot 34 \dots n-1}^2)(1 - \rho_{2n \cdot 34 \dots n-1}^2)}}$$

(см. упражнение 11, стр. 351), которое сходно с (23.4.3). Согласно этому соотношению, любой частный коэффициент корреляции может быть выражен через такие же коэффициенты с меньшим (на единицу) числом вторичных индексов. Исходя из полных коэффициентов ρ_{ik} , мы можем, таким образом, вычислить сперва все частные коэффициенты $\rho_{ij \cdot k}$ с одним вторичным индексом, затем коэффициенты $\rho_{ij \cdot kl}$ с двумя вторичными индексами и т. д.

Если известны полные и частные коэффициенты корреляции, то любая остаточная дисперсия и любой частный коэффициент регрессии могут быть вычислены с помощью следующих соотношений (см. упражнения 12—13, стр. 351):

$$(23.4.5) \quad \begin{aligned} \sigma_{1 \cdot 23 \dots n}^2 &= \sigma_1^2(1 - \rho_{12}^2)(1 - \rho_{13 \cdot 2}^2)(1 - \rho_{14 \cdot 23}^2) \dots (1 - \rho_{1n \cdot 23 \dots n-1}^2), \\ \rho_{12 \cdot 34 \dots n} &= \rho_{12 \cdot 34 \dots n} \frac{\sigma_{1 \cdot 23 \dots n}}{\sigma_{2 \cdot 34 \dots n}}, \end{aligned}$$

которые являются непосредственным обобщением (21.6.9) и (21.6.10) (аналогичные соотношения получаются путем перестановки индексов).

Из последнего равенства (23.4.5) получаем

$$(23.4.6) \quad \rho_{12 \cdot 34 \dots n}^2 = \beta_{12 \cdot 34 \dots n} \beta_{21 \cdot 34 \dots n}.$$

23.5. Сводный коэффициент корреляции. Рассмотрим остаток

$$\eta_{1 \cdot 23 \dots n} = \xi_1 - \beta_{12} \xi_2 - \dots - \beta_{1n} \xi_n = \xi_1 - \xi_1^*,$$

где $\xi_1^* = \beta_{12} \xi_2 + \dots + \beta_{1n} \xi_n$ есть наилучшая линейная оценка ξ_1 величинами ξ_2, \dots, ξ_n . Легко показать, что среди всех линейных комбинаций величин ξ_2, \dots, ξ_n величина ξ_1^* имеет наименьшую корреляцию с ξ_1 ; последняя измеряется, конечно, обычным коэффициентом корреляции. Коэффициент корреляции величин ξ_1 и ξ_1^* можно, таким образом, рассматривать как характеристику корреляции между величиной ξ_1 , с одной стороны, и *всей совокупностью величин* ξ_2, \dots, ξ_n , с другой. Будем называть его *сводным коэффициентом корреляции* между ξ_1 и (ξ_2, \dots, ξ_n) и писать

$$(23.5.1) \quad \rho_{1(23 \dots n)} = \frac{E(\xi_1 \xi_1^*)}{\sqrt{E(\xi_1^2) E(\xi_1^{*2})}}.$$

Опуская вторичные индексы у $\eta_{1 \cdot 23 \dots n}$, имеем, согласно (23.3.3) и (23.3.4),

$$E(\xi_1 \xi_1^*) = E(\xi_1 (\xi_1 - \eta_1)) = \lambda_{11} - \frac{\Lambda}{\Lambda_{11}},$$

$$E(\xi_1^{*2}) = E(\xi_1^2 - 2\xi_1 \eta_1 + \eta_1^2) = \lambda_{11} - \frac{\Lambda}{\Lambda_{11}},$$

и, следовательно,

$$(23.5.2) \quad \rho_{1(23 \dots n)} = \sqrt{1 - \frac{\Lambda}{\lambda_{11} \Lambda_{11}}} = \sqrt{1 - \frac{P}{P_{11}}}.$$

В силу (11.10.2), $\Lambda < \lambda_{11} \Lambda_{11}$, так что $E(\xi_1 \xi_1^*) \geq 0$ и

$$0 \leq \rho_{1(23 \dots n)} \leq 1.$$

При $\rho_{1(23 \dots n)} = 1$ величина ξ_1 „почти несомненно“ равна некоторой линейной комбинации величин ξ_2, \dots, ξ_n . Это означает, что при совместном распределении всех n величин вся масса заключена в некоторой гиперплоскости пространства R_n , так что это распределение несобственно, и, в соответствии с (23.5.2), $\Lambda = P = 0$. С другой

стороны, для собственного распределения из разложения (11.5.3) следует, что

$$\rho_{1(23 \dots n)}^2 = \frac{1}{P_{11}} \sum_{i,k=2}^n P_{11 \cdot ik} \rho_{1i} \rho_{1k},$$

где сумма в правой части есть, согласно параграфу 11.10, положительно-определенная квадратичная форма величин $\rho_{12}, \dots, \rho_{1n}$. Таким образом, $\rho_{1(23 \dots n)} = 0$ тогда и только тогда, когда $\rho_{12} = \dots = \rho_{1n} = 0$, т. е. когда ξ_1 некоррелировано ни с одним ξ_i ($i = 2, 3, \dots, n$).

Для численных вычислений удобно использовать соотношение (см. упражнение 13 на стр. 351)

$$(23.5.3) \quad \rho_{1(23 \dots n)}^2 = 1 - \frac{\sigma_{1 \cdot 23 \dots n}^2}{\sigma_1^2}.$$

23.6. Ортогональная средняя квадратическая регрессия. Ортогональная средняя квадратическая регрессия, рассмотренная в параграфе 21.6, может быть обобщена на любое число величин. Гиперплоскость H , проходящая через центр тяжести $m = 0$ нашего n -мерного распределения, имеет уравнение

$$\beta_1 \xi_1 + \beta_2 \xi_2 + \dots + \beta_n \xi_n = 0,$$

где β_1, \dots, β_n — направляющие косинусы перпендикуляра к этой плоскости, так что $\sum \beta_i^2 = 1$. Квадрат расстояния между H и точкой $x = (\xi_1, \dots, \xi_n)$ равен $d^2 = (\sum \beta_i \xi_i)^2$. Попробуем найти гиперплоскость H , такую, что среднее значение $E(d^2)$ принимает наименьшее возможное значение. Если такая гиперплоскость H существует, то она называется *плоскостью ортогональной средней квадратической регрессии* (см. К. Пирсон [183a]).

Для распределения, ранг которого меньше n , эта задача тривиальна, так как тогда вся масса лежит в некоторой гиперплоскости H , которая дает значение $E(d^2) = 0$. Поэтому мы можем предположить, что величина x имеет собственное распределение, откуда, согласно параграфу 11.9, следует, что все характеристические числа λ_i матрицы вторых моментов Λ положительны. Пусть x_0 — наименьшее из характеристических чисел и пусть a_1, \dots, a_n — решение однородной системы

$$\begin{aligned} (\lambda_{11} - x_0) a_1 + \lambda_{12} a_2 + \dots + \lambda_{1n} a_n &= 0, \\ \lambda_{21} a_1 + (\lambda_{22} - x_0) a_2 + \dots + \lambda_{2n} a_n &= 0, \\ \dots &\dots \\ \lambda_{n1} a_1 + \lambda_{n2} a_2 + \dots + (\lambda_{nn} - x_0) a_n &= 0, \end{aligned}$$

причем a_i не все равны нулю. Согласно параграфу 11.8, такое решение непременно существует, так как $|\Lambda - x_0 I| = 0$. Далее, мы можем, очевидно, предположить, что $\sum a_i^2 = 1$. Тогда гиперплоскость H_0 с уравнением $\sum a_i \xi_i = 0$ обладает необходимыми свойствами. Действительно, пусть d_0 есть расстояние

от точки x до H_0 , а d — расстояние до какой-нибудь другой гиперплоскости $\sum \beta_i \xi_i = 0$. Тогда, полагая $z_i = \beta_i - \alpha_i$ и имея в виду, что $\sum z_i^2 = \sum \beta_i^2 = 1$, получим

$$\begin{aligned} E(d^2) &= \sum_{i,k} \lambda_{ik} (\alpha_i + z_i) (\alpha_k + z_k) = \\ &= \sum_{i,k} \lambda_{ik} \alpha_i \alpha_k + 2 \sum_i z_i \sum_k \lambda_{ik} \alpha_k + \sum_{i,k} \lambda_{ik} z_i z_k = \\ &= E(d_0^2) + 2x_0 \sum_i \alpha_i z_i + \sum_{i,k} \lambda_{ik} z_i z_k = \\ &= E(d_0^2) + \sum_{i,k} (\lambda_{ik} - x_0 \epsilon_{ik}) z_i z_k, \end{aligned}$$

где ϵ_{ik} суть элементы единичной матрицы I . Так как x_0 есть наименьшее из характеристических чисел матрицы Λ , то матрица $\Lambda - x_0 I$, согласно параграфу 11.10, неотрицательна и, следовательно,

$$E(d^2) \geq E(d_0^2).$$

Можно показать, далее, что, если x_0 — *простой* корень некового уравнения матрицы Λ , то плоскость ортогональной регрессии H_0 определяется этим способом однозначно, тогда как, если x_0 — *кратный* корень, то существует бесконечное множество плоскостей, обладающих требуемыми свойствами.

Эти результаты становятся интуитивно ясными, если вспомнить, что, согласно (11.9.6), обратная матрица Λ^{-1} имеет характеристические числа $\frac{1}{\lambda_i}$, так что квадраты главных осей эллипсоида рассеяния (22.7.1) пропорциональны числам λ_i . Это показывает, что плоскость ортогональна средней квадратической регрессии ортогональна к наименьшей оси эллипса рассеяния и, таким образом, определена или не определена однозначно в зависимости от того, существует ли одна такая наименьшая ось или несколько.

Мы можем определить также *прямую L*, дающую *наилучшую аппроксимацию* распределения, в том смысле, что $E(\delta^2)$ становится минимальным, где δ обозначает кратчайшее расстояние между L и точкой x . Можно показать, что эта прямая совпадает с наибольшей осью эллипса рассеяния.

ГЛАВА 24

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

24.1. Характеристическая функция. Как и в двумерном случае (параграфы 21.11 и 21.12), введем сперва характеристическую функцию нормального распределения. Пусть

$$Q(t) = Q(t_1, \dots, t_n) = \sum_{j,k} \lambda_{jk} t_j t_k$$

обозначает неотрицательную квадратичную форму от $t = (t_1, \dots, t_n)$, а $m = (m_1, \dots, m_n)$ — некоторый действительный вектор. Покажем тогда, что функция

$$(24.1.1) \quad \varphi(t) = \varphi(t_1, \dots, t_n) = e^{\sum_j m_j t_j - \frac{1}{2} Q(t_1, \dots, t_n)}$$

является характеристической функцией некоторого распределения в R^n , которое мы будем называть *нормальным распределением*.

Прежде чем перейти к доказательству этого утверждения, которое будет дано в двух следующих параграфах, сделаем несколько вводных замечаний. В матричном обозначении (см. параграфы 11.2 и 11.4) выражение (24.1.1) имеет вид

$$(24.1.2) \quad \varphi(t) = e^{im't - \frac{1}{2} t' \Lambda t}$$

Разложение (22.4.1) показывает, что величины m_j и λ_{jk} , как обычно, играют роль средних значений и центральных моментов второго порядка. Далее, из (22.4.3) следует, что *любое частное распределение нормального распределения само нормально*.

Если матрица вторых моментов $\Lambda = \{\lambda_{jk}\}$ является диагональной, то характеристическая функция (24.1.1) представляется в виде произведения $\varphi_1(t_1) \dots \varphi_n(t_n)$, в котором каждый множитель есть характеристическая функция некоторого одномерного нормального распределения. Таким образом, *п некоррелированных нормально распределенных величин всегда независимы*.

Как и в двумерном случае, мы должны различать два случая соответственно тому, является ли неотрицательная форма Q определенной или полуопределенной. Очевидно, мы можем везде предполагать, что $m = 0$, так как переход к этому случаю от общего влечет за собой только сложение величины $x = (\xi_1, \dots, \xi_n)$ с постоянным вектором. Мы используем те же обозначения для моментов, коэффициентов корреляции и т. д., что и в предыдущих главах.

24.2. Собственное нормальное распределение. Если квадратичная форма Q является положительно-определенной, то обратная форма Q^{-1} существует, и мы имеем (см. 11.7)

$$Q(t) = Q(t_1, \dots, t_n) = \sum_{j,k} \lambda_{jk} t_j t_k,$$

$$Q^{-1}(x) = Q^{-1}(x_1, \dots, x_n) = \sum_{j,k} \frac{\lambda_{jk}}{\Lambda} x_j x_k.$$

(Так как матрица вторых моментов Λ симметрична, то мы вправе писать Λ_{jk} вместо Λ_{kj}). Тогда, согласно (11.12.1б),

$$\frac{1}{(2\pi)^{n/2} \sqrt{\Lambda}} \int_{R_n} e^{i \sum_j t_j x_j - \frac{1}{2} Q^{-1}(x_1, \dots, x_n)} dx_1 \dots dx_n = e^{-\frac{1}{2} Q(t_1, \dots, t_n)}$$

Это показывает, что функция

$$(24.2.1) \quad f(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\Lambda}} e^{-\frac{1}{2\Lambda} \sum_{j,k} \Lambda_{jk} x_j x_k} = \\ = \frac{1}{(2\pi)^{n/2} \sigma_1 \dots \sigma_n \sqrt{P}} e^{-\frac{1}{2P} \sum_{j,k} P_{jk} \frac{x_j}{\sigma_j} \cdot \frac{x_k}{\sigma_k}}$$

есть плотность вероятности в R_n с характеристической функцией

$$(24.2.2) \quad \varphi(t) = e^{-\frac{1}{2} \sum_{j,k} \lambda_{jk} t_j t_k}$$

Заменяя в (24.2.1) x_j на $x_j - m_j$, получим плотность вероятности общего собственного нормального распределения в R_n , характеристическая функция которого задана соотношением (24.1.1). Для этого распределения семейство подобных эллипсоидов

$$\frac{1}{2\Lambda} \sum_{j,k} \Lambda_{jk} (x_j - m_j)(x_k - m_k) = c^2,$$

порождение эллипсом концентрации (22.7.1), суть поверхности равных вероятностей. Плотность вероятности на каждой из этих поверхностей пропорциональна e^{-c^2} .

24.3. Несобственное нормальное распределение. Если неотрицательная форма Q является полуопределенной, то обратной формы не существует, и выражение (24.2.1) для плотности вероятности становится неопределенным. Как и в двумерном случае (см. параграф 21.12), находим, однако, что функция $\varphi(t) = e^{-\frac{1}{2} Q(t)}$ может быть представлена как предел последовательности функций того же типа, но с определенными формами Q_ν . (Можно, например, взять $Q_\nu = Q + \epsilon_\nu^2 \sum_j t_j^2$, где $\epsilon_\nu \rightarrow 0$.) Тогда из теоремы непрерывности (см. параграф 10.7) следует, что соответствующие собственные нормальные распределения стремятся к предельному распределению и что $\varphi(t)$ есть характеристическая функция

этого предельного распределения, которое будет называться *несобственным нормальным распределением*.

Если обозначить ранг полуопределенной формы Q через r , то $r < n$, и матрица Λ вторых моментов величин ξ_1, \dots, ξ_n имеет тоже ранг r . Тогда из параграфа 22.5 следует, что вся распределенная масса лежит в некотором линейном множестве L , размерности r . Далее, согласно параграфу 22.6, величины ξ_1, \dots, ξ_n могут быть с вероятностью, равной 1, представлены в виде линейных функций от r некоррелированных величин η_1, \dots, η_r , которые сами являются линейными функциями величин ξ_j . В следующем параграфе будет показано, что любые линейные функции нормально распределенных величин сами распределены нормально, а из параграфа 24.1 мы знаем, что некоррелированные нормально распределенные величины всегда независимы. Отсюда выводим следующую теорему:

Если n величин ξ_1, \dots, ξ_n имеют нормальное распределение ранга r , то они с вероятностью, равной 1, могут быть представлены в виде линейных функций от r независимых и нормально распределенных величин. Очевидно, что эта теорема справедлива и в случае $r = n$.

24.4. Линейное преобразование нормально распределенных величин. В дальнейшем, говоря о „нормальном распределении“ и „нормально распределенных величинах“, мы будем иметь в виду как собственные, так и несобственные нормальные распределения.

Пусть величина $x = (\xi_1, \dots, \xi_n)$ имеет в R_n нормальное распределение, такое, что $m = 0$. С помощью линейного преобразования (22.6.1) вводим новую случайную величину $y = (\eta_1, \dots, \eta_m)$, где m не обязательно равно n . Тогда в матричном обозначении $y = Cx$, где $C = C_{mn}$. Между матрицами Λ и M вторых моментов величин x и y имеем, согласно (22.6.2), соотношение $M = C\Lambda C'$, которое выполнено, даже если $m \neq 0$.

Попробуем теперь найти характеристическую функцию величины y . Согласно (24.1.2), характеристическая функция величины x в матричном обозначении имеет вид

$$\varphi(t) = E(e^{it'x}) = e^{-\frac{1}{2} t' \Lambda t}$$

Если в этом выражении заменить t переменной u , положив $t = Cu$, то, согласно (22.6.3), получим характеристическую функцию $\phi(u)$

величины y . Таким образом,

$$\phi(u) = E(e^{iu'y}) = e^{-\frac{1}{2} u'Cu} = e^{-\frac{1}{2} u'Mu}$$

Последнее выражение является характеристической функцией нормального распределения в R_m с матрицей вторых моментов M . Таким образом, взятые в любом числе линейные функции нормально распределенных величин сами распределены нормально.

Сделанное в параграфе 24.1 замечание, что частное распределение нормального распределения само нормально, содержится в этом предложении как частный случай.

24.5. Распределение суммы квадратов. В параграфе 18.1 мы изучили распределение суммы $\sum_1^n \xi_v^2$, где величины ξ_v независимы и нормальны (0,1). Это — χ^2 -распределение с n степенями свободы и плотностью вероятности $k_n(x)$, определенной равенством (18.1.3).

В дальнейшем (см. параграфы 30.1—30.3) нам понадобится более общий случай распределения $\sum \xi_v^2$, когда величины ξ_1, \dots, ξ_n распределены нормально со средними значениями, равными нулю, и с матрицей вторых моментов A , все характеристические числа (см. 11.9) которой равны 0 или 1. Предположим, что p характеристических чисел равны 0, а остальные $n-p$ равны 1. Тогда можно найти таксоортогональное преобразование $y = Cx$, заменяющее старые величины $x = (\xi_1, \dots, \xi_n)$ новыми $y = (\eta_1, \dots, \eta_n)$, что преобразованная матрица вторых моментов $M = CAC'$ есть диагональная матрица, первые $n-p$ диагональных элементов которой равны 1, а остальные p — нулю. Тогда новые величины $\eta_1, \dots, \eta_{n-p}$ независимы и нормальны (0,1), а $\eta_{n-p+1}, \dots, \eta_n$ имеют средние значения и дисперсии, равные нулю, и поэтому сами равны нулю с вероятностью 1.

Следовательно, с вероятностью 1 имеем

$$\sum_1^n \xi_v^2 = \sum_1^n \eta_v^2 = \sum_1^{n-p} \eta_v^2.$$

Итак, $\sum_1^n \xi_v^2$ распределена как сумма квадратов $n-p$ независимых величин, которые нормальны (0,1), т. е. $\sum_1^n \xi_v^2$ имеет χ^2 -распределение с $n-p$ степенями свободы и плотностью вероятности $k_{n-p}(x)$.

Рассмотрим, наконец, еще более общий случай *последовательности случайных величин* x', x'', \dots , такой, что распределение общего члена последовательности $x = (\xi_1, \dots, \xi_n)$ стремится к нормальному распределению рассмотренного выше типа. Применяя к характеристической функции величины $\sum_1^n \xi_i^2$ теорему параграфа 10.7 и многомерную форму соотношения (7.5.9), получим, что в пределе сумма квадратов $\sum_1^n \xi_i^2$ имеет χ^2 -распределение с $n - p$ степенями свободы.

24.6. Условные распределения. Пусть n величин ξ_1, \dots, ξ_n имеют собственное нормальное распределение с $m = 0$, плотность вероятности которого задана формулой (24.2.1). Условная плотность вероятности некоторого числа этих величин в предположении, что остальные величины принимают фиксированные значения, задается выражением вида (22.1.1), и легко видеть, что эта плотность непременно будет собственной нормальной плотностью вероятности. В качестве примера рассмотрим условные распределения для одной и двух величин.

Условное распределение одной величины. Условная плотность вероятности величины ξ_1 при условиях $\xi_i = x_i$ ($i = 2, \dots, n$) равна, согласно (22.1.1),

$$f(x_1 | x_2, \dots, x_n) = \frac{e^{-\frac{1}{2\Lambda} \sum_{j \neq 1} \Lambda_{jk} x_j x_k}}{\int_{-\infty}^{\infty} e^{-\frac{1}{2\Lambda} \sum_{j \neq 1} \Lambda_{jk} x_j x_k} dx_1} = \\ = A e^{-\frac{1}{2\Lambda} (\Lambda_{11} x_1^2 + 2 \sum_{k=2}^n \Lambda_{1k} x_1 x_k)} = B e^{-\frac{\Lambda_{11}}{2\Lambda} (x_1 + \sum_{k=2}^n \frac{\Lambda_{1k}}{\Lambda_{11}} x_k)^2}$$

где A и B не зависят от x_1 , но могут зависеть от x_2, \dots, x_n . Последнее выражение есть функция плотности от x_1 , и, следовательно, мы должны иметь $B = \sqrt{\frac{\Lambda_{11}}{2\pi\Lambda}}$, так что условное распределение величины ξ_1 нормально с дисперсией $\frac{\Lambda}{\Lambda_{11}}$ и средним

$$m_1(x_2, \dots, x_n) = -\frac{\Lambda_{12}}{\Lambda_{11}} x_2 - \dots - \frac{\Lambda_{1n}}{\Lambda_{11}} x_n = \beta_{12} x_2 + \dots + \beta_{1n} x_n,$$

где β суть коэффициенты регрессии, заданные соотношением (23.2.3). Таким образом, регрессия линейна, и поэтому (см. параграф 23.2) поверхность регрессии для среднего значения величины ξ_1 совпадает с плоскостью средней квадратической регрессии. Далее замечаем, что условная дисперсия

$\frac{\Lambda}{\Lambda_{11}}$ не зависит от x_2, \dots, x_n и равна остаточной дисперсии $E(\eta_{1 \cdot 23 \dots n}^2)$, заданной соотношением (23.3.4).

Условное распределение двух величин. Условная плотность вероятности величин ξ_1 и ξ_2 равна

$$f(x_1, x_2 | x_3, \dots, x_n) = \frac{e^{-\frac{1}{2\Lambda} \sum \Delta_{jk} x_j x_k}}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2\Lambda} \sum \Delta_{jk} x_j x_k} dx_1 dx_2} = \\ = Ce^{-\frac{1}{2\Lambda} (\Delta_{11} x_1^2 + 2\Delta_{12} x_1 x_2 + \Delta_{22} x_2^2) + Dx_1 + Ex_2},$$

где C, D и E не зависят от x_1 и x_2 . Введем теперь величины

$$s_1^2 = \frac{\Lambda_{22}}{\Lambda_{11+22}}, \quad s_2^2 = \frac{\Lambda_{11}}{\Lambda_{11+22}}, \quad r = -\frac{\Lambda_{12}}{\sqrt{\Lambda_{11}\Lambda_{22}}}.$$

Тогда, согласно (11.7.3), получим

$$\frac{1}{(1-r^2)s_1^2} = \frac{\Lambda_{11}\Lambda_{11+22}}{\Lambda_{11}\Lambda_{22} - \Lambda_{12}^2} = \frac{\Lambda_{11}}{\Lambda},$$

и аналогично

$$\frac{1}{(1-r^2)s_2^2} = \frac{\Lambda_{22}}{\Lambda}, \quad -\frac{r}{(1-r^2)s_1 s_2} = \frac{\Lambda_{12}}{\Lambda},$$

так что

$$\frac{1}{\Lambda} (\Lambda_{11}x_1^2 + 2\Lambda_{12}x_1 x_2 + \Lambda_{22}x_2^2) = \frac{1}{1-r^2} \left(\frac{x_1^2}{s_1^2} - \frac{2rx_1 x_2}{s_1 s_2} + \frac{x_2^2}{s_2^2} \right).$$

Сравнивая это выражение с выражением двумерной нормальной плотности вероятности в параграфе 21.12, находим, что условное распределение величин ξ_1 и ξ_2 есть собственное нормальное распределение с условными дисперсиями $\frac{\Lambda_{22}}{\Lambda_{11+22}}$ и $\frac{\Lambda_{11}}{\Lambda_{11+22}}$ и с условным коэффициентом корреляции $-\frac{\Lambda_{12}}{\sqrt{\Lambda_{11}\Lambda_{22}}}$. Заметим, что все эти три величины не зависят от x_3, \dots, x_n .

Дисперсии совпадают с дисперсиями остатков $\eta_{1 \cdot 34 \dots n}$ и $\eta_{2 \cdot 34 \dots n}$, изученными в параграфе 23.4, а условный коэффициент корреляции совпадает с частным коэффициентом корреляции $\rho_{12 \cdot 34 \dots n}$, заданным соотношением (23.4.2). Для нормального распределения последний коэффициент имеет важное значение, так как он характеризует не только корреляцию между остатками, но и корреляцию между величинами ξ_1 и ξ_2 при любых фиксированных значениях ξ_3, \dots, ξ_n .

24.7. Сложение независимых величин. Центральная предельная теорема. Сумма двух n -мерных случайных величин $\mathbf{x} = (\xi_1, \dots, \xi_n)$ и $\mathbf{y} = (y_1, \dots, y_n)$ определяется, как и в двумерном случае (см. параграф 21.11), а именно: $\mathbf{x} + \mathbf{y} = (\xi_1 + y_1, \dots, \xi_n + y_n)$. Как и в параграфе 21.11, доказывается, что *характеристическая функция суммы независимых величин есть произведение характеристических функций слагаемых*.

Выражение (24.1.1) для характеристической функции нормального распределения показывает, что *сумма любого числа нормально распределенных и независимых величин сама распределена нормально*, как это доказано в параграфе 17.3 для одномерного случая.

В параграфе 21.11 мы рассматривали сумму большого числа независимых двумерных величин, имеющих одно и то же распределение. Мы доказали, что если эту сумму разделить на квадратный корень из числа слагаемых, то распределение этой нормированной суммы стремится к некоторому нормальному распределению, когда число слагаемых стремится к бесконечности. Непосредственное обобщение доказательства этой теоремы показывает, что эта теорема справедлива для величин любого числа измерений. Это обобщение теоремы Линдеберга–Леви на случай p измерений является простейшим случаем *центральной предельной теоремы* для величин в R^n . Общая форма этой теоремы утверждает, что при некоторых условиях *сумма большого числа независимых n -мерных случайных величин распределена асимптотическиnormally*. Точные условия выполнимости этой теоремы в общем случае, когда слагаемые могут иметь различные распределения, несколько сложнее, и мы не будем здесь углубляться в их выяснение. Довольно общее предложение было найдено Крамером ([11], теорема 21a).

Упражнения к главам 21—24

1. ξ и η суть величины, имеющие конечные моменты второго порядка. Показать, что $D^2(\xi + \eta) = D^2(\xi) + D^2(\eta)$ тогда и только тогда, когда эти величины некоррелированы.

2. Пусть $\varphi_1(t)$, $\varphi_2(t)$ и $\varphi(t)$ суть характеристические функции величин ξ , η и $\xi + \eta$. В параграфе 15.12 было показано, что $\varphi(t) = \varphi_1(t)\varphi_2(t)$, если величины ξ и η независимы. Обратно, если известно, что $\varphi(t) = \varphi_1(t)\varphi_2(t)$ для всех t , то следует ли отсюда, что величины ξ и η независимы? Рассмотрите плотность вероятности $f(x, y) = \frac{1}{4}[1 + xy(x^2 - y^2)]$, $(|x| < 1, |y| < 1)$ и покажите на этом примере, что ответ на поставленный вопрос отрицателен.

3. Рассмотрите разложение (21.3.2) характеристической функции двумерного распределения. Покажите, что если распределение имеет конечные моменты всех порядков, то это разложение можно продолжить до любых степеней t и u . Используя это разложение, показать, что для нормального распределения любой центральный момент μ_{ik} чётного порядка $i+k=2n$ равен коэффициенту при $t^i u^k$ в полиноме

$$\frac{i!k!}{2^n n!} (\mu_{20} t^2 + 2\mu_{11} tu + \mu_{02} u^2)^n.$$

4. Совместное распределение величин ξ и η , нормально со средним значением, равным нулю, и коэффициентом корреляции ρ . Показать, что коэффициент корреляции величин ξ^2 и η^2 равен ρ^2 .

5. Рассмотрим две величины ξ и η с совместным распределением непрерывного типа. Пусть $\varphi(t, u)$ есть совместная характеристическая функция. Используя обозначения параграфа 21.4, имеем

$$\begin{aligned} \left(\frac{\partial^n \varphi}{\partial u^n} \right)_{u=0} &= i^n \int_{-\infty}^{\infty} e^{itx} dx \int_{-\infty}^{\infty} y^n f(x, y) dy = \\ &= i^n \int_{-\infty}^{\infty} e^{itx} E(\eta^n | \xi = x) f_1(x) dx. \end{aligned}$$

Обратно, справедлива формула, аналогичная (10.3.2):

$$E(\eta^n | \xi = x) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} e^{-itx} \left(\frac{\partial^n \varphi}{\partial u^n} \right)_{u=0} dt,$$

если только последний интеграл абсолютно сходится. Используя этот результат, вывести свойства условного среднего и условной дисперсии нормального распределения, данные в параграфе 21.12.

6. Будем использовать те же обозначения, что и в предыдущем упражнении, и предполагать, что величина η неотрицательна. Если интеграл

$$g(x) := \frac{1}{2\pi i} \int_{-\infty}^{\infty} \left(\frac{\partial \varphi}{\partial u} \right)_{u=-tx} dt$$

сходится равномерно относительно x , то он представляет плотность вероятности величины $\frac{\xi}{\eta}$. (Обобщение результата Крамера, доказанного в [11] для частного случая независимых величин ξ и η .) Используя этот результат, вывести распределения 18.2 и 18.3 и обобщить распределение Стьюдента на случай, когда величина ξ в (18.2.1) нормальна (m, σ) , где $m \neq 0$ („нецентральное“ t -распределение).

7. Найти необходимые и достаточные условия того, что три данные числа ρ_{12}, ρ_{13} и ρ_{23} могли быть коэффициентами корреляции некоторого трехмерного

распределения. Найти возможные значения c в частном случае, когда $\rho_{12} = \rho_{13} = \rho_{23} = c$.

8. Каждая из величин x , y и z имеет среднее значение 0 и стандартное отклонение 1. Величины эти удовлетворяют соотношению $ax + by + cz = 0$. Найти матрицу вторых моментов Λ и показать, что

$$a^4 + b^4 + c^4 \leq 2(a^2b^2 + a^2c^2 + b^2c^2).$$

9. Результатом некоторого случайного эксперимента может быть любое из n несовместимых событий E_1, \dots, E_n , причем вероятность наступления события E_j равна $p_j > 0$ и $\sum_1^n p_j = 1$. В ряде из N повторений нашего эксперимента

событие E_j наступает v_j раз и $\sum_1^n v_j = N$. Показать, что вероятность этого

результата равна $\frac{N!}{v_1! \dots v_n!} p_1^{v_1} \dots p_n^{v_n}$. Совместное распределение величин v_1, \dots, v_n , определенное этими вероятностями, есть обобщение биномиального распределения — так называемое *мультиномиальное распределение*. Показать, что для этого распределения $m_j = E(v_j) = Np_j$, $\lambda_{jj} = E(v_j - Np_j)^2 = Np_j(1 - p_j)$, $\lambda_{jk} = E((v_j - Np_j)(v_k - Np_k)) = -Np_j p_k$. Для матрицы вторых моментов Λ имеем $|\Lambda| = 0$ и $\Lambda_{jj} = N^{n-1} p_1 p_2 \dots p_n \neq 0$, так что ранг распределения равен $n - 1$ в соответствии с соотношением $\sum_1^n v_j = N$ между величинами v_j . Показать, что

$$\rho_{12} = -\sqrt{\frac{p_1 p_2}{(1 - p_1)(1 - p_2)}},$$

$$\rho_{12} \cdot \rho_{34} \dots \rho_{jn} = -\sqrt{\frac{p_1 p_2}{(1 - p_1 - p_3 - \dots - p_j)(1 - p_2 - p_4 - \dots - p_j)}},$$

при $j = 3, \dots, n$.

Показать, далее, что совместная характеристическая функция величин

$x_j = \frac{v_j - Np_j}{\sqrt{Np_j}}$ есть $\varphi(t_1, \dots, t_n) = e^{-t\sqrt{N} \sum_1^n \sqrt{p_j}} \left(\sum_1^n p_j e^{\frac{it_j}{\sqrt{Np_j}}} \right)^N$. При $N \rightarrow \infty$ φ стремится к пределу

$$e^{-\frac{1}{2} \left(\sum_1^n t_j^2 - (\sum_1^n t_j \sqrt{p_j})^2 \right)}.$$

Это — характеристическая функция некоторого нормального распределения в R^n . Покажите, что это распределение имеет ранг $n - 1$ и что эти величины удовлетворяют соотношению $\sum_1^n x_j \sqrt{p_j} = 0$. Найти $\rho_{12} \cdot \rho_{34} \dots \rho_{jn}$.

10. Положить в мультиномиальном распределении

$$p_j = \frac{\lambda_j}{N} \quad (j = 1, \dots, n-1) \text{ и } p_n = 1 - \frac{\lambda_1 + \dots + \lambda_{n-1}}{N}.$$

Исследовать предельное распределение при $N \rightarrow \infty$ (многомерное распределение Пуассона).

11. Показать, что остаток $\eta_{1 \cdot 2 \cdot \dots \cdot n}$, определенный соотношением (23.3.1), может быть также интерпретирован как остаток величины $\eta_{1 \cdot 2 \cdot \dots \cdot n-1}$ относительно единственной величины $\eta_{n \cdot 2 \cdot \dots \cdot n-1}$. Показать, что с помощью этого результата формула (23.4.4) для частного коэффициента корреляции может быть выведена из (23.4.3).

12. Используя результат предыдущего упражнения, доказать равенство

$$E(\eta_{1 \cdot 2 \cdot \dots \cdot n}^2) = E(\eta_{1 \cdot 2 \cdot \dots \cdot n-1}^2)(1 - \rho_{1n \cdot 2 \cdot \dots \cdot n-1}^2).$$

Оно показывает, что представление ξ_1 линейной комбинацией величин ξ_1, \dots, ξ_{n-1} может быть улучшено с помощью введения следующей величины ξ_n тогда и только тогда, когда $\rho_{1n \cdot 2 \cdot \dots \cdot n-1} \neq 0$.

13. Доказать соотношения (23.4.5) и (23.5.3).

14. Используя теорему непрерывности (параграф 10.7), доказать следующее предложение: если последовательность нормальных распределений в R_n сходится к некоторому распределению, то предельное распределение нормально. (Заметьте, что, в соответствии с параграфом 24.4, выражение „нормальное распределение“ включает как собственный, так и несобственный случай.)

15. Величины ξ_1, \dots, ξ_n имеют собственное нормальное распределение со средними значениями m_1, \dots, m_n и матрицей вторых моментов Λ . Используя (11.12.3) и заключительное замечание в параграфе 24.2, показать, что величина

$$\eta = \sum_{j,k=1}^n \frac{\Lambda_{jk}}{\Lambda} (\xi_j - m_j)(\xi_k - m_k)$$

имеет χ^2 -распределение с n степенями свободы и плотностью вероятности заданной соотношением (18.1.3).

16. Величины ξ_1, \dots, ξ_n независимы, нормально распределены и имеют одно и то же стандартное отклонение σ , в то время как их средние значения могут быть различны. Ортогональным преобразованием вводятся новые величины η_1, \dots, η_n . С помощью параграфа 24.4 показать, что величины η_i независимы, нормально распределены и имеют все то же самое стандартное отклонение σ , что и ξ_i .

ЧАСТЬ ТРЕТЬЯ

**СТАТИСТИЧЕСКИЕ
ВЫВОДЫ**

ОБЩИЕ ПОНЯТИЯ

ГЛАВА 25

ПРЕДВАРИТЕЛЬНЫЕ ПОНЯТИЯ, ОТНОСЯЩИЕСЯ К ВЫБОРУ *)

25.1. Вводные замечания. В соответствии с тем, что было сказано в главах 13—14, теорию случайных величин и распределений вероятностей, развитую во II части книги, следует рассматривать как систему математических предложений, служащую моделью явления статистической устойчивости, наблюдаемого в связи с последовательностями случайных экспериментов.

Как уже было указано в параграфе 14.6, теперь наша задача выработать методы испытания нашей математической теории на опыте и показать, как можно использовать эту теорию в проблемах получения статистических выводов из эмпирических данных. Эти вопросы и будут предметом III части книги.

Среди совокупностей статистических данных, встречающихся в приложениях, можно обнаружить некоторые классы, для теоретической обработки которых требуются различные методы. В настоящей главе мы вкратце укажем некоторые из таких наиболее важных классов. Следующая глава будет посвящена предварительному обзору принципиальных вопросов, связанных с проверкой и применением теории.

25.2. Простой случайный выбор. Рассмотрим случайный эксперимент \mathcal{E} , связанный с одномерной случайной величиной ξ . Осуществив n независимых повторений эксперимента \mathcal{E} , мы получим последовательность n наблюденных значений величины ξ , которые обозначим x_1, x_2, \dots, x_n :

Такая последовательность, представляющая собой результат n независимых повторений некоторого случайного эксперимента, является представителем простого, но чрезвычайно важного класса статистических данных. Рассматривая статистические данные, относящиеся к этому

*) Мы будем употреблять термин „выбор“ для обозначения процесса выбора, оставляя термин „выборка“ для результата выбора, т. е. для множества выбранных объектов. (Прил. перев.)

классу, мы часто будем употреблять общепринятую терминологию, заимствованную, как мы сейчас покажем, из некоторых частных приложений.

Рассмотрим случайный эксперимент \mathcal{E} такого типа: задано некоторое множество, содержащее конечное число элементов; наш эксперимент заключается в том, что мы выбираем наугад какой-нибудь элемент этого множества, регистрируем значение некоторой определенной характеристики ξ этого элемента и затем возвращаем элемент в множество. Предполагается при этом, что эксперимент организован так, что вероятность быть выбранным одинакова для всех элементов. Мы будем называть заданное множество *генеральной совокупностью*, а элементы этого множества — его *членами* или *индивидуумами* (см. параграф 13.3). Группа индивидуумов, наблюденных при n повторениях эксперимента \mathcal{E} , будет называться *случайной выборкой* из генеральной совокупности, а описанный процесс выбора — *простым случальным выбором*.

Часто мы интересуемся не индивидуумами как таковыми, а только значениями характеристической величины ξ и их распределением среди членов совокупности. В таких случаях будет удобно рассматривать генеральную совокупность, как состоящую не из индивидуумов, а из *значений величины ξ* . Последовательность n наблюденных значений x_1, \dots, x_n будет тогда рассматриваться как случайная выборка из этой совокупности значений ξ . С этой точки зрения мы можем заменить генеральную совокупность урной, содержащей билеты, по одному на каждый член совокупности, с написанными на них соответствующими значениями величины ξ . Эксперимент \mathcal{E} будет тогда заключаться в том, что мы наугад вынимаем билет, отмечаем написанное на нем значение и возвращаем билет обратно в урну.

Так как в урне имеется лишь конечное число билетов, случайная величина ξ будет иметь конечное число возможных значений, так что ее распределение будет дискретного типа (см. параграф 15.2). Однако, полагая число N билетов очень большим, можно сколь угодно точно приблизить это распределение к любому наперед заданному распределению, и если N стремится к бесконечности, то ошибку такого приближения можно заставить стремиться к нулю. Таким образом, мы можем интерпретировать любой случайный эксперимент \mathcal{E} как случайный выбор индивидуума из бесконечной генеральной совокупности (см. параграф 13.3). При этом мы представляем себе урну, содержащую бесконечное количество билетов, на каждом из которых написано некоторое число,

причем распределение этих чисел совпадает с распределением случайной величины ξ , связанной с экспериментом \mathcal{E} . Каждое осуществление эксперимента \mathcal{E} интерпретируется как вынимание билета из урны, а последовательность x_1, \dots, x_n наблюденных значений величины ξ рассматривается как случайная выборка из бесконечной совокупности чисел, написанных на билетах. Значения x_1, \dots, x_n соответственно будут называться *выборочными значениями*.

Необходимо особо подчеркнуть, что распространение идеи выбора на случай бесконечной генеральной совокупности следует рассматривать как простую иллюстрацию случайного эксперимента; мы прибегаем к ней лишь с целью введения удобной терминологии. Такие понятия, как случайный выбор индивидуумов из бесконечной совокупности, ни в коей мере не следует считать частью нашей теории.

Имея в виду эту оговорку, мы будем, однако, часто пользоваться терминологией выбора в указанном выше расширенном смысле. Множество наблюденных значений случайной величины с некоторой функцией распределения $F(x)$ будет, таким образом, часто рассматриваться как *случайная выборка из совокупности, имеющей функцию распределения $F(x)$* или, как мы иногда будем говорить, *случайная выборка из распределения, соответствующего $F(x)$* .

В дальнейшем всякий раз, когда выражения „выборка“ или „выбор“ будут употребляться без дальнейшего уточнения, всегда будет подразумеваться, что мы имеем дело с простым случайным выбором.

Все высказанное выше можно непосредственно распространить на случайные величины любого числа измерений. Тогда каждый индивидуум в нашей воображаемой бесконечной совокупности будет характеризоваться k числами, где k — размерность соответствующей случайной величины, а каждая последовательность наблюденных значений k -мерной случайной величины может интерпретироваться как случайная выборка из k -мерной бесконечной совокупности.

25.3. Распределение выборки. Рассмотрим последовательность n наблюденных значений x_1, \dots, x_n одномерной случайной величины ξ с функцией распределения $F(x)$. В соответствии с предыдущим параграфом, мы можем рассматривать x_1, \dots, x_n как множество выборочных значений, „вынутых“ из генеральной совокупности с функцией распределения $F(x)$. Выборку можно представить геометрически множеством n точек x_1, \dots, x_n на оси x .

Распределение выборки будет тогда определяться как распределение, которое получим, поместив массы, равные $\frac{1}{n}$, в каждой из точек x_1, \dots, x_n . Это дискретное распределение, имеющее n точек с центром тяжести массы (некоторые из которых, конечно, могут совпадать). Соответствующая функция распределения, которую мы будем обозначать $F^*(x)$, является ступенчатой функцией со скачками высоты $\frac{1}{n}$ в каждой точке x_i . Если обозначить через y число выборочных значений, не превосходящих x , то, очевидно,

$$(25.3.1) \quad F^*(x) = \frac{y}{n},$$

так что $F^*(x)$ представляет собой частоту события $\xi \leq x$ в нашей последовательности из n наблюдений.

Очевидно, распределение выборки однозначно определяется самой выборкой. С другой стороны, две выборки, состоящие из одинаковых значений, но полученных в разном порядке, дадут одно и то же распределение. Действительно, распределение определяется лишь положением выборочных значений на оси x , но не их порядком в выборке.

Для определенного таким образом распределения с функцией распределения $F^*(x)$ можно вычислить различные характеристики — моменты, семи-инварианты, коэффициенты асимметрии и эксцесса и т. д., по общим правилам для одномерных распределений, данным в главе 15. Эти характеристики будут называться *моментами*, *семи-инвариантами* и т. д. *выборки*; их следует отличать от соответствующих *характеристик распределения*, соответствующего случайной величине ξ и функции распределения $F(x)$. Последние характеристики будут называться также *моментами*, *семи-инвариантами* и т. д. *генеральной совокупности*.

Так, например, согласно параграфу 15.4, y -й момент *выборки* является средним арифметическим y -х степеней выборочных значений,

$$\int_{-\infty}^{\infty} x^y dF^*(x) = \frac{1}{n} \sum_1^n x_i^y,$$

тогда как соответствующий момент *совокупности* есть $a_y = \int_{-\infty}^{\infty} x^y dF(x)$.

Введенные выше определения непосредственно распространяются на выборки из многомерных совокупностей. Предположим, например,

что мы имеем выборку n пар значений $(x_1, y_1), \dots, (x_n, y_n)$ двумерной случайной величины. Эта выборка может быть представлена геометрически множеством n точек плоскости $(x_1, y_1), \dots, (x_n, y_n)$, и распределением выборки является дискретное распределение, которое получим, поместив в каждой из этих n точек массу, равную $\frac{1}{n}$. Для этого распределения можно вычислить моменты, коэффициенты регрессии и корреляции и другие характеристики по общим правилам для двумерных распределений, данным в главе 21. Эти характеристики будут моментами и т. д. выборки в отличие от соответствующих характеристик распределения (или генеральной совокупности). Обобщение на выборки из совокупностей большего числа измерений очевидно.

Распределение выборки, так же как моменты и другие характеристики этого распределения, в дальнейшем будут играть важную роль. В связи с этим мы введем в параграфе 27.1 особую систему обозначений.

25.4. Выборочные значения как случайные величины. Выборочные распределения. Чтобы получить выборку из n значений одномерной случайной величины с функцией распределения $F(x)$, мы должны осуществить n независимых повторений случайного эксперимента \mathfrak{E} , с которым эта величина связана. Такая последовательность n повторений образует составной эксперимент, дающий n независимых величин x_1, \dots, x_n , где x_i соответствует i -му повторению эксперимента \mathfrak{E} . Выборочные значения x_1, \dots, x_n , представляющие результат этого составного эксперимента, образуют составную n -мерную случайную величину (x_1, \dots, x_n) , причем все x_i суть независимые величины, имеющие одну и ту же функцию распределения $F(x)$. Значения величин x_1, \dots, x_n , наблюденные в фактически осуществленной выборке, образуют наблюденное „значение“ n -мерной случайной величины (x_1, \dots, x_n) .

Если выборочные значения рассматривать как случайные величины, то каждая функция от x_1, \dots, x_n будет, согласно параграфу 14.5, случайной величиной с некоторым распределением, однозначно определяемым совместным распределением величин x_i , т. е. функцией распределения $F(x)$. Каждый момент или любая другая характеристика выборки есть некоторая функция $g(x_1, \dots, x_n)$ выборочных значений. Следовательно, каждая выборочная характеристика является случайной величиной с распределением, однозначно определяемым функцией $F(x)$.

Если из некоторой совокупности повторно извлекать выборки, состоящие из n значений, и для каждой выборки вычислять характеристи-

стике $g(x_1, \dots, x_n)$, то последовательность получаемых таким образом значений образует последовательность наблюденных значений случайной величины $g(x_1, \dots, x_n)$. Распределение вероятностей для этой величины будет называться *выборочным распределением* соответствующей характеристики.

Эти замечания немедленно обобщаются на случай выборок из многомерных совокупностей. В том же смысле, как и выше, выборочные значения будут рассматриваться как случайные величины. Далее, любой момент, коэффициент корреляции или любая другая характеристика такой выборки являются функциями выборочных значений и, таким образом, образуют некоторые случайные величины, распределения которых однозначно задаются распределением совокупности и называются *выборочными распределениями* этих характеристик.

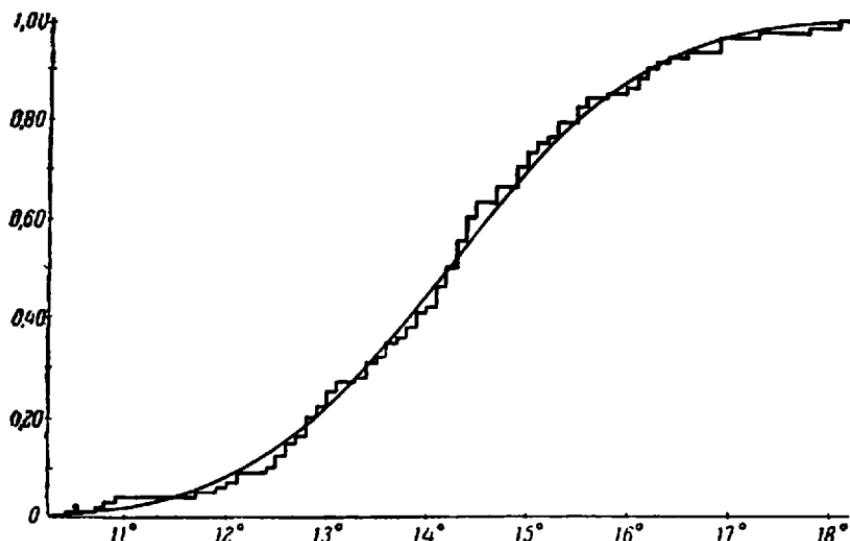
Таким образом, мы можем говорить о выборочном распределении среднего значения выборки, дисперсии, коэффициента корреляции и т. д. Свойства выборочных распределений различных важных выборочных характеристик будут изучены в главах 27—29.

25.5. Статистический аналог для распределения. В качестве примера применения понятий, введенных в предыдущем параграфе, мы рассмотрим функцию распределения $F^*(x)$ одномерной выборки, являющуюся, согласно (25.3.1), функцией от выборочных значений, содержащей переменный параметр x . Как указано в параграфе 25.3, $F^*(x)$ равна частоте события $\xi \leq x$ при n повторениях эксперимента \mathcal{E} . Согласно определению функции распределения $F(x)$ величины ξ , событие $\xi \leq x$ имеет вероятность $F(x)$. Таким образом, из теоремы Бернулли следует, как это было показано в параграфе 20.3, что $F^*(x)$ сходится по вероятности к $F(x)$ при $n \rightarrow \infty$.

Если n велико, то практически несомненно, что функция распределения выборки $F^*(x)$ приблизительно равна функции распределения совокупности $F(x)$. Следовательно, распределение выборки можно рассматривать, как некоторого рода *статистический аналог* для распределения совокупности. График $y = F^*(x)$ ступенчатой функции $F^*(x)$ мы будем называть *эмпирической функцией распределения* выборки. Таким образом, можно ожидать, что при больших значениях n эмпирическая функция распределения выборки будет давать хорошее приближение для кривой $y = F(x)$. В качестве примера мы приводим на фиг. 25 эмпирическую функцию распределения для выборки в сто сред-

них июньских температур в Стокгольме (см. таблицу 30.4.2, стр. 476) вместе с (гипотетической) нормальной функцией распределения соответствующей совокупности.

На практике выборки из непрерывных распределений часто подвергаются *группировке*. При этом индивидуальные выборочные значения не даются, а указывается лишь число выборочных значений, попавших в интервалы некоторого определенного разбиения. Затем

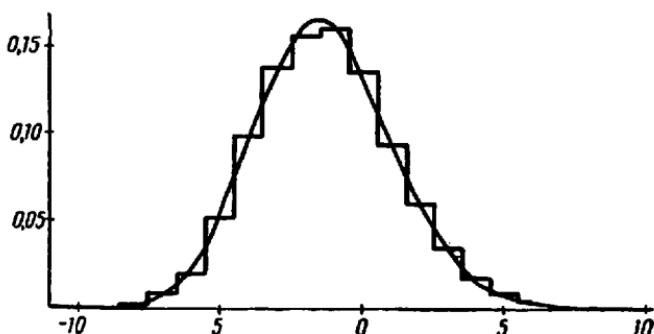


Фиг. 25. Эмпирическая функция распределения для 100 средних июньских температур (в $^{\circ}\text{C}$) в Стокгольме за 1841—1940 гг. и нормальная функция распределения

каждый интервал этого разбиения принимается за основание прямоугольника высоты $\frac{y}{nh}$, где h — длина интервала, а y обозначает число выборочных значений, попавших в этот интервал. Получающаяся при этом фигура называется *гистограммой* выборки. Площадь каждого прямоугольника гистограммы равна частоте $\frac{y}{n}$ соответствующей группы. При больших n можно ожидать, что эта площадь будет приблизительно равна вероятности для наблюденного значения попасть в соответствующий интервал, т. е. будет приблизительно равна интегралу от плотности вероятности $f(x)$, распространенному на данный интервал. Таким образом, верхняя часть контура гистограммы образует статистический

аналог для плотности вероятности, так же как эмпирическая функция распределения является статистическим аналогом для функции распределения. В качестве примера мы приводим на фиг. 26 гистограмму выборки 12 000 размеров бобов, приведенной в таблице 30.4.3, на стр. 477, вместе с (гипотетической) плотностью вероятности соответствующей совокупности, полученной с помощью разложения Эджвортса (см. формулу (17.7.5)).

Аналогичные замечания можно высказать относительно распределения выборки любого числа измерений. Позднее мы обнаружим, что между различными характеристиками распределения выборки и распределения совокупности существует соответствие такого же рода, как



Фиг. 26. Гистограмма для размеров 12 000 бобов и плотность вероятности, найденная с помощью ряда Эджвортса. Шкала по горизонтальной оси соответствует условной нумерации интервалов разбиения.

и в одномерном случае. Действительно, в параграфах 27.3 и 27.8 будет показано, что при весьма общих условиях характеристика выборки сходится по вероятности к характеристике совокупности, когда объем выборки стремится к бесконечности. В таких случаях выборочные характеристики можно рассматривать как *оценки* для соответствующих характеристик совокупности. Систематическое исследование таких оценок и их распределений вероятностей дадут в дальнейшем некоторые из наиболее мощных средств статистических выводов.

25.6. Пристрастный выбор. Таблицы случайных чисел. Когда мы имеем дело с *конечной* генеральной совокупностью, понятие простого случайного выбора имеет точный и конкретный смысл. Мы всегда можем предполагать, что постановка эксперимента удовлетворяет усло-

ниям случайного выбора индивидуумов из такой совокупности с равными шансами для всех индивидуумов, даже тогда, когда практическая реализация этих условий представляет большие трудности. На практике часто при выборе обнаруживается предпочтение некоторым индивидуумам или группам индивидуумов; тогда приходится говорить о *пристрастном выборе*. Опыт показывает, например, что такого рода «пристрастность» следует ожидать всегда, когда выбор индивидуумов из генеральной совокупности зависит от воли человека.

В план настоящей книги не входит рассмотрение вопросов, относящихся к *технике случайного выбора*, например, о мерах, устраниющих, насколько это возможно, пристрастность выбора. Мы лишь заметим, что во многих случаях полезно употреблять *таблицы случайных чисел* (см. [262], [263], [267]). Такая таблица состоит из последовательности целых чисел, представляющих результат простого случайного выбора из совокупности цифр $0, 1, \dots, 9$. Объединяя два столбца этой таблицы, мы получим последовательность чисел, образованных таким же образом из совокупности, содержащей 10^2 чисел $00, \dots, 99$, и аналогично для трех, четырех или любого большего числа столбцов.

Предположим, что мы хотим использовать такую таблицу для получения случайной выборки в 100 индивидуумов из совокупности, состоящей, скажем, из 8183 членов. Тогда сначала следует переименовать члены совокупности, присвоив им порядковые номера от 0000 до 8182, затем следует читать последовательность четырехзначных чисел из таблицы, пропуская числа, превышающие 8182, пока мы не наберем 100 чисел. Тогда наша выборка будет состоять из индивидуумов, занумерованных этими числами. Если осуществляется выбор без возвращения (см. параграф 25.7), то при чтении чисел в таблице мы должны пропускать те числа, которые уже встречались ранее.

Таблицы случайных чисел могут употребляться также для получения выборки наблюдаемых значений случайной величины с данной функцией распределения $F(x)$. Предположим, что функция $F(x)$ табулирована так, что для любого m -значного числа r можно решить относительно a , уравнение $F(a_r) = r \cdot 10^{-m}$. Тогда в нашей таблице случайных чисел мы прочтем последовательность m -значных чисел r и составим выборку значений x так, чтобы значение x , соответствующее числу r , находилось в интервале $a_r < x \leq a_{r+1}$. Таким образом мы получаем сразу *группированную выборку*: выборочные значения не определяются точно, но наш процесс дает число выборочных значений,

принадлежащих каждому интервалу (a_r, a_{r+1}) , причем вероятность попадания какого-нибудь выборочного значения в этот интервал в точности равна выражению

$$F(a_{r+1}) - F(a_r) = 10^{-m}.$$

Чем большее m мы возьмем, тем более частое получается разбиение, и тем точнее определяются выборочные значения. Подробности о таблицах случайных чисел и об их использовании читатель найдет во введениях к таблицам и в двух статьях Кендала и Бебингтона Смита [137].

25.7. Выбор без возвращения. Метод репрезентативной выборки. На практике выбор из конечной совокупности часто совершается таким образом, что вынимаемый индивидуум не возвращается в основную совокупность перед следующим выниманием. Такая последовательность выниманий, очевидно, не носит характера повторений случайного эксперимента при неизменных условиях, так как состав основной совокупности меняется от одного вынимания к другому. При этом мы говорим о *выборе без возвращения*, который следует отличать от простого случайного выбора, являющегося *выбором с возвращением*. Если основная совокупность очень велика и выборка составляет лишь очень малую часть всей совокупности, то, очевидно, различие между этими двумя видами выбора несущественно, и в предельном случае, когда совокупность становится бесконечной, а объем выборки остается конечным, это различие пропадает.

Выбор без возвращения играет важную роль в прикладной статистике. Если необходимо получить сведения о характеристиках некоторой большой совокупности, например, совокупности жителей данной страны, или деревьев в данном районе, или товаров, поставляемых данной фабрикой, и т. п., то часто бывает практически невозможно подвергнуть наблюдению или измерению каждый индивидуум такой совокупности. В таких случаях обычно употребляется метод, известный под названием *метода репрезентативной выборки*: для наблюдения извлекается выборка индивидуумов, причем стараются сделать эту выборку, насколько это возможно, показательной для всей совокупности. В таких случаях обычно применяют выбор без возвращения. Метод выбора может быть *случайным* или *преднамеренным*; в последнем случае мы выбираем индивидуумы, входящие в нашу выборку, обдуманно,

стремясь сделать выборку показательной. Часто употребляются также смешанные методы. По вопросам теории метода репрезентативной выборки мы отсылаем читателя к работе Неймана [161]. Некоторые простые случаи будут рассмотрены в параграфах 34.2 и 34.4.

ГЛАВА 26

СТАТИСТИЧЕСКИЕ ВЫВОДЫ

26.1. Вводные замечания. В параграфе 13.4 было подчеркнуто, что никакая математическая теория не изучает непосредственно вещи, с которыми мы сталкиваемся на опыте. Математическая теория всецело принадлежит к области понятий и оперирует с чисто абстрактными объектами. Однако теория служит моделью некоторой группы явлений реального мира, и абстрактные объекты и предложения теории имеют своими двойниками некоторые наблюдаемые вещи и соотношения между вещами. Для того чтобы такая модель была практически полезна, должна существовать некоторая согласованность между предложениями теории и их эмпирическими дубликатами. Если некоторое предложение имеет своим двойником какое-нибудь непосредственно наблюдаемое соотношение, то мы должны требовать, чтобы наши наблюдения действительно подтверждали наличие этого соотношения. Если такого рода согласованность обнаруживается при повторных проверках, и если мы считаем ее достаточно точной и перманентной, то теорию можно принять для использования на практике.

В настоящей главе мы рассмотрим некоторые вопросы, возникающие при применении общих принципов к математической теории вероятностей. Сначала мы рассмотрим вопрос о проверке согласованности теории с данными опыта, а затем дадим краткий обзор приложений теории для целей получения статистических выводов.

26.2. Согласованность теории с данными опыта. Критерии значимости. Понятие математической вероятности, определенное в параграфе 13.5, имеет своим эмпирическим двойником некоторые непосредственно наблюдаемые частоты. Предложение: «Вероятность события E , связанного со случайным экспериментом \mathcal{E} , равна P » имеет своим двойником, согласно параграфу 13.5, утверждение, называемое *частот-*

ной интерпретацией вероятности P и гласящее: «Практически несомненно, что частота события E в длинном ряду повторений эксперимента \mathcal{E} будет приближенно равна P ».

В соответствии с этим мы должны требовать, чтобы *всякий раз, когда теоретические выводы приводят к определенному численному значению для вероятности некоторого наблюдаемого события, наши наблюдения подтверждали бы истинность соответствующей частотной интерпретации*.

Так, например, если вероятность какого-нибудь события очень мала, то мы должны требовать, чтобы в длинном ряду наблюдений это событие осуществлялось лишь в очень малой доле всех повторений соответствующего эксперимента. Следовательно, мы должны считать практически несомненным, что при одном осуществлении этого эксперимента рассматриваемое событие не произойдет (см. параграф 13.5). Аналогично, если вероятность события очень близка к единице, то мы должны требовать, чтобы с практической несомненностью событие происходило при однократном осуществлении соответствующего эксперимента.

Во многих случаях задача проверки согласованности теории с опытными данными представляется в таком виде: мы располагаем выборкой, состоящей из n наблюденных значений некоторой случайной величины, и хотим узнать, можно ли считать, что распределение вероятностей этой величины имеет некоторые заданные свойства. Иногда гипотетическое распределение точно задано заранее: например, мы можем поставить вопрос, допустимо ли предположение, что наша выборка получена простым случайным выбором из совокупности, имеющей нормальное распределение с $m = 0$ и $\sigma = 1$ (см. параграф 17.2). В других случаях мы задаем некоторый *класс распределений* и спрашиваем, может ли наша выборка быть получена из совокупности с *некоторым распределением*, принадлежащим к данному классу.

Рассмотрим простой случай, в котором гипотетическое распределение полностью задано заранее, например, посредством его функции распределения $F(x)$. Тогда мы должны проверить *статистическую гипотезу* о том, что наша выборка получена из совокупности с этим распределением.

Мы начинаем с предположения, что подлежащая проверке гипотеза верна. Тогда, согласно параграфу 25.5, функция распределения выборки $F^*(x)$ должна служить приближением к данной функции рас-

пределения $F(x)$, если n велико. Определим некоторую неотрицательную меру отклонения F^* от F . Конечно, эту меру можно определить различными способами, но каждая мера отклонения D будет некоторой функцией от выборочных значений и потому, в соответствии с параграфом 25.4, будет иметь определенное выборочное распределение. С помощью этого выборочного распределения мы сможем вычислить вероятность $P(D > D_0)$ того, что отклонение D превышает некоторое данное число D_0 . Эту вероятность можно сделать сколь угодно малой, если выбрать достаточно большое D_0 . Выберем D_0 так, чтобы $P(D > D_0) = \epsilon$, где ϵ столь мало, что мы можем считать практически несомненным, что событие с вероятностью ϵ не произойдет при единичном опыте. Предположим теперь, что нам фактически задана некоторая выборка, состоящая из n значений. По этим значениям мы вычисляем величину D . Если окажется, что $D > D_0$, то это будет означать, что событие, обладающее вероятностью ϵ , произошло. Однако, по нашей гипотезе такое событие практически не может произойти при единичном опыте, и поэтому мы должны притти к выводу, что в этом случае наша гипотеза *опровергнута опытом*. С другой стороны, если обнаружится, что $D \leq D_0$, то мы будем готовы признать нашу гипотезу разумной интерпретацией имеющихся данных, по крайней мере до тех пор, пока это подтверждают дальнейшие опыты.

Мы привели пример рассуждения, очень часто встречающегося в статистических выводах. Нам часто придется иметь дело с более или менее сложной гипотезой о свойствах распределений вероятностей некоторых величин и будет требоваться проверить, согласуются ли с этой гипотезой имеющиеся в нашем распоряжении статистические данные. Первый путь подхода к этой проблеме таков же, как и в рассмотренном выше простейшем случае. Если гипотеза верна, то наши выборочные значения образуют статистический аналог (см. параграф 25.5) для гипотетического распределения, и в соответствии с этим мы вводим подходящую меру D отклонения выборки от этого распределения. С помощью выборочного распределения величины D мы найдем такое число D_0 , чтобы $P(D > D_0) = \epsilon$, где ϵ задается так же, как и раньше. Если в нашем конкретном случае мы получим значение $D > D_0$, то мы скажем, что такое отклонение *значимо*, и будем считать гипотезу *опровергнутой*. Если же $D \leq D_0$, то будем считать, что отклонение D вызвано случайными колебаниями, и статистические данные считаем согласующимися с нашей гипотезой.

Критерий такого общего характера будет называться *критерием значимости* для рассматриваемой гипотезы. В простейшем случае, когда выясняется согласие между распределением множества выборочных значений и теоретическим распределением, мы пользуемся специальным термином и говорим о *критерии согласия*. Вероятность ϵ , которую можно фиксировать произвольным образом, называется *уровнем значимости* критерия.

В случае, когда наша мера отклонения D превышает *предел значимости* D_0 , мы будем считать, что гипотеза опровергнута опытом. Но такое опровержение, конечно, никоим образом не равноценно *логическому* опровержению. Даже если гипотеза верна, событие $D > D_0$, имеющее вероятность ϵ , может произойти в отдельном исключительном случае. Однако, если ϵ достаточно мало, мы вправе *на практике* исключать такую возможность.

С другой стороны, получение одного значения $D \leq D_0$ не является *доказательством* правильности гипотезы. Это лишь показывает, что с точки зрения того частного критерия, к которому мы прибегли, совпадение между теорией и наблюдениями удовлетворительно. Прежде чем считать статистическую гипотезу практически обоснованной, ее следует подвергнуть повторной проверке с помощью различных критерииев.

В главах 30—31 мы будем изучать различные простые критерии значимости и дадим численные примеры их применения. В главе 35 будут подвергнуты критическому анализу общие основания такого рода критериев.

26.3. Описание. В параграфе 13.4 мы разбили приложения всякой математической теории на *описание, анализ и предсказание*. Конечно, между этими тремя рубриками нет резких границ, и сама классификация введена лишь для удобства изложения. Теперь мы вкратце дадим пояснения к некоторым важным приложениям этих трех видов.

Во-первых, теория может быть использована для чисто *описательных* целей. Когда имеется большое множество статистических данных, мы часто интересуемся некоторыми частными свойствами исследуемого явления. Сведения об этих свойствах, которые могут содержаться в множестве статистических данных, желательно получить в сжатой форме, в виде небольшого числа описательных характеристик. Для этих це-

лей можно с выгодой использовать обычные характеристики распределения выборочных значений, как, например, моменты, семи-инварианты, коэффициенты регрессии и корреляции и т. д. Использование графиков плотности вероятности для упорядочения данных, которое играло важную роль в ранней литературе по статистике, также принадлежит в основном к этой группе приложений теории.

Заменяя исходное множество данных небольшим количеством описательных характеристик, мы осуществляем, по терминологии Р. Фишера [13], [89], *сокращение данных*. Очевидно, важно, чтобы при таком сокращении данных по возможности все существенные сведения, относящиеся к исследуемому вопросу и содержащиеся в исходных данных, содержались в выбранных описательных характеристиках. Существенные свойства любой выборочной характеристики выражаются с помощью ее выборочного распределения, следовательно, перед тем, как развивать методы сокращения данных, нам необходимо систематически исследовать выборочные распределения; этой цели будут посвящены главы 27 — 29.

В большинстве случаев, однако, при статистических исследованиях не ставятся чисто описательные задачи. Действительно, описательные характеристики обычно требуются лишь для некоторых определенных целей. Например, когда мы хотим сравнить две различные совокупности данных опыта с помощью характеристик каждой из них или хотим получить оценки для характеристик тех данных, которые мы рассчитываем получить в будущем, то при этом описание наличных данных составляет лишь предварительную стадию исследования, в действительности же мы имеем здесь дело с приложениями, относящимися к одному из двух следующих типов.

26.4. Анализ. Если математическая теория испытана и принята, то ее можно воспользоваться для получения средств научного *анализа* наблюдаемых данных. В нашей области мы можем охарактеризовать этот тип приложений теории как попытку *судить об основной совокупности по выборке*. Мы имеем некоторые статистические данные, которые представляем себе как выборки из каких-то совокупностей, и стараемся использовать эти данные для того, чтобы узнать что-либо о распределении самих совокупностей. В статистической практике встречается множество разнообразных проблем такого рода. В нашем предварительном обзоре мы лишь упомянем некоторые из основ-

ных типов таких проблем, которые будут изучены более подробно в следующих главах.

В параграфе 26.2 мы уже встретились со следующим типом проблем: задана некоторая выборка наблюдаемых значений величины и спрашивается, есть ли основания предполагать, что эта выборка была получена из распределения, принадлежащего к некоторому данному классу. Можно ли говорить, например, что ошибки некоторых физических измерений имеют нормальное распределение или что распределение доходов среди населения некоторого государства следует закону Парето (см. параграф 19.3)? Распределение наличной выборки ни в каком случае не может *точно* совпадать с гипотетическим распределением, так как первое распределение дискретно, а второе непрерывно. Но вправе ли мы приписать отклонение наблюденного распределения от гипотетического случайному колебанию или же мы должны заключить, что это отклонение *значимо*, т. е. указывает на действительное различие между неизвестным нам распределением совокупности и гипотетическим распределением?

В параграфе 26.2 мы видели, как можно подойти к решению этого вопроса с помощью введения *критерия значимости*. Мы должны тогда вычислить некоторую меру отклонения D , и отклонение в конкретном случае будет считаться значимым, если D превышает некоторое заданное значение D_0 ; в противном случае отклонение приписывается случайному колебанию. В других случаях мы предполагаем, что общий характер распределений известен из предыдущих опытов и требуется получить сведения о значениях некоторых частных характеристик этих распределений. Предположим, например, что мы хотим сравнить эффективность двух различных методов лечения одной и той же болезни, причем для каждого метода имеется постоянная вероятность выздоровления. Различны ли эти две вероятности? Чтобы решить это, мы рассмотрим выборки случаев для каждого метода и сравним две частоты выздоровления. Вообще говоря, они будут различны, и перед нами встанет тот же вопрос, что и в предыдущем случае: создается ли эта разность случайными колебаниями или она значима, т. е. указывает на действительное различие между вероятностями?

Подобные, хотя часто и более запутанные задачи встречаются во многих случаях, например, в сельскохозяйственной, промышленной или медицинской статистике, когда требуется сравнить эффективность различных методов лечения или различных способов производства. Мы

рассматриваем средние значения или какие-либо другие характеристики наших выборок и спрашиваем, можно ли приписать различия между наблюденными значениями этих характеристик случайным колебаниям или их следует считать значимыми.

В таких случаях часто бывает полезно начать с гипотезы о том, что эффективность методов одинакова и что в действительности все наши выборки заимствованы из одной и той же совокупности (такое предположение иногда называется *нулевой гипотезой*). Если сделать это предположение, то часто оказывается возможным выработать критерий значимости для разностей между средними значениями или другими интересующими нас характеристиками. Если разности превышают некоторые пределы, то их следует считать значимыми, и мы должны заключить, что существует действительное различие между методами; в противном случае мы будем приписывать эти разности случайным колебаниям.

Этот тип приложений относится к области статистического *анализа причинности*. Предположим, в общем случае, что мы хотим узнать, существует ли какая-либо ощущимая причинная связь между двумя исследуемыми величинами x и y . В качестве первого подхода к этой проблеме мы можем принять нулевую гипотезу, которая в этом случае будет означать, что величины x и y независимы, и попытаемся выработать критерий значимости для этой гипотезы в указанном выше общем смысле. Предположим, например, что мы хотим проследить возможную связь между количествами x и y двух продуктов, потребляемых за год данной группой людей. По выборке наблюденных значений двумерной величины (x, y) мы можем вычислить, например, выборочный коэффициент корреляции r . Вообще говоря, этот коэффициент не будет равен нулю, тогда как по нулевой гипотезе коэффициент корреляции r соответствующего распределения равен нулю. Значима ли полученная разность или ее можно приписать случайным колебаниям? Чтобы ответить на этот вопрос, мы должны выработать критерий значимости, основанный на свойствах выборочного распределения величины r . Если r отличается от нуля значимо, то это можно считать указанием на некоторую зависимость между величинами x и y . Однако обратный вывод был бы незаконным. Даже если значение r для совокупности равно нулю, величины могут быть зависимы (см. параграф 21.7).

Различные критерии значимости, приспособленные к общим проблемам, намеченным выше, будут рассмотрены в главах 30—31.

Критерий значимости, применимый к данной проблеме, всегда можно выбрать многими различными способами. Поэтому возникает важная проблема исследования принципов, на которых основан выбор критерия, сравнения свойств различных критериев и, если это возможно, нахождения критерия, который был бы наиболее действенным для данной цели. Такие вопросы будут изучены в главе 35.

В других проблемах статистического анализа требуется, используя множество выборочных значений, получить *оценки* для различных характеристик совокупности, из которой, по предположению, извлечена данная выборка, а также получить представление о *точности* таких оценок. Простейшей проблемой этого типа является классическая задача *обращения вероятности*: задана частота события E в последовательности повторений случайного эксперимента, и требуется определить, какие выводы могут быть сделаны относительно неизвестного значения p вероятности события E . Совершенно очевидно, что в этом случае за оценку для p можно принять наблюдённую частоту; но можно ли измерить точность этой оценки или высказать какие-либо надежные вероятностные утверждения о разности между оценкой и неизвестным „истинным значением“ p ? Более сложная проблема такого же рода встречается в *теории ошибок*, когда мы располагаем совокупностью измерений величин, связанных с некоторым количеством неизвестных констант, и требуется получить оценки для значений этих констант и определить точность таких оценок. Подобные проблемы встречаются также в связи с методом *регрессии*, который имеет большое значение во многих прикладных областях. В некоторых экономических проблемах, например, предполагается, что существуют какие-то линейные или приблизительно линейные соотношения между величинами, связанными с доходами потребителей и количеством и ценами различных предметов потребления, производящихся или потребляемых на данном рынке. По имеющемуся в нашем распоряжении множеству наблюденных значений этих величин требуется получить оценки для коэффициентов в соотношениях между рассматриваемыми величинами.

Проблему оценки*) можно сформулировать в общем виде следующим образом. Мы рассматриваем случайную величину (какого-нибудь числа

*) Поскольку это не может вызвать недоразумений, мы будем употреблять один и тот же термин „оценка“ как для процесса получения оценки некоторой величины, так и для результата этого процесса. (*Прим. перев.*)

измерений), распределение которой имеет известную математическую форму, но содержит некоторое число неизвестных постоянных параметров. Нам задана выборка наблюденных значений величины и требуется, используя выборочные значения, получить оценки для параметров и определить точность этих оценок. Вообще говоря, существует бесконечное количество различных функций от выборочных значений, которые можно использовать в качестве оценок. Поэтому важно уметь сравнивать свойства различных возможных оценок для одного и того же параметра и, в частности, находить функции (если таковые существуют), которые давали бы оценки *максимальной точности*. Далее, если некоторые оценки уже вычислены, естественно спросить, можно ли высказать какие-либо надежные вероятностные утверждения относительно отклонений этих оценок от неизвестных «истинных значений» параметров. Проблемы такого рода служат предметом *теории оценок*, которой будут посвящены главы 32—34. Наконец, некоторые приложения изложенных теорий будут изложены в главах 36—37.

26.5. Предсказание. Слово „предсказание“ следует понимать здесь очень широко в смысле возможности ответа на вопросы вроде следующих:

Что случится при данных условиях?

Какие последствия можно ожидать, если действовать тем или иным способом?

Как мы должны действовать, чтобы осуществилось некоторое данное событие?

Предсказание, в этом широком смысле, является *практической целью* всякой науки.

Вопросы указанного типа часто встречаются в связи со случайными величинами. Приведем некоторые примеры.

Какое количество браков, рождений и смертей можно ожидать в данной стране в будущем году? Какой эффект можно вызвать определенным повышением или понижением цены такого-то товара? Если заданы результаты некоторых определенных испытаний выборки из партии произведенных товаров, то следует ли эту партию забраковать или выпустить в продажу с гарантией? Какие взносы и премии должно установить страховое общество, чтобы обеспечить себе устойчивый доход? С каким запасом должна быть запланирована новая

телефонная станция, чтобы свести к разумным пределам риск перегрузки линий?

Если предполагать, что для случайных величин в вопросах такого типа известны распределения вероятностей, то можно убедиться, что часто мы будем в состоянии дать хотя бы предварительный ответ на эти вопросы. Однако для полного исследования таких вопросов обычно требуется основательное знакомство с соответствующей прикладной областью. В книге, посвященной общей статистической теории, нет возможности заниматься такими исследованиями.

ВЫБОРОЧНЫЕ РАСПРЕДЕЛЕНИЯ

ГЛАВА 27

ХАРАКТЕРИСТИКИ ВЫБОРОЧНЫХ РАСПРЕДЕЛЕНИЙ

27.1. Обозначения. Рассмотрим одномерную случайную величину ξ с функцией распределения $F(x)$. Для моментов и других характеристик распределения величины ξ мы будем употреблять обозначения, введенные в главе 15. Таким образом, m и σ будут обозначать среднее значение и дисперсию этой случайной величины, a_v , μ_v и χ_v — соответственно момент, центральный момент и семи-инвариант порядка v . В дальнейшем, не оговаривая этого особо, мы будем предполагать эти величины конечными всякий раз, как они понадобятся при выводе наших формул.

Повторяя n раз случайный эксперимент, с которым связана величина ξ , мы получаем последовательность n наблюденных значений x_1, x_2, \dots, x_n нашей случайной величины. Как уже указывалось в параграфе 25.2, мы будем, пользуясь терминами простого случайного выбора, рассматривать множество значений x_1, \dots, x_n как выборку из некоторой совокупности, описываемой функцией распределения $F(x)$. *Распределение выборки* получается (см. параграф 25.3) сосредоточением в каждой точке x_i массы, равной $\frac{1}{n}$, а моменты и другие характеристики выборки определяются как характеристики этого распределения.

Во всех исследованиях относительно характеристик выборки чрезвычайно важно пользоваться ясной и согласованной системой обозначений. В этом отношении мы всюду до конца книги будем следовать, насколько это возможно, следующим трем правилам:

1. Среднее арифметическое какого-либо количества величин, например, x_1, \dots, x_n или y_1, \dots, y_n , будет обозначаться соответствующей буквой с чертой наверху: \bar{x} или \bar{y} .

2. Если некоторая характеристика совокупности (т. е. распределения случайной величины ξ) обозначается обычно греческой буквой, то та же характеристика выборки будет обозначаться соответствующей латинской буквой: s^2 вместо σ^2 , a_v вместо a_v и т. д.

3. В случаях, выходящих за рамки двух предыдущих правил, мы обычно будем обозначать выборочные характеристики той же буквой, что и соответствующую характеристику совокупности, но со звездочкой; так, например, $F^*(x)$ будет обозначать функцию распределения выборки из основной совокупности с функцией распределения $F(x)$.

Таким образом, для среднего значения и дисперсии выборки будем иметь выражения

$$(27.1.1) \quad \bar{x} = \frac{1}{n} \sum_i x_i, \quad s^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2,$$

где суммирование распространяется на все выборочные значения. Моментами a_v и центральными моментами m_v выборки будут

$$(27.1.2) \quad a_v = \frac{1}{n} \sum_i x_i^v, \quad m_v = \frac{1}{n} \sum_i (x_i - \bar{x})^v.$$

Коэффициентами асимметрии и эксцесса выборки будут, в соответствии с (15.8.1) и (15.8.2),

$$(27.1.3) \quad g_1 = \frac{m_3}{m_2^{3/2}}, \quad g_2 = \frac{m_4}{m_2^2} - 3.$$

Соотношения (15.4.4) между моментами и центральными моментами справедливы для любого распределения; в частности, они справедливы, если вместо m , a_v и μ , взять соответствующие выборочные характеристики \bar{x} , a_v и m_v .

Для функции распределения выборки мы имеем уже введенное в (25.3.1) обозначение $F^*(x)$. Аналогично, характеристической функцией выборки*) будет

$$(27.1.4) \quad \varphi^*(t) = \int_{-\infty}^{\infty} e^{itx} dF^*(x) = \frac{1}{n} \sum_i e^{itx_i},$$

а семи-инвариантами выборки будут величины, определяемые в соответствии с (15.10.2) разложением **)

$$(27.1.5) \quad \log \varphi^*(t) = \sum_1^{\infty} \frac{k_v}{v!} (it)^v.$$

*) Тогда, когда возможны недоразумения, мы будем обозначать минимую единицу жирной буквой t .

**) Здесь наши обозначения отличаются от обозначений Р. Фишера [13], который обозначает несмещенную оценку величины x_v через k_v ; мы обозначаем эту оценку через K_v (см. параграф 27.6).

Все моменты и семи-инварианты выборки конечны, и соотношения (15.10.3) — (15.10.5) между моментами и семи-инвариантами остаются справедливыми, если характеристики совокупности заменить выборочными характеристиками.

Те же самые правила будут употребляться и для выборок из многомерных совокупностей. Так, например, если заданы n пар наблюдаемых значений $(x_1, y_1), \dots, (x_n, y_n)$ из двумерного распределения, то мы будем писать (см. параграф 21.2)

$$\bar{x} = \frac{1}{n} \sum_i x_i, \quad \bar{y} = \frac{1}{n} \sum_i y_i,$$

$$m_{20} = s_1^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2,$$

(27.1.6)

$$m_{11} = r s_1 s_2 = \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y}),$$

$$m_{02} = s_2^2 = \frac{1}{n} \sum_i (y_i - \bar{y})^2.$$

В частности, величина r , определяемая соотношением

$$(27.1.7) \quad r = \frac{m_{11}}{s_1 s_2},$$

является коэффициентом корреляции выборки, соответствующим коэффициенту корреляции ρ совокупности. Так как r есть коэффициент корреляции фактического распределения (т. е. распределения выборки), то из параграфа 21.2 следует, что $-1 \leq r \leq 1$. Предельные значения $r = \pm 1$ могут достигаться лишь тогда, если все выборочные точки (x_i, y_i) расположены на одной прямой линии в плоскости x, y .

Для выборки более чем двух измерений мы будем употреблять обозначения, получаемые согласно нашим правилам, из обозначений, введенных в главах 22—23. Так, например, мы будем обозначать через s_i стандартное отклонение выборочного значения i -й величины, а через r_{ij} — коэффициент корреляции между выборочными значениями i -й и j -й величин. Далее, детерминант $|r_{ij}|$ будет обозначаться через R , а коэффициенты регрессии, частные коэффициенты корреляции и т. д. выборки (см. (23.2.3.) и (23.4.2)) — символами

$$b_{12.34\dots k} = -\frac{s_1}{s_2} \cdot \frac{R_{12}}{R_{11}},$$

$$r_{12.34\dots k} = -\frac{R_{12}}{\sqrt{R_{11} R_{22}}}.$$

где k — число измерений, а R_{ij} — алгебраическое дополнение элемента r_{ij} в детерминанте R . Как и прежде, все соотношения между характеристиками, выведенные в части II, остаются справедливыми при замене характеристик совокупности выборочными характеристиками.

Вернемся на один момент к одномерному случаю. В соответствии с параграфом 25.4, каждая характеристика $g(x_1, \dots, x_n)$ наличной выборки может рассматриваться как наблюденное значение случайной величины $g(x_1, \dots, x_n)$, где x_1, \dots, x_n — независимые величины, имеющие то же распределение, что и исходная величина ξ . Распределение случайной величины $g(x_1, \dots, x_n)$ называется *выборочным распределением* характеристики $g(x_1, \dots, x_n)$. Таким образом, мы можем говорить о выборочном распределении среднего значения \bar{x} , дисперсии s^2 и т. д.

Такие же замечания можно сделать относительно выборок любого числа измерений. Каждую выборочную характеристику можно рассматривать как наблюденное значение некоторой случайной величины, распределение которой называется выборочным распределением этой характеристики. Таким образом, мы можем говорить о выборочном распределении коэффициента корреляции r , корреляционного детерминанта R и т. д.

Для каждой выборочной характеристики g мы можем, таким образом, рассматривать ее выборочное распределение и вычислять по этому распределению моменты, семи-инварианты и т. д. Как обычно (см. параграфы 15.3 и 15.6), мы употребляем в этих случаях символы $E(g)$ и $D(g)$ для обозначения среднего значения и стандартного отклонения случайной величины $g = g(x_1, \dots, x_n)$. Далее, если мы имеем дело с некоторой характеристикой распределения величины g (например, с центральным моментом, семи-инвариантом и т. д.) с заданным стандартным обозначением (например μ , или α), то иногда мы будем употреблять то же обозначение этой характеристики, указывая в скобках случайную величину, к которой характеристика относится. Так, например, для центрального момента порядка u выборочной характеристики $g = g(x_1, \dots, x_n)$ мы будем писать

$$\mu_u(g) = E(g - E(g))^u.$$

Подобным же образом, если рассматриваются одновременно две характеристики $f(x_1, \dots, x_n)$ и $g(x_1, \dots, x_n)$, то коэффициент корреляции их совместного выборочного распределения будет обозначаться

СИМВОЛОМ

$$\rho(f, g) = \frac{\mu_1(f, g)}{\sqrt{\mu_2(f) \mu_2(g)}}.$$

Всякий раз, когда мы имеем дело с выборочными распределениями, связанными с данной совокупностью, всегда следует иметь в виду, что выборочные характеристики (\bar{x} , s , m , k , r и т. д.) рассматриваются как случайные величины, в то время как характеристики основной совокупности (m , s , μ , x , ρ и т. д.) суть фиксированные (хотя иногда неизвестные) постоянные.

27.2. Выборочное среднее значение \bar{x} . Рассмотрим одномерную выборку со значениями x_1, \dots, x_n . Считая x_i независимыми случайными величинами, каждая из которых имеет функцию распределения $F(x)$, мы получим

$$E(\bar{x}) = \frac{1}{n} \sum_i E(x_i) = m,$$

(27.2.1)

$$D^2(\bar{x}) = \frac{1}{n^2} \sum_i D^2(x_i) = \frac{\mu_2}{n}$$

Таким образом, случайная величина $\bar{x} = \frac{1}{n} \sum_i x_i$ имеет среднее значение m и дисперсию $\frac{\mu_2}{n}$ и, следовательно, стандартное отклонение $\frac{\sigma}{\sqrt{n}}$.

Тогда из теоремы Чебышева (см. параграф 20.4) немедленно следует, что выборочное среднее значение \bar{x} сходится по вероятности к среднему значению совокупности m , когда n стремится к бесконечности*).

Записывая $\bar{x} - m = \frac{1}{n} \sum (x_i - m)$ и имея в виду, что x_i независимы и что каждая разность $x_i - m$ имеет среднее значение, равное нулю, мы получаем, далее,

$$(27.2.2) \quad \begin{aligned} \mu_s(\bar{x}) &= E(\bar{x} - m)^2 = \frac{1}{n^2} E\left(\sum_i (x_i - m)\right)^2 = \\ &= \frac{1}{n^2} \sum_i E(x_i - m)^2 = \frac{\mu_2}{n^2}, \end{aligned}$$

*) Из менее элементарной теоремы Хинчина (см. параграф 20.5) следует, что указанное свойство имеет место всегда, когда существует среднее значение для совокупности m , даже если μ_2 не имеет конечного значения.

$$\begin{aligned}
 (27.2.2) \quad \mu_4(\bar{x}) &= E(\bar{x} - m)^4 = \frac{1}{n^4} E\left(\sum_i (x_i - m)\right)^4 = \\
 &= \frac{1}{n^4} \sum_i E(x_i - m)^4 + \frac{6}{n^4} \sum_{i < j} E((x_i - m)^2(x_j - m)^2) = \\
 &= \frac{\mu_4}{n^3} + \frac{3(n-1)}{n^8} \mu_2^2 = \frac{3\mu_2^2}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^8}.
 \end{aligned}$$

Старшие центральные моменты величины \bar{x} можно найти посредством аналогичных, хотя и несколько более громоздких вычислений. Таким образом, мы находим

$$\begin{aligned}
 \mu_5(\bar{x}) &= E(\bar{x} - m)^5 = \frac{10\mu_2\mu_3}{n^8} + O\left(\frac{1}{n^4}\right), \\
 \mu_6(\bar{x}) &= E(\bar{x} - m)^6 = \frac{15\mu_2^3}{n^9} + O\left(\frac{1}{n^4}\right),
 \end{aligned}$$

и вообще

$$(27.2.3) \quad E(\bar{x} - m)^{2k-1} = O\left(\frac{1}{n^k}\right), \quad E(\bar{x} - m)^{2k} = O\left(\frac{1}{n^k}\right),$$

В важном частном случае, когда распределение совокупности нормальное (m, σ) , величина \bar{x} , как было указано в параграфе 17.3, также является нормальной со средним значением m и стандартным отклонением $\frac{\sigma}{\sqrt{n}}$. Отсюда следует, что в этом случае все центральные моменты $\mu_v(\bar{x})$ нечетного порядка равны нулю, а три первых центральных момента четного порядка имеют значения

$$\mu_2(\bar{x}) = D^2(\bar{x}) = \frac{\sigma^2}{n}, \quad \mu_4(\bar{x}) = \frac{3\sigma^4}{n^3}, \quad \mu_6(\bar{x}) = \frac{15\sigma^6}{n^5}.$$

27.3. Моменты a_v . Для каждого выборочного момента $a_v = \frac{1}{n} \sum_i x_i^v$ мы получаем, в качестве прямого обобщения формул (27.2.1) и (27.2.3)

$$\begin{aligned}
 (27.3.1) \quad E(a_v) &= \frac{1}{n} \sum_i E(x_i^v) = a_v, \\
 D^2(a_v) &= \frac{1}{n^2} \sum_i D^2(x_i^v) = \\
 &= \frac{1}{n^2} \sum_i (E(x_i^{2v}) - E^2(x_i^v)) = \frac{a_{2v} - a_v^2}{n},
 \end{aligned}$$

$$E(a_v - a_v)^{2k-1} = O\left(\frac{1}{n^k}\right), \quad E(a_v - a_v)^{2k} = O\left(\frac{1}{n^k}\right).$$

В силу теоремы Хинчина (см. параграф 20.5) и первого из этих трех соотношений, выборочный момент a_n сходится по вероятности к a_∞ , при $n \rightarrow \infty$, если только момент совокупности a_∞ существует.

В силу следствия из теоремы параграфа 20.6, каждая рациональная функция или степень рациональной функции от выборочных моментов a_n сходится по вероятности к постоянной, получаемой подстановкой в эту функцию a_∞ , вместо каждого a_n , если только все эти a_n существуют и если получаемая таким образом постоянная конечна.

Отсюда следует, в частности, что все центральные моменты m_n , сечи-инварианты k_n и коэффициенты g_1 и g_2 , определяемые формулой (27.1.3), при $n \rightarrow \infty$ сходятся по вероятности к соответствующим характеристикам совокупности. Таким образом, при больших выборках все эти выборочные характеристики можно рассматривать как оценки для соответствующих характеристик совокупности. Ниже, однако, мы обнаружим, что получаемые таким образом оценки не всегда являются наилучшими из тех оценок, которые можно получить (см. параграфы 27.6 и 33.1).

Каждое среднее значение вида

$$(27.3.2) \quad E(a_n^p a_m^q \dots) = \frac{1}{n^{p+q+\dots}} E\left(\left(\sum_i x_i^n\right)^p \left(\sum_i x_i^m\right)^q \dots\right),$$

где p, q, \dots — целые числа, может быть получено непосредственным хотя часто и громоздким вычислением; при этом следует пользоваться лишь тем, что x_i — независимые величины и $E(x_i^n) = a_n$. В частном случае, когда среднее значение совокупности m равно нулю, a_n совпадают с центральными моментами μ_n . Если среди множителей в (27.3.2) встречается выборочное среднее $a_1 = \bar{x}$, то вычисления упрощаются, так как среднее значение каждого члена, содержащего одно из этих x_i в первой степени, равно нулю.

27.4. Дисперсия m_2 . Центральный выборочный момент $m_2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2$ не зависит от положения начала отсчета на шкале значений исходной случайной величины ξ . Совместив начало отсчета со средним значением совокупности, мы получим $m_2 = 0$. Если мы имеем дело с выборочными распределениями величин m_n , то всегда можно предполагать, что $m_2 = 0$, и добиться упрощения, о котором упоми-

налось в конце предыдущего параграфа. Полученная при этом формула будет справедливой независимо от значения m .

Итак, предположим, что $m = 0$, и рассмотрим выборочную дисперсию $m_2 = s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = a_2 - \bar{x}^2$. Так как $m = 0$, то из формул (27.2.1) и (27.3.1) мы получим

$$(27.4.1) \quad E(m_2) = E(a_2) - E(\bar{x}^2) = \mu_2 - \frac{\mu_2}{n} = \frac{n-1}{n} \mu_2.$$

Далее, мы имеем $m_2^2 = a_2^2 - 2\bar{x}^3 a_2 + \bar{x}^4$. Полагая всюду $m = 0$, мы получим

$$\begin{aligned} E(a_2^2) &= \frac{1}{n^2} E\left(\sum_i x_i^2\right)^2 = \frac{\mu_4 + (n-1)\mu_2^2}{n}, \\ E(\bar{x}^2 a_2) &= \frac{1}{n^3} E\left[\left(\sum_i x_i\right)^2 \sum_i x_i^2\right] = \frac{\mu_4 + (n-1)\mu_2^2}{n^2}, \\ E(\bar{x}^4) &= \frac{1}{n^4} E\left(\sum_i x_i\right)^4 = \frac{\mu_4 + 3(n-1)\mu_2^2}{n^8}, \end{aligned}$$

и отсюда, после приведения подобных членов,

$$\begin{aligned} (27.4.2) \quad E(m_2^2) &= \mu_2^2 + \frac{\mu_4 - 3\mu_2^2}{n} - \frac{2\mu_4 - 5\mu_2^2}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^8}, \\ D^2(m_2) &= E(m_2^2) - E^2(m_2) = \\ &= \frac{\mu_4 - \mu_2^2}{n} - \frac{2(\mu_4 - 2\mu_2^2)}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^8}. \end{aligned}$$

Старшие центральные моменты для m_2 можно получить таким же образом. Вычисления длинны и неинтересны, но не представляют никаких принципиальных трудностей. Мы приведем лишь главные члены третьего и четвертого моментов:

$$\begin{aligned} (27.4.3) \quad \mu_3(m_2) &= E\left(m_2 - \frac{n-1}{n}\mu_2\right)^2 = \frac{\mu_6 - 3\mu_2\mu_4 - 6\mu_3^2 + 2\mu_2^3}{n^2} + O\left(\frac{1}{n^8}\right), \\ \mu_4(m_2) &= E\left(m_2 - \frac{n-1}{n}\mu_2\right)^2 = \frac{3(\mu_4 - \mu_2^2)^2}{n^2} + O\left(\frac{1}{n^8}\right). \end{aligned}$$

Рассмотрим, наконец, смешанный второй момент (см. параграф 21.2) среднего значения \bar{x} и дисперсии m_2 выборки. При любом значении m

этот смешанный момент равен

$$\mu_{11}(\bar{x}, m_2) = E \left((\bar{x} - m) (m_2 - \frac{n-1}{n} \mu_2) \right) = E ((\bar{x} - m) m_2).$$

Так как ясно, что последнее выражение не зависит от положения начала отсчета, мы опять можем предположить, что $m = 0$, и после таких же вычислений, как и выше, получим

$$(27.4.4) \quad \begin{aligned} \mu_{11}(\bar{x}, m_2) &= E(\bar{x} m_2) = E(\bar{x} a_2) - E(\bar{x}^3) = \\ &= \frac{\mu_3}{n} - \frac{\mu_3}{n^2} = \frac{n-1}{n^2} \mu_3. \end{aligned}$$

Для каждого *симметричного* распределения мы имеем $\mu_3 = 0$, так что \bar{x} и m_2 некоррелированы. Позже мы увидим (см. параграф 29.3), что в частном случае *нормальной* совокупности \bar{x} и m_2 не только некоррелированы, но даже *независимы*. Для нормальной совокупности формулы (27.4.1) и (27.4.2) имеют вид

$$(27.4.5) \quad E(m_2) = \frac{n-1}{n} \sigma^2, \quad D^2(m_2) = \frac{2(n-1)}{n^2} \sigma^4.$$

27.5. Старшие центральные моменты и семи-инварианты. Выражения для характеристик выборочных распределений величин m_v и k_v быстро усложняются, когда v становится больше 2; мы укажем лишь немногие сравнительно простые случаи, опустив подробности вычислений. Дальнейшие сведения читатель может почерпнуть, например, из статей Чупрова [227] и Крейга [67].

С помощью таких же вычислений, как и в предыдущих параграфах, мы получаем выражения

$$(27.5.1) \quad \begin{aligned} E(m_3) &= \frac{(n-1)(n-2)}{n^3} \mu_3, \\ E(m_4) &= \frac{(n-1)(n^2-3n+3)}{n^5} \mu_4 + \frac{3(n-1)(2n-3)}{n^3} \mu_2^2. \end{aligned}$$

Для любого m_v мы имеем

$$(27.5.2) \quad m_v = \frac{1}{n} \sum_i (x_i - \bar{x})^v = a_v - \binom{v}{1} \bar{x} a_{v-1} + \binom{v}{2} \bar{x}^2 a_{v-2} - \dots$$

Как и прежде, можно предполагать, что $m = 0$, так что $E(a_v) = \mu_v$, и

$$E(\bar{x} a_{v-1}) = \frac{1}{n^3} E \left(\sum_i x_i \sum_i x_i^{v-1} \right) = \frac{\mu_v}{n}.$$

Для $1 < i \leq v$, используя неравенство Шварца (9.5.1), из формул (27.2.3) и (27.3.1) получаем

$$E^2(\bar{x}^i a_{v-i}) \leq E(\bar{x}^{2i}) E(a_{v-i}^2) = O\left(\frac{1}{n^i}\right)$$

так что $E(\bar{x}^i a_{v-i}) = O(n^{-\frac{i+2}{2}})$, и из (27.5.2) получим

$$(27.5.3) \quad E(m_v) = \mu_v + O\left(\frac{1}{n}\right).$$

Далее, согласно (27.5.2), каждая степень разности $m_v - \mu_v$ состоит из членов вида $\bar{x}^i (a_v - \mu_v) a_k a_{k_1} \dots$, и так же, как и выше, можно показать, что среднее значение такого члена имеет порядок $n^{-\frac{i+1}{2}}$. Таким образом, чтобы вычислить главную часть $E(m_v - \mu_v)^k$, достаточно взять

$$m_v - \mu_v = a_v - \mu_v - \binom{v}{1} \bar{x} a_{v-1},$$

так как все следующие члены в (27.5.2) дают добавки более высокого порядка малости. При $k=2$ мы получим, таким образом, выражение

$$(27.5.4) \quad D^2(m_v) = \frac{\mu_{2v} - 2v\mu_{v-1}\mu_{v+1} - \mu_v^2 + v^2\mu_2\mu_{v-1}}{n} + O\left(\frac{1}{n^2}\right),$$

так как, в силу (27.5.3), разность $E(m_v) - \mu_v$ имеет порядок n^{-1} . Вообще, для каждой четной степени разности $m_v - \mu_v$,

$$(27.5.5) \quad E(m_v - \mu_v)^{2k} = O\left(\frac{1}{n^k}\right).$$

Среднее значение произведения $(m_v - \mu_v)(m_p - \mu_p)$ можно вычислить таким же способом; используя снова (27.5.3), мы получим следующее выражение для смешанного второго момента между m_v и m_p :

$$(27.5.6) \quad \mu_{11}(m_v, m_p) =$$

$$= \frac{\mu_{v+p} - v\mu_{v-1}\mu_{p+1} - p\mu_{v+1}\mu_{p-1} - \mu_{v+p}^2 + v\mu_{v-1}\mu_{p-1}}{n} + O\left(\frac{1}{n^2}\right).$$

Выражения для первых семи-инвариантов выборки получаются заменой в формуле (15.10.5) моментов совокупности μ_v выборочными моментами m_v . Мы получаем при этом

$$k_1 = \bar{x}, \quad k_2 = m_2, \quad k_3 = m_3, \quad k_4 = m_4 - 3m_2^2.$$

Затем мы можем вывести выражения для среднего значения и дисперсии величин k_v с помощью данных выше формул для m_v . В частности,

выражая $E(k_i)$ через семи-инварианты совокупности x_i , мы этим способом получим

$$(27.5.7) \quad \begin{aligned} E(k_1) &= x_1, \\ E(k_2) &= \frac{n-1}{n} x_2, \\ E(k_3) &= \frac{(n-1)(n-2)}{n^2} x_3, \\ E(k_4) &= \frac{(n-1)(n^2-6n+6)}{n^3} x_4 - \frac{6(n-1)}{n^2} x_2^2. \end{aligned}$$

27.6. Несмешенные оценки. Рассмотрим выборочную дисперсию $m_2 = \frac{1}{n} \sum (x_i - \bar{x})^2$. В соответствии с параграфом 27.3, m_2 сходится по вероятности к дисперсии совокупности μ_2 , когда $n \rightarrow \infty$, так что при больших значениях n мы можем использовать m_2 в качестве оценки для μ_2 . Согласно терминологии Р. Фишера [89], [96], оценка, сходящаяся по вероятности к оцениваемой величине при неограниченном возрастании объема выборки, называется *состоятельной* оценкой. Таким образом, m_2 является состоятельной оценкой для μ_2 .

С другой стороны, формула (27.4.1) показывает, что средним значением величины m_2 является не μ_2 , а $\frac{n-1}{n} \mu_2$. Таким образом, если мы повторно извлекаем из данной совокупности выборки фиксированного объема n и вычисляем дисперсию m_2 для каждой выборки, то среднее значение всех наблюденных значений m_2 будет сходиться по вероятности не к „истинному значению“ μ_2 , а к несколько меньшему значению $\frac{n-1}{n} \mu_2$. Как оценка для μ_2 величина m_2 получает, таким образом, некоторое отрицательное *смещение*, которое можно устраниТЬ, если заменить m_2 величиной

$$M_2 = \frac{n}{n-1} m_2 = \frac{n}{n-1} s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2.$$

Действительно, тогда $E(M_2) = \frac{n}{n-1} E(m_2) = \mu_2$, и в соответствии с этим M_2 называется *несмешенной* оценкой для μ_2 . Поскольку множитель $\frac{n}{n-1}$ стремится к единице при $n \rightarrow \infty$, то и M_2 и m_2 сходятся по вероятности к μ_2 , так что M_2 является и состоятельной и несмешенной, в то время как m_2 является состоятельной, но не является несмешенной оценкой для μ_2 .

Подобным же образом, согласно параграфу 27.3, все центральные моменты m_v и семи-инварианты k_v выборки являются состоятельными оценками соответствующих величин μ_v или x_v , но из формул (27.5.1) и (27.5.7) следует, что при $v > 1$ эти оценки не являются несмешенными. Как и для m_2 , мы можем с помощью простых поправок получить оценки, которые будут и состоятельными и несмешенными. Так, например, для $v = 2, 3, 4$ мы получим следующие исправленные оценки для μ_v и x_v ,

$$M_2 = \frac{n}{n-1} m_2,$$

$$M_3 = \frac{n^2}{(n-1)(n-2)} m_3,$$

$$M_4 = \frac{n(n^2 - 2n + 3)}{(n-1)(n-2)(n-3)} m_4 - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} m_2^2$$

и

$$K_2 = \frac{n}{n-1} m_2,$$

$$K_3 = \frac{n^2}{(n-1)(n-2)} m_3,$$

$$K_4 = \frac{n^2}{(n-1)(n-2)(n-3)} [(n+1)m_4 - 3(n-1)m_2^2].$$

С помощью формул, приведенных в двух предыдущих параграфах, легко проверить, что во всех этих случаях $E(M_v) = \mu_v$ и $E(K_v) = x_v$. При больших значениях n часто бывает безразлично, используются ли M_v и K_v , или m_v и k_v , но когда n невелико, при пользовании величинами m_v и k_v может получиться существенное смещение. К вопросам, связанным со свойствами оценок, мы вернемся в главе 32.

В предыдущих параграфах мы видели, что вычисления при выводе формул для выборочных характеристик величин m_v и k_v становятся весьма утомительными, как только мы переходим от простейших случаев к более сложным. Р. Фишер [99], который ввел величины K_v (обозначенные им через k_v , см. сноску на стр. 376), обнаружил, что соответствующие выкладки для величин K_v можно значительно упростить с помощью методов комбинаторники. Эти методы далее были развиты самим Фишером, Висхартом и другими авторами. Хороший обзор этого предмета содержится в книге Кендала [19], где приведены также многочисленные ссылки на литературу.

27.7. Функции от моментов. Часто бывает необходимо получить среднее значение и дисперсию некоторой функции от выборочных мо-

ментов. Если эта функция является полиномом относительно \bar{x} и центральных моментов m_v , то проблему можно решить с помощью методов, развитых в параграфах 27.3—27.5. Аналогичным прямым методом часто можно пользоваться и тогда, когда функция содержит дробные степени \bar{x} и m_v . Рассмотрим в качестве простого примера стандартное отклонение выборки $s = \sqrt{m_2}$. Мы имеем тождество

$$\sqrt{m_2} - \sqrt{\mu_2} = \frac{m_2 - \mu_2}{2\sqrt{\mu_2}} - \frac{(m_2 - \mu_2)^2}{2\sqrt{\mu_2}(\sqrt{m_2} + \sqrt{\mu_2})^2}.$$

В силу формулы (27.4.1), первое слагаемое в правой части имеет среднее значение порядка n^{-1} . Второе слагаемое по абсолютной величине не превышает $\frac{(m_2 - \mu_2)^2}{2\mu_2^{3/2}}$, так что, в силу формул (27.4.1) и (27.4.2), среднее значение этого члена также имеет порядок n^{-1} . Таким образом,

$$(27.7.1) \quad E(\sqrt{m_2}) = \sqrt{\mu_2} + O\left(\frac{1}{n}\right).$$

С помощью аналогичных выкладок мы получаем

$$(27.7.2) \quad D^2(\sqrt{m_2}) = \frac{\mu_4 - \mu_2^2}{4\mu_2 n} + O\left(\frac{1}{n^2}\right).$$

Однако, во многих случаях мы имеем дело с функциями, содержащими отношения степеней некоторых моментов, например, коэффициенты g_1 и g_2 , коэффициент корреляции и т. д. Мы докажем теорему, охватывающую наиболее важные из этих случаев. Эта теорема будет сформулирована и доказана для случая функции $H(m_v, m_p)$ от двух центральных моментов, но ее можно непосредственно обобщить на любое число аргументов, включая среднее значение x . Теорема справедлива и тогда, когда функция H зависит только от одного аргумента. Теорема и ее доказательство распространяются также на функции от моментов многомерных выборок (см. параграф 27.8).

Рассмотрим функцию $H(m_v, m_p)$, не содержащую явно n . Мы можем рассматривать H либо как функцию от двух аргументов m_v и m_p , либо, если заменить m_v и m_p их выражениями через выборочные значения, как функцию от n величин x_1, \dots, x_n . В последнем случае функция, конечно, может содержать n явно. Докажем теперь следующую теорему:

Предположим, что выполнены следующие два условия:

1) В некоторой окрестности точки $m_v = \mu_v$, $m_p = \mu_p$ функция H непрерывна и имеет непрерывные первые и вторые производные по аргументам m_v и m_p .

2) Для всевозможных значений величин x_i выполняется неравенство $|H| < Cn^p$, где C и p — неотрицательные постоянные.

Тогда если обозначить через H_0 , H_1 и H_2 значения функции $H(m_v, m_p)$ и ее первых частных производных в точке $m_v = \mu_v$, $m_p = \mu_p$, то среднее значение и дисперсия случайной величины $H(m_v, m_p)$ будут определяться формулами

$$E(H) = H_0 + O\left(\frac{1}{n}\right),$$

$$(27.7.3) \quad D^2(H) = \mu_2(m_v) H_1^2 + 2\mu_{11}(m_v, m_p) H_1 H_2 + \\ + \mu_2(m_p) H_2^2 + O\left(\frac{1}{n^{p/2}}\right).$$

В силу формул (27.5.4) и (27.5.6), дисперсия функции H имеет, таким образом, вид $\frac{c}{n} + O(n^{-3/2})$, где c — постоянная. Имеющиеся в литературе доказательства приведенных соотношений часто неуловимо обобщены. Приведенное выше условие 2) можно значительно обобщить, но некоторые условия такого типа необходимы для справедливости теоремы. Действительно, если полностью отбросить условие 2), то, например, для каждой совокупности с $\mu_2 > 0$ функция $\frac{1}{m_2}$ должна была бы иметь среднее значение вида $\frac{1}{\mu_2} + O(n^{-1})$. Однако это, очевидно, неверно. Среднее значение величины $\frac{1}{m_2}$ не может быть конечным для всех совокупностей с распределением дискретного типа, так как для них с положительной вероятностью $m_2 = 0$. Легко показать, что аналогичные противоречия могут быть обнаружены даже для непрерывных распределений.

В параграфе 28.4 будет доказано, что функция $H(m_v, m_p)$ при больших значениях n распределена асимптотически нормально. Интересно заметить что в этом доказательстве не потребуется никаких условий, соответствующих условию 2).

Обозначим через $P(S)$ вероятностную функцию совместного распределения величин x_1, x_2, \dots, x_n . $P(S)$ есть функция множества в пространстве R_n величин x_i . Если в теореме Чебышева (15.7.1) положить $g(\xi) = (m_v - \mu_v)^{2k}$, то из формулы (27.5.5) будет следовать, что при любом $\varepsilon > 0$

$$P[(m_v - \mu_v)^{2k} \geq \varepsilon^{2k}] < \frac{A}{\varepsilon^{2kn}},$$

или

$$P[|m_v - \mu_v| \geq \varepsilon] < \frac{A}{\varepsilon^{2k} n^k},$$

где A — постоянная, не зависящая от n и ε . Соответствующий результат справедлив, конечно, и для m_p . Обозначим через Z множество всех точек в R_n , для которых удовлетворяются оба неравенства $|m_v - \mu_v| < \varepsilon$, $|m_p - \mu_p| < \varepsilon$, а через Z^* — дополнительное множество. Тогда, в соответствии со сказанным выше, мы получим

$$(27.7.4) \quad P(Z^*) < \frac{2A}{\varepsilon^{2k} n^k}, \quad P(Z) > 1 - \frac{2A}{\varepsilon^{2k} n^k}$$

Далее,

$$E(H) = \int_Z H dP + \int_{Z^*} H dP,$$

и, в силу условия 2), модуль последнего интеграла будет меньше, чем $\frac{2ACn^p}{\varepsilon^{2k} n^k}$. Если выбрать $k > p+1$, то

$$(27.7.5) \quad E(H) = \int_Z H dP + O\left(\frac{1}{n}\right).$$

Если ε достаточно мало, то по условию 1) для каждой точки множества Z мы будем иметь

$$(27.7.6) \quad \begin{aligned} I(m_v, m_p) &= H_0 + H_1(m_v - \mu_v) + H_2(m_p - \mu_p) + R, \\ R &= \frac{1}{2}[H'_{11}(m_v - \mu_v)^2 + 2H'_{12}(m_v - \mu_v)(m_p - \mu_p) + \\ &\quad + H'_{22}(m_p - \mu_p)^2], \end{aligned}$$

где H'_{ij} — значения вторых производных в некоторой промежуточной точке между (μ_v, μ_p) и (m_v, m_p) . Поэтому

$$(27.7.7) \quad \begin{aligned} \int_Z H dP &= H_0 P(Z) + H_1 \int_Z (m_v - \mu_v) dP + \\ &\quad + H_2 \int_Z (m_p - \mu_p) dP + \int_Z R dP. \end{aligned}$$

Рассмотрим слагаемые в правой части последнего равенства. В силу неравенств (27.7.4), первый член отличается от H_0 на величину порядка n^{-k} , т. е. меньшего, чем n^{-1} , так как $k > p+1 \geq 1$. Два следующих члена имеют порядок не выше n^{-1} , так как H_1 и H_2

не зависят от n , и, в силу формул (27.5.3) и (27.5.5) и неравенства Шварца (9.5.1), имеем

$$\begin{aligned} \int_Z (m_v - \mu_v) dP &= E(m_v - \mu_v) - \int_{Z^*} (m_v - \mu_v) dP = \\ &= O\left(\frac{1}{n}\right) - \int_{Z^*} (m_v - \mu_v) dP, \\ \left| \int_{Z^*} (m_v - \mu_v) dP \right| &\leq \left[\int_{Z^*} (m_v - \mu_v)^2 dP \cdot \int_{Z^*} dP \right]^{1/2} \leq \\ &\leq [E(m_v - \mu_v)^2 P(Z^*)]^{1/2} = O\left(n^{-\frac{1+k}{2}}\right), \end{aligned}$$

и аналогичные оценки для члена, содержащего m_p . Наконец, в силу условия 1), производные H'_{ij} при достаточно малом ϵ ограничены, и так же, как и выше, мы убеждаемся, что последний член в (27.7.7) имеет порядок n^{-1} . Так как первый член в (27.7.7) отличается от H_0 на величину порядка n^{-1} и справедлива формула (27.7.5), первое из соотношений (27.7.3) доказано.

Для доказательства второго соотношения (27.7.3) запишем

$$E(H - H_0)^2 = \int_Z (H - H_0)^2 dP + \int_{Z^*} (H - H_0)^2 dP.$$

Выбирая теперь $k > 2p + \frac{3}{2}$, с помощью условия 2) и уже доказанного первого соотношения (27.7.3) получим

$$D^2(H) = \int_Z (H - H_0)^2 dP + O(n^{-\frac{3}{2}}).$$

Затем представим $(H - H_0)^2$ с помощью (27.7.6) и проделаем те же выкладки, что и выше. При этом выкладки будут совершенно аналогичны сделанным выше, кроме оценки члена $\int_Z (m_v - \mu_v) R dP$, для которой получим, используя (15.5.6) и (27.7.5),

$$\begin{aligned} \left| \int_Z H'_{11} (m_v - \mu_v)^3 dP \right| &< K E(|m_v - \mu_v|^3) \leq \\ &\leq K(E(m_v - \mu_v)^4)^{3/4} = O(n^{-\frac{3}{2}}). \end{aligned}$$

Этим доказательство теоремы заканчивается.

Теперь применим формулы (27.7.3) к некоторым примерам. Рассмотрим сначала выборочные коэффициенты асимметрии и эксцесса:

$$g_1 = \frac{m_3}{m_2^{3/2}}, \quad g_2 = \frac{m_4}{m_2^2} - 3.$$

Эти функции удовлетворяют условию 1), если только $\mu_3 > 0$. Чтобы показать, что условие 2) также выполняется, запишем

$$g_1 = \sqrt{n} \frac{\sum (x_i - \bar{x})^3}{\left(\sum (x_i - \bar{x})^2\right)^{3/2}} = \sqrt{n} \sum_i \frac{(x_i - \bar{x})^3}{\sum_j (x_j - \bar{x})^2}^{3/2},$$

откуда

$$|g_1| \leq \sqrt{n} \sum_i \left(\frac{(x_i - \bar{x})^2}{\sum_j (x_j - \bar{x})^2} \right)^{3/2} \leq \sqrt{n} \sum_i \frac{(x_i - \bar{x})^2}{\sum_j (x_j - \bar{x})^2} = \sqrt{n}.$$

Аналогично доказывается, что $|g_2| \leq n$ для всех $n > 3$. Таким образом, для нахождения средних значений и дисперсий величин g_1 и g_2 можно применять формулы (27.7.3). Из формул (27.5.4) и (27.5.6) получим, с той же степенью приближения, что и формулы (27.7.3),

$$E(g_1) = \gamma_1, \quad E(g_2) = \gamma_2,$$

$$(27.7.8) \quad D^2(g_1) = \frac{4\mu_2^2\mu_6 - 12\mu_2\mu_3\mu_5 - 24\mu_2^3\mu_4 + 9\mu_3^2\mu_4 + 35\mu_2^2\mu_3^2 + 36\mu_2^5}{4\mu_2^5n},$$

$$D^2(g_2) = \frac{\mu_2^2\mu_8 - 4\mu_2\mu_4\mu_6 - 8\mu_2^2\mu_8\mu_5 + 4\mu_4^3 - \mu_2^2\mu_4^2 + 16\mu_2\mu_3^2\mu_4 + 16\mu_2^3\mu_3^2}{\mu_2^6n}.$$

Если генеральная совокупность нормальна, то эти приближенные выражения примут вид

$$(27.7.9) \quad E(g_1) = E(g_2) = 0, \quad D^2(g_1) = \frac{6}{n}, \quad D^2(g_2) = \frac{24}{n}.$$

Точные выражения для нормального случая будут даны в параграфе 29.3 (см. формулы (29.3.7)).

В качестве следующего примера рассмотрим отношение

$$V = \frac{s}{\bar{x}} = \frac{\sqrt{m_2}}{\bar{x}},$$

которое называется *коэффициентом изменчивости* выборки. Если распределение совокупности таково, что наша величина принимает только

положительные значения, то получим

$$\begin{aligned} V^2 &= \frac{\sum (x_i - \bar{x})^2}{n\bar{x}^2} = n \frac{\sum x_i^2}{(\sum x_i)^2} - 1 = \\ &= n \sum_i \left(\frac{x_i}{\sum_j x_j} \right)^2 - 1 < n \sum_i \frac{x_i}{\sum_j x_j} = n. \end{aligned}$$

так что можно применять формулы (27.7.3), заменяя в них, в соответствии с предварительным замечанием к нашей теореме, m , на \bar{x} . В силу формул (27.2.1), (27.4.2) и (27.4.4), получим, с той же степенью приближения, что и формулы (27.7.3),

$$(27.7.10) \quad \begin{aligned} E(V) &= \frac{\sigma}{m}, \\ D^2(V) &= \frac{m^2(\mu_4 - \mu_2^2) - 4m\mu_2\mu_3 + 4\mu_2^3}{4m^4\mu_2^2}. \end{aligned}$$

Нормальное распределение не обладает тем свойством, что величина принимает только положительные значения, и легко видеть, что для такого распределения величина V неограничена, так что условие 2) не выполняется. Однако можно рассмотреть нормальное распределение, усеченное в точке $x = 0$ (см. параграф 19.3), и если $\frac{\sigma}{m}$ достаточно мало, то центральные моменты такого распределения будут приблизительно равны соответствующим моментам полного нормального распределения. В этом случае приближенное выражение для дисперсии величины V принимает вид

$$(27.7.11) \quad D^2(V) = \frac{\sigma^2}{2m^2n} \left(1 + 2 \frac{\sigma^2}{m^2} \right).$$

27.8. Характеристики многомерных распределений. Формулы для выборочных характеристик, выведенные в параграфах 27.2—27.6, так же как и теорему, доказанную в параграфе 27.7, можно непосредственно распространить на характеристики многомерных выборок. При этом выкладки совершенно аналогичны произведенным выше, так что мы приведем здесь лишь некоторые формулы для двумерного случая. Обозначения, которыми мы будем пользоваться, введены в параграфе 27.1; всюду предполагается, что все встречающиеся в

формулах моменты конечны. Мы имеем

$$\begin{aligned} E(m_{ik}) &= \mu_{ik} + O\left(\frac{1}{n}\right), \\ E(m_{11}) &= \frac{n-1}{n} \mu_{11}, \quad D^2(m_{11}) = \frac{\mu_{22} - \mu_{11}^2}{n} + O\left(\frac{1}{n^2}\right), \\ \mu_{11}(m_{20}, m_{02}) &= \frac{\mu_{22} - \mu_{20}\mu_{02}}{n} + O\left(\frac{1}{n^2}\right) \\ \mu_{11}(m_{11}, m_{20}) &= \frac{\mu_{31} - \mu_{11}\mu_{20}}{n} + O\left(\frac{1}{n^2}\right). \end{aligned}$$

Выборочный коэффициент корреляции

$$r = \frac{m_{11}}{\sqrt{m_{20}m_{02}}}$$

удовлетворяет, очевидно, условиям теоремы параграфа 27.7, так как $|r| \leq 1$. Обозначая через ρ значение коэффициента корреляции совокупности, с помощью приведенных выше соотношений получим (с той же степенью приближения, что и формулы (27.7.3))

$$(27.8.1) \quad \begin{aligned} E(r) &= ? \\ D^2(r) &= \frac{\rho^2}{4n} \left(\frac{\mu_{40}}{\mu_{20}^2} + \frac{\mu_{04}}{\mu_{02}^2} + \frac{2\mu_{22}}{\mu_{20}\mu_{02}} + \frac{4\mu_{22}}{\mu_{11}^2} - \frac{4\mu_{31}}{\mu_{11}\mu_{20}} - \frac{4\mu_{13}}{\mu_{11}\mu_{02}} \right). \end{aligned}$$

В случае нормальной совокупности выражение для дисперсии сводится (см. упражнение 3, стр. 349), с точностью до величины порядка $n^{-\frac{1}{2}}$, к следующему выражению;

$$(27.8.2) \quad D^2(r) = \frac{(1-\rho^2)^2}{n}.$$

В заключение заметим, что теорема параграфа 27.3 о сходимости по вероятности выборочных характеристик справедлива без всяких изменений и в многомерном случае. Так, например, r сходится по вероятности к ρ , частный коэффициент корреляции $r_{12,34\dots k}$ выборки сходится по вероятности к $\rho_{12,34\dots k}$ и т. д.

27.9. Поправки к группировке. На практике выборки очень часто подвергают группировке (см. параграф 25.5). Предположим, что мы извлекли выборку в n значений из одномерного распределения непрерывного типа с плотностью вероятности $f(x)$, и пусть выборочные значения группированы по интервалам длины h со средними точками $\xi_i = \xi_0 + ih$, где $i = 0, \pm 1, \pm 2, \dots$. В таких случаях при вычислении моментов и других выборочных характеристик обычно предполагается, что все

выборочные значения, принадлежащие некоторому интервалу, совпадают со средней точкой этого интервала. Тогда мы в действительности производим выбор из распределения дискретного типа, в котором величина принимает каждое значение $\xi_i = \xi_0 + ih$ с вероятностью

$$p_i = \int_{\xi_i - \frac{1}{2}h}^{\xi_i + \frac{1}{2}h} f(x) dx.$$

Моменты и другие характеристики, которые мы оцениваем по нашим выборочным характеристикам, пользуясь формулами, приведенными в предыдущих параграфах этой главы, будут, таким образом, моментами этого „группированного распределения“:

$$\bar{a}_v = \sum_{-\infty}^{\infty} p_i \xi_i^v.$$

Однако во многих случаях мы в действительности хотим знать не эти моменты, а моменты данного непрерывного распределения:

$$a_v = \int_{-\infty}^{\infty} x^v f(x) dx.$$

Важно поэтому исследовать соотношения между этими двумя множествами моментов. Мы покажем, что при некоторых условиях приближенные значения моментов a_v можно получить, прибавляя некоторые поправки к групповым моментам \bar{a}_v .

Групповые моменты можно записать в виде

$$\bar{a}_v = \sum_{-\infty}^{\infty} \xi_i^v \int_{\xi_i - \frac{1}{2}h}^{\xi_i + \frac{1}{2}h} f(x) dx = \sum_{-\infty}^{\infty} g(\xi_i),$$

где $\xi_i = \xi_0 + ih$ и

$$(27.9.1) \quad g(\xi) = \xi^v \int_{\xi - \frac{1}{2}h}^{\xi + \frac{1}{2}h} f(x) dx.$$

Предполагая, что $f(x)$ непрерывна при всех x , по формуле суммиро-

жания Эйлера — Маклорена (12.2.5) получим

$$(27.9.2) \quad \bar{a}_v = \int_{-\infty}^{\infty} (\xi_0 + hy)^v dy = \int_{\xi_0 + hy - \frac{1}{2}h}^{\xi_0 + hy + \frac{1}{2}h} f(x) dx + R,$$

$$R = -h \int_{-\infty}^{\infty} P_1(y) g'(\xi_0 + hy) dy.$$

Допустим пока, что остаточным членом R можно пренебречь. Тогда, меняя порядок интегрирования, получим

$$\bar{a}_v = \int_{-\infty}^{\infty} f(x) dx \int_{\frac{x-\xi_0-1}{h}-\frac{1}{2}}^{\frac{x-\xi_0+1}{h}+\frac{1}{2}} (\xi_0 + hy)^v dy =$$

$$= \int_{-\infty}^{\infty} \frac{\left(x + \frac{h}{2}\right)^{v+1} - \left(x - \frac{h}{2}\right)^{v+1}}{h(v+1)} f(x) dx =$$

$$= \frac{1}{v+1} \sum_{i=0}^{[v/2]} \binom{v+1}{2i+1} \left(\frac{h}{2}\right)^{2i} a_{v-2i}.$$

Таким образом, групповые моменты a_v могут быть выражены линейными функциями от „истинных“ моментов a_v . Решая последовательно уравнения относительно a_v , получим

$$(27.9.3) \quad a_1 = \bar{a}_1,$$

$$a_2 = \bar{a}_2 - \frac{1}{12} h^2,$$

$$a_3 = \bar{a}_3 - \frac{1}{4} \bar{a}_1 h^2,$$

$$a_4 = \bar{a}_4 - \frac{1}{2} \bar{a}_2 h^2 + \frac{7}{240} h^4,$$

$$a_5 = \bar{a}_5 - \frac{5}{6} \bar{a}_3 h^2 + \frac{7}{48} \bar{a}_1 h^4,$$

$$a_6 = \bar{a}_6 - \frac{5}{4} \bar{a}_4 h^2 + \frac{7}{16} \bar{a}_2 h^4 - \frac{31}{1344} h^6,$$

$$\dots \dots \dots \dots \dots \dots \dots$$

Эти формулы известны под названием *поправок Шеппарда* [212]. Общая формула имеет вид (см. Вольд [245])

$$a_v = \sum_{i=0}^v \binom{v}{i} (2^{1-i} - 1) B_i \bar{a}_{v-i} h^i,$$

где B_i — числа Бернулли, определяемые формулой (12.2.2).

Если поместить начало отсчета в среднюю точку распределения, то $a_1 = \bar{a}_1 = 0$, и получаются поправки для центральных моментов:

$$(27.9.4) \quad \begin{aligned} \mu_2 &= \bar{\mu}_2 - \frac{1}{12} h^2, \\ \mu_3 &= \bar{\mu}_3, \\ \mu_4 &= \bar{\mu}_4 - \frac{1}{2} \bar{\mu}_2 h^2 + \frac{7}{240} h^4, \\ &\dots \end{aligned}$$

Эти соотношения справедливы в предположении, что остаточным членом R в (27.9.2) можно пренебречь. Предположим теперь, что заданы два целых положительных числа s и k такие, что:

- 1) $f(x)$ и ее первые $2s$ производных непрерывны для всех x .
- 2) Произведение $x^{k+2} f^{(i)}(x)$ ограничено для всех x и для $i = 0, 1, \dots, 2s$.

Тогда функция $g(\xi)$, заданная формулой (27.9.1), будет непрерывна при всех ξ вместе со своими первыми $2s+1$ производными. И легко видеть, что при $v = 1, 2, \dots, k$ и $i = 0, 1, \dots, 2s+1$ имеем

$$(27.9.5) \quad g^{(i)}(\xi) = O(\xi^{-2})$$

при $\xi \rightarrow \pm\infty$. Следовательно, можно применять формулу Эйлера — Маклорена в виде (12.2.6), так что остаточный член R можно записать следующим образом:

$$R = (-1)^{s+1} h^{2s+1} \int_{-\infty}^{\infty} P_{2s+1}(y) g^{(2s+1)}(\xi_0 + hy) dy.$$

Тогда из (12.2.1) и (27.9.5) следует, что

$$|R| < Ah^{2s+1} \int_{-\infty}^{\infty} \frac{dy}{1 + (\xi_0 + hy)^2} < Bh^{2s},$$

где A и B — постоянные, не зависящие от h . Таким образом, если длина h интервалов разбиения достаточно мала, то остаточным чле-

ном R можно пренебречь, и при применении поправок (27.9.3) или (27.9.4) к моментам любого порядка $\nu \leq k$ получающаяся ошибка будет иметь порядок h^{2s} .

Всякий раз, когда в случае распределения конечной широты график плотности вероятности $y = f(x)$ имеет соприкосновение высокого порядка с осью x в точках g и G (см. параграф 15.7), указанные выше условия 1) и 2) выполняются по крайней мере для не слишком больших значений s и k . В таких случаях на практике обнаруживается, что применение поправок Шеппарда дает обычно хорошие результаты, даже если h не очень мало. Однако всегда следует сравнить величину поправки, применяемой к моменту, со стандартным отклонением выборочного распределения этого момента. Если, как это часто случается, поправка составляет лишь малую долю от стандартного отклонения, то не имеет значения, применялась поправка или нет.

В случаях, когда график плотности вероятности не имеет соприкосновения высокого порядка с осью x на концах интервала (g, G) , обычно лучше не применять поправок Шеппарда. Для таких случаев существуют другие поправочные формулы, но применимость их, повидимому, весьма ограничена (см. Эллертон [12], стр. 231).

Лэнгдон и Оре [144] и Вольд [245], [246] дали поправки для семи-инвариантов, которые применимы при тех же условиях, что и поправки Шеппарда. Эти поправки имеют простой вид

$$x_1 = \bar{x}_1, \quad x_\nu = \bar{x}_\nu - \frac{B_\nu}{\nu} h^\nu, \quad (\nu > 1).$$

Выход поправок Шеппарда можно распространить на моменты многомерных выборок. В частности, для двумерного распределения с интервалами разбиения длины h_1 по оси x и h_2 по оси y мы имеем

$$(27.9.6) \quad \begin{aligned} \mu_{11} &= \bar{\mu}_{11}, \quad \mu_{21} = \bar{\mu}_{21}, \quad \mu_{31} = \bar{\mu}_{31} - \frac{1}{4} \bar{\mu}_{11} h_1^2, \\ \mu_{22} &= \bar{\mu}_{22} - \frac{1}{12} \bar{\mu}_{20} h_2^2 - \frac{1}{12} \bar{\mu}_{02} h_1^2 + \frac{1}{144} h_1^2 h_2^2. \end{aligned}$$

Поправки для μ_{12} и μ_{13} получаются, конечно, перестановкой индексов, а поправки для частных моментов μ_{10} и μ_{01} получаются непосредственно из (27.9.4), так что с помощью формул (27.9.6) можно получить поправки для всех моментов порядка, не превосходящего четырех.

В заключение следует заметить, что проблема поправок к группировке исследовалась также с различных других точек зрения (см. например, Фишер [89] и Кендал [136]).

ГЛАВА 28

АСИМПТОТИЧЕСКИЕ СВОЙСТВА ВЫБОРОЧНЫХ РАСПРЕДЕЛЕНИЙ

28.1. Вводные замечания. В параграфах 27.3 и 27.8 мы видели, что все обычные выборочные характеристики, являющиеся функциями от моментов, сходятся по вероятности к соответствующим характеристикам совокупности, когда объем выборки n стремится к бесконечности. В настоящей главе мы рассмотрим подробнее асимптотическое поведение выборочных распределений этих и некоторых других характеристик при больших n . Следуя замечанию, сделанному в параграфе 17.5, сначала покажем, что при весьма общих условиях характеристики, основанные на выборочных моментах, *распределены асимптотически нормально* при больших n . Затем мы рассмотрим другие классы выборочных характеристик; некоторые из них, подобно моментным характеристикам, асимптотически нормальны, в то время как другие имеют совершенно иное асимптотическое поведение.

28.2. Моменты. Рассмотрим n выборочных значений x_1, \dots, x_n , выбранных из некоторого одномерного распределения. Величина $na_v = \sum_i x_i^v$ есть сумма n независимых случайных величин x_i^v , имеющих одно и то же распределение со средним значением $E(x_i^v) = a_v$ и дисперсией $D^2(x_i^v) = a_{2v} - a_v^2$. Применяя центральную предельную теорему в формулировке Линнеберга—Леви (см. параграф 17.4), мы обнаружим, что при $n \rightarrow \infty$ функция распределения нормированной суммы

$$\frac{\sum x_i^v - na_v}{\sqrt{n(a_{2v} - a_v^2)}} = \sqrt{n} \frac{a_v - a_v}{\sqrt{a_{2v} - a_v^2}}$$

стремится к нормальной функции распределения $\Phi(x)$. В соответствии с терминологией, введенной в параграфе 17.4, каждый выборочный момент a_v является, таким образом, *асимптотически нормальным* $(a_v, \sqrt{\frac{a_{2v} - a_v^2}{n}})$. Мы замечаем, что параметры предельного нормального распределения определяются выборочными моментами, а не параметрами исходного распределения.

мального распределения совпадают со средним значением и стандартным отклонением величины a_v , заданными формулами (27.3.1). В частности, среднее значение выборки $a_1 = \bar{x}$ асимптотически нормально $\left(m, \frac{\sigma}{\sqrt{n}}\right)$, как уже было указано в параграфе 17.4.

Подобным образом, если мы рассмотрим одновременно две случайные величины $na_v = \sum x_i^v$ и $na_p = \sum x_i^p$, то, применяя двумерную форму теоремы Линдеберга — Леви (см. параграф 21.11), мы убедимся, что совместное распределение двух величин $\sqrt{n}(a_v - a_v)$ и $\sqrt{n}(a_p - a_p)$ стремится к некоторому двумерному нормальному распределению. Доказательство, очевидно, имеет общий характер, и с помощью многомерной формы теоремы Линдеберга — Леви (см. параграф 24.7) получается следующий результат:

Совместное распределение любого количества величин $\sqrt{n}(a_v - a_v)$ стремится к нормальному распределению с нулевым средним значением и вторыми моментами

$$(28.2.1) \quad \begin{aligned} \lambda_{vv} &= \sigma_v^2 = E(n(a_v - a_v)^2) = a_{2v} - a_v^2, \\ \lambda_{vp} &= E(n(a_v - a_v)(a_p - a_p)) = a_{v+p} - a_v a_p. \end{aligned}$$

Таким образом, если ввести нормированные величины z_v , положив

$$(28.2.2) \quad a_v = a_v + \frac{\sigma_v}{\sqrt{n}} z_v,$$

так что каждое z_v будет иметь среднее значение 0 и стандартное отклонение 1, то совместное распределение величин z_v будет асимптотически нормальным со смешанными вторыми моментами

$$E(z_v z_p) = \frac{\lambda_{vp}}{\sigma_v \sigma_p}.$$

Все изложенное прямо распространяется на моменты многомерных выборок.

28.3. Центральные моменты. Согласно замечаниям к формуле (27.5.2), каждый центральный момент m_v можно записать в виде

$$m_v = a_v - \bar{x} a_{v-1} + \frac{w}{n},$$

где w — случайная величина, для которой $E(w^2)$ не превышает некоторой величины, не зависящей от n . В соответствии с парагра-

фом 27.4, можно без нарушения общности полагать $m = 0$, так что $\alpha_v = \mu_v$, и

$$m_v - \mu_v = a_v - \alpha_v - v\bar{x}a_{v-1} + \frac{w}{n}.$$

Вводя нормированные величины z_v , определяемые формулой (28.2.2), получим

$$(28.3.1) \quad \sqrt{n}(m_v - \mu_v) = \sigma_v z_v - v\sigma_1 \mu_{v-1} z_1 + \frac{R}{\sqrt{n}},$$

где $R = w - v\sigma_1 \mu_{v-1} z_1 z_{v-1}$. Далее, в силу формулы (9.5.1),

$$\begin{aligned} E(|R|) &\leq E(|w|) + v\sigma_1 \mu_{v-1} E(|z_1 z_{v-1}|) \leq \\ &\leq \sqrt{E(w^2)} + v\sigma_1 \mu_{v-1} \sqrt{E(z_1^2) E(z_{v-1}^2)}, \end{aligned}$$

так что $E(|R|)$ не превышает некоторой величины, не зависящей от n , и поэтому из теоремы Чебышева (15.7.1) следует, что $\frac{R}{\sqrt{n}}$ сходится по вероятности к нулю. Применяя к выражению (28.3.1) теорему параграфа 20.6, мы обнаружим, что величина $\sqrt{n}(m_v - \mu_v)$ имеет в пределе при $n \rightarrow \infty$ то же распределение, что и линейное выражение $\sigma_v z_v - v\sigma_1 \mu_{v-1} z_1$. Однако совместное распределение величин z_v и z_1 асимптотически нормально, и каждая линейная комбинация нормально распределенных величин, в силу параграфа 24.4, сама распределена нормально.

Таким образом, каждый центральный момент выборки m_v распределен асимптотически нормально со средним значением μ_v и дисперсией

$$\frac{\sigma_v^2 - 2v\mu_{v-1}\mu_1 + v^2\sigma_1^2\mu_{v-1}^2}{n} = \frac{\mu_{2v} - 2v\mu_{v-1}\mu_{v+1} - \mu_v^2 + v^2\mu_2\mu_{v-1}^2}{n}.$$

Мы замечаем, что дисперсия предельного нормального распределения совпадает с главным членом $D^2(m_v)$ в формуле (27.5.4). Если одновременно рассматривать любое количество величин m_v , то мы тем же способом получим, используя последнюю теорему параграфа 22.6, что совместное распределение величин m_v асимптотически нормально со средними значениями μ_v и дисперсиями и смешанными вторыми моментами, совпадающими с главными членами выражений (27.5.4) и (27.5.6). Как и в предыдущем параграфе, полученный результат прямо распространяется на моменты многомерных выборок.

28.4. Функции от моментов. Как и в параграфе 27.7, ограничимся случаем функции $H(m_v, m_p)$ от двух центральных моментов одномерной выборки. Однако распространение на любое количество аргументов, на многомерные выборки и на совместное распределение любого количества функций получается немедленно. Докажем следующую теорему.

Если в некоторой окрестности точки $m_v = \mu_v, m_p = \mu_p$ функция $H(m_v, m_p)$ непрерывна и имеет непрерывные первые и вторые производные по аргументам m_v и m_p , то случайная величина $H(m_v, m_p)$ асимптотически нормальна, причем среднее значение и дисперсия предельного нормального распределения совпадают с главными членами формулы (27.7.3).

В этой теореме отсутствует требование, соответствующее условию 2) теоремы, доказанной в параграфе 27.7. Таким образом, можно утверждать, например,

что функция $\frac{1}{m_2}$ асимптотически нормальна $\left(\frac{1}{\mu_2}, \frac{\sqrt{\mu_4 - \mu_2^2}}{\mu_2^2 \sqrt{n}}\right)$, хотя в некоторых совокупностях (см. параграф 27.7) и среднее значение, и дисперсия величины $\frac{1}{m_2}$ бесконечны. В этой связи напомним замечание, сделанное в параграфе 17.4: величина может быть асимптотически нормальной, даже если ее среднее значение и дисперсия не существуют или не стремятся к среднему значению и дисперсии предельного нормального распределения.

Как и в параграфе 27.7, рассмотрим множество Z всех точек (x_1, \dots, x_n) таких, что $|m_v - \mu_v| < \varepsilon$ и $|m_p - \mu_p| < \varepsilon$. Однако теперь мы будем считать величину ε зависящей от n , именно, положим $\varepsilon = n^{-\frac{1}{4}}$. Тогда, пользуясь обозначениями параграфа 27.7 и полагая $k = 1$, получим

$$P(Z) > 1 - \frac{2A}{\varepsilon^{2n}} = 1 - 2An^{-\frac{1}{2n}}.$$

Если n достаточно велико, то в каждой точке множества Z справедливо разложение (27.7.6), которое можно записать в виде

$$\sqrt{n}(H - H_0) = H_1 \sqrt{n}(m_v - \mu_v) + H_2 \sqrt{n}(m_p - \mu_p) + R\sqrt{n},$$

где $|R\sqrt{n}| < K\varepsilon^2\sqrt{n} = Kn^{-\frac{1}{2}}$. Таким образом, неравенство $|R\sqrt{n}| < Kn^{-\frac{1}{2}}$ удовлетворяется с вероятностью $\geq P(Z) > 1 - 2An^{-\frac{1}{2n}}$, следовательно, $R\sqrt{n}$ сходится по вероятности к нулю. Согласно теореме параграфа 20.6, величины $\sqrt{n}(H - H_0)$ и $H_1 \sqrt{n}(m_v - \mu_v) + H_2 \sqrt{n}(m_p - \mu_p)$ в пределе при $n \rightarrow \infty$ имеют одно и то же распределение. Однако, в силу предыдущего параграфа, вторая из этих вели-

чин асимптотически нормальна именно с тем средним значением и с той дисперсией, которые требуются нашей теоремой. Поэтому теорема доказана.

Из этой теоремы следует, что любая выборочная характеристика, образованная из моментов, при больших значениях n распределена приближенно нормально около соответствующей характеристики совокупности, с дисперсией вида $\frac{c}{n}$, при одном лишь условии, что главные члены формул (27.7.3) дают конечные значения для среднего значения и дисперсии предельного распределения.

Это остается справедливым и для выборок любого числа измерений. Так, например, коэффициенты асимметрии и эксцесса (см. параграф 15.8), коэффициенты регрессии (см. параграфы 21.6 и 23.2), обобщенная дисперсия (см. параграф 22.7) и полный, частный и сводный коэффициенты корреляции (см. параграфы 21.7, 23.4 и 23.5) — все распределены асимптотически нормально около соответствующих коэффициентов совокупности.

В этой связи необходимо сделать следующее важное замечание. Вообще говоря, постоянная c в выражении для дисперсии должна иметь положительное значение. Однако в конкретных случаях c может быть нулем, и тогда дисперсия будет меньшего порядка, чем n^{-1} . Из доказательства теоремы видно, что в этом случае величина $\sqrt{n}(H - H_0)$ сходится по вероятности к нулю, что можно выразить, сказав, что H асимптотически нормальна с нулевой дисперсией, пока рассматриваются члены порядка n^{-1} . Может случиться, однако, что некоторое выражение вида $n^p(H - H_0)$ с $p > \frac{1}{2}$ будет иметь определенное предельное распределение, но не обязательно нормальное. Пример такого явления встретится в параграфе 29.12 в связи с распределением сводного коэффициента корреляции, в частном случае, когда соответствующая характеристика совокупности равна нулю.

28.5. Квантили. Рассмотрим выборку в n значений из одномерного распределения непрерывного типа с функцией распределения $F(x)$ и плотностью вероятности $f(x) = F'(x)$. Пусть $\zeta = \zeta_p$ обозначает квадиль порядка p этого распределения (см. параграф 15.6), т. е. (по предложению единственного) корень уравнения $F(\zeta) = p$, где $0 < p < 1$. Будем предполагать, что в некоторой окрестности точки $x = \zeta_p$ плотность вероятности $f(x)$ непрерывна и имеет непрерывную производную $f'(x)$.

Обозначим, далее, через z_p соответствующую квантиль выборки. Если np не есть целое число, и если мы расположим выборочные значения $x_1 \leq x_2 \leq \dots \leq x_n$ в порядке возрастания, то существует единственная квантиль z_p , равная выборочному значению $x_{\mu+1}$, где $\mu = [np]$ обозначает наибольшее целое число, не превосходящее np . Если же np — целое число, то мы имеем неопределенный случай (см. параграфы 15.5—15.6), и z_p может иметь любое значение из интервала (x_{np}, x_{np+1}) . Во избежание тривиальных усложнений, будем в дальнейшем предполагать, что np не есть целое число.

Обозначим через $g(x)$ плотность вероятности случайной величины $z = z_p$. Вероятность $g(x)dx$ того, что z попадает в бесконечно малый интервал $(x, x + dx)$, совпадает с вероятностью того, что $\mu = [np]$ выборочных значений не превосходят x , а $n - \mu - 1$ выборочных значений превосходят $x + dx$, так что оставшееся значение попадает между x и $x + dx$. Поэтому

$$g(x)dx = \binom{n}{\mu} (n - \mu) (F(x))^\mu (1 - F(x))^{n-\mu-1} f(x)dx.$$

Чтобы изучить поведение распределения величины z при больших n , рассмотрим случайную величину $y = \sqrt{\frac{pq}{n}} f(\zeta)(z - \zeta)$, где $q = 1 - p$. Согласно формуле (15.1.2), y имеет плотность вероятности

$$\frac{1}{f(\zeta)} \sqrt{\frac{pq}{n}} g\left(\zeta + \sqrt{\frac{pq}{n}} \frac{x}{f(\zeta)}\right) = A_1 A_2 A_3,$$

где для каждого фиксированного x при $n \rightarrow \infty$ имеем (см. (16.4.8))

$$A_1 = \sqrt{\frac{pq}{n}} \binom{n}{\mu} p^n q^{n-\mu} \frac{n-\mu}{q} \rightarrow \frac{1}{\sqrt{2\pi}},$$

$$A_2 = \frac{f\left(\zeta + \sqrt{\frac{pq}{n}} \frac{x}{f(\zeta)}\right)}{f(\zeta)} \rightarrow 1,$$

$$A_3 = \left(\frac{F(t)}{p}\right)^\mu \left(\frac{1-F(t)}{q}\right)^{n-\mu-1},$$

где $t = \zeta + \sqrt{\frac{pq}{n}} \frac{x}{f(\zeta)}$. Но $F(\zeta) = p$, и поэтому

$$F(t) = p + x \sqrt{\frac{pq}{n}} + \frac{1}{2} x^2 \frac{pq}{n} \frac{f'(\zeta)}{f^2(\zeta)} + o\left(\frac{1}{n}\right).$$

Подставляя это в выражение для A_3 , после некоторых вычислений получим

$$A_3 \rightarrow e^{-\frac{x^2}{2}},$$

так что плотность вероятности величины y стремится к нормальной плотности вероятности $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. При этом видно, что A_1, A_2, A_3 равномерно ограничены в любом интервале $a < x < b$, так что, согласно (5.3.6), вероятность неравенства $a < y < b$ стремится к пределу

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

Отсюда следует, что выборочная квантиль z_p асимптотически нормальна $(\zeta, \frac{1}{f(\zeta)} \sqrt{\frac{pq}{n}})$, где $\zeta = \zeta_p$ — соответствующая квантиль совокупности. В частности, медиана выборки асимптотически нормальна $(\zeta, \frac{1}{2f(\zeta)} \sqrt{\frac{\pi}{2n}})$, где $\zeta = \zeta_{0.5}$ — медиана совокупности.

Для нормального распределения с параметрами m и σ медианой является m , и мы имеем $f(m) = \frac{1}{\sigma \sqrt{2\pi}}$. Таким образом, медиана z выборки n значений из нормального распределения асимптотически нормальна $(m, \sigma \sqrt{\frac{\pi}{2n}})$.

С другой стороны, известно, что среднее значение \bar{x} такой выборки z точности нормально $(m, \frac{\sigma}{\sqrt{n}})$. Если $n \rightarrow \infty$, то и z и \bar{x} сходятся по вероятности к m , и при больших значениях n в качестве оценки для m можно употреблять или z , или \bar{x} . Однако последнюю оценку следует считать более точной, так как стандартное отклонение $\frac{\sigma}{\sqrt{n}}$, соответствующее оценке \bar{x} , меньше, чем стандартное отклонение $\sigma \sqrt{\frac{\pi}{2n}} = 1.2533 \frac{\sigma}{\sqrt{n}}$, соответствующее оценке z . Систематическое сравнение точности различных оценок для характеристик совокупности будет предпринято в теории оценок (см. главу 32).

Рассмотрим теперь совместное распределение двух квантилей z' и z'' порядков p_1 и p_2 , где $p_1 < p_2$. С помощью таких же вычислений, как и выше, можно показать, что это распределение асимптотически

нормально. Средними значениями предельного нормального распределения являются соответствующие квантили совокупности ζ' и ζ'' , а асимптотическими выражениями для моментов второго порядка $\mu_2(z')$, $\mu_{11}(z', z'')$, $\mu_3(z'')$ будут

$$\frac{p_1 q_1}{n f^2(\zeta')}, \quad \frac{p_1 q_2}{n f(\zeta') f(\zeta'')}, \quad \frac{p_2 q_2}{n f^2(\zeta'')}.$$

Если положить, в частности, $p_1 = \frac{1}{4}$, $p_2 = \frac{3}{4}$, то ζ' и ζ'' будут верхней и нижней квартилями совокупности, и мы обнаружим, что семиинтерквартильная широта (см. параграф 15.6) выборки, $\frac{1}{2}(z'' - z')$, распределена асимптотически нормально со средним значением $\frac{1}{2}(\zeta'' - \zeta')$ и стандартным отклонением

$$\frac{1}{8\sqrt{n}} \sqrt{\frac{3}{f^2(\zeta')} - \frac{2}{f(\zeta') f(\zeta'')} + \frac{3}{f^2(\zeta'')}}.$$

Для нормальной (m, σ) совокупности среднее значение семи-интерквартильной широты равно 0,6745 σ , а стандартное отклонение $0,7867 \frac{\sigma}{\sqrt{n}}$.

28.6. Экстремальные значения и широта выборки. Пока мы рассмотрели лишь такие выборочные характеристики, распределения которых в больших выборках стремятся к нормальному распределению. Теперь обратимся к группе характеристик, имеющих совершенно иное поведение.

У одномерной выборки, состоящей из n значений, всегда имеются два конечных и однозначно определенных **экстремальных значения** *) и также конечная *широта*, являющаяся разностью между этими экстремальными значениями.

Вообще, расположим n выборочных значений в порядке их величин и рассмотрим значение, стоящее на u -м месте от начала или от конца этой последовательности (u -ое *нижнее* или *верхнее значение*). При $u=1$ получаются, конечно, экстремальные значения.

*) Если, например, два наивысших значения равны, то каждое из них можно считать верхним экстремальным значением, и аналогично в других случаях.

Часто бывает важно знать выборочные распределения экстремальных значений, v -х значений, широты и других подобных характеристик выборки. Сейчас мы рассмотрим некоторые свойства этих распределений.

Ограничимся случаем, когда совокупность имеет распределение непрерывного типа с функцией распределения F и плотностью вероятности $f = F'$. Обозначим через x v -е верхнее значение выборки в n элементов из этой совокупности. Дифференциал вероятности $g_v(x) dx$ выборочного распределения величины x совпадает с вероятностью того, что среди n выборочных значений $n - v$ значений $< x$ и $v - 1$ значений $> x + dx$, так что остающееся значение попадает между x и $x + dx$. Поэтому

$$(28.6.1) \quad g_v(x) dx = n \binom{n-1}{v-1} (F(x))^{n-v} (1 - F(x))^{v-1} f(x) dx.$$

Если ввести новую величину ξ , положив

$$(28.6.2) \quad \xi = n(1 - F(x)),$$

то имеем $0 \leq \xi \leq n$, и плотностью вероятности $h_v(\xi)$ этой новой величины будет

$$(28.6.3) \quad h_v(\xi) = \binom{n-1}{v-1} \left(\frac{\xi}{n}\right)^{v-1} \left(1 - \frac{\xi}{n}\right)^{n-v}$$

при $0 \leq \xi \leq n$ и $h_v(\xi) = 0$ вне интервала $(0, n)$. При $n \rightarrow \infty$ $h_v(\xi)$ для любого $\xi \geq 0$ сходится к пределу

$$(28.6.4) \quad \lim_{n \rightarrow \infty} h_v(\xi) = \frac{\xi^{v-1}}{\Gamma(v)} e^{-\xi}.$$

Далее, $h_v(\xi)$ равномерно ограничена для всех n в каждом конечном интервале значений ξ , так что, согласно (5.3.6), величина ξ в пределе при $n \rightarrow \infty$ распределена с плотностью вероятности (28.6.4), которая есть частный случай для (12.3.3).

Аналогично, если y обозначает v -е нижнее значение нашей выборки и если ввести новую величину η , положив

$$(28.6.5) \quad \eta = n F(y),$$

то мы убедимся, что η имеет плотность вероятности $h_v(\eta)$, и, таким образом, в пределе при $n \rightarrow \infty$ плотность вероятности $\frac{\eta^{v-1}}{\Gamma(v)} e^{-\eta}$.

Можно рассмотреть также совместное распределение v -го верхнего значения x и v -го нижнего значения y . Введя величины ξ и η формулами

(28.6.2) и (28.6.5), таким же образом, как и выше, убеждаемся, что совместная плотность вероятности для ξ и η имеет вид

$$(28.6.6) \quad \frac{1}{n^2} \frac{n!}{[(v-1)!]^2 (n-2v)!} \left(\frac{\xi}{n} \right)^{v-1} \left(\frac{\eta}{n} \right)^{v-1} \left(1 - \frac{\xi}{n} - \frac{\eta}{n} \right)^{n-2v},$$

где $\xi > 0$, $\eta > 0$, $\xi + \eta < n$ и $2v < n$. При $n \rightarrow \infty$ это выражение стремится к

$$(28.6.7) \quad \frac{\xi^{v-1}}{\Gamma(v)} e^{-\xi} \cdot \frac{\eta^{v-1}}{\Gamma(v)} e^{-\eta},$$

так что в пределе ξ и η независимы.

Если функция распределения $F(x)$ задана, то иногда можно точно решить уравнения (28.6.2) и (28.6.5) относительно x и y . Тогда получим v -е значения x и y , выраженные через вспомогательные величины ξ и η , распределения которых известны. Если же точное решение невозможно, то часто можно получить асимптотическое решение для больших значений n . В таких случаях известные распределения величин ξ и η можно использовать для нахождения предельных форм распределений v -х значений, широты и т. д.

1. Прямоугольное распределение. Пусть величина, из значений которой извлекается выборка, равномерно распределена (см. параграф 19.1) по интервалу (a, b) . Если в выборке в n значений из этого распределения x и y суть v -е верхнее и нижнее значения, то формулы (28.6.2) и (28.6.5) дадут

$$x = b - \frac{b-a}{n} \xi, \quad y = a + \frac{b-a}{n} \eta,$$

где ξ и η имеют совместную плотность вероятности (28.6.6) с предельной формой (28.6.7). Поэтому получаем

$$E(x) = b - \frac{v}{n+1} (b-a), \quad D^2(x) = \frac{v(n-v+1)}{(n+1)^2(n+2)} (b-a)^2,$$

и аналогичные выражения для y . Далее имеем

$$(28.6.8) \quad E\left(\frac{x+y}{2}\right) = \frac{a+b}{2}, \quad D^2\left(\frac{x+y}{2}\right) = \frac{v}{2(n+1)(n+2)} (b-a)^2.$$

откуда видно, что среднее арифметическое v -х значений x и y является состоятельной и несмещенной оценкой (см. параграф 27.6) для среднего значения $\frac{a+b}{2}$ всего распределения. Наконец, для разности

$x - y$ имеет

$$(28.6.9) \quad E(x - y) = \left(1 - \frac{2v}{n+1}\right)(b - a),$$

$$D^2(x - y) = \frac{2v(n-2v+1)}{(n+1)^2(n+2)}(b-a)^2.$$

При $v=1$ разность $x - y$ является, конечно, широтой выборки.

2. Треугольное распределение. В случае треугольного распределения (см. параграф 19.1) по отрезку (a, b) , уравнения (28.6.2) и (28.6.5) при $x > \frac{a+b}{2}$, $y < \frac{a+b}{2}$ дают

$$x = b - (b-a) \sqrt{\frac{\xi}{2n}}, \quad y = a + (b-a) \sqrt{\frac{\eta}{2n}}.$$

Рассмотрим только частный случай $v=1$, когда x и y — экстремальные значения выборки. Тогда получим

$$(28.6.10) \quad E\left(\frac{x+y}{2}\right) = \frac{a+b}{2},$$

$$D^2\left(\frac{x+y}{2}\right) = \frac{4-\pi}{16n}(b-a)^2 + O\left(\frac{1}{n^2}\right),$$

$$E(x-y) = \left(1 - \sqrt{\frac{\pi}{2n}}\right)(b-a) + O\left(\frac{1}{n^{1/2}}\right);$$

$$D^2(x-y) = \frac{4-\pi}{4n}(b-a)^2 + O\left(\frac{1}{n^2}\right).$$

3. Распределение Коши. Для распределения, заданного плотностью вероятности (19.2.1), величина (28.6.2) имеет вид

$$\hat{z} = \frac{n\lambda}{\pi} \int_{-\infty}^{\infty} \frac{dt}{t^2 + (\mu - \lambda t)^2} = \frac{n}{\pi} \operatorname{arcctg} \frac{x-\mu}{\lambda},$$

или

$$x = \mu + \lambda \operatorname{ctg} \frac{\pi \xi}{n} = \mu + \frac{\lambda n}{\pi \hat{z}} + O\left(\frac{1}{n}\right),$$

где ξ имеет предельное распределение (28.6.4). Остаточный член склоняется по вероятности к нулю и поэтому из параграфа 20.6 следует, что v -е верхнее значение x в пределе распределено так же, как и $\mu + \frac{\lambda n}{\pi} v$,

где $v = \frac{1}{\xi}$ имеет плотность вероятности $\frac{1}{\Gamma(v)} v^{-v-1} e^{-\frac{1}{v}}$. Аналогично v -е нижнее значение y распределено так же, как и $\mu - \frac{\lambda n}{\pi} v$, где v в пре-

деле независима от v и имеет такое же, как v , распределение. В случае $v=1$ средние значения величин x и y бесконечны. При $v=2$ мы имеем

$$(28.6.11) \quad E\left(\frac{x+y}{2}\right) = \mu, \quad D^2\left(\frac{x+y}{2}\right) = \frac{1}{2(v-1)^2(v-2)} \left(\frac{\ln}{\pi}\right)^2 + O(n).$$

Мы замечаем, что дисперсия не стремится к нулю при $n \rightarrow \infty$. Соответственно этому, $\frac{x+y}{2}$ не сходится по вероятности к μ , так что $\frac{x+y}{2}$ не является состоятельной оценкой (см. параграф 27.6) для μ .

4. Распределение Лапласа. В случае плотности вероятности (19.2.4) для v -го верхнего значения x мы получаем, если $x > \mu$,

$$x = \mu + \lambda \log \frac{n}{2} - \lambda \log \xi,$$

где ξ имеет предельное распределение (28.6.4). Подставляя v вместо $-\log \xi$, получим

$$x = \mu + \lambda \log \frac{n}{2} + \lambda v,$$

где $v = -\log \xi$ имеет в пределе плотность вероятности

$$j_v(v) = \frac{1}{\Gamma(v)} e^{-v} v^{v-1}$$

Аналогично, v -е нижнее значение имеет вид

$$y = \mu - \lambda \log \frac{n}{2} - \lambda w,$$

где w в пределе независима от v и имеет плотность вероятности $j_v(w)$. В частном случае $v=1$ имеем (см. следующий пример)

$$(28.6.12) \quad E\left(\frac{x+y}{2}\right) = \mu, \quad D^2\left(\frac{x+y}{2}\right) = \frac{\lambda^2 \pi^2}{12} + O\left(\frac{1}{n}\right),$$

и мы замечаем, что, как и в предыдущем случае, $\frac{x+y}{2}$ не является состоятельной оценкой для μ .

5. Нормальное распределение. Рассмотрим сначала нормальное распределение с нормированными параметрами $m=0$ и $s=1$. Если x есть v -е верхнее значение выборки в n значений из этого распределения, то (28.6.2) имеет вид

$$\xi = \frac{n}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt.$$

Требуется найти асимптотическое решение x этого уравнения при большом n . Интегрируя по частям, приведем уравнение к виду

$$\frac{\xi \sqrt{2\pi}}{n} = \frac{1}{x} e^{-\frac{x^2}{2}} \left(1 + O\left(\frac{1}{x^2}\right) \right).$$

Предполагая, что ξ ограничено, после некоторых вычислений получим

$$x = \sqrt{2 \log n} - \frac{\log \log n + \log 4\pi}{2 \sqrt{2 \log n}} - \frac{\log \xi}{\sqrt{2 \log n}} + O\left(\frac{1}{\log n}\right),$$

откуда следует, что остаточный член сходится по вероятности к нулю.

Рассматривая общий случай нормального распределения с произвольными параметрами m и σ , воспользуемся лишь заменой x на $\frac{x-m}{\sigma}$.

Заменяя также $-\log \xi$ на v , мы таким образом получим, что верхнее y -е значение x может быть представлено в виде

$$(28.6.13) \quad x = m + \sigma \sqrt{2 \log n} - \sigma \frac{\log \log n + \log 4\pi}{2 \sqrt{2 \log n}} + \frac{\sigma}{\sqrt{2 \log n}} v,$$

где величина $v = -\log \xi$ имеет в пределе при $n \rightarrow \infty$ плотность вероятности

$$(28.6.14) \quad j_v(v) = \frac{1}{\Gamma(v)} e^{-v} v^{1-v},$$

с которой мы уже встречались в предыдущем примере. Аналогично для y -го нижнего значения y получим выражение

$$(28.6.15) \quad y = m - \sigma \sqrt{2 \log n} + \sigma \frac{\log \log n + \log 4\pi}{2 \sqrt{2 \log n}} - \frac{\sigma}{\sqrt{2 \log n}} w,$$

где w в пределе независима от v и имеет плотность вероятности $j_w(w)$.

Таким образом, при больших значениях n y -е значения x и y получаются с помощью простого линейного преобразования из величин, имеющих предельное распределение, определяемое плотностью вероятности (28.6.14). Графики плотности вероятности $w = j_w(w)$ для некоторых значений y изображены на фиг. 27.

Мы замечаем, что предельное распределение, с точностью до нормирования, имеет тот же вид, что и в предыдущем примере. Прямое обобщение вышенесложенных методов показывает, что та же самая плотность вероятности $j_w(w)$ встречается во всех случаях, когда плотность вероятности генеральной совокупности при больших значениях $|x|$ имеет асимптотическое выражение

$$f(x) \sim Ae^{-B|x|^p},$$

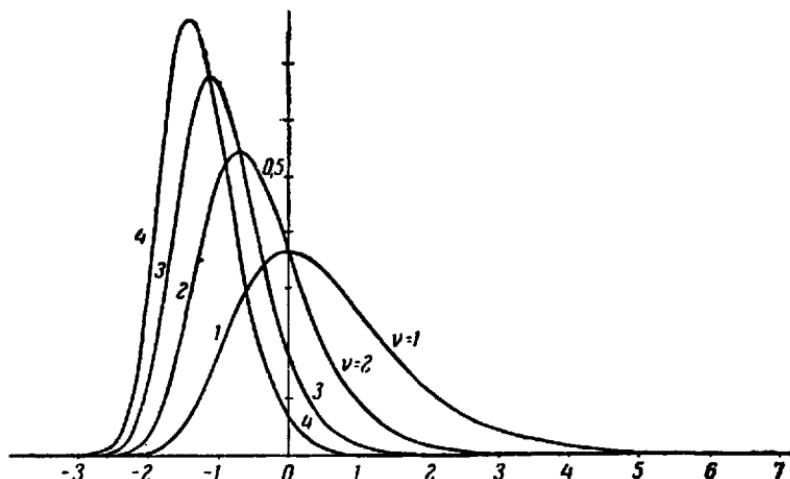
где A , B и p — положительные постоянные.

Модой величины, имеющей плотность вероятности $j_v(v)$, является $-\log v$, в то время как среднее значение и дисперсия этой величины даются выражениями

$$E(v) = \int_{-\infty}^{\infty} v j_v(v) dv = -\frac{1}{\Gamma(v)} \int_0^{\infty} \xi^{v-1} \log \xi e^{-\xi} d\xi = C - S_1,$$

$$D^2(v) = \int_{-\infty}^{\infty} v^2 j_v(v) dx - (C - S_1)^2 =$$

$$= \frac{1}{\Gamma(v)} \int_0^{\infty} \xi^{v-1} \log^2 \xi e^{-\xi} d\xi - (C - S_1)^2 = \frac{\pi^2}{6} - S_2,$$



Фиг. 27. Плотность вероятности $u = f_v(v)$ при $v = 1, 2, 3, 4$

получающимися с помощью (12.5.6) и (12.5.7). Здесь C обозначает постоянную Эйлера, определяемую формулой (12.2.7), а величины S_1 и S_2 имеют вид

$$S_1 = \frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{v-1}, \quad S_2 = \frac{1}{1^2} + \frac{1}{2^2} + \dots + \frac{1}{(v-1)^2}.$$

Отсюда получаем для ν -го верхнего значения x :

$$E(x) = m + \sigma \left(\sqrt{2 \log n} - \frac{\log \log n + \log 4\pi + 2(S_1 - C)}{2\sqrt{2 \log n}} + O\left(\frac{1}{\log n}\right) \right), \quad (28.6.16)$$

$$D^2(x) = \frac{\sigma^2}{2 \log n} \left(\frac{\pi^2}{6} - S_2 \right) + O\left(\frac{1}{\log^2 n}\right),$$

и аналогичные выражения для y -го нижнего значения. Далее, получаем

$$(28.6.17) \quad E\left(\frac{x+y}{2}\right) = m,$$

$$D^2\left(\frac{x+y}{2}\right) = \frac{\sigma^2}{4 \log n} \left(\frac{\pi^2}{6} - S'_1\right) + O\left(\frac{1}{\log^2 n}\right),$$

так что в этом случае $\frac{x+y}{2}$ является состоятельной оценкой для m , хотя дисперсия стремится к нулю лишь как $(\log n)^{-1}$, т. е. далеко не так быстро, как n^{-1} . Для разности $x - y$ между y -ми значениями получим

$$(28.6.18) \quad E(x-y) = 3 \left(\frac{4 \log n - \log \log n - \log 4\pi - 2(S_1 - C)}{1/2 \log n} + O\left(\frac{1}{\log n}\right) \right),$$

$$D^2(x-y) = \frac{\sigma^2}{\log n} \left(\frac{\pi^2}{6} - S_2\right) + O\left(\frac{1}{\log^2 n}\right).$$

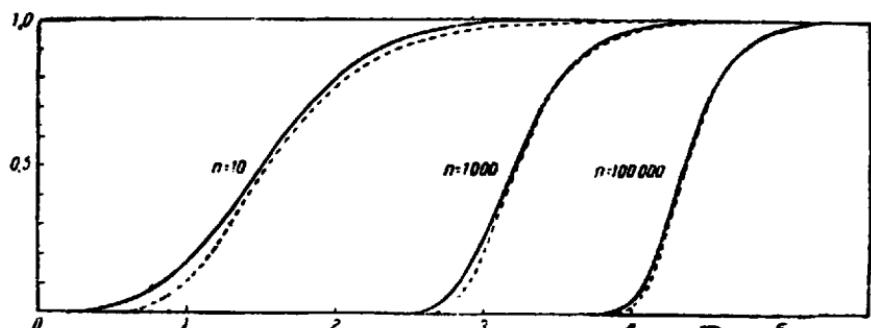
Таким образом, можно получить состоятельную оценку для σ , умножая $x - y$ на подходящую постоянную, и дисперсия этой оценки при заданном большом значении n будет приблизительно пропорциональна величине

$$\frac{\pi^2}{6} - S_2 = \sum_v^{\infty} \frac{1}{v^2}.$$

Предельные формы, исследованные выше в связи с нормальным распределением и распределением Лапласа, получены частично Р. Фишером и Типпеттом [110] и частично Гумбелем [120], в статьях которых можно найти дальнейшие сведения о свойствах этих распределений и их статистических приложениях.

В предельных выражениях для случая нормального распределения остаточные члены имеют такой же порядок, как отрицательная степень $\log n$. Поскольку $\log n$ стремится к бесконечности медленнее любой степени n , приближение к предельным формам здесь значительно медленнее, чем, например, в случае приближения к нормальности распределения какой-нибудь моментной характеристики. Точные распределения экстремальных значений и широты выборки в случае совокупности с нормальным распределением были изучены многими авторами, имеются также соответствующие таблицы. Читатель может обратиться

к таблицам К. Пирсона [264] и к статьям Ирвина [131], Типпетта [226], Е. Пирсона и Дэни [196], Е. Пирсона и Гартли [197]. На фиг. 28 приведено точное распределение для наибольшего члена выборки в сравнении с соответствующими распределениями, вычисленными по предельным выражениям (28.6.13) --- (28.6.14).



Фиг. 28. Функция распределения для верхнего экстремального значения выборки в n значений из нормальной совокупности с $\mu=0$ и $\sigma=1$, точная и вычисляемая по приближенной формуле (пунктир)

ГЛАВА 29

ТОЧНЫЕ ВЫБОРОЧНЫЕ РАСПРЕДЕЛЕНИЯ

29.1. Постановка проблемы. В двух предыдущих главах мы показали, как вычислять моменты и различные другие характеристики выборочных распределений, и изучили асимптотическое поведение распределений при неограниченном возрастании объема выборки. Однако ясно, что знание *точного* вида выборочного распределения имеет значительно большее значение, чем знание некоторого количества моментных характеристик распределения и предельного выражения при больших значениях n . Асимптотические выражения иногда резко расходятся с истинными распределениями, в особенности если мы имеем дело с *малыми выборками*, как это часто бывает в приложениях; в таких случаях знание точного вида распределения особенно желательно.

Предположим, что мы имеем дело с выборкой в n наблюдаемых значений из одномерного распределения с функцией распределения $F(x)$ и мы хотим отыскать выборочное распределение некоторой выборочной характеристики $g(x_1, \dots, x_n)$. Задача заключается в том,

чтобы отыскать распределение заданной функции $g(x_1, \dots, x_n)$ от n независимых случайных величин x_1, \dots, x_n , имеющих одну и ту же функцию распределения $F(x)$.

Теоретически эта задача была решена в параграфе 14.5, где было показано, что если заданы функции F и g , то всегда существует единственное решение. Задачу часто можно решить *численно*, вычисляя таблицы с помощью приближенных формул. Однако, если требуется, чтобы решение было выражено явно через известные функции, то положение резко меняется. При современном состоянии нашей науки такое решение можно получить лишь в сравнительно редких случаях.

Одним из случаев, в которых результат может быть получен с известной общностью, является простой случай *среднего значения* одномерной выборки $\bar{x} = \frac{1}{n} \sum_i x_i$. В главах 16—19 мы видели (см. параграфы 16.2, 16.5, 17.3, 18.1, 19.2), что многие распределения удовлетворяют так называемой *теореме сложения*, которая дает точное выражение для функции распределения $G_n(x)$ суммы $x_1 + \dots + x_n$, где x_i — независимые величины с одной и той же заданной функцией распределения $F(x)$. Тогда функцией распределения среднего значения \bar{x} является $G_n(nx)$, так что можно отыскать точное выборочное распределение среднего значения, если только распределение генеральной совокупности удовлетворяет теореме сложения. Приведем несколько примеров:

Если функция распределения $F(x)$ нормальна (m, σ), то, как мы видели в параграфе 17.3, среднее значение \bar{x} нормально $(m, \frac{\sigma}{\sqrt{n}})$.

Если $F(x)$ соответствует распределению Коши, то, согласно параграфу 19.2, \bar{x} имеет ту же самую функцию распределения $F(x)$, что и генеральная совокупность.

Если генеральная совокупность имеет распределение Пуассона с параметром λ , то среднее значение \bar{x} имеет возможные значения $0, \frac{1}{n}, \frac{2}{n}, \dots$, и из (16.5.4) следует, что $P(\bar{x} = \frac{v}{n}) = \frac{(n\lambda)^v}{v!} e^{-n\lambda}$.

Кроме случая среднего значения (см. Ирвин [132]) относительно точного вида выборочных распределений известно очень мало результатов общего характера. Лишь в одном частном случае, именно в случае *выбора из нормальной генеральной совокупности* (любого числа

измерений), удается систематически исследовать проблему и добиться достаточно полных результатов. В настоящей главе мы будем заниматься именно этим случаем.

Некоторые отдельные результаты, относящиеся к описанному здесь кругу идей, были получены на ранней стадии развития статистической теории Гельмлертом, К. Пирсоном и Стюартом. Однако впервые этот предмет был систематически изучен Р. Фишером, который дал строгие доказательства полученных ранее результатов и установил точный вид распределений в новых чрезвычайно важных случаях. В своей работе, посвященной этим проблемам, Фишер широко использует методы многомерной аналитической геометрии. Позже к проблемам описанного нами типа были применены и другие методы, как, например, использование характеристических функций, некоторые преобразования переменных и т. д. В дальнейшем мы приведем примеры использования различных методов.

29.2. Лемма Фишера. Степени свободы. При изучении выборочных распределений, связанных с нормально распределенными величинами, часто оказывается полезным следующее преобразование, полученное Р. Фишером [97]. Предположим, что x_1, \dots, x_n — независимые случайные величины, каждая из которых нормальна $(0, \sigma)$. Рассмотрим ортогональное преобразование (см. параграф 11.9)

$$(29.2.1) \quad y_i = c_{i1}x_1 + c_{i2}x_2 + \dots + c_{in}x_n, \quad (i = 1, 2, \dots, n).$$

вводящее вместо x_1, \dots, x_n новые величины y_1, \dots, y_n . Согласно параграфу 24.4, совместное распределение величин y_i нормально, и мы получим (см. упражнение 16, стр. 351) $E(y_i) = 0$ и

$$E(y_i y_k) = \sigma^2 \sum_{j=1}^n c_{ij} c_{kj} = \begin{cases} \sigma^2 & \text{при } i = k, \\ 0 & \text{при } i \neq k, \end{cases}$$

так что новые величины y_i некоррелированы. Тогда из параграфа 24.1 следует, что эти величины даже независимы. Таким образом, преобразованные величины y_i независимы и нормальны $(0, \sigma)$.

Геометрический смысл этого результата очевиден. Преобразование (29.2.1) соответствует (см. параграф 11.9) вращению координатной системы около начала координат, и наш результат означает, что рассмотренное здесь нормальное распределение частного вида в R_n инвариантно относительно такого вращения.

Теперь предположим, что сначала задано лишь некоторое число $p < n$ линейных функций y_1, y_2, \dots, y_p , где $y_i = c_{i1}x_1 + \dots + c_{in}x_n$, а c_{ij} удовлетворяют условиям ортогональности

$$\sum_{j=1}^n c_{ij}c_{kj} = \begin{cases} 1 & \text{при } i=k, \\ 0 & \text{при } i \neq k, \end{cases}$$

для $i = 1, 2, \dots, p; k = 1, 2, \dots, p$. Согласно параграфу 11.9, можно найти $n-p$ строк чисел c_{i1}, \dots, c_{in} , где $i = p+1, \dots, n$, таких, что полная матрица $C_{nn} = \{c_{ik}\}$ будет ортогональной. Рассмотрим квадратичную форму величин x_1, \dots, x_n

$$(29.2.2) \quad Q(x_1, \dots, x_n) = \sum_1^n x_i^2 - y_1^2 - \dots - y_p^2.$$

Если к этой форме применить ортогональное преобразование (29.2.1), то $\sum_1^n x_i^2$, согласно параграфу 11.9, преобразуется в $\sum_1^n y_i^2$, и мы получим

$$Q = y_{p+1}^2 + \dots + y_n^2.$$

Таким образом, Q равна сумме квадратов $n-p$ независимых нормальных $(0, \sigma)$ величин, которые, кроме того, не зависят от y_1, \dots, y_p . Используя (18.1.8), получим следующую лемму, доказанную Р. Фишером [97]:

Величина Q , определяемая формулой (29.2.2), независима от y_1, \dots, y_p и имеет плотность вероятности

$$\frac{1}{\sigma^2} k_{n-p} \left(\frac{x}{\sigma^2} \right) = \frac{1}{2^{\frac{n-p}{2}} \sigma^{n-p} \Gamma \left(\frac{n-p}{2} \right)} x^{\frac{n-p}{2}-1} e^{-\frac{x}{2\sigma^2}},$$

где $k_n(x)$ — плотность вероятности (18.1.3) распределения χ^2 .

Число $n-p$ есть *ранг* формы Q (см. параграф 11.6), т. е. наименьшее число независимых величин, к которым может быть приведена форма Q с помощью неособенного линейного преобразования. В статистических приложениях это число свободных величин в данной задаче обычно называется, в соответствии с терминологией, введенной Р. Фишером, числом *степеней свободы* задачи или распределения случайных величин, рассматриваемых в этой задаче.

Так, например, величина $\chi^2 = \sum_1^n \xi_i^2$ и ее плотность вероятности $k_n(x)$, рассмотренные в параграфе 18.1, имеют n степеней свободы, так как квадратичная форма χ^2 имеет ранг n . Соответствующее распределение называется *распределением χ^2 с n степенями свободы*.

Аналогично, рассмотренная выше форма $Q = \sum_1^n x_i^2 - y_1^2 - \dots - y_p^2$ ранга $n-p$ имеет $n-p$ степеней свободы, и доказанный выше результат означает, что величина $\frac{Q}{\sigma^2}$ имеет *распределение χ^2 с $n-p$ степенями свободы*.

Такая же терминология часто употребляется и для других распределений. В случае распределения Стьюдента принято говорить, что плотность вероятности $s_n(x)$, определяемая формулой (18.2.4), соответствует *распределению Стьюдента с n степенями свободы*, так как квадратичная форма в знаменателе величины t , определяемой формулой (18.2.1), имеет ранг n . Для z -распределения Фишера (см. параграф 18.3) мы будем делать различие между m степенями свободы в *числителе* формулы (18.3.1) и n степенями свободы в ее *знаменателе*.

29.3. Совместное распределение величины \bar{x} и s^2 в выборках из нормального распределения. В параграфе 29.1 мы уже указали, что среднее значение \bar{x} выборки в n элементов из совокупности с нормальным (m, σ) распределением само нормально $(m, \frac{\sigma}{\sqrt{n}})$. Теперь рассмотрим распределение выборочной дисперсии $s^2 = m_2 = \frac{1}{n} \sum (x_i - \bar{x})^2$ и, одновременно, совместное распределение величин \bar{x} и s^2 . Можно предполагать без ограничения общности, что среднее значение совокупности m равно нулю, так как это не меняет s^2 и эквивалентно прибавлению к некоторой постоянной.

Таким образом, предположим, что каждое x_i нормально $(0, \sigma)$, и рассмотрим тождество (см. (11.11.2))

$$(29.3.1) \quad ns^2 = \sum_1^n (x_i - \bar{x})^2 = \sum_1^n x_i^2 - n\bar{x}^2.$$

Выражение $n\bar{x}^2 = \left(\frac{x_1}{\sqrt{n}} + \dots + \frac{x_n}{\sqrt{n}} \right)^2$ есть квадрат линейной

формы $c_1x_1 + \dots + c_nx_n$, для которой $c_1^2 + \dots + c_n^2 = 1$. Поэтому можно применить лемму предыдущего параграфа, положив в (29.2.2) $p=1$ и $y_1 = \sqrt{n} \bar{x}$. Возвращаясь к общему случаю среднего значения совокупности m , получим следующую теорему, впервые строго доказанную Р. Фишером [97]:

Среднее значение \bar{x} и дисперсия s^2 нормальной выборки независимы, причем \bar{x} нормально $\left(m, \frac{c}{\sqrt{n}}\right)$, а $\frac{ns^2}{c^2}$ имеет распределение χ^2 с $n-1$ степенями свободы.

Можно показать, что независимость \bar{x} и s^2 имеет место только в случае нормальной генеральной совокупности (см. Джейри [115] и Лукас [150]). С другой стороны, в параграфе 27.4 мы видели, что \bar{x} и s^2 некоррелированы, если только третий центральный момент μ_3 генеральной совокупности равен нулю.

Из доказанной нами теоремы следует, что несмещенная оценка (см. параграф 27.6) для дисперсии, т. е. величина $\frac{n}{n-1}s^2$, имеет плотность вероятности $\frac{n-1}{\sigma^2} k_{n-1} \left(\frac{(n-1)x}{\sigma^2} \right)$. Сравнивая последнюю с плотностью вероятности величины $\frac{1}{n} \sum_1^n x_i^2$, данной в таблице в конце параграфа 18.1, мы увидим, что величина $\frac{n}{n-1}s^2 = \frac{1}{n-1} \sum_1^n (x_i - \bar{x})^2$ распределена так же, как среднее арифметическое $n-1$ квадратов независимых нормальных $(0, \sigma)$ величин, в соответствии с тем фактом, что распределение имеет $n-1$ степеней свободы.

Среднее значение и дисперсия величины $s^2 = m_2$ уже были получены ранее (см. формулу (27.4.5)). С помощью формулы (18.1.5) мы получаем следующее общее выражение для моментов:

$$(29.3.2) \quad E(m_2) = \frac{(n-1)(n+1)(n+3)\dots(n+2v-3)}{n^v} \sigma^{2v}.$$

Отсюда выводятся выражения для коэффициентов асимметрии и эксцесса:

$$\gamma_1(m_2) = \frac{2\sqrt{2}}{\sqrt{n-1}}, \quad \gamma_2(m_2) = \frac{12}{n-1}.$$

Для стандартного отклонения выборки $s = \sqrt{m_2}$ получаем, с помощью

соответствующей общей теоремы и используя формулу Стирлинга (12.5.3),

$$(29.3.3) \quad E(s) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \sqrt{\frac{2}{n}} s = s + O\left(\frac{1}{n}\right),$$

$$D^2(s) = \left(\frac{n-1}{n} - \frac{\Gamma_2\left(\frac{n}{2}\right)}{\Gamma_2\left(\frac{n-1}{2}\right)} \frac{2}{n} \right) s^2 = \frac{s^2}{2n} + O\left(\frac{1}{n^2}\right),$$

в согласии с общими выражениями (27.7.1) и (27.7.2).

Вследствие большого значения теоремы о совместном распределении величин \bar{x} и s^2 дадим для нее еще одно доказательство, комбинируя использование некоторых преобразований переменных с геометрическими соображениями. Как и прежде, при доказательстве будем предполагать, что $m=0$.

Рассмотрим n -мерное пространство выборок R_n величин x_1, \dots, x_n . Наша выборка изображается в этом пространстве переменной точкой — выборочной точкой $X = X(x_1, \dots, x_n)$. Пусть XR — перпендикуляр, опущенный из точки X на линию $x_1 = x_2 = \dots = x_n$. Тогда R имеет координаты $(\bar{x}, \dots, \bar{x})$, так что квадрат расстояния OR между точкой R и началом координат O равен $n\bar{x}^2$, и, следовательно.

$$\overline{XR}^2 = \overline{OX}^2 - \overline{OR}^2 = \sum_1^n x_i^2 - n\bar{x}^2 = ns^2.$$

Совместное распределение величин x_i рассматривается обычным образом, как распределение единичной массы по R_n , и дифференциал вероятности этого распределения равен

$$dP = \frac{1}{(2\pi)^{\frac{n}{2}} s^n} e^{-\frac{1}{2s^2} \sum_1^n x_i^2} dx_1 \dots dx_n.$$

Повернем координатные оси так, чтобы одна из них совпала с прямой OR . Этот поворот выражается ортогональным преобразованием $y_i = \sum_1^n c_{ij} x_j$, где одна из величин y_i , например, y_n , равна $\sqrt{n}\bar{x} =$

$= \frac{x_1}{\sqrt{n}} + \dots + \frac{x_n}{\sqrt{n}}$. Отсюда получим $\sum_1^n x_i^2 = \sum_1^n y_i^2 = n\bar{x}^2 + \sum_1^{n-1} y_i^2$, и, далее, $\sum_1^n y_i^2 = ns^2$. Детерминант преобразования равен или $+1$,

или — 1, и по формуле (22.2.3) имеем

$$\begin{aligned} dP &= \frac{1}{(2\pi)^{\frac{n}{2}} s^n} e^{-\frac{n}{2s^2}(\bar{x}^2 + s^2)} dy_1 \dots dy_{n-1} dy_n = \\ &= \frac{\sqrt{n}}{(2\pi)^{\frac{n}{2}} s^n} e^{-\frac{n}{2s^2}(\bar{x}^2 + s^2)} dy_1 \dots dy_{n-1} d\bar{x}. \end{aligned}$$

Сделаем замену переменных

$$(29.3.4) \quad y_i = \sqrt{n} s z_i, \quad (i = 1, 2, \dots, n-1),$$

которая означает, что мы принимаем длину $XR = \sqrt{n}s$ за единицу масштаба. Однако при последней замене переменных вместо $n-1$ величин y_i получаем n новых величин s и z_1, \dots, z_{n-1} . Соответственно этому, между новыми переменными существует соотношение, которое можно найти, возведя в квадрат и складывая $n-1$ уравнений (29.3.4). При этом получим

$$(29.3.5) \quad \sum_{i=1}^{n-1} z_i^2 = 1,$$

так что одна из величин z_i , например, z_{n-1} , может быть выражена как функция от $n-2$ остальных z_i , так что в (29.3.4) старые переменные y_1, \dots, y_{n-1} заменяются новыми переменными s и z_1, \dots, z_{n-1} .

Так как $\frac{\partial z_{n-1}}{\partial z_i} = -\frac{z_i}{z_{n-1}}$, то якобиан J преобразования имеет вид

$$J = \begin{vmatrix} \sqrt{n} z_1 & \sqrt{n} s & 0 & \dots & 0 \\ \sqrt{n} z_2 & 0 & \sqrt{n} s & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \sqrt{n} z_{n-2} & 0 & 0 & \dots & \sqrt{n} s \\ \sqrt{n} z_{n-1} - \sqrt{n} s \frac{z_1}{z_{n-1}} - \sqrt{n} s \frac{z_2}{z_{n-1}} - \dots - \sqrt{n} s \frac{z_{n-2}}{z_{n-1}} & & & & \end{vmatrix} =$$

$$= \frac{n^{\frac{n-1}{2}} s^{n-2}}{z_{n-1}} \begin{vmatrix} z_1 & 1 & 0 & \dots & 0 \\ z_2 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ z_{n-2} & 0 & 0 & \dots & 1 \\ z_{n-1}^2 - z_1 - z_2 - \dots - z_{n-2} & & & \dots & \end{vmatrix} =$$

$$= (-1)^{n-1} \frac{n^{\frac{n-1}{2}} s^{n-2}}{z_{n-1}} = \pm \frac{n^{\frac{n-1}{2}} s^{n-2}}{\sqrt{1 - z_1^2 - \dots - z_{n-2}^2}}$$

Для любой системы значений $(y_1, \dots, y_{n-1}) \neq (0, \dots, 0)$ мы получаем по формулам (29.3.4) и (29.3.5) однозначно определенную систему значений z_1, \dots, z_{n-2} и s , для которой $s > 0$. С другой стороны, любой системе значений z_1, \dots, z_{n-2} и s , для которой $\sum_1^{n-2} z_i^2 < 1$ и $s > 0$, соответствуют два значения z_{n-1} , отличающиеся знаком и определяемые из формулы (29.3.5), именно $z_{n-1} = \pm \sqrt{1 - z_1^2 - \dots - z_{n-2}^2}$, а поэтому две системы значений y_i , именно $y_1, \dots, y_{n-2}, \pm y_{n-1}$. Обе эти системы дают одно и то же значение дифференциала вероятности dP и модуля $|J|$ якобиана, так что с помощью замечания к параграфу 22.2 получаем выражение

$$\begin{aligned} dP &= \frac{2n^{n/2}}{(2\pi)^{n/2}} e^{-\frac{n}{2s^2}(\bar{x}^2 + s^2)} \frac{s^{n-2}}{\sqrt{1 - z_1^2 - \dots - z_{n-2}^2}} d\bar{x} ds dz_1 \dots dz_{n-2} = \\ &= \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n\bar{x}^2}{2s^2}} d\bar{x} \cdot \frac{2 \left(\frac{n}{2}\right)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} s^{n-2} e^{-\frac{n}{2s^2}s^2} ds \times \\ &\quad \times \frac{\Gamma\left(\frac{n-1}{2}\right)}{\frac{n-1}{2}} \frac{dz_1 \dots dz_{n-2}}{\sqrt{1 - z_1^2 - \dots - z_{n-2}^2}}. \end{aligned}$$

Элемент вероятности dP получается здесь в виде произведения трех множителей, именно элементов вероятности для \bar{x} и s и совместного элемента вероятности величин z_1, \dots, z_{n-2} . Таким образом, мы видим (см. (22.1.2)), что \bar{x} и s не зависят не только друг от друга, но также и от составной величины (z_1, \dots, z_{n-2}) , и что распределения величин \bar{x} и s те же самые, что и распределения, даваемые приведенной выше теоремой *).

Для дальнейших целей в заключение укажем, что в общем случае, когда среднее значение совокупности m не равно нулю, приведенное

*). Этот же результат можно получить с помощью преобразования $x_i = \bar{x} + sz_i$, которое использовали для этой и для других целей Беренс [60], Стеффенсен [218], Раш и Хальд [206].

выше преобразование элемента вероятности можно записать в виде

$$(29.3.6) \quad dP = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - m)^2} dx_1 \dots dx_n =$$

$$= \frac{\sqrt{n}}{\sigma \sqrt{2\pi}} e^{-\frac{n}{2\sigma^2} (\bar{x} - m)^2} d\bar{x} \cdot \frac{2 \left(\frac{n}{2}\right)^{\frac{n-1}{2}}}{\pi^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)} s^{n-2} e^{-\frac{n}{2\sigma^2} s^2} ds \times$$

$$\times \frac{\Gamma\left(\frac{n-1}{2}\right)}{\pi^{\frac{n-1}{2}}} \frac{dz_1 \dots dz_{n-1}}{\sqrt{1 - z_1^2 - \dots - z_{n-1}^2}}.$$

Рассмотрим результат этого преобразования для выражения

$$\frac{m_v}{m_2^{\frac{v+2}{2}}} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^v, \quad (v > 2).$$

С помощью тождества (29.3.1) легко показать, что каждое $x_i - \bar{x}$ преобразуется в линейную комбинацию величин y_1, \dots, y_{n-1} . Тогда из формул 29.3.4 и 29.3.5 следует, что $m_v m_2^{-\frac{v+2}{2}}$ есть функция только от z_1, \dots, z_{n-1} . Таким образом, три величины \bar{x}, s и $m_v m_2^{-\frac{v+2}{2}}$ независимы (см. Джейри [116]).

Следуя Джейри, можно использовать этот результат для получения точных выражений (впервые полученных Фишером [101]) для среднего значения и дисперсия коэффициентов $g_1 = m_v m_2^{-\frac{v+2}{2}}$ и $g_2 = m_v m_2^{-\frac{v+2}{2}} - 3$, в естественных асимптотических выражений (27.7.9). Действительно, из теоремы о независимости следует, что

$$E(m_v^p m_2^{-\frac{v+2}{2}}) E(m_2^{\frac{v+2}{2}}) = E(m_v^p),$$

так что среднее значение величины $(m_v m_2^{-\frac{v+2}{2}})^p$ можно вычислить по $E(m_v^p)$ и $E(m_2^{\frac{v+2}{2}})$.

Таким образом, получим

$$(29.3.7) \quad E(g_1) = 0, \quad E(g_2) = -\frac{6}{n+1},$$

$$D^2(g_1) = \frac{6(n-2)}{(n+1)(n+3)},$$

$$D^2(g_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}.$$

Таким образом, g_2 получается с отрицательным смещением порядка $n-1$, а g_1 является несмещенной. Если вместо g_1 и g_2 мы рассмотрим аналогичные величины

$$(29.3.8) \quad G_1 = \frac{K_3}{K_2^{\eta_2}} = \frac{\sqrt{n(n-1)}}{n-2} g_1,$$

$$G_2 = \frac{K_4}{K_2^2} = \frac{n-1}{(n-2)(n-3)} \left[(n+1) g_2 + 6 \right],$$

где K_v — несмешенные оценки семи-инвариантов по Фишеру (см. параграф 27.6), то смещение пропадает, и получим

$$(29.3.9) \quad E(G_1) = E(G_2) = 0,$$

$$D^2(G_1) = \frac{6n(n-1)}{(n-2)(n+1)(n+3)},$$

$$D^2(G_2) = \frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}.$$

29.4. Стьюдентово отношение. Рассмотрим величины $\sqrt{n}(\bar{x} - m)$

и $\frac{n}{n-1}s^2$ в том случае, когда распределение генеральной совокупности нормально (m , σ). В соответствии с предыдущим параграфом, эти две величины независимы, причем $\sqrt{n}(\bar{x} - m)$ нормальна $(0, \sigma)$, а $\frac{n}{n-1}s^2$ распределена так же, как среднее арифметическое $n-1$ квадратов независимых нормальных $(0, \sigma)$ величин. По определению распределения Стьюдента в параграфе 18.2, отношение

$$(29.4.1) \quad t = \frac{\sqrt{n}(\bar{x} - m)}{\sqrt{\frac{n}{n-1}s^2}} = \sqrt{n-1} \frac{\bar{x} - m}{s}$$

имеет *распределение Стьюдента с $n-1$ степенями свободы*. Таким образом, t имеет плотность вероятности

$$s_{n-1}(x) = \frac{1}{\sqrt{(n-1)\pi}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}$$

Это можно, конечно, доказать и непосредственно. Предположив, для простоты, что $m=0$, заменим выборочные величины x_1, \dots, x_n новыми переменными y_1, \dots, y_n с помощью ортогонального преобразо-

вания, при котором $y_1 = \sqrt{n} \bar{x} = \frac{x_1}{\sqrt{n}} + \dots + \frac{x_n}{\sqrt{n}}$. Тогда $ns^2 = \sum_1^n x_i^2 - n\bar{x}^2 = \sum_2^n y_i^2$, так что

$$t = \frac{y_1}{\sqrt{\frac{1}{n-1} \sum_2^n y_i^2}},$$

причем, согласно параграфу 29.2, величины y_i независимы и нормальны $(0, \sigma)$. Поэтому можно непосредственно применить доказательство формул (18.2.1) — (18.2.4).

Если в первом выражении для t в формуле (29.4.1) заменить величину $\frac{n}{n-1}s^2$ ее средним значением σ^2 , то получим величину $\sqrt{n} \frac{\bar{x} - m}{\sigma}$, которая, очевидно, нормальна $(0, 1)$. Из параграфа 20.6 следует, что разность $t = \sqrt{n} \frac{\bar{x} - m}{\sigma}$ сходится по вероятности к пулю при $n \rightarrow \infty$. В соответствии с формулой (20.2.2), плотность вероятности для t стремится к $\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ при $n \rightarrow \infty$.

Величина t , определяемая формулой (29.4.1), известна под изнанием *стъюдентова отношения**). Ее распределение впервые было получено Стъюдентом [221], результаты которого были затем строго доказаны Р. Фишером [97].

Как уже было указано в параграфе 18.2, плотность вероятности s_{n-1} , так же как и сама величина t , не содержит σ . Как только известно m , можно вычислить t по выборочным значениям и сравнить вычисленное значение t с теоретическим распределением. Таким путем мы получаем практически важный критерий значимости для *отклонения выборочного среднего значения \bar{x} от некоторого гипотетического значения среднего значения совокупности m* (см. параграфы 31.2 и 31.3, пример 4).

Чрезвычайно большое практическое значение имеет применение распределения Стъюдента в критерии значимости *разности между двумя средними значениями* (Р. Фишер [97]; см. параграф 31.2).

* Стъюдент рассматривал величину $z = \frac{t}{\sqrt{n-1}} = \frac{\bar{x} - m}{s}$.

Относящееся к этой проблеме выборочное распределение получается следующим образом.

Предположим, что даны две независимые выборки x_1, \dots, x_{n_1} и y_1, \dots, y_{n_2} , извлеченные из некоторой нормальной совокупности. Без ограничения общности можно предположить, что $m=0$. Обозначим $\bar{x} = \frac{1}{n_1} \sum_{i=1}^{n_1} x_i$ и $s_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (x_i - \bar{x})^2$ среднее значение и дисперсию первой выборки, а \bar{y} и s_2^2 — соответствующие характеристики второй выборки. Заменим все $n_1 + n_2$ величин $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}$ новыми величинами $z_1, \dots, z_{n_1+n_2}$ с помощью ортогонального преобразования, при котором $z_1 = \sqrt{n_1} \bar{x}$ и $z_2 = \sqrt{n_2} \bar{y}$. Тогда квадратичная форма

$$Q = n_1 s_1^2 + n_2 s_2^2 = \sum_{i=1}^{n_1} x_i^2 + \sum_{i=1}^{n_2} y_i^2 - n_1 \bar{x}^2 - n_2 \bar{y}^2$$

преобразуется в $Q = \sum_{i=3}^{n_1+n_2} z_i^2$, откуда видно, что ранг (или число степеней свободы) формы Q равен $n_1 + n_2 - 2$. Если определить случайную величину u соотношением

$$(29.4.2) \quad u = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \frac{\bar{x} - \bar{y}}{\sqrt{Q}} = \\ = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \frac{\bar{x} - \bar{y}}{\sqrt{n_1 s_1^2 + n_2 s_2^2}},$$

то эта величина u при нашем ортогональном преобразовании выражается так:

$$u = \frac{\sqrt{\frac{n_2}{n_1 + n_2}} z_1 - \sqrt{\frac{n_1}{n_1 + n_2}} z_2}{\sqrt{\frac{1}{n_1 + n_2 - 2} \sum_{i=3}^{n_1+n_2} z_i^2}} = \frac{w}{\sqrt{\frac{1}{n_1 + n_2 - 2} \sum_{i=3}^{n_1+n_2} z_i^2}},$$

где w и $z_3, \dots, z_{n_1+n_2}$ — независимые и нормальные $(0, \sigma)$ величины. Теперь можно снова применить рассуждения параграфа 18.2, из которых будет следовать, что величина u имеет распределение Стьюдента с $n_1 + n_2 - 2$ степенями свободы, так что u имеет плотность вероят-

ности $s_{n_1+n_2-2}(x)$. Этот результат справедлив, очевидно, независимо от значения m . Можно заметить, что в этом случае при величина a , ни соответствующая плотность вероятности не содержит какого-нибудь из параметров m и σ при пределении генеральной совокупности. Таким образом, можно вычислить a непосредственно по выборочным значениям и сравчить полученное значение a с теоретическим распределением (см. параграфы 31.2 и 31.3, пример 4).

Рассмотрим квадратичную форму $ns^2 = \sum_1^n (x_i - \bar{x})^2 = \sum_1^n x_i^2 - n\bar{x}^2$ относительно n выборочных переменных x_1, \dots, x_n , предполагая, что среднее значение совокупности m равно нулю. Заменим x_i новыми переменными y_i с помощью ортогонального преобразования, при котором первыми двумя новыми переменными будут

$$y_1 = \sqrt{n} \bar{x} = \frac{x_1}{\sqrt{n}} + \frac{x_2}{\sqrt{n}} + \dots + \frac{x_n}{\sqrt{n}},$$

$$y_2 = \sqrt{\frac{n}{n-1}} (x_1 - \bar{x}) = \sqrt{\frac{n-1}{n}} x_1 - \frac{x_2}{\sqrt{n(n-1)}} - \dots - \frac{x_n}{\sqrt{n(n-1)}},$$

мы преобразуем форму ns^2 к виду $\sum_2^n y_i^2$. Следовательно, величина

$$(29.4.3) \quad t = \frac{x_1 - \bar{x}}{s},$$

которая выражает отклонение выборочного значения x_1 от выборочного среднего значения \bar{x} , измеренное стандартным отклонением выборки s , принятым за единицу масштаба, после нашего преобразования примет вид

$$t = -\sqrt{\frac{y_2}{\frac{1}{n-1} \sum_2^n y_i^2}}.$$

Но величины y_2, \dots, y_n независимы и нормальны $(0, s)$, так что, в силу формул (18.2.6) и (18.2.7), величина t имеет плотность вероятности (см. Томсон [225] и Арлей [53])

$$(29.4.4) \quad \frac{1}{V(n-1)\pi} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \left(1 - \frac{x^2}{n-1}\right)^{\frac{n-4}{2}}, \quad (|x| < V\sqrt{n-1}).$$

Тогда величина $\frac{t V\sqrt{n-2}}{V\sqrt{n-1-t^2}}$, согласно параграфу 18.2, имеет распределение Стьюдента с $n-2$ степенями свободы. Из определения величины t следует,

что эти результаты справедливы независимо от значения m . Каждое относительное отклонение $\frac{x_i - \bar{x}}{s}$ имеет, конечно, такое же распределение, как и т. Эти результаты имеют большое значение в вопросе о критерии для браковки наблюдаемых выборочных значений, лежащих далеко от среднего значения.

Вообще, если рассмотреть средние арифметические $\bar{x}_k = \frac{x_1 + \dots + x_k}{k}$,

где $1 \leq k \leq n$, и положить $t_k = \frac{\bar{x}_k - \bar{x}}{s}$, то величина $t_k \sqrt{\frac{k(n-1)}{n-k}}$ имеет плотность вероятности (29.4.4), и, следовательно, величина

$$(29.4.5) \quad t = \frac{t_k \sqrt{k(n-1)}}{\sqrt{n-k-k^2/k}}$$

имеет распределение Стьюдента с $n-2$ степенями свободы (Томсон [225]). Это можно использовать для проверки значимости разности между средним значением подгруппы и средним значением всей совокупности (см. параграф 31.3, пример 5).

29.5. Лемма. Теперь приступим к изучению выборочных распределений, связанных с *многомерной нормальной генеральной совокупностью*. В этом предварительном параграфе мы докажем некоторые результаты, полученные Висхартом [240] и Бартлеттом [241], которые потребуются нам в дальнейшем. Пусть $A = \begin{Bmatrix} a_{11} & \dots & a_{1k} \\ \dots & \dots & \dots \\ a_{k1} & \dots & a_{kk} \end{Bmatrix}$, где

$a_{ji} = a_{ij}$ — положительно-определенная матрица (см. параграф 11.10)

с постоянными элементами, а $X = \begin{Bmatrix} x_{11} & \dots & x_{1k} \\ \dots & \dots & \dots \\ x_{k1} & \dots & x_{kk} \end{Bmatrix}$, где $x_{ji} = x_{ij}$,

— переменная матрица. В силу симметрии X содержит, конечно, лишь $\frac{k(k+1)}{2}$ различных переменных x_{ij} . Детерминанты этих матриц будем обозначать $A = |a_{ij}|$ и $X = |x_{ij}|$.

Рассмотрим теперь $\frac{k(k+1)}{2}$ -мерное пространство $R_{\frac{k(k+1)}{2}}$ вели-

чин x_{ij} , где $k \geq 1$. Пусть S обозначает множество всех точек этого пространства, в которых соответствующая матрица X положительно определена, а S^* — дополнительное множество. Для любого $n > k$

определен следующую функцию от переменных x_{ij} :

$$(29.5.1) \quad f_n(x_{11}, \dots, x_{kk}) = \begin{cases} C_{kn} A^{\frac{n-1}{2}} X^{\frac{n-k-2}{2}} e^{-\sum_{i,j} a_{ij} x_{ij}} & \text{в } S, \\ 0 & \text{в } S^*, \end{cases}$$

где C_{kn} — постоянные, зависящие от k и n , но не от a_{ij} или x_{ij} . Сумма в показателе при e распространяется на $i=1, 2, \dots, k$ и $j=1, 2, \dots, k$.

Теперь покажем, что постоянные C_{kn} можно определить так, чтобы $f_n(x_{11}, \dots, x_{kk})$ была плотностью вероятности некоторого распределения в $R_{k \frac{(k+1)}{2}}$. Действительно, за C_{kn} следует взять

$$(29.5.2) \quad C_{kn} = \frac{1}{\pi^{\frac{k(k-1)}{4}} \Gamma\left(\frac{n-1}{2}\right) \Gamma\left(\frac{n-2}{2}\right) \dots \Gamma\left(\frac{n-k}{2}\right)}.$$

При $k=1$ формулы (29.5.1) — (29.5.2) дадут

$$f_n(x) = \frac{a^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} x^{\frac{n-3}{2}} e^{-ax}, \quad (x > 0, a > 0),$$

и эта функция является, очевидно, плотностью вероятности в R_1 .

При $k > 1$ следует показать, что C_{kn} могут быть определены таким образом, чтобы интеграл от f_n по всему пространству $R_{k \frac{(k+1)}{2}}$

равнялся 1. Сначала рассмотрим частный случай, когда A — диагональная матрица (см. параграф 11.1), так что $a_{ij} = 0$ при $i \neq j$. Поскольку A положительно-определенна, $a_{ii} > 0$ при $i = 1, \dots, k$. В любой точке множества S мы имеем $x_{ii} > 0$ при $i = 1, \dots, k$. Приведя для любого x_{ij} с $i \neq j$ подстановку

$$(29.5.3) \quad x_{ij} = y_{ij} \sqrt{x_{ii} x_{jj}},$$

получим $y_{ji} = y_{ij}$ и $X = DYD$, где D обозначает диагональную матрицу с элементами $\sqrt{x_{11}}, \sqrt{x_{22}}, \dots, \sqrt{x_{kk}}$, а

$$Y = \begin{pmatrix} 1 & y_{12} & \dots & y_{1k} \\ y_{21} & 1 & \dots & y_{2k} \\ \dots & \dots & \dots & \dots \\ y_{k1} & y_{k2} & \dots & 1 \end{pmatrix}.$$

Обозначив через Y детерминант матрицы Y , таким образом, имеем $X = x_{11}x_{22} \dots x_{kk}Y$. Если X положительно определена, то это же верно и для Y , и обратно. Якобиан преобразования (29.5.8) равен

$(x_{11}x_{22} \dots x_{kk})^{\frac{k-1}{2}}$, так что имеем

$$\begin{aligned} & \int_S X^{\frac{n-k-2}{2}} e^{-\sum_1^k a_{ii}x_{ii}} dx_1 dx_2 \dots dx_{kk} = \\ & = \int_0^\infty \dots \int_0^\infty (x_{11}x_{22} \dots x_{kk})^{\frac{n-3}{2}} e^{-\sum_1^k a_{ii}x_{ii}} dx_1 dx_{22} \dots \\ & \quad \dots dx_{kk} \int_{S'} Y^{\frac{n-k-2}{2}} dy_{12} \dots dy_{k-1,k}, \end{aligned}$$

где интегрирование по y_{ij} распространяется на множество S' всех y_{ij} , для которых Y положительно определена. Очевидно, интеграл по переменным y_{ij} , который мы обозначим J_k , зависит только от k и n , так что весь интеграл по множеству S сводится к

$$\frac{\left[\Gamma\left(\frac{n-1}{2}\right) \right]^k J_k}{(a_{11}a_{22} \dots a_{kk})^{\frac{n-1}{2}}} = \frac{H_{kn}}{A^{\frac{n-1}{2}}},$$

где H_{kn} зависят только от k и n . Если в (29.5.1) положить $C_{kn} = H_{kn}^{-1}$, то интеграл от $f_n(x_{11}, \dots, x_{kk})$ по всему пространству $R_{\frac{k(k+1)}{2}}$ будет равен 1, так что f_n (которая, очевидно, неотрицательна) является плотностью вероятности некоторого распределения в $R_{\frac{k(k+1)}{2}}$.

Чтобы завершить доказательство в случае, когда $a_{ij} = 0$ при $i \neq j$, остается проверить выражение (29.5.2) для C_{kn} . Из вышеприведенного следует, что надо доказать равенство

$$J_k = \int_S Y^{\frac{n-k-2}{2}} dy_{12} \dots dy_{k-1,k} = \pi^{\frac{k(k-1)}{4}} \prod_{i=1}^k \frac{\Gamma\left(\frac{n-i}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

при $2 \leq k < n$. Это можно доказать по индукции, и мы лишь наметим общий путь доказательства. При $k = 2$ наше соотношение сводится к соотношению

$$J_2 = \int_{-1}^{+1} (1-y^2)^{\frac{n-4}{2}} dy = V \pi \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)},$$

которое может быть проверено непосредственно, так как подстановка $y^2 = z$ превращает этот интеграл в бета-функцию (см. параграф 2.4). Предположим теперь, что наше соотношение доказано для некоторого значения k , и рассмотрим J_{k+1} . Раскрывая детерминант, стоящий под знаком интеграла, по формуле (11.5.3) получим для J_{k+1} выражение

$$\int_{S'} dy_1 \dots dy_{k-1,k} \int \left(Y - \sum_{i,j=1}^k y_{ij} y_{l,k+1} y_{j,k+1} \right)^{\frac{n-k-3}{2}} dy_{l,k+1} \dots dy_{k,k+1},$$

где интегрирование по переменным $y_{l,k+1}$ распространяется на все значения этих переменных, для которых $\sum_{i,j=1}^k y_{ij} y_{l,k+1} < Y$. Последний интеграл можно вычислить таким же способом, как и интегралы (11.12.3) — (11.12.4), и получим

$$J_{k+1} = J_k \frac{\Gamma\left(\frac{n-k-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{k}{\pi^{\frac{k}{2}}} = \pi^{\frac{k(k+1)}{2}} \prod_{l=1}^{k+1} \frac{\Gamma\left(\frac{n-l}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}.$$

Таким образом, наше соотношение остается справедливым для $k+1$, чем доказательство закончено.

В общем случае, когда A — любая положительно определенная матрица, рассмотрим преобразование

$$(29.5.4) \quad C'AC = B, \quad C'XC = Y,$$

где C — ортогональная матрица, такая, что B — диагональная матрица (см. параграф 11.9). Множество S в пространстве переменных x преобразуется в аналогичное множество S_1 в пространстве переменных y . Из приведенного выше доказательства следует, что функция

$$(29.5.5) \quad g_n(y_{11}, \dots, y_{kk}) = \begin{cases} C_{kn} B^{\frac{n-1}{2}} Y^{\frac{n-k-2}{2}} e^{-\sum_{i,j} b_{ij} y_{ij}} & \text{в } S_1 \\ 0 & \text{в } S_1^c \end{cases}$$

есть плотность вероятности в пространстве переменных y (заметим, что $b_{ij} = 0$ при $i \neq j$). Поскольку детерминант матрицы C равен ± 1 , имеем $A = B$ и $X = Y$, и далее непосредственной подстановкой можно убедиться, что $\sum_{i,j} a_{ij} x_{ij} = \sum_{i,j} b_{ij} y_{ij}$. Таким образом, если в распределении (29.5.5) произвести преобразование случайных величин, определяемое формулами (29.5.4), то, в соответствии с параграфом 22.2, мы получим преобразованное распределение с плотностью вероятности $f_n(x_{11}, \dots, x_{kk})$. Таким образом, f_n есть плотность вероятности, и наше утверждение доказано.

В частном случае $k=2$ мы имеем три переменных x_{11} , x_{22} и $x_{12}=x_{21}$. Множество S есть область, определяемая неравенствами $x_{11}>0$, $x_{22}>0$, $x_{12}^2 < x_{11}x_{22}$. В S имеем

$$(29.5.6) \quad f_n(x_{11}, x_{12}, x_{22}) = C_{2n} (\sigma_{11}\sigma_{22} - \sigma_{12}^2)^{\frac{n-1}{2}} (x_{11}x_{22} - x_{12}^2)^{\frac{n-4}{2}} e^{-\sigma_{11}x_{11} - \sigma_{22}x_{22} - 2\sigma_{12}x_{12}},$$

где (см. параграф (12.4.4))

$$C_{2n} = \frac{1}{\sqrt{\pi} \Gamma\left(\frac{n-1}{2}\right) \Gamma\left(\frac{n-2}{2}\right)} = \frac{2^{n-3}}{\pi \Gamma(n-2)}.$$

Вне S плотность вероятности равна нулю.

Рассмотрим также характеристическую функцию $\varphi_n(t_{11}, \dots, t_{kk})$, соответствующую плотности вероятности $f_n(x_{11}, \dots, x_{kk})$, определяемой формулой (29.5.1). Пусть $T=\{t_{ij}\}$ обозначает симметричную матрицу переменных t_{ij} , и положим

$$\epsilon_{ij} = \begin{cases} 1 & \text{при } i=j, \\ \frac{1}{2} & \text{при } i \neq j. \end{cases}$$

Поскольку в S^* $f_n=0$, характеристической функцией, соответствующей плотности вероятности f_n , является

$$\varphi_n(t_{11}, \dots, t_{nn}) = \int_S e^{\sum_{i,j} t_{ij} x_{ij}} f_n(x_{11}, \dots, x_{kk}) dx_{11} dx_{12} \dots dx_{kk}.$$

(Чтобы избежать недоразумений, мы пишем здесь i — обозначение единицы — жирным шрифтом, как уже было указано в параграфе 27.1.) При $t_{ii}=0$ этот интеграл равен 1, так что имеем

$$\int_S X^{\frac{n-k-2}{2}} e^{-\sum_{i,j} a_{ij} x_{ij}} dx_{11} dx_{12} \dots dx_{kk} = \frac{1}{C_{kn} A^{\frac{-1}{2}}}.$$

Заменив здесь a_{ij} на $a_{ij} - i\epsilon_{ij} t_{ij}$ и обозначая через A^* детерминант $|a_{ij} - i\epsilon_{ij} t_{ij}|$, получим окончательное выражение

$$(29.5.7) \quad \varphi_n(t_{11}, \dots, t_{kk}) = \left(\frac{A}{A^*} \right)^{\frac{n-1}{2}}$$

для характеристической функции, соответствующей распределению (29.5.1) *).

29.6. Выбор из двумерного нормального распределения. В фундаментальной работе 1915 года Р. Фишер [88] дал точные выражения для некоторых выборочных распределений, связанных с двумерной нормальной генеральной совокупностью. Мы докажем некоторые из результатов Фишера, используя метод характеристических функций, впервые примененный к таким задачам Романовским [208], [209]. При этом будет показано, что полученные распределения представляют собой частные случаи распределений, рассмотренных в предыдущем параграфе.

Рассмотрим собственное нормальное распределение двух величин (см. параграф 21.12). Без ограничения общности можно предполагать, что моменты первого порядка равны нулю, так что в обычных обозначениях плотность вероятности имеет вид

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_1^2}-\frac{2\rho xy}{\sigma_1\sigma_2}+\frac{y^2}{\sigma_2^2}\right)} = \\ = \frac{1}{2\pi\sqrt{M}} e^{-\frac{1}{2M}(\mu_{20}x^2+2\mu_{12}xy+\mu_{02}y^2)},$$

где $M = \mu_{20}\mu_{02} - \mu_{11}^2 = \sigma_1^2\sigma_2^2(1 - \rho^2)$ — детерминант матрицы вторых моментов $M = \begin{Bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{Bmatrix}$. По выборке в n наблюденных пар значений $(x_1, y_1), \dots, (x_n, y_n)$ вычислим моментные характеристики первого и второго порядка (см. (27.1.6))

$$(29.6.1) \quad \begin{aligned} x &= \frac{1}{n} \sum_i x_i, \quad \bar{y} = \frac{1}{n} \sum_i y_i, \\ m_{20} &= s_1^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2 = \frac{1}{n} \sum_i x_i^2 - \bar{x}^2, \\ m_{11} &= rs_1s_2 = \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n} \sum_i x_i y_i - \bar{x}\bar{y}, \\ m_{02} &= s_2^2 = \frac{1}{n} \sum_i (y_i - \bar{y})^2 = \frac{1}{n} \sum_i y_i^2 - \bar{y}^2. \end{aligned}$$

*) Ингам [130] показал прямым методом, что характеристическая функция (29.5.7) дает, в соответствии с формулой обращения (10.6.3), плотность вероятности (29.5.1).

Постараемся отыскать совместное распределение пяти случайных величин \bar{x} , \bar{y} , m_{20} , m_{11} и m_{02} . Характеристическая функция этого распределения есть функция от пяти аргументов t_1 , t_2 , t_{20} , t_{11} , t_{02} , именно

$$(29.6.2) \quad E(e^{t_1\bar{x}+t_2\bar{y}+t_{20}m_{20}+t_{11}m_{11}+t_{02}m_{02}}) = \\ = \frac{1}{(2\pi)^n M^{n/2}} \int e^{\theta} dx_1 \dots dx_n dy_1 \dots dy_n,$$

где

$$\Omega = i(t_1\bar{x} + \dots + t_{02}m_{02}) - \frac{1}{2M} \sum_1^n (\mu_{02}x_i^2 - 2\mu_{11}x_iy_i + \mu_{20}y_i^2),$$

а интегрирование распространяется на $2n$ -мерное пространство значений величин $x_1, \dots, x_n, y_1, \dots, y_n$.

Заменим x_1, \dots, x_n новыми переменными ξ_1, \dots, ξ_n с помощью ортогонального преобразования, при котором $\xi_1 = \sqrt{n}\bar{x}$, и применим преобразование с такой же матрицей к величинам y_1, \dots, y_n , которые при этом заменятся новыми величинами η_1, \dots, η_n , причем $\eta_1 = \sqrt{n}\bar{y}$. Тогда имеем

$$\sum_1^n x_i^2 = \sum_1^n \xi_i^2, \quad \sum_1^n x_i y_i = \sum_1^n \xi_i \eta_i, \quad \sum_1^n y_i^2 = \sum_1^n \eta_i^2, \\ nm_{20} = \sum_2^n \xi_i^2, \quad nm_{11} = \sum_2^n \xi_i \eta_i, \quad nm_{02} = \sum_2^n \eta_i^2,$$

и отсюда

$$\Omega = i \frac{t_1 \xi_1 + t_2 \eta_1}{\sqrt{n}} - \frac{1}{2M} (\mu_{02} \xi_1^2 - 2\mu_{11} \xi_1 \eta_1 + \mu_{20} \eta_1^2) - \\ - \frac{1}{n} \sum_2^n \left[\left(\frac{n}{2} \frac{\mu_{02}}{M} - it_{20} \right) \xi_i^2 + 2 \left(-\frac{n}{2} \frac{\mu_{11}}{M} - \frac{1}{2} it_{11} \right) \xi_i \eta_i + \right. \\ \left. + \left(\frac{n}{2} \frac{\mu_{20}}{M} - it_{02} \right) \eta_i^2 \right].$$

Подставляя это выражение для Ω в формулу (29.6.2), приведем $2n$ -кратный интеграл к произведению n двойных интегралов, каждый из которых может быть вычислен непосредственно с помощью формул (11.12.1) и (11.12.2). Тогда совместная характеристическая функция (29.6.2) примет вид

$$(29.6.3) \quad e^{-\frac{1}{2n} \left(\mu_{02} t_1^2 + 2\mu_{11} t_1 t_2 + \mu_{20} t_2^2 \right)} \cdot \left(\frac{A}{A^*} \right)^{\frac{n-1}{2}},$$

где

$$A = \begin{vmatrix} \frac{n\mu_{02}}{2M} & -\frac{n\mu_{11}}{2M} \\ -\frac{n\mu_{11}}{2M} & \frac{n\mu_{20}}{2M} \end{vmatrix} = \frac{n^2}{4M},$$

$$A^* = \begin{vmatrix} \frac{n\mu_{02}}{2M} - it_{20} & -\frac{n}{2}\frac{\mu_{11}}{M} - \frac{1}{2}it_{11} \\ -\frac{n}{2}\frac{\mu_{11}}{M} - \frac{1}{2}it_{11} & \frac{n\mu_{20}}{2M} - it_{02} \end{vmatrix}$$

Совместная характеристическая функция (29.6.3) есть произведение двух множителей, первый из которых содержит лишь переменные t_1 и t_2 , а второй — лишь t_{20} , t_{11} и t_{02} . Первый множитель является, согласно (21.12.2), характеристикой функцией некоторого нормального распределения с нулевым средним значением *) и матрицей вторых моментов $n^{-1}M$. С другой стороны, второй множитель есть частный случай характеристической функции (29.5.7). Действительно, если в предыдущем параграфе положить $k=2$ и

$$A = \begin{Bmatrix} \frac{n\mu_{02}}{2M} - \frac{n\mu_{11}}{2M} \\ -\frac{n\mu_{11}}{2M} \end{Bmatrix} = \frac{n}{2} M^{-1}, \quad T = \begin{Bmatrix} t_{20} & t_{11} \\ t_{11} & t_{02} \end{Bmatrix},$$

то характеристическая функция (29.5.7) превратится во второй сомножитель формулы (29.6.3).

Поэтому соответствующее распределение есть частный случай распределения (29.5.1) при $k=2$ (который уже был дан формулой (29.5.6)), с заменой величин x_{11} , x_{12} , x_{22} соответственно на t_{20} , t_{11} , t_{02} . Таким образом, в силу параграфа 22.4 мы получили следующую теорему:

Составные случайные величины (\bar{x}, \bar{y}) и (t_{20}, t_{11}, t_{02}) независимы. Совместное распределение величин \bar{x} и \bar{y} нормально и имеет те же моменты первого порядка, что и распределение генеральной совокупности, и матрицу вторых моментов $n^{-1}M$. Совместное распределение величин t_{20}, t_{11}, t_{02} имеет плотность ве-

*) Если, вообще, рассмотреть распределение генеральной совокупности с произвольными средними значениями, то, очевидно, мы получим здесь те же самые средние значения, что и у генеральной совокупности.

вероятности f_n , задаваемую формулой

$$(29.6.4) \quad f_n(m_{20}, m_{11}, m_{02}) = \frac{n^{n-1}}{4\pi\Gamma(n-2)} \frac{\left(\frac{m_{20}m_{02} - m_{11}^2}{M}\right)^{\frac{n-4}{2}}}{\frac{n-1}{2}} e^{-\frac{n}{2M}(\nu_{20}m_{20} - 2\nu_{11}m_{11} + \nu_{02}m_{02})}$$

в области $m_{20} > 0$, $m_{02} > 0$, $m_{11}^2 < m_{20}m_{02}$, причем $f_n = 0$ вне этой области.

Средние значения и матрицу вторых моментов для пяти выборочных моментов можно вычислить по характеристической функции (29.6.3). При этом получим, например,

$$E(m_{20}) = \frac{n-1}{n} \mu_{20}, \quad E(m_{11}) = \frac{n-1}{n} \mu_{11}, \quad E(m_{02}) = \frac{n-1}{n} \mu_{02},$$

в соответствии с параграфами 27.4 и 27.8.

29.7. Коэффициент корреляции. Введем новую переменную r в совместное распределение (29.6.4) величин m_{20} , m_{11} и m_{02} , полагая $m_{11} = r \sqrt{m_{20}m_{02}}$, так что r — коэффициент корреляции выборки. Тогда в силу формулы (22.2.3), получим следующее выражение для совместной плотности вероятности величин m_{20} , m_{02} и r :

$$\sqrt{m_{20}m_{02}} f_n(m_{20}, r\sqrt{m_{20}m_{02}}, m_{02}) = \frac{n^{n-1}}{4\pi\Gamma(n-2) M^{\frac{n-1}{2}}} \frac{\frac{n-3}{2} \frac{n-3}{2}}{m_{20}^{\frac{n-3}{2}} m_{02}^{\frac{n-3}{2}}} (1-r^2)^{\frac{n-4}{2}} e^{-\frac{n}{2M}(\nu_{20}m_{20} - 2\nu_{11}r\sqrt{m_{20}m_{02}} + \nu_{02}m_{02})},$$

где $m_{20} > 0$, $m_{02} > 0$, $r^2 < 1$. Частная плотность вероятности для r получится интегрированием совместной плотности вероятности по m_{20} и m_{02} от 0 до $+\infty$. Интегрирование можно произвести точно, если разложить в степенной ряд множитель $e^{\frac{n}{M} \nu_{11}r\sqrt{m_{20}m_{02}}}$, и мы получим при этом плотность вероятности для выборочного коэффициента корреляции r :

$$(29.7.1) \quad f_n(r) = \frac{2^{\frac{n-3}{2}}}{\pi(n-3)!} (1-\rho^2)^{\frac{n-1}{2}} (1-r^2)^{\frac{n-4}{2}} \sum_{v=0}^{\infty} \Gamma^2\left(\frac{n+v-1}{2}\right) \frac{(2\rho r)^v}{v!}$$

при $-1 < r < 1$. Входящий в это выражение степенной ряд можно преобразовать различными способами. Например, с помощью простых

вычислений можно получить:

$$\int_0^1 \frac{x^{n-2}}{(1-\rho x)^{n-1} \sqrt{1-x^2}} dx = \frac{2^{n-3}}{(n-2)!} \sum_{v=0}^{\infty} \Gamma^2 \left(\frac{n+v-1}{2} \right) \frac{(2\rho)^v}{v!},$$

и отсюда получается следующее выражение для плотности вероятности величины r :

$$(29.7.2) \quad f_n(r) = \frac{n-2}{\pi} (1-\rho^2)^{\frac{n-1}{2}} (1-r^2)^{\frac{n-4}{2}} \int_0^1 \frac{x^{n-2}}{(1-\rho x)^{n-1} \sqrt{1-x^2}} dx.$$

Распределение для r было получено Р. Фишером [88]. Укажем замечательное свойство этого распределения: оно зависит только от объема выборки n и от коэффициента корреляции совокупности ρ .

При $n=2$ плотность вероятности $f_n(r)$ обращается в нуль, в соответствии с тем фактом, что коэффициент корреляции, вычисленный по выборке, состоящей только из двух значений, необходимо равен ± 1 , так что в этом случае распределение принадлежит к дискретному типу. При $n=3$ плотность вероятности U-образна, с бесконечными ординатами в точках $r=\pm 1$. При $n=4$ получается прямоугольное распределение, если $\rho=0$, и J-образное распределение в случае $\rho \neq 0$. При $n > 4$ распределение унимодально, с модой в точке $r=0$, если $\rho=0$, и около точки $r=\rho$, если $\rho \neq 0$. Некоторые примеры изображены на фиг. 29 а и 29 б.

Распределение для r было детально изучено различными авторами (см., например, Сопер и др. [216] и Романовский [208]), подробные таблицы опубликованы Дэвидом [261]. Известны различные точные и приближенные формулы для характеристик этого распределения. Любой момент для r можно, конечно, непосредственно вычислить, исходя из формулы (29.7.1), но мы ограничимся здесь асимптотическими формулами для $E(r)$ и $D^2(r)$ при больших n , которые уже даны в (27.8.1) и (27.8.2).

Для практических целей часто предпочтительно употребляется преобразование

$$(29.7.3) \quad z = \frac{1}{2} \log \frac{1+r}{1-r}, \quad \zeta = \frac{1}{2} \log \frac{1+\rho}{1-\rho},$$

введенное Р. Фишером [13], [90]. Фишер показал, что величина z уже для небольших значений n распределена приблизительно нормально,

со средним значением и дисперсией, заданными приближенными выражениями

$$(29.7.4) \quad E(z) = \zeta + \frac{\rho}{2(n-1)}, \\ D^2(z) = \frac{1}{n-3}.$$

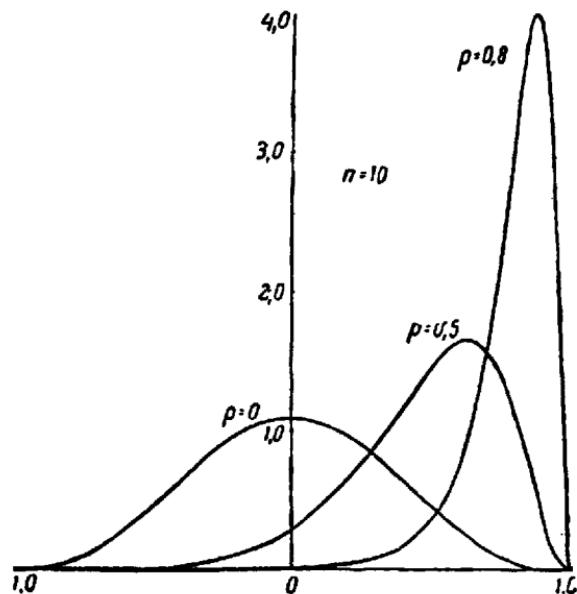
Таким образом, в первом приближении форма z -распределения не зависит от значения ρ , в то время как распределение для r значительно изменяет форму при изменении ρ . В этом отношении поучительно сравнить иллюстрации для распределений величин r и z , приведенные на фиг. 29 и фиг. 30. (См. ниже параграф 31.3, пример 6.)

В частном случае $\rho=0$ плотность вероятности (29.7.1), в силу формулы (12.4.4), сводится к

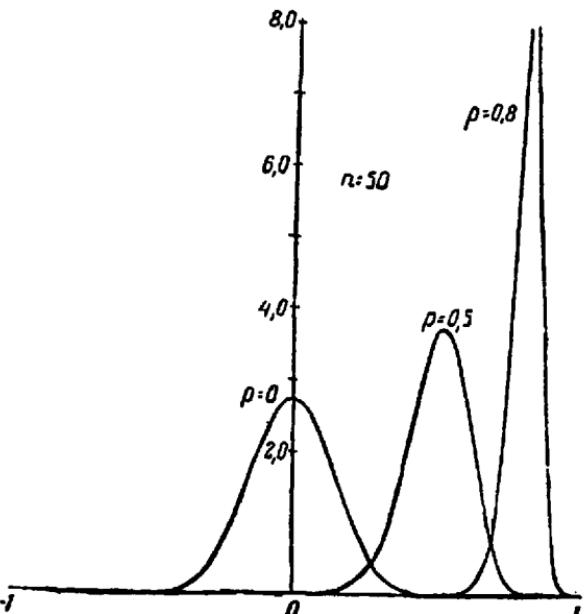
$$(29.7.5) \quad f_n(r) =$$

$$= \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \frac{(1-r^2)^{\frac{n-4}{2}}}{\sqrt{\pi}}$$

как это предполагал

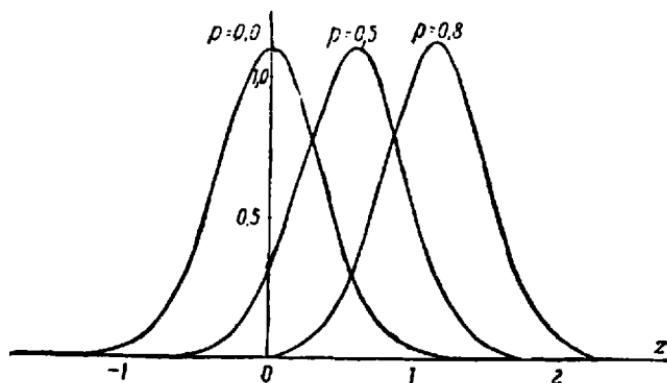


Фиг. 29а. Графики плотности вероятности для коэффициента корреляции r в выборках из нормальной совокупности. $n = 10$

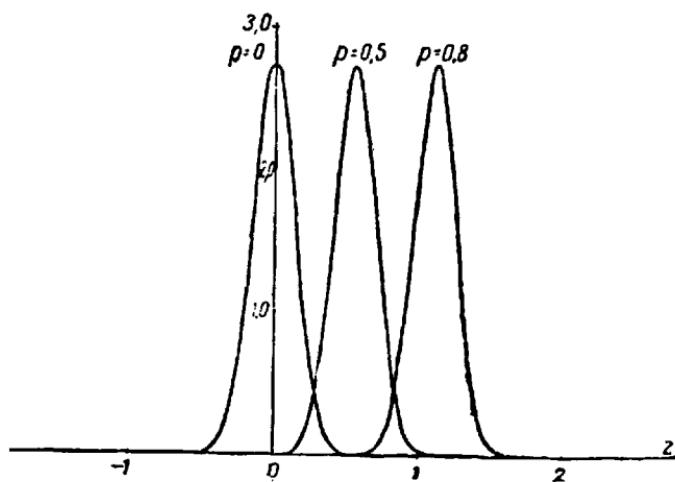


Фиг. 29б. Графики плотности вероятности для коэффициента корреляции r в выборках из нормальной совокупности. $n = 50$

Стьюарт [222] еще в 1908 г. Мы уже встречались с этой плотностью вероятности в связи с формулами (18.2.7) и (29.4.4). Согласно параграфу 18.2, преобразованная величина $t = \sqrt{n-2} \frac{r}{\sqrt{1-r^2}}$



Фиг. 30а. Графики плотности вероятности для $z = \frac{1}{2} \log \frac{1+r}{1-r}$ в выборках из нормальной совокупности. $n = 10$



Фиг. 30б. Графики плотности вероятности для $z = \frac{1}{2} \log \frac{1+r}{1-r}$ в выборках из нормальной совокупности. $n = 50$

в этом случае имеет распределение Стьюдента с $n-2$ степенями свободы. Если t_p обозначает p -процентное значение t при $n-2$ степенях свободы (см. параграф 18.2), то мы получаем с вероятностью

p^0 значение t такое, что $|t| > t_p$; последнее неравенство эквивалентно (см. параграф 31.3, пример 7) следующему:

$$(29.7.6) \quad |r| > \frac{t_p}{\sqrt{\frac{t_p^2}{t_p^2 + n - 2}}}.$$

29.8. Коэффициенты регрессии. Коэффициенты регрессии для генеральной совокупности

$$\beta_{21} = \frac{\mu_{11}}{\mu_{20}} = \frac{\rho \beta_2}{c_1}, \quad \beta_{12} = \frac{\mu_{11}}{\mu_{02}} = \frac{\rho \beta_1}{c_2}$$

были определены в параграфе 21.6. В соответствии с общими правилами параграфа 27.1, соответствующие коэффициенты регрессии для выборки будут обозначаться через

$$(29.8.1) \quad b_{21} = \frac{m_{11}}{m_{20}} = \frac{rs_2}{s_1}, \quad b_{12} = \frac{m_{11}}{m_{02}} = \frac{rs_1}{s_2}.$$

Достаточно рассмотреть выборочное распределение одной из этих величин, например, b_{21} . Тогда распределение для b_{12} получится перестановкой индексов.

Заменим m_{11} в совместном распределении величин m_{20} , m_{11} , m_{02} (формула (29.6.4)) новой величиной b_{21} , положив $m_{11} = m_{20}b_{21}$. Тогда можно непосредственно осуществить интегрирование сначала по всем значениям m_{02} , для которых $m_{02} > m_{20}b_{21}^2$, и затем по всем положительным значениям m_{20} , и мы получим следующее выражение для плотности вероятности выборочного коэффициента регрессии b_{21} :

$$(29.8.2) \quad \frac{1}{V^\pi} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{M^{\frac{n-1}{2}}}{\mu_{20}^{\frac{n-2}{2}} (\mu_{20}b_{21}^2 - 2\mu_{11}b_{21} + \mu_{02})^{n/2}}.$$

Это распределение впервые было получено К. Пирсоном [185] и Романовским [210]. Если ввести здесь новую величину

$$(29.8.3) \quad t = \frac{\mu_{20} V \sqrt{n-1}}{V M} (b_{21} - \beta_{21}) = \frac{c_1 V \sqrt{n-1}}{c_2 V \sqrt{1-\rho^2}} (b_{21} - \beta_{21}),$$

где $M = \mu_{20}\mu_{02} - \mu_{11}^2$, то обнаружится, что t имеет распределение Стьюдента с $n-1$ степенями свободы.

Мы видели, что распределение для r содержало лишь параметр совокупности, прямо соответствующий самой величине. Распределение

же величины b_{21} не обладает этим привлекательным свойством. Действительно, плотность вероятности (29.8.2) содержит все три момента μ_{20} , μ_{11} и μ_{02} , и если мы хотим вычислить величину t по формуле (29.8.3), чтобы проверить какое-нибудь гипотетическое значение величины β_{21} , то мы должны ввести гипотетические значения всех трех моментов. Чтобы избавиться от этого неудобства, рассмотрим величину

$$(29.8.4) \quad t' = \frac{s_1 \sqrt{n-2}}{s_2 \sqrt{1-r^2}} (b_{21} - \beta_{21}),$$

где характеристики совокупности s_1 , s_2 и r , входящие в (29.8.3), заменены соответствующими выборочными характеристиками s_1 , s_2 и r , а множитель $\sqrt{n-1}$ заменен на $\sqrt{n-2}$. Если эту величину t' ввести вместо m_{02} в совместное распределение (29.6.4), то можно непосредственно осуществить интегрирование по переменным m_{11} и m_{20} , и мы получим интересный результат: t' имеет распределение Стьюдента с $n-2$ степенями свободы (Бартлет [54]). Замена характеристик совокупности выборочными характеристиками привела, таким образом, к потере одной степени свободы. Если требуется проверить гипотетическое значение величины β_{21} , то теперь можно вычислить t' непосредственно по фактической выборке, и мы получим, таким образом, критерий значимости для отклонения наблюденного значения b_{21} от гипотетического β_{21} (см. параграф 31.3, пример 6).

29.9. Выбор из k -мерного нормального распределения. Результаты параграфа 29.6 можно обобщить на случай k -мерной нормальной генеральной совокупности. Рассмотрим собственное k -мерное нормальное распределение (см. параграф 24.2). Без ограничения общности можно считать моменты первого порядка равными нулю, так что плотность вероятности имеет вид (см. параграф 24.2.1)

$$(29.9.1) \quad \frac{1}{(2\pi)^{k/2} \sqrt{\Lambda}} e^{-\frac{1}{2\Lambda} \sum_{i,j} \Lambda_{ij} x_i x_j} = \frac{1}{(2\pi)^{k/2} s_1 \dots s_k \sqrt{P}} e^{-\frac{1}{2P} \sum_{i,j} P_{ij} \frac{x_i^2}{s_i^2} \frac{x_j^2}{s_j^2}},$$

где $\Lambda = \{\lambda_{ij}\}$ — матрица вторых моментов, а $P = \{p_{ij}\}$ — корреляционная матрица распределения (см. параграф 22.3). Λ и P — соответствующие детерминанты. Повсюду в этом параграфе индексы i и j будут пробегать значения от 1 до k .

Предположим теперь, что мы извлекли выборку в n наблюденных точек из этого распределения. Обозначим y -ю точку выборки через $(x_{1y}, x_{2y}, \dots, x_{ky})$, где $y = 1, 2, \dots, n$, и предположим, что $n > k$. Затем вычислим выборочные моменты первого и второго порядка. В соответствии с общими правилами параграфа 27.1 и обозначениями для соответствующих моментов совокупности, введенными в параграфе 22.3, будем обозначать выборочные моменты следующим образом:

$$(29.9.2) \quad \bar{x}_i = \frac{1}{n} \sum_{v=1}^n x_{iv}, \quad l_{ii} = s_i^2 = \frac{1}{n} \sum_{v=1}^n (x_{iv} - \bar{x}_i)^2,$$

$$l_{ij} = r_{ij} s_i s_j = \frac{1}{n} \sum_{v=1}^n (x_{iv} - \bar{x}_i)(x_{jv} - \bar{x}_j).$$

Существует k выборочных средних значений \bar{x}_i и k дисперсий $l_{ii} = s_i^2$. Далее, поскольку $l_{ji} = l_{ij}$, существует $\frac{k(k-1)}{2}$ различных смешанных вторых моментов l_{ij} с $i \neq j$. Общее число различных величин l_{ij} равно, таким образом, $\frac{k(k+1)}{2}$.

Матрицы $L = \{l_{ij}\}$ и $R = \{r_{ij}\}$ суть матрица вторых моментов и корреляционная матрица выборки; соответствующие детерминанты обозначим через $L = |l_{ij}|$ и $R = |r_{ij}|$.

Совместное распределение всех величин \bar{x}_i и l_{ij} можно получить таким же образом, как и соответствующее распределение в параграфе 29.6. Непосредственно обобщая формулы (29.6.2), получим для совместной характеристической функции величин \bar{x}_i и l_{ij} выражение

$$(29.9.3) \quad \frac{1}{(2\pi)^{\frac{kn}{2}} \Lambda^{\frac{n}{2}}} \int e^{iz} dx_1 \dots dx_k,$$

$$\Omega = i \sum_i t_i \bar{x}_i + i \sum_{i,j} \epsilon_{ij} t_i l_{ij} - \frac{1}{2\Lambda} \sum_{v=1}^n \sum_{i,j} \Lambda_{ij} x_{iv} x_{jv},$$

где интегрирование распространяется на kn -мерное пространство величин x_i , ($i = 1, \dots, k$; $v = 1, \dots, n$), и, так же, как и в параграфе 29.5, мы обозначили $\epsilon_{ij} = 1$ при $i = j$ и $\epsilon_{ij} = \frac{1}{2}$ при $i \neq j$.

При любом i заменим n величин x_{i1}, \dots, x_{in} системой n новых величин $\xi_{i1}, \dots, \xi_{in}$ с помощью ортогонального преобразования,

при котором $\xi_{ii} = \sqrt{n} \bar{x}_i$, используя одну и ту же матрицу преобразования при всех значениях i . Тогда для всех i и j будем иметь

$$\sum_{v=1}^n x_{iv} x_{jv} = \sum_{v=1}^n \xi_{iv} \xi_{jv}, \quad n l_{ij} = \sum_{v=1}^n x_{iv} x_{jv} - n \bar{x}_i \bar{x}_j = \sum_{v=1}^n \xi_{iv} \xi_{jv},$$

и отсюда

$$(29.9.4) \quad \Omega = \frac{i}{\sqrt{n}} \sum_i t_i \xi_{ii} - \frac{1}{2\Lambda} \sum_{i,j} \Lambda_{ij} \xi_{ii} \xi_{jj} - \\ - \frac{1}{n} \sum_{v=2}^n \sum_{i,j} \left(\frac{n \Lambda_{ij}}{2 \lambda} - i \epsilon_{ij} t_{ij} \right) \xi_{iv} \xi_{jv}.$$

Вводя это выражение для Ω в (29.9.3), можно преобразовать интеграл таким же образом, как соответствующий интеграл в (29.6.2), и совместная характеристическая функция (29.9.3) примет вид

$$(29.9.5) \quad e^{-\frac{1}{2n} \sum_{i,j} \lambda_{ij} t_i t_j} \cdot \left(\frac{A}{A^*} \right)^{\frac{n-1}{2}},$$

где A и A^* обозначают детерминанты матриц

$$A = \left\{ \frac{n \Lambda_{ij}}{2 \Lambda} \right\} = \frac{n}{2} \Lambda^{-1},$$

и

$$A^* = \left\{ \frac{n \Lambda_{ij}}{2 \Lambda} - i \epsilon_{ij} t_{ij} \right\}.$$

Таким образом, в частности, $A = \left(\frac{n}{2}\right)^k \Lambda^{-1}$. Так же как и в параграфе 29.6, совместная характеристическая функция есть произведение двух множителей, первый из которых является характеристической функцией некоторого нормального распределения, а второй имеет вид (29.5.7), так что соответствует распределению вида (29.5.1) с $A = \frac{1}{2} n \Lambda^{-1}$ и с матрицей переменных $X = L = \{l_{ij}\}$. Обозначая через S множество всех точек $\frac{k(k+1)}{2}$ -мерного пространства переменных t_{ij} , для которых симметричная матрица L положительно определена, получим следующее собщение теоремы параграфа 29.6:

Составные случайные величины $(\bar{x}_1, \dots, \bar{x}_k)$ и $(l_{11}, l_{12}, \dots, l_{kk})$ независимы, совместное распределение величин $\bar{x}_1, \dots, \bar{x}_k$ нормально.

имеет те же моменты первого порядка, что и распределение генеральной совокупности, и матрицу вторых моментов $n^{-1} \Lambda$. Совместное распределение $\frac{k(k+1)}{2}$ различных величин l_{ij} имеет плотность вероятности

$$(29.9.6) \quad f_n(l_{11}, l_{12}, \dots, l_{kk}) = C_{kn} \left(\frac{n^k}{2^k \Lambda} \right)^{\frac{n-1}{2}} L^{\frac{n-k-2}{2}} e^{-\frac{n}{2\Lambda} \sum_{i,j} \Lambda_{ij} l_{ij}}$$

для каждой точки множества S , и $f_n = 0$ в дополнительном множестве S^* . Постоянны C_{kn} даются формулой (29.5.2).

Эта теорема была впервые доказана Висхартом [240] с помощью геометрических методов, предложенных Р. Фишером, и затем Висхартом и Бартлеттом [241] методом характеристических функций. Мы рекомендуем читателю также статью Симонсена [213a].

29.10. Обобщенная дисперсия. Детерминант $L = |l_{ij}|$ представляет собой обобщенную дисперсию выборки (см. параграф 22.7). Следуя Уилксу [232], покажем, как можно определить моменты величины L . Относительно точного распределения для L мы отсылаем читателя к Кульбаку [143].

Интеграл от плотности вероятности f_n в (29.9.6) по множеству S равен, очевидно, единице. Множество S инвариантно относительно любого преобразования вида $w_{ij} = a l_{ij}$, где $a > 0$. Полагая $a = n$ и обозначив $W = |w_{ij}|$, получим

$$\int_S W^{\frac{n-k-2}{2}} e^{-\frac{1}{2\Lambda} \sum_{i,j} \Lambda_{ij} w_{ij}} dw_{11} dw_{12} \dots dw_{kk} = \frac{(2^k \Lambda)^{\frac{n-1}{2}}}{C_{kn}} .$$

Поскольку это равенство справедливо при всех значениях $n > k$, можно заменить n на $n + 2v$, и, возвращаясь к переменным l_{ij} , мы получим

$$\int_S L^{\frac{n-k-2}{2} + v} e^{-\frac{n}{2\Lambda} \sum_{i,j} \Lambda_{ij} l_{ij}} dl_{11} dl_{12} \dots dl_{kk} = \left(\frac{2^k \Lambda}{n^k} \right)^{\frac{n-1}{2} + v} \cdot \frac{1}{C_{k,n+2v}} .$$

После умножения на $C_{kn} \left(\frac{n^k}{2^k \Lambda} \right)^{\frac{n-1}{2}}$, вспоминая формулы (29.9.6) и

(29.5.2), получим

$$E(L^v) = \left(\frac{2^k \Lambda}{n^k}\right)^v \frac{C_{kn}}{C_{k,n+2}} = \left(\frac{2^k \Lambda}{n^k}\right)^v \prod_{i=1}^k \frac{\Gamma\left(\frac{n-i}{2} + v\right)}{\Gamma\left(\frac{n-i}{2}\right)}$$

при $n + 2^v > k$, т. е. для всех $v > -\frac{n-k}{2}$. В частности, имеем

$$E(L) = \frac{(n-1)(n-2)\dots(n-k)}{n^k} \Lambda,$$

$$D^2(L) = \frac{k(2n+1-k)}{(n-k)(n-k+1)} \cdot \frac{(n-1)^2\dots(n-k)^2}{n^{2k}} \Lambda^2.$$

Для одномерного распределения ($k=1$) имеем $L = l_{11} = m_2$ и $\Lambda = \sigma^2$ и тогда приведенное выше выражение для $E(L^v)$ сводится к формуле (29.3.2).

29.11. Обобщенное Стьюентово отношение. Рассмотрим теперь выборку из k -мерного нормального распределения с произвольными и средними значениями m_1, m_2, \dots, m_k и обозначим через l'_{ij} смешанные моменты относительного среднего значения совокупности:

$$(29.11.1) \quad l'_{ij} = \frac{1}{n} \sum_{v=1}^n (x_{iv} - m_i)(x_{jv} - m_j) = l_{ij} + (\bar{x}_i - m_i)(\bar{x}_j - m_j),$$

где \bar{x}_i и l_{ij} даются формулами (29.9.2).

Существует $\frac{k(k+1)}{2}$ различных величин l'_{ij} . Если обозначить $\xi_{iv} = x_{iv} - m_i$, то совместная характеристическая функция величин l'_{ij} примет вид

$$\frac{1}{(2\pi)^{\frac{kn}{2}} \Lambda^{\frac{n}{2}}} \int e^{2i} d\xi_{11} \dots d\xi_{kn},$$

где

$$\begin{aligned} \Omega' &= i \sum_{i,j} \epsilon_{ij} t_{ij} l'_{ij} - \frac{1}{2\Lambda} \sum_{v=1}^n \sum_{i,j} \Lambda_{ij} \xi_{iv} \xi_{jv} = \\ &= -\frac{1}{n} \sum_{v=1}^n \sum_{i,j} \left(\frac{n \Lambda_{ij}}{2\Lambda} - i \epsilon_{ij} t_{ij} \right) \xi_{iv} \xi_{jv}. \end{aligned}$$

Сравнивая это с (29.9.3)–(29.9.5), мы обнаружим, что характеристической функцией величин l'_{ij} является $\left(\frac{A}{A^*}\right)^{n/2}$, где A и A^* обозна-

чают те же детерминанты, что и в (29.9.5). Отсюда следует, что совместную характеристическую функцию величин I'_{ij} можно получить, если заменить в (29.9.6) n на $n+1$ и отбросить два множителя $\frac{n^k}{2^k \Lambda}$ и $\frac{n}{2\Lambda}$, происходящие из матрицы A .

Полагая $L' = |I'_{ij}|$, получим с помощью такого же преобразования, как и в предыдущем параграфе,

$$E(L'^{\mu}) = \left(\frac{2^k \Lambda}{n^k}\right)^{\mu} \prod_{i=1}^k \frac{\Gamma\left(\frac{n+1-i}{2} + \mu\right)}{\Gamma\left(\frac{n+1-i}{2}\right)}$$

для любого $\mu > -\frac{n+1-k}{2}$. С другой стороны, в соответствии с (29.11.1), L' есть функция от случайных величин I_{ij} и $\xi_i = \bar{x}_i - m_i$, и совместной плотностью вероятности всех этих величин является, согласно теореме параграфа 29.9, функция

$$g(\xi, I) = \frac{n^{k/2}}{(2\pi)^{k/2} V \Lambda} e^{-\frac{n}{2\Lambda} \sum_{i,j} I_{ij} \xi_i \xi_j} f_n(I),$$

где $f_n(I) = f_n(I_{11}, I_{12}, \dots, I_{kk})$ дается формулой (29.9.6). Таким образом, можно записать

$$E(L'^{\mu}) = \int L'^{\mu} g(\xi, I) d\xi dI,$$

где интегрирование производится по всему множеству S (определенному в параграфе 29.6) относительно переменных I_{ij} и от $-\infty$ до $+\infty$ по каждому ξ_i . Теперь можно применить здесь снова преобразование предыдущего параграфа, положив $w_{ij} = nI_{ij}$ и $\eta_i = \sqrt{n} \xi_i$ и затем заменив n на $n+2$. Приравнивая два выражения для $E(L'^{\mu})$, получим для любых $\nu > 0$ и $\mu > -\nu - \frac{n+1-k}{2}$

$$E(L^{\nu} L'^{\mu}) = \left(\frac{2^k \Lambda}{n^k}\right)^{\mu+\nu} \cdot \prod_{i=1}^k \frac{\Gamma\left(\frac{n-i}{2} + \nu\right)}{\Gamma\left(\frac{n-i}{2}\right)} \cdot \frac{\Gamma\left(\frac{n+1-i}{2} + \mu + \nu\right)}{\Gamma\left(\frac{n+1-i}{2} + \nu\right)}.$$

Полагая здесь $\mu = -\nu$, получим

$$E\left(\frac{L}{L'}\right)^{\nu} = \frac{\Gamma\left(\frac{n-k}{2} + \nu\right)}{\Gamma\left(\frac{n-k}{2}\right)} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2} + \nu\right)}.$$

Таким образом, согласно формуле (18.4.4), величина $\frac{L}{L'}$ имеет те же моменты, что и бэта-распределение с плотностью вероятности

$$(29.11.2) \quad \beta\left(x; \frac{n-k}{2}, \frac{k}{2}\right) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right)\Gamma\left(\frac{k}{2}\right)} x^{\frac{n-k}{2}-1} (1-x)^{\frac{k}{2}-1}, \quad (0 < x < 1)$$

Поскольку распределение с конечной широтой однозначно определяется своими моментами (см. параграф 15.4), то величина $\frac{L}{L'}$ имеет плотность вероятности (29.11.2). С другой стороны, из формулы (29.11.1) получаем

$$L' = L + \sum_{i,j} L_{ij} (\bar{x}_i - m_i)(\bar{x}_j - m_j),$$

$$\frac{L}{L'} = \frac{1}{1 + \sum_{i,j} \frac{L_{ij}}{L} (\bar{x}_i - m_i)(\bar{x}_j - m_j)},$$

где L_{ij} — алгебраическое дополнение элемента l_{ij} в детерминанте L . Квадратичная форма в знаменателе неотрицательна, так как L — матрица вторых моментов некоторого распределения, именно, распределения выборки. Если ввести теперь новую величину T , полагая

$$(29.11.3) \quad T^2 = (n-1) \sum_{i,j} \frac{L_{ij}}{L} (x_i - m_i)(\bar{x}_j - m_j),$$

где $T \geq 0$, то получим

$$\frac{L}{L'} = \frac{1}{1 + \frac{T^2}{n-1}},$$

и с помощью простого преобразования формулы (29.11.2) получим плотность вероятности для величины T в виде

$$(29.11.4) \quad \frac{2\Gamma\left(\frac{n}{2}\right)}{(n-1)^{k/2}\Gamma\left(\frac{n-k}{2}\right)\Gamma\left(\frac{k}{2}\right)} \cdot \frac{x^{k-1}}{\left(1 + \frac{x^2}{n-1}\right)^{n/2}}, \quad (x > 0).$$

При $k=1$ это выражение превращается в положительную половину обычного стьюдентова распределения (18.2.4) с $n-1$ степенями

свободы. Распределение величины T было получено Хотеллингом [126], а приведенное выше доказательство принадлежит Уилксу [232].

Таким же образом, как обычное стьюдентово отношение t можно использовать для критерия значимости отклонения наблюденного среднего значения \bar{x} от некоторого гипотетического значения m , обобщенное стьюдентово отношение T дает критерий для совместного отклонения выборочных средних значений $\bar{x}_1, \dots, \bar{x}_k$ от некоторой гипотетической системы значений m_1, \dots, m_k .

В параграфе 29.4 мы показали, как можно изменить Стьюдентово отношение, чтобы получить критерий для разности между двумя средними значениями. Аналогичное видоизменение можно применить и к обобщенному отношению T .

Предположим, что заданы две выборки соответственно в n_1 и n_2 индивидуумов из одной и той же k -мерной нормальной совокупности, и пусть \bar{x}_{1i}, l_{1ij} и \bar{x}_{2i}, l_{2ij} — средние значения, дисперсии и смешанные вторые моменты двух выборок. Пусть, далее H обозначает матрицу

$$H = \{n_1 l_{1ij} + n_2 l_{2ij}\} = n_1 L_1 + n_2 L_2,$$

а H и H_{ij} — соответствующие детерминант и алгебраические дополнения. Полагая

$$(29.11.5) \quad U^2 = \frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2} \sum_{i,j} \frac{H_{ij}}{H} (\bar{x}_{1i} - \bar{x}_{2i})(\bar{x}_{1j} - \bar{x}_{2j}),$$

где $U = 0$, мы сможем показать теми же методами, как и выше, что U имеет плотность вероятности (29.11.4), в которой n заменено на $n_1 + n_2 - 1$. Выражение (29.11.5) не содержит параметров генеральной совокупности, так что U можно непосредственно вычислить по выборке и использовать как критерий совместного расхождения между двумя системами выборочных средних \bar{x}_{1i} и \bar{x}_{2j} . При $k = 1$, как легко видеть, U^2 обращается в величину u^2 , определяемую формулой (29.4.2).

29.12. Коэффициенты регрессии. В случае двумерного распределения мы видели, что величина (29.8.4), связанная простым образом с выборочным коэффициентом регрессии, имеет t -распределение с $n - 2$ степенями свободы. Этот результат был обобщен Бартлеттом [54] на случай распределений любого числа измерений.

Заменяя в (28.2.3) и (28.4.5) характеристики совокупности выборочными характеристиками, получим для коэффициента регрессии $b_{12 \cdot 34 \dots k}$ выражение

$$b_{12 \cdot 34 \dots k} = -\frac{s_1}{s_2} \cdot \frac{R_{12}}{R_{11}} = r_{12 \cdot 34 \dots k} \frac{s_{1 \cdot 34 \dots k}}{s_{2 \cdot 34 \dots k}},$$

где остаточные дисперсии s могут быть вычислены по выборочным коэффициентам r , как показывает первое соотношение (28.4.5).

Если $\beta_{12 \cdot 34 \dots k}$ обозначает значение коэффициента регрессии для совокупности, то величина

$$(29.12.1) \quad t = \sqrt{n-k} \frac{s_{1 \cdot 34 \dots k}}{s_{1 \cdot 34 \dots k}} (b_{12 \cdot 34 \dots k} - \beta_{12 \cdot 34 \dots k})$$

имеет распределение Стьюдента с $n-k$ степенями свободы. Таким же образом, как и в случае (29.8.4), можно получить критерий значимости для отклонения наблюденного значения коэффициента регрессии b от какого-либо гипотетического значения β (см. параграф 31.3, пример 7).

29.13. Частные и сводные коэффициенты корреляции. Теперь мы приступим к некоторым дальнейшим применением распределения (29.9.6), ограничиваясь частным случаем, когда k величин в нормальной генеральной совокупности независимы. В этом случае $\lambda_{ij}, \rho_{ij}, \Lambda_{ij}$ обращаются при $i \neq j$ в нуль, так что матрица вторых моментов Λ — диагональная матрица, а корреляционная матрица R — единичная матрица (см. параграф 22.3).

В совместном распределении величин I_{ij} (29.9.6) заменим I_{ij} с $i \neq j$ выборочными коэффициентами корреляции r_{ij} с помощью подстановки $I_{ij} = r_{ij} \sqrt{I_{ii} I_{jj}}$. Тогда получим $L = I_{11} I_{22} \dots I_{kk} R$, где $R = |r_{ij}|$ — детерминант корреляционной матрицы выборки R . Якобианом преобразования (см. аналогичное преобразование (29.5.3)) является $(I_{11} \dots I_{kk})^{\frac{k-1}{2}}$, и совместная плотность вероятности величин I_{ii} и r_{ij} , согласно формуле (22.2.3), в рассматриваемом здесь частном случае принимает вид

$$C_{kn} \left(\frac{n^k}{2\lambda_{11}\lambda_{22} \dots \lambda_{kk}} \right)^{\frac{n-1}{2}} (I_{11} I_{22} \dots I_{kk})^{\frac{n-3}{2}} R^{\frac{n-k-2}{2}} e^{-\frac{n}{2} \sum_i \frac{I_{ii}}{\lambda_{ii}}}$$

при $I_{ii} > 0$ и для всех значений r_{ij} , при которых матрица R является положительно определенной. Для всех других значений переменных плотность вероятности равна нулю.

Теперь можно непосредственно интегрировать по каждому I_{ii} от 0 до ∞ . Вспоминая значение (29.5.2) постоянной C_{kn} , мы получим совместную плотность вероятности выборочных коэффициентов корреляции r_{ij} :

$$(29.13.1) \quad \frac{\left(\Gamma \left(\frac{n-1}{2} \right) \right)^{k-1}}{\pi^{\frac{k(k-1)}{4}} \Gamma \left(\frac{n-2}{2} \right) \dots \Gamma \left(\frac{n-k}{2} \right)} R^{\frac{n-k-2}{2}},$$

В соответствии с терминологией Фриша [113], детерминант R есть квадрат *коэффициента разброса* выборки (см. параграф 22.7). Моменты величины R можно определить методом параграфа 29.10. Обозначая через B_{kn} множитель при $R^{\frac{n-k-2}{2}}$ в формуле (29.13.1), получим, например,

$$(29.13.2) \quad E(R) = \frac{B_{kn}}{B_{k,n+1}} = \frac{(n-2)(n-3)\dots(n-k)}{(n-1)^{k-1}},$$

$$D^2(R) = \frac{k(k-1)}{n^3} + O\left(\frac{1}{n^3}\right).$$

Частный коэффициент корреляции между выборочными значениями величин x_1 и x_2 получается, после исключения остающихся переменных x_3, x_4, \dots, x_k , в силу формулы (23.4.2), в виде

$$(29.13.3) \quad r_{12 \cdot 34 \dots k} = -\frac{R_{12}}{\sqrt{R_{11} R_{22}}}.$$

где R_{ij} — алгебраические дополнения в R . В рассматриваемом здесь частном случае некоррелированной генеральной совокупности соответствующее значение характеристики совокупности $r_{12 \cdot 34 \dots k}$ равно, конечно, нулю.

Чтобы отыскать распределение величины $r_{12 \cdot 34 \dots k}$, будем рассматривать (29.13.3) как результат замены r_{12} новой величиной $r_{12 \cdot 34 \dots k}$, в то время как все r_{ij} , кроме r_{12} , остаются переменными. R_{11} и R_{22} не содержат r_{12} , так что, используя обозначения, аналогичные обозначениям параграфа 11.5, можно записать (29.13.3) в виде

$$r_{12 \cdot 34 \dots k} = \frac{R_{11 \cdot 22}}{\sqrt{R_{11} R_{22}}} r_{12} + Q,$$

где Q не содержит r_{12} . Отсюда видно, что существует взаимно-однозначное соответствие между этими двумя множествами величин. Якобианом преобразования является

$$\frac{\partial r_{12}}{\partial r_{12 \cdot 34 \dots k}} = \frac{\sqrt{R_{11} R_{22}}}{R_{11 \cdot 22}}.$$

Далее, из формул (11.7.3) и (29.13.3) получим

$$R = \frac{R_{11} R_{22}}{R_{11 \cdot 22}} (1 - r_{12 \cdot 34 \dots k}^2).$$

Произведя в (29.13.1) замену (29.13.3), мы обнаружим, что совместной плотностью вероятности величины $r_{12 \cdot 34 \dots k}$ и всех r_{ij} , отличных от r_{12} ,

является

$$C \frac{(R_{11}R_{22})^{\frac{n-k-1}{2}}}{R_{11 \cdot 22}^{\frac{n-k}{2}}} (1 - r_{12 \cdot 34 \dots k}^2)^{\frac{n-k-2}{2}},$$

где C — некоторая постоянная. Эта функция есть произведение двух множителей, один из которых зависит только от $r_{12 \cdot 34 \dots k}$, а другой — только от r_{ij} . Поскольку величина $r_{12 \cdot 34 \dots k}$ изменяется в интервале $(-1, 1)$, постоянный множитель в ее плотности вероятности можно легко определить, и, в силу (22.1.2), мы получим следующую теорему:

Частный коэффициент корреляции $r_{12 \cdot 34 \dots k}$ независим от всех r_{ij} , отличных от r_{12} , и имеет плотность вероятности

$$(29.13.4) \quad \frac{1}{V\pi} \frac{\Gamma\left(\frac{n-k+1}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right)} (1-x^2)^{\frac{n-k-2}{2}}, \quad (-1 < x < 1).$$

Заметим, что полный коэффициент корреляции r_{12} , согласно формуле (29.7.5), имеет в настоящем случае плотность вероятности

$$\frac{1}{V\pi} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} (1-x^2)^{\frac{n-4}{2}}$$

Чтобы перейти от распределения величины r_{12} к распределению величины $r_{12 \cdot 34 \dots k}$, мы, таким образом, должны лишь заменить n на $n-(k-2)$, т. е. вычесть из n число исключаемых переменных. Р. Фишер [93] показал, что это свойство сохраняется даже в общем случае, когда величины в генеральной совокупности не независимы.

В случае независимости (см. параграф 29.7) величина $t = V^{n-k} \frac{r}{\sqrt{1-r^2}}$, где $r = r_{12 \cdot 34 \dots k}$, имеет распределение Стьюдента с $n-k$ степенями свободы. Следовательно, неравенство

$$(29.13.5) \quad |r_{12 \cdot 34 \dots k}| > \frac{t_p}{V \frac{t_p^2}{t_p^2 + n-k}}$$

(где t_p есть p -процентное значение t при $n-k$ степенях свободы) имеет вероятность $p^0/0$ (см. параграф 31.3, пример 7).

Сводным коэффициентом корреляции $r_{1(2 \dots k)}$ между выборочными значениями величин x_1 и (x_2, \dots, x_k) является, согласно (23.5.2),

неотрицательный квадратный корень

$$(29.13.6) \quad r_{1(2 \dots k)} = \sqrt{1 - \frac{R}{R_{11}}}$$

Соответствующая характеристика совокупности $r_{1(2 \dots k)}$ в нашем случае некоррелированной нормальной генеральной совокупности равна нулю. Постараемся отыскать распределение величины $r_{1(2 \dots k)}$.

В совместном распределении (29.13.1) величин r_{ij} заменим $k-1$ переменных $r_{12}, r_{13}, \dots, r_{1k}$ k новыми переменными $r = r_{1(2 \dots k)}$ и z_2, \dots, z_k с помощью соотношений (29.13.6) и

$$r_{1i} = z_i r \quad (i = 2, 3, \dots, k).$$

Тогда, в силу (11.5.3), мы имеем следующее соотношение между новыми величинами:

$$\sum_{i,j=2}^k R_{11,ij} z_i z_j = R_{11};$$

с помощью этого соотношения одну из величин z_i , например z_2 , можно выразить как функцию от остальных z_i и от r_{ij} с $i > 1, j > 1$. Якобианом нашего преобразования является

$$\begin{vmatrix} \frac{\partial r_{12}}{\partial r} & \frac{\partial r_{12}}{\partial z_3} & \cdots & \frac{\partial r_{12}}{\partial z_k} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial r_{1k}}{\partial r} & \frac{\partial r_{1k}}{\partial z_3} & \cdots & \frac{\partial r_{1k}}{\partial z_k} \end{vmatrix} = \begin{vmatrix} z_2 & r \frac{\partial z_2}{\partial z_3} & r \frac{\partial z_2}{\partial z_4} & \cdots & r \frac{\partial z_2}{\partial z_k} \\ z_3 & r & 0 & \cdots & 0 \\ z_4 & 0 & r & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ z_k & 0 & 0 & \cdots & r \end{vmatrix} = r^{k-2} Q',$$

где Q' не содержит r . Далее, из (29.13.6) мы получаем $R = R_{11}(1 - r^2)$ и, используя это при описанной выше замене переменных в (29.13.1), получим для совместной плотности вероятности новых величин выражение вида

$$r^{k-2} (1 - r^2)^{\frac{n-k-2}{2}} Q'',$$

где Q'' не содержит r .

Таким образом, сводный коэффициент корреляции $r_{1(2 \dots k)}$ независим от всех r_{ij} с $i > 1, j > 1$ и имеет плотность вероятности

$$(29.13.7) \quad \frac{2 \Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{k-1}{2}\right) \Gamma\left(\frac{n-k}{2}\right)} x^{k-2} (1 - x^2)^{\frac{n-k-2}{2}}, \quad (0 < x < 1).$$

Квадрат r^2 имеет бэта-распределение с плотностью вероятности

$$(29.13.8) \quad \beta\left(x; \frac{k-1}{2}, \frac{n-k}{2}\right) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{k-1}{2}\right) \Gamma\left(\frac{n-k}{2}\right)} x^{\frac{k-3}{2}} (1-x)^{\frac{n-k-2}{2}}$$

Распределение величины r было найдено Р. Фишером [94], который решил также в [98] более общую задачу нахождения этого распределения в случае произвольной нормальной генеральной совокупности. В этом общем случае плотность вероятности для r можно представить в виде произведения функции (29.13.7) и степенного ряда, содержащего характеристику совокупности $\rho_{1(2\dots k)}$, так же, как это было сделано в случае обычного коэффициента корреляции (см. (29.7.1)).

В заключение рассмотрим поведение распределения величины r^2 , при больших значениях n . Величина nr^2 имеет плотность вероятности

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{k-1}{2}\right) \Gamma\left(\frac{n-k}{2}\right)} \frac{1}{n} \left(\frac{x}{n}\right)^{\frac{k-3}{2}} \left(1 - \frac{x}{n}\right)^{\frac{n-k-2}{2}},$$

которая при $n \rightarrow \infty$ стремится к пределу

$$(29.13.9) \quad \frac{1}{\frac{k-1}{2} \Gamma\left(\frac{k-1}{2}\right)} x^{\frac{k-3}{2}} e^{-\frac{x}{2}};$$

последнее выражение является плотностью вероятности распределения χ^2 с $k-1$ степенями свободы (см. параграф 31.3, пример 7).

Таким образом, распределение величины r^2 не стремится к нормальному при $n \rightarrow \infty$. Соответственно, из (29.13.8) получаем

$$E(r^2) = \frac{k-1}{n-1}, \quad D^2(r^2) := \frac{2(k-1)(n-k)}{(n-1)^2(n+1)} = O\left(\frac{1}{n^2}\right),$$

так что здесь мы имеем пример исключения, упомянутого в конце параграфа 28.4, когда дисперсия имеет порядок, меньший чем n^{-1} , и теорема о сходимости к нормальному распределению неприменима. Однако это имеет место лишь в рассмотренном здесь частном случае, когда характеристика совокупности ρ равна нулю. Если $\rho \neq 0$, то дисперсия величины r^2 будет иметь порядок n^{-1} и распределение будет приближаться кциальному при $n \rightarrow \infty$.

ГЛАВА 30

КРИТЕРИИ СОГЛАСИЯ И АНАЛОГИЧНЫЕ КРИТЕРИИ

30.1. Критерий χ^2 в случае полностью определенного гипотетического распределения. Теперь мы приступаем к вопросу проверки согласия между теорией вероятностей и фактическими наблюдениями. В настоящем параграфе мы будем находиться в обстановке, описанной в параграфе 26.2, когда имеется выборка в n наблюденных значений некоторой величины (любого числа измерений), и мы хотим узнать, разумно ли рассматривать эту величину как случайную величину, имеющую некоторое данное распределение вероятностей.

Будем называть *гипотезой H* гипотезу о том, что наши данные образуют выборку в n значений случайной величины с данной вероятностной функцией $P(S)$. Мы предполагаем здесь, что $P(S)$ *полностью определена*, так что в ее выражении не содержится никаких неизвестных параметров, и вероятность $P(S)$ может быть вычислена для любого заданного множества S . Требуется выработать метод для проверки того, можно ли считать, что наши данные согласуются с гипотезой H .

Если гипотеза H справедлива, то распределение выборки (см. параграф 25.3), которое является простым дискретным распределением масс $\frac{1}{n}$, сосредоточенных в n наблюденных точках, можно рассматривать как *статистический аналог* (см. параграф 25.5) для генерального распределения, определенного функцией $P(S)$. Из-за случайных колебаний эти два распределения, как правило, не будут совпадать, но можно ожидать, что при больших значениях n распределение выборки будет служить приближением для генерального распределения. Как уже было указано в параграфе 26.2, естественно ввести некоторую *меру расхождения* этих двух распределений и базировать свой критерий на свойствах выборочного распределения этой меры.

Такие меры расхождения можно конструировать различными способами, но наиболее употребительна мера, связанная с важным

критерием χ^2 , введенным К. Пирсоном [183]. Предположим, что пространство значений изучаемой величины разбито на конечное число r частей S_1, \dots, S_r , без общих точек, и пусть p_1, \dots, p_r — соответствующие значения заданной вероятностной функции, так что $p_i = P(S_i)$ и $\sum_1^r p_i = 1$. Предположим, что все $p_i > 0$. Эти r частей S_i могут быть, например, r группами, на которые разбиты наши выборочные значения для целей табулирования. Пусть соответствующие групповые частоты в выборке суть v_1, \dots, v_r , так что каждому множеству S_i принадлежат v_i выборочных значений, и при этом $\sum_1^r v_i = n$.

Первой нашей задачей является отыскание подходящей меры расхождения между распределением выборки и гипотетическим распределением. Любое множество S_i содержит в первом распределении массу $\frac{v_i}{n}$, во втором распределении — массу p_i . В соответствии с общим принципом наименьших квадратов (см. параграф 15.6), примем за такую меру расхождения выражение вида $\sum_1^r c_i \left(\frac{v_i}{n} - p_i \right)^2$, где коэффициенты c_i могут быть выбраны более или менее произвольно. К. Пирсон показал, что если положить $c_i = \frac{n}{p_i}$, то получится мера расхождения с чрезвычайно простыми свойствами. Таким способом получается выражение

$$\chi^2 = \sum_1^r \frac{(v_i - np_i)^2}{np_i} = \sum_1^r \frac{v_i^2}{np_i} - n.$$

Таким образом, χ^2 просто выражается через *наблюденные частоты* v_i и *ожидаемые частоты* np_i для всех r групп.

Исследуем теперь выборочное распределение величины χ^2 , *все время предполагая, что гипотеза H верна*. При этом окажется, что

$$(30.1.1) \quad E(\chi^2) = r - 1,$$

$$D^2(\chi^2) = 2(r - 1) + \frac{1}{n} \left(\sum_1^r \frac{1}{p_i} - r^2 - 2r + 2 \right)$$

Теперь мы докажем следующую теорему К. Пирсона [183], которая показывает, что при увеличении объема выборки выборочное распределение величины χ^2 стремится к предельному распределению, совершенно не зависящему от гипотетической вероятностной функции $P(S)$.

При $n \rightarrow \infty$ выборочное распределение величины χ^2 стремится к распределению, определяемому плотностью вероятности

$$(30.1.2) \quad k_{r-1}(x) = \frac{1}{2^{\frac{r-1}{2}} \Gamma\left(\frac{r-1}{2}\right)} x^{\frac{r-3}{2}} e^{-\frac{x}{2}}, \quad (x > 0),$$

изученной в параграфе 18.1. Употребляя терминологию, введенную в параграфах 18.1 и 29.2, можно сказать, что в пределе величина χ^2 имеет распределение χ^2 с $r-1$ степенями свободы.

В каждом из n наблюдений, дающих n наблюденных точек нашей выборки, мы получаем с вероятностью p_i результат, принадлежащий множеству S_i . Для всякого множества неотрицательных целых чисел v_1, \dots, v_r , таких, что $\sum_i v_i = n$, вероятность того, что в течение n наблюдений мы получим ровно v_i раз результат, принадлежащий множеству S_i , где $i = 1, \dots, r$, равна (см. упражнение 9, стр. 350)

$$\frac{n!}{v_1! \dots v_r!} p_1^{v_1} \dots p_r^{v_r};$$

это выражение является общим членом в разложении полинома $(p_1 + \dots + p_r)^n$. Таким образом, совместным распределением r групповых частот v_1, \dots, v_r является простое обобщение биномиального распределения, известное под названием *мультиномимального распределения*. Совместной характеристической функцией величин v_1, \dots, v_r является

$$(p_1 e^{it_1} + \dots + p_r e^{it_r})^n,$$

что можно непосредственно доказать, прямо обобщая вывод соответствующего выражения (16.2.3) в биномиальном случае. Положим

$$(30.1.3) \quad x_i = \frac{v_i - np_i}{\sqrt{np_i}}, \quad (i = 1, 2, \dots, r);$$

тогда x_i удовлетворяют тождеству

$$\sum_i x_i \sqrt{p_i} = 0 \text{ и } \chi^2 = \sum_i x_i^2.$$

Далее, совместной характеристической функцией величин x_1, \dots, x_r является

$$\varphi(t_1, \dots, t_r) = e^{-n\sqrt{\pi} \sum_i t_i \sqrt{p_i}} \left(p_1 e^{\frac{it_1}{\sqrt{np_1}}} + \dots + p_r e^{\frac{it_r}{\sqrt{np_r}}} \right)^n$$

Из разложения Маклорена для этой функции с помощью нетрудных вычислений выводятся выражения (30.1.1). Далее, для любых фиксированных t_1, \dots, t_r получим

$$\begin{aligned}\log \varphi(t_1, \dots, t_r) &= n \log [1 + \frac{t}{\sqrt{n}} \sum_1^r t_i \sqrt{p_i} - \frac{1}{2n} \sum_1^r t_i^2 + O(n^{-3/2})] = \\ &= -t \sqrt{n} \sum_1^r t_i \sqrt{p_i} = -\frac{1}{2} \sum_1^r t_i^2 + \frac{1}{2} \left(\sum_1^r t_i \sqrt{p_i} \right)^2 + O(n^{-1/2}),\end{aligned}$$

так что характеристическая функция стремится к пределу

$$\lim_{n \rightarrow \infty} \varphi(t_1, \dots, t_r) = e^{-\frac{1}{2} \left[\sum_1^r t_i^2 - \left(\sum_1^r t_i \sqrt{p_i} \right)^2 \right]} = e^{-\frac{1}{2} Q(t_1, \dots, t_r)}$$

Квадратичная форма $Q(t_1, \dots, t_r) = \sum_1^r t_i^2 - \left(\sum_1^r t_i \sqrt{p_i} \right)^2$ имеет матрицу $\Lambda = I - pp'$, где I обозначает единичную матрицу (см. параграф 11.1), а p — вектор-столбец (см. параграф 11.2) $p = (\sqrt{p_1}, \dots, \sqrt{p_r})$. Заменив t_1, \dots, t_r новыми переменными u_1, \dots, u_r с помощью ортогонального преобразования, при котором $u_r = \sum_1^r t_i \sqrt{p_i}$, получим (см. параграф 11.11)

$$Q(t_1, \dots, t_r) = \sum_1^r t_i^2 - \left(\sum_1^{r-1} t_i \sqrt{p_i} \right)^2 = \sum_1^{r-1} u_i^2.$$

Отсюда следует, что $Q(t_1, \dots, t_r)$ неотрицательна и имеет ранг $r-1$ (см. параграф 11.6) и что матрица Λ имеет $r-1$ характеристических чисел (см. параграф 11.9), равных 1, в то время как r -е характеристическое число равно нулю.

Таким образом, при $n \rightarrow \infty$ совместная характеристическая функция величин x_1, \dots, x_r стремится к выражению $e^{-\frac{1}{2} Q}$, являющемуся характеристической функцией некоторого несобственного нормального распределения (см. параграф 24.3) ранга $r-1$, в котором вся масса сосредоточена на гиперплоскости $\sum x_i \sqrt{p_i} = 0$. Из теоремы непрерывности (см. параграф 10.7) следует, что в пределе величины x_1, \dots, x_r имеют несобственное нормальное распределение с нулевыми средними значениями и матрицей вторых моментов Λ . Из параграфа 24.5 следует, что в пределе величина $\chi^2 = \sum_1^r x_i^2$ имеет распределение χ^2 с $r-1$ степенями свободы. Таким образом, теорема К. Пирсона доказана.

С помощью этой теоремы введем теперь критерий для рассматриваемой выше гипотезы H . Пусть χ_p^2 обозначает p -процентное значение χ^2 при $r-1$ степенях свободы (см. параграф 18.1 и таблицу III). Тогда по теореме К. Пирсона, вероятность $P = P(\chi^2 > \chi_p^2)$ при больших n будет приблизительно равна $p\%$. Фиксируем теперь такое малое p , чтобы можно было считать практически несомненным, что при одном испытании событие с вероятностью $p\%$ не произойдет (см. параграф 26.2). Предположим, далее, что n столь велико, что для практических целей вероятность P можно отождествить с ее предельным значением $p\%$. *Если гипотеза H верна, то практически невозможно встретить в единственной выборке значение χ^2 , превышающее χ_p^2 .*

Если в наличной выборке обнаруживается значение $\chi^2 > \chi_p^2$, то мы скажем, что наша выборка обнаруживает *значимое отклонение* от гипотезы H , и мы должны забраковать эту гипотезу, по крайней мере до тех пор, пока мы не получим дальнейших данных. Вероятность того, что такое положение возникнет в случае, когда гипотеза H на самом деле справедлива, т. е. была забракована неправильно, в точности есть вероятность $P = P(\chi^2 > \chi_p^2)$, приближенно равная $p\%$. Поэтому будем говорить, что нами выработан p -процентный *уровень значимости*.

Если, с другой стороны, мы обнаружим значение $\chi^2 \leq \chi_p^2$, то это значение можно считать *совместным* с гипотезой H . Очевидно, отдельный результат такого рода нельзя считать достаточным доказательством истинности гипотезы. Чтобы получить такое доказательство, мы должны повторно применить критерий к новым данным аналогичного характера. Следует применить также и другие критерии, если это возможно.

Когда критерий χ^2 применяется на практике и все ожидаемые частоты $np_i \geq 10$, то предельное распределение χ^2 , приведенное в таблице III на стр. 610, дает, как правило, значение χ_p^2 , соответствующее заданному $P = \frac{p}{100}$, с достаточным для обычных целей приближением. Если некоторые из $np_i < 10$, то обычно бывает целесообразно перед применением критерия объединить маленькие группы, чтобы каждая группа содержала по крайней мере 10 ожидаемых результатов. Если наблюдений так мало, что этого сделать нельзя, то таблицами χ^2 не следует пользоваться, но некоторую информацию можно все же извлечь из значений $E(\chi^2)$ и $D(\chi^2)$, вычисляемых по формулам (30.1.1).

Таблица III применима лишь тогда, если число степеней свободы ≤ 30 . При большем числе степеней свободы обычно достаточно пользоваться теоремой Р. Фишера (см. параграф 20.2) о том, что величина $\sqrt{2\chi^2}$ при n степенях свободы распределена приблизительно нормально, со средним значением $\sqrt{2n-1}$ и с дисперсией, равной 1.

30.2. Примеры. В практических приложениях различных критериев значимости часто употребляются 5-, 1- и 0,1-процентные уровни значимости. Конечно, в каждом данном случае уровень значимости выбирается в зависимости от конкретных условий. В численных примерах, которые будут приведены в этой книге, будем называть значения, превосходящие 5-процентный предел, но не 1-процентный предел, *почти значимыми*, значение между 1- и 0,1-процентными пределами — *значимыми* и значения, превышающие 0,1-процентный предел, — *высоко значимыми*. Конечно, эта терминология чисто условна.

Пример 1. В последовательности n независимых испытаний событие E произошло v раз. Совместны ли эти данные с гипотезой о том, что событие E имеет в каждом испытании заданную вероятность $p = 1 - q$?

Данные можно рассматривать как выборку в n значений величины, равной 1 или 0, в зависимости от того, произошло или не произошло событие E . Гипотеза H заключается в том, что эти две альтернативы имеют фиксированные вероятности p и q . Таким образом, мы имеем две группы данных с наблюденными частотами v и $n - v$, причем соответствующие ожидаемые частоты равны np и nq . Поэтому

$$(30.2.1) \quad \chi^2 = \frac{(v-np)^2}{np} + \frac{(n-v-nq)^2}{nq} = \frac{(v-np)^2}{npq}.$$

Согласно теореме предыдущего параграфа, эта величина χ^2 при больших n приближенно подчиняется распределению χ^2 с одной степенью свободы. Это согласуется с тем фактом (см. параграфы 16.4 и 18.1), что нормированная величина $\frac{v-np}{\sqrt{npq}}$ асимптотически нормальна $(0,1)$, так что ее квадрат в пределе имеет плотность вероятности $k_1(x)$. Соответственно, процентные значения величины χ^2 при одной степени свободы, данные в таблице III, суть квадраты соответствующих значений для нормального распределения, данных в таблице II.

При $n = 4040$ бросаниях монеты Бюффон получил $v = 2048$ выпадений герба и $n - v = 1992$ выпадений решетки. Совместимо ли это с гипотезой о том, что существует постоянная вероятность $p = \frac{1}{2}$?

выпадения герба? Здесь $\chi^2 = \frac{(v - np)^2}{npq} = 0,776$, что лежит ниже 5-процентного значения величины χ^2 с одной степенью свободы, равного (согласно таблице III) 3,841, так что данные можно считать совместными с гипотезой. Соответствующее значение $P = P(\chi^2 \geq 0,776)$ равно приблизительно 0,38, так что мы имеем вероятность около 38% получить отклонения от ожидаемого результата по крайней мере такое же большое, как фактически наблюденное.

Пример 2. Предположим, что в нашем распоряжении имеется k независимых множеств результатов наблюдений, содержащих соответственно по n_1, \dots, n_k результатов, причем событие E произошло соответственно v_1, \dots, v_k раз. Гипотезу о том, что наступление события E имеет постоянную вероятность p , можно проверить различными способами.

Вся совокупность наших данных состоит из $n = \sum n_i$ результатов наблюдений с $v = \sum v_i$ наступлениями события E , так что мы получим первый критерий, вычисляя величину $\chi^2 = \frac{(v - np)^2}{npq}$. Далее, величина $\chi_i^2 = \frac{(v_i - n_i p)^2}{n_i p q}$ дает самостоятельный критерий для i -го множества результатов наблюдений.

Тогда $\chi_1^2, \dots, \chi_k^2$ независимы и при больших n имеют асимптотически одинаковое распределение, именно распределение χ^2 с одной степенью свободы. По теореме сложения (см. (18.1.7)), сумма $\sum \chi_i^2$ имеет в пределе распределение χ^2 с k степенями свободы, что дает совместный критерий для всех наших значений χ_i^2 .

Наконец, если числа n_i велики, то $\chi_1^2, \dots, \chi_k^2$ можно считать выборкой в k наблюденных значений величины с плотностью вероятности $k_1(x)$ и можно применить критерий χ^2 для оценки отклонения выборки от этого гипотетического распределения.

Пример 3. В заключение рассмотрим пример, в котором гипотетическое распределение принадлежит к непрерывному типу. Эйткен ([2], стр. 49) приводит следующие распределения для показаний часов в двух выборках по 500 часов, выставленных в витринах часовщиков (час 0 обозначает промежуток времени от 0 ч. до 1 ч., 1 — от 1 ч. до 2 ч. и т. д.).

Согласно гипотезе о том, что показания часов равномерно распределены по интервалу $(0, 12)$, ожидаемое число случаев каждого класса должно равняться $\frac{500}{12} = 41,67$, откуда мы получаем $\chi_1^2 = 10,000$ для

Таблица 30.2.1

Выборка	Чис												Всего
	0	1	2	3	4	5	6	7	8	9	10	11	
1	41	34	54	39	49	45	41	33	37	41	47	39	500
2	36	47	41	47	49	45	32	37	40	41	37	48	500

первой выборки и $\chi^2_1 = 8,032$ для второй выборки, причем для составной выборки из всех 1000 случаев получается $\chi^2 = 9,464$. В каждом случае имеется $12 - 1 = 11$ степеней свободы, и по таблице III мы убеждаемся, что совпадение хорошее. Можно рассмотреть также сумму $\chi^2_1 + \chi^2_2 = 18,032$, имеющую 22 степени свободы, и это также покажет хорошее совпадение.

30.3. Критерий χ^2 . в случае, когда по выборке оцениваются некоторые параметры. Случай вполне определенного гипотетического распределения встречается в приложениях очень редко. Значительно чаще встречаются случаи, когда гипотетическое распределение содержит некоторое количество неизвестных параметров, относительно значений которых мы можем располагать лишь теми сведениями, какие могут быть извлечены из самой выборки. При этом нам задается вероятностная функция $P(S; \alpha_1, \dots, \alpha_s)$ известной математической формы, но содержащая s неизвестных параметров $\alpha_1, \dots, \alpha_s$. Гипотеза H , которую надлежит проверить, заключается в том, что наша выборка извлечена из совокупности, имеющей распределение с вероятностной функцией P при *некоторых* значениях параметров α_j .

Как и в параграфе 30.1, предположим, что наша выборка разбита на r групп, соответствующих r непресекающимся множествам S_1, \dots, S_r ; обозначим наблюдаемую группу частот через v_1, \dots, v_r , а соответствующие вероятности — через $p_i(\alpha_1, \dots, \alpha_s) = P(S_i; \alpha_1, \dots, \alpha_s)$, $i = 1, 2, \dots, r$.

Если бы „истинные значения“ параметров α_j были известны, оставалось бы лишь вычислить величину

$$(30.3.1) \quad \chi^2 := \sum_{i=1}^r \frac{[v_i - np_i(\alpha_1, \dots, \alpha_s)]^2}{np_i(\alpha_1, \dots, \alpha_s)}$$

и применить критерий, описанный в параграфе 30.1, так что не потребовалось бы никаких дальнейших исследований.

Однако в настоящем случае значения параметров α_j неизвестны и должны быть оценены по выборке. Тогда если заменить в (30.3.1) неизвестные постоянные α_j их оценками, вычисленными по выборке, то p_i уже не будут постоянными, а будут функциями от выборочных значений, и мы не сможем применять теорему параграфа 30.1 о предельном распределении для χ^2 . Как уже было указано в параграфе 26.4, существует, вообще говоря, бесконечное количество различных возможных методов оценки параметров α_j , так что следует ожидать, что свойства выборочного распределения величины χ^2 будут в той или иной степени зависеть от избранного метода.

Проблема нахождения предельного распределения для χ^2 при этих усложненных условиях впервые была рассмотрена Р. Фишером ([91], [95]), который показал, что в этом случае необходимо видоизменить предложение К. Пирсоном предельное распределение (30.1.2). Для одного важного класса методов оценок изменение, предложенное Р. Фишером, имеет очень простой вид: *необходимо лишь уменьшить число степеней свободы предельного распределения (30.1.2) на столько единиц, сколько число параметров, оцениваемых по выборке.*

Мы выберем здесь один особенно важный метод оценки и приведем детальный вывод соответствующего предельного распределения величины χ^2 . В параграфе 33.4 будет показано, что существует целый класс методов оценки, приводящих к тому же предельному распределению.

Естественно попытаться определить „найлучшие“ значения параметров α_j так, чтобы сделать величину χ^2 , определяемую формулой (30.3.1), сколь возможно малой. Это — *метод оценки по минимуму χ^2 .* При этом методе следует решить относительно $\alpha_1, \alpha_2, \dots, \alpha_s$ уравнения

$$(30.3.2) \quad -\frac{1}{2} \frac{\partial f^2}{\partial \alpha_j} = \sum_{i=1}^r \left(\frac{v_i - np_i}{p_i} + \frac{(v_i - np_i)^2}{2np_i^2} \right) \frac{\partial p_i}{\partial \alpha_j} = 0,$$

где $j = 1, 2, \dots, s$, и затем подставить найденные при этом значения α_j в формулу (30.3.1). Предельное распределение величины χ^2 для этого метода оценки было изучено Нейманом и Е. Пирсоном [170], которые использовали методы многомерной геометрии, аналогичные методам, внедренным Р. Фишером. Мы отсылаем читателя по этому вопросу также к статье Шеппарда [213].

Систему уравнений (30.3.2) часто очень трудно решать даже в простейших случаях. Однако можно показать, что при больших n влиянием второго члена в скобках можно пренебречь. Если при дифференцировании χ^2 по a_j , знаменатель второго члена в (30.3.1) считать постоянным, то (30.3.2) заменяется системой

$$(30.3.3) \quad \sum_{i=1}^s \frac{v_i - n \cdot v_i}{p_i} \frac{\partial p_i}{\partial a_j} = 0, \quad (j = 1, 2, \dots, s),$$

иметь дело с которой обычно гораздо проще. Метод оценки, заключающийся в определении a_j из этой системы уравнений, называется *видоизмененным методом минимума χ^2* . Оба метода при довольно общих условиях дают одинаковое предельное распределение χ^2 при больших значениях n , но мы рассмотрим здесь лишь более простой метод, основанный на уравнениях (30.3.3).

В силу условия а) нижеследующей теоремы, уравнения (30.3.3) приводятся к виду

$$(30.3.3a) \quad \sum_{i=1}^r \frac{v_i}{p_i} \frac{\partial p_i}{\partial a_j} = 0;$$

эти уравнения можно записать также в виде $\frac{\partial L}{\partial a_j} = 0$, где $L = p_1^{v_1} \cdots p_r^{v_r}$.

Метод оценки, заключающийся в определении таких значений a_j , для которых величина L принимает наибольшее возможное значение, называется *методом максимума правдоподобия*; этот метод, предложенный Р. Фишером, будет изучаться ниже, в главе 33. С точки зрения задачи, рассматриваемой в этом параграфе, видоизмененный метод минимума χ^2 совпадает с методом максимума правдоподобия. Однако последний метод применим также к задачам значительно более общего характера.

Вследствие важности вопроса мы дадим вывод предельного распределения для χ^2 при наиболее общих условиях, в предположении, что параметры a_j оцениваются с помощью видоизмененного метода минимума χ^2 . Сначала приведем подробную формулировку подлежащей доказательству теоремы.

Пусть заданы r функции $p_1(a_1, \dots, a_s), \dots, p_r(a_1, \dots, a_s)$ от $s < r$ переменных a_1, \dots, a_s , удовлетворяющие для всех точек некоторого невырожденного интервала A в s -мерном пространстве значений a_j следующим условиям:

- a) $\sum_{i=1}^r p_i(a_1, \dots, a_s) = 1$;
- b) $p_i(a_1, \dots, a_s) > c^2 > 0$ при всех i ;

c) все p_i имеют непрерывные производные $\frac{\partial p_i}{\partial a_j}$ и $\frac{\partial^2 p_i}{\partial a_j \partial a_k}$;

d) матрица $D = \left\{ \frac{\partial p_i}{\partial a_j} \right\}$, где $i = 1, \dots, r$ и $j = 1, \dots, s$, имеет ранг s .

Пусть возможные результаты некоторого случайногo эксперимента Σ разбиты на r непересекающихся групп, и предположим, что вероятность получения результата, принадлежащего к i -й группе, равна $p_i^0 = p_i(a_1^0, \dots, a_s^0)$, где $a_0 = (a_1^0, \dots, a_s^0)$ — внутренняя точка интервала A . Пусть y_i обозначает число результатов, принадлежащих к i -й группе, которое было получено в последовательности n повторений эксперимента Σ , так что $\sum_1^n y_i = n$.

Тогда уравнения (30.3.3) видоизмененного метода минимума χ^2 имеют в точности одну систему решений $\alpha = (a_1, \dots, a_s)$, такую, что α сходится по вероятности к a_0 при $n \rightarrow \infty$. Значение χ^2 , получаемое при подстановке этих значений α , в (30.3.1), в пределе при $n \rightarrow \infty$ имеет распределение χ^2 с $r - s - 1$ степенями свободы.

Доказательство этой теоремы довольно сложно, и мы разобьем его на две части. В первой части будет показано, что уравнения (30.3.3) имеют единственное решение α , сходящееся по вероятности (см. параграф 20.3) к a_0 . Во второй части доказательства мы рассмотрим величины

$$(30.3.4) \quad y_i = \frac{y_i - np_i(a_1, \dots, a_s)}{\sqrt{np_i(a_1, \dots, a_s)}}, \quad (i = 1, \dots, r),$$

где $\alpha = (a_1, \dots, a_s)$ — решение уравнений (30.3.3), существование которого будет уже установлено. Мы покажем, что при $n \rightarrow \infty$ совместное распределение величин y_i стремится к некоторому несобственному нормальному распределению, аналогичному предельному распределению величин x_i , определяемых формулой (30.1.3). Как и в соответствующем доказательстве в параграфе 30.1, предельное распределение для

$\chi^2 = \sum_1^r y_i^2$ будет непосредственно получено из параграфа 24.5.

В доказательстве будем все время предполагать, что индекс i про-бегает значения $1, 2, \dots, r$, а индексы j и k — значения $1, 2, \dots, s$.

Сначала введем некоторые матричные обозначения и приведем уравнения (30.3.3) к матричному виду. Обозначив через $\left(\frac{\partial p_i}{\partial a_j} \right)_0$ значение

производной $\frac{\partial p_i}{\partial a_j}$ в точке a_0 , мы сможем записать (30.3.3) в виде

$$(30.3.5) \sum_k (a_k - a_k^0) \sum_l \frac{1}{p_i^0} \left(\frac{\partial p_i}{\partial a_j} \right)_0 \left(\frac{\partial p_l}{\partial a_k} \right)_0 = \sum_l \frac{v_l - np_l^0}{n p_i^0} \left(\frac{\partial p_l}{\partial a_j} \right)_0 + \omega_j(a),$$

где

$$(30.3.6) \omega_j(a) = \sum_l \frac{v_l - np_l^0}{n} \left[\frac{1}{p_l} \frac{\partial p_l}{\partial a_j} - \frac{1}{p_l^0} \left(\frac{\partial p_l}{\partial a_j} \right)_0 \right] - \sum_l (p_l - p_l^0) \left[\frac{1}{p_l} \frac{\partial p_l}{\partial a_j} - \frac{1}{p_l^0} \left(\frac{\partial p_l}{\partial a_j} \right)_0 \right] - \sum_l \frac{1}{p_l^0} \left(\frac{\partial p_l}{\partial a_j} \right)_0 \left[p_l - p_l^0 - \sum_k \left(\frac{\partial p_l}{\partial a_k} \right)_0 (a_k - a_k^0) \right]$$

Обозначим через B матрицу порядка $r \times s$

$$B = \begin{Bmatrix} \frac{1}{\sqrt{p_1^0}} \left(\frac{\partial p_1}{\partial a_1} \right)_0 & \dots & \frac{1}{\sqrt{p_s^0}} \left(\frac{\partial p_1}{\partial a_s} \right)_0 \\ \dots & \dots & \dots \\ \frac{1}{\sqrt{p_r^0}} \left(\frac{\partial p_r}{\partial a_1} \right)_0 & \dots & \frac{1}{\sqrt{p_r^0}} \left(\frac{\partial p_r}{\partial a_s} \right)_0 \end{Bmatrix}$$

Согласно параграфу 11.1, $B = P_0 D_0$, где P_0 — диагональная матрица с элементами $\frac{1}{\sqrt{p_1^0}}, \dots, \frac{1}{\sqrt{p_r^0}}$ по диагонали, а D_0 — матрица, получаемая из

матрицы $D = \left\{ \frac{\partial p_l}{\partial a_j} \right\}$, если положить $a_j = a_j^0$. Поэтому, в силу условия d), матрица B имеет ранг s (см. параграф 11.6). По аналогии с (30.1.3) положим

$$(30.3.7) \quad x_i = \frac{v_i - np_i^0}{\sqrt{np_i^0}}$$

и обозначим через a , a_0 , $\omega(a)$ и x векторы-столбцы (см. параграф 11.2)

$$\begin{aligned} a &= (a_1, \dots, a_s), \\ a_0 &= (a_1^0, \dots, a_s^0), \\ \omega(a) &= (\omega_1(a), \dots, \omega_s(a)), \\ x &= (x_1, \dots, x_r). \end{aligned}$$

первые три из которых можно считать матрицами порядка $s \times 1$, а четвертый — матрицей порядка $r \times 1$.

В матричных обозначениях систему уравнений (30.3.5), где $j = 1, 2, \dots, s$, можно переписать следующим образом (см. параграф 11.3)

$$B' B (a - a_0) = n^{-1/2} B' x + \omega(a).$$

$B'B$ — симметрическая матрица порядка $s \times s$, которая, согласно параграфу 11.9, невырождена, так что существует (см. параграф 11.7) обратная матрица $(B'B)^{-1}$, и мы получаем *)

$$(30.3.8) \quad \alpha = \alpha_0 + n^{-\frac{1}{2}} (B'B)^{-1} B' x + (B'B)^{-1} \omega(\alpha).$$

Таким образом, это матричное уравнение эквивалентно основной системе уравнений (30.3.8).

При каждом фиксированном i случайная величина y_i имеет среднее значение np_i^0 и стандартное отклонение $\sqrt{np_i^0(1-p_i^0)}$; согласно неравенству Бьеенеме — Чебышева (15.7.2), вероятность соотношения $|y_i - np_i^0| \geq \lambda \sqrt{n}$ не превышает $\frac{p_i^0(1-p_i^0)}{\lambda^2} < \frac{p_i^0}{\lambda^2}$. Следовательно, вероятность того, что $|y_i - np_i^0| \geq \lambda \sqrt{n}$ хотя бы при одном значении i не превышает $\lambda^{-2} \sum_i p_i^0 = \lambda^{-2}$, и, обратно, с вероятностью большей, чем $1 - \lambda^{-2}$, имеем

$$(30.3.9) \quad |y_i - np_i^0| < \lambda \sqrt{n} \text{ для всех } i = 1, \dots, r.$$

Теперь будем предполагать, не оговаривая этого в дальнейшем, что y_i удовлетворяют соотношениям (30.3.9). Под λ будем понимать функцию от n , такую, что сама она стремится к бесконечности вместе с n , а $\frac{\lambda^2}{\sqrt{n}}$ стремится к нулю. Например, можно положить $\lambda = n^q$,

$0 < q < \frac{1}{4}$. Все результаты, полученные при этих предположениях, будут справедливы с вероятностью, превышающей $1 - \lambda^{-2}$ и поэтому стремящейся к 1 при $n \rightarrow \infty$.

Согласно условию б), получим из (30.3.7)

$$(30.3.10) \quad |x_i| < \frac{\lambda}{c}.$$

Далее, если $\alpha' = (\alpha'_1, \dots, \alpha'_s)$ и $\alpha'' = (\alpha''_1, \dots, \alpha''_s)$ — какие-либо точки из интервала A , то после некоторых вычислений, используя усло-

*) Заметим, что здесь нельзя писать $(B'B)^{-1} = B^{-1}(B')^{-1}$, так как по предположению $s < r$ и матрица B не квадратная, так что обратная матрица B^{-1} не определена. Если положить $s = r$, то мы увидим, что условия теоремы а) и д) несовместны. В этом случае, если предполагается, что условия а) — с) удовлетворены, то матрицы D, B и $B'B$ должны быть вырожденными, так что обратная матрица $(B'B)^{-1}$ не определена и (30.3.8) не имеет смысла.

вия б) и с) и разлагая формулу (30.3.6) в ряд Тэйлора, получим

$$(30.3.11) \quad |\omega_j(\alpha') - \omega_j(\alpha'')| \leq K_1 |\alpha' - \alpha''| \left(|\alpha' - \alpha_0| + |\alpha'' - \alpha_0| + \frac{\lambda}{\sqrt{n}} \right).$$

В правой части этого неравенства $|\alpha - b|$ обозначает расстояние (см. параграф 3.1) между точками a и b в s -мерном пространстве значений α_j , а K_1 есть постоянная, не зависящая от α' , α'' , j и n .

Теперь определим последовательность векторов $\alpha_v = (\alpha_1^{(v)}, \dots, \alpha_s^{(v)})$, полагая для $v = 1, 2, \dots$

$$(30.3.12) \quad \alpha_v = \alpha_0 + n^{-\frac{1}{2s}} (B'B)^{-1} B'x + (B'B)^{-1} \omega(\alpha_{v-1}).$$

Покажем, что последовательность $\alpha_1, \alpha_2, \dots$ сходится к определенному пределу α , который, очевидно, будет решением уравнения (30.3.8). Согласно (30.3.6), имеем $\omega(\alpha_0) = 0$, так что

$$(30.3.13) \quad \alpha_1 - \alpha_0 = n^{-\frac{1}{2s}} (B'B)^{-1} B'x,$$

а для $v > 0$

$$(30.3.14) \quad \alpha_{v+1} - \alpha_v = (B'B)^{-1} [\omega(\alpha_v) - \omega(\alpha_{v-1})].$$

Обе матрицы $(B'B)^{-1} B'$ и $(B'B)^{-1}$ не зависят от n . Если обозначить через g верхнюю границу абсолютных значений элементов этих двух матриц, то, во-первых, из (30.3.13) и (30.3.10) будет следовать, что каждый элемент вектора $\alpha_1 - \alpha_0$ удовлетворяет неравенству

$$|x_j^{(1)} - x_j^{(0)}| < \frac{rg}{c} \cdot \frac{\lambda}{\sqrt{n}},$$

так что

$$|\alpha_1 - \alpha_0| < K_2 \frac{\lambda}{\sqrt{n}},$$

где K_2 не зависит от n . Таким же образом из (30.3.14) и (30.3.11) следует, что

$$|\alpha_{v+1} - \alpha_v| \leq K_3 |\alpha_v - \alpha_{v-1}| \left(|\alpha_v - \alpha_0| + |\alpha_{v-1} - \alpha_0| + \frac{\lambda}{\sqrt{n}} \right)$$

для всякого $v > 0$, причем K_3 не зависит ни от v , ни от n . Из последних двух неравенств по индукции следует, что для всех достаточно больших n и для всех $v = 0, 1, 2, \dots$ имеем

$$(30.3.15) \quad |\alpha_{v+1} - \alpha_v| \leq K_2 [(4K_2 + 1) K_3]^v \left(\frac{\lambda}{\sqrt{n}} \right)^{v+1}.$$

По предположению, α_0 — внутренняя точка интервала A ; поэтому из последнего неравенства следует, что при достаточно большом n все векторы $\alpha_1, \alpha_2, \dots$ (рассматриваемые как точки в пространстве значений α) принадлежат к A , и последовательность $\alpha_1, \alpha_2, \dots$ сходится к определенному пределу

$$(30.3.16) \quad \alpha = \alpha_0 + (\alpha_1 - \alpha_0) + (\alpha_2 - \alpha_1) + \dots$$

Этот предел, как уже было указано, является решением уравнения (30.3.8), а следовательно, и исходных уравнений (30.3.3). Из (30.3.15) следует, что $\alpha \rightarrow \alpha_0$ при $n \rightarrow \infty$. Более того, α есть единственное решение уравнения (30.3.8), стремящееся к α_0 при $n \rightarrow \infty$. Действительно, если α' — другое решение, стремящееся к α_0 , то

$$\alpha' - \alpha = (B'B)^{-1}(\omega(\alpha') - \omega(\alpha));$$

рассуждая так же, как и выше, мы видим, что

$$|\alpha' - \alpha| \leq K_3 |\alpha' - \alpha| \left(|\alpha' - \alpha_0| + |\alpha - \alpha_0| + \frac{\lambda}{\sqrt{n}} \right),$$

где выражение в скобках стремится к нулю при $n \rightarrow \infty$; это, очевидно, возможно лишь в том случае, когда $\alpha' = \alpha$ для всех достаточно больших n .

Все это было доказано в предположении, что удовлетворяются соотношения (30.3.9), т. е. все доказанное справедливо с вероятностью, превышающей $1 - \lambda^{-2}$ и, следовательно, стремящейся к 1 при $n \rightarrow \infty$. Таким образом, нами установлено существование в точности одного решения уравнения (30.3.8) или системы уравнений (30.3.3), сходящегося по вероятности к α_0 , чем заканчивается первая часть доказательства.

Предполагая попрежнему, что удовлетворяются соотношения (30.3.9), мы получим из (30.3.8), (30.3.13) и (30.3.16):

$$(B'B)^{-1}\omega(\alpha) = \alpha - \alpha_1 = (\alpha_2 - \alpha_1) + (\alpha_3 - \alpha_2) + \dots$$

Из (30.3.15) следует, что каждая компонента вектора $(B'B)^{-1}\omega(\alpha)$ не превышает $K' \frac{\lambda^2}{n}$, где K' не зависит от n , так что (30.3.8) можно переписать в виде

$$(30.3.17) \quad \alpha - \alpha_0 = n^{-\frac{1}{2}} (B'B)^{-1} B' x + \frac{K' \lambda^2}{n} \theta_1,$$

где $\theta_1 = (\theta'_1, \dots, \theta'_s)$ обозначает некоторый вектор-столбец, для которого $|\theta'_j| \leq 1$ при $j = 1, \dots, s$.

Рассмотрим теперь величины y_i , определяемые формулой (30.3.4). Предполагая, что соотношения (30.3.9) удовлетворяются, мы получим с помощью (30.3.7), (30.3.10) и (30.3.17):

$$\begin{aligned} y_i &= \frac{v_i - np_i^0}{\sqrt{np_i^0}} - \sqrt{n} \frac{p_i - p_i^0}{\sqrt{p_i^0}} + \frac{v_i - np_i}{\sqrt{n}} \left(\frac{1}{\sqrt{p_i}} - \frac{1}{\sqrt{p_i^0}} \right) = \\ &= x_i - \sqrt{\frac{n}{p_i^0}} \sum_j \left(\frac{\partial p_i}{\partial x_j} \right)_0 (x_j - a_j^0) + O\left(\frac{\lambda^2}{\sqrt{n}}\right). \end{aligned}$$

Выражая это соотношение в матричных обозначениях, получим

$$y = x - \sqrt{n} B (a - a_0) + \frac{K\lambda^2}{\sqrt{n}} \theta_2,$$

где $y = (y_1, \dots, y_r)$, $\theta_2 = (\theta_1, \dots, \theta_r)$, причем $|y_i| \leq 1$, а K — постоянная, не зависящая от n . Подставляя сюда выражение (30.3.17) для $a - a_0$, получим

$$\begin{aligned} (30.3.18) \quad y &= x - B(B'B)^{-1}B'x + \frac{K\lambda^2}{\sqrt{n}} \theta = \\ &= [I - B(B'B)^{-1}B']x + \frac{K\lambda^2}{\sqrt{n}} \theta, \end{aligned}$$

где I — единичная матрица порядка $r \cdot r$, а $\theta = (\theta_1, \dots, \theta_r)$, где $|\theta_i| \leq 1$; K — постоянная, не зависящая от n .

Далее мы уже не будем предполагать, что соотношения (30.3.9) удовлетворяются, и определим вектор $z = (z_1, \dots, z_r)$, положив

$$y = Ax + z,$$

где A обозначает симметричную матрицу

$$A = I - B(B'B)^{-1}B'$$

Тогда из (30.3.18) будет следовать, что с вероятностью большей, чем $1 - \lambda^{-2}$, $|z_i| \leq K \frac{\lambda^2}{\sqrt{n}}$ при всех i , так что z сходится по вероятности к нулю. Далее, в параграфе 30.1 было показано, что величины x_1, \dots, x_r в пределе при $n \rightarrow \infty$ распределены нормально с нулевыми средними значениями и матрицей вторых моментов $A = I - pp'$, где $p = (\sqrt{p_1^0}, \dots, \sqrt{p_r^0})$. Из последнего предложения параграфа 22.6 следует, что предельное распределение величины y получается с помощью линейного преобразования $y = Ax$, где

$x = (x_1, \dots, x_r)$ имеет нормальное предельное распределение с матрицей вторых моментов Λ ранга $r - 1$.

В силу параграфа 24.4, совместное предельное распределение величин y_1, \dots, y_r — нормальное с нулевыми средними значениями и матрицей вторых моментов

$$\Lambda \Lambda' = [I - B(B'B)^{-1}B'] [I - pp'] [I - B(B'B)^{-1}B']'$$

Согласно условию а), j -м элементом вектора $B'p$ является

$$\sum_i \left(\frac{\partial p_j}{\partial x_i} \right)_0 = 0,$$

так что вектор $B'p$ тождественно равен нулю. Поэтому после перемножения мы убеждаемся в том, что матрица вторых моментов предельного распределения для y сводится к

$$(30.3.19) \quad \Lambda \Lambda' = I - pp' - B(B'B)^{-1}B'.$$

Остается лишь показать, что эта симметричная матрица порядка $r \cdot r$ имеет $r - s - 1$ характеристических чисел, равных 1, а остальные характеристические числа этой матрицы равны нулю, так что влияние последнего члена в (30.3.19) заключается в понижении ранга матрицы на s единиц. Тогда из параграфа 24.5 будет следовать, что сумма квадратов

$$\chi^2 = \sum_i y_i^2$$

имеет в пределе распределение χ^2 с $r - s - 1$ степенями свободы, и наша теорема будет доказана.

Для этой цели сначала заметим, что, согласно параграфу 11.9, s характеристических чисел x_j симметричной матрицы $B'B$ положительны. Полагая $x_j = \mu_j^2$, где $\mu_j > 0$, и обозначив через M диагональную матрицу с элементами μ_1, \dots, μ_s по диагонали, можно, согласно параграфу 11.9, найти ортогональную матрицу C порядка $s \cdot s$, такую, что $C'B'BC = M^2$, откуда $(B'B)^{-1} = (CM^2C')^{-1} = C'M^{-1} \cdot M^{-1}C'$. Отсюда следует, что

$$(30.3.20) \quad B(B'B)^{-1}B' = BCM^{-1} \cdot M^{-1}C'B' = HH',$$

где $H = BCM^{-1}$ — матрица порядка $r \cdot s$, такая, что

$$H'H = M^{-1}C'B'BCM^{-1} = M^{-1}M^2M^{-1} = I,$$

причем I обозначает здесь единичную матрицу порядка $s \cdot s$. Последнее соотношение означает, что s столбцов матрицы H удовлетворяют

условиям ортогональности (11.9.2). Выше было показано, что $B'p=0$, откуда $H'p=M^{-1}C'B'p=0$. Таким образом, если дополнить матрицу H столбцом с элементами $\sqrt{p_1^0}, \dots, \sqrt{p_r^0}$, то $s+1$ столбцов новой матрицы H_1 , будут удовлетворять условиям ортогональности. Поскольку $s < r$, то, согласно параграфу 11.9, можно найти ортогональную матрицу K порядка $r \cdot r$, последние $s+1$ столбцов которой будут совпадать с матрицей H_1 .

Тогда $K'p$ будет матрицей порядка $r \cdot 1$, т. е. вектором-столбцом, и по правилу перемножения матриц мы получим $K'p=(0, \dots, 0, 1)$. Таким образом, произведение $K'pp'K=(0, \dots, 0, 1) \cdot \{0, \dots, 0, 1\}$ есть матрица порядка $r \cdot r$, все элементы которой равны нулю, кроме последнего элемента главной диагонали, равного единице. Аналогичным образом, произведение $K'H'HK$ есть матрица порядка $r \cdot r$, все элементы которой равны нулю, кроме s диагональных элементов, непосредственно примыкающих к последнему элементу и равных единице.

Согласно (30.3.20), матрица вторых моментов (30.3.19) принимает вид $I - pp' - HH'$. Из вышеуказанного следует, что преобразованная матрица $K'(I - pp' - HH')K$ — диагональная матрица, первые $r-s-1$ диагональных элементов которой равны 1, а остальные диагональные элементы равны нулю. Этим доказано наше утверждение относительно характеристических чисел матрицы вторых моментов (30.3.19). Как указано выше, на этом заканчивается доказательство нашей теоремы.

С помощью этой теоремы теперь можно ввести критерий для гипотезы H в точности таким же образом, как и в простейшем случае, рассмотренном в параграфе 30.1. В следующем параграфе будут приведены некоторые примеры применения этого критерия.

30.4. Примеры. Рассмотрим применение критерия χ^2 к двум особенно важным случаям, именно к распределению Пуассона и к нормальному распределению. Другие простые распределения можно исследовать аналогичным образом.

Пример 1. *Распределение Пуассона.* Предположим, что требуется проверить гипотезу о том, что данная выборка в n значений x_1, \dots, x_n извлечена из некоторого распределения Пуассона с неизвестным значением параметра λ . Каждое x_μ есть некоторое неотрицательное целое число i ; разобьем x_μ по их значениям на r групп,

объединив малочисленные данные отдельно для больших и малых i . Предположим, что при этом мы получили

$$v_k \text{ наблюдений с } x \leq k,$$

$$v_i \quad \text{при } x = i, \text{ где } i = k+1, \dots, k+r-2,$$

$$v_{k+r-1} \quad \text{при } x \geq k+r-1$$

Если обозначить $\tilde{\omega}_i = P(x=i) = \frac{\lambda^i}{i!} e^{-\lambda}$, то соответствующими вероятностями будут

$$p_k = P(x \leq k) = \sum_0^k \tilde{\omega}_i,$$

$$p_i = P(x=i) = \tilde{\omega}_i \text{ при } i = k+1, \dots, k+r-2,$$

$$p_{k+r-1} = P(x \geq k+r-1) = \sum_{k+r-1}^{\infty} \tilde{\omega}_i.$$

Чтобы оценить неизвестный параметр λ с помощью видоизмененного метода минимума χ^2 , следует решить систему (30.3.3) или эквивалентную ей систему (30.3.3а). Поскольку неизвестный параметр только один, то $s=1$, система сводится к одному уравнению, и (30.3.3а) дает

$$v_k \frac{\sum_0^k \left(\frac{i}{\lambda} - 1\right) \tilde{\omega}_i}{\sum_0^k \tilde{\omega}_i} + \sum_{k+1}^{k+r-2} \left(\frac{i}{\lambda} - 1\right) v_i + v_{k+r-1} \frac{\sum_{k+r-1}^{\infty} \left(\frac{i}{\lambda} - 1\right) \tilde{\omega}_i}{\sum_{k+r-1}^{\infty} \tilde{\omega}_i} =$$

Это уравнение имеет единственный корень $\lambda = \lambda^*$, равный

$$\lambda^* = \frac{1}{n} \left[v_k \frac{\sum_0^k i \tilde{\omega}_i}{\sum_0^k \tilde{\omega}_i} + \sum_{k+1}^{k+r-2} i v_i + v_{k+r-1} \frac{\sum_{k+r-1}^{\infty} i \tilde{\omega}_i}{\sum_{k+r-1}^{\infty} \tilde{\omega}_i} \right]$$

Здесь второй член в скобках равен сумме всех x_μ таких, что $k < x_\mu < k+r-1$, а первый и последний члены приближительно равны сумме всех x_μ , которые $\leq k$ или $\geq k+r-1$ соответственно. Таким образом, оценкой λ^* для λ может служить среднее арифметическое выборочных значений

$$\lambda^* = \frac{1}{n} \sum_1^n x_\mu = \bar{x}$$

Полагая в теореме предыдущего параграфа $s = 1$, мы обнаружим, что предельное распределение χ^2 в этом случае имеет $r - 2$ степеней свободы.

Таблица 30.4.1.

Применение критерия χ^2 к распределению Пуассона

t	Пример 1a		Пример 1b		Пример 1c				
	число чистых периодов с i-частичными	np_i	$(y_i - np_i)^2 / np_i$	число чистых периодов с i-кратными частичами	np_i	$(y_i - np_i)^2 / np_i$	число чистых периодов с i-цветками	np_i	$(y_i - np_i)^2 / np_i$
0	57	54,399	0,1244	—	—	—	—	—	—
1	203	210,523	0,2688	—	—	—	—	—	—
2	383	407,361	1,4568	—	—	—	—	—	—
3	525	525,496	0,0005	—	—	—	5	—	—
4	532	508,418	1,0938	1	—	—	2	—	—
5	408	393,515	0,5332	3	—	—	10	25,0217	2,5717
6	273	253,817	1,4498	5	—	—	19	19,1330	0,0010
7	139	140,325	0,0125	8	15,7955	0,0919	20	24,1934	0,7268
8	45	67,882	7,7132	13	11,4043	0,2233	42	26,7639	8,6736
9	27	29,189	0,1642	14	15,0930	0,0792	27	26,3178	0,0177
10	10	17,075	0,0677	15	17,9773	0,4931	25	23,2913	0,1254
11	4	—	—	15	19,4661	1,0247	23	18,7389	0,9689
12	2	—	—	21	19,3217	0,1458	11	13,8199	0,5754
13	—	—	—	18	17,7032	0,0050	5	22,7171	1,9861
14	—	—	—	17	15,0616	0,2495	6	—	—
15	—	—	—	16	11,9599	1,3648	4	—	—
16	—	—	—	9	8,9034	0,0010	—	—	—
17	—	—	—	6	16,3140	0,3282	—	—	—
18	—	—	—	3	—	—	—	—	—
19	—	—	—	2	—	—	—	—	—
20	—	—	—	2	—	—	1	—	—
21	—	—	—	1	—	—	—	—	—
Всего	2 608	2 608,000	12,8849	169	169,0000	4,0065	200	200,0000	15,6466
	$\bar{x} = 3,870$ $\chi^2 = 12,885$ $P = 0,17$			$\bar{x} = 11,911$ $\chi^2 = 4,006$ $P = 0,91$			$\bar{x} = 8,850$ $\chi^2 = 15,647$ $P = 0,03$		

В таблице 30.4.1 приведены три численных примера применения нашего критерия. Пример 1а дает число α -частиц, излучаемых диском в течение 2608 периодов по 7,5 секунды, согласно наблюдениям Резерфорда и Гейгера (см. [2], стр. 77). Пример 1б дает числа красных кровяных шариков в 169 отделениях гематометра, наблюденные Н. Хольмбергом. Пример 1с дает числа цветков на 200 растениях *Primula veris*, подсчитанные М.-Л. Крамер (1928 г.).

В соответствии с правилами, приведенными в параграфе 30.1, отдельные группы в каждой выборке следует объединить так, чтобы каждая группа содержала по крайней мере 10 результатов наблюдений. Так, в примере 1б наблюденные частоты в группах $i=7$ и $i=17$ равны соответственно $1+3+5+8=17$ и $6+3+2+2+1=14$. Совпадение оказалось хорошим в примере 1а, очень хорошим в 1б, но в 1с обнаружилось «почти значимое» отклонение от гипотетического распределения Пуассона, вызванное главным образом чрезмерным числом растений с восемью цветками.

Рассмотренные в этих примерах величины принадлежат к классу величин, которые часто хорошо согласуются с распределением Пуассона. Если данные обнаруживают значимое отклонение от распределения Пуассона, то совпадение можно значительно улучшить, введя гипотезу о том, что параметр λ сам есть случайная величина, имеющая распределение пирсонова типа III с плотностью вероятности $\frac{a^x}{\Gamma(x)} x^{x-1} e^{-ax}$, где a и x — положительные параметры. При этом получается *отрицательно-биномиальное распределение* (см. упражнение 21, стр. 286), которое имеет интересные приложения, например, к статистике несчастных случаев и заболеваний (Гринвуд и Юл [119], Эггенбергер [81], Ньюболд [159а]) и к задачам, связанным с количеством индивидуумов данной породы в выборках из биологических совокупностей (Энерот [31а], Фишер, Корбет и Вильямс [111]). В статистике несчастных случаев введение величины λ можно интерпретировать как способ учета *различных степеней риска* для членов данной совокупности. Аналогичные интерпретации возможны и в других случаях. Этот вопрос можно рассматривать также с точки зрения *случайных процессов* (см. Лундберг [152]).

Пример 2. Нормальное распределение. Пусть выборка в n значений x_1, \dots, x_n разбита на r групп так, что i -я группа, содержащая y_i значений, принадлежит интервалу $(\xi_i - \frac{1}{2}h, \xi_i + \frac{1}{2}h)$, где $\xi_i = \xi_1 + (i-1)h$. Нам надлежит проверить гипотезу о том, что эта выборка извлечена из некоторой нормальной совокупности с неизвестными значениями параметров m и σ . Если эта гипотеза верна, то

вероятности p_i , соответствующая i -й группе выборочных значений, равняется

$$p_i = \frac{1}{\sigma \sqrt{2\pi}} \int e^{-\frac{(x-m)^2}{2\sigma^2}} dx,$$

где интегрирование производится по i -му интервалу разбиения. Для двух крайних групп ($i=1$ и $i=r$) за интервалы разбиения следует принять соответственно $(-\infty, \xi_1 + \frac{1}{2}h)$ и $(\xi_r - \frac{1}{2}h, +\infty)$

Тогда, обозначая для краткости $g(x) = e^{-\frac{(x-m)^2}{2\sigma^2}}$, получим

$$\frac{\partial p_i}{\partial m} = \frac{1}{\sigma^3 \sqrt{2\pi}} \int (x-m) g(x) dx,$$

$$\frac{\partial p_i}{\partial \sigma} = \frac{1}{\sigma^4 \sqrt{2\pi}} \int (x-m)^2 g(x) dx - \frac{p_i}{\sigma}.$$

Тогда уравнения (30.3.3а) после некоторых простых преобразований примут вид

$$m = \frac{1}{n} \sum_i v_i \frac{\int x g(x) dx}{\int g(x) dx},$$

$$\sigma^2 = \frac{1}{n} \sum_i v_i \frac{\int (x-m)^2 g(x) dx}{\int g(x) dx};$$

при этом все интегралы распространены на соответствующие интервалы разбиения. Сначала предположим, что выборка разбита на группы так, что крайние группы не содержат наблюденных значений. Тогда $v_1 = v_r = 0$. При малых значениях h можно получить приближенное решение, просто заменяя функции под знаками интегралов их значениями в средних точках ξ_i соответствующих интервалов разбиения. При этом мы получим оценки m^* и σ^* , определяемые выражениями

$$m^* = \frac{1}{n} \sum_i v_i \xi_i, \quad \sigma^{*2} = \frac{1}{n} \sum_i v_i (\xi_i - m^*)^2.$$

Таким образом, m^* и σ^{*2} совпадают со средним значением \bar{x} и дисперсией s^2 группированной выборки, вычисленными по обычному правилу (см. параграф 27.9), согласно которому все выборочные значения некоторой группы перемещаются в среднюю точку соответствующего

интервала разбиения. Чтобы получить лучшее приближение, можно разложить функции, стоящие под знаками интегралов, в ряды Тэйлора в окрестностях средних точек ξ_i . При малых h после некоторых вычислений убеждаемся, что приведенные выше формулы надлежит исправить следующим образом:

$$m^* = \frac{1}{n} \sum_i v_i \xi_i + O(h^4), \quad \sigma^{*2} = \frac{1}{n} \sum_i v_i (\xi_i - m^*)^2 - \frac{h^2}{12} + O(h^4).$$

Итак, пренебрегая членами порядка h^4 , можно использовать в качестве оценки для m среднее значение группированной выборки, а для дисперсии необходимо применить поправку Шеппарда (см. параграф 27.9).

Такая процедура может дать удовлетворительное приближение, даже если h не очень мало, а крайние группы фактически не пусты, но содержат лишь малую часть всей выборки. На практике удобно объединять крайние группы выборки в соответствии с правилом параграфа 30.1 так, чтобы каждая группа содержала хотя бы 10 выборочных значений. В качестве оценок для m и σ^2 следует взять значения \bar{x} и s^2 , вычисленные по первоначальной группировке до какого-либо объединения групп, причем к s^2 должна быть применена поправка Шеппарда. Если r — число групп после объединения, фактически использованное при вычислении χ^2 , то предельное распределение для χ^2 имеет $r - 3$ степени свободы, так как два параметра мы определяем по выборке.

Если генеральное распределение нормально, то асимптотические выражения для средних значений и дисперсий выборочных характеристик g_1 и g_2 даются формулами (27.7.9), а соответствующие точные выражения — формулами (29.3.7). Дальнейший признак нормальности распределения заключается в сравнении величин g_1 и g_2 , вычисленных по фактической выборке, с соответствующими средними значениями и дисперсиями.

В таблице 30.4.2 приведен результат подбора нормальных кривых к распределениям средних температур июня и июля в Стокгольме в течение $n = 100$ лет (1841—1940).

В исходных данных приведены значения, округленные до ближайшего десятичного знака, так что точными интервалами разбиения являются (12,45; 12,95) и т. д. Мы пользуемся здесь несколько меньшими группами, чем это обычно рекомендуется. Оба значения χ^2 показывают удовлетворительное согласие с гипотезой о нормальном

распределении. В таблице приведены также значения g_1 и g_2 . Соответствующие гипотезе точные выражения (29.3.7) дают, в обоих случаях

$$E(g_1) = 0, D(g_1) = 0,238; \quad E(g_2) = -0,059, \quad D(g_2) = 0,55,$$

так что никакое из наблюдаемых значений не отличается значимо от соответствующего среднего значения.

Таблица 30.4.2

**Распределение средних температур июня и июля
в Стокгольме (1841—1940 гг.)**

Температура (в °C)	Июнь		Июль	
	Частота наблюденная	ожидаемая	Температура (в °C)	Частота наблюденная
≤12,4	10	12,89	≤14,9	11
12,5—12,9	12	7,89	15,0—15,4	7
13,0—13,4	9	10,20	15,5—15,9	8
13,5—13,9	10	11,93	16,0—16,4	13
14,0—14,4	19	12,62	16,5—16,9	14
14,5—14,9	10	12,08	17,0—17,4	13
15,0—15,4	9	10,46	17,5—17,9	6
15,5—15,9	6	8,19	18,0—18,4	9
16,0—16,4	7	5,81	18,5—18,9	7
≥16,5	8	7,93	≥19,0	12
Сумма	100	100,00	Сумма	100
				100,00
$\bar{x} = 14,23$ $g_1 = 0,098$ $\chi^2 = 7,85$ (7 степ. своб.) $P = 0,35$			$\bar{x} = 16,98$ $g_1 = 0,332$ $\chi^2 = 3,34$ (7 степ. своб.) $P = 0,85$	

На фиг. 25, стр. 361, приведена диаграмма эмпирического закона распределения для июня (по выборке в 100 выборочных значений) вместе с соответствующей нормальной кривой.

Если значения g_1 или g_2 значимы, то совпадение с нормальной кривой часто можно значительно улучшить, используя разложения

Шарлье или Эджвортса, изучавшиеся в параграфах 17.6 — 17.7. При этом следует иметь в виду, что для каждого дополнительного параметра, определяемого по выборке, число степеней свободы следует уменьшить на единицу.

В таблице 30.4.3 приведено распределение размеров $n = 12\,000$ бобов *Phaseolus vulgaris* (данные Иогансена, заимствованные нами у Шарлье [9], стр. 73). Согласно гипотезе о нормальном распределении, имеем

$$E(g_1) = 0, D(g_1) = 0,0224 \text{ и } E(g_2) = -0,0005, D(g_2) = 0,0447,$$

Таблица 30.4.3
Размеры бобов ($\xi_1 = 6,825 \text{ мм}$, $h = 0,25 \text{ мм}$)

Номер группы i	Наблюдаемая частота χ_i	Ожидаемая частота μ_i		
		нормальное распределение	первое приближение	второе приближение
1	32	67,6	17,5	26,6
2	103	132,2	98,3	90,4
3	239	309,8	291,5	277,2
4	624	617,3	648,9	636,8
5	1 187	1 045,7	1 142,2	1 141,1
6	1 650	1 505,8	1 630,4	1 639,9
7	1 883	1 842,3	1 918,1	1 931,6
8	1 930	1 919,9	1 892,4	1 906,2
9	1 638	1 697,9	1 587,3	1 599,5
10	1 130	1 277,3	1 158,8	1 163,5
11	737	817,0	752,4	745,1
12	427	444,2	441,9	427,3
13	221	205,3	235,6	223,8
14	110	80,7	112,7	109,1
15	57	27,0	47,5	49,7
16	32	10,0	24,5	32,2
Сумма	12000	12000,0	12000,0	12000,0
$\bar{x} = 8,512$ $s = 0,6163$ $g_1 = -0,2878$ $g_2 = 0,1953$	$\chi^2 = 196,5$ (13 степ. своб.) $P < 0,001$	$\chi^2 = 34,3$ (12 степ. своб.) $P < 0,001$	$\chi^2 = 14,9$ (11 степ. своб.) $P = 0,19$	

так что фактические значения g_1 и g_2 , приведенные в таблице, значительно отличаются от значений, которые мы ожидаем при гипотезе о нормальном распределении.

В таблице приведены также ожидаемые частоты и соответствующие значения χ^2 , вычисленные по каждой из трех гипотез о том, что, в соответствии с (17.7.3) или (17.7.5)*), плотность вероятности нормированной величины $\frac{x - \bar{x}}{s}$ имеет вид:

$$\text{а) нормальная: } \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

$$\text{б) первое приближение: } \varphi(x) = \frac{g_1}{3!} \varphi^{(3)}(x);$$

$$\text{в) второе приближение: } \varphi(x) = \frac{g_1}{3!} \varphi^{(3)}(x) + \frac{g_2}{4!} \varphi^{(4)}(x) + \frac{10g_1^2}{6!} \varphi^{(6)}(x).$$

В первых двух случаях отклонения выборок от гипотетических распределений высоко значимы, поскольку $P < 0,001$; но в третьем случае $P = 0,19$, так что совпадение удовлетворительное. На фиг. 26, стр. 362, показана гистограмма этого распределения и дан график плотности вероятности для второго приближения. Более подробные исследования этого и других примеров приведены у Крамера [70].

30.5. Таблицы сопряженности признаков. Предположим, что n элементов выборки расположены в таблицу с двумя входами типа таблицы 30.5.1, по двум переменным признакам (количественным или качественным).

Таблица такого рода называется *таблицей сопряженности признаков*, и часто требуется проверить гипотезу о том, что переменные признаки, по которым построена таблица, *независимы*. Обозначим через p_{ij} вероятность того, что выбранный случайным образом индивидуум принадлежит к i -й строке и j -му столбцу этой таблицы. Тогда гипотеза о независимости эквивалентна (см. (21.1.4)) гипотезе о том, что существует $r+s$ постоянных $p_{i\cdot}$ и $p_{\cdot j}$, таких, что

$$p_{ij} = p_{i\cdot} p_{\cdot j},$$

$$\sum_i p_{i\cdot} = \sum_j p_{\cdot j} = 1.$$

*). Тем же методом, как и выше, можно показать, что в качестве оценок для коэффициентов γ_1 и γ_2 следует использовать g_1 и g_2 , вычисленные по группированной выборке с использованием поправок Шеппарда.

Таблица 30.5.1

Признаки	1 2 ... s	Сумма
1	v ₁₁ v ₁₂ ... v _{1s}	v _{1.}
2	v ₂₁ v ₂₂ ... v _{2s}	v _{2.}
.
.
.
r	v _{r1} v _{r2} ... v _{rs}	v _{r.}
Сумма	v _{..1} v _{..2} ... v _{..s}	n

В соответствии с этой гипотезой, совместное распределение двух признаков содержит $r+s-2$ неизвестных параметра, так как с помощью двух последних соотношений две из $r+s$ постоянных, например, $p_r.$ и $p_{..s}$, можно выразить через остальные $r+s-2$ постоянные.

Чтобы применить к этой проблеме критерий χ^2 , следует вычислить

$$\chi^2 = \sum_{i,j} \frac{(v_{ij} - np_i p_{..j})^2}{np_i p_{..j}},$$

где суммирование распространяется на все rs групп в таблице сопряженности признаков, и затем заменить здесь $p_i.$ и $p_{..j}$ их оценками, полученными из уравнений (30.3.3) или (30.3.3а), которые в этом случае имеют вид

$$\sum_j \left(\frac{v_{ij}}{p_i.} - \frac{v_{rj}}{p_r.} \right) = 0, \quad (i = 1, \dots, r-1),$$

$$\sum_i \left(\frac{v_{ij}}{p_{..j}} - \frac{v_{is}}{p_{..s}} \right) = 0, \quad (j = 1, \dots, s-1).$$

Решением этих уравнений является

$$p_i. = \frac{v_{..i}}{n}, \quad p_{..j} = \frac{v_{..j}}{n},$$

так что оценками, которые надлежит использовать, являются просто частоты, вычисленные по частным суммам. После подстановки этих оценок вместо $p_i.$ и $p_{..j}$ выражение для χ^2 принимает вид

$$(30.5.1) \quad \chi^2 = n \sum_{i,j} \frac{\left(\frac{v_{ij}}{p_i.} - \frac{v_{..i} v_{..j}}{n} \right)^2}{\frac{v_{..i} v_{..j}}{n}} = n \left(\sum_{i,j} \frac{v_{ij}^2}{v_{..i} v_{..j}} - 1 \right).$$

Поскольку мы имеем здесь rs групп и $r+s-2$ параметров, определяемых по выборке, предельное распределение для χ^2 имеет $rs-(r+s-2)-1=(r-1)(s-1)$ степеней свободы. Точные выражения для среднего значения и дисперсии величины χ^2 , определяемой формулой (30.5.1), были получены различными авторами (см. Холден [123], где приведены дальнейшие ссылки на литературу). Предполагая, что гипотеза о независимости верна, получим

$$(30.5.2) \quad E(\chi^2) = \frac{n}{n-1} (r-1)(s-1)$$

Дисперсия имеет сложное выражение, которое мы здесь не приводим.

Большое значение χ^2 будет означать, что отклонение от гипотезы о независимости значимо, но из этого нельзя извлечь прямых сведений о *степени зависимости* или *связанности* признаков. С другой стороны, величина

$$f^2 = \frac{\chi^2}{n} = \sum_{i,j} \frac{\left(\frac{v_{ij}}{n} - \frac{v_i}{n} \frac{v_j}{n} \right)^2}{\frac{v_i}{n} \frac{v_j}{n}}$$

является выборочной характеристикой, соответствующей *средней квадратической сопряженности* ψ^2 , определяемой формулой (21.9.6). Если q — наименьшее из чисел r и s , то из параграфа 21.9 следует, что

$$0 \leq \frac{f^2}{q-1} = \frac{\chi^2}{n(q-1)} \leq 1.$$

Верхний предел 1 достигается тогда и только тогда, когда каждая строка (при $r \geq s$) или каждый столбец (при $r \leq s$) содержат лишь один отличный от нуля элемент. Таким образом, $\frac{\chi^2}{n(q-1)}$ можно считать мерой связанности признаков, обнаруженной в выборке. Распределение этой меры получается, конечно, с помощью простой замены переменной в распределении для χ^2 . (О других мерах связанности см., например, учебник Юла и Кендала [43], гл. 3—4.)

По шведской переписи населения в марте 1936 г. из совокупности всех супружеских пар в районах Швеции была получена выборка в 25 263 пары, вступивших в брак в течение пяти последних лет. В таблице 30.5.2 приведено распределение годовых доходов и количества детей у супружеских пар в этой выборке. По формуле (30.5.1) получаем $\chi^2 = 568,5$ с $(5-1)(4-1) = 12$ степенями свободы, так что

отклонение от гипотезы о независимости высоко значимо. С другой стороны, мера связности равна $\frac{\chi^2}{n(q-1)} = 0,00750$, что указывает лишь на очень слабую зависимость между признаками.

Таблица 30.5.2

Распределение супружеских пар по годовому доходу и количеству детей

Количество детей	Доходы (за единицу принято 1000 крон)				Сумма
	0-1	1-2	2-3	>3	
0	2 161	3 577	2 184	1 636	9 558
1	2 755	5 081	2 222	1 052	11 110
2	936	1 753	649	306	3 635
3	225	419	96	38	778
≥ 4	39	98	31	14	182
Сумма	6 116	10 928	5 173	3 016	25 263

В частном случае, когда $r=s=2$, таблица сопряженности признаков становится четырехзначной таблицей, и выражение (30.5.1) обращается в

$$(30.5.3) \quad \chi^2 = n \frac{(v_{11}v_{22} - v_{12}v_{21})^2}{v_1v_2v_{-1}v_{-2}},$$

так что $f^2 = \frac{\chi^2}{n}$ соответствует выражению (21.9.7) для φ^2 . Если признаки количественны, то f^2 совпадает с квадратом коэффициента корреляции выборки (см. (21.9.7) и (21.7.3)). В случае четырехзначной таблицы в предельном распределении для χ^2 имеется лишь $(2-1)(2-1)=1$ степень свободы и $q-1=1$.

В таблице 30.5.3 приведено распределение цвета волос на голове и бровей у 46 542 шведских призывников, согласно данным Лундборга и Линдерса [26]. По формуле (30.5.3) получаем $\chi^2 = 19 288$ и $f^2 = 0,414$, что указывает на заметную зависимость между признаками.

Когда в четырехзначной таблице ожидаемые частоты $\frac{v_i v_j}{n}$ малы, то приближение, получаемое с помощью обычных таблиц χ^2 -может быть улучшено, если вычислить χ^2 по первому выражению (30.5.1) и перед возвышением в квадрат уменьшить абсолютное зна-

чение каждой разности $v_{ij} - \frac{v_{i..}v_{.j}}{n}$ на $\frac{1}{2}$. Это называется *поправкой Иэйтса* [250].

Таблица 30.5.3

Цвет волос у шведских призывников

Цвет бровей	Цвет волос на голове		Сумма
	светлые или рыжие	темные	
Светлые или рыжие	30 472	3 238	33 710
Темные	3 364	9 468	12 832
Сумма	33 836	12 706	46 542

30.6. χ^2 как критерий однородности. Таблица сопряженности признаков 30.5.1 представляет собой совокупный результат последовательности n повторений случайного эксперимента, причем результаты классифицируются по двум переменным признакам. Однако во многих случаях встречаются таблицы формально такого же вида, но совершенно иной природы.

Предположим, что осуществлено s последовательных рядов экспериментов, состоящих соответственно из n_1, \dots, n_s единичных наблюдений, причем числа n_j не случайны, а должны рассматриваться как заданные. В каждом эксперименте наблюдается некоторый переменный признак, и результаты каждого ряда наблюдений разбиваются по значениям этого признака на r групп. Количество результатов наблюдений в i -й группе j -го ряда будем обозначать через v_{ij} . Тогда наши данные будут располагаться в таблице такого же вида, как таблица 30.5.1, причем суммы по столбцам $v_{.j}$ в этой таблице обозначены здесь через n_j . В этом случае, однако, таблица представляет собой не результат одной единственной последовательности наблюдений, как в случае таблицы 30.5.1, а результат s независимых рядов наблюдений, каждому из которых соответствует в нашей таблице один столбец.

В таких случаях часто бывает необходимо проверить гипотезу о том, что s выборок, представляемых столбцами таблицы, *извлечены из одной и той же совокупности*, или, иначе говоря, что в этом отношении наши данные *однородны*. Такая гипотеза эквивалентна гипотезе о том, что существует r постоянных p_1, \dots, p_r таких, что $\sum_i p_i = 1$,

и вероятность принадлежности отдельного результата к i -й группе во всех s последовательностях равна p_i .

Чтобы проверить эту гипотезу, вычислим χ^2 по той же формуле (30.5.1), что в предыдущем случае. Незначительно изменив доказательство теоремы параграфа 30.3, убеждаемся, что если гипотеза верна, то χ^2 имеет обычное предельное распределение с тем же числом степеней свободы $(r - 1)(s - 1)$.

Последнее утверждение в отличие от соответствующего утверждения в предыдущем параграфе не является прямым следствием общей теоремы параграфа 30.3, а требует самостоятельного доказательства. Действительно, эту теорему можно обобщить на случай, когда рассматриваются s независимых выборок по n_1, \dots, n_s элементов, разбитых на одинаковое число r групп, и с помощью видоизмененного метода минимума χ^2 , примененного к выражению

$$\chi^2 = \sum_{i,j} \frac{(v_{ij} - n_j p_i)^2}{n_j p_i}, \text{ определяется некоторое число } t \text{ неизвестных параметров.}$$

Непосредственное обобщение доказательства теоремы, приведенной в параграфе 30.3, показывает, что χ^2 имеет обычное предельное распределение с $(r - 1)s - t$ степенями свободы. В рассматриваемом выше случае мы имеем дело с гипотезой о том, что все s выборок извлечены из одной и той же совокупности, без дальнейшего уточнения вида распределения этой совокупности, так что параметрами являются сами вероятности p_i . Благодаря соотношению $\sum_i p_i = 1$ имеется лишь $t = r - 1$ параметров, так что мы получаем $(r - 1)(s - 1)$ степеней свободы.

Обобщая теорему параграфа 30.3, можно использовать χ^2 также для проверки гипотезы о том, что заданные s выборок извлечены из одной и той же совокупности *заданного типа*, например, имеющей распределение Пуассона, нормальное распределение и т. п. В таком случае применение видоизмененного метода минимума χ^2 показывает, что параметры распределения отыскиваются так же, как и в случае одной единственной выборки с групповыми частотами, равными суммам строк v_i , в данной таблице. Доказательство этого утверждения предоставляем читателю в качестве упражнения.

В частном случае $r = 2$ таблицу можно записать следующим образом:

v_1	v_2	\dots	v_s	$\sum_j v_j$
$n_1 - v_1$	$n_2 - v_2$	\dots	$n_s - v_s$	$n - \sum_j v_j$
n_1	n_2	\dots	n_s	n

Здесь мы имеем s последовательностей наблюдений, в каждом из которых некоторое событие E осуществляется соответственно y_1, \dots, y_s раз, и спрашивается, есть ли основание предполагать, что событие E во всех этих наблюдениях имеет одну и ту же постоянную, хотя и неизвестную, вероятность p . Оценкой для p здесь должна служить частота события E во всей совокупности данных: $p^* = 1 - q^* = \frac{1}{n} \sum_j y_j$, и по формуле (30.5.1)*) получаем

$$(30.6.1) \quad \chi^2 = \sum_j \frac{(y_j - n_j p^*)^2}{n_j p^* q^*} = \frac{1}{p^* q^*} \sum_j \frac{y_j^2}{n_j} - n \frac{p^*}{q^*}$$

с $s-1$ степенями свободы. Величина $Q = \sqrt{\frac{n-1}{n(s-1)} \chi^2}$ совпадает с введенным Лексисом *коэффициентом расхождения*. В соответствии с (30.5.2) имеем $E(Q^2) = 1$ (см., например, Чупров [227а] и Крамер [10], стр. 105—123).

В таблице 30.6.1 приведено число детей, родившихся в Швеции в течение $s=12$ месяцев 1935 г. Оценкой для вероятности рождения мальчика является $p^* = \frac{45682}{88273} = 0,5175082$. По формуле (30.6.1) получаем $\chi^2 = 14,986$ с 11 степенями свободы; этому соответствует $P = 0,18$, и можно считать, что данные совместимы с гипотезой о постоянной вероятности.

Таблица 30.6.1

**Распределение пола детей, родившихся в Швеции
в 1935 г.**

Месяцы Пол ребенка \	1	2	3	4	5	6	7	8	9	10	11	12	Итого
Мальчики	3743	3550	4017	4173	4117	3944	3964	3797	3712	3512	3392	3761	45682
Девочки	3537	3407	3866	3711	3775	3665	3621	3596	3491	3391	3160	3371	42591
Общее число	7280	6957	7883	7884	7892	7609	7585	7393	7203	6903	6552	7132	88273

*). См. также (30.2.1).

В заключение рассмотрим случай $s = 2$. В этом случае мы имеем две независимые выборки и хотим узнать, извлечены ли они из одной и той же совокупности. Таблицу можно записать в виде

μ_1	ν_1	$\mu_1 + \nu_1$
μ_2	ν_2	$\mu_2 + \nu_2$
\cdot	\cdot	\cdot
μ_r	ν_r	$\mu_r + \nu_r$
m	n	$m + n$

Здесь имеется $r - 1$ степеней свободы, и (30.5.1) дает (см. К. Пирсон [186] и Р. Фишер [91]):

$$(30.6.2) \quad \chi^2 = mn \sum_i \frac{1}{\mu_i + \nu_i} \left(\frac{\mu_i}{m} - \frac{\nu_i}{n} \right)^2.$$

ТАБЛИЦА 30.6.2
Распределение доходов по шведской переписи
1930 г.

Доходы (за единицу принято 1000 крон)	Все рабочие и служащие в про- мышленности				Заводские мастера		
	взрослые группы		$\hat{\omega}_I$	взрослые группы		$\hat{\omega}_I$	
	$40-50$ μ_i	$50-60$ ν_i		$40-50$ μ_i	$50-60$ ν_i		
0 — 1	7 831	7 558	0,50886997	71	54	0,56800000	
1 — 2	26 740	20 685	0,56383764	430	324	0,57029178	
2 — 3	35 572	24 186	0,59526758	1 072	894	0,54526958	
3 — 4	20 009	12 280	0,61968472	1 609	1 202	0,57239417	
4 — 6	11 527	6 776	0,62978747	1 178	903	0,56607400	
> 6	6 919	4 222	0,62103940	158	112	0,58518519	
	108 598	75 707	0,58922981	4 518	3 489	0,56425328	
	$\chi^2 = 840,62$ (5 степ. своб.) $P < 0,001$				$\chi^2 = 4,27$ (5 степ. своб.) $P = 0,51$		

Если положить $\hat{\omega}_I = \frac{\mu_I}{\mu_I + \nu_I}$ и $\hat{\omega} = \frac{m}{m + n}$, то выражение (30.6.2) сводится к следующему выражению, полученному Снедекором ([35], стр. 173) и часто удобному для практических подсчетов:

$$(30.6.3) \quad \chi^2 = \frac{(m + n)^2}{mn} \left(\sum_i \frac{\mu_i^2}{\mu_i + \nu_i} - \frac{m^2}{m + n} \right) = \frac{1}{\hat{\omega}(1 - \hat{\omega})} \left(\sum_i \mu_i \hat{\omega}_I - m \hat{\omega} \right).$$

В таблице 30.6.2 приведены некоторые распределения доходов по шведской переписи 1930 года. Если сравнивать распределения доходов у возрастных групп 40—50 и 50—60 лет всех промышленных рабочих и служащих, то (30.6.2) даст $\chi^2 = 840,62$ с 5 степенями свободы, что обнаруживает высоко значимую разность между распределениями. Очевидно, что в этом случае числа \hat{b}_i имеют тенденцию возрастать при возрастании доходов. Однако, если рассмотреть более однородную группу заводских мастеров, то этой тенденции не обнаруживается, и сравнение распределений доходов у двух возрастных групп дает $\chi^2 = 4,27$ и $P = 0,51$, так что последние две выборки можно считать извлеченными из одной и той же совокупности.

30.7. Критерий для процента смертности. Предположим, что при изучении смертности получены следующие данные для двух различных классов лиц (проживающих в различных районах, имеющих различные профессии и т. п.):

Возрастная группа	Класс А		Класс В	
	количество подвергнувшихся риску	количество умерших	количество подвергнувшихся риску	количество умерших
1	n_1	d_1	n'_1	d'_1
2	n_2	d_2	n'_2	d'_2
.
.
.
r	n_r	d_r	n'_r	d'_r

Требуется проверить, отличаются ли значимо друг от друга полученные из этих данных последовательности чисел $\frac{d_i}{n_i}$ и $\frac{d'_i}{n'_i}$, характеризующих смертность. Для каждой возрастной группы можно образовать таблицу следующего типа:

	Класс А	Класс В
Количество умерших	d_1	d'_1
Количество оставшихся в живых	$n_i - d_i$	$n'_i - d'_i$

Вычислим по формуле (30.6.2.) соответствующее значение χ^2 с одной степенью свободы:

$$\chi_i^2 = \frac{n_i n'_i (n_i + n'_i)}{(d_i + d'_i) (n_i + n'_i - d_i - d'_i)} \left(\frac{d_i}{n_i} - \frac{d'_i}{n'_i} \right)^2$$

Различные χ_i^2 независимы. Таким образом, если предположить, что смертность в обеих совокупностях одинакова, то сумма $\chi^2 = \sum_i \chi_i^2$ будет иметь обычное предельное распределение с r степенями свободы и будет служить критерием для высказанной гипотезы (см. К. Пирсон и Точер [187], Р. Фишер [91], Валунд [228]).

В таблице 30.7.1 приведены некоторые данные из наблюдений Г. Берга [61] над туберкулезом. Требуется установить, существует ли какая-либо значимая разница в смертности мужчин и женщин в течение первого года после установления заболевания. Суммарное χ^2 равняется 22,2 и имеет 10 степеней свободы, чему соответствует $P = 0,014$, так что, по принятой нами терминологии (см. параграф 30.2), отклонение „почти значимо“. Из значений χ_i^2 , приведенных в последнем

ТАБЛИЦА 30.7.1

Смертность среди больных, страдающих легочным туберкулезом в открытой форме, в течение первого года после заболевания

Возрастная группа	Мужчины			Женщины			χ^2
	количество заболевших n_i	количество умерших d_i	смертность в %	количество заболевших n'_i	количество умерших d'_i	смертность в %	
15—19	400	156	38,4	500	174	34,8	1,25
20—24	695	204	29,4	816	246	30,	0,11
25—29	585	169	28,9	619	184	29,7	0,09
30—34	454	128	28,2	433	150	34,6	4,22
35—39	274	82	29,9	257	92	35,8	2,10
40—44	221	68	30,8	194	83	42,8	6,43
45—49	153	41	26,8	94	39	41,5	5,75
50—54	110	34	30,9	58	20	34,5	0,23
55—59	69	36	52,2	29	13	44,8	0,45
≥60	89	43	48,3	47	28	59,6	1,57
Всего	3 056	961	—	3 047	1 029	—	22,20

столбце таблицы, видно, что основную часть χ^2 составляют значения, соответствующие возрастам 30—50 лет, причём смертность среди женщин значительно выше, чем среди мужчин.

30.8. Дальнейшие критерии согласия. Как уже указывалось в параграфе 30.1, критерий χ^2 всегда полезно дополнять другими критериями. Во многих случаях простой просмотр знаков и абсолютных величин разностей между наблюденными и ожидаемыми частотами обнаруживает систематические отклонения от проверяемой гипотезы, даже и тогда, когда χ^2 имеет незначимую величину.

Если критерий χ^2 применяется к сравнительно небольшой выборке, то приходится прибегать к группировке с большими интервалами разбиения и, следовательно, жертвовать большой долей сведений, содержащихся в выборке. В таких случаях желательно применить критерий, базирующийся на индивидуальных выборочных значениях. Рассмотрим вкратце один критерий такого типа.

Пусть требуется проверить гипотезу о том, что выборка в n наблюденных значений x_1, \dots, x_n извлечена из совокупности с заданной функцией распределения $F(x)$. Функцией распределения выборки (см. параграф 25.3) является $F^*(x) = \frac{v}{n}$, где v — число выборочных значений, не превышающих x . Поскольку при любом фиксированном x функция F^* сходится по вероятности к F (см. параграф 25.3), можно принять за меру отклонения нашей выборки от гипотезы интеграл

$$\int_{-\infty}^{\infty} [F^*(x) - F(x)]^2 dK(x),$$

где функция $K(x)$ может быть выбрана более или менее произвольно. Критерии, базирующиеся на мерах такого рода, впервые были введены Крамером [10], [70] и Мизесом [27]. Следуя Смирнову [215], положим здесь $K(x) = F(x)$ и получим, таким образом, интеграл

$$\omega^2 = \int_{-\infty}^{\infty} [F^*(x) - F(x)]^2 dF(x).$$

Если выборочные значения x_1, \dots, x_n расположены в порядке возрастания, то при любой непрерывной $F(x)$ будем иметь

$$\omega^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_1^n \left[F(x_v) - \frac{2v-1}{2n} \right]^2,$$

Если известны индивидуальные выборочные значения, то можно легко вычислить точное значение ω^2 . Если же имеется в распоряжении лишь группированная выборка, то можно найти приближенное значение, например, при обычном предположении, что x , расположены в средних точках интервалов разбиения.

Как было указано в параграфе 25.5, $F^*(x)$ есть частота события с вероятностью $F(x)$ в последовательности n испытаний. Поэтому $E(F^* - F)^2 = \frac{F(1-F)}{n}$. С помощью этого замечания можно получить среднее значение и дисперсию величины ω^2 , которые оказываются не зависящими от $F(x)$:

$$E(\omega^2) = \frac{1}{6n}, \quad D^2(\omega^2) = \frac{4n-3}{180n^3}.$$

Сравнивая значение ω^2 , найденное по фактической выборке, со средним значением и дисперсией, вычисленными по этим формулам, получим критерий для нашей гипотезы. Выборочное распределение величины ω^2 , которое также не зависит от $F(x)$, было подвергнуто дальнейшему исследованию Смирновым [215]; он показал, что при $n \rightarrow \infty$ величина $n\omega^2$ имеет некоторое предельное распределение, не являющееся нормальным и не зависящее от n (см. пример с $n\omega^2$ в параграфе 29.13). Было бы желательно обобщить эту теорию на случай, когда гипотетическая $F(x)$ полностью не задана, а содержит некоторые параметры, подлежащие оценке по выборке.

Другие важные критерии согласия были предложены, например, Нейманом [164] и Е. Пирсоном [191].

ГЛАВА III

КРИТЕРИИ ЗНАЧИМОСТИ ДЛЯ ПАРАМЕТРОВ

31.1. Критерии, основанные на стандартных ошибках. В приложениях часто требуется использовать множество выборочных значений для проверки гипотезы о том, что некоторый параметр соответствующей совокупности, например, среднее значение, коэффициент корреляции и т. п., имеет наперед заданное значение. В других случаях мы располагаем несколькими независимыми выборками, и требуется установить, являются ли разности между наблюденными значениями некоторой выборочной характеристики значимыми, т. е. указывающими

на действительную разность между соответствующими параметрами совокупности.

В главе 28 мы видели, что некоторые важные классы выборочных характеристик в больших выборках асимптотически нормальны, со средними значениями и дисперсиями, определяемыми по параметрам совокупности. Поэтому можно получить критерий значимости для гипотез вышеуказанного типа, следуя общим правилам, изложенным в параграфе 26.2 (см. также параграф 35.1).

Так, например, если выборка в n значений x_1, \dots, x_n извлечена из некоторой (не обязательно нормальной) совокупности со средним значением m и стандартным отклонением σ , то из параграфов 17.4 и 28.2 известно, что среднее значение выборочных значений асимптотически нормально $(m, \frac{\sigma}{\sqrt{n}})$. Предположим на минуту, что σ известно и проверяется гипотеза, что m имеет некоторое фиксированное значение m_0 . Если эта гипотеза верна, то \bar{x} асимптотически нормально $(m_0, \frac{\sigma}{\sqrt{n}})$. Обозначив через λ_p p -процентное значение нормального отклонения (см. параграф 17.2), можно получить при большом n отклонение $|\bar{x} - m_0|$, превышающее $\frac{\lambda_p \sigma}{\sqrt{n}}$, с вероятностью, приблизительно равной $p^0/0$. Работая с p -процентным уровнем значимости, мы, таким образом, должны забраковать нашу гипотезу, если $|\bar{x} - m_0|$ превышает этот предел, тогда как меньшее отклонение следует считать совместимым с гипотезой.

На практике σ обычно неизвестно. Однако, согласно параграфу 27.3, стандартное отклонение выборки s сходится по вероятности к σ при $n \rightarrow \infty$. Поэтому при большом n вероятность того, что разность $|s - \sigma|$ превышает некоторое малое число, сама очень мала. Таким образом, в нашем критерии можно просто заменить σ на s и действовать так, как если бы проверялась гипотеза о том, что \bar{x} нормально $(m_0, \frac{s}{\sqrt{n}})$, где s — известное нам значение, вычисленное по выборке.

Если наблюденное отклонение $|\bar{x} - m_0|$ превысит $\frac{\lambda_p s}{\sqrt{n}}$, то с точки зрения p -процентного уровня значимости мы должны будем забраковать гипотезу $m = m_0$, тогда как меньшее отклонение следует считать совместимым с этой гипотезой.

Такой же метод можно применять и в более общих случаях. Рассмотрим какую-нибудь выборочную характеристику z , распределение которой в больших выборках асимптотически нормально. Заменим каждый неизвестный параметр совокупности в выражении для дисперсии асимптотически нормального распределения величины z соответствующей известной выборочной характеристикой и сохраним лишь главный член этого выражения при большом n . Полученное при этом выражение $d(z)$ будет называться *стандартной ошибкой* величины z в больших выборках. Если требуется проверить гипотезу о том, что $E(z)$ имеет некоторое фиксированное значение z_0 , то будем считать, что z распределена нормально с известным стандартным отклонением $d(z)$. Если отклонение $|z - z_0|$ превысит $\lambda_p d(z)$, то гипотеза $E(z) = z_0$ будет забракована с точки зрения p -процентного уровня значимости; в противном случае гипотеза будет подтверждена.

Все выражения для стандартных отклонений выборочных характеристик и асимптотически нормальных распределений этих характеристик, выведенные в главах 27—28, могут быть преобразованы в стандартные ошибки изложенным выше способом. Так, например, в силу формул (27.2.1), (27.4.2) и (27.7.2), стандартными ошибками выборочного среднего значения \bar{x} , выборочной дисперсии $s^2 = m_2$ и стандартного отклонения выборки $s = \sqrt{m_3}$ являются

$$d(\bar{x}) = \frac{s}{\sqrt{n}}, \quad d(s^2) = \frac{\sqrt{m_4 - s^4}}{\sqrt{n}}, \quad d(s) = \frac{\sqrt{m_4 - s^4}}{2s\sqrt{n}}.$$

Если предполагается, что совокупность нормальна, то для стандартных ошибок можно получить более простые выражения. Так, например, в силу параграфа 28.5, стандартной ошибкой медианы нормальной выборки будет

$$s \sqrt{\frac{\pi}{2n}} = 1,2533 \frac{s}{\sqrt{n}}.$$

Если значение выборочной характеристики z вычислено, то на практике обычно указывают степень точности, приписывая к вычисленному значению $\pm d(z)$. Так, например, выборочное среднее значение записывается в виде $\bar{x} \pm \frac{s}{\sqrt{n}}$. Для частоты события с постоянной вероятностью p в последовательности n испытаний имеем, согласно формуле (16.2.2), $E\left(\frac{v}{n}\right) = p$ и $D\left(\frac{v}{n}\right) = \sqrt{\frac{pq}{n}}$, так что стандартной ошиб-

кой является $\sqrt{\frac{v(n-v)}{n^3}}$, и, следовательно, частота записывается в виде $\frac{v}{n} \pm \sqrt{\frac{v(n-v)}{n^3}}$. Соответствующая величина в процентах $\tilde{w} = 100 \frac{v}{n}$ записывается в виде $\tilde{w} \pm \sqrt{\frac{\tilde{w}(100-\tilde{w})}{n}}$.

Если заданы две независимые выборки, то разность между их средними значениями или какими-либо другими характеристиками можно исследовать с помощью стандартных ошибок. Если средние значения \bar{x} и \bar{y} рассматриваются как нормальные, соответственно $(m_1, \frac{s_1}{\sqrt{n_1}})$ и $(m_2, \frac{s_2}{\sqrt{n_2}})$, то разность $\bar{x} - \bar{y}$ нормальна $(m_1 - m_2, \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}})$ и всякую гипотезу о значении разности $\bar{x} - \bar{y}$ можно проверить указанным выше способом. В частности, гипотеза $m_1 = m_2$ будет забракована с точки зрения p -процентного уровня значимости, если $|\bar{x} - \bar{y}| > \lambda_p \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, и будет подтверждена, если $|\bar{x} - \bar{y}| \leq \lambda_p \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

Все указанные выше методы действительны лишь при условии, что наши выборки „велики“. При этом условии предполагалось: а) что выборочные распределения наших характеристик нормальны, б) что некоторые характеристики совокупности можно заменить соответствующими значениями, вычисленными по выборке. На практике часто довольно затруднительно выяснить, достаточно ли велики наши выборки, чтобы указанные два предположения оправдывались. Однако можно все же указать некоторые практические правила. Если мы имеем дело со средними значениями, то уже при $n > 30$ получаются обычно хорошие результаты. Для дисперсий, медиан, коэффициентов асимметрии и эксцесса, коэффициентов корреляции в окрестности значения $p=0$ рекомендуется требовать, чтобы n было, по крайней мере, около 100. Для коэффициентов корреляции, значительно отличающихся от нуля, даже выборки в 300 индивидуумов не всегда дают удовлетворительное приближение.

Но даже в случаях, когда n меньше, чем требуется по указанным только что правилам, или когда выборочное распределение не стремится к нормальному, часто все же с помощью стандартных ошибок можно

получить некоторые сведения, хотя при этом всегда рекомендуется особенная осторожность в выводах. Если выборочное распределение значительно отличается от нормального, то таблицы нормального распределения не дают удовлетворительного приближения для вероятности отклонения, превышающего фиксированное значение. В этих случаях всегда можно использовать неравенство (15.7.2), которое дает для любого распределения верхний предел $\frac{1}{k^2}$ вероятности того, что отклонение от среднего значения превышает умноженное на k стандартное отклонение. Однако в большинстве встречающихся на практике случаев этот верхний предел слишком велик. Так, например, из {15.7.4} следует, что этот предел можно существенно понизить для всех унимодальных и умеренно асимметричных распределений. Это же следует из неравенства, приведенного в упражнении 6 на стр. 284, если только коэффициент γ_2 для распределения не слишком велик. Если имеются основания предполагать, что выборочное распределение именно такого типа, то отклонение, превышающее установленное стандартное отклонение, можно, как правило, считать значимым. Если n недостаточно велико, то рекомендуется использовать полные выражения для стандартных отклонений (конечно, если эти выражения известны), а не только их главные члены. Далее, при этом следует использовать несмещенные оценки (см. параграф 27.6) для значений параметров совокупности; например, для стандартной ошибки среднего значения вместо $\sqrt{\frac{s}{n}}$ брать $\sqrt{\frac{s}{n-1}}$. Однако лучше всего использовать в этих случаях критерии, основанные на точных распределениях. Такие критерии будут рассмотрены в следующем параграфе.

31.2. Критерии, основанные на точных распределениях. Если точные выборочные распределения интересующих нас характеристик известны, то приближенные методы оценок, изложенные в предыдущем параграфе, можно заменить точными методами. Как было замечено в параграфе 29.1, такое положение возникает главным образом в тех случаях, когда производится выбор из нормальных совокупностей.

Предположим, например, что задана выборка в n значений из нормальной совокупности с неизвестными параметрами m и s и требуется проверить гипотезу о том, что m имеет некоторое заранее заданное значение. Если эта гипотеза верна, то выборочное среднее

значение \bar{x} будет в точности нормально $(m, \frac{\sigma}{\sqrt{n}})$ и нормированная величина $\sqrt{n} \frac{\bar{x} - m}{\sigma}$ будет нормальна $(0,1)$. Приближенный метод предыдущего параграфа заключается в замене неизвестного параметра σ его оценкой, вычисленной по выборке, — при малом n предпочтительно оценкой $\sqrt{\frac{n}{n-1}} s$. Полученное при этом выражение $t = \sqrt{n-1} \frac{\bar{x} - m}{s}$ рассматривается как нормальное $(0,1)$. Это выражение t совпадает со *студентовским отношением* (см. параграф 29.4), которое, как мы видели, имеет распределение Стьюдента с $n-1$ степенями свободы. Если t_p обозначает p -процентное значение величины t с $n-1$ степенями свободы (см. параграф 18.2), то вероятность получить отклонение t , такое, что $|t| > t_p$, будет в точности равна $p\%$. Таким образом, гипотетическое значение m следует забраковать с точки зрения p -процентного уровня значимости, если $|t| > t_p$; в противном же случае это значение m будет принято.

При $n \rightarrow \infty$ t -распределение приближается к нормальному (см. параграф 20.2); это предельное распределение приведено в последней строке таблицы IV. Из этой таблицы видно, что нормальное распределение достаточно хорошо аппроксимирует t -распределение уже при $n \geq 30$. Однако при малом n вероятность большого отклонения от среднего значения у t -распределения существенно больше (см. фиг. 20, стр. 266).

Если мы хотим проверить, значимо ли отличаются друг от друга средние значения \bar{x} и \bar{y} двух независимых нормальных выборок, то можно принять „нулевую гипотезу“, согласно которой обе выборки извлечены из одной и той же нормальной совокупности. В параграфе 29.4 было показано, что если эта гипотеза верна, то величина

$$(31.2.1) \quad u = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \cdot \frac{\bar{x} - \bar{y}}{\sqrt{\frac{n_1 s_1^2}{n_1 + n_2} + \frac{n_2 s_2^2}{n_1 + n_2}}}$$

имеет t -распределение с $n_1 + n_2 - 2$ степенями свободы. Если заданы средние значения и дисперсии выборок, то u можно вычислить непосредственно по выборкам. Если при этом $|u|$ превысит p -процентное значение величины t с $n_1 + n_2 - 2$ степенями свободы, то наши данные будут значимо отклоняться от нулевой гипотезы, с точки зрения p -процентного уровня значимости. Если имеются основания предполагать, что совокупности в действительности нормальны и что стандартные

отклонения этих совокупностей σ_1 и σ_2 равны друг другу, то отказ от нулевой гипотезы будет значить, что средние значения этих совокупностей m_1 и m_2 различны (см. параграф 35.5).

Очевидно, что совершенно так же можно действовать и в случае любой функции z от выборочных значений, если только точное распределение этой величины z известно. Мы выдвигаем вероятностную гипотезу, что наблюденное значение величины z с большой вероятностью принадлежит окрестности некоторого фиксированного значения z_0 . Если эта гипотеза H верна и z имеет некоторое известное распределение, то по этому распределению можно найти p -процентное значение отклонения $|z - z_0|$, т. е. величину h_p , такую, что вероятность отклонения $|z - z_0| > h_p$ в точности равна $p\%$. Работая с p -процентным уровнем значимости и придерживаясь правил параграфа 26.2, мы должны будем забраковать гипотезу H , если по фактической выборке получится отклонение $|z - z_0|$, превышающее h ; меньшее отклонение мы должны будем считать совместимым с гипотезой H (см. параграф 35.1).

Если мы имеем дело с выборками, извлеченными из нормальных совокупностей, то критерии значимости для различных параметров могут быть основаны на точных выборочных распределениях, выведенных в главе 29. На практике величины, встречающиеся в различных областях статистики, очень часто могут считаться хотя бы приближенно-нормальными (см. параграф 17.8). В таких случаях критерии, выведенные для нормального случая, обычно дают удовлетворительное приближение. Действительно, мы уже видели, что даже значительные отклонения генеральной совокупности от нормального распределения не очень сильно влияют на выборочные распределения различных важных характеристик. По этому вопросу мы отсылаем читателя к некоторым экспериментальным исследованиям Е. Пирсона [190] и к диссертации Квензеля [200] о некоторых выборочных распределениях, связанных с совокупностями Шарлье типа А. Продолжение этих исследований было бы очень желательным.

31.3. Примеры. Теперь покажем некоторые применения критериев, рассмотренных в двух предыдущих параграфах. Сначала рассмотрим случаи, в которых выборки настолько велики, что совершенно законно использовать критерии, основанные на стандартных ошибках. Затем перейдем к различным случаям выборок малого или умеренного объема.

При этом для обозначения различных степеней значимости мы будем пользоваться терминологией, введенной в параграфе 30.2.

Пример 1. В таблице 31.3.1 приведено распределение 928 570 детей, родившихся в Норвегии в течение 1871—1900 г. по полу и по

Таблица 31.3.1
Количество детей, родившихся в Норвегии в 1871—1900 гг.

Возраст отца <i>x</i>	Возраст матери <i>y</i>							Сумма
	<20	20—25	25—30	30—35	35—40	40—45	≥45	
<i>Мальчики</i>								
<20	377	974	555	187	93	25	6	2 217
20—25	2 173	18 043	11 173	3 448	1 022	258	30	36 147
25—30	1 814	26 956	43 082	16 760	4 564	973	123	94 272
30—35	700	14 252	38 505	41 208	14 475	3 243	287	112 670
35—40	238	4 738	17 914	32 240	31 573	8 426	836	95 965
40—45	103	1 791	6 586	16 214	24 770	18 079	2 171	69 714
45—50	47	695	2 593	5 952	12 453	13 170	4 006	38 916
50—55	21	311	995	2 503	4 492	6 322	2 574	17 218
55—60	5	133	412	925	1 790	2 141	1 086	6 492
60—65	10	57	190	408	736	822	348	2 571
65—70	6	25	68	173	266	283	131	952
≥75	2	12	46	59	119	113	48	399
Всего	5 495	67 987	122 119	120 077	96 353	53 855	11 646	477 533
<i>Девочки</i>								
<20	319	861	504	206	91	22	3	2 006
20—25	2 133	16 990	10 643	3 193	979	242	45	34 225
25—30	1 793	25 147	40 817	15 637	4 305	943	96	88 738
30—35	707	13 254	36 745	38 619	13 669	3 018	292	106 304
35—40	236	4 676	17 165	30 453	29 858	7 883	772	91 043
40—45	101	1 670	6 278	15 323	23 803	16 933	1 941	66 099
45—50	33	640	2 384	5 603	11 764	12 336	3 823	36 588
50—55	16	284	964	2 469	4 221	5 815	2 480	16 249
55—60	12	120	406	874	1 726	2 000	1 079	6 217
60—65	6	54	171	381	591	750	325	2 278
65—70	3	29	87	154	277	247	114	911
≥70	1	18	30	67	108	115	40	379
Всего	5 365	63 743	116 194	112 979	91 392	50 354	11 010	451 037

возрасту родителей (данные заимствованы у Виксела [231]). Требуется, используя эти данные, исследовать зависимость, если таковая имеется, пола потомства от возраста родителей.

В качестве первого подхода к решению этой проблемы вычислим процент случаев рождения мальчиков и соответствующую стандартную ошибку для четырех больших возрастных групп, указанных в таблице 31.3.2.

Между числами в этой таблице нет значимых разностей. Наибольшая разность — между числами 51 589 и 51 111 — составляет $0,478 \pm 0,222$. Здесь наблюдаемая разность в 2,15 раза больше соответствующей стандартной ошибки, т. е., по нашей условной терминологии, эта разность лишь „почти значима“. Тем не менее, таблица может внушить предположение, что избыток мальчиков имеет тенденцию увеличиваться при уменьшении разности возрастов родителей $x - y$.

Чтобы исследовать вопрос более тщательно, рассмотрим возрасты x и y родителей ребенка как наблюденные значения двумерной случайной величины. Тогда таблица 31.3.1 будет давать совместные распределения величин x и y в двух выборках по $n_1 = 477\,533$ и $n_2 = 451\,037$ значений (n_1 — количество родившихся мальчиков, n_2 — количество родившихся девочек). Если соотношение полов среди новорожденных изменяется в зависимости от возрастов родителей, то распределения (x, y) должны быть различны для мальчиков и девочек, так что две выборки нельзя будет считать извлеченными из одной и той же совокупности.

ТАБЛИЦА 31.3.2
Процент рождений мальчиков

Возраст отца x	Возраст матери y	
	< 30	> 30
< 35	$51,409 \pm 0,090$	$51,589 \pm 0,122$
> 35	$51,111 \pm 0,186$	$51,430 \pm 0,031$

В таблице 31.3.3 приведены групповые моменты для обеих выборок и исправленные моменты с учетом поправок (27.9.4) и (27.9.6). Прежде всего мы замечаем, что распределение значимо отличается от нормального распределения. Рассмотрим, например, частное

Т а б л и ц а 31.3.3
Выборочные моменты для таблицы 31.3.1
(За единицу принята длина интервала разбиения 5 лет.)

Центральные моменты	Мальчики		Девочки	
	центр. моменты групповые	центр. моменты исправленные	центр. моменты групповые	центр. моменты исправленные
m_{20}	2,9127	2,8294	2,9036	2,8203
m_{11}	1,4140	1,4140	1,4085	1,4085
m_{02}	1,7956	1,7123	1,7929	1,7096
m_{80}	3,0699	3,0699	3,0391	3,0391
m_{03}	0,4588	0,4588	0,4588	0,4588
m_{40}	28,6579	27,2307	28,4535	27,0309
m_{81}	10,3527	9,9992	10,2509	9,8988
m_{22}	7,7285	7,3431	7,6970	7,3126
m_{18}	5,8110	5,4575	5,8020	5,4499
m_{04}	7,5250	6,6564	7,5260	6,6587

распределение возраста x отцов у мальчиков. По гипотезе о том, что это распределение нормально, мы получаем с помощью исправленных моментов $g_1 = 0,6450 \pm 0,0035$ и $g_2 = 0,4015 \pm 0,0071$, причем стандартные ошибки вычислены по формуле (27.7.9). Отклонение от нуля в обоих случаях высоко значимо, так что гипотеза о нормальности явно несостоятельна *).

Т а б л и ц а 31.3.4
Выборочные характеристики для таблицы 31.3.1
(За единицу принят один год.)

Характеристика	Для мальчиков	Для девочек	Разность $\times 10^3$
\bar{x}	$35,699 \pm 0,0122$	$35,703 \pm 0,0125$	$+4 \pm 17,5$
\bar{y}	$32,128 \pm 0,0095$	$32,116 \pm 0,0097$	$-12 \pm 13,6$
$\bar{x} - \bar{y}$	$3,571 \pm 0,0095$	$3,587 \pm 0,0097$	$+16 \pm 13,6$
s_1	$8,410 \pm 0,0094$	$8,397 \pm 0,0097$	$-13 \pm 13,5$
s_2	$6,543 \pm 0,0053$	$6,533 \pm 0,0055$	$-5 \pm 7,6$
r	$0,6424 \pm 0,00097$	$0,6414 \pm 0,00101$	$-1,0 \pm 1,40$

*) По Викселю, это распределение приблизительно логарифмически-нормально (см. параграф 17.5).

В таблице 31.3.4 приведены значения некоторых важнейших выборочных характеристик для количества рождений мальчиков и девочек, а также разности между соответствующими характеристиками этих двух выборок. Стандартные ошибки вычислены, в соответствии с правилами параграфа 31.1, по общим формулам (27.2.1), (27.7.2) и (27.8.1); таким образом, более простые выражения (27.8.2) и (29.3.3), соответствующие случаю нормальной совокупности, не были применены для вычисления стандартных ошибок. Для разности $\bar{x} - \bar{y}$ получаем

$$D^2(\bar{x} - \bar{y}) = \frac{s_1^2 - 2rs_1s_2 + s_2^2}{n},$$

и, следовательно, квадрат стандартной ошибки имеет вид

$$d^2(\bar{x} - \bar{y}) = \frac{s_1^2 - 2rs_1s_2 + s_2^2}{n}.$$

Из таблицы 31.3.4 видно, что между характеристиками нет значимых разностей. В частности, среднее значение разности возрастов родителей $x - y$ для девочек незначимо превышает соответствующее значение для мальчиков, так что предположение, которое было вызвано рассмотрением таблицы 31.3.2, не подтверждается дальнейшим анализом.

Наконец, для проверки того, можно ли считать две выборки в таблице 31.3.1 извлеченными из одной и той же совокупности, можно непосредственно применить критерий χ^2 . Действительно, в каждой из двух выборок имеется $12 \cdot 7 = 84$ группы, так что всю таблицу 31.3.1 можно считать таблицей с 84 столбцами и 2 строками, аналогичной таблицам, рассмотренным в параграфе 30.6, и однородность данных в этой таблице можно проверять с помощью метода χ^2 , используя для вычисления значения χ^2 формулы (30.6.2) или (30.6.3). Объединив все группы с отцами старше 60 лет и с матерями старше 40 лет, имеем таблицу с 60 столбцами и 2 строками, для которой получим $\chi^2 = 51,97$ с $(60 - 1)(2 - 1) = 59$ степенями свободы. Согласно фишеровскому методу приближения (см. параграф 20.2), $\sqrt{2\chi^2} = 10,20$ будет наблюденным значением нормальной величины со средним значением $\sqrt{117} = 10,82$ и с единичным стандартным отклонением. Вероятность получить значение χ^2 по крайней мере столь же большое, как и фактически наблюденное значение, по таблице I (стр. 603) равняется приблизительно $1 - \Phi(10,20 - 10,82) = 0,73$, так что совпадение очень хорошее, и наши

данные оказываются совместимыми с гипотезой о том, что обе выборки извлечены из одной и той же совокупности.

Таким образом, анализ данных, приведенных в таблице 31.3.1, не обнаружил никакого значимого влияния возраста родителей на пол детей.

Пример 2. Распределения различных измерений тела в совокупности людей одной и той же расы обычно хорошо совпадают с нормальной кривой, а малые отклонения достаточно хорошо описываются первыми членами рядов Шарлье или Эджвортса (см., например, (17.7.5)). По этому вопросу мы отсылаем читателя к статье Крамера [70], в которой приведены подробные примеры.

В таких случаях стандартные ошибки выборочных характеристик можно вычислять по упрощенным формулам, справедливым для случая нормальной генеральной совокупности. Так, например, в силу формулы (29.3.3), стандартную ошибку для s можно принять равной $\frac{s}{\sqrt{2n}}$, стандартную ошибку коэффициента изменчивости V можно вычислять по формуле (27.7.11) и т. д. Хульктранц [128] вычислил выборочные характеристики для роста шведских призывников в среднем возрасте 19 лет 8 месяцев по измерениям в 1915/16 и 1924/25 гг. Мы приводим эти данные в таблице 31.3.5. Из этой таблицы обнаруживается чрезвычайно значимое увеличение среднего значения и медианы за промежуток между измерениями в 9 лет. С другой стороны, стандартное

ТАБЛИЦА 31.3.5

Выборочные характеристики для роста шведских призывников

Характеристики	1915/16 гг.	1924/25 гг.	Разность $\times 10^4$
n	80 084	89 337	
Среднее \bar{x} (см)	$171,80 \pm 0,022$	$172,58 \pm 0,020$	$+78 \pm 3,0$
Медиана (см)	$171,81 \pm 0,027$	$172,55 \pm 0,025$	$+74 \pm 3,7$
Стандартное отклонение (см)	$6,15 \pm 0,015$	$6,04 \pm 0,014$	$-11 \pm 2,1$
Семи-интерквартильная широта (см)	$4,05 \pm 0,017$	$4,02 \pm 0,016$	$-3 \pm 2,3$
$100 V = 100 \frac{s}{x}$	$3,58 \pm 0,0090$	$3,50 \pm 0,0083$	$-8 \pm 1,2$

отклонение и коэффициент изменчивости обнаруживают высоко значимое уменьшение, а уменьшение семи-интерквартильной широты незначимо.

Эти результаты хорошо согласуются с имеющимися в нашем распоряжении дальнейшими данными измерений роста шведских призывников. В течение последних 100 лет средний рост призывников неуклонно увеличивался, а стандартное отклонение уменьшалось.

Согласно таблице 31.3.5, увеличение среднего роста за 9 лет по наблюденным выборкам равняется $0,78 \pm 0,030$ см. Какие заключения из этого факта можно сделать относительно неизвестного увеличения Δt среднего значения совокупности t ? В действительности мы имеем наблюденное значение 0,78 см величины, распределенной приблизительно нормально, с неизвестным средним значением Δt и со стандартным отклонением, приблизительно равным 0,030 см. Предположим, что слово „приблизительно“ в обоих случаях можно опустить, и обозначим, как обычно, через λ_p p -процентное значение нормального отклонения (см. параграф 17.2). Рассмотрим гипотезу $\Delta t = c$, где c — заданное число. Если мы работаем с p -процентным уровнем значимости, то эту гипотезу, очевидно, можно считать согласующейся с данными, когда c лежит в пределах $0,78 \pm 0,030 \lambda_p$; в других случаях от этой гипотезы следует отказаться. Значения $0,78 \pm 0,030 \lambda_p$ называются p -процентными доверительными пределами для Δt , а интервал, определяемый этими пределами, называется p -процентным доверительным интервалом. К этим понятиям мы вернемся в главе 34.

Пример 3. Достигнение исключительно высоких или низких уровней воды в озерах или реках часто имеет большое практическое значение. Для средних юньских уровней воды в озере Вэнэрн за $n = 124$ года (1807 — 1930), по данным Линдквиста [149], среднее значение равняется $\bar{x} = 4454,5$ см над уровнем моря, а стандартное отклонение $s = 48,51$ см. Распределение хорошо совпадает с нормальной кривой. Объединяя исходные данные (которых мы здесь не приводим) в девять групп с интервалом разбиения длины $h = 20$ см, получаем $\chi^2 = 3,728$. При $9 - 2 - 1 = 6$ степенях свободы этому значению χ^2 соответствует $P = 0,71$, так что согласие с гипотетическим распределением очень хорошее.

Если обозначить через x , y — верхнее значение нормальной выборки в n значений, а через z , w — нижнее значение этой выборки, то среднее значение и стандартное отклонение для x , можно вычислить

по формуле (28.6.16), а соответствующие выражения для y_v найдутся путем небольшого видоизменения этой формулы. Заменяя в этих выражениях параметры совокупности m и σ приведенными выше выборочными значениями \bar{x} и s и пренебрегая поправочными членами, мы получим средние значения и стандартные ошибки, приведенные в таблице 31.3.6, в которой содержатся также предельные июньские уровни, фактически наблюдавшиеся в течение указанного периода.

Таблица 31.3.6

Предельные июньские уровни воды в озере Вэнерн (1807—1930 гг.)

v	x_v наблюденное	$E(x_v)$ прибл.	$d(x_v)$	Разность в единицах стандартной ошибки	y_v наблюденное	$E(y_v)$ прибл.	$d(y_v)$	Разность в единицах стандартной ошибки
1	4566	4582,1	20,04	-0,80	4350	4326,9	20,04	+1,15
2	4548	4566,5	12,55	-1,47	4356	4342,6	12,55	+1,07
3	4546	4558,7	9,82	-1,29	4360	4350,4	9,82	+0,98
4	4535	4553,4	8,32	-2,21	4366	4355,6	8,32	+1,25
5	4535	4549,5	7,35	-1,97	436	4359,5	7,35	+0,88

Абсолютная величина разности между наблюденными значениями и соответствующими средними значениями ни в одном случае не превышает значений, которые можно приписать случайным колебаниям. Однако мы замечаем, что все x_v меньше соответствующих средних значений $E(x_v)$, а все y_v больше значений $E(y_v)$. Это объясняется отчасти корреляцией между величинами x_v (и между y_v), а отчасти тем, что значительные ошибки оказывают влияние на приближенные средние значения, так как мы имеем дело со сравнительно небольшим значением $n = 124$.

Если допустить, что распределение останется неизменным в течение, скажем, 500 лет, то мы получим тем же способом, как и выше, среднее значение для максимального x_1 в течение всего периода $m = 4603,5$ см и стандартную ошибку 17,6 см. Поэтому представляется совершенно невероятным, чтобы в течение рассматриваемого периода уровень воды превысил $4603,5 + 4 \cdot 17,6 = 4673,9$ см.

Пример 4. В таблице 31.3.7 приведены данные, заимствованные из классической работы Стьюдента [221] о t -распределении. По этим

данним требуется проверить, существует ли значимая разница между действием двух снотворных средств А и В. Если предположить, что разность между дополнительными часами сна, вызванными действием этих двух наркотиков, распределена нормально, то последний столбец в таблице 31.3.7 будет представлять выборку в $n = 10$ значений из нормальной совокупности. Согласно обычной нулевой гипотезе о том, что разницы между действием обоих наркотиков не существует, среднее значение этой нормальной совокупности $m_0 = 0$. Если нулевая гипотеза верна, то стьюдентово отношение $t = \sqrt{9} \frac{\bar{z} - 0}{s_0}$ имеет t -распределение с 9 степенями свободы (см. параграф 31.2). По наблюденным значениям получаем $t = 4,06$, чему соответствует значение P , заключающееся между 0,01 и 0,001 (см. таблицу IV, стр. 611). Таким образом, отклонение от нуля значимо, и нулевая гипотеза опровергнута.

В этом случае значение $n = 10$ невелико, так что можно ожидать, что приближенный критерий, основанный на стандартной ошибке величины \bar{z} , не даст особенно точного результата. Если применить этот критерий, используя для стандартной ошибки оценку $\frac{s_0}{\sqrt{10 - 1}}$, то мы должны будем считать то же значение, что и выше, именно $\sqrt{9} \frac{\bar{z} - 0}{s_0} = 4,06$, наблюденным значением величины, которая, согласно нулевой гипотезе, нормальна (0,1). По таблице II (стр. 609) этому наблюденному значению соответствует вероятность $P < 0,0001$. Если сравнить вероятность $P < 0,0001$ со значением P , получающимся для точного критерия, то обнаружится, что ошибка, возникающая при использовании приближенного критерия, ведет к преувеличению значимости отклонения.

Если для проверки действия двух снотворных средств в экспериментах, зарегистрированных таблицей 31.3.7, были бы использованы две различные группы пациентов, по десять человек в каждой группе, то полученные данные можно исследовать также и другими способами (см. Р. Фишер [13], стр. 123—125). Предположим, что длительность сна после приема каждого из снотворных средств имеет нормальное распределение с одинаковыми в обоих случаях стандартными отклонениями. Тогда выборки x и y будут независимыми выборками из нормальных совокупностей с одной и той же σ , и требуется проверить нулевую гипотезу о том, что средние значения m_1 и m_2 этих совокупностей равны. Величина u , определяемая формулой (31.2.1),

в которой следует положить $n_1 = n_2 = 10$, имеет в нашем случае t -распределение с 18 степенями свободы, и по таблице 31.3.7 получаем $\alpha = 1,86$, чему соответствует $P = 0,08$, так что мы не обнаруживаем какой-либо значимой разницы между действием двух снотворных средств.

Таблица 31.3.7

Дополнительные часы сна у десяти пациентов, вызванные действием двух снотворных средств А и В

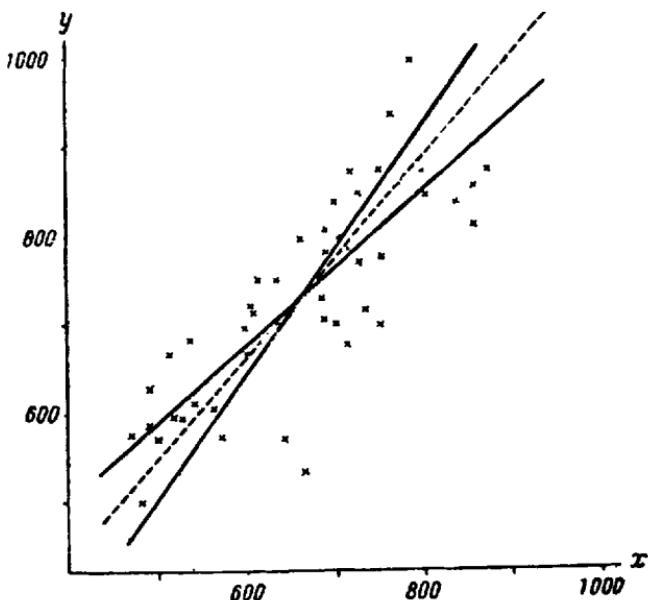
Пациент	Наркотик А x	Наркотик В y	Разность $z = x - y$
1	1,9	0,7	1,2
2	0,8	-1,6	2,4
3	1,1	-0,2	1,3
4	0,1	-1,2	1,3
5	-0,1	-0,1	0,0
6	4,4	3,4	1,0
7	5,5	3,7	1,8
8	1,6	0,8	0,8
9	4,6	0,0	4,6
10	3,4	2,0	1,4
	$\bar{x} = 2,33$ $s_x = 1,899$	$\bar{y} = 0,75$ $s_y = 1,697$	$\bar{z} = 1,58$ $s_z = 1,167$

В случаях, когда можно предполагать, что столбцы x и y независимы, применимы оба вышеописанных метода, и если один из критериев обнаружит явно значимую разность, то мы должны отказатьься от нулевой гипотезы, даже если второй критерий не обнаруживает какой-либо значимой разности. Однако в интересующем нас случае с таблицей 31.3.7 существует очевидная корреляция между столбцами x и y , созданная тем, что соответствующие друг другу числа относятся к одному и тому же пациенту, так что применение второго метода незаконно.

Пример 5. Среднее значение и стандартное отклонение для юльских температур в Стокгольме за $n = 100$ лет (1841—1940) (см. таблицу 30.4.2) равняются $\bar{x} = 16,982$ и $s = 1,6145$. Средние значения за 30 первых и 30 последних лет этого периода соответ-

ственno равны 16,893 и 17,463. Значимо ли отличаются эти групповые средние значения от общего среднего значения 16,982?

По формуле (29.4.5) получаем $t = -0,36$ для $k = 30$ первых лет и $t = 1,97$ для 30 последних лет, в обоих случаях с $n - 2 = 98$ степенями свободы. Оба значения лежат ниже 5-процентного предела значимости, так что примененный критерий не обнаруживает никаких значимых колебаний летней температуры в течение столетия.



Фиг. 31. Цены на картофель в 46 районах Швеции, декабрь 1936 (x) и декабрь 1937 (y). Линии регрессии и линия ортогональной регрессии (пунктир).

Пример 6. На фиг. 31 изображено распределение цен на картофель (в эре за 100 кг) в декабре 1936 г. (x) и в декабре 1937 г. (y) в $n = 46$ различных местах в Швеции по данным официальной статистики. Обычные характеристики выборки имеют значения

$$\bar{x} = 660,57; \bar{y} = 732,59; s_1 = 106,86; s_2 = 120,91; r = 0,7928; b_{12} = 0,7007; b_{21} = 0,8971.$$

Предположим, что значения (x, y) образуют выборку из нормальной совокупности и требуется получить сведения о неизвестных значениях коэффициента регрессии β_{21} и коэффициента корреляции r этой совокупности.

В соответствии с (29.8.4), величина $t = \frac{s_1\sqrt{n-2}}{s_2\sqrt{1-r^2}}(b_{21} - \beta_{21})$ имеет распределение Стьюдента с $n-2$ степенями свободы. Подставляя сюда приведенные выше значения выборочных характеристик, мы сможем проверить гипотезу о том, что β_{21} равняется какому-нибудь заданному числу c . Если мы работаем с p -процентным уровнем значимости, то гипотезу следует считать согласующейся с данными, когда c лежит в пределах

$$b_{21} \pm \frac{s_2\sqrt{1-r^2}}{s_1\sqrt{n-2}} t_p,$$

где t_p означает p -процентное значение t при $n-2$ степенях свободы; в других случаях гипотезу следует забраковать. Указанные пределы являются p -процентными доверительными пределами для β_{21} (см. выше пример 2). В нашем конкретном случае для β_{21} получаются следующие доверительные пределы:

$$\begin{aligned} p = 5\% &— 0,687 \text{ и } 1,107, \\ p = 1\% &— 0,617 \text{ и } 1,177, \\ p = 0,1\% &— 0,530 \text{ и } 1,264. \end{aligned}$$

Для выборочного коэффициента корреляции $r = 0,7928$ получаем по формулам (27.8.1) и (27.8.2) приближенное среднее значение ρ и стандартную ошибку

$$d(r) = \frac{1-r^2}{\sqrt{n}} = 0,0548.$$

Если выборочное распределение для r достаточно близко к нормальному, то это распределение можно использовать для проверки гипотезы о том, что ρ равно какому-либо заданному числу. Однако, если ρ значительно отличается от нуля, то выборочное распределение для r стремится к нормальному довольно медленно, и при $n = 46$ следует ожидать, что получаемые с помощью стандартной ошибки результаты не очень точны. Поэтому рекомендуется использовать точные таблицы распределения величины r (Дэвид [261]) или логарифмическое преобразование (29.7.3) — (29.7.4), предложенное Р. Фишером. В последнем случае величину $z = \frac{1}{2} \log \frac{1+r}{1-r}$ следует считать распределенной нормально, со средним значением $\frac{1}{2} \log \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)}$ и стандартным

отклонением $\frac{1}{\sqrt{n-3}}$, так что величина

$$\lambda = \sqrt{n-3} \left[\frac{1}{2} \log \frac{1+r}{1-r} - \left(\frac{1}{2} \log \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)} \right) \right]$$

нормальна $(0, 1)$. Таким образом, работая с p -процентным уровнем значимости, мы должны считать данные совместимыми с каким-нибудь гипотетическим значением ρ , если

$$\frac{1}{2} \log \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)}$$

лежит в пределах

$$\frac{1}{2} \log \frac{1+r}{1-r} \pm \frac{\lambda_p}{\sqrt{n-3}},$$

где λ_p — p -процентное значение нормального отклонения; в других случаях гипотетическое значение ρ следует забраковать. Если r известно, то указанные пределы можно вычислить при любом p , и тогда соответствующие значения ρ можно получить численным решением уравнения вида $\frac{1}{2} \log \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)} = k$. Эти значения будут p -процентными доверительными пределами для ρ . В нашем конкретном случае для ρ получаются следующие доверительные пределы:

$$p = 5\% \text{ — } 0,6486 \text{ и } 0,8783,$$

$$p = 1\% \text{ — } 0,5913 \text{ и } 0,8980,$$

$$p = 0,1\% \text{ — } 0,5164 \text{ и } 0,9171.$$

Пример 7. В таблице 31.3.8 приведены заимствованные из официальных отчетов значения следующих четырех величин за $n=30$ лет (1913—1942):

x_1 — средний урожай озимой пшеницы в килограммах с 10^4 м^2 в 20 сельских районах Кальмара (Швеция).

x_2 — средняя температура воздуха в Кальмаре за предыдущую зиму (октябрь — март) в градусах Цельсия.

x_3 — средняя температура воздуха в Кальмаре за текущий вегетационный период (апрель — сентябрь).

x_4 — количество осадков в миллиметрах за вегетационный период, среднее для трех метеорологических станций в районе.

В этом случае имеет смысл рассматривать величины x_2 , x_3 и x_4 как причины, оказывающие то или иное влияние на количество

Таблица 31.3.8

Урожай пшеницы, температура и количество осадков в районе Кальмара

Год	Урожай пшеницы x_1	Зимняя температура x_2	Летняя температура x_3	Количество осадков x_4	Наилучшая линейная оценка для x_1 \hat{x}_1^*
1913	1990	2,7	12,8	230	2125
1914	1950	3,1	13,7	268	2295
1915	1630	1,9	12,0	188	1899
1916	1720	1,3	11,7	315	2058
1917	1560	1,0	12,7	180	1794
1918	1680	1,6	12,0	261	2004
1919	1980	2,3	12,2	216	2017
1920	2180	1,7	12,8	346	2223
1921	2370	3,1	13,1	131	1995
1922	1790	1,1	11,8	256	1918
1923	2400	1,6	11,2	327	2100
1924	1410	0,1	11,8	320	1913
1925	2570	3,7	13,2	382	2580
1926	2180	1,1	12,5	279	1996
1927	2150	2,5	12,2	351	2313
1928	2530	0,8	10,5	324	1956
1929	2100	0,8	10,9	196	1718
1930	2330	3,6	12,4	381	2529
1931	1850	1,6	10,7	273	1970
1932	2230	1,9	12,5	289	2123
1933	2510	2,2	11,9	338	2234
1934	2600	3,0	13,5	267	2271
1935	2480	3,2	12,3	372	2453
1936	1940	2,8	12,3	357	2370
1937	2770	2,1	13,5	358	2332
1938	2570	3,3	12,9	202	2154
1939	2510	3,8	13,4	311	2461
1940	1420	-1,1	11,3	172	1434
1941	810	-0,4	11,3	194	1572
1942	1990	-2,4	11,2	261	1434

урожая x_1 . Требуется исследовать природу причинной зависимости между этими величинами. Если данных так мало, как в нашем примере, то нельзя надеяться достичь очень точных результатов и придется удовлетвориться лишь некоторыми общими сведениями о значимости или незначимости различных возможных связей между нашими величинами.

Будем предполагать, что совместное распределение наших четырех величин нормально. Корреляционная матрица выборки $R = \{r_{ij}\}$ имеет вид

$$\left\{ \begin{array}{cccc} 1 & 0,59107 & 0,41082 & 0,46120 \\ 0,59107 & 1 & 0,67028 & 0,31838 \\ 0,41082 & 0,67028 & 1 & 0,10720 \\ 0,46120 & 0,31838 & 0,10720 & 1 \end{array} \right\}$$

Детерминант $R = |r_{ij}|$ есть квадрат коэффициента разброса выборки (см. параграф 22.7). Если x_i независимы, то по формуле (29.13.2) получим $E(R) = 0,806$, а $D(R)$ приближенно равно 0,115. По приведенной выше матрице R получаем значение $R = 0,273$, так что зависимость между нашими величинами явно существует.

Значимость различных r_{ij} можно оценить с помощью распределения (29.7.5), которое имеет место, когда x_i независимы. Согласно (29.7.6), гипотеза о независимости величин x_i и x_j должна быть за-бракована с точки зрения p -процентного уровня значимости, если $|r_{ij}|$ превышает значение $\frac{t_p}{\sqrt{t_p^2 + v}}$, где t_p обозначает p -процентное значе-

ние t при $v = n - 2$ степенях свободы. Таблица величины $\frac{t_p}{\sqrt{t_p^2 + v}}$ при различных значениях n и p приводится у Фишера и Иэйтса [262]. Для обычных 5-, 1- и 0,1-процентного уровней значимости значениями величины $\frac{t_p}{\sqrt{t_p^2 + v}}$ являются:

$$\sqrt{\frac{t_p}{t_p^2 + v}}$$

Число степеней свободы	$p = 5\%$	$p = 1\%$	$p = 0,1\%$
$v = 26$	0,3740	0,4786	0,5880
$v = 27$	0,3673	0,4706	0,5790
$v = 28$	0,3609	0,4629	0,5703

Для наших r_{ij} имеем $v = n - 2 = 28$ степеней свободы, так что все r_{ij} , кроме r_{24} и r_{34} , превышают 5-процентный уровень значимости.

r_{13} лежит между 5- и 1-процентными уровнями значимости, r_{14} почти равняется 1-процентному уровню значимости, а r_{12} и r_{23} даже превосходят 0,1-процентный уровень значимости. Интересно отметить, что r_{12} значительно превосходит r_{13} : температура предыдущей зимы, по-видимому, влияет на урожай сильнее, чем температура лета.

Частные коэффициенты корреляции $r_{ij \cdot k}$ можно вычислить по формуле (23.4.3), и мы получаем следующие значения:

$$\begin{aligned} r_{12 \cdot 3} &= 0,4666 & r_{13 \cdot 2} &= 0,0244 & r_{14 \cdot 2} &= 0,3570 \\ r_{12 \cdot 4} &= 0,5281 & r_{13 \cdot 4} &= 0,4096 & r_{14 \cdot 3} &= 0,4602 \end{aligned}$$

Для пределов значимости величин $r_{ij \cdot k}$ имеем, согласно формуле (29.13.5), выражения такого же вида, как и для величин r_{ij} , с $\nu = n - 3 = 27$ степенями свободы. Таким образом, из шести приведенных выше коэффициентов $r_{ij \cdot k}$ лишь $r_{12 \cdot 4}$ превышает 1-процентный уровень значимости, хотя $r_{12 \cdot 3}$ и $r_{14 \cdot 3}$ очень близки к этому уровню. Если сравнить, например, $r_{18} = 0,41082$ со значениями для $r_{13 \cdot 2}$ и $r_{13 \cdot 4}$, то обнаружится, что пренебрежение влиянием зимней температуры x_2 сводит корреляцию между урожаем x_1 и летней температурой x_3 к совершенно незначимой величине $r_{13 \cdot 2} = 0,0244$, а пренебрежение влиянием количества осадков x_4 практически не меняет этой корреляции. С другой стороны, сравнение величины $r_{12} = 0,59108$ со значениями $r_{12 \cdot 3}$ и $r_{12 \cdot 4}$ показывает, что корреляция между урожаем и зимней температурой при пренебрежении влиянием летней температуры или количества осадков уменьшается несущественно. Такие же выводы, как для r_{12} , получаются и для r_{14} . Эти соображения заставляют предположить, что действительно важными факторами являются зимняя температура x_2 и количество осадков x_4 , а влияние летней температуры x_3 объясняется главным образом тем, что x_3 довольно сильно коррелирована с x_2 ($r_{23} = 0,67028$).

Частные коэффициенты корреляции с двумя вторичными индексами вычисляются по формуле (23.4.4). Находим

$$r_{12 \cdot 34} = 0,3739, \quad r_{13 \cdot 24} = 0,0848, \quad r_{14 \cdot 23} = 0,3650,$$

так что эти значения подтверждают высказанное выше предположение, хотя ни одно из этих значений, строго говоря, не является значимым. Здесь мы имеем $\nu = n - 4 = 26$ степеней свободы, и 5-процентный предел значимости для $r_{ij \cdot kl}$ равняется 0,3740.

Рассмотрим теперь сводные коэффициенты корреляции. С помощью формулы (23.5.3) получаем

$$r_{1(23)} = 0,5914, \quad r_{1(24)} = 0,6575, \quad r_{1(34)} = 0,5872, \quad r_{1(234)} = 0,6606.$$

Сравнение $r_{12} = 0,5911$ с $r_{1(23)} = 0,5914$ подтверждает уже полученные результаты, так как показывает, что сведения о x_8 практически ничего не прибавляют к сведениям об урожае x_1 , если x_2 уже известно. Аналогично, сводный коэффициент корреляции $r_{1(24)}$ лишь незначительно меньше, чем $r_{1(23)}$.

Если величины x_1, \dots, x_k независимы, то согласно (29.13.9), произведение $n r^2_{1(2\dots k)}$ при больших n приближенно имеет распределение χ^2 с $k - 1$ степенями свободы. В нашем конкретном случае $n r^2_{1(34)} = 10,344$ с 2 степенями свободы, а $n r^2_{1(234)} = 13,092$ с 3 степенями свободы. Поскольку и $r_{1(23)}$ и $r_{1(24)}$ превышают $r_{1(34)}$, все четыре сводных коэффициента корреляции, приведенные выше, значимо отличаются от нуля.

Наконец, вычислим частные коэффициенты регрессии:

$$\begin{aligned} b_{12,34} &= 133,65, & \text{соответствующее } t &= 2,055, \\ b_{13,24} &= 44,87, & " &= 0,434, \\ b_{14,23} &= 1,9963, & " &= 1,999, \end{aligned}$$

где значения t вычислены по формуле (29.12.1) при гипотезе о том, что соответствующие значения для совокупности $\beta_{1i,jk}$ равны нулю. t имеет 26 степеней свободы, так что, согласно таблице IV (стр. 611), ни одно из трех вычисленных значений коэффициентов регрессии не является значимым, хотя b_{12} и b_{14} очень близки к 5-процентному пределу значимости. Если наблюденные значения b отождествить с неизвестными значениями для совокупности, то это будет означать, например, что повышение средней зимней температуры на один градус в среднем вызовет повышение урожая на 134 кг с 10^4 м² при неизменных летней температуре и количестве осадков, а повышение на один градус летней температуры создает увеличение урожайности лишь на 45 кг с 10^4 м².

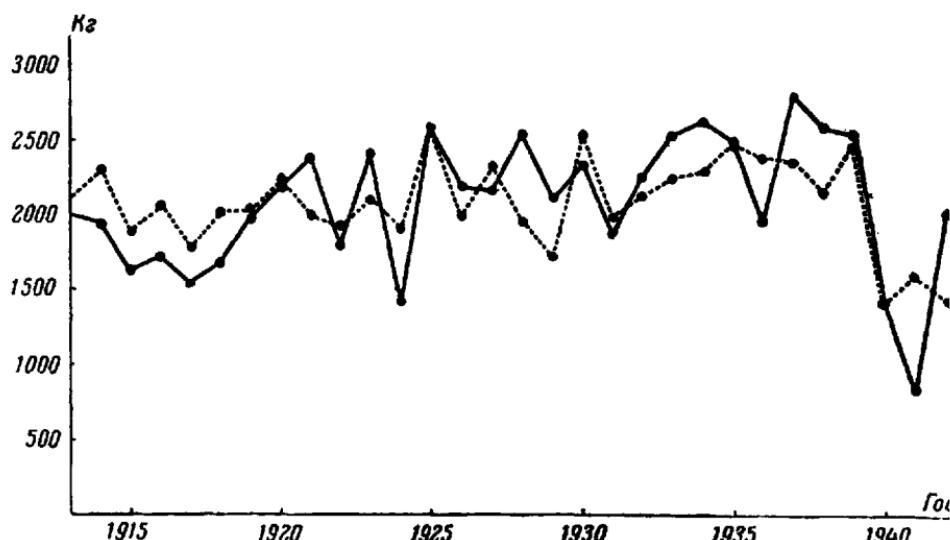
Уравнение выборочной плоскости регрессии для x_1 дает наилучшую линейную оценку наблюденного значения x_1 по значениям x_2, x_3, x_4 :

$$x_1^* = 133,65 x_2 + 44,87 x_3 + 1,9963 x_4 + 730,9.$$

Значения x_1^* , вычисленные по этой формуле, приведены в последнем столбце таблицы 31.3.8. Значения x_1 и x_1^* изображены также на фиг. 32.

Следует помнить, что во всех рассмотренных выше критериях повсюду предполагалось, что мы имеем дело с выборками, полученными методом *простого случайного выбора* (см. параграф 25.2). Это

означает, что выборочные значения предполагаются *взаимно независимыми*. Однако во многих приложениях такое предположение незаконно. Такие случаи часто наблюдаются, например, при анализе статистических *временных рядов*. Предварительный план этой книги подразумевает



Фиг. 32. Графики урожайности пшеницы x_1 и наилучшей линейной оценки \hat{x}_1 (пунктир).

вал включение главы, посвященной анализу временных рядов, основанному на математической теории *случайных процессов*. Однако автору пришлось уделить много места для обобщения теории на случай много-dimensionalных пространств, и это, к сожалению, помешало полной реализации предварительного плана. Исследование статистических временных рядов читатель может найти в диссертации Вольда [246а].

ГЛАВЫ 32—34

ТЕОРИЯ ОЦЕНОК*)

ГЛАВА 32

КЛАССИФИКАЦИЯ ОЦЕНОК

32.1. Постановка проблемы. В предыдущих главах мы неоднократно встречались с проблемой оценки некоторых параметров совокупности при помощи множества выборочных значений. Теперь мы приступим к более систематическому исследованию этой проблемы.

Основой теории оценок послужил ряд фундаментальных работ Р. Фишера ([89], [96], [103], [104] и др.). В главах 32—33 мы изложим некоторые из основных идей, введенных Фишером, дополняя кое-где его результаты. В этой главе мы займемся классификацией и исследованием свойств различных видов оценок. Глава 33 будет посвящена рассмотрению некоторых общих методов оценки, в частности, важного *метода максимума правдоподобия*, предложенного Р. Фишером. В главе 34 будет исследована возможность использования оценок для получения надежных выводов относительно значений параметров.

Допустим, что нам задана выборка из совокупности, распределение которой имеет известную математическую форму, но содержит некоторое количество неизвестных параметров. Всегда существует бесконечное число функций от выборочных значений, которые можно предложить в качестве оценок для параметров. Поэтому возникает следующий вопрос: *Как наилучшим образом использовать данные для получения оценок?* А в связи с этим вопросом немедленно возникает следующий: *Что подразумевать под „наилучшими” оценками?*

Соблазнительно было бы назвать наилучшей оценкой такую, которая наиболее близка к истинному значению оцениваемого параметра. Однако следует иметь в виду, что каждая оценка является функцией от выборочных значений, так что ее следует считать наблюдаемым

*) Значительная часть излагаемых в этих главах вопросов является в высшей степени дискуссионной, и в литературе встречается множество противоположных мнений относительно достоинств различных рассматриваемых здесь понятий и методов.

значением некоторой случайной величины. Поэтому мы никак не можем предсказать индивидуальное значение оценки в данном частном случае, так что о качестве оценки следует судить не по индивидуальным ее значениям, а лишь по распределению ее значений, которые обнаруживаются в длинном ряду испытаний, т. е. по *выборочному распределению оценки*. Если основная часть массы в этом распределении сосредоточена в малой окрестности истинного значения оцениваемого параметра, то с большой вероятностью можно считать, что оценка отличается от истинного значения лишь на малую величину. С этой точки зрения оценка будет тем „лучше“, чем *меньшее рассеяние от истинного значения* показывает выборочное распределение оценки, и поставленный выше вопрос можно сформулировать более точно следующим образом: *Как следует использовать данные, чтобы получить оценки с минимальным рассеянием?* Этот вопрос будет служить отправной точкой нашего исследования.

В части II мы видели, что рассеяние (или, наоборот, концентрацию) распределения можно измерять различными способами и что выбор между различными мерами в большой степени произволен. Такой же произвол, конечно, будет иметь место в выборе различных оценок. Каждой мере рассеяния соответствует некоторая „наилучшая“ оценка, т. е. оценка, имеющая наименьшее возможное рассеяние с точки зрения этой выбранной меры.

В дальнейшем мы будем рассматривать исключительно меры рассеяния и концентрации, связанные с *дисперсией* и ее многомерными обобщениями. Во-первых, этот выбор оправдывается общими соображениями в пользу принципа наименьших квадратов, выдвинутыми в параграфе 15.6. Далее, в том важном случае, когда выборочные распределения оценок хотя бы приблизительно нормальны, любая имеющая смысл мера рассеяния определяется вторыми моментами, так что при этом возможна лишь единственная мера. Изложение этого вопроса с некоторых других точек зрения читатель сможет найти в статьях Питмена [198], [199] и Джейри [116а].

Сначала удобно рассмотреть случай выборок из совокупности, распределение которой содержит лишь один неизвестный параметр. Этому случаю посвящены параграфы 32.2 — 32.5. В параграфах 32.6 — 32.7 будут рассмотрены случаи с многими неизвестными параметрами. Параграф 32.8 будет посвящен некоторому важному общению изложенной теории.

32.2. Две леммы. В этом параграфе мы докажем две леммы, которые понадобятся в дальнейшем. Каждая из этих лемм связана с одним из двух простейших типов распределений и является частным случаем одной общей теоремы, которую мы здесь не приводим.

Лемма 1. Предположим, что функция $g(x; a)$ при каждом фиксированном значении a , принадлежащем некоторому невырожденному интервалу A , является плотностью вероятности, имеющей первый момент $\phi(a)$ и конечный, второй момент. Пусть, далее, почти всюду на оси x для всех a из A существует частная производная $\frac{dg}{da}$, причем $\left| \frac{dg}{da} \right| < G_0(x)$, где G_0 , а также xG_0 — функции, интегрируемые на $(-\infty, +\infty)$.

Тогда для всех a из A существует производная $\frac{d\phi}{da}$, причем

$$(32.2.1) \quad \int_{-\infty}^{\infty} (x-a)^2 g(x; a) dx \cdot \int_{-\infty}^{\infty} \left(\frac{\partial \log g}{\partial a} \right)^2 g(x; a) dx \geq \left(\frac{d\phi}{da} \right)^2.$$

Знак равенства при фиксированном значении a достигается тогда и только тогда, когда существует число k , не зависящее от x , но, может быть, зависящее от a , такое, что

$$(32.2.2) \quad \frac{\partial \log g}{\partial a} = k(x-a)$$

для почти всех x , удовлетворяющих неравенству $g(x; a) > 0$.

По предположению для всех a из A имеем

$$(32.2.3) \quad \int_{-\infty}^{\infty} g(x; a) dx = 1, \quad \int_{-\infty}^{\infty} x g(x; a) dx = \phi(a),$$

и оба эти интеграла удовлетворяют условиям дифференцируемости по параметру под знаком интеграла (см. параграф 7.3), так что $\frac{d\phi}{da}$ существует и имеет вид*).

$$\frac{d\phi}{da} = \int_{-\infty}^{\infty} x \frac{\partial g}{\partial a} dx = \int_{-\infty}^{\infty} (x-a) \frac{\partial g}{\partial a} dx = \int_{-\infty}^{\infty} (x-a) \sqrt{g} \frac{\partial \log g}{\partial a} \sqrt{g} dx.$$

*). Если $g(x; a) = 0$ для всех x , принадлежащих некоторому интервалу, то и $\frac{\partial g}{\partial a} = 0$, так как в противном случае g имела бы отрицательные значения. Тогда выражение $\frac{\partial \log g}{\partial a} \sqrt{g} = \frac{1}{\sqrt{g}} \frac{\partial g}{\partial a}$ должно было бы равняться нулю.

Соотношение (32.2.1) получается немедленно применением неравенства Шварца (9.5.1)*)

Знак равенства в (9.5.1) достигается тогда и только тогда, когда существуют две постоянные u и v , не равные одновременно нулю, и такие, что $ug(x) + vh(x) = 0$ для почти всех (P) значений x . Поскольку $(x - a)\sqrt{g}$ не может равняться нулю почти всюду на оси x , то при фиксированном значении a знак равенства в (32.2.1) достигается тогда и только тогда, когда

$$\frac{\partial \log g}{\partial x} \sqrt{g} = k(x - a) \sqrt{g}$$

для почти всех x , причем k не зависит от x . Тем самым лемма 1 полностью доказана.

Приведем два примера, в которых выполняется условие (32.2.2). Легко проверить, что в обоих этих примерах в (32.2.1) достигается знак равенства.

Пример 1. *Нормальное распределение со средним значением a и постоянным стандартным отклонением.* Полагая

$$g(x; a) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

где σ не зависит ни от x , ни от a , мы получим $\phi(a) = a$ и $\frac{\partial \log g}{\partial a} = \frac{x-a}{\sigma^2}$ для всех x и a .

Пример 2. *Распределение χ^2 .* Согласно (18.1.6), плотность вероятности $k_n(v)$ распределения χ^2 имеет первый момент n . Таким образом, плотность вероятности $g(x; a) = \frac{n}{a} k_n\left(\frac{nx}{a}\right)$, где $a > 0$, имеет первый момент $\phi(a) = a$ и из (18.1.3) получается $\frac{\partial \log g}{\partial a} = \frac{n}{2a^2}(x - a)$ для всех $x > 0$ и $a > 0$.

Лемма 2. *Предположим, что при каждом значении a , принадлежащем некоторому невырожденному интервалу A , конечная или счетная последовательность функций $p_1(a), p_2(a), \dots$ представляет собой вероятности некоторого распределения дискретного типа, причем соответствующие точки сосредоточения массы u_1, u_2, \dots не зависят от a . Пусть, далее, это распределение имеет первый момент $\phi(a)$ и конечный второй момент и производные $p'_i(a)$ существуют для*

*⁴⁾ Я обязан профессору Л. Альфорсу одним замечанием, упростившим мое первоначальное доказательство.

всех i и для каждого a из A , причем ряд $\sum u_i p'_i(a)$ сходится абсолютно и равномерно в A . Тогда для всех a из A существует производная $\frac{d\psi}{da}$, причем,

$$(32.2.4) \quad \sum_i (u_i - a)^2 p_i(a) \cdot \sum_i \left(\frac{d \log p_i}{da} \right)^2 p_i(a) \geq \left(\frac{d\psi}{da} \right)^2.$$

Знак равенства при фиксированном значении a достигается тогда и только тогда, когда существует число k , не зависящее от i , но, может быть, зависящее от a , такое, что

$$(32.2.5) \quad \frac{d \log p_i}{da} = k(u_i - a),$$

для всех i , удовлетворяющих условию $p_i(a) > 0$.

Эта лемма вполне аналогична лемме 1 и доказывается таким же способом с помощью следующих соотношений, соответствующих соотношениям (32.2.3):

$$\sum_i p_i(a) = 1, \quad \sum_i u_i p_i(a) = \psi(a).$$

Как и в предыдущей лемме, мы приведем два примера, в которых выполняется условие (32.2.5). Легко проверить, что в обоих этих примерах в (32.2.4) достигается знак равенства.

Пример 3. Для биномиального распределения с $p = \frac{a}{n}$ имеем $u_i = i$ и $p_i := \binom{n}{i} \left(\frac{a}{n}\right)^i \left(1 - \frac{a}{n}\right)^{n-i}$, где $i = 0, 1, \dots, n$. Поэтому среднее значение равняется $\psi(a) = np = a$ и $\frac{d \log p_i}{da} = \frac{i}{a} - \frac{n-i}{n-a} = \frac{n}{a(n-a)}(u_i - a)$.

Пример 4. Если $n \rightarrow \infty$ при фиксированном a , то биномиальное распределение стремится к распределению Пуассона с $u_i = i$ и $p_i = \frac{a^i}{i!} e^{-a}$.

Здесь мы имеем $\psi(a) = a$ и $\frac{d \log p_i}{da} = \frac{u_i - a}{a}$.

32.3. Минимум дисперсии оценки. Эффективные оценки. Предположим, что каждому значению параметра a , принадлежащему к иерархическому интервалу A , соответствует некоторая функция распределения $F(x;a)$. Пусть x_1, \dots, x_n — выборка в n значений из

совокупности с функцией распределения $F(x; \alpha)$, где α — какое-то значение из интервала A , и требуется оценить неизвестное „истинное значение“ параметра α . Для обозначения какой-нибудь функции от выборочных значений, предлагаемой в качестве оценки для α , будем употреблять общий символ $\alpha^* = \alpha^*(x_1, \dots, x_n)^*$.

В параграфах 32.3 — 32.4 мы будем считать объем выборки n фиксированным числом ≥ 1 . В параграфе 32.5 мы перейдем к вопросам, относящимся к асимптотическому поведению наших оценок при больших n .

В соответствии с терминологией, введенной в параграфе 27.6, α^* называется *несмешенной* оценкой для α , если $E(\alpha^*) = \alpha$. Как было показано на некоторых простых примерах в параграфе 27.6, часто можно с помощью простой поправки исключить смещение оценки и получить несмешенную оценку. Однако в общем случае оценка имеет некоторое *смещение* $b(\alpha)$, зависящее от α , так что

$$E(\alpha^*) = \alpha + b(\alpha).$$

Можно показать, что при некоторых общих условиях среднее квадратичное отклонение $E(\alpha^* - \alpha)^2$ ограничено снизу некоторым положительным числом, зависящим только от функции распределения $F(x; \alpha)$, объема выборки n и смещения $b(\alpha)$. В частном случае, когда α^* является несмешенной оценкой, каково бы ни было истинное значение α из A , смещение $b(\alpha)$ тождественно равно нулю, и дисперсия $D^*(\alpha^*)$ ограничена снизу некоторым числом, зависящим только от F и α .

Мы ограничимся доказательством этой теоремы для случая, когда функция распределения принадлежит к одному из двух простейших типов.

1. Непрерывный тип. Рассмотрим распределение непрерывного типа с плотностью вероятности $f(x; \alpha)$, где α может иметь любое значение из интервала A . Значения x_1, \dots, x_n , получаемые при

*) Здесь важно указать на различие в значениях символов α^* и α . По определению, α^* есть функция от выборочных значений x_1, \dots, x_n , которые рассматриваются как случайные величины. Таким образом, α^* сама есть *случайная величина*, имеющая некоторое выборочное распределение. С другой стороны, α есть *переменная величина в смысле обычного анализа*, которая для совокупности, соответствующей данной выборке, может принимать произвольное фиксированное, хотя и, может быть, неизвестное, значение из интервала A .

независимых выборках из этого распределения, суть независимые случайные величины, имеющие одну и ту же плотность вероятности $f(x; \alpha)$. Каждая конкретная выборка представляется определенной точкой $\mathbf{x} = (x_1, \dots, x_n)$ в пространстве выборок R_n переменных x_1, \dots, x_n , и дифференциал вероятности совместного распределения имеет вид

$$L(x_1, \dots, x_n; \alpha) dx_1 \dots dx_n = f(x_1; \alpha) \dots f(x_n; \alpha) dx_1 \dots dx_n.$$

Совместная плотность вероятности $L = f(x_1; \alpha) \dots f(x_n; \alpha)$ называется функцией правдоподобия для выборки (см. параграф 33.2).

Пусть $\alpha^* = \alpha^*(x_1, \dots, x_n)$ — однозначная функция от x_1, \dots, x_n , не зависящая от α , непрерывная и имеющая непрерывные частные производные $\frac{\partial \alpha^*}{\partial x_i}$ во всех точках \mathbf{x} , кроме, может быть, некоторых точек, лежащих на конечном числе гиперповерхностей. Постараемся использовать α^* в качестве оценки для α и предположим, что $E(\alpha^*) = \alpha + b(\alpha)$, так что $b(\alpha)$ — смещение оценки α^* .

Уравнение $\alpha^* = c$ при различных значениях c определяет семейство гиперповерхностей в R_n , и каждую точку в R_n можно однозначно определить значением α^* , соответствующим той гиперповерхности этого семейства, на которой лежит рассматриваемая точка, и $n - 1$ «локальными» координатами ξ_1, \dots, ξ_{n-1} , определяющими положение точки на этой гиперповерхности. Рассмотрим теперь преобразование, при котором старые переменные x_1, \dots, x_n заменяются новыми переменными $\alpha^*, \xi_1, \dots, \xi_{n-1}$. Выбирая «локальные» координаты ξ_i так, чтобы это преобразование удовлетворило условиям А) и В) параграфа 22.2, мы сможем записать совместную плотность вероятности новых величин в виде

$$f(x_1; \alpha) \dots f(x_n; \alpha) |J|,$$

где J — якобиан преобразования, а x_i следует заменить их выражениями через новые величины.

Случайная величина α^* имеет некоторое распределение, зависящее, вообще говоря, от параметра α ; обозначим соответствующую плотность вероятности через $g(\alpha^*; \alpha)$. Далее, совместное условное распределение величин ξ_1, \dots, ξ_{n-1} , соответствующее заданному значению α^* имеет плотность вероятности, которую мы обозначим через

$$h(\xi_1, \dots, \xi_{n-1} | \alpha^*; \alpha).$$

Согласно (22.1.1), имеем

$$(32.3.1) \quad f(x_1; a) \dots f(x_n; a) |J| = g(a^*; a) h(\xi_1, \dots, \xi_{n-1} | a^*; a),$$

и преобразование дифференциала вероятности можно записать, в соответствии с (22.2.3), в виде

$$(32.3.2) \quad f(x_1; a) \dots f(x_n; a) dx_1 \dots dx_n = \\ = g(a^*; a) h(\xi_1, \dots, \xi_{n-1} | a^*; a) da^* d\xi_1 \dots d\xi_{n-1}.$$

Предположим теперь, что для каждого a из A при почти всех значениях $x, a^*, \xi_1, \dots, \xi_{n-1}$ существуют частные производные $\frac{\partial f}{\partial a}, \frac{\partial g}{\partial a}$ и $\frac{\partial h}{\partial a}$, причем

$$\left| \frac{\partial f}{\partial a} \right| < F_0(x), \quad \left| \frac{\partial g}{\partial a} \right| < G_0(a^*), \quad \left| \frac{\partial h}{\partial a} \right| < H_0(\xi_1, \dots, \xi_{n-1}; a^*),$$

где F_0, G_0, a^*G_0 и H_0 — функции, интегрируемые соответственно по всему пространству x, a^*, a^* и ξ_1, \dots, ξ_{n-1} . В таком случае будем говорить, что мы имеем дело со *случаем регулярной оценки непрерывного типа*, причем a^* будет называться *регулярной оценкой* для a . Теперь мы приступим к доказательству основной теоремы, которую можно сформулировать следующим образом:

В каждом случае регулярной оценки непрерывного типа среднее квадратичное отклонение оценки a^ от истинного значения a удовлетворяет неравенству*

$$(32.3.3) \quad E(a^* - a)^2 \geq \frac{\left(1 + \frac{db}{da}\right)^2}{n \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial a}\right)^2 f(x; a) dx}.$$

При любом a из A знак равенства достигается тогда и только тогда, когда удовлетворяются следующие два условия, коль скоро $g(a^*; a) > 0$:

А) Плотность вероятности $h(\xi_1, \dots, \xi_{n-1} | a^*; a)$ не зависит от a .

Б) $\frac{\partial \log f}{\partial a} = k(a^* - a)$, где k не зависит от a^* , но может зависеть от a .

В частном случае, когда a^* является несмещенной оценкой, каково бы ни было значение a из A , мы имеем $b(a) \equiv 0$, и (32.3.3)

обращается в

$$(32.3.3a) \quad D^2(a^*) \geq \frac{1}{n \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial a} \right)^2 f dx}.$$

Согласно нашему предположению относительно функций f и h , соотношения

$$\int_{-\infty}^{\infty} f(x; a) dx = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h(\xi_1, \dots, \xi_{n-1} | a^*; a) d\xi_1 \dots d\xi_{n-1} = 1$$

удовлетворяют условиям дифференцируемости по параметру a под знаком интегралов (см. параграф 7.3). Получающиеся при этом соотношения можно записать в виде

$$(32.3.4) \quad \begin{aligned} & \int_{-\infty}^{\infty} \frac{\partial \log f}{\partial a} f(x; a) dx = \\ & = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\partial \log h}{\partial a} h(\xi_1, \dots, \xi_{n-1} | a^*; a) d\xi_1 \dots d\xi_{n-1} = 0. \end{aligned}$$

Беря логарифмические производные по a от обеих частей равенства (32.3.1), мы получим, поскольку якобиан J не зависит от a ,

$$(32.3.5) \quad \sum_1^n \frac{\partial \log f(x_i; a)}{\partial a} = \frac{\partial \log g}{\partial a} + \frac{\partial \log h}{\partial a}.$$

Возведем обе части этого равенства в квадрат, умножим соответственно на правую и левую части равенства (32.3.2) и проинтегрируем полученные выражения по всему пространству. Согласно (32.3.4), все слагаемые, содержащие произведения двух различных производных, обращаются в нуль, и мы получим

$$(32.3.6) \quad \begin{aligned} & n \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial a} \right)^2 f(x; a) dx = \int_{-\infty}^{\infty} \left(\frac{\partial \log g}{\partial a} \right)^2 g(a^*; a) da^* + \\ & + \int_{-\infty}^{\infty} g da^* \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\frac{\partial \log h}{\partial a} \right)^2 h d\xi_1 \dots d\xi_{n-1} \geq \\ & \geq \int_{-\infty}^{\infty} \left(\frac{\partial \log g}{\partial a} \right)^2 g(a^*; a) da^*. \end{aligned}$$

Приведенное здесь доказательство этого неравенства принадлежит Дюге [76]. Знак равенства здесь достигается тогда и только тогда, когда $\frac{\partial h}{\partial a} = 0$ почти во всех точках, в которых $g > 0$, т. е. если выполняется условие А).

Наконец, плотность вероятности $g(a^*; a)$ удовлетворяет условиям леммы 1 предыдущего параграфа при $\phi(a) = a + b(a)$, и, применив лемму к неравенству (32.3.6), мы получаем доказательство нашей теоремы.

Интеграл в знаменателях правых частей равенств (32.3.3) и (32.3.3a) можно представить одним из следующих эквивалентных выражений:

$$E \left(\frac{\partial \log f}{\partial a} \right)^2 = \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial a} \right)^2 f dx = \int_{-\infty}^{\infty} \frac{1}{f} \left(\frac{\partial f}{\partial a} \right)^2 dx.$$

Легко видеть, что доказанная только что теорема верна и в случае выборок из *многомерной* совокупности с плотностью вероятности $f(x_1, \dots, x_k; a)$, содержащей неизвестный параметр a .

Рассмотрим теперь случай регулярной и несмещенной оценки a^* . В этом случае правая часть равенства (32.3.3a) представляет собою наименьшее возможное значение дисперсии $D^2(a^*)$. Отношение этого минимального значения к фактическому значению $D^2(a^*)$ называется *эффективностью* оценки a^* и обозначается $e(a^*)$. Очевидно, что $0 \leq e(a^*) \leq 1$. Если в (32.3.3a) достигается знак равенства, то $D^2(a^*)$ достигает своего наименьшего возможного значения, и $e(a^*) = 1$. В этом случае мы будем говорить, что a^* — *эффективная оценка*^{*}). Это понятие было введено Р. Фишером [89], [96].

Из доказанной выше теоремы следует, что регулярная несмещенная оценка эффективна тогда и только тогда, когда выполнены

*). Как правило, этот термин употребляется для описания поведения оценки в больших выборках, т. е. при неограниченном возрастании n . Однако здесь удобнее проводить различие между *эффективной оценкой*, под которой понимается оценка с минимальной для данного конечного объема n выборки дисперсией, и *асимптотически эффективной оценкой* (см. параграф 32.5), которая имеет аналогичное свойство в случае выборок неограниченного возрастающего объема. *Эффективная оценка* существует лишь при весьма ограничительных условиях (см. параграф 32.4), в то время как существование *асимптотически эффективной оценки* можно доказать при некоторых общих условиях регулярности (см. параграф 33.3).

условия А) и В). Это делается очевидным, если записать $e(a^*)$ в виде

$$(32.3.7) \quad e(a^*) = \frac{\min D^2(a^*)}{D^2(a^*)} = \frac{1}{n E\left(\frac{\partial \log g}{\partial a}\right)^2 D^2(a^*)} = \\ = \frac{E\left(\frac{\partial \log g}{\partial a}\right)^2}{n E\left(\frac{\partial \log f}{\partial a}\right)^2} \cdot \frac{1}{E\left(\frac{\partial \log g}{\partial a}\right)^2 D^2(a^*)}.$$

Оба множителя в последнем выражении не превышают единицы, а для эффективности необходимо, чтобы оба множителя равнялись единице. Но первый множитель равен единице тогда и только тогда, когда выполняется условие А) доказанной выше теоремы, а второй множитель обращается в единицу тогда и только тогда, когда выполняется условие В). Если эффективная оценка существует, то ее всегда можно получить с помощью *метода максимума правдоподобия*, предложенного Р. Фишером (см. параграф 33.2).

Пусть теперь a_1^* — эффективная оценка, а a_2^* — какая-нибудь регулярная несмещенная оценка с эффективностью $e > 0$. Покажем, что *коэффициент корреляции между a_1^* и a_2^* равняется $\rho(a_1^*, a_2^*) = \sqrt{e}$* . Действительно, регулярная несмещенная оценка $a^* = (1 - k)a_1^* + ka_2^*$ имеет дисперсию

$$D^2(a^*) = \left((1 - k)^2 + \frac{2\rho k(1 - k)}{\sqrt{e}} + \frac{k^2}{e} \right) D^2(a_1^*) = \\ = \left(1 + 2k\frac{\rho - \sqrt{e}}{\sqrt{e}} + k^2 \frac{e - 2\rho\sqrt{e} + 1}{e} \right) D^2(a_1^*),$$

и если $\rho \neq \sqrt{e}$, то коэффициент при $D^2(a_1^*)$ всегда можно сделать меньшим 1, придавая числу k достаточно малое положительное или отрицательное значение. Но тогда получим $D^2(a^*) < D^2(a_1^*)$, и a^* будет иметь эффективность > 1 , что невозможно.

В частности, при $e = 1$ мы имеем $\rho = 1$, так что две эффективные оценки a_1^* и a_2^* имеют одно и то же среднее значение a , одну и ту же дисперсию и коэффициент корреляции $\rho = 1$. Тогда из параграфа 21.7 будет следовать, что вся масса в совместном распределении величин a_1^* и a_2^* сосредоточена на прямой $a_1^* = a_2^*$. Таким образом, *две эффективные оценки одного и того же параметра „почти всегда“ равны друг другу*.

В этом параграфе мы приведем несколько примеров эффективных оценок (примеры 1—2 для непрерывного случая, примеры 3—4 для дискретного случая). Читатель проверит, что во всех случаях для эффективных оценок выполняются условия А) и В). Чтобы убедиться в этом, — мы говорим здесь о непрерывном случае, в дискретном случае все может быть сделано аналогично, — следует прежде всего найти плотность вероятности $g(x^*; a)$ исследуемой оценки, и тогда выполнение условия В) будет доказано с помощью примеров параграфа 32.2. Далее следует ввести подходящие вспомогательные переменные ξ_1, \dots, ξ_{n-1} и вычислить условную плотность вероятности h по формуле (32.3.1); тогда останется лишь проверить, что h не зависит от a . Во всех примерах, кроме четвертого, мы имеем дело лишь с регулярными оценками. Читатель должен проверить это детально хотя бы в некоторых случаях.

Пример 1. Среднее значение нормальной совокупности. Полагая

$$f(x, m) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

где $a = m$ — параметр, подлежащий оценке, а σ — известная постоянная, мы можем выбрать в качестве A любой конечный интервал и получим

$$E\left(\frac{\partial \log f}{\partial m}\right)^2 = \int_{-\infty}^{\infty} \left(\frac{x-m}{\sigma^2}\right)^2 f dx = \frac{1}{\sigma^2}.$$

Следовательно, дисперсия любой регулярной несмешанной оценки m^* удовлетворяет неравенству $D^2(m^*) \geq \frac{\sigma^2}{n}$. В частности, для оценки $m^* = \bar{x} = \sum \frac{x_i}{n}$ получаем, согласно параграфу 27.2, $E(\bar{x}) = m$ и $D^2(\bar{x}) = \frac{\sigma^2}{n}$, так что *среднее значение есть эффективная оценка для m* .

Мы видели выше, что некоторые другие возможные оценки для m , как, например, выборочная медиана (см. параграф 28.5) и среднее y -х верхнего и нижнего значений выборки (см. (28.6.17)) имеют большую дисперсию, чем \bar{x} .

Поучительно рассмотреть различные другие функции от выборочных значений, которые можно использовать в качестве несмешанных оценок для m ; при этом окажется, что их дисперсия всегда по меньшей мере равна $\frac{\sigma^2}{n}$. Приведем здесь простой пример такого рода. Рассмотрим выборку в $n = 3$ значения из указанной выше нормальной совокупности и расположим выборочные значения в порядке возрастания величины: $x_1 \leq x_2 \leq x_3$. Можно было бы думать, что взвешенное среднее значение

$$z = cx_1 + (1 - 2c)x_2 + cx_3$$

при некотором выбранном подходящим образом значении c будет „лучшей“ оценкой для m , чем простое среднее арифметическое значение, получающееся при $c = \frac{1}{3}$. Однако $E(z) = m$ и

$$D^2(z) = \frac{\sigma^2}{3} + \frac{3\sigma^2}{\pi} (2\pi - 3\sqrt{3}) \left(c - \frac{1}{3}\right)^2,$$

так что дисперсия для z достигает наименьшего значения в точности при $c = \frac{1}{3}$. Доказательство формулы для $D^2(z)$ и проверку выполнения условий регулярности оценки мы предоставляем читателю в качестве упражнения.

Пример 2. Дисперсия нормальной совокупности. Полагая

$$f(x, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}},$$

где $\sigma = \sigma^2$ — параметр, подлежащий оценке, а m — известная постоянная, мы можем выбрать в качестве A любой конечный интервал $a < \sigma^2 < b$, $a > 0$, и получим

$$E\left(\frac{\partial \log f}{\partial \sigma^2}\right)^2 = \int_{-\infty}^{\infty} \left(\frac{(x-m)^2}{2\sigma^4} - \frac{1}{2\sigma^2}\right)^2 f dx = \frac{1}{2\sigma^4}.$$

Следовательно, дисперсия любой регулярной несмещенной оценки для σ^2 будет по меньшей мере равна $\frac{2\sigma^4}{n}$. Исправляя за счет смещения выборочную

дисперсию s^2 (см. 27.6), мы получим выражение $\frac{n}{n-1} s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$, которое, согласно (27.4.5), является несмещенной оценкой для σ^2 , имеющей дисперсию $\frac{2\sigma^4}{n-1}$.

Очевидно, эта оценка *не* является эффективной, а имеет эффективность $\frac{n-1}{n} < 1$. С другой стороны, рассмотрим оценку $s_0^2 = \frac{1}{n} \sum (x_i - m)^2$. Это законно, поскольку m — известная постоянная.

Легко видеть, что s_0^2 имеет среднее значение σ^2 и дисперсию $\frac{2\sigma^4}{n}$, т. е. является эффективной оценкой для σ^2 .

Пример 3. Стандартное отклонение нормальной совокупности. Если в распределении предыдущего примера вместо дисперсии σ^2 считать подлежащим оценке параметром стандартное отклонение σ , то получим

$$E\left(\frac{\partial \log f}{\partial \sigma}\right)^2 = \int_{-\infty}^{\infty} \left(\frac{(x-m)^2}{\sigma^3} - \frac{1}{\sigma}\right)^2 f dx = \frac{2}{\sigma^2}.$$

Следовательно, дисперсия любой регулярной несмешенной оценки для σ будет по меньшей мере равна $\frac{\sigma^2}{2n}$. Рассмотрим, например, выражение

$$s' = \sqrt{\frac{n}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} s,$$

где s — стандартное отклонение выборки. По формуле (29.3.3) получаем $E(s') = \sigma$ и

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

так что эффективность $e(s')$ стремится к 1 при $n \rightarrow \infty$. Однако при малом n эффективность значительно меньше единицы. Полагая, например, $n=2$, получим $e(s') = \frac{1}{2(\pi-2)} = 0,4380$, а при $n=3$ $e(s') = \frac{\pi}{6(4-\pi)} = 0,6100$.

Аналогично убеждаемся, что выражение

$$s'_0 = \sqrt{\frac{n}{2}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} s_0,$$

где s_0 — величина, определенная в примере 2, является несмешенной оценкой для σ , имеющей дисперсию

$$D^2(s'_0) = \left(\frac{n}{2} \frac{\Gamma^2\left(\frac{n}{2}\right)}{\Gamma^2\left(\frac{n+1}{2}\right)} - 1 \right) \sigma^2 = \frac{\sigma^2}{2n} + O\left(\frac{1}{n^2}\right).$$

Эффективность $e(s'_0)$ стремится к 1 при $n \rightarrow \infty$. При $n=2$ имеем $e(s'_0) = \frac{\pi}{4(4-\pi)} = 0,9151$, а при $n=3$ $e(s'_0) = \frac{4}{3(3\pi-8)} = 0,9358$, т. с. значительно больше, чем эффективность оценки s' .

Для среднего отклонения $s_1 = \frac{1}{n} \sum |x_i - m|$ получим после простых выкладок

$$E\left(\sqrt{\frac{\pi}{2}} s_1\right) = \sigma, \quad D^2\left(\sqrt{\frac{\pi}{2}} s_1\right) = (\pi-2) \frac{\sigma^2}{2n},$$

так что $\sqrt{\frac{\pi}{2}} s_1$ является несмешенной оценкой для σ с эффективностью $\frac{1}{\pi-2} = 0,8760$.

Пример 4. Нерегулярный случай. Если плотность вероятности имеет точки разрыва, расположение которых зависит от параметра, то условия регулярности обычно нарушаются. В таких случаях часто можно найти несмешанные оценки „ненормально высокой“ точности, т. е. такие, что их дисперсия меньше, чем нижний предел для регулярных оценок, определяемый формулой (32.3.3а).

Рассмотрим, например, плотность вероятности $f(x; a)$, равную e^{x-a} при $x \geq a$ и 0 при $x < a$. В точке $x = a$ производная $\frac{\partial f}{\partial a}$ не существует, так что мы имеем нерегулярный случай. Как мы видели в параграфе 7.3, равенство $\int f dx = 1$ в этом случае нельзя дифференцировать обычным образом; действительно, $\int \frac{\partial f}{\partial a} dx = 1$. Если перейти от (32.3.5) к (32.3.6), то все n^2 слагаемых в левой части будут равны 1. Предполагая, что функции g и h удовлетворяют нашим условиям, мы получим вместо $D^2(a^*) \geq \frac{1}{n}$, что следовало бы из (32.3.3а), лишь более слабое неравенство $D^2(a^*) \geq \frac{1}{n^2}$.

В частности, для оценки $a^* = \min x_i - \frac{1}{n}$, где $\min x_i$ обозначает наименьшее из выборочных значений, мы получим плотность вероятности $nf(na^*, na - 1)$, так что $E(a^*) = a$, $D^2(a^*) = \frac{1}{n^3}$. Таким образом, a^* есть несмешанная оценка, дисперсия которой при всех $n > 1$ меньше, чем предел, определяемый формулой (32.3.3а).

Другой пример такого рода получим, рассматривая прямоугольное распределение и используя среднее значение или разность между крайними значениями выборки в качестве оценки для среднего значения или широты совокупности. В соответствии с формулами (28.6.8) и (28.6.9), в обоих случаях дисперсия имеет порядок n^{-2} и потому при большом n становится меньшей, чем предел, определяемый формулой (32.3.3а).

2. Дискретный тип. Рассмотрим дискретное распределение с точками сосредоточения массы u_1, u_2, \dots и соответствующими вероятностями $p_1(a), p_2(a), \dots$, где a может иметь любое значение из интервала A и u_i не зависит от a . Этот случай, во многом аналогичный предыдущему, будет рассмотрен лишь вкратце. Как и в предыдущем случае, мы рассматриваем оценку $a^* = a^*(x_1, \dots, x_n)$ со средним значением $E(a^*) = a + b(a)$.

Вероятность того, что *выборочная точка* в R_n с координатами x_1, \dots, x_n занимает положение M , определяемое равенствами $x_1 = u_{i_1}, \dots, x_n = u_{i_n}$, равна $p_{i_1}(a) \dots p_{i_n}(a)$. Однако точку M можно

определить также и другими n координатами, именно значением a^* , которое a^* имеет в точке M , и $n - 1$ новыми координатами y_1, \dots, y_{n-1} , определяющими положение точки M на гиперповерхности $a^* = a^*_y$. Если обозначить через $q_y(a)$ вероятность того, что a^* имеет значение a^* , и через $r_{y_1, \dots, y_{n-1} | y}(a)$ — условную вероятность множества значений y_1, \dots, y_{n-1} , соответствующих точке M , при данном y , то получим следующее соотношение, аналогичное формуле (32.3.2):

$$(32.3.8) \quad p_{l_1}(a) \dots p_{l_n}(a) = q_y(a) r_{y_1, \dots, y_{n-1} | y}(a).$$

Определим теперь случай регулярной оценки дискретного типа следующим условием: при всяком a из A существуют все производные $p'_i(a)$, $q'_y(a)$ и $r'_{y_1, \dots, y_{n-1} | y}(a)$, причем ряды $\sum_i p'_i(a)$ и т. п., соответствующие аналогичным интегралам в непрерывном случае, сходятся абсолютно и равномерно в A . В этом случае будем также называть a^* регулярной оценкой для a .

В каждом случае регулярной оценки дискретного типа справедливо следующее неравенство, аналогичное неравенству (32.3.3):

$$(32.3.9) \quad E(a^* - a)^2 \geq \frac{\left(1 + \frac{db}{da}\right)^2}{n \sum_i \left(\frac{d \log p_i}{da}\right)^2 p_i(2)}.$$

При любом a из A знак равенства достигается тогда и только тогда, когда выполняются следующие два условия, коль скоро $q_y(a) > 0$:

А) Условная вероятность $r_{y_1, \dots, y_{n-1} | y}(a)$ не зависит от a ;

Б) $\frac{d \log q_y}{da} = k(a^* - a)$, где k не зависит от y , но может зависеть от a .

В частном случае, когда a^* является несмещенной оценкой, каково бы ни было значение a из A , имеем $b(a) = 0$, и (32.3.9) обращается в

$$(32.3.9a) \quad D^2(a^*) \geq \frac{1}{n \sum_i \left(\frac{d \log p_i}{da}\right)^2 p_i}.$$

Эта теорема доказывается так же, как и соответствующая теорема для непрерывного случая: берем логарифмическую производную от обеих частей равенства (32.3.8), возводим в квадрат, умножаем

соответственно на левую и правую части равенства (32.3.8) и затем суммируем по всевозможным выборочным точкам M . Доказательство завершается применением леммы 2 предыдущего параграфа.

Как и в непрерывном случае, несмешенная оценка будет называться **эффективной**, если в (32.3.9а) достигается знак равенства. Определение **эффективности** оценки и замечания относительно корреляции между различными оценками распространяются на дискретный случай с очевидными видоизменениями.

Выражения (32.3.3а) и (32.3.9а) суть частные случаи более общего неравенства

$$D^2(a^*) \geq \frac{1}{\int\limits_{-\infty}^{\infty} \left(\frac{d \frac{\partial F}{\partial a}}{dF} \right)^2 dF},$$

которое справедливо при некоторых общих условиях даже для функций распределения $F(x; a)$, не принадлежащих к одному из двух простейших типов. Интеграл, стоящий в знаменателе правой части этого неравенства, известен под названием **интеграла Хеллингера** (см., например, Гобсон [17], I, стр. 609). Мы не будем здесь заниматься этим общим случаем, а перейдем к рассмотрению дальнейших примеров эффективных оценок.

Пример 5. Для **биномиального распределения** имеем $p_i = \binom{N}{i} p^i q^{N-i}$, где $a = p$ — параметр, подлежащий оценке, а N — известное целое число и $q = 1 - p$. Тогда

$$\sum_i \left(\frac{d \log p_i}{dp} \right)^2 p_i = \sum_0^N \left(\frac{i}{p} - \frac{N-i}{q} \right)^2 p_i = \frac{N}{pq}.$$

Таким образом, дисперсия любой регулярной несмешенной оценки p^* , вычисляемой по выборке в n значений, по меньшей мере равна $\frac{pq}{nN}$. В частности, для оценки $p^* = \bar{x} = \frac{1}{nN} \sum x_i$ получаем $E(p^*) = p$ и $D^2(p^*) = \frac{pq}{nN}$, так что эта оценка является эффективной.

Пример 6. Для **распределения Пуассона** с параметром λ имеем $p_i = \frac{\lambda^i}{i!} e^{-\lambda}$ и

$$\sum_i \left(\frac{d \log p_i}{d\lambda} \right)^2 p_i = \sum_0^\infty \left(\frac{i}{\lambda} - 1 \right)^2 p_i = \frac{1}{\lambda}.$$

Таким образом, дисперсия любой регулярной несмешенной оценки по меньшей

мере равна $\frac{\lambda}{n}$. В частности, для оценки $\lambda^* = \bar{x} = \sum \frac{x_i}{n}$ имеем $E(\lambda^*) = \lambda$ и $D^2(\lambda^*) = \frac{\lambda}{n}$, так что эта оценка является эффективной.

32.4. Достаточные оценки. Для того, чтобы регулярная несмешенная оценка a^* была эффективной, т. е. имела минимальную дисперсию, необходимо и достаточно выполнение обоих условий А) и В) предыдущего параграфа. Если же требовать выполнения лишь одного условия А), то получится более широкий класс оценок. В настоящем параграфе мы рассмотрим этот класс оценок, ограничиваясь распределениями непрерывного типа, так как дискретный случай исследуется совершенно аналогично.

Условие А) в непрерывном случае сводится к тому, чтобы условная плотность вероятности $h(\xi_1, \dots, \xi_{n-1} | a^*; a)$ не зависела от a , коль скоро $g(a^*; a) > 0$. Это означает, что распределение массы в бесконечно тонкой области, ограниченной двумя соседними гиперповерхностями a^* и $a^* + da^*$, должно быть независимым от a . В таком случае можно считать, что оценка a^* подытоживает все существенные сведения о параметре a , содержащиеся в выборке. Действительно, если известно значение a^* , соответствующее нашей выборке, скажем, a_0^* , то выборочная точка M должна лежать на гиперповерхности $a^* = a_0^*$, а условное распределение на этой гиперповерхности не зависит от a , так что дальнейшее уточнение положения точки M не даст никаких новых сведений о параметре a . Пользуясь терминологией Р. Фишера [89], [96], будем называть a^* достаточной оценкой. Поскольку якобиан J в (32.3.1) не зависит от a , то оценка a^* будет достаточной тогда и только тогда, когда

$$(32.4.1) \quad f(x_1; a) \dots f(x_n; a) = g(a^*; a) H(x_1, \dots, x_n),$$

где H не зависит от a .

Из самого характера условий А) и В) вытекает, что ожидать существования эффективных или достаточных оценок можно лишь для некоторых специальных классов совокупностей. Если существование таких оценок установлено, то их можно вычислять с помощью метода максимума правдоподобия (см. параграф 33.2), и между ними обнаруживаются важные связи.

Дальнейшие сведения об условиях существования и других свойствах эффективных и достаточных оценок читатель может найти в статьях

Фишера [89], [96], [103], [104] и др., Неймана [162], Неймана и Е. Пирсона [173], Куимена [141], Дармута [74], Дюге [76] и др.

В примерах 1, 2, 5 и 6 предыдущего параграфа были рассмотрены различные конкретные эффективные оценки. Все они, *a fortiori*, являются достаточными. Это можно исподвольно показать в каждом случае, рассматривая преобразование, заменяющее исходные выборочные переменные оценкой α^* и $n - 1$ новыми переменными, выбранными подходящим образом; следует лишь проверить, что удовлетворяется условие А). Мы рекомендуем читателю провести эти преобразования подробно. (См. также аналогичный случай в параграфе 32.6, пример 1.) Оценка s_0' , определенная в примере 3 параграфа 32.3 является примером регулярной несмещенной оценки, удовлетворяющей условию А), но не удовлетворяющей условию В), т. е. примером достаточной, но не эффективной оценки. Другие примеры такого рода будут приведены в параграфе 33.3 (см. пример 3). Таким образом, класс достаточных оценок является действительно более широким, чем класс эффективных оценок.

Приведенное выше определение достаточной оценки применимо к классу регулярных несмещенных оценок, но его можно исподвольно распространить на класс всех регулярных оценок как несмещенных, так и смешанных. После этого обобщения из определения достаточных оценок будет немедленно следовать, что свойство оценки быть достаточной инвариантно относительно преобразования параметра. Таким образом, если α^* есть достаточная оценка параметра α и если α заменяется новым параметром $\varphi(\alpha)$, то $\varphi(\alpha^*)$ будет достаточной оценкой для $\varphi(\alpha)$. Соответствующей теоремы для эффективных оценок не существует.

32.5. Асимптотически эффективные оценки. В предыдущих параграфах мы считали объем выборки n фиксированным числом ≥ 1 . Предположим теперь, что регулярная несмещенная оценка $\alpha^* = \alpha^*(x_1, \dots, x_n)$ определена для всех достаточно больших значений n , и рассмотрим асимптотическое поведение α^* при $n \rightarrow \infty$.

Если α^* сходится по вероятности к α при $n \rightarrow \infty$, то α^* является *состоятельной оценкой* для α (см. параграф 27.6). В главах 27—29 мы видели (см., например, параграфы 27.7 и 28.4), что во многих важных случаях стандартное отклонение оценки α^* имеет при больших n порядок $n^{-1/2}$, так что $D(\alpha^*) \sim cn^{-1/2}$, где c — некоторая постоянная. Если α^* — несмещенная оценка со стандартным отклонением такого вида, то очевидно, что α^* — состоятельная оценка (см. параграф 20.4). Далее, в этом случае эффективность $e(\alpha^*)$, определяемая формулой (32.3.7), стремится при $n \rightarrow \infty$ к определенному пределу:

$$(32.5.1) \quad \lim_{n \rightarrow \infty} e(\alpha^*) = e_0(\alpha^*) = \frac{1}{c^2 E \left(\frac{\partial \log f}{\partial \alpha} \right)^2}$$

В дискретном случае получается аналогичное выражение. Этот предел называется *асимптотической эффективностью* оценки a^* . Очевидно, $0 \leq e_0(a^*) \leq 1$.

Рассмотрим, далее, важный случай оценки a^* , смещенной или несмещенной, которая при больших n асимптотически нормальна $(a, \frac{c}{\sqrt{n}})$.

В параграфе 28.4 мы видели, что этот случай может иметь место даже тогда, когда $E(a^*)$ или $D(a^*)$ не существуют. Однако, если n велико, то для практических целей можно считать распределение величины a^* эквивалентным нормальному распределению со средним значением α и стандартным отклонением $\frac{c}{\sqrt{n}}$ и, соответственно, даже в этом

случае называть величину $e_0(a^*)$, определяемую правой частью равенства (32.5.1), асимптотической эффективностью оценки a^* .

Если $e_0(a^*) = 1$, то будем называть a^* *асимптотически эффективной оценкой* параметра α . При довольно общих условиях асимптотически эффективная оценка может быть найдена с помощью метода максимума правдоподобия (см. параграф 33.3).

Пример 1. Для *нормального распределения* медиана выборки может быть использована в качестве оценки среднего значения m , и, в силу параграфа 28.5, эта оценка имеет асимптотическую эффективность $\frac{2}{\pi} = 0,6366$. Та-

ким образом, если мы оцениваем m , вычисляя медиану для выборки, например, в $n = 10\,000$ элементов, то получаемая оценка имеет такую же точность, как оценка, получаемая вычислением среднего значения \bar{x} по выборке лишь в $\frac{2n}{\pi} = 6366$ элементов. Тем не менее, медиана иногда употребляется на практике, так как ее можно вычислить очень быстро.

В качестве оценки для m можно использовать также среднее значение v -х верхнего и нижнего значений выборки. В силу формулы (28.6.17), асимптотическая эффективность такой оценки равна нулю.

Если в нормальном распределении известно m и требуется оценить дисперсию σ^2 или стандартное отклонение σ , то можно использовать различные оценки, связанные с выборочной дисперсией s^2 . Мы уже встречались с асимптотически эффективными оценками такого рода в примерах 2–3 параграфа 32.3. В качестве оценки для σ можно использовать также разность между v -ми верхним и нижним значениями выборки, умноженную на подходящую постоянную. В соответствии с (28.6.18), асимптотическая эффективность такой оценки будет равна нулю. Поэтому использование такой оценки в больших выборках приведет к „потере информации“ большей даже, чем в случае рассмотренной выше медианы выборки. Тем не менее, на практике часто употребляются оценки для m и для σ , основанные на v -х значениях выборки,

так как эти оценки вычисляются очень просто, а при небольших значениях n потеря информации незначительна (см. статьи, упомянутые в этой связи в параграфе 28.6).

Пример 2. Для *распределения Коши* с плотностью вероятности $f(x; \mu) = \frac{1}{\pi [1 + (x - \mu)^2]}$ имеем

$$E\left(\frac{\partial \log f}{\partial \mu}\right)^2 = \frac{4}{\pi} \int_{-\infty}^{\infty} \frac{(x - \mu)^2 dx}{[1 + (x - \mu)^2]^3} = \frac{1}{2}.$$

Таким образом, дисперсия любой регулярной несмешенной оценки для μ по меньшей мере равна $\frac{2}{n}$. Согласно параграфу 19.2, выборочное среднее значение \bar{x} имеет ту же самую плотность вероятности $f(x; \mu)$, так что среднее значение не является состоятельной оценкой для μ . Это же справедливо относительно среднего арифметического n -х верхнего и нижнего значений выборки (см. (28.6.11)). С другой стороны, выборочная медиана, согласно параграфу 28.5, асимптотически нормальна $\left(\mu; \frac{\pi}{12\sqrt{n}}\right)$ и имеет, таким образом, асимптотическую эффективность $\frac{2}{n} : \frac{\pi^2}{4n} = \frac{8}{\pi^2} = 0,8106$.

32.6. Случай двух неизвестных параметров. Покажем вкратце, как можно обобщить понятия и теоремы предыдущих параграфов на случаи нескольких неизвестных параметров. При этом достаточно привести точные формулировки результатов для непрерывных распределений, так как соответствующие результаты в дискретном случае получаются по аналогии. Для упрощения записей мы ограничимся в дальнейшем случаем *несмешенных* оценок.

В настоящем параграфе мы рассмотрим распределение с двумя неизвестными параметрами α и β , имеющее заданную плотность вероятности $f(x; \alpha, \beta)$. По выборке в n значений x_1, \dots, x_n , извлеченных из этого распределения, мы вычисляем две функции $\alpha^* = \alpha^*(x_1, \dots, x_n)$ и $\beta^* = \beta^*(x_1, \dots, x_n)$, которые будут являться несмешенными оценками, соответственно, для α и β . Затем рассмотрим преобразование в пространстве выборок R_n , при котором старые переменные x_1, \dots, x_n заменяются новыми переменными $\alpha^*, \beta^*, \xi_1, \dots, \xi_{n-2}$. Для этого преобразования имеем следующие соотношения, аналогичные

формулам (32.3.1) и (32.3.2):

$$J \prod_{i=1}^n f(x_i; \alpha, \beta) = g(\alpha^*, \beta^*; \alpha, \beta) h(\xi_1, \dots, \xi_{n-2} | \alpha^*, \beta^*; \alpha, \beta),$$

$$\prod_{i=1}^n f(x_i; \alpha, \beta) dx_i =$$

$$= g(\alpha^*, \beta^*; \alpha, \beta) h(\xi_1, \dots, \xi_{n-2} | \alpha^*, \beta^*; \alpha, \beta) d\alpha^* d\beta^* d\xi_1 \dots d\xi_{n-2}.$$

Здесь g — совместная плотность вероятности величин α^* и β^* , а h — условная плотность вероятности величин ξ_1, \dots, ξ_{n-2} при данных значениях α^* и β^* ; наконец, J — якобиан, не зависящий от α и β .

Случай регулярной оценки определяется теперь как случай, в котором плотности вероятности f , g и h удовлетворяют условиям регулярности, установленным в параграфе 32.3, по обоим параметрам α и β .

Действуя так же, как в параграфе 32.3, и бери вместо частных производных по α полный дифференциал по α и β , мы получим (см. Дюге [76]):

$$(32.6.1) \quad n \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial \alpha} d\alpha + \frac{\partial \log f}{\partial \beta} d\beta \right)^2 f d\alpha \geq$$

$$\geq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial \log g}{\partial \alpha} d\alpha + \frac{\partial \log g}{\partial \beta} d\beta \right)^2 g d\alpha^* d\beta^*,$$

причем знак равенства достигается тогда и только тогда, когда условная плотность вероятности h не зависит от α и β , коль скоро $g > 0$. Если это условие выполняется, то можно считать, что оценки α^* и β^* подыскивают всю содержащуюся в выборке существенную информацию относительно α и β . Обобщая определение, данное в параграфе 32.4, мы будем называть α^* и β^* *совместно-достаточными оценками* для α и β .

Правая и левая части равенства (32.6.1) суть квадратичные формы относительно $d\alpha$ и $d\beta$. В силу однородности, такое же неравенство между значениями квадратичных форм будет справедливо, если заменить $d\alpha$ и $d\beta$ какими-нибудь величинами u и v , так что (32.6.1) можно записать в виде

$$(32.6.2) \quad n \left[E \left(\frac{\partial \log f}{\partial \alpha} \right)^2 u^2 + 2E \left(\frac{\partial \log f}{\partial \alpha} \frac{\partial \log f}{\partial \beta} \right) uv + E \left(\frac{\partial \log f}{\partial \beta} \right)^2 v^2 \right] \geq$$

$$\geq E \left(\frac{\partial \log g}{\partial \alpha} \right)^2 u^2 + 2E \left(\frac{\partial \log g}{\partial \alpha} \frac{\partial \log g}{\partial \beta} \right) uv + E \left(\frac{\partial \log g}{\partial \beta} \right)^2 v^2.$$

Рассмотрим теперь неравенство (32.2.1), выражающее основной результат леммы 1 параграфа 32.2, и допустим, что $\Phi(a) = a$. Тогда (32.2.1) можно записать в виде неравенства между двумя квадратичными формами от одной переменной:

$$E \left(\frac{\partial \log g}{\partial x} \right)^2 u^2 \geq E(a^* - a)^2,$$

где $g = g(a^*; a)$ — плотность вероятности со средним значением $E(a^*) = a$, а квадратичная форма в правой части неравенства — обратная к форме $E(a^* - a)^2 u^2$. Выраженная в такой форме лемма 1 может быть обобщена на плотности вероятности, содержащие несколько параметров (см. Крамер [72]; подробного доказательства такой обобщенной леммы мы здесь не приводим). В случае двух параметров обобщенная лемма утверждает, что правая часть неравенства (32.6.2) по меньшей мере равна квадратичной форме, обратной к форме

$$E(a^* - a)^2 u^2 + 2E[(a^* - a)(\beta^* - \beta)]uv + E(\beta^* - \beta)^2 v^2 = \\ = \sigma_1^2 u^2 + 2\rho\sigma_1\sigma_2 uv + \sigma_2^2 v^2,$$

где σ_1, σ_2 и ρ обозначают стандартные отклонения и коэффициент корреляции для величин a^* и β^* , так что

$$(32.6.3) \quad E \left(\frac{\partial \log g}{\partial x} \right)^2 u^2 + 2E \left(\frac{\partial \log g}{\partial x} \frac{\partial \log g}{\partial \beta} \right) uv + E \left(\frac{\partial \log g}{\partial \beta} \right)^2 v^2 \geq \\ \geq \frac{1}{1-\rho^2} \left(\frac{u^2}{\sigma_1^2} - \frac{2\rho uv}{\sigma_1\sigma_2} + \frac{v^2}{\sigma_2^2} \right).$$

Эллипс рассеяния для совместного распределения величин a^* и β^* имеет уравнение (см. (21.10.1))

$$(32.6.4) \quad \frac{1}{1-\rho^2} \left(\frac{(u-a)^2}{\sigma_1^2} - \frac{2\rho(u-a)(v-\beta)}{\sigma_1\sigma_2} + \frac{(v-\beta)^2}{\sigma_2^2} \right) = 4.$$

Таким образом, неравенства (32.6.2) и (32.6.3) означают, что фиксированный эллипс

$$(32.6.5) \quad n \left[E \left(\frac{\partial \log f}{\partial x} \right)^2 (u - a)^2 + 2E \left(\frac{\partial \log f}{\partial x} \frac{\partial \log f}{\partial \beta} \right) (u - a)(v - \beta) + E \left(\frac{\partial \log f}{\partial \beta} \right)^2 (v - \beta)^2 \right] = 4$$

лежит целиком внутри эллипса рассеяния любой пары регулярных несмещенных оценок a^*, β^* . Это является обобщением неравенства (30.3.3а) на случай двух параметров.

Если в обоих соотношениях (33.6.2) и (32.6.3) достигается знак равенства, то мы будем говорить, что α^* и β^* суть *совместно-эффективные оценки* для α и β . В этом случае два эллипса (32.6.4) и (32.6.5) совпадают, и совместное распределение величин α^* и β^* имеет *меньшее рассеяние* (см. параграф 21.10), чем распределение любой неэффективной пары оценок.

Рассмотрим теперь какую-нибудь пару совместно-эффективных оценок α_0^* и β_0^* . Дисперсии этих оценок и коэффициент корреляции между ними можно получить, отыскав квадратичную форму, обратную к форме в левой части равенства (32.6.5):

$$\mathbf{D}^2(\alpha_0^*) = \frac{1}{n\Delta} E \left(\frac{\partial \log f}{\partial \beta} \right)^2, \quad \mathbf{D}^2(\beta_0^*) = \frac{1}{n\Delta} E \left(\frac{\partial \log f}{\partial \alpha} \right)^2,$$

$$\rho(\alpha_0^*, \beta_0^*) = - \frac{E \left(\frac{\partial \log f}{\partial \alpha} \frac{\partial \log f}{\partial \beta} \right)}{\sqrt{E \left(\frac{\partial \log f}{\partial \alpha} \right)^2 E \left(\frac{\partial \log f}{\partial \beta} \right)^2}},$$

где

$$\Delta = E \left(\frac{\partial \log f}{\partial \alpha} \right)^2 E \left(\frac{\partial \log f}{\partial \beta} \right)^2 - E^2 \left(\frac{\partial \log f}{\partial \alpha} \frac{\partial \log f}{\partial \beta} \right)$$

Отсюда получаем, например,

$$\mathbf{D}^2(\alpha_0^*) = \frac{1}{1 - \rho^2(\alpha_0^*, \beta_0^*)} \cdot \frac{1}{n E \left(\frac{\partial \log f}{\partial \alpha} \right)^2}.$$

Если $E \left(\frac{\partial \log f}{\partial \alpha} \frac{\partial \log f}{\partial \beta} \right) \neq 0$, то дисперсия величины α_0^* превышает дисперсию эффективной оценки для случая, когда α — единственный неизвестный параметр (см. (32.3.3а)). В случае двух неизвестных параметров часто бывает необходимо оценить лишь один из параметров, например α , и тогда возникает вопрос, нельзя ли найти другую пару регулярных несмещенных оценок, для которой $\mathbf{D}^2(\alpha^*) < \mathbf{D}^2(\alpha_0^*)$, причем безразлично, сколь велико при этом $\mathbf{D}^2(\beta^*)$.

Однако, поскольку эллипс (32.6.5) лежит целиком внутри эллипса (32.6.4), максимальное значение абсциссы для всех точек первого эллипса равно, самое большее, соответствующему максимуму для второго эллипса. Поэтому после некоторых выкладок мы получаем неравенство

$$\mathbf{D}^2(\alpha^*) = \sigma_1^2 \geq \frac{1}{n\Delta} E \left(\frac{\partial \log f}{\partial \beta} \right)^2 = \mathbf{D}^2(\alpha_0^*),$$

из которого следует, что *невозможно найти оценку для α , которая была бы „лучше“, чем $\hat{\alpha}_0^*$.*

Отношение двумерной дисперсии (см. параграф 22.7) пары совместно-эффективных оценок $\hat{\alpha}_0^*, \hat{\beta}_0^*$ к соответствующей дисперсии любой пары регулярных несмещенных оценок α^*, β^* мы будем называть *совместной эффективностью* оценок α^* и β^* и обозначать через $e(\alpha^*, \beta^*)$. Эта величина равна квадрату отношения площадей эллипсов (32.6.5) и (32.6.4), т. е., согласно (11.12.3),

$$e(\alpha^*, \beta^*) = \frac{1}{n^2 \lambda \sigma_1^2 \sigma_2^2 (1 - \rho^2)}.$$

Понятия *асимптотической эффективности* и *асимптотически-эффективной оценки* (см. параграф 32.5) непосредственно переносятся на рассматриваемый случай двух оцениваемых параметров.

Как и в параграфе 32.3, все полученные выше результаты остаются верными для выборок из многомерной совокупности с плотностью вероятности $f(x_1, \dots, x_n; \alpha, \beta)$, содержащей два неизвестных параметра.

Пример 1. Если в нормальном распределении неизвестны оба параметра $\alpha = m$ и $\beta = \sigma^2$, то имеем (см. параграф 32.3, примеры 1–2)

$$E\left(\frac{\partial \log f}{\partial m}\right)^2 = \frac{1}{\sigma^2}, \quad E\left(\frac{\partial \log f}{\partial m} \frac{\partial \log f}{\partial \sigma^2}\right) = 0; \quad E\left(\frac{\partial \log f}{\partial \sigma^2}\right)^2 = \frac{1}{2\sigma^4},$$

так что в этом случае оптимальный эллипс (32.6.5) имеет уравнение

$$\frac{(u - m)^2}{\sigma^2} + \frac{(v - \sigma^2)^2}{2\sigma^4} = \frac{4}{n}.$$

Следовательно, этот фиксированный эллипс лежит внутри эллипса рассеяния совместного распределения любой пары регулярных несмещенных оценок для m и σ^2 . В частности, для пары оценок $\alpha^* = \bar{x}$ и $\beta^* = \frac{n}{n-1}s^2$ соотношение (29.3.6) представляет собой преобразование, с помощью которого выборочные переменные x_1, \dots, x_n заменяются новыми переменными \bar{x} , s и z_1, \dots, z_{n-2} . Последний множитель в выражении для плотности вероятности новых величин есть условная плотность вероятности величин z_1, \dots, z_{n-2} и не зависит от неизвестных параметров m и σ^2 (а также от \bar{x} и s , но последнее не имеет значения для нашей задачи). Отсюда следует, что \bar{x} и $\frac{n}{n-1}s^2$ быть совместно-достаточными оценками для m и σ^2 . Далее, имеем

$$D^2(\bar{x}) = \frac{\sigma^2}{n}, \quad D^2\left(\frac{n}{n-1}s^2\right) = \frac{2\sigma^4}{n-1}, \quad \rho\left(\bar{x}, \frac{n}{n-1}s^2\right) = 0.$$

Таким образом, эллипс рассеяния для \bar{x} и $\frac{n}{n-1}s^2$ имеет уравнение

$$\frac{(u-m)^2}{\sigma^2} + \frac{n-1}{n} \frac{(v-\sigma^2)^2}{2\sigma^4} = \frac{4}{n}$$

Квадрат отношения площадей эллипсов дает для совместной эффективности оценок значение $\frac{n-1}{n}$. При $n \rightarrow \infty$ эффективность стремится к единице, так что \bar{x} и $\frac{n}{n-1}s^2$ суть асимптотически-эффективные оценки для m и σ^2 .

Это же справедливо, конечно, и для \bar{x} и s^2 , хотя s^2 не является несмешенной.

Пример 2. Рассмотрим двумерную нормальную плотность вероятности (21.12.1) с известными значениями σ_1 , σ_2 и ρ , в то время как $\alpha = m_1$ и $\beta = m_2$ — неизвестные параметры. По выборке в n пар значений $(x_1, y_1), \dots, (x_n, y_n)$ мы образуем оценки $\hat{\alpha} = \bar{x}$ и $\hat{\beta} = \bar{y}$. Легко показать, что в этом случае эллипс рассеяния оценок \bar{x} и \bar{y} совпадает с фиксированным эллипсом (32.6.5), уравнение которого имеет вид

$$\frac{n}{1-\rho^2} \left(\frac{(u-m_1)^2}{\sigma_1^2} - \frac{2\rho(u-m_1)(v-m_2)}{\sigma_1\sigma_2} + \frac{(v-m_2)^2}{\sigma_2^2} \right) = 4.$$

Таким образом, \bar{x} и \bar{y} суть совместно-эффективные (и *a fortiori* совместно-достаточные) оценки для m_1 и m_2 .

32.7. Случай нескольких неизвестных параметров. Результаты предыдущих параграфов можно обобщить на распределения, содержащие любое количество неизвестных параметров. Если $\alpha_1^*, \dots, \alpha_k^*$ — какие-нибудь регулярные несмешенные оценки для k неизвестных параметров $\alpha_1, \dots, \alpha_k$, то таким же способом, как и при $k=2$, можно показать, что k -мерный эллипсоид

$$(32.7.1) \quad n \sum_{i,j=1}^k E \left(\frac{\partial \log f}{\partial \alpha_i} \frac{\partial \log f}{\partial \alpha_j} \right) (u_i - \alpha_i)(u_j - \alpha_j) = k+2$$

лежит целиком внутри эллипса рассеяния (см. параграф 22.7) совместного распределения величин $\alpha_1^*, \dots, \alpha_k^*$. В предельном случае, когда оба эти эллипса совпадают, мы будем говорить, что $\alpha_1^*, \dots, \alpha_k^*$ суть *совместно-эффективные оценки* для $\alpha_1, \dots, \alpha_k$. Таким образом, распределение множества совместно-эффективных оценок имеет *меньшее рассеяние* (см. параграф 22.7), чем распределение любого множества неэффективных оценок. Матрица вторых моментов для множества совместно-

эффективных оценок — обратна матрице квадратичной формы в левой части равенства (32.7.1), как это было показано для случая двух параметров. Понятия достаточности, эффективности и т. д. вводятся так же, как и в случае $k=2$.

Рассмотрим, например, двумерную нормальную плотность вероятности с пятью неизвестными параметрами $m_1, m_2, \mu_{20}, \mu_{11}$ и μ_{02} . По выборке в n пар значений $(x_1, y_1), \dots, (x_n, y_n)$ мы получаем несмешанные оценки $\bar{x}, \bar{y}, \frac{n}{n-1}m_{20}, \frac{n}{n-1}m_{11}$ и $\frac{n}{n-1}m_{02}$ (см. параграф 29.6). Матрицу вторых моментов для совместного распределения этих пяти оценок можно получить, например, с помощью выражения (29.6.3) для совместной характеристической функции оценок. Далее, коэффициенты в уравнении (32.7.1) оптимального эллипса можно найти, подставляя выражение для плотности вероятности в (32.7.1) и осуществляя интегрирование. С помощью простых, хотя и несколько громоздких вычислений мы убедимся, что совместная эффективность этих пяти оценок равна $\left(\frac{n-1}{n}\right)^3$. Эта величина стремится к единице при $n \rightarrow \infty$, так что наши оценки асимптотически-эффективны.

32.8. Обобщение. В настоящей главе мы имели дело с проблемой оценки некоторых параметров по множеству значений, извлеченных независимым образом из фиксированного распределения. Однако наши методы применимы и в более общих условиях. Рассмотрим, например, следующую проблему:

Величины x_1, \dots, x_n имеют в R_n совместное распределение с плотностью вероятности $f(x_1, \dots, x_n; \alpha)$, где функция f задана, но содержит неизвестный параметр α . Известна наблюденная точка $x = (x_1, \dots, x_n)$, и требуется найти «наилучшую возможную» оценку $\alpha^* = \alpha^*(x_1, \dots, x_n)$ для α по наблюденным координатам x_i .

В частном случае, когда совместная плотность вероятности имеет вид $f(x_1; \alpha) \dots f(x_n; \alpha)$, поставленная проблема уже решена в параграфе 32.3, где x_i рассматривались как независимые величины с одним и тем же распределением. Общая проблема включает также случаи, когда x_i коррелированы или когда они образуют несколько независимых выборок из различных распределений. Но и в этом общем случае мы будем называть точку $x = (x_1, \dots, x_n)$ *выборочной точкой*, принадлежащей *пространству выборок* R_n .

Рассмотрим такое же преобразование переменных в пространстве выборок, как в (32.3.1) и (32.3.2). Однако теперь мы должны ввести

в формулы этого преобразования общее выражение для совместной плотности вероятности, так что, например, равенство (32.3.2) будет иметь вид

$$f(x_1, \dots, x_n; \alpha) dx_1 \dots dx_n = \\ = g(\alpha^*; \alpha) h(\xi_1, \dots, \xi_{n-1} | \alpha^*; \alpha) d\alpha^* d\xi_1 \dots d\xi_{n-1}.$$

Все рассуждения параграфов 32.3 — 32.5 (непрерывный случай) повторяются почти без изменения, и таким способом понятия несмешанных, эффективных, достаточных оценок и т. д. обобщаются на рассматриваемый общий случай. Так, например, обобщенная форма неравенства (32.3.3а) для дисперсии несмешанной оценки имеет вид

$$D^2(\alpha^*) \geq \left[\int \dots \int \left(\frac{\partial \log f(x_1, \dots, x_n; \alpha)}{\partial \alpha} \right)^2 f(x_1, \dots, x_n; \alpha) dx_1 \dots dx_n \right]^{-1} \\ = \left[E \left(\frac{\partial \log f}{\partial \alpha} \right)^2 \right]^{-1},$$

и если здесь достигается знак равенства, то мы называем α^* *эффективной* оценкой. Если условная плотность вероятности h не зависит от α , мы называем α^* *достаточной* оценкой, и т. д.

Такое же обобщение можно, очевидно, осуществить и для дискретных распределений, а также для распределений, содержащих несколько неизвестных параметров.

ГЛАВА 33

МЕТОДЫ НАХОЖДЕНИЯ ОЦЕНОК

33.1. Метод моментов. Теперь мы перейдем к рассмотрению некоторых общих методов нахождения оценок для параметров распределения по выборочным значениям.

Самым первым общим методом, предложенным для этой цели, является *метод моментов*, введенный К. Пирсоном ([180], [182], [184] и другие работы) и интенсивно используемый им и его школой. Этот метод заключается в приравнивании определенного количества выборочных моментов к соответствующим моментам распределения, являющимся функциями от неизвестных параметров. Рассматривая количество моментов, равное числу подлежащих оценке параметров, и решая полученные уравнения относительно этих параметров, мы получаем искомые оценки. На практике этот метод часто приводит к сравнительно простым вычислениям.

Оценки, получаемые таким способом с помощью множества выборочных значений, суть функции от выборочных моментов, и некоторые свойства выборочных распределений этих оценок можно получить из результатов глав 27—28. Так, например, мы видели (см., в частности, параграфы 27.7 и 28.4), что при довольно общих условиях распределение оценки такого рода при больших n асимптотически нормально и что среднее значение такой оценки отличается от истинного значения параметра на величину порядка n^{-1} , а стандартное отклонение асимптотически имеет вид $\frac{c}{\sqrt{n}}$. Смещение такой оценки часто удается исключить с помощью простых поправок, и при этом получается несмещенная оценка для параметра (см. параграф 27.6).

Таким образом, при некоторых общих условиях метод моментов позволяет найти оценки, для которых существует асимптотическая эффективность, определенная в параграфе 32.5 (или соответствующая величина в случае нескольких параметров). Однако Р. Фишер [89] указал, что получаемая в методе моментов асимптотическая эффективность часто значительно меньше единицы, так что оценки, найденные с помощью метода моментов, с точки зрения эффективности не являются „наилучшими“ из возможных, т. е. в больших выборках они имеют не наименьшую возможную дисперсию. Тем не менее, метод моментов часто очень удобен для практических целей. Иногда оценки, получаемые с помощью метода моментов, можно принять в качестве первого приближения, по которому можно определять другими методами дальнейшие оценки более высокой эффективности.

В частном случае нормального распределения метод моментов дает для неизвестных параметров m и s^2 оценки \bar{x} и s^2 . Исключая смещение, мы получаем несмещенные и асимптотически эффективные (см. параграф 32.6, пример 1) оценки \bar{x} и $\frac{n}{n-1} s^2$. Фишер [89] показал, что в этом отношении нормальное распределение является исключением среди распределений системы Пирсона (см. параграф 19.4), так как в других распределениях этой системы асимптотическая эффективность, как правило, получается меньше 1. Некоторые примеры мы приведем в параграфе 33.3.

33.2. Метод максимума правдоподобия. Наиболее важным с теоретической точки зрения общим методом нахождения оценок является *метод максимума правдоподобия*. В частных случаях этот метод применялся еще Гауссом [16], но как общий метод для нахождения оценок он впервые был предложен Р. Фишером в краткой статье [87].

в 1912 г. и получил свое дальнейшее развитие в ряде его работ [89], [96], [103], [104] и т. д.). Важные результаты были получены и другими авторами; в этом отношении мы особенно рекомендуем читателю работу Дюге [76].

Используя обозначения параграфа 32.3, определим *функцию правдоподобия* L для выборки в n значений из совокупности *непрерывного* типа соотношением

$$(33.2.1a) \quad L(x_1, \dots, x_n; \alpha) = f(x_1; \alpha) \dots f(x_n; \alpha);$$

в *дискретном* случае положим

$$(33.2.1b) \quad L(x_1, \dots, x_n; \alpha) = p_{i_1}(\alpha) \dots p_{i_n}(\alpha).$$

Если выборочные значения фиксированы, то функция правдоподобия L является функцией от единственного переменного аргумента α . Метод максимума правдоподобия заключается в том, что в качестве оценки для неизвестного параметра совокупности α принимается такое значение α , при котором L достигает наибольшего возможного значения. Поскольку $\log L$ достигает максимума при том же значении α , что и L , следует решать относительно α *уравнение правдоподобия*.

$$(33.2.2) \quad \frac{\partial \log L}{\partial \alpha} = 0.$$

При этом следует отбросить все корни вида $\alpha = \text{const}$ и считать *решением* лишь тот корень, который действительно зависит от выборочных значений x_1, \dots, x_n . Каждое решение уравнения правдоподобия будем называть *оценкой максимального правдоподобия* для α .

В этом параграфе мы рассмотрим некоторые свойства метода максимума правдоподобия для выборок фиксированного объема n ; в следующем параграфе будет исследовано асимптотическое поведение оценок максимального правдоподобия при больших значениях n . Важность метода выясняется следующими двумя предложениями:

Если для параметра α существует эффективная оценка α^ , то уравнение правдоподобия имеет единственное решение α^* .*

Если для параметра α существует достаточная оценка α^ , то каждое решение уравнения правдоподобия является функцией от α^* .*

Эти утверждения достаточно доказать для непрерывного случая; доказательство для дискретного случая получится с помощью очевидных видоизменений. Если существует эффективная оценка α^* ,

то условия А) и В), установленные в связи с формулой (32.3.3а), выполнены, и мы получаем по формуле (32.3.5)

$$\frac{\partial \log L}{\partial \alpha} = \sum_1^n \frac{\partial \log f(x_i; \alpha)}{\partial \alpha} = \frac{\partial \log g}{\partial \alpha} = k(\alpha^* - \alpha),$$

где k не зависит от выборочных значений, но может зависеть от α . В соответствии с тем, что мы условились понимать под решением уравнения правдоподобия (32.2.2), это уравнение имеет единственное решение $\alpha = \alpha^*$.

Далее, если существует достаточная оценка α^* , то удовлетворяется условие А) параграфа 32.3, и, в силу формулы (32.3.5), уравнение правдоподобия принимает вид

$$\frac{\partial \log L}{\partial \alpha} = \frac{\partial \log g(x^*, \alpha)}{\partial \alpha} = 0.$$

Функция g зависит лишь от двух аргументов x^* и α , так что каждое решение будет функцией от α^* .

Приведенные выше определения и теоремы можно непосредственно обобщить на случай нескольких неизвестных параметров и на случай выборок из многомерных распределений. Так, например, для непрерывного распределения с двумя неизвестными параметрами α и β функцией правдоподобия является $L(x_1, \dots, x_n; \alpha, \beta) = \prod f(x_i; \alpha, \beta)$, и оценки максимального правдоподобия для α и β получаются при совместном решении уравнений $\frac{\partial \log L}{\partial \alpha} = 0$, $\frac{\partial \log L}{\partial \beta} = 0$ относительно α и β . Если существует пара совместно-эффективных оценок α^* и β^* , то уравнения правдоподобия имеют единственное решение $\alpha = \alpha^*$, $\beta = \beta^*$.

Метод максимума правдоподобия можно применять даже при общих условиях, рассмотренных в параграфе 32.8. В этом случае метод заключается в том, что в качестве оценки берется то значение α , при котором совместная плотность вероятности $f(x_1, \dots, x_n; \alpha)$ принимает максимальное возможное значение при фиксированных x_i .

В следующем параграфе мы приведем некоторые примеры.

33.3. Асимптотические свойства оценок максимального правдоподобия. Теперь мы исследуем асимптотическое поведение оценок максимального правдоподобия при больших значениях n . Сначала рассмотрим случай единственного неизвестного параметра α .

Мы покажем, что при некоторых общих условиях уравнение правдоподобия (33.2.2) имеет решение, сходящееся по вероятности при $n \rightarrow \infty$ к истинному значению параметра a . Это решение является асимптотически-нормальной и асимптотически-эффективной оценкой для a .

Как и выше, достаточно привести доказательство для случая непрерывного распределения с плотностью вероятности $f(x; a)$. Воспользуемся методом доказательства, предложенным Дюге [76]. Предположим, что выполнены следующие условия:

1) При каждом a , принадлежащем некоторому невырожденному интервалу A , для почти всех x существуют производные $\frac{\partial \log f}{\partial x}$, $\frac{\partial^2 \log f}{\partial x^2}$ и $\frac{\partial^3 \log f}{\partial x^3}$.

2) При каждом a из A имеем $\left| \frac{\partial f}{\partial a} \right| < F_1(x)$, $\left| \frac{\partial^2 f}{\partial a^2} \right| < F_2(x)$ и $\left| \frac{\partial^3 \log f}{\partial a^3} \right| < H(x)$, где функции F_1 и F_2 интегрируемы на $(-\infty, \infty)$ и $\int_{-\infty}^{\infty} H(x) f(x, a) dx < M$, причем M не зависит от a .

3) При каждом a из A интеграл $\int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial a} \right)^2 f dx$ конечен и положителен.

Обозначим через a_0 неизвестное истинное значение параметра a в распределении, из которого производится выбор, и предположим, что a_0 есть внутренняя точка интервала A . Покажем сначала, что уравнение правдоподобия имеет решение, сходящееся по вероятности к a_0 . Указывая индексом 0, что следует положить $a = a_0$, для каждого a из A получаем

$$\frac{\partial \log f}{\partial a} = \left(\frac{\partial \log f}{\partial x} \right)_0 + (a - a_0) \left(\frac{\partial^2 \log f}{\partial x^2} \right)_0 + \frac{1}{2} \theta (a - a_0)^2 H(x),$$

где $|\theta| < 1$. Таким образом, после умножения на $\frac{1}{n}$ уравнение правдоподобия (33.2.2) можно записать в виде

$$(33.3.1) \quad \frac{1}{n} \frac{\partial \log L}{\partial a} = B_0 + B_1(a - a_0) + \frac{1}{2} \theta B_2(a - a_0)^2 = 0,$$

где, записывая f_t вместо $f(x_i, \alpha)$, имеем

$$(33.3.2) \quad B_0 = \frac{1}{n} \sum_1^n \left(\frac{\partial \log f_t}{\partial \alpha} \right)_0, \quad B_1 = \frac{1}{n} \sum_1^n \left(\frac{\partial^2 \log f_t}{\partial \alpha^2} \right)_0,$$

$$B_2 = \frac{1}{n} \sum_1^n H(x_t).$$

Величины B_i суть функции от случайных величин x_1, \dots, x_n , и остается показать, что с вероятностью, стремящейся к 1 при $n \rightarrow \infty$, уравнение (33.3.1) имеет корень, заключенный в пределах $\alpha_0 \pm \delta$, где δ — сколь угодно малое положительное число.

Рассмотрим поведение величин B_i при больших значениях n . Из условий 1) и 2) следует (см. (32.3.4)), что для каждого α из A

$$\int_{-\infty}^{\infty} \frac{\partial f}{\partial \alpha} dx = \int_{-\infty}^{\infty} \frac{\partial^2 f}{\partial \alpha^2} dx = 0,$$

и поэтому

$$(33.3.3) \quad E \left(\frac{\partial \log f}{\partial \alpha} \right)_0 = \int_{-\infty}^{\infty} \left(\frac{1}{f} \frac{\partial f}{\partial \alpha} \right)_0 f(x; \alpha_0) dx = 0,$$

$$E \left(\frac{\partial^2 \log f}{\partial \alpha^2} \right)_0 = \int_{-\infty}^{\infty} \left[\frac{1}{f} \frac{\partial^2 f}{\partial \alpha^2} - \left(\frac{1}{f} \frac{\partial f}{\partial \alpha} \right)^2 \right]_0 f(x; \alpha_0) dx =$$

$$= -E \left(\frac{\partial \log f}{\partial \alpha} \right)_0^2 = -k^2,$$

где, согласно условию 3), $k > 0$. Таким образом, в силу формулы (33.3.2), величина B_0 есть среднее арифметическое n независимых случайных величин, имеющих одно и то же распределение с нулевым средним значением. Из теоремы Хинчина (см. параграф 20.5) следует, что B_0 сходится по вероятности к нулю. Таким же образом убеждаемся, что B_1 сходится по вероятности к $-k^2$, а B_2 сходится по вероятности к неотрицательному значению $E H(x) < M$.

Пусть теперь δ и ϵ — фиксированные произвольно малые положительные числа, а $P(S)$ — совместная вероятностная функция для случайных величин x_1, \dots, x_n . При достаточно больших n , скажем, для

всех $n > n_0 = n_0(\delta, \varepsilon)$, имеем

$$P_1 = P(|B_0| \geq \delta^2) < \frac{1}{3} \varepsilon,$$

$$P_2 = P\left(B_1 \geq -\frac{1}{2} k^2\right) < \frac{1}{3} \varepsilon,$$

$$P_3 = P(|B_2| \geq 2M) < \frac{1}{3} \varepsilon.$$

Пусть S — множество всех точек $x = (x_1, \dots, x_n)$, для которых удовлетворяются все три неравенства

$$|B_0| < \delta^2, \quad B_1 < -\frac{1}{2} k^2, \quad |B_2| < 2M.$$

Дополнительное множество S^* состоит из всех точек x , для которых не выполняется хотя бы одно из этих трех неравенств, так что, согласно (6.2.2), $P(S^*) \leq P_1 + P_2 + P_3 < \varepsilon$. Поэтому $P(S) > 1 - \varepsilon$. Таким образом, вероятность попадания точки x в множество S , совпадающая с P -мерой множества S , превышает $1 - \varepsilon$, если $n > n_0(\delta, \varepsilon)$.

При $\alpha = a_0 \pm \delta$ правая часть равенства (33.3.1) принимает значение $B_0 \pm B_1 \delta + \frac{1}{2} \delta B_2 \delta^2$. В каждой точке x , принадлежащей S , сумма первого и третьего слагаемых в этом выражении по абсолютной величине меньше, чем $(M+1)\delta^2$, а $B_1 \delta < -\frac{1}{2} k^2 \delta$. Если $\delta < \frac{\frac{1}{2} k^2}{M+1}$, то знак всего выражения при $\alpha = a_0 \pm \delta$ определяется вторым слагаемым, так что $\frac{\partial \log L}{\partial \alpha} > 0$ при $\alpha = a_0 - \delta$ и $\frac{\partial \log L}{\partial \alpha} < 0$ при $\alpha = a_0 + \delta$. Далее, по условию 1), функция $\frac{\partial \log L}{\partial \alpha}$ для почти всех $x = (x_1, \dots, x_n)$ есть непрерывная функция от α из A . Таким образом, при произвольно малых δ и ε уравнение правдоподобия имеет, с вероятностью, превышающей $1 - \varepsilon$, корень, заключенный в пределах $a_0 \pm \delta$, если только $n > n_0(\delta, \varepsilon)$. Следовательно, первая часть доказательства закончена.

Пусть, далее, $\alpha^* = \alpha^*(x_1, \dots, x_n)$ есть решение уравнения правдоподобия, существование которого уже установлено. Из (33.3.1) и (33.3.2) получаем

$$(33.3.4) \quad k \sqrt{n} (\alpha^* - a_0) = \frac{\frac{1}{k \sqrt{n}} \sum_{i=1}^n \left(\frac{\partial \log f_i}{\partial \alpha} \right)_0}{-\frac{B_1}{k^2} - \frac{1}{2} \delta B_2 \frac{\alpha^* - a_0}{k^2}}$$

Из вышесказанного следует, что знаменатель дроби в правой части этого равенства сходится по вероятности к 1. Далее, согласно формуле (33.3.3), $\left(\frac{\partial \log f}{\partial a}\right)_0$ есть случайная величина со средним значением 0 и стандартным отклонением k . В силу теоремы Линдеберга — Леви (см. параграф 17.4), сумма $\sum_1^n \left(\frac{\partial \log f_i}{\partial a}\right)_0$ асимптотически нормальна $(0, k\sqrt{n})$, и, следовательно, числитель дроби в правой части равенства (33.3.4) асимптотически нормален $(0, 1)$.

Наконец, из теоремы сходимости параграфа 20.6 следует, что $k\sqrt{n} (a^* - a_0)$ — величина асимптотически нормальная $(0, 1)$, так что a^* асимптотически нормальна $(a_0, \frac{c}{\sqrt{n}})$, где $\frac{1}{c^2} = k^2 = E \left(\frac{\partial \log f}{\partial a}\right)_0^2$. По формуле (32.5.1) асимптотическая эффективность оценки a^* равняется

$$e_0(a^*) = \frac{1}{c^2 E \left(\frac{\partial \log f}{\partial a}\right)_0^2} = 1,$$

так что наша теорема доказана. Соответствующая теорема для дискретного распределения доказывается аналогичным образом.

В случае нескольких неизвестных параметров следует ввести условия, служащие непосредственным обобщением условий 1) — 3). Тогда можно доказать, таким же способом, как и выше, используя многомерную форму теоремы Линдеберга — Леви (см. параграфы 21.11 и 24.7), что уравнения правдоподобия имеют систему решений, являющихся асимптотически нормальными и совместно асимптотически эффективными оценками для параметров.

Пример 1. Логарифм функции правдоподобия для выборки в n значений из нормального распределения с неизвестными параметрами m и s^2 имеет вид

$$\log L = -\frac{1}{2s^2} \sum (x_i - m)^2 - \frac{1}{2} n \log s^2 - \frac{1}{2} n \log 2\pi,$$

и метод максимума правдоподобия приводит к уравнениям

$$\frac{\partial \log L}{\partial m} = \frac{1}{s^2} \sum (x_i - m) = 0,$$

$$\frac{\partial \log L}{\partial s^2} = \frac{1}{2s^4} \sum (x_i - m)^2 - \frac{n}{2s^2} = 0.$$

Отсюда получаем следующие оценки максимального правдоподобия:

$$m^* = \frac{1}{n} \sum x_i = \bar{x}, \quad (s^*)^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = s^2;$$

эти оценки совпадают с оценками, получаемыми с помощью метода моментов. Мы уже видели (см. параграфы 28.4 и 32.6, пример 1), что эти оценки асимптотически нормальны и асимптотически эффективны.

Пример 2. Рассмотрим распределение типа III (см. параграф 19.4)

$$f(x; \lambda) = \frac{1}{\Gamma(\lambda)} x^{\lambda-1} e^{-x}, \quad (x > 0, \lambda > 0)$$

с неизвестным параметром λ . К каждому конечному интервалу $a < \lambda < b$ ($a > 0$) можно применить неравенство (32.3.3а), и при этом обнаружится, что нижний предел дисперсии регулярной несмещенной оценки для λ по выборке в n значений имеет вид (см. параграф 12.3)

$$\frac{1}{n E \left(\frac{d \log f}{d \lambda} \right)^2} = \frac{1}{n E \left(\log x - \frac{a \log \Gamma(\lambda)}{d \lambda} \right)^2} = \frac{1}{n \frac{d^2 \log \Gamma(\lambda)}{a \lambda^2}}.$$

Чтобы оценить параметр λ с помощью метода моментов, приравниваем среднее значение выборки \bar{x} первому моменту величины λ в распределении и получаем $\lambda^* = \bar{x}$. При этом $E(\lambda^*) = \lambda$ и $D^2(\lambda^*) = \frac{1}{n}$. Отсюда, в силу формул (32.3.7) и (12.5.4), следует, что эффективность оценки λ^* не зависит от n и представляется выражением

$$e(\lambda^*) = \frac{1}{\lambda \frac{a^2 \log \Gamma(\lambda)}{d \lambda^2}} = \frac{1}{1 + \frac{1}{2\lambda} + 2\lambda \int_0^\infty \frac{P_1(x)}{(\lambda+x)^3} dx}.$$

Это выражение всегда меньше 1 и стремится к нулю при $\lambda \rightarrow 0$. С другой стороны, метод максимума правдоподобия приводит к уравнению

$$\frac{1}{n} \frac{d \log L}{d \lambda} = \frac{1}{n} \sum \log x_i - \frac{d \log \Gamma(\lambda)}{a \lambda} = 0,$$

и оценкой максимального правдоподобия является единственный положительный корень этого уравнения $\lambda = \lambda^{**}$. Согласно доказанной выше общей теореме, λ^{**} асимптотически нормальна $\left[\lambda, \left(n \frac{a^2 \log \Gamma(\lambda)}{a \lambda^2} \right)^{-1/2} \right]$, и асимптотическая эффективность оценки λ^{**} равна 1. Это нетрудно обнаружить и непосредственно, так как величина $\log x$ имеет среднее значение $\frac{d \log \Gamma(\lambda)}{a \lambda}$ и дисперсию $\frac{d^2 \log \Gamma(\lambda)}{a \lambda^2}$, так что, по теореме Линнеберга—Леви (см. параграф 17.4), величина

$\Sigma \log x_i$ асимптотически нормальна $\left[\frac{d \log \Gamma(\lambda)}{a \lambda}, \left(\frac{1}{n} \frac{d^2 \log \Gamma(\lambda)}{a \lambda^2} \right)^{1/2} \right]$.

Пример 3. В распределении типа III

$$f(x; a) = \frac{a^a}{\Gamma(a)} x^{a-1} e^{-ax}, \quad (x > 0, a > 0),$$

будем считать λ заданной положительной постоянной, а a — неизвестным параметром. Тогда будем иметь

$$E \left(\frac{\partial \log f}{\partial a} \right)^2 = E \left(\frac{\lambda}{a} - x \right)^2 = \frac{\lambda}{a^2}.$$

В этом случае и метод моментов и метод максимума правдоподобия приводят к одной и той же оценке $\frac{\lambda}{x}$ для a . Исключая смещение, получим несмещенную оценку $a^* = \frac{n\lambda - 1}{n\bar{x}}$, которая имеет плотность вероятности

$$g(a^*; a) = \frac{a^{n\lambda} (n\lambda - 1)^{n\lambda}}{\Gamma(n\lambda)} \left(\frac{1}{a^*} \right)^{n\lambda + 1} e^{-\frac{a(n\lambda - 1)}{a^*}},$$

что можно без труда вывести, например, с помощью характеристической функции (12.3.4). Предполагая, что $n\lambda > 2$, мы получим $E(a^*) = a$, $D^2(a^*) :=$
 $= \frac{a^2}{n\lambda - 2}$ и

$$E \left(\frac{\partial \log g}{\partial a} \right)^2 = E \left(\frac{n\lambda}{a} - \frac{n\lambda - 1}{a^*} \right)^2 = E \left(\frac{n\lambda}{a} - \sum x_i \right)^2 = \frac{n\lambda}{a^2}$$

Таким образом, в этом случае $n E \left(\frac{\partial \log f}{\partial a} \right)^2 = E \left(\frac{\partial \log g}{\partial a} \right)^2$, так что в (32.3.6) достигается знак равенства, и, следовательно, условие А) теоремы (32.3.3) выполняется. Отсюда следует, что a^* является достаточной оценкой для a , что можно проверить также и непосредственно с помощью формулы (32.4.1). С другой стороны, условие В) не выполняется, так как $\frac{\partial \log g}{\partial a}$ не может быть представлено в виде $k(a^* - a)$. Эффективностью оценки a^* является

$$e(a^*) = \frac{1}{n E \left(\frac{\partial \log f}{\partial a} \right)^2 D^2(a^*)} = \frac{n\lambda - 2}{n\lambda} < 1,$$

так что a^* не является эффективной ни при каком конечном n (см. параграф 32.3). Заставляя n стремиться к бесконечности, видим, что a^* является асимптотически-эффективной оценкой.

33.4. Метод минимума χ^2 . Рассмотренный в параграфе 30.3 метод минимума χ^2 применим лишь в случае группированного непрерывного распределения или дискретного распределения. Оценки, получаемые этим методом, при больших n асимптотически эквивалентны оценкам, полученным с помощью более простого видоизмененного метода минимума χ^2 , выражаемого уравнениями (30.3.3) или (30.3.3а), и в параграфе 30.3 уже было отмечено, что в рассматриваемых случаях последний метод совпадает с методом максимума правдоподобия.

Основная теорема о предельном распределении χ^2 для случая, когда некоторые параметры оцениваются по выборке, была доказана в параграфе 30.3 в предположении, что оценки находятся с помощью видоизмененного метода минимума χ^2 . Однако там же было указано, что имеется целый класс методов нахождения оценок, приводящих к тому же самому предельному распределению для χ^2 . Теперь мы докажем это утверждение.

Асимптотические выражения для оценок, получаемых с помощью видоизмененного метода минимума χ^2 , были приведены в явной форме (30.3.17) для общего случая s неизвестных параметров a_1, \dots, a_s . Предположим, что выполнены условия 1) — 3) предыдущего параграфа или аналогичные условия для дискретного распределения. Тогда из предыдущего параграфа следует, что оценки (30.3.17) асимптотически нормальны (это уже было показано в параграфе 30.3) и асимптотически эффективны.

Во всех множествах асимптотически нормальных и асимптотически эффективных оценок для параметров имеются члены порядка $n^{-1/2}$, такие же, как и в (30.3.17). Однако из вывода предельного распределения для χ^2 следует, что это предельное распределение полностью определяется членами порядка $n^{-1/2}$ в (30.3.17). Действительно, по формулам (30.3.1) и (30.3.4) получаем $\chi^2 = \sum_1^s y_i^2$, и (30.3.18) показывает, что предельное распределение для $y = (y_1, \dots, y_n)$ определяется именно указанными членами.

Таким образом, *теорема параграфа 30.3 о предельном распределении величины χ^2 справедлива для любого множества асимптотически нормальных и асимптотически-эффективных оценок параметров.*

ГЛАВА 34

ДОВЕРИТЕЛЬНЫЕ ОБЛАСТИ

34. 1. Вводные замечания. Предположим, что множество выборочных значений используется для получения оценок некоторого количества неизвестных параметров в распределении известного математического вида, причем известны выборочные распределения оценок, так что можно вычислять соответствующие средние значения, дисперсии и т. д.

Можем ли мы в такой обстановке высказать какие-либо вероятностные утверждения относительно истинных значений параметров? Возможно ли, например, указать для какого-нибудь параметра два предела и утверждать, что с некоторой определенной вероятностью истинное значение параметра должно лежать между этими пределами?

В ранней литературе по этому предмету вероятностные утверждения такого типа вводились с помощью известной *теоремы Байеса*; одной из типичных задач, решаемых таким способом, была классическая проблема *обращения вероятности* (см. параграф 34.2, пример 2). Однако эти приложения теоремы Байеса часто подвергались резкой критике, и появилось стремление избегать рассуждений такого рода и пересмотреть вопрос с совершенно иных точек зрения. Сделанные в этом направлении попытки можно разбить на две группы, одна из которых связана с теорией *доверительных вероятностей*, предложенной Р. Фишером (см., например, [14], [100], [102], [105]—[109]), а другая — с теорией *доверительных интервалов*, построенной Нейманом (см., например, [30], [161], [163], [165]—[167]). Мы ограничимся здесь в основном кратким изложением теории Неймана.

В следующем параграфе будет рассмотрен случай единственного неизвестного параметра, для которого мы проведем сравнение старого метода исследования с помощью теоремы Байеса и современной теории. В параграфе 34.3 мы перейдем к более общим случаям, и, наконец, в параграфе 34.4 будут рассмотрены некоторые примеры.

34.2. Единственный неизвестный параметр. Рассмотрим выборку в n значений x_1, \dots, x_n из распределения, содержащего лишь один неизвестный параметр α . Предположим сначала, что это распределение — непрерывного типа, имеющее плотность вероятности $f(x; \alpha)$. Предположим для простоты, что $f(x; \alpha)$ определена для всех значений параметра α . Пусть $\alpha^* = \alpha^*(x_1, \dots, x_n)$ — оценка для α и $g(\alpha^*; \alpha)$ — плотность вероятности этой оценки.

Вычислив значение α^* по фактической выборке, мы задаем вопрос: можно ли высказать какое-либо разумное вероятностное утверждение относительно неизвестного значения параметра α в распределении, из которого извлечена выборка? Этот вопрос будет рассмотрен с двух совершенно различных точек зрения.

1. Классический метод. В некоторых случаях законно предполагать, что фактическое значение параметра α в совокупности, из

которой извлекается выборка, определяется случайным экспериментом. Случаи такого рода встречаются, например, в статистике массового производства, где α обозначает некоторую неизвестную характеристику большой партии произведенных товаров, которую требуется оценить по небольшой выборке. Отдельная партия товаров рассматривается как индивидуум, извлеченный из совокупности аналогичных партий товаров, и значения параметра α испытывают случайные колебания, вызываемые изменениями производственного процесса и качества сырья. Извлечение одного индивидуума из этой совокупности партий товаров является случайным экспериментом, который и определяет фактическое значение параметра α . Аналогичные случаи встречаются, например, в некоторых вопросах генетики.

В таких случаях параметр α сам является случайной величиной, имеющей некоторое априорное распределение. Предположим, что это распределение характеризуется известной плотностью вероятности $\tilde{\omega}(\alpha)$. Функция $\tilde{\omega}(\alpha)$ является частной плотностью вероятности для α в совместном распределении величин α и α^* , а $g(\alpha^*; \alpha)$ является условной плотностью вероятности для α^* при заданном значении α . Обратно, условной плотностью вероятности для α при заданном значении α^* будет, согласно формуле (21.4.10),

$$h(\alpha | \alpha^*) = \frac{\tilde{\omega}(\alpha) g(\alpha^*; \alpha)}{\int_{-\infty}^{\infty} \tilde{\omega}(\alpha) g(\alpha^*; \alpha) d\alpha}.$$

Это соотношение выражает теорему Байеса для нашего конкретного случая. Поэтому величина

$$(34.2.1) \quad P(k_1 < \alpha < k_2 | \alpha^*) = \int_{k_1}^{k_2} h(\alpha | \alpha^*) d\alpha$$

представляет собой условную вероятность события $k_1 < \alpha < k_2$ при заданном значении α^* . Эта вероятность обычно называется апостериорной вероятностью события $k_1 < \alpha < k_2$, в отличие от априорной

вероятности этого события, равной $\int_{k_1}^{k_2} \tilde{\omega}(\alpha) d\alpha$.

В силу параграфов 14.3 и 21.4, апостериорная вероятность (34.2.1) допускает частотную интерпретацию, которая заключается

в следующем. Рассмотрим последовательность большого количества независимых испытаний, каждое из которых заключается в извлечении партии товаров из совокупности всех партий и, затем, в извлечении выборки в n значений из выбранной партии (мы используем терминологию, связанную с рассмотренным выше примером, но наши рассуждения имеют, очевидно, общий характер). По выборке вычисляем оценку a^* и, далее, предполагаем, что можно исследовать все товары в полной партии, так что соответствующее значение a может быть определено непосредственно. Таким образом, результатом каждого испытания будет пара наблюденных значений величин a и a^* . Из последовательности всех испытаний выберем подпоследовательность, образованную из тех случаев, в которых наблюденное значение a^* принадлежит к некоторой малой окрестности заданного значения a_0^* . Частота события $k_1 < a < k_2$ в этой подпоследовательности будет, с точностью, определяемой случайными колебаниями, дана значением априорной вероятности (34.2.1) при $a^* = a_0^*$.

Вышеизложенное является прямой частотной интерпретацией априорной вероятности. С помощью небольшого видоизменения этих рассуждений можно получить результат, имеющий большее формальное сходство с приведенной ниже теорией доверительных интервалов. Пусть задано число ε , такое, что $0 < \varepsilon < 1$. Для каждого заданного a^* можно определить в (34.2.1) такие пределы $k_1 = k_1(a^*, \varepsilon)$ и $k_2 = k_2(a^*, \varepsilon)$, что вероятность $P(k_1 < a < k_2 | a^*)$ будет иметь значение $1 - \varepsilon$. (Читатель может обратиться для наглядности к фиг. 33 на стр. 555, в которой вместо c_1 и c_2 следует взять k_1 и k_2). Рассмотрим снова вышеописанную последовательность всех испытаний и вычислим пределы $k_1 = k_1(a^*, \varepsilon)$ и $k_2 = k_2(a^*, \varepsilon)$ по выборке, полученной в каждом испытании. Интервал (k_1, k_2) будет зависеть от a^* , так что, вообще говоря, последовательные испытания будут давать различные интервалы. Будем считать в каждом испытании осуществление события $k_1 < a < k_2$ „успехом“ и осуществление обратного события — „неудачей“. Тогда вероятность успеха постоянно равна $1 - \varepsilon$ и, соответственно (см. параграф 16.6), частота успехов в длинном ряду испытаний будет, с точностью, определяемой случайными колебаниями, равняться $1 - \varepsilon$. Практические выводы из этого результата, в случаях, когда применение изложенного метода законно, будут аналогичны выводам, излагаемым ниже.

2. Метод доверительных интервалов. В случае, когда имеются определенные основания считать a случайной величиной с известным распределением вероятностей, применение предыдущего метода совершенно законно и приводит к точным вероятностным утверждениям о значении a , соответствующем заданной выборке. Однако в большинстве

встречающихся на практике случаев эти условия не выполняются. Как правило, α является просто неизвестной постоянной, и нет никаких оснований полагать, что истинное значение этой постоянной определяется какой-либо последовательностью действий, схожей со случным экспериментом. Часто даже имеются указания противоположного характера, как, например, в случаях, когда значения α для различных совокупностей испытывают систематическое изменение во времени или в пространстве. Более того, даже если законно считать α случайной величиной, обычно не имеется достаточных сведений об априорном распределении этой величины.

Таким образом, весьма желательно найти подход к нашей проблеме без каких бы то ни было гипотез о случайной или неслучайной природе параметра α . Авторы, упомянутые в предыдущем параграфе, разработали некоторые методы в этом направлении, и мы теперь постараемся показать, как можно решить поставленную проблему с помощью метода *доверительных интервалов*, предложенного Нейманом (см. также Уилкс [42], [234]). В настоящем параграфе мы изложим этот метод при некоторых упрощающих предположениях; более общие случаи будут рассмотрены в следующем параграфе.

Теперь будем считать α переменной величиной в смысле обычного анализа, т. е. значение этой величины в совокупности, из которой извлечена наша выборка, предполагается постоянным, хотя и неизвестным. Полученные результаты будут справедливы независимо от того, определяется ли значение α случным экспериментом или нет, так что излагаемый метод фактически является более общим, чем предыдущий.

Как и раньше, рассмотрим выборку в n значений из распределения с плотностью вероятности $f(x; \alpha)$ и обозначим через $g(\alpha^*; \alpha)$ плотность вероятности оценки $\alpha^* = \alpha^*(x_1, \dots, x_n)$. Обозначим, далее, через $P(S; \alpha)$ совместную вероятностную функцию выборочных величин x_1, \dots, x_n и зададим число ϵ , такое, что $0 < \epsilon < 1$.

При каждом фиксированном α плотность вероятности $g(\alpha^*; \alpha)$ задает распределение вероятностей величины α^* , которое можно интерпретировать как распределение единичной массы по вертикальной прямой в плоскости (α, α^*) , проходящей через точку $(\alpha, 0)$ (см. фиг. 33).

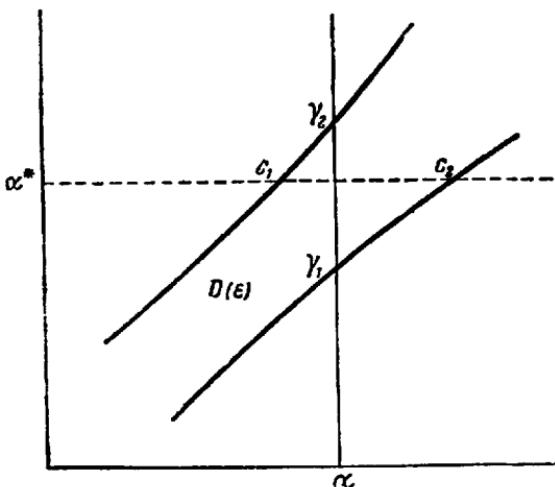
Предположим теперь, что для каждого значения α определены два числа $\gamma_1 = \gamma_1(\alpha, \epsilon)$ и $\gamma_2 = \gamma_2(\alpha, \epsilon)$, такие, что количество массы, попадающей на интервал $\gamma_1 < \alpha^* < \gamma_2$ соответствующей вертикали, т. е.

вероятность события $\gamma_1 < \alpha^* < \gamma_2$ при значении оцениваемого параметра, равном α , равно

$$(34.2.2) \quad P(\gamma_1 < \alpha^* < \gamma_2) = \int_{\gamma_1}^{\gamma_2} g(\alpha^*; \alpha) d\alpha^* = 1 - \epsilon.$$

Очевидно, такие числа γ_1 и γ_2 всегда можно отыскать, и притом бесконечным количеством способов, так как эти числа можно спределить из соотношений

$$\int_{-\infty}^{\gamma_1} g d\alpha^* = \epsilon_1 \quad \text{и} \quad \int_{\gamma_2}^{\infty} g d\alpha^* = \epsilon_2,$$



Фиг. 33. Доверительные интервалы для единственного неизвестного параметра

где ϵ_1 и ϵ_2 — любые положительные числа, сумма которых равна ϵ .

Если извлекается выборка в n значений из распределения, соответствующего какому-либо значению α , то событие $\gamma_1 < \alpha^* < \gamma_2$ всегда имеет вероятность $1 - \epsilon$. Числа γ_1 и γ_2 зависят от α , и если α меняется, то точки (α, γ_1) и (α, γ_2) описывают в плоскости (α, α^*) две кривых, как показано на фиг. 33. Предположим, что каждая из этих кривых пересекается с прямой, параллельной оси α , лишь в одной точке. Обозначим точки пересечения этих кривых с горизонтальной прямой, проходящей через точку $(0, \alpha^*)$, буквами $c_1 = c_1(\alpha^*, \epsilon)$ и $c_2 = c_2(\alpha^*, \epsilon)$, и пусть $D(\epsilon)$ обозначает область, заключенную между нашими двумя

кривыми. Рассмотрим три соотношения

$$(34.2.3) \quad \begin{aligned} (\alpha, \alpha^*) &\subset D(\epsilon), \\ \gamma_1(\alpha, \epsilon) &< \alpha^* < \gamma_2(\alpha, \epsilon), \\ c_1(\alpha^*, \epsilon) &< \alpha < c_2(\alpha^*, \epsilon). \end{aligned}$$

При любом фиксированном значении α каждое из этих соотношений удовлетворяется на некотором множестве точек $x = (x_1, \dots, x_n)$ в пространстве выборок. Однако все три соотношения совершенно эквивалентны, так как они выражают один и тот же факт, что точка (α, α^*) принадлежит области $D(\epsilon)$. Таким образом, три соответствующих множества точек в пространстве выборок совпадают, и, следовательно, при любом значении α из (34.2.2) получаем

$$(34.2.4) \quad P(c_1 < \alpha < c_2; \alpha) = 1 - \epsilon.$$

Оба соотношения (34.2.2) и (34.2.4) дают значение функции множества $P(S; \alpha)$ для некоторого множества S в пространстве выборок, определяемого двумя различными, но эквивалентными способами, именно, двумя последними соотношениями (34.2.3). Первое из этих соотношений означает, что случайная величина α^* принимает значение, заключенное между постоянными пределами γ_1 и γ_2 . С другой стороны, последнее соотношение (34.2.3) означает, что случайная величина $c_1(\alpha^*, \epsilon)$ имеет значение меньше α , а случайная величина $c_2(\alpha^*, \epsilon)$ имеет значение больше α , или, другими словами, что переменный интервал (c_1, c_2) покрывает фиксированную точку α . В соответствии с (34.2.4), вероятность этого события равна $1 - \epsilon$, каково бы ни было значение α .

Рассмотрим теперь последовательность независимых испытаний, каждое из которых заключается в извлечении выборки в n значений из совокупности с плотностью вероятности $f(x; z)$, причем значения α , соответствующие последовательным испытаниям, могут сохраняться постоянными или меняться совершенно произвольным образом, случайно или неслучайно. По каждому множеству выборочных значений вычисляем величины $c_1 = c_1(\alpha^*, \epsilon)$ и $c_2 = c_2(\alpha^*, \epsilon)$, используя заранее заданное значение ϵ . Вообще говоря, в различных испытаниях c_1 и c_2 будут иметь различные значения. Каждое испытание будем считать „успехом“, если интервал (c_1, c_2) покрывает соответствующую точку α , и „неудачей“ в обратном случае. В силу формулы (34.2.4), вероятность успеха постоянно равна $1 - \epsilon$, и, следовательно (см. параграф 16.6), частота

успехов в длинной последовательности испытаний будет, с точностью, определяемой случайными колебаниями, равняться $1 - \epsilon$.

Предположим теперь, что мы постоянно придерживаемся следующего порядка действий. Прежде всего, выбираем раз и навсегда некоторое малое число ϵ , скажем, $\epsilon = \frac{p}{100}$. Всякий раз, когда выборка уже извлечена и вычислены соответствующие пределы c_1 и c_2 , будем утверждать, что неизвестное значение a в соответствующей совокупности заключено между c_1 и c_2 . В соответствии с изложенным выше, вероятность того, что это утверждение неверно, всегда равняется $\epsilon = \frac{p}{100}$. Таким образом, в длинном ряду испытаний наше утверждение будет неверным лишь приблизительно в $p\%$ случаев, а в остальных случаях будет правильным.

Интервал (c_1, c_2) будем называть доверительным интервалом для параметра a , соответствующим коэффициенту доверия $1 - \epsilon$ или доверительному уровню $\epsilon = \frac{p}{100}$. Числа c_1 и c_2 являются соответствующими доверительными пределами.

Сравнивая этот способ исследования со способом, основанным на теореме Байеса, мы видим, что метод доверительных интервалов не требует никаких гипотез о случайной или неслучайной природе величины a . С другой стороны, из самой общности метода доверительных интервалов вытекает, что этот метод не приводит к вероятностным утверждениям такого типа: „Вероятность того, что a заключено между такими-то фиксированными пределами, равняется $1 - \epsilon$ “. Действительно, такое утверждение не имеет смысла, если a не является случайной величиной. Утверждения, которые следуют из метода доверительных интервалов, являются утверждениями типа соотношения (34.2.4), или, в словесной форме, такого типа: „Вероятность того, что такие-то пределы (которые могут меняться от выборки к выборке) заключают между собой значение параметра a , соответствующее фактической выборке, равняется $1 - \epsilon$ “. Как указано выше, из этого утверждения можно вывести правило действий, связанное с постоянным риском ошибки, равным ϵ , где ϵ может быть любым фиксированным числом.

Следует заметить, что заданному ϵ соответствует не единственная система доверительных интервалов. Так же, как мы могли брать различные оценки для одного и того же параметра a , мы можем рассматривать различные системы доверительных интервалов, приво-

дящие к различным правилам действий, связанным с одним и тем же риском ошибки. Это — ни в коей мере не противоречие. Как мы видели выше, доверительные интервалы, получаемые с помощью данного правила, могут меняться от выборки к выборке, и естественно, что для данной выборки разные правила могут привести к различным интервалам (см. ниже, пример 1).

Очевидно, в наших интересах выработать такие правила, которые при данных условиях приводили бы к возможно более *коротким* интервалам. Предположим, например, что мы имеем дело с несмещеными и приблизительно нормально распределенными оценками a^* . Полосу $D(\epsilon)$ в фиг. 33 можно сделать возможно более узкой, выбирая для a^* оценку с *минимальной дисперсией*. Таким образом, классы эффективных и асимптотически эффективных оценок, изученные в главе 32, при весьма общих условиях приводят к кратчайшим или асимптотически кратчайшим доверительным интервалам. Мы не имеем возможности входить здесь в дальнейшее рассмотрение этого вопроса и отсылаем читателя к статьям Неймана [165] и Уилкса [233].

В заключение заметим, что приведенные выше определения и выводы применимы даже в случае *дискретного* распределения, содержащего единственный неизвестный параметр. Однако в этом случае необходимо сделать одно важное изменение. Если распределение на вертикали, проходящей через точку $(a, 0)$ на фиг. 33, имеет точки сосредоточения массы, то пределы γ_1 и γ_2 не всегда можно определить так, чтобы $P(\gamma_1 < a^* < \gamma_2; a) = 1 - \epsilon$, как это требуется формулой (34.2.2). Мы должны будем удовлетвориться выбором таких γ_1 и γ_2 , чтобы $P(\gamma_1 < a^* < \gamma_2; a) \geq 1 - \epsilon$, что, очевидно, всегда возможно. Тогда полоса $D(\epsilon)$ и доверительный интервал (c_1, c_2) определяются так же, как и в непрерывном случае. В этом случае риск ошибки при утверждении, что a принадлежит (c_1, c_2) , не точно равняется ϵ , а *не превосходит* ϵ . За этим исключением, все рассуждения в дискретном случае совершенно аналогичны рассуждениям, приведенным выше.

Пример 1. Пусть требуется оценить среднее значение m нормальной совокупности с известным стандартным отклонением σ . Заменяя на фиг. 33 γ и a^* на m и m^* , рассмотрим сначала эффективную оценку $m^* = \bar{x} = \sum \frac{v_i}{n}$, которая нормальна $\left(m, \frac{\sigma}{\sqrt{n}}\right)$. Для доверительного уровня $\epsilon = \frac{p}{100}$ пределы γ_1 и γ_2 на фиг. 33 можно положить равными $m \pm \frac{\lambda_p \sigma}{\sqrt{n}}$, где λ_p — p -процентное

значение нормального отклонения. Тогда границами области $D(\epsilon)$ будут прямые линии $\bar{x} = m \pm \frac{\lambda p^\sigma}{\sqrt{n}}$. Соотношения

$$m - \frac{\lambda p^\sigma}{\sqrt{n}} < \bar{x} < m + \frac{\lambda p^\sigma}{\sqrt{n}},$$

$$\bar{x} - \frac{\lambda p^\sigma}{\sqrt{n}} < m < \bar{x} + \frac{\lambda p^\sigma}{\sqrt{n}},$$

очевидно, эквивалентны, так что пределы c_1 и c_2 равны $\bar{x} \pm \frac{\lambda p^\sigma}{\sqrt{n}}$. Таким образом, правило, заключающееся в утверждении, что как только выборка фактически получена, неизвестное среднее значение m заключено между пределами $\bar{x} \pm \frac{\lambda p^\sigma}{\sqrt{n}}$, дает постоянный риск ошибки, равный $p\%$.

Действительно, указанный доверительный интервал уже был рассмотрен в параграфе 31.3 (пример 2), и мы видели, что если пользоваться p -процентным уровнем значимости, то гипотеза о равенстве среднего значения распределения заранее заданному числу c должна считаться совместимой с наличными данными, когда c заключено между доверительными пределами

$\bar{x} \pm \frac{\lambda p^\sigma}{\sqrt{n}}$, а в противном случае должна быть отвергнута.

С другой стороны, предположим, что рассматривается неэффективная оценка $m^* = z$, где z — медиана выборки. Согласно параграфу 28.5, z асимптотически нормальна $(m, \frac{k\sigma}{\sqrt{n}})$, где $k = \sqrt{\frac{\pi}{2}} = 1,2533$. Предположим, для простоты вывода, что ошибкой аппроксимации можно пренебречь, так что распределение z в точности нормально. Каждое из эквивалентных соотношений

$$m - \frac{k\lambda p^\sigma}{\sqrt{n}} < z < m + \frac{k\lambda p^\sigma}{\sqrt{n}} \quad \text{и} \quad z - \frac{k\lambda p^\sigma}{\sqrt{n}} < m < z + \frac{k\lambda p^\sigma}{\sqrt{n}}$$

имеет вероятность $p\%$, и, следовательно, в этом случае мы получаем p -процентные доверительные пределы $z \pm \frac{k\lambda p^\sigma}{\sqrt{n}}$. Таким образом, мы по данной выборке получили различные доверительные интервалы для m , соответственно правилам, основанным на \bar{x} или на z . Тем не менее, риск ошибки в обоих случаях одинаковый, если используется одно и то же значение ϵ . Очевидно, первое правило всегда дает более короткий интервал, чем второе.

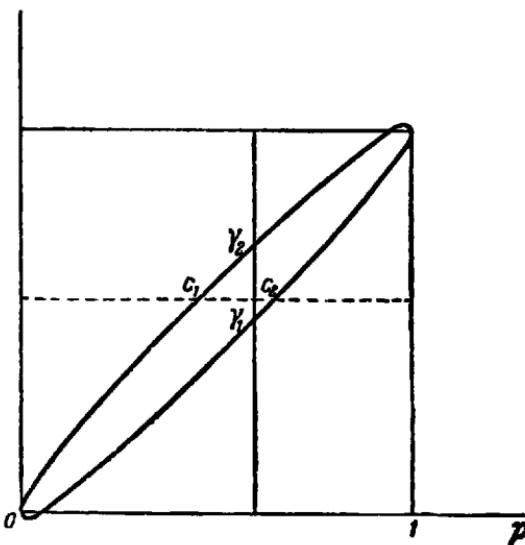
П р и м е р 2. Предположим, что осуществлено n повторений случайного эксперимента, причем событие E произошло v раз. Требуется оценить неизвестную вероятность p события F . Эта классическая проблема *обращения вероятности* в большинстве учебников решается с помощью теоремы Байеса.

Применим здесь к этой проблеме теорию доверительных интервалов и рассмотрим эффективную оценку (см. параграф 32.3, пример 5) $p^* = \frac{v}{n}$, которая

асимптотически нормальна $\left(p, \sqrt{\frac{pq}{n}} \right)$, где $q = 1 - p$. Полагая пределы γ_1 и γ_2 равными $p \pm \lambda \sqrt{\frac{pq}{n}}$ и считая, как и в предыдущем примере, что распределение в точности нормально, вместо фиг. 33 получим фиг. 34.

Область $D(z)$ ограничена здесь кривыми $p^* = p \pm \lambda \sqrt{\frac{pq}{n}}$, образующими две половины эллипса; λ есть 100 ε -процентное значение нормального откло-

$$p^* = \frac{\nu}{n}$$



Фиг. 34. Доверительные интервалы для неизвестной вероятности. $n = 100$, $\varepsilon = 0,05$

нения. Тот факт, что точка (p, p^*) расположена внутри эллипса, можно выразить иначе, сказав, что p^* лежит в пределах $p \pm \lambda \sqrt{\frac{pq}{n}}$, или эквивалентным утверждением, что p лежит в пределах

$$(34.2.5) \quad \frac{n}{n+\lambda^2} \left(p^* + \frac{\lambda^2}{2n} \pm \lambda \sqrt{\frac{p^* q^*}{n} + \frac{\lambda^2}{4n^2}} \right).$$

Последние пределы определяют 100 ε -процентный доверительный интервал для p .

Конечно, этот результат — приближенный, так как в действительности p^* имеет дискретное распределение, лишь приближенно нормальное. Е. Пирсон и Клоннер [195] построили графики, основанные на точном распределении и

позволяющие определять доверительные интервалы 5- и 1-процентного уровней. Пирсон и Клоппер указывают, что их графики можно использовать, например, для определения значения n , которое необходимо взять, чтобы достичь желаемой степени точности в оценке вероятности p . Предположим, например, что p примерно равно 50%, и что требуется получить доверительный интервал, по длине не превышающий δ . С помощью приближенного решения (34.2.5) получим, полагая $p^* = \frac{1}{2}$,

$$\frac{\lambda}{\sqrt{n + \lambda^2}} \leq \delta \text{ или } n \geq \lambda^2 \frac{1 - \delta^2}{\delta^2}.$$

Взяв, например, $\delta = \varepsilon = 0,01$, мы получим $n > 66340$.

Пример 3. Предположим, что задана совокупность из N индивидуумов (N — конечное число), Np из которых обладают некоторым признаком A , а остальные $Nq = N - Np$ этим признаком не обладают. Требуется оценить неизвестную пропорцию p методом *репрезентативной выборки* (см. параграф 25.7). Извлечем *бесповторно* случайную выборку в n индивидуумов и отметим в этой выборке число v индивидуумов, обладающих признаком A . В общепринятых учебниках по теории вероятностей (см., например, Крамер [10], стр. 38) показывается, что

$$E\left(\frac{v}{n}\right) = p, \quad D^2\left(\frac{v}{n}\right) = \frac{N-n}{N-1} \cdot \frac{pq}{n}.$$

Далее, величина $p^* = \frac{v}{n}$ распределена приблизительно нормально, если n и $N - n$ велики. Принимая p^* за оценку для p , мы, как и выше, предположим что ошибкой, вызванной допущением о нормальности распределения, можно пренебречь. Тогда вероятность того, что p^* лежит между пределами $p \pm \lambda \sqrt{\frac{N-n}{N-1} \frac{pq}{n}}$, равна ε (здесь λ имеет то же значение, что и в предыдущем примере). Таким образом, мы получим доверительные пределы для неизвестной пропорции p , просто беря в (34.2.5) $\frac{N-1}{N-n} \cdot n$ вместо n .

34.3. Общий случай. Развитая в предыдущем параграфе теория доверительных интервалов легко обобщается на более общие случаи. Рассмотрим распределение непрерывного типа, содержащее k неизвестных параметров a_1, \dots, a_k , и предположим, что из этого распределения извлечена выборка в n элементов.

Как обычно, будем считать выборочные значения координатами точки $x = (x_1, \dots, x_n)$ в n -мерном пространстве выборок R_n , и, аналогично, множество параметров распределения будем представлять точкой $a = (a_1, \dots, a_n)$ в k -мерном пространстве параметров P_k .

Для простоты допустим, что распределение определено для всех точек α в P_k , и обозначим совместную вероятностную функцию величин x_1, \dots, x_n через $P(S; \alpha)$, где S — множество из пространства выборок R_n .

Для дальнейших исследований нет необходимости предполагать, что величины x_1, \dots, x_n независимы и имеют одно и то же распределение. Действительно, при том же обобщении, что и в параграфе 32.8, можно через $P(S; \alpha)$ обозначить любую n -мерную вероятностную функцию непрерывного типа, определенную для всех точек $\alpha = (\alpha_1, \dots, \alpha_k)$ из пространства параметров.

Для каждой точки α из пространства параметров P_k можно определить такое множество $S(\alpha)$ точек x в R_n , что

$$(34.3.1) \quad P[x \subset S(\alpha); \alpha] = 1 - \varepsilon,$$

где ε — заранее заданное число. Множество $S(\alpha)$ соответствует интервалу $\gamma_1 < \alpha^* < \gamma_2$ на фиг. 33 *), а соотношение (34.3.1) соответствует (34.2.2). Далее, области $D(\varepsilon)$ на фиг. 33 соответствует множество D всех точек (α, x) пространства-произведения $P_k \cdot R_n$, для которых выполняется соотношение $x \subset S(\alpha)$. Для каждой точки x в R_n рассмотрим множество $\Sigma(x)$ всех точек α из P_k таких, что $(\alpha, x) \subset D$. Тогда $\Sigma(x)$ соответствует интервалу $c_1 < \alpha < c_2$ на фиг. 33, и три соотношения

$$(\alpha, x) \subset D, \quad x \subset S(\alpha), \quad \alpha \subset \Sigma(x)$$

эквивалентны по тем же причинам, что и соответствующие соотношения (34.2.3). Отсюда получаем аналог для (34.2.4):

$$(34.3.2) \quad P[\alpha \subset \Sigma(x); \alpha] = 1 - \varepsilon.$$

Дальнейшие рассуждения совершенно аналогичны рассуждениям в предыдущем частном случае. Если мы повторно извлекаем выборки в n значений из распределений данного типа, то соответствующие точки α из пространства параметров можно сохранять неизменными или считать меняющимися совершенно произвольным образом, и если для каждой выборки мы утверждаем, что некоторая фиксированная точка α пространства параметров принадлежит множеству $\Sigma(x)$, соответствующему этой выборке, то в каждом случае будем иметь вероятность

*) Можно считать, что фиг. 33 относится к выборке единственного наблюденного значения α^* из распределения с плотностью вероятности $g(\alpha^*; \alpha)$.

$\epsilon = \frac{p}{100}$ ошибочности этого утверждения. Следовательно, в длинном ряду таких утверждений мы будем ошибаться лишь приблизительно в $p\%$ всех случаев.

Для точки a из пространства параметров будем называть множество $\Sigma(x)$ доверительной областью, соответствующей коэффициенту доверия $1 - \epsilon$ или доверительному уровню $\epsilon = \frac{p}{100}$. В частности, если множеством $\Sigma(x)$ служит интервал в P_k , определяемый единственным соотношением вида

$$(34.3.3) \quad c_1(x, \epsilon) < a_r < c_2(x, \epsilon),$$

где r — один из индексов $1, \dots, k$, а c_1 и c_2 не зависят от a_1, \dots, a_k , то будем называть $\Sigma(x)$ доверительным интервалом для параметра a_r . Последнее определение, очевидно, включает в себя определение предыдущего параграфа как частный случай. В более общем случае, когда $\Sigma(x)$ — цилиндрическое множество (см. параграф 3.5), основанием которого является некоторое множество в подпространстве параметров a_1, \dots, a_r , где $r < k$, то будем говорить, что $\Sigma(x)$ есть доверительная область для параметров a_1, \dots, a_r .

Замечания предыдущего параграфа относительно обобщения на распределения с точками сосредоточения массы применимы и в настоящем общем случае. Наконец, обобщение на случай выборок из многомерных распределений получается непосредственно.

34.4. Примеры. В примере 6 параграфа 31.3 мы уже встречались с некоторыми доверительными интервалами для коэффициентов регрессии и корреляции в случае выборок из двумерного нормального распределения. Теперь рассмотрим несколько дальнейших примеров, которые послужат поводом для некоторых общих замечаний.

Пример 1. Среднее значение нормального распределения. Если x_1, \dots, x_n — множество выборочных значений из нормального распределения с неизвестными параметрами m и s , то величина (см. параграф 29.4)

$$t = \sqrt{n-1} \frac{\bar{x} - m}{s}$$

имеет распределение Стьюдента с $n-1$ степенями свободы с соответствующей плотностью вероятности $s_{n-1}(t)$. Для всякого интервала (t', t'') соотношение

$$(34.4.1) \quad t' < \sqrt{n-1} \frac{\bar{x} - m}{s} < t''$$

имеет, таким образом, вероятность $\int_{t'}^{t''} s_{n-1}(t) dt$, не зависящую от параметров m и σ , и, выбирая подходящим образом t' и t'' , можно сделать эту вероятность равной любому наперед заданному числу $1 - \varepsilon$.

Предположим, что t' и t'' фиксированы. Тогда для каждой точки (m, σ) из пространства параметров соотношение (34.4.1) определяет множество точек x из пространства выборок, отвечающее множеству $S(a)$ предыдущего параграфа. Однако (34.4.1) можно записать также в следующем эквивалентном виде:

$$(34.4.2) \quad \bar{x} - t'' \frac{s}{\sqrt{n-1}} < m < \bar{x} + t' \frac{s}{\sqrt{n-1}}.$$

Для каждой фиксированной точки $x = (x_1, \dots, x_n)$ из пространства выборок это соотношение определяет интервал из пространства параметров, не зависящий от σ , т. е. являющийся частным случаем интервала (34.3.3), в котором вместо a_x стоит m . Согласно определению предыдущего параграфа, (34.4.2) является *доверительным интервалом для среднего значения m* , и мы получаем следующее соотношение, отвечающее общему соотношению (34.3.2):

$$(34.4.3) \quad P\left(\bar{x} - t'' \frac{s}{\sqrt{n-1}} < m < \bar{x} + t' \frac{s}{\sqrt{n-1}}; m, \sigma\right) = \int_{t'}^{t''} s_{n-1}(t) dt.$$

Таким образом, если мы повторно извлекаем выборки в n значений из нормальных совокупностей, причем значения m и σ , соответствующие исследовательным выборкам, остаются постоянными или изменяются произвольным образом и если для каждой выборки мы вычислим доверительные пределы $\bar{x} - t'' \frac{s}{\sqrt{n-1}}$ и $\bar{x} + t' \frac{s}{\sqrt{n-1}}$, то частота случаев, в которых m будет заключено между этими пределами, в длинном ряде наблюдений будет приблизительно равна $\int_{t'}^{t''} s_{n-1}(t) dt$.

Каждый выбор чисел t' и t'' будет, согласно (34.4.3), давать правило вычисления доверительных интервалов для параметра m , соответствующих коэффициенту доверия $\int_{t'}^{t''} s_{n-1}(t) dt$. Полагая, например,

$t' = -t_p$ и $t'' = t_p$, где t_p — p -процентное значение величины t с $n-1$ степенями свободы, мы получим доверительные пределы

$$\bar{x} \pm t_p \frac{s}{\sqrt{n-1}},$$

соответствующие коэффициенту доверия $1 - \frac{p}{100}$ или доверительному уровню $p^0\%$.

Рассмотрим выборку в $n=10$ значений из предположительно нормальной совокупности, приведенную в последнем столбце таблицы 31.3.7. Среднее значение и стандартное отклонение выборки соответственно равны 1,58 и 1,167. Отсюда получаем, согласно последнему правилу, доверительные пределы для неизвестного среднего значения совокупности m : $1,58 \pm 0,389 t_p$. Для доверительного уровня $p=5\%$ получается доверительный интервал $0,70 < m < 2,46$, для $p=1\%$ — интервал $0,32 < m < 2,84$.

Выбирая t' и t'' иначе, мы получим другие правила вычисления доверительных интервалов для m . Предположим, например, что нам заранее задан интервал (a, b) . Извлекаем выборку в n значений из нормальной совокупности и обозначим полученную выборочную точку через \bar{x}_0 , среднее значение выборки через \bar{x}_0 и стандартное отклонение — через s_0 . По этим значениям \bar{x}_0 и s_0 определим t' и t'' так, чтобы

$$\bar{x}_0 - t' \frac{s_0}{\sqrt{n-1}} = a, \quad \bar{x}_0 + t'' \frac{s_0}{\sqrt{n-1}} = b.$$

Как и любые другие значения, t' и t'' , определенные таким способом, соответствуют некоторому правилу вычисления доверительных интервалов для m , и для нашей наличной выборки \bar{x}_0 это правило приводит в точности к данному интервалу (a, b) . Решая приведенные выше уравнения относительно t' и t'' , найдем, что соответствующий коэффициент доверия равен

$$(34.4.4) \quad \int_{\frac{\sqrt{n-1}(\bar{x}_0-a)}{s_0}}^{\frac{\sqrt{n-1}(\bar{x}_0-b)}{s_0}} s_{n-1}(t) dt.$$

Если выборка \bar{x}_0 известна, то число (34.4.4) можно вычислить для любого интервала (a, b) . Таким образом можно говорить, что при оценке параметра m с помощью выборочных характеристик \bar{x} и s

*наблюденная выборка x_0 приписывается любому заданному интервалу (a, b) коэффициент доверия, определяемый формулой (34.4.4) *).*

Однако последнее утверждение следует употреблять очень осторожно, не отступая от его точного смысла. Мы не говорим, что существует вероятность (34.4.4) попадания t в интервал (a, b) . Как уже было указано в параграфе 34.2, такое утверждение имеет смысл лишь в случае, когда t есть случайная величина. В действительности мы утверждаем лишь то, что существует правило вычисления доверительных интервалов для t , которое в частном случае выборки x приводит к заданному интервалу (a, b) , причем это правило относится к коэффициенту доверия (34.4.4).

В рассмотренном выше случае выборки в $n = 10$ значений из таблицы 31.3.7 мы с помощью таблицы IV (стр. 611) обнаруживаем, что интервал

$$0,5 < t < 2,5 \text{ имеет коэффициент доверия } \int_{-2,37}^{+2,78} s_9(t) dt = 0,97.$$

Как и в параграфе 34.2, следует заметить, что приведенная выше система доверительных интервалов и коэффициентов доверия не единственна. Если, например, заменить \bar{x} и s медианой и средним отклонением выборки, то получится другая система правил.

Пример 2. Разность между средними значениями двух нормальных распределений. Пусть x_1, \dots, x_{n_1} и y_1, \dots, y_{n_2} — две независимые выборки со средними значениями \bar{x} и \bar{y} и стандартными отклонениями s_1 и s_2 . Предположим, что эти выборки извлечены из нормальных совокупностей со средними значениями m_1 и m_2 и стандартными отклонениями σ_1 и σ_2 . Предположим, что все четыре параметра неизвестны, и требуется оценить разность $m_1 - m_2$ между средними значениями совокупностей. Эта проблема неоднократно дискутировалась в литературе (см., например, Бартлет [55], [56], Беренс [60], Фишер [105]—[109], Нейман [167], Велч [229]).

В параграфе 31.2 был рассмотрен вопрос о том, значимо ли $m_1 - m_2$ отличается от нуля, причем для упрощения предполагалось равенство параметров σ_1 и σ_2 . Если вместо $\bar{x} - \bar{y}$ в величине u , определяемой формулой (31.2.1), взять $\bar{x} - \bar{y} - (m_1 - m_2)$ и предположить, что $\sigma_1 = \sigma_2$, то полученная величина будет иметь распределение Стьюдента с $n_1 + n_2 - 2$ степенями свободы. Отсюда мы получаем, так же,

* В этом месте мы, возможно, выходим за границы теории Неймана [167]. Это же замечание относится к соответствующему месту в примере 2.

как и в предыдущем примере, следующие доверительные пределы для неизвестной разности $m_1 - m_2$:

$$(34.4.5) \quad \bar{x} - \bar{y} \pm t_p \sqrt{\frac{(n_1 + n_2)(n_1 s_1^2 + n_2 s_2^2)}{n_1 n_2 (n_1 + n_2 - 2)}}.$$

Здесь следует брать t_p с $n_1 + n_2 - 2$ степенями свободы. Теперь мы сделаем некоторые замечания относительно общего случая, когда σ_1 и σ_2 могут иметь любые значения.

Любой точке $(m_1, m_2, \sigma_1, \sigma_2)$ пространства параметров соответствует совместное распределение $n_1 + n_2$ величин x_i и y_i , представляемых переменной точкой в $(n_1 + n_2)$ -мерном пространстве R . Пусть заданы четыре постоянные k_1, k_2, c_1, c_2 , удовлетворяющие единственному условию $k_1 < k_2$. Для каждой точки пространства параметров соотношение

$$k_1 < c_1 \frac{\bar{x} - m_1}{s_1} + c_2 \frac{\bar{y} - m_2}{s_2} < k_2$$

спределяет некоторое множество S точек $(x, y) = (x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2})$ в пространстве R . Поскольку случайные величины

$$t = \sqrt{n_1 - 1} \frac{\bar{x} - m_1}{s_1} \text{ и } u = \sqrt{n_2 - 1} \frac{\bar{y} - m_2}{s_2}$$

независимы и имеют распределения Стьюдента соответственно с n_1 и n_2 степенями свободы, вероятность того, что выборочная точка (x, y) принадлежит множеству S , равна

$$(34.4.6) \quad J = \int \int s_{n_1-1}(t) s_{n_2-1}(u) dt du,$$

где интегрирование распространяется на область, определяемую соотношением

$$k_1 < \frac{c_1 t}{\sqrt{n_1 - 1}} + \frac{c_2 u}{\sqrt{n_2 - 1}} < k_2.$$

Величина J не зависит от параметров, и множество S соответствует множеству $S(\alpha)$ параграфа 34.3. Соотношение, определяющее множество S , можно записать в эквивалентном виде

$$(34.4.7) \quad \frac{c_1 \bar{x}}{s_1} + \frac{c_2 \bar{y}}{s_2} - k_2 < \frac{c_1 m_1}{s_1} + \frac{c_2 m_2}{s_2} < \frac{c_1 \bar{x}}{s_1} + \frac{c_2 \bar{y}}{s_2} - k_1.$$

Для каждой фиксированной точки (x, y) это соотношение задает некоторое цилиндрическое множество $\Sigma(x, y)$ в четырехмерном пространстве величин $(m_1, m_2, \sigma_1, \sigma_2)$, основанием которого является полоса, ограниченная двумя параллельными прямыми в подпространстве (m_1, m_2) . Таким образом, в соответствии с параграфом 34.3, множество $\Sigma(x, y)$ является доверительной областью для m_1 и m_2 с коэффициентом доверия J , определяемым формулой (34.4.6).

Каждый выбор постоянных k_1, k_2, c_1 и c_2 порождает, по (34.4.7), некоторую доверительную область для m_1 и m_2 , причем соответствующий коэффициент доверия дается формулой (34.4.6). Подходящим выбором этих постоянных можно добиться того, чтобы коэффициент доверия принял наперед заданное значение $1 - \epsilon$.

Как и в предыдущем примере, предположим теперь, что задан интервал (a, b) и извлечены две выборки x_0 и y_0 . По значениям $\bar{x}_0, \bar{y}_0, s_1^0$ и s_2^0 , вычисленным для этих выборок, определим k_1, k_2, c_1 и c_2 так, чтобы

$$c_1 = s_1^0, \quad c_2 = -s_2^0, \quad k_1 = \bar{x}_0 - \bar{y}_0 - b; \quad k_2 = \bar{x}_0 - \bar{y}_0 - a.$$

Как и любые другие значения наших постоянных, эти значения соответствуют некоторому правилу определения доверительных областей для m_1 и m_2 . Подставляя наши значения постоянных в (34.4.7), мы обнаружим, что в случае выборок x_0 и y_0 это правило приводит к области

$$a < m_1 - m_2 < b,$$

а область интегрирования в выражении (34.4.6) для коэффициента доверия имеет вид

$$(34.4.8) \quad \bar{x}_0 - \bar{y}_0 - b < \frac{s_1^0 t}{\sqrt{n_1 - 1}} - \frac{s_2^0 u}{\sqrt{n_2 - 1}} < \bar{x}_0 - \bar{y}_0 - a.$$

Таким образом, существует правило определения доверительных областей для m_1 и m_2 , которое в частном случае выборок x_0 и y_0 приводит к области $a < m_1 - m_2 < b$, причем это правило связано с коэффициентом доверия J , заданным формулой (34.4.6), где интегрирование распространяется на область (34.4.8). В этом смысле можно говорить, что выборки x_0 и y_0 приписывают области $a < m_1 - m_2 < b$ коэффициент доверия J .

Отсюда можно вывести критерий значимости, предложенный Беренсом и Фишером ([60] и [105] — [109]). Пусть заданы две выборки со

средними значениями \bar{x} и \bar{y} и стандартными отклонениями s_1 и s_2 , и пусть θ такой угол, что

$$\frac{s_1}{\sqrt{n_1 - 1}} = r \sin \theta, \quad \frac{s_2}{\sqrt{n_2 - 1}} = r \cos \theta,$$

где

$$r = \sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}.$$

Рассмотрим интеграл J в (34.4.6), распространенный по области

$$t \sin \theta - u \cos \theta > d,$$

и определим d так, чтобы $J = \delta$, где δ — заданное число, удовлетворяющее неравенству $0 < \delta < 1$. При фиксированном δ величина d есть функция от n_1 , n_2 и θ , которую можно вычислить, если значения аргументов известны. Если $\bar{x} - \bar{y} > dr$, то область $m_1 \leq m_2$, в соответствии с вышесказанным, имеет коэффициент доверия меньший, чем δ . Аналогично, если $\bar{x} - \bar{y} < -dr$, то область $m_1 \geq m_2$ также имеет коэффициент доверия меньший, чем δ . Если δ достаточно мало, то средние значения \bar{x} и \bar{y} следует считать отличающимися друг от друга значимо, когда $|\bar{x} - \bar{y}| > dr$. Для применения этого критерия имеются таблицы (см. [223] и [262]).

Пример 3. Среднее значение конечной совокупности (см. параграф 34.2, пример 3). Предположим, что имеется совокупность, состоящая из большого, но конечного числа N индивидуумов, среди которых распределен некоторый признак x . Для среднего значения, дисперсии и других характеристик величины x в полной совокупности мы используем обычные обозначения: m , s^2 , γ_1 и т. д. Требуется оценить неизвестное среднее значение совокупности m с помощью метода репрезентативной выборки (см. параграф 25.7). Извлечем беспорядочно случайную выборку в n индивидуумов и обозначим через $\bar{x} = \sum \frac{x_i}{n}$ и $s^2 = \sum \frac{(x_i - \bar{x})^2}{n}$ среднее значение и дисперсию n наблюдений выборочных значений величины x . Тогда имеем (см., например, Нейман [160], Хагстрем [121a])

$$E(\bar{x}) = m, \quad D^2(\bar{x}) = \frac{N-n}{N-1} \frac{s^2}{n}, \quad E(s^2) = \frac{N}{N-1} \frac{n-1}{n} s^2,$$

$$D^2(s^2) = \frac{N(N-n)}{(N-1)^2(N-2)(N-3)} \frac{(n-1)s^4}{n^3} \left[2nN^2 - 6(n+1)(N-1) + (nN - N - n - 1)(N-1)\gamma_2 \right],$$

где $\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$ — коэффициент эксцесса (см. параграф 15.8) для совокупности.

Если и n и $N-n$ велики, то \bar{x} приблизительно нормально, так что величина $\sqrt{\frac{(N-1)n}{N-n}}(\bar{x}-m)$ приблизительно нормальна $(0, s)$. Формулы для среднего значения и дисперсии величины s^2 можно записать в виде

$$E\left(\frac{q^2}{\sigma^2}\right) = \frac{N(n-1)}{N-n} \left[1 + O\left(\frac{1}{N}\right) \right],$$

$$D^2\left(\frac{q^2}{\sigma^2}\right) = \frac{2N(n-1)}{N-n} \left[1 + \frac{n-1}{2n} \gamma_2 + O\left(\frac{1}{N}\right) \right].$$

где $q^2 = \frac{Nns^2}{N-n}$. Если предположить, что можно пренебречь эксцессом совокупности γ_2 , то из (18.1.6) будет следовать, что при большом N величина $\frac{q^2}{\sigma^2}$ имеет приблизительно такое же среднее значение и дисперсию, как распределение χ^2 с $\frac{N(n-1)}{N-n}$ степенями свободы. Хотя в этом случае точное распределение для s^2 или q^2 неизвестно, можно в первом приближении предположить, что величина

$$t = \frac{\sqrt{\frac{(N-1)n}{N-n}}(\bar{x}-m)}{\sqrt{\frac{N-n}{N(n-1)}q^2}} = \sqrt{\frac{(N-1)(n-1)}{N-n}} \cdot \frac{\bar{x}-m}{s}$$

имеет распределение Стьюдента (18.2.4), где вместо n взято $\frac{N(n-1)}{N-n}$. Для неизвестного среднего значения совокупности мы тогда получим, как и в примере 1, p -процентные доверительные пределы

$$\bar{x} \pm t_p s \sqrt{\frac{N-n}{(N-1)(n-1)}}.$$

КРИТЕРИИ ЗНАЧИМОСТИ, II

ГЛАВА 35

ОБЩАЯ ТЕОРИЯ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ *)

35.1. Выбор критерия значимости. В предварительном обзоре проблем статистических выводов, приведенном в главе 26, введение критерия значимости для статистической гипотезы было описано (см. параграфы 26.2 и 26.4) в следующих общих выражениях: если требуется проверить, согласуется ли множество выборочных значений с заданной гипотезой H , то мы рассматриваем распределение выборки и вычисляем некоторую подходящую меру $D \geq 0$ отклонения этого распределения от гипотетического распределения; по выборочному распределению величины D мы определяем критическое значение D_0 , такое, что если гипотеза H верна, то $P(D > D_0) = \epsilon$, где ϵ — заранее заданный уровень значимости. Если в конкретном случае мы обнаружим отклонение $D > D_0$, то гипотеза H отвергается, в то время как появление значения $D \leq D_0$ считается совместимым с гипотезой H , которая тогда принимается.

Приняв это правило, мы с вероятностью, равной ϵ , можем отвергнуть в действительности справедливую гипотезу H . Поскольку ϵ можно выбрать произвольно, эту вероятность можно сделать сколь угодно малой.

Описанный общий принцип, лежащий в основе всех рассмотренных в главах 30—31 частных критериев, несомненно, во многом опирается на интуицию. При данной гипотезе появление очень большого отклонения D имеет очень малую вероятность. Если такое отклонение обнаружено в нашем конкретном случае, то мы, естественно, склонны считать гипотезу опровергнутой опытом. С другой стороны, появление какого-либо „умеренного“ отклонения принадлежит к разряду событий, которых следует ожидать, если гипотеза верна.

Исследуем, однако, этот принцип более подробно. Предположим, например, что D имеет непрерывное распределение с плотностью веро-

*) См. список в начале гл. 32.

ятности, подобной распределению χ^2 при $n > 2$ (см. фиг. 19, стр. 260). Тогда справедливо, что при гипотезе H вероятность большого отклонения, скажем, $D > D_0$, мала. Действительно, эта вероятность равна площади под графиком плотности вероятности, расположенной *справа* от ординаты точки D_0 , и D_0 всегда можно задать так, чтобы такая площадь была равна любому наперед заданному $\epsilon > 0$. Но равным образом верно и то, что появление очень *малого* отклонения, скажем, $D < D_1$, также имеет малую вероятность, так как D_1 можно задать так, чтобы площадь под кривой *слева* от ординаты точки D_1 была равна ϵ . Если мы условимся отвергать гипотезу H , когда $D < D_1$, и принимать эту гипотезу в остальных случаях, то мы получим ту же самую вероятность ϵ отвергнуть в действительности верную гипотезу. Вообще можно бесконечным количеством способов выбрать такое точечное множество S , что D имеет значение из S с вероятностью $P(D \in S) = \epsilon$, если гипотеза H верна. Для какого-нибудь такого множества S рассмотрим критерий, по которому гипотеза H отвергается, когда D имеет значение из S , и принимается в остальных случаях. Вероятность ошибочно отвергнуть в действительности справедливую гипотезу H всегда будет равна ϵ , так что с этой точки зрения все критерии, основанные на различных множествах S , эквивалентны. Интуитивно же мы склоняемся к критерию D_0 , где S есть интервал больших отклонений $D > D_0$ и определенно противимся критерию D_1 , согласно которому мы отвергаем гипотезу как раз в тех случаях, когда отклонения малы. Однако можно ли привести какие-нибудь разумные доводы в пользу некоторых определенных множеств S сравнительно с остальными множествами?

В качестве примера, рассмотрим критерий χ^2 . В главе 30 мы называли наблюденное значение χ^2 значимым, если оно превосходило $p (= 100\%)$ -процентное значение χ_{p}^2 . Это, очевидно, соответствует упомянутому выше критерию D_0 , и множеством S здесь является интервал $\chi^2 > \chi_p^2$. Если число n степеней свободы велико, то $\sqrt{2\chi^2}$ можно считать нормальным ($\sqrt{2n-1}, 1$), и . то же самое множество S можно приблизенно представить интервалом $\sqrt{2\chi^2} > \sqrt{2n-1} + \lambda_{2p}$ или $\chi^2 > \frac{1}{2}(\sqrt{2n-1} + \lambda_{2p})^2$, (λ_{2p} — $2p$ -процентное значение нормального отклонения); при этом площадь под *правым* концом приблизительно нормальной кривой равна $\frac{p}{100} = \epsilon$. Однако в последнем случае имеет смысл принимать во внимание оба конца нормальной кривой и считать χ^2 значимым, если $|\sqrt{2\chi^2} - \sqrt{2n-1}| > \lambda$. В этом случае множество S состоит из

двух интервалов $\chi^2 < \frac{1}{2}(\sqrt{2n-1} - \lambda_p)^2$ и $\chi^2 > \frac{1}{2}(\sqrt{2n-1} + \lambda_p)^2$. В обоих случаях вероятность того, что наша гипотеза будет несправедливо отвергнута, равна ϵ .

Далее, мера отклонения D определяется вовсе не однозначно. Например, согласие гипотетического распределения с выборкой можно измерять величинами χ^2 , ω^2 и т. д. Аналогичным образом, отклонение нормальной выборки от гипотезы о том, что среднее значение совокупности равно m , можно измерять, например, величинами $|\bar{x} - m|$ или $|z - m|$, где \bar{x} и z — среднее значение и медиана выборки, и т. д. Для каждой меры отклонения Δ можно бесконечным числом способов найти такое точечное множество Σ , что $P(\Delta \subset \Sigma) = \epsilon$, если гипотеза H верна. Критерий, согласно которому мы отвергаем гипотезу H , когда Δ принимает значение из Σ , и принимаем эту гипотезу в остальных случаях, попрежнему отвечает заданной вероятности ϵ забраковать в действительности верную гипотезу H .

Очевидно, возникает важная проблема научиться сравнивать различные возможные критерии для данной гипотезы. Можно ли присвоить разумный смысл утверждению, что один из двух критериев, соответствующих одному и тому же значению ϵ , „лучше“ или „более эффективен“, чем другой?

В течение последних лет этой проблеме был посвящен ряд работ Неймана, Е. Пирсона и их последователей. Мы отсылаем читателя к серии фундаментальных статей [170]—[173] Неймана и Пирсона и к общему изложению этой теории Нейманом [168], где можно найти многочисленные библиографические указания.

Основную идею теории Неймана — Пирсона можно вкратце описать следующим образом. Если критерий значимости применяется на практике, то в каждом случае возникает следующая альтернатива: можно решить или отвергнуть или принять предложенную гипотезу H и действовать в соответствии с этим решением *). И в том, и в другом случае наше решение может оказаться ошибочным, так как можно отвергнуть гипотезу H , когда она в действительности верна, и принять

*) Конечно, существует и третья возможность: можно решить оставить гипотезу под сомнением и отложить дальнейшие действия до тех пор, пока не будут получены новые данные. Однако, здесь мы рассматриваем случай, когда мы уже располагаем всеми доступными нам данными, и необходимо решить, в каком направлении действовать.

гипотезу, когда она неверна *). Повидимому, совершенно разумно исходить из того принципа, что при выборе критерия мы должны стараться, *насколько это возможно, уменьшить возможность совершения обоих видов ошибок*.

Для признания критерия гипотезы H „хорошим“ необходимо потребовать, чтобы при этом критерии вероятность отвергнуть гипотезу H , если она в действительности верна, была мала, а вероятность отвергнуть H , когда она ошибочна, была велика. Из двух критериев, соответствующих одной и той же вероятности ε отвергнуть в действительности правильную гипотезу H , следует предпочтеть тот, который дает большую вероятность отвергнуть H , когда эта гипотеза ошибочна.

Теперь мы покажем некоторые применения этого общего принципа. Нам придется ограничиться лишь кратким изложением некоторых из наиболее элементарных черт этой важной теории, которая в настоящее время интенсивно развивается.

35.2. Простые и сложные гипотезы. Рассмотрим n случайных величин x_1, \dots, x_n с совместным распределением в R_n непрерывного типа, определяемым вероятностной функцией $P(S; \alpha) = P(S; \alpha_1, \dots, \alpha_k)$ известного математического вида, но содержащей k неизвестных параметров $\alpha_1, \dots, \alpha_k$, или соответствующей плотностью вероятности $f(x, \alpha) = f(x_1, \dots, x_n; \alpha_1, \dots, \alpha_k)$.

Если, в частности, x_i — независимые величины, имеющие одно и то же распределение, то мы имеем обычный случай выборки в n значений из этого распределения. Однако, как было указано в аналогичном случае в параграфе 32.8 (см. также параграф 34.8), приведенные выше определения пригодны и в более общих случаях, как, например, в случае, когда x_i состоят из нескольких независимых выборок, извлеченных, может быть, из неодинаковых распределений. И в общем случае мы будем называть точку $x = (x_1, \dots, x_n)$ и пространство R_n соответственно

*) Этой возможностью двойной ошибки наша проблема отличается от того, что мы имели в теории нахождения оценок. Например, когда мы утверждали, что неизвестное значение некоторого параметра принадлежит такому-то доверительному интервалу, то это утверждение могло быть или правильным, или ошибочным, но существовала лишь одна возможность ошибиться, именно указать интервал, который на самом деле не содержит истинного значения параметра.

выборочной точкой и пространством выборок. Параметры α , изображаются точкой $\alpha = (\alpha_1, \dots, \alpha_k)$ в пространстве параметров P_k .

Предположим теперь, что выборочная точка x определена посредством одного случайного эксперимента, соответствующего составной величине (x_1, \dots, x_n) . Гипотеза о том, что неизвестная параметрическая точка α в распределении этой составной величины принадлежит заданному точечному множеству ω из пространства параметров P_k , будет коротко называться *гипотезой H* . Если ω состоит из единственной точки α_0 , то будем называть гипотезу *простой*, в других случаях гипотеза будет называться *сложной*. Очевидно, простая гипотеза полностью задает распределение, а сложная гипотеза оставляет распределение более или менее неопределенным.

Каждая параметрическая точка α , считаемая *a priori* возможной, будет называться *допустимой точкой*, соответствующей *допустимой гипотезе*. Множество Ω всех допустимых точек может совпадать со всем пространством параметров P_k или составлять лишь часть пространства P_k .

Если x_i образуют выборку независимых значений из собственного нормального распределения, о параметрах m и σ которого не имеется никаких сведений, то множество Ω допустимых гипотез состоит из точек полу-плоскости $\sigma > 0$. Гипотеза $m=0, \sigma=1$ является простой, а гипотеза $m=0$ (без указания значения σ) — сложной.

35.3. Критерии для простых гипотез. Наиболее мощные критерии. Предположим, что требуется проверить простую гипотезу H_0 , согласно которой неизвестная параметрическая точка α совпадает с данной точкой α_0 . Критерий для этой гипотезы будет заключаться в правиле, согласно которому H_0 должна быть отвергнута, когда наблюденная точка x принадлежит некоторому множеству S из R_n , а в остальных случаях H_0 должна быть принята. Множество S будет называться *критическим множеством* критерия; сам критерий, опирающийся на критическое множество S , часто будет коротко называться *критерием S* .

Если критическое множество S фиксировано, то вероятность отвергнуть гипотезу H_0 совпадает с вероятностью $P(S; \alpha_0)$ попадания выборочной точки во множество S . Эта вероятность есть функция от k переменных $\alpha_1, \dots, \alpha_k$; условимся называть ее *функцией мощности*

критерия. В соответствии с общей установкой, высказанной в параграфе 35.1, мы должны попытаться построить критерий так, чтобы его функция мощности была *мала*, если H_0 верна (т. е. если α совпадает с α_0) и *велика*, если H_0 неверна (т. е. если α — какая-либо допустимая точка, отличная от α_0).

Поскольку распределение величины x непрерывно, всегда существует бесконечное количество множеств S таких, что $P(S; \alpha_0) = \varepsilon$, где ε — заранее заданный уровень значимости. Если за критическое множество критерия избрать какое-либо из этих множеств S , то вероятность отвергнуть H_0 , когда эта гипотеза на самом деле верна, будет равняться ε ; будем коротко говорить, что мы имеем дело с критерием *уровня* ε . Среди всех критериев уровня ε теперь требуется отыскать такой критерий, при котором вероятность отвергнуть гипотезу H_0 , когда эта гипотеза в действительности неверна, будет как можно большей, т. е. такой критерий, функция мощности которого $P(S; \alpha)$ принимала бы возможно большие значения для любой допустимой точки $\alpha \neq \alpha_0$.

Пусть α_1 — фиксированная допустимая точка, отличная от α_0 . Так как мы можем произвольно изменять значения плотности вероятности на множествах меры нуль, то всегда можно предполагать, что $f(x, \alpha_0)$ и $f(x, \alpha_1)$ конечны и определены для всех x . Тогда для любого $c \geq 0$ будет вполне определено множество X всех точек x , для которых

$$(35.3.1) \quad f(x, \alpha_1) \geq c f(x, \alpha_0).$$

При возрастании c от 0 до ∞ функция $\psi(c) = P(X; \alpha_0)$ никогда не возрастает. Далее, $\psi(0) = 1$, и легко обнаружить, что $0 \leq \psi(c) \leq \frac{1}{c}$, так что $\psi(c) \rightarrow 0$ при $c \rightarrow \infty$. Чтобы избежать тривиальных затруднений, предположим *), что существует такое значение c , для которого $\psi(c) = \varepsilon$.

*). Всегда существует такое значение c , что $\psi(c - 0) \geq \varepsilon$, $\psi(c + 0) \leq \varepsilon$. Исключительный случай, когда $\psi(c)$ фактически не принимает значения ε , можно включить в наше доказательство, слегка изменив определение множества X . Действительно, $\psi(c - 0) - \psi(c + 0)$ есть интеграл от $f(x, \alpha_0)$ по множеству Z всех точек x , таких, для которых в (35.3.1) достигается знак равенства. Исключая из множества X надлежащим образом выбранную часть множества Z , всегда можно получить множество, удовлетворяющее соотношению (35.3.2). Во всех точках этого видоизмененного множества X удовлетворяется соотношение (35.3.1), а для всех точек дополнительного множества имеем $f(x, \alpha_1) \leq c f(x, \alpha_0)$. Очевидно, этого достаточно для получения вывода (35.3.6).

Для соответствующего множества X имеем

$$(35.3.2) \quad P(X; \alpha_0) = \int_X f(x, \alpha_0) dx = \varepsilon$$

Пусть теперь S — критическое множество какого-нибудь критерия уровня ε , так что

$$(35.3.3) \quad P(S; \alpha_0) = \int_S f(x, \alpha_0) dx = \varepsilon$$

Покажем теперь, что

$$(35.3.4) \quad P(X; \alpha_1) \geq P(S; \alpha_1).$$

Таким образом, среди всех критериев уровня ε критерий X дает наибольшее возможное значение вероятности забраковки гипотезы H_0 , когда верна другая гипотеза H_1 , согласно которой $\alpha = \alpha_1$. В соответствии с этим, критерий X будет называться наиболее мощным критерием для гипотезы H_0 относительно гипотезы H_1 , среди всех критериев уровня ε .

Из формул (35.3.2) и (35.3.3) получаем

$$(35.3.5) \quad P(X - SX; \alpha_0) = \varepsilon - P(SX; \alpha_0) = P(S - SX; \alpha_0).$$

Из определения (35.3.1) множества X следует, что для любого x , не принадлежащего множеству X , $c f(x, \alpha_0) > f(x, \alpha_1)$. Поэтому

$$(35.3.6) \quad P(X - SX; \alpha_1) \geq c P(X - SX; \alpha_0) = c P(S - SX; \alpha_0) \geq P(S - SX; \alpha_1).$$

Прибавляя к обеим частям последнего неравенства $P(SX; \alpha_1)$, получим (35.3.4).

Может случиться, что для всех допустимых точек $\alpha_1 \neq \alpha_0$ мы получим одно и то же множество X . В таком случае будем говорить, что среди всех критериев уровня ε критерий X является равномерно наиболее мощным критерием для гипотезы H_0 относительно всего множества Ω допустимых гипотез. Если равномерно наиболее мощный критерий существует, то ясно, что его можно считать лучшим сравнительно с любым другим критерием того же самого уровня ε . К сожалению, такое положение встречается очень редко.

Рассмотрим случай, когда x_i суть n выборочных значений из распределения, содержащего единственный неизвестный параметр α , и предложим, что

для α существует достаточная оценка α^* . По параграфу 32.4, совместную плотность вероятности $f(x_1, \dots, x_n; \alpha)$ в этом случае можно записать в виде произведения $g(\alpha^*; \alpha) H(x_1, \dots, x_n)$, где H не зависит от α . Если $H > 0$, то (35.3.1) принимает вид $g(\alpha^*; \alpha) \geq cg(\alpha^*; \alpha_0)$. Если удовлетворены некоторые общие условия регулярности, то множеством X будет служить область, ограниченная гиперповерхностью $g(\alpha^*, \alpha_1) = cg(\alpha^*; \alpha_0)$, и это уравнение эквивалентно некоторому количеству уравнений вида $\alpha^* = \text{const}$. Если для различных взаимно исключающих друг друга гипотез α_1 граничной гиперповерхностью множества X всегда служит один и тот же представитель семейства $\alpha^* = \text{const}$, то равномерно наиболее мощный критерий существует. Однако на примерах можно показать (см. Нейман и Пирсон [173]), что это свойство выполняется не всегда. Поэтому даже в рассматриваемом простом случае мы не можем утверждать существование равномерно наиболее мощного критерия без введения дополнительных ограничений. (См., далее, Нейман [165], где этот вопрос рассматривается в связи с упомянутой в параграфе 34.2 проблемой нахождения кратчайших доверительных интервалов.)

Еще более простым случаем, в котором наши рассуждения приводят к полному решению задачи, является случай наличия лишь двух различных гипотез. В этом случае совместную плотность вероятности величин x_i можно записать в виде $(1 - \alpha)f_0(x) + \alpha f_1(x)$, где f_0 и f_1 — заданные плотности вероятности, а допустимыми значениями α являются 0 и 1. Подлежащей проверке гипотезой H_0 является гипотеза $\alpha = 0$, т. е. гипотеза о том, что наблюденные выборочные значения извлечены из распределения с плотностью вероятности f_0 , тогда как единственной допустимой альтернативой является f_1 . Тогда нам следует отыскать множество X всех таких точек x , для которых $f_1 \geq cf_0$, где c определяется условием $\int_X f_0(x) dx = \varepsilon$. Критерий, согласно которому гипотеза H_0

отвергается, когда наблюденная выборочная точка принадлежит множеству X и принимается в остальных случаях, будет наиболее мощным критерием уровня ε . Этот критерий можно применить, например, к проблемам следующего типа (см. Квензель, Эссен—Меллер [203]): предположим, что измеряются некоторые характеристики x_i двух людей A и B , причем требуется проверить гипотезу, что A является отцом B . Если известны распределения величин x_i среди детей всех лиц, имеющих те же характеристики, что и A а также среди всей совокупности людей, причем плотности вероятности этих распределений мы обозначим соответственно f_0 и f_1 , то по нашей гипотезе выборочные значения характеристик лица B должны быть извлечены из совокупности с плотностью вероятности f_0 , поскольку альтернативной гипотезой является f_1 . Эта гипотеза может быть проверена с помощью вышеописанного критерия.

Другой пример будет приведен в следующем параграфе.

35.4. Несмещенные критерии. Ограничимся теперь случаем единственного неизвестного параметра α . Пусть допустимые значения α образуют интервал A , и предположим, что почти для всех $x = (x_1, \dots, x_n)$

плотность вероятности $f(x, \alpha)$ имеет во всех внутренних точках α интервала A частную производную $\frac{\partial f}{\partial \alpha} = f_1(x, \alpha)$, такую, что $|f_1(x, \alpha)| < F(x)$, где $F(x)$ — функция, интегрируемая по R_n . Тогда, согласно параграфу 7.3, производная

$$(35.4.1) \quad \frac{\partial P(S; \alpha)}{\partial \alpha} = \int_S f_1(x, \alpha) dx$$

существует для каждого множества S в R_n и для каждого α из A .

Предположим, что рассматривается простая гипотеза H_0 , согласно которой $\alpha = \alpha_0$, где α_0 — внутренняя точка интервала A , и обозначим через S критическое множество некоторого критерия с уровнем ϵ . Тогда функция мощности $P(S; \alpha)$ будет функцией от α , такой, что $P(S; \alpha_0) = \epsilon$. Если для некоторого допустимого $\alpha_1 \neq \alpha_0$ мы получим $P(S; \alpha_1) < \epsilon$, то это будет означать, что *менее вероятно, что мы отвергнем гипотезу H_0 в случае, когда верна противоположная гипотеза H_1 , согласно которой $\alpha = \alpha_1$, чем в случае, когда верна сама гипотеза H_0* . Критерий, обладающий таким, очевидно, нежелательным, свойством, мы будем называть *смещенным* критерием.

С другой стороны, если $P(S; \alpha) \geq \epsilon$ для всех допустимых точек α то критерий и его критическое множество S будут называться *несмешенными*. Поскольку $P(S; \alpha_0) = \epsilon$ и производная (35.4.1) существует для всех α из A , в случае несмешенного критерия имеем

$$(35.4.2) \quad \left(\frac{\partial P(S; \alpha)}{\partial \alpha} \right)_0 = 0,$$

В обобщение определения (35.3.1) рассмотрим теперь множество X всех таких точек x , для которых

$$(35.4.3) \quad f(x, \alpha_1) \geq c f(x, \alpha_0) + c_1 f_1(x, \alpha_0),$$

где $\alpha_1 \neq \alpha_0$ есть внутренняя точка интервала A , а постоянные $c \geq 0$ и c_1 определяются так, чтобы удовлетворялись следующие условия *)

$$(35.4.4) \quad \begin{aligned} P(X; \alpha_0) &= \int_X f(x; \alpha_0) dx = \epsilon, \\ \left(\frac{\partial P(X; \alpha)}{\partial \alpha} \right)_0 &= \int_X f_1(x; \alpha_0) dx = 0. \end{aligned}$$

*) Такими же рассуждениями, как для (35.3.2), можно показать, что постоянные c и c_1 , обладающие указанными свойствами, могут быть определены всегда, кроме некоторых исключительных случаев, в которых, однако, можно изменить определение множества X таким же образом, как это было указано в предыдущей сноске, т. е. исключить из X некоторое подмножество множества Z всех точек x , для которых в (35.4.3) достигается знак равенства.

Критическое множество S любого несмешенного критерия уровня ϵ удовлетворяет соотношению (35.3.5), и из формул (35.4.2) и (35.4.4) получаем аналогичное соотношение

$$\left(\frac{\partial P(X - SX; \alpha)}{\partial \alpha} \right)_0 = - \left(\frac{\partial P(SX; \alpha)}{\partial \alpha} \right)_0 = \left(\frac{\partial P(S - SX; \alpha)}{\partial \alpha} \right)_0$$

Таким же способом, как и в параграфе 35.3, затем получаем

$$P(X; \alpha_1) \geq P(S; \alpha_1).$$

Может случиться, что для всех допустимых точек $\alpha_1 \neq \alpha_0$ мы получим одно и то же множество X . Это будет означать, что критерий X является несмешенным и среди всех несмешенных критериев дает наибольшее возможное значение вероятности отвергнуть гипотезу H_0 , когда верна любая противоположная гипотеза $\alpha = \alpha_1^*$). Такой критерий X будет называться *наиболее мощным несмешенным критерием для гипотезы H_0* .

Рассмотрим случай выборки в n значений x_1, \dots, x_n из нормального распределения с известным стандартным отклонением σ и неизвестным средним значением m , и пусть требуется проверить гипотезу H_0 , согласно которой $m = m_0$. Сначала постараемся найти условия существования *равномерно наиболее мощного критерия*, соответствующего заданному уровню ϵ . Для любого $m_1 \neq m_0$ соотношение (35.3.1) принимает вид

$$(35.4.5) \quad \frac{f(\mathbf{x}, m_1)}{f(\mathbf{x}, m_0)} = e^{-\frac{1}{2\sigma^2} \sum [(x_i - m_1)^2 - (x_i - m_0)^2]} = e^{M\lambda - \frac{1}{2}M} \geq c,$$

где $M = \sqrt{n} \frac{m_1 - m_0}{\sigma}$, $\lambda = \sqrt{n} \frac{\bar{x} - m_0}{\sigma}$. Предположим сначала, что $m_1 > m_0$.

Тогда $M > 0$, и если мы положим

$$c = e^{M\lambda_{2p} - \frac{1}{2}M^2}$$

где $p = 100\epsilon$ и λ_{2p} есть $2p$ -процентное значение нормального отклонения, то неравенство (35.4.5) будет выполняться на множестве X всех таких точек $\mathbf{x} = (x_1, \dots, x_n)$, для которых $\lambda \geq \lambda_{2p}$ или $\bar{x} \geq m_0 + \frac{\lambda_{2p}^2}{\sqrt{n}}$. Очевидно, это множество не зависит от m_1 , и вероятность того, что \mathbf{x} принадлежит множеству X , согласно гипотезе H_0 , равняется $\frac{p}{100} = \epsilon$, так что условие (35.3.2) выполняется. Таким образом, критерий, основанный на множестве X и заключающийся

*). Это утверждение является некоторым видоизменением предложенного полученного Нейманом и Пирсоном [172].

в том, что H_0 отвергается, когда $\bar{x} \geq m_0 + \frac{\lambda_2 p^\sigma}{\sqrt{n}}$, является равномерно наименее мощным критерием для H_0 относительно множества всех гипотез, по которым $m_1 > m_0$.

Для всех $m_1 < m_0$ таким же способом получим равномерно наименее мощный критерий, основанный на критическом множестве X , определяемом соотношением $\bar{x} \leq m_0 - \frac{\lambda_2 p^\sigma}{\sqrt{n}}$. Однако, если множество допустимых гипотез содержит значения m , лежащие как справа, так и слева от точки m_0 , мы более не получим одного и того же множества X для всех допустимых m_1 . Отсюда следует, что в таком случае равномерно наименее мощного критерия не существует.

Рассмотрим функцию мощности критерия, основанного на критическом множестве $\bar{x} \geq m_0 + \frac{\lambda_2 p^\sigma}{\sqrt{n}}$. Функция мощности равна вероятности того, что выборочная точка принадлежит этому множеству, если истинное среднее значение есть m . Указанная вероятность равна $1 - \Phi(z)$, где $z = \lambda_2 p + \sqrt{n} \frac{m_0 - m}{\sigma}$; она монотонно возрастает вместе с m и при $m = m_0$ принимает значение ε . Таким образом, при $m > m_0$ функция мощности $> \varepsilon$, так что мы имеем вероятность $> \varepsilon$ отвергнуть гипотезу H_0 , когда истинное среднее значение превышает m_0 . С другой стороны, если $m < m_0$, то функция мощности $< \varepsilon$; это означает, что критерий является смещенным. Конечно, подобными же свойствами обладает также критерий, основанный на множестве $\bar{x} \leq m_0 - \frac{\lambda_2 p^\sigma}{\sqrt{n}}$.

Теперь рассмотрим наилучший несмещенный критерий, используя тот же, что и выше, уровень $\varepsilon = \frac{p}{100}$. Условие (35.4.3) имеет здесь вид

$$e^{\frac{M\lambda - \lambda_2 M^2}{2}} \geq c + c'_1 \lambda,$$

где $c'_1 = c_1 \frac{\sqrt{n}}{\sigma}$. Постоянные c и c_1 всегда можно выбрать так, чтобы знак равенства здесь достигался, когда $\lambda = \pm \lambda_p$ и тогда множество X будет состоять из всех тех точек x , для которых $|\lambda| \geq \lambda_p$ или $|\bar{x} - m_0| \geq \frac{\lambda_p \sigma}{\sqrt{n}}$. Очевидно, это множество X удовлетворяет обоим условиям (35.4.4). Таким образом, обычный критерий, по которому гипотеза H_0 отвергается, когда абсолютное отклонение $|\bar{x} - m_0|$ превышает $\frac{\lambda_p \sigma}{\sqrt{n}}$, является наименее мощным несмещенным критерием для H_0 .

Функция мощности этого критерия равняется $\Phi(z') + 1 - \Phi(z'')$, где $z' = -\lambda_p + \sqrt{n} \frac{m_0 - m}{\sigma}$, $z'' = \lambda_p + \sqrt{n} \frac{m_0 - m}{\sigma}$, а m — истинное среднее

значение. Легко видеть, что эта функция достигает минимума при $t = t_0$ и значение этого минимума равно ε . При $t \neq t_0$ функция мощности всегда превосходит ε и стремится к 1 при $t \rightarrow \pm \infty$. В соответствии с вышеуказанным, график функции мощности лежит целиком выше соответствующего графика для любого другого несмещенного критерия. Функция мощности предыдущего критерия, основанного на множестве $\bar{x} \geq m_0 + \frac{\lambda_2 p^\sigma}{\sqrt{n}}$, при $t > t_0$ превышает рассматриваемую функцию мощности, но при $t < t_0$ оказывается меньше последней и даже стремится к нулю при $t \rightarrow -\infty$.

В обычных критериях, основанных на использовании стандартных ошибок (см. параграф 31.1), предполагалось, что исследуемую величину z с практической достаточной точностью можно считать распределенной нормально с известным стандартным отклонением $d(z)$. Если на основании одного наблюдения значения z мы проверяем гипотезу о том, что среднее значение $E(z)$ имеет некоторое определенное значение z_0 , и если реальные условия задачи позволяют ограничить множество Ω допустимых гипотез, например областью $E(z) \geq z_0$, то из вышеизложенного вытекает, что следует воспользоваться критерием, по которому наша гипотеза (при уровне $10\varepsilon = p\% = \alpha$) отвергается, когда $z \geq z_0 + \lambda_2 p^\sigma d(z)$. Такое положение иногда встречается на практике, например, когда мы имеем дело с данными относительно эффективности некоторого метода, который, повидимому, не ухудшает, а, может быть, улучшает качество производящейся продукции. С другой стороны, если мы не можем *a priori* ввести ограничение такого „одностороннего“ типа, то следует пользоваться лишь обычным критерием, основанным на абсолютном отклонении $|z - z_0|$.

35.5. Критерии для сложных гипотез. При переходе от простых гипотез к сложным теория значительно усложняется; мы ограничимся здесь лишь некоторыми краткими замечаниями, отсылая читателя за дальнейшими сведениями к оригиналным работам, упомянутым в параграфе 35.1.

Пользуясь общими обозначениями, введенными в параграфе 35.2, рассмотрим гипотезу H о том, что неизвестная параметрическая точка α принадлежит заданному множеству ω , являющемуся подмножеством множества Ω всех допустимых точек. Как и в случае простой гипотезы, критерий для гипотезы H будет представлять собой правило, по которому H отвергается, если наблюденная выборочная точка x принадлежит некоторому критическому множеству S , и принимается в остальных случаях. В соответствии с общим пожеланием, высказанным в параграфе 35.1, следует стараться определить множество S так, чтобы функция мощности $P(s; \alpha)$ была мала, когда α принадлежит к ω , и велика, когда α принадлежит подмножеству допустимых точек $\Omega - \omega$.

В некоторых случаях можно отыскать множество S (и даже семейство множеств), такое, чтобы $P(S; \alpha)$ постоянно равнялась заданному уровню ε для всех α из ω . Тогда мы будем говорить, что S подобно пространству выборок*) относительно множества ω . Критерий S , т. е. критерий, основанный на критическом множестве S , всегда приводит к вероятности ε отвергнуть гипотезу H , когда она на самом деле верна, каково бы ни было значение α из ω , и в соответствии с этим мы говорим, что критерий S есть критерий уровня ε .

Предположим теперь, что можно отыскать такой критерий X уровня ε , что при любом α из $\Omega - \omega$ и при любом критерии S уровня ε выполняется неравенство $P(X; \alpha) \geq P(S; \alpha)$. По аналогии с параграфом 35.3, будем говорить, что среди всех критериев уровня ε критерий X является равномерно наиболее мощным критерием для гипотезы H относительно множества допустимых гипотез $\Omega - \omega$.

Подобным же образом, если для критерия S уровня ε имеем $P(S; \alpha) \geq \varepsilon$ при всех допустимых α , то критерий будет называться несмешенным. Наиболее мощным несмешенным критерием является такой критерий X уровня ε , что $P(X; \alpha) \geq P(S; \alpha)$ для всех α из $\Omega - \omega$ и для всех несмешенных критериев S уровня ε .

Общие условия, при которых существуют эти классы критериев, и методы, с помощью которых такие критерии могут быть найдены, до сих пор известны лишь частично. Приведем лишь некоторые примеры без доказательств.

Рассмотрим n выборочных значений x_1, \dots, x_n из нормального распределения с неизвестными параметрами m и s . Пусть требуется проверить гипотезу о том, что $m = m_0$, без указания значения s . Множеством Ω всех допустимых гипотез в плоскости (m, s) является (см. параграф 35.2) полуплоскость $s > 0$, и множество ω состоит из той части прямой $m = m_0$, которая лежит в этой полуплоскости.

Обозначим через T какое-нибудь множество действительных чисел, для которого $\int s_{n-1}(t) dt = \varepsilon$, где $s_{n-1}(t)$ — плотность вероятности Стьюдентова

отношения $t = \sqrt{n-1} \frac{\bar{x} - m_0}{s}$, и пусть S обозначает множество всех точек x из R_n , для которых соответствующее значение t принадлежит множеству T . Тогда для каждого ε имеем $P(S; m_0, \varepsilon) = P(t \in T) = \varepsilon$, откуда следует, что S подобно пространству выборок относительно заданного множества ω .

*). Введение этого термина оправдывается тем, что множество $S = R_n$ удовлетворяет поставленному условию с $\varepsilon = 1$.

Если множество допустимых точек $\Omega - \omega$ ограничивается точками с $t > t_0$ и если в качестве S берется множество всех x , для которых $t > t_{2p}$, где $p = 100\epsilon$, то можно показать (см. статьи, упомянутые в параграфе 35.1), что критерий S является равномерно наиболее мощным. Подобным же образом для точек $t < t_0$ критерий, основанный на множестве $t < -t_{2p}$, является равномерно наиболее мощным. Если допустимые точки содержат значения t , лежащие как справа, так и слева от значения t_0 , то равномерно наиболее мощного критерия не существует, но критерий, по которому гипотеза H отвергается при $|t| > t_p$, является наиболее мощным несмещенным критерием уровня ϵ . Все эти результаты аналогичны результатам, доказанным в параграфе 35.4 для случая, когда σ известно.

Случай разности между средними значениями двух нормальных распределений исследовали, исходя из функции мощности, Уэлч [229] и Хэу [127]. Ими было обнаружено, что критерий $|t| > t_p$, использованный в параграфах 31.2 и 31.3 (пример 4), является единственным удовлетворительным критерием для гипотезы $\mu_1 = \mu_2$, когда известно, что $\sigma_1 = \sigma_2$. Этот критерий может оказаться сильно смещенным, если допустимые гипотезы содержат случай $\sigma_1 \neq \sigma_2$.

ГЛАВА 36

ДИСПЕРСИОННЫЙ АНАЛИЗ

36.1. Изменчивость средних значений. *Дисперсионный анализ* объединяет группу статистических приемов, предложенных Р. Фишером ([13], [14]) в связи с некоторыми экспериментальными задачами в различных областях биологических исследований, в особенности в сельскохозяйственной статистике. Однако область применения этих приемов значительно шире, и они с успехом применяются в самых разнообразных экспериментах.

Предположим, что результатом эксперимента являются наблюденные значения x_1, \dots, x_n некоторых величин и что эти значения можно считать независимо извлеченными из нормальных распределений с постоянным, хотя и неизвестным стандартным отклонением σ . С другой стороны, средние значения t_i этих распределений могут меняться под влиянием некоторых факторов, входящих в эксперимент, как, например, под влиянием различных способов обработки, различия видов животных или растений, неоднородности почвы и т. п. Целью эксперимента является исследование этой изменчивости средних значений так что может потребоваться проверить различные гипотезы относительно этих величин, как, например, нулевую гипотезу (см. параграф 26.4) о том, что различие в способе обработки или вид животных не

влияет на средние значения, и т. п. Конечно, может потребоваться также отыскать оценки для некоторых средних значений или функций от средних значений.

При общей нулевой гипотезе о том, что все x_i имеют одно и то же среднее значение, несмешенной оценкой для неизвестной дисперсии σ^2 является, как известно, сумма квадратов отклонений от среднего значения выборки $\sum (x_i - \bar{x})^2$, поделенная на соответствующее число степеней свободы (именно, на $n - 1$). Основная идея дисперсионного анализа заключается в разбиении этой суммы квадратов отклонений на несколько компонент, каждая из которых соответствует действительной или предполагаемой причине изменчивости средних значений. Эти компоненты располагаются так, чтобы доставить критерии для различных гипотез о поведении средних значений или оценки для различных интересующих нас функций от средних значений.

В следующем параграфе мы подробно изучим этот метод на простом частном случае, а затем перейдем к более общим случаям.

36.2. Простая группировка величин. Рассмотрим простейший случай, в котором наблюденные величины разбиваются на r группы, причем i -я группа содержит n_i величин, предполагаемых нормальными (m_i, σ), где σ не зависит от i . Требуется исследовать свойства величин m_i и в первую очередь проверить нулевую гипотезу о том, что все m_i равны друг другу, т. е. что распределения всех групп одинаковы. В частном случае $r = 2$ эта задача сводится к задаче о разности между двумя средними значениями, уже рассматривавшейся в параграфах 31.2 и 34.4.

Пусть x_{ij} обозначает j -ю величину в i -й группе, $\bar{x}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} =$ среднее арифметическое величин i -й группы и $\bar{x} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} x_{ij} =$ $= \frac{1}{n} \sum_{i=1}^r n_i \bar{x}_{i\cdot}$ — среднее арифметическое всех $n = \sum_1^r n_i$ величин. Тогда имеем тождество

$$\sum (x_{ij} - \bar{x})^2 = \sum (x_{ij} - \bar{x}_{i\cdot})^2 + \sum (\bar{x}_{i\cdot} - \bar{x})^2,$$

где каждая сумма распространяется на все $n = \sum_1^r n_i$ величин. Таким

образом, полная сумма квадратов отклонений от общего среднего значения \bar{x} разбивается на две компоненты, именно: 1) сумму квадратов отклонений каждой величины от соответствующего группового среднего значения („сумма квадратов внутри групп“) и 2) сумму квадратов отклонений групповых средних значений от общего среднего значения („сумма квадратов между группами“). Это тождество имеет очевидное сходство с тождеством (21.9.1), употреблявшимся для определения корреляционного отношения.

Переписывая это тождество более подробно и одновременно меняя порядок слагаемых в правой части, получаем

$$(36.2.1) \quad \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = \sum_{i=1}^r n_i (\bar{x}_{i\cdot} - \bar{x})^2 + \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i\cdot})^2,$$

или, короче,

$$Q = Q_1 + Q_2;$$

тогда Q , Q_1 и Q_2 будут квадратичными формами от x_{ij} . Как известно (см. параграфы 11.11 и 29.3), с помощью ортогонального преобразования Q можно привести к виду $\sum_1^{n-1} y_i^2$, следовательно, эта форма имеет ранг $n-1$. Далее, Q_1 есть сумма квадратов r линейных форм $L_i = \sqrt{n_i}(\bar{x}_{i\cdot} - \bar{x})$, удовлетворяющих тождеству $\sum_1^r \sqrt{n_i} L_i = 0$, так что, согласно параграфу 11.6, Q_1 имеет ранг $\leq r-1$. Аналогично, Q_2 есть сумма квадратов $n-r$ линейных форм $L_{ij} = x_{ij} - \bar{x}_{i\cdot}$, удовлетворяющих r независимым соотношениям $\sum_{j=1}^{n_i} L_{ij} = 0$ ($i = 1, 2, \dots, r$), следовательно, Q_2 имеет ранг $\leq n-r$. Но, по параграфу 11.6, ранг Q не превосходит суммы рангов Q_1 и Q_2 , и отсюда следует, что ранги двух последних форм в точности равны, соответственно, $r-1$ и $n-r$, так что мы имеем следующее соотношение между рангами форм (36.2.1):

$$n-1 = (r-1) + (n-r).$$

Отсюда мы заключаем, согласно параграфу 11.11, что существует ортогональное преобразование n величин x_{ij} в новые величины y_1, \dots, y_n , такое, что три члена в (36.2.1) преобразуются в соответствующие

члены соотношения

$$\sum_{i=1}^{n-1} y_i^2 = \sum_{i=1}^{r-1} y_i^2 + \sum_{i=r}^{n-1} y_i^2.$$

По нашей гипотезе, величины x_{ij} независимы и распределены нормально с общим стандартным отклонением σ , и, следовательно, по параграфу 24.4 (см. также упражнение 16, стр. 351), то же справедливо для величин y_i . Таким образом, Q_1 и Q_2 независимы.

Предположим сначала, что верна нулевая гипотеза, т. е. что $m_i = m$ для всех i . Полагая $x_{ij} = m + \xi_{ij}$, мы получим независимые и нормальные $(0, \sigma)$ величины ξ_{ij} . Вводя это преобразование в Q , Q_1 и Q_2 и обозначая через $\bar{\xi}_i$ и $\bar{\xi}$ средние арифметические, соответствующие \bar{x}_i и \bar{x} , преобразуем три формы в идентичные выражения, в которых вместо буквы x повсюду будет стоять ξ . При указанном выше ортогональном преобразовании величины ξ_{ij} заменяются новыми величинами η_1, \dots, η_n , независимыми и нормальными $(0, \sigma)$. При этом Q , Q_1 и Q_2 примут соответственно вид $\sum_{i=1}^{n-1} \eta_i^2$, $\sum_{i=1}^{r-1} \eta_i^2$ и $\sum_{i=r}^{n-1} \eta_i^2$. В силу параграфа 18.1, мы видим, что $\frac{Q}{\sigma^2}$, $\frac{Q_1}{\sigma^2}$ и $\frac{Q_2}{\sigma^2}$ имеют распределение χ^2 , соответственно, с $n-1$, $r-1$ и $n-r$ степенями свободы. Полагая

$$s^2 = \frac{1}{n-1} Q = \frac{1}{n-1} \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2,$$

$$s_1^2 = \frac{1}{r-1} Q_1 = \frac{1}{r-1} \sum_{i=1}^r n_i (\bar{x}_i - \bar{x})^2,$$

$$s_2^2 = \frac{1}{n-r} Q_2 = \frac{1}{n-r} \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2,$$

мы таким образом, имеем

$$E(s^2) = E(s_1^2) = E(s_2^2) = \sigma^2.$$

Дисперсионное отношение $e^{2s^2} = \frac{s_1^2}{s_2^2}$ можно записать в виде

$$e^{2s^2} = \frac{s_1^2}{s_2^2} = \frac{\frac{1}{r-1} Q_1}{\frac{1}{n-r} Q_2} = \frac{\frac{1}{r-1} \sum_{i=1}^{r-1} \eta_i^2}{\frac{1}{n-r} \sum_{i=r}^{n-1} \eta_i^2}$$

Так как η_i независимы и нормальны $(0, \sigma)$, величина z имеет распределение, полученное Р. Фишером; оно определяется плотностью вероятности (18.3.5), где вместо m и n следует взять соответственно $r-1$ и $n-r$. В частности, среднее значение и стандартное отклонение величины e^{2z} даются формулой (18.3.4). Имеются таблицы пределов значимости для e^{2z} и z при различных значениях уровня значимости $\epsilon = \frac{p}{100}$ (Фишер [13], Фишер — Иэйтс [262], Снедекор [35], Боннье-Теден [8]). „Критерий z “, введенный Фишером, заключается в том, что нулевая гипотеза отвергается с точки зрения p -процентного уровня значимости, когда $|z| > z_p$, где z_p определяется так, что $P(|z| > z_p) = \epsilon = \frac{p}{100}$.

Нулевая гипотеза, очевидно, является сложной гипотезой (см. параграф 35.2), связанной с параметрами m_1, \dots, m_r и σ , именно гипотезой, что все m_i равны одному и тому же значению m , о котором никаких предположений не делается. Каковы бы ни были значения m и σ , вероятность отвергнуть нулевую гипотезу в случае, когда эта гипотеза в действительности верна, равняется $P(|z| > z_p) = \epsilon$. Таким образом критическое множество, соответствующее критерию z , подобно пространству выборок, и критерий z есть критерий уровня ϵ в соответствии с определением, приведенным в параграфе 35.5.

Вычисления обычно располагают в такую таблицу:

Изменчивость	Число степеней свободы	Сумма квадратов	Среднее квадратов
Межгруппы	$r-1$	$Q_1 = \sum_{i=1}^r n_i (x_i - \bar{x})^2$	$s_1^2 = \frac{Q_1}{r-1}$
Интрегруппы	$n-r$	$Q_2 = \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$	$s_2^2 = \frac{Q_2}{n-r}$
Сумма	$n-1$	$Q = \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$	$s^2 = \frac{Q}{n-1}$

Каждое из трех чисел в столбце „средних квадратов“ при нулевой гипотезе дает несмешенную оценку для дисперсии совокупности σ^2 ,

и критерий z можно считать критерием совместности независимых оценок s_1^2 и s_2^2 .

Теперь перейдем к рассмотрению случая, когда нулевая гипотеза неверна, т. е. когда групповые средние значения не равны друг другу. Полагая $x_{ij} = m_i + \xi_{ij}$, где ξ_{ij} — независимые и нормальные $(0, \sigma)$ величины, получим

$$\begin{aligned} (\bar{x}_{ij} - \bar{x})^2 &= (\bar{\xi}_{ij} - \bar{\xi})^2 + 2(m_i - \bar{m})(\bar{\xi}_{ij} - \bar{\xi}) + (m_i - \bar{m})^2 \\ (x_{ij} - \bar{x}_i)^2 &= (\xi_{ij} - \bar{\xi}_{ij})^2, \end{aligned}$$

где $\bar{\xi}_{ij}$ и $\bar{\xi}$ определяются так же, как и выше, а $\bar{m} = \frac{1}{n} \sum_1^n n_i m_i$

Подставляя эти выражения в Q_1 и Q_2 , обнаружим, во-первых, что Q_2 имеет то же самое распределение, как и в случае, когда нулевая гипотеза верна. Далее, получаем (см. Эрвин [133])

$$E(s_1^2) = E\left(\frac{1}{r-1} Q_1\right) = \sigma^2 + \frac{1}{r-1} \sum_1^r n_i (m_i - \bar{m})^2,$$

$$E(s_2^2) = E\left(\frac{1}{n-r} Q_2\right) = \sigma^2,$$

или

$$E\left(\frac{r-1}{n}(s_1^2 - s_2^2)\right) = \frac{1}{n} \sum_1^r n_i (m_i - \bar{m})^2.$$

Правую часть последнего равенства можно считать мерой изменчивости неизвестных групповых средних значений m_i . Таким образом, несмещенную оценку для этой меры дает число $\frac{(r-1)(s_1^2 - s_2^2)}{n}$, которое можно вычислить по исходным данным.

Наконец, для любого заданного $i \neq j$ величина $\bar{x}_{ij} - \bar{x}_j$ нормальна

$$\left(m_i - m_j, \sigma \sqrt{\frac{n_i + n_j}{n_i n_j}}\right).$$

$$\bar{x}_{ij} - \bar{x}_j = (\bar{x}_i - \bar{x}) - (\bar{x}_j - \bar{x})$$

и замечая, что указанное выше ортогональное преобразование, заменяющее x_{ij} на y_{ij} , превращает каждую разность $\bar{x}_{ij} - \bar{x}$ в линейную комбинацию величин y_1, \dots, y_r , убеждаемся, что $\bar{x}_{ij} - \bar{x}_j$ независимы от Q_2 . Отсюда следует, что величина

$$t = \sqrt{\frac{n_i n_j}{n_i + n_j}} \cdot \frac{\bar{x}_{ij} - \bar{x}_j - (m_i - m_j)}{s_2}$$

имеет распределение Стьюдента с $n - r$ степенями свободы. Работая с p -процентным уровнем значимости, мы получаем, таким образом (см. параграф 34.4), доверительные пределы для разности между двумя неизвестными групповыми средними значениями:

$$(36.2.2) \quad \bar{x}_{i\cdot} - \bar{x}_{j\cdot} \pm t_p s_3 \sqrt{\frac{n_i + n_j}{n_i + n_j}}.$$

В частном случае, когда имеется лишь две группы ($r = 2$), эти пределы совпадают с доверительными пределами, указанными в формуле (34.4.5) (следует учесть различие в обозначениях!). Если $r > 2$, то мы, конечно, можем применять формулу (34.4.5) для получения доверительных пределов разности $m_i - m_j$, основываясь исключительно на наблюденных значениях i -й и j -й групп. Однако t_p будет тогда иметь лишь $n_i + n_j - 2$ степеней свободы, так что (36.2.2) со своими $n - r$ степенями свободы будет при том же самом значении p давать, вообще говоря, меньшее значение t_p , т. е. более короткий доверительный интервал.

Если нулевая гипотеза верна, то функция мощности (см. параграфы 35.3 и 35.5) критерия z будет иметь значение ϵ . Поведение функции мощности, когда нулевая гипотеза неверна, исследовал Тэнг [224], который опубликовал таблицы численных значений этой функции. Эти таблицы можно применять также и в более общих случаях, которые будут рассмотрены в следующем параграфе.

Наши x_{ij} суть n случайных величин, совместное распределение которых содержит $r + 1$ неизвестных параметров m_1, \dots, m_r и σ^2 . Совместной плотностью вероятности является

$$f = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{ij} - m_i)^2}$$

Проблема оценки параметров по выборке, состоящей из одного наблюденного значения каждого x_{ij} , является примером обобщенной проблемы нахождения оценок, рассмотренной в параграфе 32.8. Соотношения $E(\bar{x}_{i\cdot}) = m_i$ и $E(s_2^2) = \sigma^2$ показывают, что числа $\bar{x}_{1\cdot}, \dots, \bar{x}_r$ и s_2^2 являются *несмещенными* оценками для параметров. С помощью соотношения (32.4.1),енным образом обобщенного, как это указано в 32.6 – 32.8, мы обнаруживаем, что эти числа являются также *совместно-достаточными* оценками. Далее, с помощью некоторых вычислений можно убедиться, что *совместная эффективность* этих оценок равна $\frac{n-r}{n}$.

36.3. Обобщение^{*)}. Предыдущие рассмотрения можно обобщить на случаи, когда наблюденные величины расположены в более сложную систему групп и подгрупп различных порядков. Вообще говоря, величины при этом снабжаются двумя или большим числом индексов, но для целей настоящего параграфа достаточно сохранить для величин x_1, \dots, x_n обозначение параграфа 36.1. Как и прежде, мы предполагаем, что величины x_i независимы и нормальны (m_i, σ), причем σ не зависит от i .

При каждой системе группировки, используемой в данной частной задаче, можно рассмотреть суммы квадратов отклонений, более или менее подобные суммам Q_1 и Q_2 предыдущего параграфа, и таким способом часто удается получить соотношение такого же вида, как и (36.2.1):

$$(36.3.1) \quad \sum_1^n (x_i - \bar{x})^2 = Q_1 + Q_2 + \dots + Q_k.$$

где Q_k — суммы квадратов некоторых линейных форм от x_i , причем выполняется соответствующее соотношение между рангами r , квадратичных форм Q :

$$n - 1 = r_1 + r_2 + \dots + r_k.$$

Как и в предыдущем параграфе, мы видим (см. параграф 11.11), что существует ортогональное преобразование, превращающее Q_1, \dots, Q_k в суммы соответственно r_1, \dots, r_k квадратов y_i^2 , причем никакие две Q_i не содержат общей переменной y_i . Из независимости величин y_i вытекает, что и Q_1, \dots, Q_k независимы.

Предположим теперь, что требуется проверить гипотезу H , согласно которой неизвестные средние значения m_i удовлетворяют некоторым линейным уравнениям. Разложение полной суммы квадратов (36.3.1) часто можно устроить таким образом, что если гипотеза H верна, то две из форм Q_i , скажем, Q_1 и Q_2 , обращаются в нуль при замене всех x_i соответствующими m_i . Так, например, в случае, рассмотренном в предыдущем параграфе, проверяется гипотеза $m_1 = m_2 = \dots = m_r$, где m_i — групповые средние, и если гипотеза H

^{*)} Я воспользовался здесь неопубликованной рукописью Г. Андерсона, любезно предоставленной в мое распоряжение автором. Изложение этой теории с аналогичных, но более общих точек зрения, читатель найдет в работах Колоджейчика [140a] и Тэнга [224].

верна; то, как легко видеть, формы Q_1 и Q_2 , определяемые формулой (36.2.1), обе обращаются в нуль, если все переменные заменить их средними значениями.

Предполагая, что гипотеза H верна и что разложение (36.3.1) устроено описанным выше образом, заменим в Q_1 и Q_2 x_i на $m_i + \xi_i$. Если неотрицательная квадратичная форма $q(x_1, \dots, x_n)$ обращается в нуль в точке (m_1, \dots, m_n) , то легко видеть, что в этой точке равны нулю также все производные $\frac{\partial q}{\partial x_i}$. Таким образом, мы получаем тождество $Q_v(x_1, \dots, x_n) = Q_v(\xi_1, \dots, \xi_n)$ при $v=1, 2$. Указанное выше ортогональное преобразование превращает Q_1 в сумму r_1 квадратов $\sum_1^{r_1} \eta_i^2$, где η_i — независимые и нормальные $(0, \sigma)$ величины; аналогично преобразуется форма Q_2 . Таким образом, при гипотезе H величины Q_1 и Q_2 независимы и имеют распределения χ^2 соответственно с r_1 и r_2 степенями свободы. Если ввести средние квадратов

$$s_1^2 = \frac{Q_1}{r_1}, \quad s_2^2 = \frac{Q_2}{r_2}$$

и дисперсионное отношение

$$e^{2z} = \frac{s_1^2}{s_2^2},$$

то получим $E(s_1^2) = E(s_2^2) = \sigma^2$, и z будет иметь распределение Фишера, приведенное в параграфе 18.3. Таким образом, для проверки гипотезы H можно использовать критерий z .

Если гипотеза H неверна, то можно вывести такие же результаты, как и в предыдущем параграфе.

36.4. Случайные блоки. Теперь покажем применение изложенной теории к некоторым случаям, имеющим большое практическое значение. Будем пользоваться терминологией, заимствованной из сельскохозяйственных приложений, но такие же экспериментальные приемы могут быть использованы во многих областях исследовательской работы, например, в биологии и др.

Рассмотрим эксперимент, состоящий в том, что сравнивается эффективность r различных способов обработки земли, засеваемой некоторой сельскохозяйственной культурой. На имеющейся площади мы

выделяем s участков („блоков“) одинаковой площади; каждый блок разбиваем на r равных делянок, каждую из которых обрабатываем одним из r различных способов, распределив их случайным образом. Тогда каждый блок будет содержать по одной делянке, обработанной каждым способом, и каждым способом будет обработано s различных делянок.

Пусть x_{ij} обозначает вес урожая, собранного с делянки, обработанной i -м способом и принадлежащей к j -му блоку. Предположим, что x_{ij} независимы и нормальны (m_{ij} , σ) и положим

$$\bar{x}_{i \cdot} = \frac{1}{s} \sum_{j=1}^s x_{ij}, \quad \bar{x}_{\cdot j} = \frac{1}{r} \sum_{i=1}^r x_{ij}, \quad \bar{x} = \frac{1}{rs} \sum_{i=1}^r \sum_{j=1}^s x_{ij}.$$

Таким образом, $\bar{x}_{i \cdot}$ и $\bar{x}_{\cdot j}$ являются выборочными средними значениями соответственно для i -го способа обработки и для j -го блока, а \bar{x} есть общее среднее значение. Тогда легко проверить тождество

$$(36.4.1) \quad \sum_{i=1}^r \sum_{j=1}^s (x_{ij} - \bar{x})^2 = s \sum_{i=1}^r (\bar{x}_{i \cdot} - \bar{x})^2 + r \sum_{j=1}^s (\bar{x}_{\cdot j} - \bar{x})^2 + \\ + \sum_{i=1}^r \sum_{j=1}^s (x_{ij} - \bar{x}_{i \cdot} - \bar{x}_{\cdot j} + \bar{x})^2 = Q_1 + Q_2 + Q_3$$

и соответствующее соотношение между рангами

$$rs - 1 = (r - 1) + (s - 1) + (r - 1)(s - 1).$$

Отсюда, так же как и в предыдущем параграфе, заключаем, что Q_1 , Q_2 и Q_3 независимы.

Q_1 и Q_2 называются суммами квадратов отклонений соответственно „между способами обработки“ и „между блоками“, а Q_3 обычно называется, по причинам, которые будут изложены ниже, „суммой квадратов, созданной ошибкой“. Численные значения можно расположить в таблицу так же, как это было показано в параграфе 36.2.

Различие средних значений m_{ij} может быть вызвано неоднородностью почвы и различием в способах обработки. Благодаря случайному распределению способов обработки в каждом блоке можно предполагать, что влияние неоднородности почвы *внутри каждого блока* оказывается лишь в случайных частях величин x_{ij} . Таким образом, любая разность между двумя m_{ij} , относящимися к одному и тому же блоку, вызывается различием в способах обработки, и мы

предполагаем, что $m_{ij} = f_i + b_j$, где f_i зависит только от способа обработки, а b_j — только от участка. Будем коротко называть f_i „эффектом обработки“ и b_j — „эффектом блока“. При этих предположениях мы увидим, что Q_8 обратится в нуль, если все x_{ij} заменить их средними значениями, так что $\frac{Q_8}{\sigma^2}$ имеет распределение χ^2 с $(r-1)(s-1)$ степенями свободы. Следовательно, среднее квадратов $s_3^2 = \frac{Q_8}{(r-1)(s-1)}$ является несмешенной оценкой для σ^2 , что и оправдывает введенную выше терминологию.

Проверим гипотезу H о том, что между способами обработки делянок нет различия. Если гипотеза H верна, то можно при всех i положить $f_i = 0$, так что $m_{ij} = b_j$ будет зависеть лишь от номера участка j . В этом случае обе формы Q_1 и Q_8 обращаются в нуль, если все x_{ij} заменить их средними значениями. Вводя среднее квадратов $s_1^2 = \frac{Q_1}{r-1}$, можно, согласно предыдущему параграфу, проверить гипотезу H применением критерия z к дисперсионному отношению $e^{2z} = \frac{s_1^2}{s_3^2}$.

Если гипотеза H неверна, то можно показать, как и в параграфе 36.2, что число $\frac{r-1}{r_s}(s_1^2 - s_3^2)$ является несмешенной оценкой для дисперсии $\frac{1}{r} \sum (f_i - \bar{f})^2$ среди неизвестных эффектов обработки. Далее, для каждого $i \neq j$ мы получаем доверительные пределы для неизвестной разности $f_i - f_j$ между эффектами i -го и j -го способов обработки:

$$\bar{x}_{i \cdot} - \bar{x}_{j \cdot} \pm t_p s_8 \sqrt{\frac{2}{s}}.$$

Здесь следует взять t_p с $(r-1)(s-1)$ степенями свободы.

Если нам пришлось отвергнуть гипотезу H , может потребоваться проверить другую гипотезу H_1 , согласно которой различие между способами обработки создается исключительно за счет одного способа, скажем, способа с номером $i = 1$, в то время как между остальными способами различий нет. Если гипотеза H_1 верна, то можно положить $f_2 = \dots = f_r = 0$, в то время как f_1 может отличаться от нуля.

Пусть $\bar{x}_{(2 \dots r)}$ обозначает общее выборочное среднее значение для способов обработки $2, \dots, r$:

$$\bar{x}_{(2 \dots r)} = \frac{1}{(r-1)s} \sum_{i=2}^r \sum_{j=1}^s x_{ij}.$$

Сумма квадратов „между способами обработки“ Q_1 , появившаяся в формуле (36.4.1), может быть разбита на части в соответствии с тождеством

$$Q_1 = \frac{(r-1)s}{r} (\bar{x}_1 - \bar{x}_{(2 \dots r)})^2 + s \sum_{i=2}^r (\bar{x}_i - \bar{x}_{(2 \dots r)})^2 = Q'_1 + Q''_1,$$

которое дает соотношение между рангами

$$r-1 = 1 + (r-2).$$

Q'_1 и Q''_1 можно считать суммами квадратов соответственно „между группой 1 и объединением групп $2, \dots, r$ “ и „между группами $2, \dots, r$ “. Введя это выражение в (36.4.1), мы обнаружим, что если гипотеза H_1 верна, то обе формы Q'_1 и Q''_1 обращаются в нуль, когда все величины x_{ij} заменяются их средними значениями. Вводя средний квадрат $s''_1 = \frac{Q''_1}{r-2}$, мы, таким образом, сможем проверить гипотезу H_1 , применяя критерий z к дисперсионному отношению $e^{2z} = \frac{s''_1}{s''_3}$.

Для неизвестного эффекта обработки f_1 получаем доверительные пределы

$$\bar{x}_1 - \bar{x}_{(2 \dots r)} \pm t_p s_3 \sqrt{\frac{r}{(r-1)s}},$$

где, как и прежде, t_p имеет $(r-1)(s-1)$ степеней свободы.

Другие гипотезы подобного рода, связанные со свойствами различных эффектов обработки, могут быть проверены аналогичными методами. Требуемые неравенства находятся, как правило, без труда.

36.5. Латинские квадраты. С помощью метода случайных блоков мы пытаемся исключить влияние неоднородности почвы, чтобы осуществить несмещеннное сравнение эффективности способов обработки

(или влияния вида животных и т. п., в зависимости от случая), которые применялись в нашем эксперименте. Более полное исключение влияния неоднородности обычно достигается с помощью метода латинских квадратов.

Рассмотрим r^2 делянок, занумерованных посредством двух индексов в квадратную схему. Допустим, что к этим делянкам были применены r различных способов обработки так, что каждый способ был применен к одной делянке каждого столбца и каждой строки. Предположим, что для нашего эксперимента случайно выбрано одно такое расположение способов обработки из многочисленных возможных расположений, удовлетворяющих поставленному условию и известных под названием латинских квадратов *). Обозначим через x_{ij} вес урожая, собранного с делянки в i -й строке и j -м столбце, и пусть $\bar{x}_{i\cdot}$ и $\bar{x}_{\cdot j}$ — средние значения строк и столбцов, \bar{x}_h — среднее значение для делянок, обработанных h -м способом, и \bar{x} — общее среднее значение. В этом случае имеем тождество

$$\begin{aligned} \sum_i \sum_j (x_{ij} - \bar{x})^2 &= r \sum_h (\bar{x}_h - \bar{x})^2 + r \sum_i (\bar{x}_{i\cdot} - \bar{x})^2 + \\ &+ r \sum_j (\bar{x}_{\cdot j} - \bar{x})^2 + \sum_i \sum_j (x_{ij} - \bar{x}_h - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + 2\bar{x})^2 = \\ &= Q_1 + Q_2 + Q_3 + Q_4, \end{aligned}$$

где все суммирования производились по значениям индексов от 1 до r , а в каждом слагаемом из Q_4 индекс h должен соответствовать способу обработки, применявшемуся на делянке (i, j) . Соотношение между рангами имеет здесь следующий вид:

$$r^2 - 1 = (r - 1) + (r - 1) + (r - 1) + (r - 1)(r - 2).$$

Предположим теперь, что среднее значение $E(x_{ij}) = m_{ij}$ состоит из „эффекта обработки“ f_h и из другой части, вызванной неоднородностью почвы и состоящей из „эффекта строки“ r_i и „эффекта столбца“ c_j . Тогда $m_{ij} = f_h + r_i + c_j$, и, как и прежде, мы обнаруживаем, что Q_4 имеет распределение χ^2 с $(r - 1)(r - 2)$ степенями свободы, так что средний квадрат $s_4^2 = \frac{Q_4}{(r - 1)(r - 2)}$ является несмешанной оценкой для общей дисперсии σ^2 величин x_{ij} . Табличное расположение данных имеет здесь следующий вид:

* Таблицы таких расположений приведены в книге Фишера и Иэйтса [262].

Изменчивость	Число степеней свободы	Сумма квадратов	Среднее квадратов
Между способами обработки	$r - 1$	Q_1	$s_1^2 = \frac{Q_1}{r - 1}$
Межлу строками	$r - 1$	Q_2	$s_2^2 = \frac{Q_2}{r - 1}$
Межлу столбцами	$r - 1$	Q_3	$s_3^2 = \frac{Q_3}{r - 1}$
Ошибка	$(r - 1)(r - 2)$	Q_4	$s_4^2 = \frac{Q_4}{(r - 1)(r - 2)}$
Сумма	$r^2 - 1$	Q	—

Гипотезу о том, что не существует различий между способами обработки, можно проверить, применяя критерий χ^2 к дисперсионному отношению $e^{2\chi^2} = \frac{s_1^2}{s_2^2}$. В случае, когда эта гипотеза отвергнута, можно оценить дисперсию среди эффектов обработки и разность между двумя эффектами обработки такими же методами, как и в предыдущем параграфе. Таким же образом можно проверять и другие гипотезы, связанные со свойствами величин f_h .

Мы рассмотрели здесь лишь простейшие случаи дисперсионного анализа. За другими сведениями по теории схем экспериментов и по вопросу об обобщении на случай одновременного анализа нескольких величин („анализ смешанных вторых моментов“) мы отсылаем читателя к книгам Фишера [13], [14], Снедекора [35], Бонье и Тедена [8].

ГЛАВА 37

НЕКОТОРЫЕ ПРОБЛЕМЫ РЕГРЕССИИ

37.1. Проблемы, содержащие неслучайные величины. В практических приложениях очень часто встречаются проблемы, в которых изучается случайная величина y , зависящая от некоторого количества *нестационарных* переменных x_1, \dots, x_n . В экономической и социальной статистике значения x_i появляются, как правило, просто как заданные в наших статистических данных неслучайные числа. С другой стороны, в экспериментальных работах значения x_i часто могут быть произвольно выбраны экспериментатором. В обоих случаях x_i играют роль переменных

параметров, входящих в распределение величины y , и наши статистические данные состоят из множества наблюденных значений y , каждое из которых соответствует известным значениям переменных x_i . Кроме *известных* параметров x_i , распределение величины y может, конечно, содержать также некоторые *неизвестные* параметры.

Предположим, например, что мы исследуем количество u товара A , потребляемого на данном рынке, и цены x_1, \dots, x_n самого товара A и некоторого количества других товаров. Может казаться законным считать u случайной величиной с распределением, определяемым ценами x_1, \dots, x_n , причем явления, обусловливающие значения цен, могут и не походить на случайный эксперимент. Тогда x_i следует считать просто заданными числами в наших данных. С другой стороны, предположим, что исследуется влияние качества сырья и применяемого технологического процесса, характеризующихся величинами x_1, \dots, x_n , на производительность u некоторой фабрики. Тогда можно выбирать подходящим образом системы значений величин x_i и наблюдать соответствующие значения u . Как и прежде, u будет здесь считаться случайной величиной, распределение которой содержит параметры x_i .

Теория средней квадратической регрессии, развитая в главах 21 и 23, применима, с некоторыми видоизменениями, и к настоящему случаю. Для случая же, когда зависимая величина u распределена нормально и ее среднее значение является линейной функцией от величин x_i , Фишером [92], [97] и Бартлеттом [54] было показано, что некоторые коэффициенты регрессии имеют выборочные распределения, сходные с распределениями, выведенными в параграфах 29.8 и 29.12; на этих распределениях можно основывать критерии значимости таким же образом, как это было показано в параграфе 31.3 (примеры 6—7). Некоторые результаты Фишера и Бартлетта мы изложим в ближайших двух параграфах.

37.2. Простая регрессия. Задана выборка, состоящая из n наблюденных пар значений $(x_1, y_1), \dots, (x_n, y_n)$. Для выборочных моментов будем употреблять обычные обозначения \bar{x}, \bar{y}, m_{20} и т. д., введенные равенствами (27.1.6) и (27.1.7). Однако теперь мы предположим, что x — неслучайная величина и что при каждом фиксированном x случайная величина y имеет нормальное распределение со средним значением $a + \beta(x - \bar{x})$ и стандартным отклонением σ , где a, β и σ — неизвестные параметры, не содержащие x . Таким образом, выборочные моменты,

содержащие только x_i , как, например, \bar{x} , $m_{20} = s_1^2$ и т. д., следует считать не случайными величинами, а просто заданными постоянными. С другой стороны, все величины, зависящие от y_i , как, например, \bar{y} , $m_{02} = s_2^2$, $m_{11} = rs_1s_2$ и т. д., суть случайные величины. Величины y_i мы будем предполагать независимыми.

Оценки максимального правдоподобия (см. параграф 33.2) для α , β и σ находятся с помощью минимизирования совместной плотности вероятности величин y_i , которая равняется

$$f = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - \alpha - \beta(x_i - \bar{x})]^2}.$$

Можно показать, что оценками для α и β являются значения этих параметров, при которых сумма квадратов в показателе достигает своего наименьшего возможного значения. Отсюда получаем оценки

$$\alpha^* = \bar{y}, \quad \beta^* = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{m_{11}}{s_1^2},$$

а оценка максимального правдоподобия для σ определяется равенством

$$\sigma^{*2} = \frac{1}{n} \sum_{i=1}^n [y_i - \alpha^* - \beta^*(x_i - \bar{x})]^2 = s_2^2(1 - r^2).$$

Как линейные функции от y_i величины α^* и β^* обе распределены нормально, и мы получаем

$$E(\alpha^*) = \alpha, \quad D^2(\alpha^*) = \frac{\sigma^2}{n},$$

$$E(\beta^*) = \beta, \quad D^2(\beta^*) = \frac{\sigma^2}{s_1^2 n}.$$

Далее, имеем тождество

$$(37.2.1) \quad \begin{aligned} & \sum_{i=1}^n [y_i - \alpha^* - \beta^*(x_i - \bar{x})]^2 = \\ & = \sum_{i=1}^n [y_i - \alpha - \beta(x_i - \bar{x})]^2 - n(\alpha^* - \alpha)^2 - s_1^2 n (\beta^* - \beta)^2. \end{aligned}$$

Полагая

$$\eta_i = y_i - \alpha - \beta(x_i - \bar{x}),$$

убеждаемся, что величины η_i независимы и нормальны $(0, \sigma)$ и что линейные формы

$$\zeta_1 = \sqrt{n}(\alpha^* - \alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i,$$

$$\zeta_2 = s_1 \sqrt{n}(\beta^* - \beta) = \frac{1}{s_1 \sqrt{n}} \sum_{i=1}^n (x_i - \bar{x}) \eta_i$$

удовлетворяют условиям ортогональности (11.9.1). Записывая тождество (37.2.1) в виде

$$n\sigma^{*2} = \sum_1^n \eta_i^2 - \zeta_1^2 - \zeta_2^2,$$

мы сможем применить лемму Фишера (см. параграф 29.2) и убедимся, что α^*, β^* и σ^* независимы и что $\frac{n\sigma^{*2}}{\sigma^2}$ распределена подобно χ^2 с $n-2$ степенями свободы. Следовательно, согласно параграфу 18.2, величины

$$\sqrt{n-2} \frac{\alpha^* - \alpha}{\sigma^*} \quad \text{и} \quad s_1 \sqrt{n-2} \frac{\beta^* - \beta}{\sigma^*}$$

имеют распределение Стьюдента с $n-2$ степенями свободы. Относительно коэффициента регрессии β^* этот результат формально совпадает с уже полученным (см. (29.8.4)) результатом для случая, когда обе величины x и y суть случайные величины с нормальным совместным распределением.

Поскольку s_1, α^*, β^* и σ^* известны, можно использовать полученный результат для проверки любых гипотетических значений α и β и для вывода доверительных пределов для этих параметров таким же образом, как это было показано в случае среднего значения обычного нормального распределения в параграфах 31.2 и 34.4. В частности, мы обнаруживаем, что коэффициент регрессии β значимо отличается от нуля с точки зрения p -процентного уровня значимости, если $|\beta^*| > \frac{t_p \sigma^*}{s_1 \sqrt{n-2}}$,

где t_p следует взять с $n-2$ степенями свободы.

Наконец, нам может понадобиться оценка для неизвестной ординаты линии регрессии

$$Y = \alpha + \beta(X - \bar{x})$$

в какой-нибудь заданной точке X . Можно показать, что величина

$$t = \frac{\sqrt{n-2}}{\sqrt{1 + \left(\frac{X-\bar{x}}{s_1}\right)^2}} \cdot \frac{\alpha^* + \beta^*(X - \bar{x}) - Y}{\sigma^*}$$

имеет распределение Стьюдента с $n-2$ степенями свободы, так что p -процентными доверительными пределами для Y служат

$$(37.2.2) \quad \alpha^* + \beta^*(X - \bar{x}) \pm t_p \frac{\sigma^*}{\sqrt{n-2}} \sqrt{1 + \left(\frac{X-\bar{x}}{s_1}\right)^2}.$$

37.3. Множественная регрессия. Теперь перейдем к случайной величине y , среднее значение которой есть линейная функция от k неслучайных величин x_1, \dots, x_k . Предположим, что задана выборка в n независимо наблюденных точек $(y_v, x_{v1}, \dots, x_{vk})$, где $v = 1, 2, \dots, n$. Для выборочных моментов будем употреблять обозначения, введенные в параграфах 27.1 и 29.9, полагая, например, в соответствии с (29.9.2),

$$l_{ij} = \frac{1}{n} \sum_{v=1}^n (x_{vi} - \bar{x}_i)(x_{vj} - \bar{x}_j), \quad (i, j = 1, 2, \dots, k),$$

и, далее, рассматривая y как величину x_n ,

$$l_{0j} = \frac{1}{n} \sum_{v=1}^n (y_v - \bar{y})(x_{vj} - \bar{x}_j).$$

Через L и L_{ij} обозначим соответственно детерминант

$$L = \begin{vmatrix} l_{11} & \dots & l_{1k} \\ \dots & \dots & \dots \\ l_{k1} & \dots & l_{kk} \end{vmatrix}$$

и его алгебраические дополнения. Будем предполагать, что $L \neq 0$.

Предположим теперь, что при любых фиксированных значениях x_1, \dots, x_k случайная величина y имеет нормальное распределение со средним значением

$$(37.3.1) \quad E(y) = \alpha + \beta_1(x_1 - \bar{x}_1) + \dots + \beta_k(x_k - \bar{x}_k)$$

и со стандартным отклонением σ . Можно показать, что оценками максимального правдоподобия α^* и β_i^* ($i = 1, 2, \dots, k$) являются значения α и β_i , при которых сумма

$$\sum_{v=1}^n [y_v - \alpha - \beta_1(x_{v1} - \bar{x}_1) - \dots - \beta_k(x_{vk} - \bar{x}_k)]^2$$

достигает своего наименьшего возможного значения. Отсюда получаем оценки

$$(37.3.2) \quad \alpha^* = \bar{y}, \quad \beta_i^* = \frac{1}{L} \sum_{j=1}^k l_{0j} L_{ij}, \quad (i = 1, \dots, k),$$

а оценкой максимального правдоподобия для σ^2 является

$$(37.3.3) \quad \sigma^{*2} = \frac{1}{n} \sum_{v=1}^n [y_v - a^* - \beta_1^*(x_1, -\bar{x}_1) - \dots - \beta_k^*(x_k, -\bar{x}_k)]^2 = s_{0.12\dots k}^2,$$

где $s_{0.12\dots k}^2$ — выборочное значение остаточной дисперсии (см. параграфы 23.3 и 29.12) величины y относительно x_1, \dots, x_k . Будем предполагать, что эта величина положительна, т. е. что наблюденные значения y не могут быть точно представлены линейной функцией от x_i .

Как линейные функции от y , величины a^* и β_i^* имеют нормальные распределения. С помощью некоторых вычислений мы получим следующие средние значения и центральные моменты второго порядка:

$$(37.3.4) \quad \begin{aligned} E(a^*) &= a, \quad E(\beta_i^*) = \beta_i, \\ E(a^* - a)^2 &= \frac{\sigma^2}{n}, \quad E[(a^* - a)(\beta_i^* - \beta_i)] = 0, \\ E[(\beta_i^* - \beta_i)(\beta_j^* - \beta_j)] &= \frac{\sigma^2}{n} \frac{L_{ij}}{L}, \end{aligned}$$

где $i, j = 1, \dots, k$. Отсюда получим, в частности, в силу формулы (23.3.4),

$$D^2(\beta_1^*) = E(\beta_1^* - \beta_1)^2 = \frac{\sigma^2}{n} \frac{L_{11}}{L} = \frac{\sigma^2}{ns_{1.23\dots k}^2}$$

и аналогичные выражения для $D^2(\beta_i^*)$, $i = 2, \dots, k$.

Далее, можно показать, что величина σ^* независима от величин a^* и β_i^* и что $\frac{n\sigma^{*2}}{\sigma^2}$ имеет распределение χ^2 с $n - k - 1$ степенями свободы.

В частном случае, когда матрица $L = \begin{Bmatrix} l_{11} & \dots & l_{1k} \\ \vdots & \ddots & \vdots \\ l_{k1} & \dots & l_{kk} \end{Bmatrix}$ — диагональная, это можно доказать, непосредственно обобщая метод, использованный в предыдущем параграфе. Действительно, выражения $\sqrt{n}(a^* - a)$ и $s_i \sqrt{n}(\beta_i^* - \beta_i)$ суть линейные формы от переменных $\eta_v = y_v - a - \beta_1(x_1, -\bar{x}_1) - \dots - \beta_k(x_k, -\bar{x}_k)$, и если L есть диагональная матрица, то эти формы при $i = 1, \dots, k$ удовлетворяют условиям ортогональности, так что лемму Фишера можно применять так же, как и выше. В общем случае сначала следует заменить величины

x_i новыми величинами x'_i с помощью ортогонального преобразования, при котором матрица вторых моментов для новых величин будет диагональной (см. параграф 22.6). Применяя к величинам β_i контрагредиентное преобразование (см. параграф 11.7), мы сможем закончить доказательство так же, как и в рассмотренном частном случае.

Отсюда следует, что величины

$$\sqrt{n-k-1} \frac{\alpha^* - \alpha}{\sigma^*} = \sqrt{n-k-1} \frac{\alpha^* - \alpha}{s_{0.12\dots k}},$$

$$s_{1.23\dots k} \sqrt{n-k-1} \frac{\beta_1^* - \beta_1}{\sigma^*} = \sqrt{n-k-1} \frac{s_{1.23\dots k}}{s_{0.12\dots k}} (\beta_1^* - \beta_1)$$

и аналогичные величины с $\beta_2^*, \dots, \beta_k^*$ все имеют распределение Стьюдента с $n-k-1$ степенями свободы. Относительно β_i^* этот результат непосредственно соответствует результату (29.12.1). Как и прежде, мы можем теперь вывести критерии значимости и доверительные пределы для неизвестных параметров α и β_i .

Мы можем получить также совместный критерий для множества гипотетических значений коэффициентов регрессии β_1, \dots, β_k . Из выражений (37.3.4) для моментов нормально распределенных величин β_i^* следует (см. упражнение 15, стр. 351), что величина

$$\frac{n}{\sigma^2} \sum_{i,j=1}^k I_{ij} (\beta_i^* - \beta_i) (\beta_j^* - \beta_j)$$

имеет распределение χ^2 с k степенями свободы. Следовательно, величина

$$e^{2z} = \frac{n-k-1}{k\sigma^2} \sum_{i,j=1}^n I_{ij} (\beta_i^* - \beta_i) (\beta_j^* - \beta_j)$$

распределена так же, как дисперсионное отношение (см. параграфы 18.3 и 36.2) с k степенями свободы в числителе и $n-k-1$ степенями свободы в знаменателе. Если задано множество гипотетических значений β_1, \dots, β_k , то число e^{2z} можно вычислить по нашим данным, так что для проверки предложенных значений β_i можно использовать таблицы распределения z .

Предположим, в частности, что требуется проверить гипотезу о равенстве нулю всех коэффициентов регрессии: $\beta_1 = \beta_2 = \dots = \beta_k = 0$.

Из (37.3.2) и (37.3.3), после некоторых вычислений и используя (23.5.2) и (23.5.3), получим

$$\sum_{i,j=1}^k l_{ij} \beta_i^* \beta_j^* = l_{00} r_{0(12\dots k)}^2,$$

$$\sigma^{*2} = s_{0(12\dots k)}^2 = l_{00} (1 - r_{0(12\dots k)}^2),$$

где $r_{0(12\dots k)}$ — сводный коэффициент корреляции между выборочными значениями величин y и (x_1, \dots, x_k) . Тогда из параграфов 18.3 и 18.4 следует, что $r_{0(12\dots k)}^2$ имеет бета-распределение с плотностью вероятности $\beta\left(x; \frac{k}{2}, \frac{n-k-1}{2}\right)$. Таким образом, указанный выше критерий в этом частном случае совпадает с основанным на распределении (29.13.8) критерием для гипотезы о том, что сводный коэффициент корреляции значимо отличается от нуля.

Для ординаты $a + \sum_1^k \beta_i (X_i - \bar{x}_i)$ линии регрессии в какой-нибудь данной точке (X_1, \dots, X_k) мы получаем в качестве неопределенного обобщения формулы (37.2.2) p -процентные доверительные пределы

$$(37.3.5) \quad a^* + \sum_1^k \beta_i^* (X_i - \bar{x}_i) \pm \pm t_p \frac{\sigma^*}{\sqrt{n-k-1}} \sqrt{1 + \sum_{i,j=1}^k \frac{l_{ij}}{L} (X_i - \bar{x}_i)(X_j - \bar{x}_j)},$$

где t_p следует брать с $n-k-1$ степенями свободы. В заключение сделаем три замечания, имеющие большое значение во многих приложениях:

1. Не будем предполагать, что y имеет нормальное распределение, а предположим лишь, что при любых фиксированных значениях x_i среднее значение y определяется линейным выражением (37.3.1), а стандартное отклонение всегда равняется s . При этих более общих предположениях можно показать, что оценки (37.3.2) для параметров a и β_i являются наилучшими (т. е. имеют наименьшие дисперсии) среди всех несмещенных оценок, являющихся линейными функциями от наблюденных значений y_i . Дисперсии и смешанные вторые моменты этих оценок уже даны формулами (37.3.4), а $\frac{n\sigma^{*2}}{n-k-1}$ является несмешенной оценкой для s . Далее, наилучшей линейной несмешенной оценкой для ординаты линии регрессии в какой-нибудь данной точке (X_1, \dots, X_k)

является $a^* + \sum_1^k \beta_i^*(X_i - \bar{x}_i)$, и стандартная ошибка этой оценки равна коэф-

фициенту при t_p в формуле (37.3.5). Это утверждение эквивалентно классической теореме о наименьших квадратах, полученной Марковым и другими авторами. Доказательство читатель сможет найти, например, у Неймана и Дэвида [169].

2. Рассмотренные в настоящем параграфе величины x_i могут быть любыми величинами, зависящими друг от друга или независимыми, но удовлетворяющими единственному условию $L \neq 0$, которое означает, что n точек (x_1, \dots, x_k) не могут лежать на одной гиперплоскости в k -мерном пространстве переменных x_i . В частности, все x_i могут быть функциями от одной независимой переменной x . Предположим, например, что x_i есть полином $p_i(x)$ степени i . Тогда поставленная выше проблема является проблемой параболической регрессии (см. параграф 21.6). Если полиномы $p_i(x)$ удовлетворяют условиям ортогональности, которые в этом случае имеют вид, указанный в примере 3 параграфа 12.6, то рассмотренная выше матрица L является диагональной и все вычисления значительно упрощаются, как мы видели это в аналогичном случае, рассмотренном в параграфе 21.6.

3. Если условие $L \neq 0$ не выполняется, то дисперсии и смешанные вторые моменты величин \hat{y}_i^* становятся бесконечными или неопределенными, как это видно из (37.3.4). Если L очень мало, но не равно точно нулю, то точки (x_1, \dots, x_k) лежат "почти" на одной гиперплоскости. В этом случае в наших формулах для доверительных интервалов и т. п. могут появиться очень большие коэффициенты или коэффициенты, являющиеся отношениями очень малых чисел. Небольшие ошибки в данных или в вычислениях, небольшие отклонения от нормальности и т. д. будут при этом оказывать большое влияние на результат, так что здесь рекомендуется особенная осторожность. Такое явление часто имеет место, когда x_i связаны очень сильно, что часто случается в экономических данных. Методы, которые следует применять для анализа регрессии с данными такого рода, неоднократно дискутировались многими авторами, особенно в связи с проблемами, которые будут рассмотрены в следующем параграфе. Мы рекомендуем читателю в этом отношении, например, обширную работу Шульца [34] и статьи Фриша [113], [114] и Вольда [247], [248].

37.4. Дальнейшие проблемы регрессии. В некоторых приложениях теории регрессии, например, в психологии и в экономике, мы имсем дело со множеством случайных величин x_1, \dots, x_m , которые можно представить в виде

$$(37.4.1) \quad x_1 = a_{11}u_1 + \dots + a_{1n}u_n + v_1,$$

где $m > n$ и $u_1, \dots, u_n, v_1, \dots, v_m$ суть $m+n$ некоррелированных случайных величин, причем $A = A_{mn} = \{a_{ij}\}$ — матрица ранга n .

В психологии, в *факторном анализе* человеческих способностей, величины x_1, \dots, x_m представляют собой измерения m различных данных способностей некоторого лица, u_1, \dots, u_n — более или менее „общие“ факторы интеллигентности и v_1, \dots, v_m — „специальные“ факторы, связанные с отдельными способностями. Основные проблемы в этих случаях обычно связаны с возможностью представления данного множества величин x_i в виде (37.4.1) и с существованием и количеством „общих“ факторов u_j .

С другой стороны, в некоторых экономических проблемах имеются теоретические основания предполагать, что рассматриваемые величины удовлетворяют некоторым линейным (или приближенно линейным) соотношениям. Однако часто эти величины нельзя наблюдать непосредственно, так как появляются „ошибки“ или „возмущения“. Вместо „систематических составляющих“ вышеуказанных величин x_i :

$$x'_i = a_{i1}u_1 + \dots + a_{in}u_n,$$

между которыми имеется $m-n$ линейных соотношений, мы можем наблюдать лишь сами величины x'_i , определяемые формулами (37.4.1), где v_i представляют собой „возмущения“. Основные проблемы здесь связаны с нахождением оценок для коэффициентов линейных соотношений между систематическими составляющими x'_i .

Проблемы такого рода настолько тесно связаны с частными прикладными областями, что мы не можем заниматься их полным исследованием в нашей книге. Приложения к психологии, относящиеся к этому кругу идей, были впервые исследованы Спирменом. В этом отношении мы отсылаем читателя, например, к обзорам Спирмена [35a] и Томсона [37a]. Описанные выше экономические проблемы вызвали построение Фришем [114] *конфлюэнтного анализа*, который получил дальнейшее развитие и был приведен в связь с теорией выборок и оценок в работах Купмана [142], Рейерсоля [207] и других авторов.

Выведем здесь лишь одно простое свойство матрицы вторых моментов Λ для множества величин x_1, \dots, x_m , которые можно представить в виде (37.4.1). Без ограничения общности можно предположить, что единицами измерения величин x_i , u_j и v_i служат их средние значения и что при всех j $E(u_j^2) = 1$. Положим $E(v_i^2) = \sigma_i^2$ и обозначим

через Σ диагональную матрицу с диагональными элементами $\sigma_1, \dots, \sigma_m$. Тогда, согласно параграфу 22.6, имеем

$$\Lambda = AA' + \Sigma^2.$$

Если все σ_i положительны, то матрица вторых моментов Λ имеет ранг m , так что распределение величин x_1, \dots, x_m собственное (см. параграф 22.5). С другой стороны *), матрица AA' имеет ранг $n < m$. Отсюда следует, что в матрице вторых моментов Λ любой минор порядка $\geq n+1$, не содержащий ни одного элемента главной диагонали **), равен нулю. Непосредственно видно, что этим же свойством обладает корреляционная матрица $P = \{\rho_{ij}\}$ величин x_1, \dots, x_m . Эта теорема была получена Терстоном.

Рассмотрим частный случай $n=1$ (в приложениях к психологии этот случай соответствует наличию лишь одного „общего“ фактора). Тогда коэффициенты корреляции ρ , соответствующие любым четырем различным индексам h, i, j, k , удовлетворяют *тетраэдному соотношению*:

$$\begin{vmatrix} \rho_{hj} & \rho_{hk} \\ \rho_{ij} & \rho_{ik} \end{vmatrix} = \rho_{hj}\rho_{ik} - \rho_{hk}\rho_{ij} = 0.$$

*) По 11.6, ранг AA' не превосходит n , и легко видеть, что в этой матрице имеется хотя бы один минор порядка n , не равный нулю.

**) Следует указать, что миноры, удовлетворяющие этим условиям, существуют лишь тогда, когда $2n+2 \leq m$.

ТАБЛИЦА 1

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ (см. гл. 17)

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad \varphi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

$\varphi^{(v)}(x) = (-1)^v H_v(x) \varphi(x)$, где $H_v(x)$ — полином Эрмита степени v . При отрицательных значениях x функции вычисляются с помощью соотношений
 $\Phi(-x) = 1 - \Phi(x)$, $\varphi(-x) = \varphi(x)$, $\varphi^{(v)}(-x) = (-1)^v \varphi^{(v)}(x)$.

x	$\Phi(x)$	$\varphi(x)$	$\varphi'(x)$	$\varphi''(x)$	$\varphi^{(3)}(x)$	$\varphi^{(4)}(x)$	$\varphi^{(5)}(x)$	$\varphi^{(6)}(x)$
0,0	0,50000	0,39894	-0,00000	-0,39894	+0,00000	+1,19683	-0,00000	-5,98413
0,1	0,53983	0,39695	0,03970	0,39298	0,11869	1,16708	0,59146	5,77625
0,2	0,57926	0,39104	0,07821	0,37540	0,23150	1,07990	1,14197	5,17112
0,3	0,61791	0,38139	0,11442	0,34706	0,33295	0,94130	1,61420	4,22226
0,4	0,65542	0,36827	0,14731	0,30935	0,41835	0,76707	1,97770	3,01241
0,5	0,69146	0,35207	0,17603	0,26405	0,48409	0,55010	2,21141	1,64481
0,6	0,72757	0,33322	0,19993	0,21326	0,52783	0,32309	2,30517	-0,23237
0,7	0,75804	0,31225	0,21858	0,15925	0,54863	+0,09371	2,26012	+1,11354
0,8	0,78814	0,28969	0,23175	0,10429	0,54694	-0,12468	2,08800	2,29382
0,9	0,81594	0,26609	0,23948	-0,05056	0,52445	0,32034	1,80951	3,23026
1,0	0,84134	0,24197	0,24197	0,00000	0,48394	0,48394	1,45182	3,87153
1,1	0,86413	0,21785	0,23964	+0,04575	0,42895	0,60909	1,04580	4,19585
1,2	0,88493	0,19419	0,23302	0,08544	0,36352	0,69255	0,62301	4,21034
1,3	0,90320	0,17137	0,22278	0,11824	0,29184	0,73413	-0,21300	3,94753
1,4	0,91924	0,14973	0,20962	0,14374	0,21800	0,73642	+0,15897	3,45953
1,5	0,93319	0,12952	0,19428	0,16190	0,14371	0,70425	0,47355	2,81094
1,6	0,94520	0,11092	0,17747	0,17304	0,07809	0,64405	0,71813	2,07125
1,7	0,95543	0,09405	0,15988	0,17775	+0,01759	0,56316	0,88702	1,30785
1,8	0,96407	0,07895	0,14211	0,17685	-0,03411	0,46915	0,98090	+0,58014
1,9	0,97128	0,06562	0,12467	0,17126	0,07605	0,36928	1,00583	-0,06467
2,0	0,97725	0,05349	0,10798	0,16197	0,10798	0,26996	0,97184	0,59390
2,1	0,98214	0,04398	0,09237	0,14998	0,13024	0,17646	0,89150	0,98987
2,2	0,98610	0,03547	0,07804	0,13622	0,14369	0,09274	0,77844	1,24885
2,3	0,98928	0,02833	0,06515	0,12152	0,14920	-0,02141	0,64604	1,37863
2,4	0,99180	0,02239	0,05375	0,10660	0,14834	+0,03623	0,50642	1,39654
2,5	0,99379	0,01753	0,04382	0,09202	0,14242	0,07997	0,36974	1,32421
2,6	0,99534	0,01358	0,03532	0,07824	0,13279	0,11053	0,24376	1,18645
2,7	0,99653	0,01042	0,02814	0,06355	0,12071	0,12926	0,13381	1,00761
2,8	0,99744	0,00792	0,02216	0,05414	0,10727	0,13793	+0,04287	0,80970
2,9	0,99813	0,00595	0,01726	0,04411	0,09339	0,13850	-0,02810	0,61102
3,0	0,99865	0,00443	0,01330	0,02545	0,07977	0,12926	0,07977	0,42546
3,1	0,99903	0,00327	0,01013	0,02813	0,06694	0,12313	0,11395	0,26242
3,2	0,99931	0,00238	0,00763	0,02203	0,05523	0,11066	0,13319	0,12712
3,3	0,99952	0,00172	0,00568	0,01704	0,04485	0,09690	0,14036	-0,02130
3,4	0,99966	0,00123	0,00419	0,01301	0,03586	0,08290	0,13840	+0,05607
3,5	0,99977	0,00087	0,00305	0,00882	0,02825	0,06943	0,13000	0,10784
3,6	0,99984	0,00061	0,00220	0,00732	0,02194	0,05703	0,11755	0,13862
3,7	0,99989	0,00042	0,00157	0,00539	0,01680	0,04599	0,10297	0,15102
3,8	0,99993	0,00029	0,00111	0,00392	, 0 69	0,03646	0,08777	0,15124
3,9	0,99995	0,00020	0,00077	0,00282	0,00946	0,02842	0,07302	0,14264
4,0	0,99997	0,00013	-0,00054	+0,00201	-0,00696	+0,02181	-0,05912	+0,12861

ТАБЛИЦА II
НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ (см. параграф 17.2)

Вероятность того, что наблюденное значение нормально распределенной величины ξ отличается от среднего значения m в ту или иную сторону более чем на взятое λ раз стандартное отклонение σ , равняется

$$P = P(|\xi - m| > \lambda\sigma) = 2[1 - \Phi(\lambda)] = \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{\lambda} e^{-\frac{t^2}{2}} dt.$$

Значение $\lambda = \lambda_p$, соответствующее $P = \frac{p}{100}$, называется p -процентным значением нормального отклонения.

λ_p как функция от p		p как функция от λ_p	
$p = 100P$	λ_p	λ_p	$p = 100P$
100	0,0000	0,0	100,000
95	0,0627	0,2	84,148
90	0,1257	0,4	68,916
85	0,1891	0,6	54,851
80	0,2533	0,8	42,371
75	0,3186	1,0	31,731
70	0,3833	1,2	23,014
65	0,4538	1,4	16,151
60	0,5244	1,6	10,960
55	0,5978	1,8	7,186
50	0,6745	2,0	4,550
45	0,7534	2,2	2,781
40	0,8416	2,4	1,640
35	0,9346	2,6	0,932
30	1,0364	2,8	0,511
25	1,1503	3,0	0,270
20	1,2816	3,2	0,137
15	1,4395	3,4	0,067
10	1,6449	3,6	0,032
5	1,9690	3,8	0,014
1	2,5758	4,0	0,006
0,1	3,2905		
0,01	3,8906		

ТАБЛИЦА III
РАСПРЕДЕЛЕНИЕ χ^2 (см. параграф 18.1)

Плотность вероятности $k_n(x)$ распределения χ^2 с n степенями свободы определяется формулой (18.1.3).

p -процентное значение χ_p^2 величины χ^2 с n степенями свободы определяется так, чтобы вероятность для наблюдаемого значения χ^2 превысить χ_p^2 равнялась

$$P = \frac{p}{100} = P(\chi^2 > \chi_p^2) = \int_{\chi_p^2}^{\infty} k_n(x) dx.$$

Число степеней свободы n	χ_p^2 как функция от n и $p = 100 P$													
	$p=99$	98	95	90	80	70	50	30	20	10	5	2	1	0,1
1	0,000	0,001	0,004	0,016	0,064	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,341	16,268
4	0,297	0,429	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,465
5	0,554	0,752	1,145	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086	20,517
6	0,872	1,134	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	1,239	1,564	2,167	2,833	3,822	4,671	6,246	8,38	9,803	12,017	14,067	16,622	18,475	24,322
8	1,646	2,032	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,567	18,168	20,090	26,125
9	2,088	2,532	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	2,558	3,059	3,940	4,865	6,179	7,967	9,342	11,781	13,442	15,887	18,307	21,161	23,209	29,588
11	3,053	3,609	4,575	5,578	6,959	8,148	10,311	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	3,571	4,178	5,226	6,304	7,807	9,024	11,340	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	4,107	4,765	5,892	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,332	25,472	27,688	34,528
14	4,660	5,368	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
15	5,229	5,935	7,261	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,697
16	5,812	6,614	7,962	9,312	11,152	12,624	15,328	18,418	20,465	23,542	26,296	29,533	32,000	39,252
17	6,408	7,255	8,672	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,790
18	7,015	7,906	9,390	10,865	12,857	14,440	17,328	20,601	22,760	25,989	28,869	32,346	34,805	42,312
19	7,633	8,567	10,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191	43,820
20	8,260	9,237	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566	45,315
21	8,897	9,915	11,591	13,240	15,445	17,182	20,337	23,828	26,171	29,615	32,671	36,243	38,932	46,797
22	9,542	10,600	12,338	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289	48,268
23	10,196	11,293	13,091	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	38,968	41,628	49,728
24	10,856	11,992	13,848	15,659	18,062	19,943	23,337	27,096	29,553	33,195	36,415	40,270	42,580	51,179
25	11,524	12,697	14,611	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	41,566	44,314	52,620
26	12,198	13,409	15,379	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	42,856	45,642	54,052
27	12,879	14,125	16,151	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	44,140	46,963	55,476
28	13,565	14,847	16,928	18,979	21,588	23,647	27,326	31,391	34,027	37,916	41,337	45,419	48,278	56,893
29	14,236	15,574	17,708	19,768	22,475	24,577	28,326	32,461	35,139	39,087	42,557	46,693	49,588	58,302
30	14,953	16,306	18,493	20,599	23,364	25,508	29,336	33,533	36,250	40,256	43,773	47,962	50,892	59,703

ТАБЛИЦА IV

t-РАСПРЕДЕЛЕНИЕ (см. параграф 18.2)

Плотность вероятности $s_n(x)$ t-распределения с n степенями свободы определяется формулой (18.2.4).

p-процентное значение t_p величины t с n степенями свободы определяется так, чтобы вероятность P для наблюденного значения t превзойти по абсолютной величине t_p равнялась

$$P = \frac{p}{100} = P(|t| > t_p) = 2 \int_{t_p}^{\infty} s_n(x) dx.$$

Число степеней свободы n	t_p как функция от n и $p = 100P$												
	$p=90$	80	70	60	50	40	30	20	10	5	2	1	0,1
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,883	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	0,134	0,271	0,414	0,569	0,741	0,941	1,192	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,153	1,476	2,015	2,571	3,365	4,032	6,859
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,913	2,447	3,143	3,707	5,559
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,830	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,119	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,795	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,693	0,873	1,083	1,353	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,691	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,723	2,080	2,518	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,519	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,069	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,061	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,758	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,433	2,704	3,551
60	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,480
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

К Н И Г И

1. Aitken A. C. Determinants and Matrices. University Mathematical Texts, 1. Edinburgh and London, 1944 (3-е изд.).
2. — Statistical Mathematics. University Mathematical Texts, 2. Edinburgh and London, 1944 (3-е изд.).
3. Bochner M. Introduction to higher Algebra. New York, 1908. *На русском языке*: Бочнер М. Введение в высшую алгебру. Москва, 1935.
4. Bochner S. Vorlesungen über Fourier'sche Integrale. Leipzig, 1932.
5. Bohr H. and Mollerup J. Lærebog i matematisk Analyse, I—IV. Kobenhavn, 1938—1942 (2-е изд.).
6. Borel É. Leçons sur la théorie des fonctions. Paris, 1914 (2-е изд.).
7. Borel É. и др. Traité du calcul des probabilités et de ses applications. Paris, 1924 и далее.
8. Bonnier G. and Tedin O. Biologisk variationsanalys. Stockholm, 1940.
9. Charlier C. V. L. Vorlesungen über die Grundzüge der mathematischen Statistik. Lund, 1931.
- 9a. — Application de la théorie des probabilités à l'astronomie. Paris, 1931. ([7], т. II, ч. IV.)
10. Crainér H. Sannolikhetskalkylen och några av dess användningar. Stockholm, 1927.
11. — Random Variables and Probability Distributions. Cambridge Tracts in Mathematics, No. 36. Cambridge, 1937. *На русском языке*: Крамер Г. Случайные величины и распределения вероятностей. Москва, 1947.
12. Elderton W. P. Frequency Curves and Correlation. Cambridge, 1938 (3-е изд.).
13. Fisher R. A. Statistical Methods for Research Workers. Edinburgh — London, 1941 (8-е изд.).
14. — The Design of Experiments. Edinburgh — London, 1937 (2-е изд.).
15. Fréchet M. Recherches théoriques modernes sur la théorie des probabilités. Paris, 1937—1938. ([7], т. I, ч. III.)
16. Gauss C. F. Werke, Bd. 4, Göttingen, 1880.
17. Hobson E. W. The Theory of Functions of a Real Variable, I—II. Cambridge, 1923—1927.
18. Jeffreys H. Theory of Probability. Oxford, 1939.
19. Kendall M. G. The Advanced Theory of Statistics, I. London, 1943. См. также Yule G. U.
20. Keynes J. M. A Treatise on Probability. London, 1921.
21. Колмогоров А. Н. Основные понятия теории вероятностей. Москва, 1931. *На немецком языке*: Колмогоров А. Grundbegriffe der Wahrscheinlichkeitsrechnung. Berlin, 1933.
22. Laplace P. S. Théorie analytique des probabilités. Paris, 1812 (1-е изд.), 1814 (2-е изд.), 1820 (3-е изд.).
23. Lebesgue H. Leçons sur l'intégration et la recherche des fonctions primitives. Paris, 1928 (2-е изд.). *На русском языке*: Лебег А. Интегрирование и отыскание примитивных функций. Москва, 1934.
24. Lévy P. Calcul des probabilités. Paris, 1925.
25. — Théorie de l'addition des variables aléatoires. Paris, 1937.

26. Lundborg H. and Linders F. J. The Racial Characters of the Swedish Nation. Uppsala, 1926.
27. Mises R. v. Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik. Leipzig — Wien, 1931.
28. — Wahrscheinlichkeit, Statistik und Wahrheit. Wien, 1936. (2-е изд.). На русском языке: Мизес Р. Вероятность, статистика и истина. Москва.
29. Moivre A. de. Miscellanea Analytica. 1733. II дополнение.
Mollerup J., см. Вонг H.
30. Neyman J. Lectures and Conferences on Mathematical Statistics. Washington, 1938.
31. Poincaré H. Calcul des probabilités. Paris, 1912 (2-е изд.).
32. Poisson S. D. Recherches sur la probabilité des jugements etc. Paris, 1837.
33. Saks S. Theory of the Integral. Warszawa, 1937 (2-е изд.). Готовится к печати русский перевод.
34. Schultz H. The Theory and Measurement of Demand. Chicago, 1938.
35. Snedecor G. W. Statistical Methods. Ames, Iowa, 1940.
- 35a. Spearman C. The Abilities of Man. London, 1927.
36. Szegő G. Orthogonal Polynomials. New York, 1939.
Tedin O. см. Bonnier G.
37. Thiele T. N. Theory of Observations. London, 1903.
- 37a. Thomson G. H. The Factorial Analysis of Human Ability. London, 1939.
38. Titchmarsh E. C. Introduction to the Theory of Fourier Integrals. Oxford, 1937. На русском языке: Тичмарш Е. Введение в теорию интегралов Фурье. Москва, 1947.
39. Uspensky J. V. Introduction to Mathematical Probability. New York, 1937.
40. Vallée-Poussin Ch. J. de la. Intégrales de Lebesgue, fonctions d'ensembles, classes de Baire. Paris, 1934 (2-е изд.).
41. Wiener N. The Fourier Integral and certain of its Applications. Cambridge, 1933.
42. Wilks S. S. The Theory of Statistical Inference. Ann Arbor, 1937.
43. Yule G. U. and Kendall M. G. An Introduction to the Theory of Statistics. London, 1940 (12-е изд.).

СТАТЬИ*)

50. Aitken A. C. On the graduation of data by the orthogonal polynomials of least squares. *Proc. R. Soc. Edinburgh*, 53 (1933), 54.
53. Arley N. On the distribution of relative errors from a normal population of errors. *K. Danske Vid. Selsk., Mat.-fys. Medd.* 18, № 3, (1940).
Babington Smith B. см. Kendall M. G.
54. Bartlett M. S. On the theory of statistical regression. *Proc. R. Soc. Edinburgh*, 53 (1933), 260.
55. — The information available in small samples. *PCPS*, 32 (1936), 560.
56. — Complete simultaneous fiducial distributions. *AMS*, 10 (1939), 129.
— См. Wishart J.

*) Сокращение названий журналов: *AE*—Annals of Eugenics; *AMS*—Annals of Mathematical Statistics; *B*—Biometrika; *CR*—Comptes Rendus de l'Académie des Sciences, Paris; *JRS*—Journal of the Royal Statistical Society; *MA*—Mathematische Annalen; *M*—Metron; *PCPS*—Proceedings of the Cambridge Philosophical Society; *PRS*—Proceedings of the Royal Society, London, ser. A; *PTRS*—Philosophical Transactions of the Royal Society, London, ser. A; *SA*—Skandinavisk Aktuarieridskrift; *TAMS*—Transactions of the American Mathematical Society.

60. Behrens W. U. Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen. *Landwirtsch. Jahrbücher*, 68 (1929), 897.
 61. Berg G. The prognosis of open pulmonary tuberculosis. *Acta Tuberculosis Scand.*, Suppl. IV (1939).
 62. Bergström H. On the central limit theorem. *SA* (1944), 139; (1945), 106.
 63. Бергштром С. Н. Распространение предельной теоремы вероятностей из суммы зависимых случайных величин. *Усп. матем. наук*, 10 (1944). *На французском языке: Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. MA*, 97 (1927), 1.
 63a. Bortkiewicz L. v. Das Gesetz der kleinen Zahlen. Leipzig (1898).
 64. Cantelli F. P. La tendenza ad un limite nel senso del calcolo delle probabilità. *Rend. Circ. Mat. Palermo*, 16 (1916), 191.
 65. Charlier C. V. L. Researches into the theory of probability. *K. Fysiogr. Sällsk. Handl.*, B 16 (1906).
 Clopper C. J. см. Pearson E. S.
 66. Cochran W. G. The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. *PCPS*, 30 (1933—1934), 178.
 Corbett A. S. см. Fisher R. A.
 67. Craig C. C. An application of Thiele's semi-invariants to the sampling problem. *M*, 7, № 4 (1928), 3.
 68. Cramér H. Sur quelques points du calcul des probabilités. *Proc. London Math. Soc.*, 23 (1925), LVIII.
 69. — On some classes of series used in mathematical statistics. *Sixth Scandinavian Congr. of Math.*, Kobenhavn, 1925.
 70. — On the composition of elementary errors. *SA* (1928), 12, 141.
 71. — On the representation of a function by certain Fourier integrals. *TAMS*, 46 (1939), 191.
 72. — Contributions to the theory of statistical estimation. *SA* (1946).
 73. Darmois G. Sur les lois de probabilités à estimation exhaustive. *CR*, 200 (1935), 1265.
 74. David F. N. см. Neyman J.
 Davies O. L. см. Pearson E. S.
 75. Doob J. L. Probability as measure. *AMS*, 12 (1941), 206.
 75a. — Probability and Statistics. *TAMS*, 36 (1934), 759.
 75b. — Statistical estimation. *TAMS*, 39 (1935), 410.
 76. Dugué D. Application des propriétés de la limite au sens du calcul des probabilités à l'étude de diverses questions d'estimation. *Journ. de l'Ec. Polytechn.* (1937), 305.
 80. Edgeworth F. Y. The law of error. *PCPS*, 20 (1905), 36.
 81. Eggenberger F. Die Wahrscheinlichkeitsansteckung. *Mitt. d. Ver. schweizerischer Vers.-Math.* (1924), 31.
 81a. Enero O. Om främnden vid fläckskådd samt om sambandet mellan plantalant och slutenhetsrad vid självsådd. *Norrl. Skogsav.-förb. Tidskr.* (1945), 161.
 82. Esscher F. On graduation according to the method of least squares by means of certain polynomials. *Förslär.-A.-B. Skandias Festskr.* (1930), II, 107.
 83. Esseen C. G. Fourier analysis of distribution functions. A mathematical study of the Laplace-Gauss law. *Acta Math.*, 77 (1945), 1.
 Essen-Möller E. см. Quensel C. E.
 84. Feller W. Sur les axiomatics du calcul des probabilités et leurs relations avec les expériences. *Actualités scientifiques et industrielles*, № 735 (1938), 7.
 85. — Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Math. Zeitschr.*, 40 (1935), 521.

86. Feller W. Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *II. Math. Zeitschr.*, 42 (1937), 301.
87. Fisher R. A. On an absolute criterion for fitting frequency curves. *Mess. of Math.*, 41 (1912), 155.
88. — Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *B*, 10 (1915), 507.
89. — On the mathematical foundations of theoretical statistics. *PTRS*, 222 (1921), 309.
90. — On the "probable error" of a coefficient of correlation deduced from a small sample. *M*, 1, № 4 (1921), 1.
91. — On the interpretation of χ^2 from contingency tables, and the calculation of P. *JRS*, 85 (1922), 87.
92. — The goodness of fit of regression formulae and the distribution of regression coefficients. *JRS*, 85 (1922), 597.
93. — The distribution of the partial correlation coefficient. *M*, 3 (1924), 329.
94. — On a distribution yielding the error functions of several well-known statistics. *Proc. Intern. Math. Congr. Toronto* (1924), 805.
95. — The conditions under which χ^2 measures the discrepancy between observation and hypothesis. *JRS*, 87 (1924), 442.
96. — Theory of statistical estimation. *PCPS*, 22 (1925), 700.
97. — Applications of Student's distribution. *M*, 5, № 3 (1925), 90.
98. — The general sampling distribution of the multiple correlation coefficient. *PRS*, 121 (1928), 654.
99. — Moments and product moments of sampling distributions. *Proc. London Math. Soc.*, 30 (1929), 199.
100. — Inverse probability. *PCPS*, 26 (1930), 528.
101. — The moments of the distribution for normal samples of measures of departure from normality. *PRS*, 130 (1930), 16.
102. — The concepts of inverse probability and fiducial probability referring to unknown parameters. *PRS*, 139 (1933), 343.
103. — Two new properties of mathematical likelihood. *PRS*, 144 (1934), 285.
104. — The logic of inductive inference. *JRS*, 98 (1935), 39.
105. — The fiducial argument in statistical inference. *AE*, 6 (1935), 391.
106. — On a point raised by M. S. Bartlett on fiducial probability. *AE*, 7 (1937), 370.
107. — The comparison of samples with possibly unequal variances. *AE*, 9 (1939), 174.
108. — A note on fiducial inference. *AMS*, 10 (1940), 383.
109. — The asymptotic approach to Behrens' integral with further tables for the d test of significance. *AE*, 11 (1941).
110. Fisher R. A. and Tippett L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *PCPS*, 24 (1928), 180.
111. Fisher R. A., Corbet A. S. and Williams C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. of Animal Ecology*, 12 (1943), 42.
112. Fréchet M. Sur la convergence "en probabilité". *M*, 8, № 4 (1930), 3.
- 112a. — Sur l'extension de certaines évaluations statistiques au cas de petits échantillons. *Rev. Inst. Intern. de Statistique* (1943), 182.
113. Frisch R. Correlation and scatter in statistical variables. *Nord. Statist. Tidsskr.*, 8 (1929), 36.
114. — Statistical confluence analysis by means of complete regression systems. Oslo (1934).
115. Geary R. C. Distribution of Student's ratio for non-normal samples. *JRS*, Suppl. 3 (1936).
116. — A general expression for the moments of certain symmetrical functions of normal samples. *B*, 25 (1933), 184.

- 116a. Geary R. C. Comparison of the concepts of efficiency and closeness for consistent estimates of a parameter. *P.*, 33 (1944), 123.
 117. Гнеденко Б. В. (Gnedenko B.) О характеристических функциях. *Бюл. Мос. Гос. Университета*, 1, № 5 (1937).
 118. Gram J. P. Om Rækkeudviklinger bestemte ved Hjælp af de mindste Kvadraters Methode. Kobenhavn, 1879.
 119. Greenwood M. and Yule G. U. An Inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease, or of repeated accidents. *JRS*, 83 (1920), 255.
 120. Gumbel E. J. Les valeurs extrêmes des distributions statistiques. *Ann. Inst. Henri Poincaré*, 5 (1936), 115.
 121. Hagstrom K. G. La loi de Pareto et la réassurance. *SA* (1925), 65.
 121a. — Alcune formule appartenenti alla statistica rappresentativa. *Giorn. Ist. Italiano d. Attuari*, 3 (1932), 147.
 122. — Inkostutjämningen i Sverige. *Skand. Bankens Kvart.-skr.*, April (1944). Hald A. см Rasch G.
 123. Haldane J. B. S. The mean and variance of χ^2 , when used as a test of homogeneity, when expectations are small. *P.*, 31 (1940), 346.
 124. Hartley H. O. The range in normal samples. *P.*, 32 (1942), 334.
 — см. Pearson F. S.
 125. Helmert F. R. Über die Wahrscheinlichkeit von Potenzsummen der Beobachtungsfehler etc. *Z. f. Math. u. Phys.*, 21 (1876).
 126. Hotelling H. The generalization of Student's ratio. *AMS*, 2 (1931), 360.
 127. Hsu P. L. Contribution to the theory of Student's *t*-test as applied to the problem of two samples. *Statist. Reserach Nem.*, 2 (1938), 1.
 128. Hultkrantz J. V. Über die Zunahme der Körpergrösse in Schweden in den Jahren 1840—1926. Uppsala, 1927.
 130. Ingham A. E. An integral which occurs in statistics. *PCPS*, 29 (1932), 271.
 131. Irwin J. O. The further theory of Francis Galton's individual difference problem. *B.*, 17 (1925), 100.
 132. — On the frequency distribution of the means of samples from populations of certain of Pearson's types. *M.*, 8, № 4 (1930), 51.
 133. — Mathematical theorems involved in the analysis of variance. *JRS*, 94 (1931), 284.
 134. Jordan C. Approximation and graduation according to the principle of least squares by orthogonal polynomials. *AMS*, 3 (1932), 257.
 135. Kapteyn J. C. Skew frequency curves in biology and statistics. Groningen, 1903, 1916.
 136. Kendall M. G. The conditions under which Sheppard's corrections are valid. *JRS*, 101 (1938), 592.
 137. — and Babington Smith B. Randomness and random sampling numbers. *JRS*, 101 (1938), 147, *JRS*, Suppl. 6 (1939), 51.
 139. Хинчин А. Я. (Khintchine A.) Sur la loi des grands nombres. *CR*, 188 (1929), 477.
 140. — Sul dominio di attrazione della legge di Gauss. *Giorn. Ist. Italiano d. Attuari*, 6 (1935), 378.
 140a. Kolodziejczyk S. On an important class of statistical hypotheses. *E*, 27 (1935), 161.
 141. Koopman B. O. On distributions admitting a sufficient statistic. *TAMS*, 39 (1936), 359.
 142. Koopmans T. Linear regression analysis of economic time series. *Netherlands Econ. Inst., Haarlem*, 1937.
 143. Kullback S. An application of characteristic functions to the distribution problem of statistics. *AMS*, 5 (1934), 263.

144. Langdon W. H. and Ore O. Semi-invariants and Sheppard's correction. *Annals of Math.*, 31 (1930), 230.
145. Lévy P. Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchaînées. *Journ. Math. pures appl.*, 14 (1935), 347.
146. Ляпунов А. М. (Lyapounoff A.) Sur une proposition de la théorie des probabilités. *Full. Acad. Sc. St.-Pétersbourg*, 13 (1900), 359.
147. — Nouvelle forme du théorème sur la limite de probabilité. *Mém. Acad. Sc. St.-Pétersbourg*, 14, № 5 (1901).
148. Lindeberg J. W. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Zeitschr.*, 15 (1922), 211.
149. Lindquist R. A treatise on reliable predictions of water conditions. Stockholm, 1932.
150. Lukacs E. A characterization of the normal distribution. *AMS*, 13 (1942), 91.
151. Lukomski J. On some properties of multidimensional distributions. *AMS*, 10 (1939), 236.
152. Lundberg O. On random processes and their application to sickness and accident statistics. (*Dissert.*) Stockholm, Uppsala, 1940.
154. Madow W. G. Limiting distributions of quadratic and bilinear forms. *AMS*, 11 (1940), 125.
155. Mendel G. Versuche mit Pflanzenhybriden. *Verhandl. naturforsch. Ver. Prunn*, 4 (1865).
156. Mises R. v. Grundlagen der Wahrscheinlichkeitsrechnung. *Math. Zeitschr.*, 4 (1919), 1.
157. — Deux nouveaux théorèmes de limite dans le calcul des probabilités. *Revue Fac. Sc. Istanbul*, 1 (1935), 61.
158. — Les lois de probabilité pour les fonctions statistiques. *Ann. Inst. Henri Poincaré*, 6 (1936), 185.
159. — On the foundations of probability and statistics. *AMS*, 12 (1941), 191.
- 159a. Newbold E. Practical applications of the statistics of repeated events, particularly to industrial accidents. *JRS*, 90 (1927), 487.
160. Neyman J. Contributions to the theory of small samples drawn from a finite population. *P*, 17 (1925), 472.
161. — On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *JRS*, 97 (1934), 558.
162. — Su un teorema concernente le cosiddette statistiche sufficienti. *Giorn. Ist. Italiano. d. Attuari*, 6 (1935), 320.
163. — On the problem of confidence intervals. *AMS*, 6 (1935) p. 111.
164. — "Smooth test" f. r goodness of fit. *SA* (1937), 149.
165. — Outline of a theory of statistical estimation based on the classical theory of probability. *PIRS*, 236 (1937), 333.
166. — L'estimation statistique traitée comme un problème classique de probabilité. *Actualités scientifiques et industrielles*, № 739 (1938), 25.
167. — Fiducial argument and the theory of confidence intervals. *B*, 32 (1941), 128.
168. — Basic ideas and some recent results of the theory of testing statistical hypotheses. *JRS*, 105 (1942), 292. Русский перевод в *Усп. матем. наук*, 10 (1944).
169. Neyman J. and David F. N. Extension of the Markoff theorem on least squares. *Statist. Research. Mem.*, 2 (1938), 105.
170. Neyman J. and Pearson E. S. On the use and interpretation of certain test criteria for purposes of statistical inference. *B*, 20 A (1928), 175, 263.
171. Neyman J. and Pearson E. S. On the problem of the most efficient tests of statistical hypotheses. *PIRS*, 231 (1933), 289.
172. — Contributions to the theory of testing statistical hypotheses. *Statist. Research. Mem.*, 1 (1936), 1, 2 (1938), 25.

173. Neyman J. and Pearson E. S. Sufficient statistics and uniformly most powerful tests of statistical hypotheses. *Statist. Research Mem.*, 1 (1930), 113. Ориг. см. Langdon W. H.
180. Pearson K. Contributions to the mathematical theory of evolution. *PTRS*, 185 (1894), 71.
181. — Contributions etc., II: Skew variation in homogeneous material. *PTRS*, 186 (1895), 343.
182. — Contributions etc., IV: On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *PTRS*, 191 (1898), 229.
183. — On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag.*, V, 50 (1900), 157.
- 183a. — On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, VI, 2 (1901), 559.
184. — On the systematic fitting of curves to observations and measurements. *B*, 1 (1902), 265, 2 (1902), 1.
185. — Researches on the mode of distribution of the constants of samples taken at random from a bivariate normal population. *PRS*, 112 (1926), 1.
186. — On the probability that two independent distributions of frequency are really samples from the same population. *B*, 8 (1911), 250.
187. Pearson K. and Tocher J. F. On criteria for the existence of differential death-rates. *B*, 11 (1916), 159.
190. Pearson E. S. The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. *B*, 20 A (1928), 356.
191. — The probability integral transformation for testing goodness of fit and combining independent tests of significance. *B*, 30 (1938), 134.
 — См. также Neyman J.
195. Pearson E. S. and Clopper C. J. The use of confidence or fiducial limits illustrated in the case of the binomial. *B*, 26 (1934), 404.
196. Pearson E. S. and Davies O. L. Method of estimating from samples the population standard deviation. *JRS Suppl.*, 1 (1934), 76.
197. Pearson E. S. and Hartley H. O. The probability integral of the range in samples of n observations from a normal population. *B*, 32 (1942), 301.
198. Pitman E. J. G. The «closest» estimates of statistical parameters. *PCPS*, 33 (1937), 212.
199. — The estimation of the location and scale parameters of a continuous population of any given form. *B*, 30 (1939), 391.
200. Quensel C. E. The distributions of the second moment and of the correlation coefficient in samples from populations of type A. *Kungl. Fyslogr. Sällsk. Handl.*, 49, no. 4 (1938).
201. — Inkomstfördelning och skattetryck. Publ. by Sveriges Industriförbund, Stockholm, 1944.
202. — On the logarithmico-normal distribution. *SA* (1945), 141.
203. Quensel C. E. and Essen-Möller E. Zur Theorie des Vaterschaftsnachweises auf Grund von Ähnlichkeitsbefunden. *Zeitschr. f. gerichtl. Medizin*, 31 (1939), 70.
205. Radon J. Theorie und Anwendung der absolut additiven Mengenfunktionen. *Sitzungsber. Akad. Wien*, 122 (1913), 1295. Русский перевод в *Усп. матем. наук*, 1 (1936).
206. Rasch G. and Hald A. Nogle Anvendelser af Transformationsmetoden i den normale Fordelings Teori. *Festskrift til Prof. J. F. Steffensen*, Kobenhavn (1943), 52.
207. Reiersøl O. Confluence analysis by means of instrumental sets of variables. *Arkiv för matematik etc.*, 32 A, no. 4 (1945).

208. Романовский В. И. (Romanovsky V). Sur certaines espérances mathématiques et sur l'erreur moyenne du coefficient de corrélation. *CR*, 180 (1925), 1897.
209. — On the moments of standard deviations and of correlation coefficient in samples from normal population. *M*, 5, no. 4. (1925), 3.
210. — On the distribution of the regression coefficient in samples from normal population. *Известия АН СССР*, 20 (1926), 613.
211. Sheppard W. F. On the application of the theory of error to cases of normal distribution and normal correlation. *PTRS*, 192 (1898), 101..
212. — On the calculation of the most probable values of frequency constants for data arranged according to equidistant divisions of a scale. *Proc. London Math. Soc.*, 29 (1898), 353.
213. — The fit of a for nula for discrepant observations. *PTRS*, 228 (1929).
- 213a. Simonsen W. On distributions of functions of samples from a normally distributed infinite population. *SA*, (1944), 235; (1945), 20.
214. Слутский Е. Е. (Slutsky E.) Über stochastische Asymptoten und Grenzwerte. *M*, 5, no. 3. (1925), 3.
215. Смирнов Н. В. (Smirnov N.) Sur la distribution de ω^2 . *CR*, 202 (1936), 449.
216. Soper H. E. and others. On the distribution of the correlation coefficient in small samples. *B*, 11 (1917), 328.
217. Steffensen J. F. Factorial moments and discontinuous frequency-functions. *SA* (1923), 73.
218. — Free functions and the Student-Fisher theorem. *SA* (1936), 108.
220. Stieltjes T. J. Extrait d'une lettre adressée à M. Hermite. *Bull. Sc. Math.*, 2-e série, 13 (1889), 170.
221. Student*. The probable error of a mean. *B*, 6 (1908), 1.
222. — Probable error of a correlation coefficient. *B*, 6 (1908), 302.
223. Sukhatme P. V. On Fisher and Behrens' test of significance for the difference in means of two normal samples. *Sankhyā*, 4 (1938), 39.
224. Tang P. C. The power function of the analysis of variance tests with tables and illustrations of their use. *Statist. Research Mem.*, 2 (1938), 126.
225. Thompson W. R. On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *AMS*, 6 (1935), 214.
226. Tippeit L. H. C. On the extreme individuals and the range of samples taken from a normal population. *B*, 17 (1925), 364.
— см. Fisher R. A.;
Tocher J. F. см. Pearson K.
227. Tschuprow A. A. On the mathematical expectation of the moments of frequency distributions. *B*, 12 (1919), 140, 185; *B*, 13 (1921), 233.
- 227a. — Zur Theorie der Stabilität statistischer Reihen. *SA* (1918, 1919).
228. Wahlund S. Demographic studies in the nomadic and the settled population of northern Lapland. Uppsala, 1932.
229. Welch B. L. The significance of the difference between two means when the population variances are unequal. *B*, 29 (1938), 350.
230. Wicksell S. D. On the genetic theory of frequency. *Arkiv för matematik etc.*, 12 (1917), 20.
231. — Sex proportion and parental age. *Kungl. Fysiogr. Sällsk. Handl.*, 37, no 6 (1926).
232. Wilks S. S. Certain generalizations in the analysis of variance. *B*, 24 (1932), 471.
233. — Shortest average confidence intervals from large samples. *AMS*, 9 (1933), 166.
234. — Fiducial distributions in fiducial inference. *AMS*, 9 (1938), 272.
Williams C. B. см. Fisher R. A.
240. Wishart J. The generalized product moment distribution in samples from a normal multivariate population. *B*, 20 A (1928), 32.

241. Wishart J. and Bartlett M. S. The generalized product moment distribution in a normal system. *PCPS*, 29 (1932), 260.
245. Wold H. Sulla correzione di Sheppard. *Giorn. Ist. Italiano d. Attuari*, 5 (1934), 304.
246. — Sheppard's corrections formulae in several variables. *SA* (1934), 248.
- 246a. — A study in the analysis of stationary time series. (*Diss.*) Stockholm, Uppsala, 1938.
247. — Efterfrågan på Jordbruksprodukter och dess känslighet för pris-och inkomstförändringar. *Statens Off. Utr.*, no. 16, (1940).
248. — A theorem on regression coefficients obtained from successively extended sets of variables. *SA* (1945), 181.
250. Yates F. Contingency tables involving small numbers and the χ^2 test. *JRS*, Suppl. 1 (1934), 217.
251. Yule G. U On the theory of correlation for any number of variables treated by a new system of notation. *PRS*, 79 (1907), 182.
— см. также Greenwood M.

ТАБЛИЦЫ

260. British Association, Mathematical Tables, Vol. 7: Tables of the Probability integral by W. F. Sheppard (1939).
261. David F. N. Tables of the Correlation Coefficient. London, 1938.
262. Fisher R. A and Yates F. Statistical Tables. Edinburgh and London, 1943 (2-е изд.).
263. Kendall M. G. and Babington Smith B., Tables of Random Sampling Numbers. Tracts for computers, no. 24, 1940.
264. Pearson K. Tables for Statisticians and Biometricalians. I, 1924 (2-е изд.), II, 1931.
265. — Tables of the Incomplete Γ function. 1922.
266. — Tables of the Incomplete B -function.
267. Tippett L. H. C. Random Sampling Numbers. Tracts for computers, no. 15, 1927.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

(Цифры обозначают страницы)

I

- Анализ смешанных вторых моментов (analysis of covariance) 597
- Вероятность (probability) 169
- апостериорная (probability a posteriori) 552
 - априорная (probability a priori) 552
 - доверительная (fiducial probability) 551
 - условная (conditional probability) 181
- Вероятностная функция (probability function) 71
- Временные ряды (time series) 512
- Выбор (sampling) 355
- без возвращения (sampling without replacement) 364, 561, 569:
 - с возвращением (sampling with replacement) 364
 - простой случайный (simple random sampling) 356
 - пристрастный (biased sampling) 363
- Выборка (sample) 165, 356
- группированная (grouped sample) 361, 393
- Выборки среднее значение (sample mean) 379
- пространство (sample space) 419, 519, 539, 561
 - характеристики (sample characteristics) 375
- Выборочные значения (sample values) 357
- Выборочная точка (sample point) 419
- Генеральная совокупность (population) 165, 356
- Гистограмма (histogram) 361
- Грама-Шарлье ряд (Gram-Charlier series) 247, 286
- Дециль (decile) 204
- Дисперсия (variance) 203, 381
- обобщенная (generalized variance) 332, 443
 - остаточная (residual variance) 307, 336
 - условная (conditional variance) 296, 298
- Дисперсионный анализ (analysis of variance) 268, 584
- Дисперсионное отношение (variance ratio) 587
- Доверительная область (confidence region) 550, 563
- Доверительный интервал (confidence interval) 501, 553, 555, 565
- предел (confidence limit) 501, 507, 557
 - уровень (confidence level) 557
- Значимое отклонение (significant deviation) 457
- Значимости пределы (significance limits) 368
- уровень (significance level) 368, 457
- Квантиль (quantile) 204
- Квартиль (quartile) 204
- Композиция (composition) 214
- Конфлюэнтийный анализ (confluence analysis) 606
- Корреляция (correlation) 293, 306, 332
- Коррелированные величины (correlated values) 307
- Корреляционное отношение (correlation ratio) 310
- Коэффициент асимметрии (coefficient of skewness) 201, 211, 391
- доверия (confidence coefficient) 557
 - изменчивости (coefficient of variation) 391
 - корреляции (coefficient of correlation) 306, 325, 393, 435, 448
 - полный (total coefficient of correlation) 336
 - сводный (multiple coefficient of correlation) 339, 450
 - частный (partial coefficient of correlation) 337, 347
 - разброса (scatter coefficient) 332, 449
 - расхождения (coefficient of divergence) 486
 - регрессии (coefficient of regression) 302, 333, 346, 439, 447
 - эксцесса (coefficient of excess) 208, 211, 391
- Кривая равных вероятностей (equiproability curve), 318

- Кривая регрессии** (regression curve) 299, 300
— плотности (frequency curve) 192
Критерий значимости (test of significance) 368, 370, 453, 489, 571
— наиболее мощный (most powerful test) 577, 580
— несмешанный (unbiased test) 578, 583
— равномерно наиболее мощный (uniformly most powerful test) 577, 580
Критерий однородности (test of homogeneity) 482
— согласия (test of goodness of fit) 368, 453, 488
Критерий χ^2 (χ^2 -test) 453, 460
Критическое множество (critical set) 575
Латинские квадраты (latin squares) 595
Матрица вторых моментов (moment matrix) 292, 326
Матрица коэффициентов корреляции (correlation matrix) 326
Медиана (median) 201, 404
Метод максимума правдоподобия (maximum likelihood method) 462, 541
— минимума χ^2 (χ^2 minimum method) 463, 549
— моментов (method of moments) 540
— репрезентативной выборки (representative method) 364, 561, 569
Мода (mode) 202
Момент (moment) 86, 101, 196, 325, 380, 398
— абсолютный (absolute moment) 197
— групповой (raw or grouped moment) 394
— смешанный (product or mixed moment) 291
— факториальный (factorial moment) 234
— центральный (central moment) 197, 291, 325, 383, 399
Независимые величины (independent variables) 182, 187, 211, 212, 314, 348
— испытания (independent trials) 231
— повторения (эксперимента) (independent repetitions) 184
— события (independent events) 182
Некоррелированные величины (uncorrelated variables) 307, 326
Нормированная величина (standardized variable) 203, 209
Нулевая гипотеза (null hypothesis) 371
Остаток (residual) 336
Оценка (estimation, estimate) 362, 372, 513
Оценка асимптотически-эффективная (asymptotically efficient estimate) 531, 537
— достаточная (sufficient estimate) 530, 532, 540
— максимального правдоподобия (maximum likelihood estimate) 542, 543
— несмешанная (unbiased estimate) 385
— состоятельная (consistent estimate) 385, 531
— эффективная (efficient estimate) 517, 522, 529, 536, 538, 540
Плоскость регрессии (regression plane) 333
Плотность вероятности (probability density) 72, 97, 188, 322
Поверхность регрессии (regression surface) 332
Поправка Иэтса (Yates' correction) 484
Поправка Шеппарда (Sheppard's correction) 396
Распределение (distribution) 71, 95, 175
— бимодальное (bimodal distribution) 202
— вероятностей (probability distribution) 175, 176, 187, 188
— выборки (distribution of a sample) 357
— выборочное (sampling distribution) 360
— мультимодальное (multimodal distribution) 202
— несобственное (singular distribution) 327
— совместное (joint or simultaneous distribution) 177
— собственное (non-singular distribution) 327
— унимодальное (unimodal distribution) 202
— условное (conditional distribution) 179, 187, 295
— усеченное (truncated distribution) 273
— частное (marginal distribution) 97, 178, 288, 321
Распределение биномиальное (binomial distribution) 218, (529 b)
— отрицательно-биномиальное (negative binomial distribution) 287, 473
— бета (Beta distribution) 269, 279
— Коши (Cauchy's distribution) 273, 403, 533
— Лапласа (Laplace's distribution) 273, 409
— мультикомиальное (multimodal distribution) 350, 455

- Распределение нормальное (normal distribution) 116, 233, 317, 341, 409, 478, 525, 532, 547, 558, 563, 566, 580, 583
- асимптотически-нормальное (asymptotically normal distribution) 239
- логарифмически-нормальное (logarithmico-normal distribution) 246, 286
- Парето (Pareto's distribution) 275
- Пуассона (Poisson's distribution) 239, 277, 351, 470, 529
- Стьюдента (Student's distribution t) 264, 278
- Фишера (Fisher's distribution z) 268
- χ^2 (χ^2 -distribution) 259, 277
- прямоугольное (rectangular distribution) 270, 407
- треугольное (triangular distribution) 272, 408
- Рассеяние (dispersion) 203
- Регрессия (regression) 299, 332, 597, 605
- Регрессия линейная (linear regression) 300, 301, 333, 346
- ортогональная (orthogonal regression) 304, 340
- параболическая (parabolic regression) 305, 607
- Семи-инвариант (semi-invariant) 210, 216, 327, 383
- Семи-интерквартильная широта (semi-interquartile range) 204
- Случайные блоки (randomized blocks) 592
- Случайная величина (random variable) 173, 187
- Случайный процесс (random process) 473, 512
- Случайный эксперимент (random experiment) 157
- Случайные числа (random sampling numbers) 363
- Смешанный момент второго порядка (covariance) 291, 326
- Смещение (bias) 335, 423
- Составная величина (combined variable) 177
- Составной эксперимент (combined experiment) 176
- Сокращение данных (reduction of data) 369
- Среднее (mean) 194, 197, 200
- Среднее значение (mean value) 194, 290, 325
- условное (conditional mean value) 295, 297, 298, 332
- Среднее отклонение (mean deviation) 204
- Средняя квадратическая регрессия (mean square regression) 301, 333
- Средняя квадратическая связанность (mean square contingency) 311, 312, 483
- Стандартное отклонение (standard deviation) 203
- Стандартная ошибка (standard error) 491
- Статистическая устойчивость (statistical regularity) 164, 177
- Степени свободы (degrees of freedom) 259, 264, 417
- Степень связности (degree of dependence or association) 480
- Стьюдентово отношение (Student's ratio) 424
- обобщенное (generalized Student's ratio) 447
- Сходимость по вероятности (convergence in probability) 279, 330
- Таблица сопряженности признаков (contingency table) 478
- Точка сосредоточения массы (discrete mass point) 71, 96
- Уравнение правдоподобия (likelihood equation) 542
- Факторный анализ (factor analysis) 606
- Функция плотности (frequency function) 72, 97, 188, 322
- правдоподобия (likelihood function) 519, 542
- распределения (distribution function) 71, 95, 175, 188, 288, 321
- Частота (frequency, frequency ratio) 163.
- Широта (range) 204, 405
- Эджворта ряд (Edgeworth's series) 254
- Экстремальные значения (extreme values) 405
- Эксцесс (excess) 208
- Элемент вероятности (probability element) 189, 322
- Эллипс рассеяния (ellipse of concentration) 312, 535
- Эллипсоид рассеяния (ellipsoid of concentration) 331
- Эмпирическая функция распределения (sum polygon) 360
- Эффективность (efficiency) 522, 529
- асимптотическая (asymptotic efficiency) 532, 537

- Аддитивный класс множеств** 20, 42
Алгебраическое дополнение 126
Борелевские множества 24, 25, 27, 42,
44, 45, 174
Бета-функция 145
Вековое уравнение 131
Вектор 123
Внутренняя точка 23, 26
Гамма-функция 143
Гиперплоскость 27
Гиперповерхность 27
Главная диагональ 122
Детерминант 125
Измеримые множества 40, 42, 44
Интеграл Лебега 47, 48, 49, 56, 59, 61
— Лебега — Стильбеса 63, 78, 80, 81,
82, 85, 100
— Римана 48, 88, 104
— Римана — Стильбеса 88, 104,
Хеллингера 529
Интегралы Дирихле 139
— Фурье 105
Интервал 21, 22, 26
— непрерывности 72, 95
Лемма Бореля 30
— Фишера 413
Матрица 129
— вырожденная 127
— диагональная 122
— единичная 123
— квадратная 122
— невырожденная 127
— нулевая 123
— обратная 128
— ортогональная 130
— преобразования 124
— присоединенная 128
— симметричная 122
— транспонированная 122
Мера внешняя 37
— внутренняя 37
— Лебега 40, 45, 62, 91
Множество 13
— дополнительное 17
— линейное 21, 27
— ограниченное 22, 26
— прямоугольное 29
— пустое 14
— счетное 13, 19, 36
— цилиндрическое 28
Минор 126
— главный 126
- Неравенство Бъенэме — Чебышева** 206
— Ляпунова 283
— треугольника 26
— Шварца 104
- Окрестность** 22, 26
Ортогональное преобразование 130
- Подматрица** 122
Подмножество 14
Подпространство 28
Полиномы Лагерра 152
— ортогональные 149, 306
— Эрмита 151, 248
Постоянная Эйлера 143
Почти всюду 53
Преобразование линейное 124, 309,
329, 344
— ортогональное 130
Произведение пространств 29, 99
Пространство 14
Суммы Дарбу 46, 77, 88, 100, 102
Теорема Ба'еса 551, 552
— Бернулли 221
— Больцано — Вейерштрасса 23
— единственностии 109, 118
— Линдеберга — Леви 240, 316, 348
— Ляпунова 241
— Муавра 225
— непрерывности 112, 119
— сложения 220, 230, 238, 261, 414
— Хицчина 281
— Чебышева 206, 280
Форма билинейная 124
— квадратичная 125
— неотрицательная 132
— положительно определенная 132
— полуопределенная 132
— обратная 128
Формула Стирлинга 148
— Эйлера — Маклорена 142
Функция *B*-измеримая 49, 101
— интегрируемая 47, 49, 56, 59, 80
— множества 33, 61, 62, 63, 70, 91,
92, 93, 174
— ступенчатая 66, 69
— характеристическая 105—115, 117,
119, 209, 293, 327
- Характеристические числа** 131
Центральная предельная теорема 238,
348
- Числа Бернулли** 141

ДОПОЛНЕНИЕ КО ВТОРОМУ ИЗДАНИЮ

A. V. Прохоров

Место книги Г. Крамера в системе подготовки специалистов по математической статистике в значительной мере определяется тем, что в ней систематически и на высоком уровне строгости проведено математическое обоснование таких классических разделов, как теория оценок и теория проверки статистических гипотез. Выбор проблем (который сегодня кажется умеренным) и методов их решения делают книгу прекрасным введением в математическую статистику, предназначенным для тех читателей, которые хотят научиться использовать доступный математический аппарат для доказательства классических результатов, а затем углублять свои знания, изучая монографии и статьи, посвященные избранным проблемам и основанные на более специализированном теоретическом фундаменте.

Характер настоящего дополнения к книге Крамера в основном определяется следующими соображениями. Прежде всего мы отдаем себе отчет в том, что задача объединения всех разделов, методов и результатов, развитых в математической статистике за последние тридцать лет, и изложения их с полной логической ясностью на общей теоретической основе вряд ли может быть решена в дополнении к этой книге (даже в виде сводки или обзора всех накопленных результатов). Однако как учебное руководство книга Крамера нуждается в дополнении справочно-библиографического характера для тех хорошо подготовленных читателей, которые интересуются литературой по специальным вопросам математической статистики, затронутым в книге недостаточно глубоко или получившим решение лишь в недавнее время. Поэтому мы решили ограничиться материалом справочного характера. При этом мы не претендовали на полноту даже списка литературы, но старались указывать те источники, к достоинствам которых относится и достаточно обширная библиография по соответствующим вопросам. Весь справочный материал естественным образом разделяется на три основные части соответственно построению книги. Мы надеемся, что обновление библиографического списка позволит читателям ориентироваться в широком круге вопросов математической статистики.

Часть I. Математическое введение

Главы 1—12. Все приведенные в этих главах сведения из теории множеств, теории меры и интеграла Лебега, линейной алгебры, теории интеграла Фурье полностью удовлетворяют тем задачам математической статистики, которыми ограничил себя Крамер. Основным в этой части разделом является введение в теорию меры и интегрирования и связанные с этим вопросы, имеющие отношение к теории вероятностей.

Общеизвестно, что серьезное изучение теории вероятностей и математической статистики невозможно без обращения к теории меры, однако следует понимать, что роль теории меры непосредственно определяется характером рассматриваемых задач. Лаконизм Крамера при изложении теории меры объясняется, в частности, тем, что автор ограничивается кругом вероятностных моделей, приводящих к распределениям в n -мерном евклидовом пространстве (эти распределения в конкретных задачах оказываются, как правило, дискретными или непрерывными). Эти модели соответствуют занимающему и до сих пор центральное место в математической статистике случаю повторной выборки фиксированного объема. В то же время строгий вывод фундаментальных результатов в таких областях математической статистики, как последовательный анализ (или, более общим образом, теория статистических решений) и теория оптимальной остановки наблюдений, см. ниже) и статистика случайных процессов, требует неизбежного обращения к глубоким фактам как общей теории меры, так и теории меры в топологических пространствах. Можно добавить, что при асимптотическом изучении использующих выборки фиксированного объема непараметрических критериев (типа критериев Колмогорова—Смирнова) порою целесообразно, а порою неизбежно обращение к таким понятиям, как сходимость мер в функциональных пространствах и т. д.

Читатели, которые хотели бы получить более обширные сведения из теории меры и интегрирования, могут обратиться к книгам Халмоса (1953), Сакса (1949), Колмогорова и Фомина (1968), Данфорда и Шварца (1962). Читателям, интересующимся теорией меры лишь в связи с теорией вероятностей, необходимо читать соответствующие разделы в книгах Дуба (1956), Лозва (1962), Невё (1969) (некоторые дополнительные ссылки см. ниже в тексте, относящемся ко второй части).

Информация об интеграле Фурье, полезная для изучающих теорию вероятностей, имеет отношение к свойствам характеристических функций и формулам обращения; поэтому необходимые ссылки будут сделаны во второй части добавления.

Глава, посвященная матрицам, детерминантам и квадратичным формам, может быть дополнена материалом книги Рао (1968), гл. 1.

Часть II. Случайные величины и распределения вероятностей

Глава 13. Обсуждение вопроса о соотношении между частотой и вероятностью и о роли вероятностных моделей может быть дополнено положениями и иллюстрациями из статьи Колмогорова (1956) (см. также Колмогоров (1971) и Нейман (1955)).

Глава 14. Построения этой главы уместно дополнить одним замечанием из монографии Гнеденко и Колмогорова (1949). Поскольку эта монография стала библиографической редкостью, приведем это важное замечание полностью. Определяя понятие вероятностной меры в R_n , авторы пишут: «Такое понятие n -мерного распределения вероятностей выбрано в качестве аксиоматической основы всех дальнейших рассмотрений в книге Крамера. Однако даже в узких рамках исключительного интереса к конечномерным распределениям этот подход не лишен недостатков. В этом можно убедиться, например, прочитав параграфы 14.2 и 14.5 книги Крамера. Его аксиома 3 заставляет предполагать, что основным объектом, свойства которого должны быть фиксированы при помощи аксиом, является какая-то не названная непосредственно совокупность всех случайных величин. В параграфе 14.5 даже «доказывается», что любая измеримая по Борелю функция случайной величины сама является случайной величиной. Но остается неясным, имеются ли при этом в виду те случайные величины, к которым относятся аксиомы 1, 2 и 3, или случайные величины в каком-то новом смысле. Можно было бы, конечно, избежать этих неясностей за счет еще более сильного ограничения: всегда отправляться от некоторого исходного распределения вероятностей $P_\xi(A)$ для некоторого запаса «основных» случайных величин $\xi_1, \xi_2, \dots, \xi_n$ и строго отличать от них «производные» случайные величины $\eta = f(\xi_1, \dots, \xi_n)$, для которых законы распределения вычисляются исходя из основного распределения $P_\xi(A)$. Более широкая и естественная перспектива и возможность все рассматриваемые случайные величины считать равноправными открываются лишь при более общем подходе к делу, развитом в книге Колмогорова (1933, 1974)». Как показывает опыт развития теории вероятностей и математической статистики, аксиоматика, предложенная А. Н. Колмогоровым, охватывает почти все задачи (исключение составляет лишь случай, когда ситуация описывается в терминах одних лишь условных вероятностей). С другой стороны, выяснилось, что для исключения некоторых патологических случаев целесообразно ограничить общее определение понятия распределения вероятностей. Первый шаг в этом направлении был сделан в упомянутой монографии Гнеденко и Колмогорова, где введено понятие совершенной меры и дано определение: «Вероятность $P(A)$ есть совершенная мера, удовлетворяющая условию нормированности». Однако, ограничение совершенными мерами еще

не снимало всех трудностей, и при дальнейшем изучении был выделен класс так называемых «плотных» (*tight*) мер; см. обзор Прохорова (1961) и книги Партаасарата (1967), Биллингслея (1968) и Топсё (1970).

Главы 15–20. Эти главы посвящены числовым характеристикам случайных величин, важнейшим классам одномерных дискретных и непрерывных распределений, а также предельным теоремам (закону больших чисел и центральной предельной теореме), применения которых в математической статистике весьма многочисленны.

Общая теория характеристических функций основана на свойствах интеграла Фурье. Литература, посвященная интегралу Фурье, насчитывает множество обстоятельных пособий, однако читателям, имеющим чисто вероятностные интересы, достаточно обратиться к книгам, содержащим все необходимые сведения об интеграле Фурье в связи с распределениями вероятностей; см., например, Гнеденко и Колмогоров (1949), Рамачандран (1967), Лукач (1970). Некоторые предельные теоремы для случайных величин с неотрицательными целочисленными значениями доказываются методом производящих функций; см., например, Феллер (1967).

Известна принципиальная роль неравенства Чебышева или его обобщения (см. формулу (15.7.1)) в теории вероятностей. Попытки уточнить это неравенство для некоторых специальных классов распределений объяснялись практическими соображениями. Были открыты также экспоненциальные варианты неравенства Чебышева для сумм взаимно независимых случайных величин; см. Бернштейн (1946, 1964), Карлин и Стадден (1966). Известны и асимптотические формулы для вероятностей так называемых «больших уклонений» (кстати, первые такие формулы были получены самим Крамером (1938)). Обзор важнейших результатов в этой области сосредоточен в монографиях Ибрагимова, Линника (1965) и Петрова (1972).

Теорема Муавра—Лапласа об аппроксимации биномиального распределения является самой первой предельной теоремой теории вероятностей. Важность этой теоремы подчеркивают также исследования, посвященные проблеме улучшения сходимости; см. Бернштейн (1943) и Феллер (1945).

Следует отметить, что все варианты центральной предельной теоремы доказываются Крамером по стандартному принципу: устанавливается сходимость характеристических функций нормированных сумм к характеристической функции нормального распределения, а затем используется известная теорема непрерывности Крамера — Леви (стр. 112) [вопрос о скорости сходимости обсуждается на стр. 254].

Проблема оценки скорости сходимости в центральной предельной теореме решается посредством построения разложений вероятностных распределений, основанных на нормальном распределении. Известны два типа таких разложений: разложения по ортого-

нальным полиномам Грама—Шарлье и асимптотические разложения Эджвортса. О принадлежащих Крамеру, Берри, Эссеену и их последователям результатах в области асимптотических разложений можно прочитать в монографиях Крамера (1947), Гнеденко и Колмогорова (1949), Петрова (1972). Библиография по вопросу асимптотической аппроксимации распределений имеется в книге Уоллеса (1958).

Большую роль в теории оценок и в теории проверки гипотез играют точные и асимптотические распределения функций («статистик» или «критериев») от случайных величин с известным совместным распределением вероятностей. Относительно точных распределений возникает вопрос, в какой мере они характеризуют распределение исходных случайных величин. Так, известно, что совместное распределение величин $\frac{x_i - \bar{x}}{s}$, $i = 1, 2, \dots, n$ (см. обозначения на стр. 376) в выборке из нормальной совокупности равномерно на $(n - 2)$ -мерной единичной сфере. При $n \geq 6$ верно и обратное, т. е. это свойство характеризует исходное нормальное распределение. Подобно этому и независимость \bar{x} и s^2 (см. стр. 417) характеризует нормальность исходного распределения. Известны глубокие результаты, показывающие, насколько сильным является налагаемое *a priori* ограничение независимости тех или иных статистик; известен и ряд других так называемых характеризационных теорем (это направление было начато и существенно развито Ю. В. Линником; с его современным состоянием можно познакомиться по монографии Кагана, Линника, Рао (1972)).

В асимптотической теории распределений «статистик» большое значение имеют различные «уллучшающие сходимость» преобразования случайных величин (см., например, замечания на стр. 278 о χ^2 и на стр. 436 о z -преобразовании Фишера). Некоторые употребительные преобразования перечислены в обзоре Хойла (1973). Ряд важных преобразований указан в работах Большева (1959, 1963), где, в частности, получена аппроксимация биномиального распределения, более точная, чем известные пуассоновское и нормальное приближения, и, кроме того, уточнены аппроксимации распределений некоторых статистик. О практическом применении этих преобразований см. также вводную статью к таблицам Большева и Смирнова (1965).

Для рассматриваемых Крамером задач математической статистики достаточно знакомства лишь с двумя типами сходимости последовательностей случайных величин: по распределению и по вероятности. О соотношении различных типов сходимости случайных величин и о весьма важном понятии сходимости с вероятностью единицы см. книги Лоэва (1962), Невё (1969) и Дюгэ (1972).

Главы 21—24. В разделе, посвященном распределениям в двумерном и многомерном евклидовых пространствах, Крамер делает ос-

новной упор на характеристику зависимости между случайными величинами; см. также аналитическую теорию корреляции и регрессии в обстоятельно написанной книге Кендалла и Стьюарта (1973). Варианты многомерной центральной предельной теоремы и обсуждение предельных проблем, возникающих при суммировании случайных векторов, приведены в обзоре Прохорова (1972).

Полезный справочный материал, необходимый при изучении частей I и II, содержит книга Прохорова и Розанова (1973).

Часть III. Статистические выводы

Глава 26. Об общих вопросах теории статистических выводов см. также статью Неймана (1955).

Главы 27—29. Основой практического использования статистических выводов служит теория распределений выборочных характеристик, т. е. распределений различных функций от случайных величин, входящих в выборку. Формально простая задача нахождения точных формул для функций распределения статистик оказывается в большинстве случаев технически трудно разрешимой. Подробные отчеты о методах получения точных выборочных распределений, наиболее часто используемых в приложениях теории оценок и теории проверки гипотез, полученных в явном виде или табулированном, содержатся в соответствующих разделах книг Большева и Смирнова (1965), Кендалла и Стьюарта (1966), Уилкса (1967). Решение многих задач математической статистики основывается на использовании предельных теорем для некоторых универсальных распределений. Однако точность соответствующих приближенных формул зачастую оказывается недостаточной для практических нужд. В асимптотической теории выборочных распределений в последние годы были предприняты попытки найти более точные аппроксимационные формулы для некоторых статистических распределений. Для повышения точности аппроксимации были использованы сходящиеся и асимптотические разложения распределений; основные результаты в этом направлении (касающиеся полиномиального, пуассоновского, гипергеометрического распределений, распределений Стьюдента, F , χ^2 , распределений непараметрической статистики) суммированы в работе Калинина (1968). В работе Моленара (1970) указаны весьма точные приближения для некоторых дискретных распределений и даны рекомендации, позволяющие использовать полученные результаты для таких непрерывных распределений, как F и χ^2 . Тем же целям служат улучшающие сходимость преобразования случайных величин. Некоторые статистические приложения так называемых асимптотически пирсоновских преобразований указаны в работах Большева (1961, 1963).

Следует выделить лишь частично затронутую в главе 28 задачу точного и асимптотического распределений членов вариационного

ряда или так называемых порядковых статистик, которая возникает в непараметрической статистике; подробное изложение вопроса см. в книгах Фрэзера (1957), Гумбеля (1965); см. также библиографии Сэвиджа (1962), Уолца (1962), остальные ссылки будут сделаны ниже.

О наиболее важных распределениях выборочных характеристик в случае многомерного статистического анализа можно узнать из книг Андерсона (1963), Кендалла и Стьюарта (1966 а), Кульбака (1967).

В связи с общими задачами теории проверки гипотез возникает проблема классификации и выделения семейств распределений, обладающих достаточными статистиками. О классах распределений, допускающих так называемые подобные зоны (зоны, подобные пространству выборок), в частности об экспоненциальном классе распределений, см. Леман (1964), Линник (1966).

Глава 30. Эта глава о критериях согласия посвящена в основном классическому критерию χ^2 К. Пирсона: имеется вывод предельного распределения для статистики χ^2 и даны многочисленные примеры использования критерия в различных ситуациях (см. также Ван дер Варден (1960)). Обзор результатов по χ^2 -критерию можно найти в работах Кокрена (1952, 1954), Ватсона (1959), Ланкастера (1969) и Никулина (1973). Об оптимальных свойствах критериев типа χ^2 для непрерывных распределений см. статью Чубисова (1971).

Из других критериев согласия упомянут лишь непараметрический ω^2 -критерий Крамера — Мизеса — Смирнова. Мощность и эффективность этого критерия была исследована Чубисовым (1961). Необходимые сведения о других непараметрических критериях, основанных на функционалах от разности функций эмпирического и теоретического распределений, таких, как критерий Колмогорова и Смирнова, содержатся в работах Гнеденко (1969), Смирнова (1970), Боровкова (1962), Гихмана, Гнеденко, Смирнова (1956), вводной части таблиц Большева и Смирнова (1965). Оптимальность некоторых непараметрических критериев обсуждается в работе Боровкова, Сычевой (1968). Общей теории так называемых ранговых критериев посвящена книга Гаека и Шидака (1971). Современное состояние теории непараметрических критериев отражено в обзорах Пури (1970, 1972), Хёфдинга (1968).

Главы 32—34. Значительные достижения теории оценок связаны с понятиями состоятельности, эффективности, достаточности. Возможности использования теории достаточных статистик для отыскания несмещенных оценок с минимальной дисперсией изучались в работах Гиршика, Мостеллера, Сэвиджа (1946), Блекуэлла (1947), Рао (1949), Лемана и Шеффе (1950), Колмогорова (1950). Фундаментальная теорема, принадлежащая Рао, Блекуэллу, Колмогорову, в частности, утверждает: если семейство зависящих от параметра $\theta \in \Theta$ распределений имеет достаточную статистику $T = T(X)$,

$X = (X_1, \dots, X_n)$, то для каждой оценки $\hat{\theta}$ ($E_{\theta}|\hat{\theta}| < \infty$) можно найти оценку $\theta^* = \theta^*(T(X))$, зависящую от X только через $T(X)$, такую, что $E_{\theta}\theta^* = E_{\theta}\hat{\theta}$, $\hat{\theta} \in \Theta$, при этом $D_{\theta}\theta^* \leq D_{\theta}\hat{\theta}$, если $E_{\theta}|\hat{\theta}|^2 < \infty$. Подобного рода теоремы во многих конкретных случаях позволяют строить наилучшие несмешанные оценки. Вопросам теоретического исследования характеристических свойств и критериев существования достаточных статистик посвящены работы Халмоса и Сэвиджа (1949), Дынкина (1951). В книге Кульбака (1967) обсуждаются информационные свойства достаточных статистик.

Вывод важнейшего в теории оценок неравенства Крамера — Рао (полученного также Дармуа и Фреше и называемого иначе неравенством информации) может быть дополнен ссылкой на работу Рао М. (1961), в которой развит общий подход к проблеме получения нижних границ для дисперсий оценок (см. также простое доказательство неравенства у Большева (1961)).

Непосредственное обобщение теоремы Крамера о достаточных условиях состоятельности и асимптотической эффективности оценок максимального правдоподобия дано Кулдорфом (1966), который исследовал проблему оценки параметров распределения по группированным выборкам (уточнение и многомерное обобщение этих результатов см. в работе Бодина (1970)). Исследованию асимптотических свойств оценок максимального правдоподобия посвящены работы Ле Кама (1956), Бахадура (1960), Рао (1962), Вольфовича (1965), Вейса и Вольфовича (1966, 1967). Обсуждение общих вопросов оценивания по методу максимального правдоподобия см. в обширном обзоре Нордена (1972, 1973). Дополнительные сведения о других способах получения оценок можно получить из книг Рао (1968), Кендалла и Стьюарта (1973), Линника (1962) (последняя полностью посвящена методу наименьших квадратов). Асимптотические методы теории оценивания в рамках общей теории статистических решений подверглись анализу в работе Ле Кама (1974), которая содержит также исторический обзор основных результатов и солидную библиографию. В работах Ибрагимова и Хасьминского (1972, 1973а, 1973б) получены новые результаты об асимптотических свойствах широкого класса оценок.

Заметную роль в задачах математической статистики стал играть так называемый эмпирический байесовский подход, предназначенный для построения оценок в тех случаях, когда оцениваемый параметр является случайной величиной с неизвестным (априорным) распределением; см. Роббинс (1956), Нейман (1962), Большев (1970).

Обзор основных фактов теории интервального оценивания имеется в книгах Уилкса (1967) и Кендалла и Стьюарта (1973). Вопрос о сопоставлении доверительного (Ю. Нейман) и фидуциального (Р. Фишер) методов рассматривается в работах Берн-

штейна (1964) и Неймана (1961). Связь между свойствами (такими, как несмещенность, инвариантность и пр.) доверительных множеств и параметрических критериев проверки гипотез обсуждается в книге Лемана (1964). О задаче одновременного оценивания посредством доверительных интервалов, возникающей в регрессионном анализе, см., например, Уилкс (1967), гл. 10, и Шеффе (1963).

Об использовании техники доверительных интервалов для построения так называемых толерантных интервалов можно прочитать в книге Кендалла и Стьюарта (1973). Вывод толерантных границ для нормального распределения имеется в работе Вальда и Вольфовича (1946), см. также таблицы Оуэна (1966).

Важным для приложений методом получения оценок является метод статистических испытаний (метод Монте-Карло), численный метод решения математических задач (например, вычисление интегралов, решение уравнений) посредством моделирования случайных величин. С обоснованием и вычислительными процедурами метода статистических испытаний можно познакомиться по монографии Ермакова (1971).

Глава 35. Теория статистической проверки гипотез, основные принципы которой были сформулированы Нейманом и Пирсоном, за последние двадцать лет была развита настолько, что стала неотъемлемой частью учебных руководств по математической статистике. По существу в дополнение к программе, намеченной Крамером, можно рекомендовать более широкое ознакомление с ролью фундаментальной леммы Неймана — Пирсона и со свойствами критерия отношения правдоподобия в случае выборок фиксированного объема (см., например, соответствующие разделы в книгах Уилкса (1967) и Рао (1968)). По поводу проверки сложных гипотез, в частности условий существования и построения оптимальных в том или ином смысле критериев, см. книгу Кендалла и Стьюарта (1973), а также статью Неймана (1959). Синтез многих важнейших результатов в области проверки гипотез принадлежит Леману (1964). Совершенно иначе построена книга Неймана (1968), вводящая в круг идей современной теории проверки гипотез при минимальном использовании математического аппарата.

Известная проблема Беренса — Фишера имеет солидную литературу, см., например, Брени (1955) и Тьюки (1957). Предлагаемое здесь (стр. 566—569, пример 2) решение соответствует «фициальной» концепции Р. Фишера, а не принятой в книге концепции Ю. Неймана. Проблема существования отвечающих принципам Неймана «подобных» критериев изучается в книге Линника (1966). Построению критерия, приближенно подобного пространству выборок, посвящены работы Пагуровой (1968 а, б); см. также Большев и Смирнов (1965) и Рао (1968), гл. 7.

Главы 36—37. Отметим, что разделы книги Крамера, посвященные дисперсионному и регрессионному анализу, изложены не-

сколько схематично. В настоящее время под влиянием практических потребностей эти направления математической статистики в достаточной мере обеспечены литературой, отражающей как теоретические, так и прикладные аспекты их развития. Постановка задач в этой области и полезные литературные ссылки имеются в книгах Уилкса (1967), гл. 10, и Рао (1968), гл. 4. Наиболее полно дисперсионный и регрессионный анализ изложен в монографиях Шеффе (1963), Плэкета (1960) и Уильямса (1959); широко обсуждаются вопросы оптимального планирования экспериментов, которое имеет целью увеличение точности статистических оценок; при регрессионном анализе, например, задача планирования состоит в указании таких значений контролируемых переменных, при которых некая заданная функция от неизвестных параметров обладает теми или иными экстремальными свойствами. Обширная библиография по планированию экспериментов имеется в книге Гринвуда и Хартли (1961). Подробный анализ темы планирования, посвященный практическим задачам регрессионного и конфлюэнтного анализа, имеется в работе Клепикова и Соколова (1964) [ср. Фёдоров (1969)]. С набором стандартных приемов регрессионного и дисперсионного анализа можно познакомиться по книге Хальда (1956).

В последние годы были развиты статистические методы анализа выборок из многомерных совокупностей. Многомерный статистический анализ, в значительной мере применимый к выборкам из нормальных совокупностей, кроме многомерных обобщений классических разделов теории оценок и теории проверки гипотез, включает в себя специфические проблемы классификации наблюдений (дискриминантный анализ) и оценки размерности множества наблюдаемых переменных (метод главных компонент, факторный анализ). Проблематика этого направления и основные результаты в наибольшей мере освещены в монографиях Андерсона (1963), Кендалла и Стьюарта (1966 б), Демпстера (1969), Кришнайя (1969); см. также библиографию Андерсон, Дас Гупта, Штьян (1969). Введением в задачи многомерного анализа могут служить соответствующие разделы книг Уилкса (1967) и Рао (1968). Математической теории факторного анализа посвящена книга Лоули и Максвелла (1967), включающая как характеристику моделей, так и примеры их использования.

При работе над книгой Крамер ограничил свою программу изложением и математическим обоснованием сложившихся разделов математической статистики; в частности, рассматриваемые Крамером методы оценки параметров и проверки гипотез основаны на предположении, что объем выборки фиксирован заранее. Идея определения числа испытаний в ходе эксперимента в зависимости от результатов уже проведенных наблюдений на простом примере различия двух гипотез привела к возникновению последовательного статистического анализа. Вальдом (1960) была создана теория

последовательного анализа и даны применения последовательного критерия отношения вероятностей (см. также Вальд и Вольфович (1948)). Простейшее представление о последовательном оценивании и последовательных критериях можно получить в книгах Уилкса (1967) и Рао (1968); см. также Джексон (1960), Джонсон (1961). Развитие методов последовательного анализа способствовало созданию общей теории статистических решающих функций; см. монографии Вальда (1950), Блекуэлла и Гиршика (1958), Чернова и Мозеса (1962), Де Гроота (1974). В то же время изучение последовательных статистических процедур в связи со случайными процессами привело к появлению теории оптимальной остановки наблюдений (раздел теории управляемых случайных процессов); см. Ширяев (1961), Чжоу, Роббинс, Зигмунд (1971).

Из непараметрических последовательных методов получения оценок следует также отметить метод стохастической аппроксимации, который является в определенном смысле вероятностным аналогом итерационных методов решения задач численного анализа. Наиболее полно задачи и результаты этого метода изложены в монографиях Вазана (1972), Невельсона и Хасьминского (1972).

Круг вопросов, которыми занимается современная математическая статистика, весьма широк. Некоторое представление об этом дает «Библиография статистических библиографий», выпускаемая Ланкастером (1968, 1969—1974).

СПИСОК ЛИТЕРАТУРЫ

- Андерсон (Anderson T. W.)**
 (1963) Введение в многомерный статистический анализ (перев. с англ.), М., Физматгиз.
 (1971) The statistical analysis of time series, New York. (Готовится русский перевод: Андерсон Т., Статистический анализ временных рядов, М., «Мир».)
- Андерсон, Дас Гупта, Штыян (Anderson T. W., Das Gupta S., Styan G. P. R.)**
 (1969) A bibliography of multivariate statistical analysis, London, Oliver and Boyd.
- Арлей, Бух (Arley N., Buch K.)**
 (1951) Введение в теорию вероятностей и математическую статистику (перев. с англ.), М., ИЛ.
- Бахадур (Bahadur R.)**
 (1960) On the asymptotic efficiency of tests and estimates, *Sankhya, Indian J. Statist.*, 22, № 3—4, 229—252.
- Бернштейн С. Н.**
 (1943) Возврат к вопросу о точности предельной формулы Лапласа, *Известия АН СССР*, сер. матем., 7, 3—16.
 (1946) Теория вероятностей, изд. 4, М.—Л., Гостехиздат.
 (1964) Собрание сочинений, том IV, М., «Наука», 386—393, 566—569.
- Биллингслий (Billingsley P.)**
 (1968) Convergence of probability measures, New York, John Wiley.
- Блекуэлл (Blackwell D.)**
 (1947) Conditional expectation and unbiased sequential estimation, *Ann. Math. Stat.*, 18, 105—110.

- Блекуэлл, Гиршик (Blackwell D., Girshik M. A.)
(1958) Теория игр и статистических решений (перев. с англ.), М., ИЛ.
- Бодин Н. А.
(1970) Оценка параметров распределения по группированным выборкам, *Тр. МИАН СССР*, CXI, 110—154, Л., «Наука».
- Большев Л. Н.
(1959) О преобразованиях случайных величин, *Теория вероятн. и ее примен.*, 4, № 2, 136—149.
(1961) Уточнение неравенства Рао — Крамера, *Теория вероятн. и ее примен.*, 6, № 3, 319—326.
(1963) Асимптотически пирсоновские преобразования, *Теория вероятн. и ее примен.*, 8, № 2, 129—155.
(1970) Applications of the empirical Bayes approach, *Actes, Congres intern. Math.*, v. 3, 241—247.
- Большев Л. Н., Смирнов Н. В.
(1965) Таблицы математической статистики, М., «Наука».
- Боровков А. А.
(1962) К задаче о двух выборках, *Известия АН СССР*, сер. математ., 26, 605—624.
- Боровков А. А., Сычева Н. М.
(1968) О некоторых асимптотически оптимальных непараметрических критериях, *Теор. вероятн. и ее. примен.*, 13, № 3, 385—418.
- Брени (Brégné H.)
(1955) L'état actuel du problème de Behrens—Fisher, *Trab. estatist.*, 6, № 2, 111—131.
- Вазан (Wasan M. T.)
(1972) Стохастическая аппроксимация (перев. с англ.), М., «Наука».
- Вальд (Wald A.)
(1950) Statistical decision functions, New York, John Wiley.
(1960) Последовательный анализ (перев. с англ.), М., Физматгиз.
- Вальд, Вольфович (Wald A., Wolfowitz J.)
(1946) Tolerance limits of a normal distribution, *Ann. Math. Stat.*, 17, 208—215.
(1948) Optimum character of the sequential probability ratio test, *Ann. Math. Stat.*, 19, 326—339.
- Ван дер Варден (van der Waerden B.)
(1960) Математическая статистика (перев. с нем.), М., ИЛ.
- Уотсон (Watson G. S.)
(1959) Some recent results in chi-square goodness-of-fit test, *Biometrics*, 15, № 3, 440—468.
- Вейс, Вольфович (Weiss L., Wolfowitz J.)
(1966) Generalized maximum likelihood estimators, *Теория вероятн. и ее примен.*, 11, № 1, 68—93.
(1967) Maximum probability estimators, *Ann. Inst. Statist. Math.*, 19, 193—206.
- Вольфович (Wolfowitz J.)
(1965) Asymptotic efficiency of the maximum likelihood estimator, *Теория вероятн. и ее примен.*, 10, № 2, 267—281.
- Гаек, Шидак (Hájek J., Sidák Z.)
(1971) Теория ранговых критериев (перев. с англ.), М., «Наука».
- Гиршик, Мостеллер, Сэвидж (Girshik M. A., Mosteller F., Savage L. J.)
(1946) Unbiased estimates for certain binomial sampling problems with applications, *Ann. Math. Stat.*, 17, 13—23.
- Гихман И. И., Гнеденко Б. В., Смирнов Н. В.
(1956) Непараметрические методы статистики, Труды III Всесоюз. матем. съезда, том III, М., Изд-во АН СССР, 320—334.
- Гнеденко Б. В., Колмогоров А. Н.
(1949) Предельные распределения для сумм независимых случайных величин, М., Гостехиздат.

- Гнеденко Б. В.
(1969) Курс теории вероятностей, изд. 5. М., «Наука».
- Гринвуд, Хартли (Greenwood J. A., Hartley H. O.)
(1961) Guide to tables in mathematical statistics, Princeton University Press.
- Гумбель (Gumbel E. J.)
(1965) Статистика экстремальных значений (перев. с англ.), М., «Мир».
- Даунфорд, Шварц (Dunford N., Schwartz J.)
(1962) Линейные операторы. Общая теория (перев. с англ.), М., ИЛ.
- De Groot (De Groot M.)
(1974) Оптимальные статистические решения (перев. с англ.), М., «Мир».
- Демпстер (Dempster A. P.)
(1969) Elements of continuous multivariate analysis, Addison—Wesley Publishing Company.
- Джексон (Jackson J. E.)
(1960) Bibliography on sequential analysis, *J. Amer. Statist. Ass.*, 55, 516.
- Джонсон (Johnson N. L.)
(1961) Sequential analysis: a survey, *J. Roy. Statist. Soc.*, A, 124, 372.
- Дуб (Doob J. L.)
(1956) Вероятностные процессы (перев. с англ.), М., ИЛ.
- Дыкин Е. Б.
(1951) Необходимые и достаточные статистики для семейства распределений вероятностей, УМН, 6, 68—90.
- Дюгэ (Dugué D.)
(1972) Теоретическая и прикладная статистика (перев. с фр.), М., «Наука».
- Ермаков С. М.
(1971) Метод Монте-Карло и смежные вопросы, М., «Наука».
- Ибрагимов И. А., Линник Ю. В.
(1965) Независимые и стационарно связанные величины, М., «Наука».
- Ибрагимов И. А., Хасминский Р. З.
(1972) Асимптотическое поведение некоторых статистических оценок в гладком случае, *Теория вероятн. и ее примен.*, 17, № 3, 469—486.
(1973 а) Асимптотическое поведение некоторых статистических оценок в гладком случае, *Теория вероятн. и ее примен.*, 18, № 1, 78—93.
(1973 б) О моментах обобщенных байесовских оценок и оценок максимального правдоподобия, *Теория вероятн. и ее примен.*, 18, № 3, 535—546.
- Каган А. М., Линник Ю. В., Рао С. Р.
(1972) Характеризационные задачи математической статистики, М., «Наука».
- Калинин В. М.
(1968) Предельные свойства вероятностных распределений, *Tr. МИАН СССР*, CIV, 88—134, Л., «Наука».
- Карлин, Стадден (Karlin S., Studden W. J.)
(1966) Tschebyshoff systems: with applications in analysis and statistics, New York, Intersci. Publs.
- Кендалл (Kendall M. G.)
(1962) Rank correlation methods (3ed), London, Griffin.
- Кендалл, Сьюард (Kendall M. G., Stuart A.)
(1966 а) Том 1. Теория распределений (перев. с англ.), М., «Наука».
(1973) Том 2. Статистические выводы и связи (перев. с англ.), М., «Наука».
(1966 б) Design and analysis, and time-series, London, Griffin. (Готовится русский перевод: Кендалла М., Сьюард А. Многомерный статистический анализ и временные ряды, М., «Наука».)
- Клепиков Н. П., Соколов С. Н.
(1964) Анализ и планирование экспериментов методом максимума правдоподобия, М., «Наука».
- Колмогоров А. Н.
(1933) Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin. (Русский перевод: Колмогоров А. Н. Основные понятия теории вероятностей, М.—Л., ОНТИ, 1936.)

- (1950) Несмешенные оценки, *Изв. АН СССР*, сер. матем., 14, № 4, 303—326.
- (1956) Теория вероятностей, сб. «Математика, ее содержание, методы и значение», том 2, гл. XI, М., Изд-во АН СССР.
- (1971) Вероятность, Большая Советская Энциклопедия, том 4, стр. 544.
- (1974) Основные понятия теории вероятностей, 2-ое изд., М., «Наука».
- Колмогоров А. Н., Фомин С. В.**
- (1968) Элементы теории функций и функционального анализа, М., «Наука».
- Кокрен (Cochran W. G.)**
- (1952) The χ^2 test of goodness-of-fit, *Ann. Math. Stat.*, 23, 315—345.
- (1954) Some methods for strengthening the common tests, *Biometrics*, 10, 417.
- Крамер (Cramer H.)**
- (1938) Sur un nouveau théorème-limite de la théorie des probabilités, Paris, Actual. sci. et ind., n 736. (Русский перевод: Крамер Г., Об одной новой предельной теореме теории вероятностей, *УМН*, 10 (1944), 166—178.)
- (1947) Случайные величины и распределения вероятностей (перев. с англ.), М., ИЛ.
- Кришнайя (Krishnaiah P. R.)**
- (1969) Multivariate Analysis — II. Proceed. of the Second International Symposium on Multivariate Analysis, Academic Press.
- Кулдорф (Kulldorff G.)**
- (1966) Введение в теорию оценивания (перев. с англ.), М., «Наука».
- Кульбак (Kullback S.)**
- (1967) Теория информации и статистика (перев. с англ.), М., «Наука».
- Ланкастер (Lancaster H. O.)**
- (1968) Bibliography of statistical bibliographies, London, Oliver and Boyd.
- (1969) The chi-squared distribution, New York, John Wiley.
- (1969—1974) Bibliography of statistical bibliographies, *Int. Stat. Rev.*: a second list, 37, 57—67; a third list, 38, 258—267; a fourth list, 39, 64—73; a fifth list, 40, 73—81; a sixth list, 41, 375—379; a seventh list, 42, 67—70.
- Ле Кам (Le Cam L.)**
- (1956) On the asymptotic theory of estimation and testing hypotheses, Proc. Third Berkeley Symposium on Math. Stat. and Prob., Vol. 1, 129—156, University of California Press.
- (1974) Notes on asymptotic methods in statistical decision theory, Publ. Centre de Recherches Mathématiques, Université de Montréal.
- Леман (Lehmann E. L.)**
- (1964) Проверка статистических гипотез (перев. с англ.), М., «Наука».
- Леман, Шеффе (Lehmann E. L., Scheffé H.)**
- (1950) Completeness, similar regions and estimates, *Sankhyā*, 10, 305—340.
- Линник Ю. В.**
- (1962) Метод наименьших квадратов и основы теории обработки наблюдений, изд. 2, М., Физматгиз.
- (1963) On the Berens — Fisher problem, Ottawa, Proc. Int. Stat. Conference.
- (1966) Статистические задачи с мешающими параметрами, М., «Наука».
- Лоули, Максвелл (Lawley D. N., Maxwell A. E.)**
- (1967) Факторный анализ как статистический метод (перев. с англ.), М., «Мир».
- Лоэв (Loève M.)**
- (1962) Теория вероятностей (перев. с англ.), М., ИЛ.
- Лукач (Lukacs E.)**
- (1970) Characteristic functions (2 ed), London, Griffin.
- Моленар (Molenar W.)**
- (1970) Approximations to the Poisson, binomial and hypergeometric distribution functions (Math. Centre Tracts, 31), Amsterdam Math. Centrum.
- Невельсон М. Б., Хасминский Р. З.**
- (1972) Стохастическая аппроксимация и рекуррентное оценивание, М., «Наука».

- Невё (Neveu J.)**
(1969) Математические основы теории вероятностей (перев. с фр.), М., «Мир».
- Нейман (Neyman J.)**
(1955) The problem of inductive inference, *Communications on pure and applied mathematics*, 8, 13—46.
(1959) Optimal asymptotic tests of composite hypotheses, *Probability and Stat: The H. Cramér volume*, Upsala, 213—234.
(1961) Silver jubilee of my dispute with Fisher, *J. Oper. Res. Soc. of Japan*, 3, 145—154.
(1962) Two breakthroughs in the theory of statistical decision making, *Int. Stat. Rev.*, 30, № 1, 11—27.
(1968) Вводный курс теории вероятностей и математической статистики (перев. с англ.), М., «Наука».
- Никулин М. С.**
(1973) Критерий хи-квадрат для непрерывных распределений с параметрами сдвига и масштаба, *Теория вероятн. и ее примен.*, 18, № 3, 583—591.
- Норден (Norden R. H.)**
(1972) A survey of maximum likelihood estimation, *Int. Stat. Rev.*, 40, № 3, Part 1, 329—354.
(1973) A survey of maximum likelihood estimation, *Int. Stat. Rev.*, 41, № 1, Part 2, 39—58.
- Оуэн (Owen D. B.)**
(1966) Сборник статистических таблиц, М., ВЦ АН СССР.
- Пагурова В. И.**
(1968 а) О сравнении средних значений в двух нормальных выборках. *Теория вероятн. и ее примен.*, 13, № 3, 561—568.
(1968 б) Критерий сравнения средних значений по двум нормальным выборкам, вып. 5, М., Изд. ВЦ АН СССР.
- Парасарати (Parathasarathy K. R.)**
(1967) Probability measures on metric spaces, New York, Academic Press.
- Петров В. В.**
(1972) Суммы независимых случайных величин, М., «Наука».
- Плэкет (Plackett R. L.)**
(1960) Principles of regression analysis, Oxford, Clarendon Press.
- Прохоров Ю. В.**
(1961) Method of characteristic functionals, Proc. 4th Berkeley Sympos. Math. Statist. and Probability, 1960, Vol. 2, Berkeley — Los Angeles, Univ. California Press, 403—419.
(1972) Многомерные распределения: неравенства и предельные теоремы, «Итоги науки», М., Изд-во ВИНИТИ.
- Прохоров Ю. В., Розаинов Ю. А.**
(1973) Теория вероятностей (изд. 2), М., «Наука».
- Пури (Puri M. L.)**
(1970) Nonparametric techniques in statistical inference, London, Cambridge Univ. Press.
(1972) Some aspects of nonparametric inference, *Int. Stat. Rev.*, 40, № 3, 299—327.
- Рамачандран (Ramachandran B.)**
(1967) Advanced theory of characteristic function, Statistic Publ. Soc., Calcutta.
(Готовится русский перевод: Рамачандран Б. Теория характеристических функций, М., «Наука».)
- Rao (Rao C. R.)**
(1949) Sufficient statistics and minimum variance estimates, *Proc. Cambr. Phil. Soc.*, 45, 213—218.
(1961) Apparent anomalies and irregularities in maximum likelihood estimation. *Bull. Inst. Internat. Statist.*, XXXVIII, 4-e livre, Tokyo.

- (1962) Efficient estimates and optimum inference procedures in large samples, *J. Roy. Statist. Soc.*, **B24**, №1, 46—63.
- (1968) Линейные статистические методы и их применения (перев. с англ.), М., «Наука».
- Pao (Rao M. M.)**
- (1961) Theory of lower bounds for risk functions in estimation, *Math. Ann.*, **148**, № 5, 379—398.
- Роббинс (Robbins H.)**
- (1951) Asymptotically subminimax solutions of compound statistical decision problems, Proceedings of the Second Berkeley Symposium on Statistics and Probability, 1950, 131—148, Berkeley and Los Angeles, University of California Press.
- (1956) An empirical Bayes' approach to statistics, Proceedings of the Third Berkeley Symposium on Statistics and Probability, I, 1955, 157—164, Berkeley and Los Angeles, University of California Press.
- Сакс (Saks S.)**
- (1949) Теория интеграла (перев. с англ.), М., ИЛ.
- Сархан, Гринберг (Sarhan A. E., Greenberg B. G.)**
- (1970) Введение в теорию порядковых статистик (перев. с англ.), М., «Статистика».
- Смирнов Н. В.**
- (1970) Теория вероятностей и математическая статистика, Избранные труды, М., «Наука».
- Соле (Soler J. L.)**
- (1972) Основные структуры математической статистики (перев. с франц.), М., «Мир».
- Спренд (Sprent P.)**
- (1969) Models in regression and related topics, London, Methuen.
- Сэвидж (Savage I. R.)**
- (1962) Bibliography of nonparametric statistics, Harvard Univ. Press.
- Таблицы**
- (1952) Tables of the binomial probability distribution, National Bureau of Standards, Washington.
- (1955) A million random digits with 100.000 normal deviates, Rand Corporation, Glencoe, Illinois.
- (1959) Tables of the bivariate normal distribution function, National Bureau of Standards, Washington.
- Топсё (Tópsoe F.)**
- (1970) Topology and measure, Berlin, Springer Verlag.
- Тьюки (Tukey J. W.)**
- (1957) Some examples with fiducial relevances, *Ann. Math. Stat.*, **28**, 687—695.
- Уилкс (Wilks S. S.)**
- (1967) Математическая статистика (перев. с англ.), М., «Наука».
- Уильямс (Williams E. J.)**
- (1959) Regression analysis, New York, John Wiley.
- Уоллес (Wallace D. L.)**
- (1958) Asymptotic approximations to distributions, *Ann. Math. Stat.*, **29**, 635—654.
- Уолш (Walsh J. E.)**
- (1962) Handbook of nonparametric statistics, investigations of randomness, moments, percentiles and distributions, Nostrand, London.
- Феллер (Feller W.)**
- (1945) On the normal approximation to the binomial distribution, *Ann. Math. Stat.*, **16**, 319—329.
- (1967) Введение в теорию вероятностей и ее приложения (перев. с англ.), Тома 1 и 2, М., «Мир».
- Фёдоров В. В.**
- (1969) Теория оптимальных экспериментов, вып. 7, Изд. МГУ.

- Фишер (Fisher R. A.)**
(1935) *The design of experiments*, Edinburg, Oliver and Boyd.
(1958) Статистические методы для исследователей (перев. с англ.), М., Госстатиздат.
- Фрэзер (Fraser D. A. S.)**
(1957) *Nonparametric methods in statistics*, New York, John Wiley.
- Халмос (Halmos P. R.)**
(1953) Теория меры (перев. с англ.), М., ИЛ.
- Халмос, Савидж (Halmos P. R., Savage L. J.)**
(1949) Application of the Radon — Nikodym theory of sufficient statistics, *Ann. Math. Stat.*, 20, 225—241.
- Хальд (Hald A.)**
(1956) Математическая статистика с техническими приложениями (перев. с англ.), М., ИЛ.
- Хёффдинг (Hoeffding W.)**
(1968) Some recent developments in nonparametric statistics, *Rev. of the Inter. Stat. Institute*, 36, № 2, 176—183.
- Хойл (Hoyle M. N.)**
(1973) Transformations — An Introduction and a Bibliography, *Int. Stat. Rev.*, 41, № 2, 203—223.
- Чжоу, Роббинс, Зигмунд (Chow Y. S., Robbins H., Siegmund D.)**
(1971) Great expectations: The theory of optimal stopping. Boston, Houghton Mifflin Co.
- Чернов, Леман (Chernoff H., Lehmann E. L.)**
(1954) The use of maximum likelihood estimates in χ^2 tests for goodness-of-fit. *Ann. Math. Stat.*, 25, 579.
- Чернов, Моzes (Chernoff H., Moses L. E.)**
(1962) Элементарная теория статистических решений (перев. с англ.), М., «Сов. радио».
- Чибисов Д. М.**
(1961) Об асимптотической мощности и эффективности критерия ϕ^2 , *ДАН СССР*, 138, № 2, 322—325.
(1971) Некоторые критерии типа хи-квадрат для непрерывных распределений. *Теория вероятн. и ее примен.*, 16, № 1, 3—20.
- Шеффе (Scheffé H.)**
(1963) Дисперсионный анализ (перев. с англ.), М., Физматгиз.
- Ширяев А. Н.**
(1961) Статистический последовательный анализ, гл. 4, М., «Наука».
- Юл и Кендал (Yule S. U., Kendall M. G.)**
(1960) Теория статистики (перев. с англ.), М., Госстатиздат.

О ГЛАВЛЕНИЕ

Предисловие ко второму русскому изданию	5
Предисловие к первому русскому изданию	6
Из предисловия автора	9

ЧАСТЬ ПЕРВАЯ

МАТЕМАТИЧЕСКОЕ ВВЕДЕНИЕ

Главы 1—3

Точечные множества

Глава 1. Общие свойства множеств	13
1. Множества. — 2. Подмножества. Пространство. — 3. Операции над множествами. — 4. Последовательности множеств. — 5. Монотонные последовательности. — 6. Аддитивные классы множеств.	
Глава 2. Линейные точечные множества	21
1. Интервалы. — 2. Некоторые свойства множеств в R_1 . — 3. Борелевские множества.	
Глава 3. Точечные множества в пространстве R_n	25
1. Интервалы. — 2. Некоторые свойства множеств из R_n . — 3. Борелевские множества. — 4. Линейные множества. — 5. Подпространства. Произведение пространств.	
Литература к главам 1—3	29

Главы 4—7

Теория меры и интегрирования в R_1

Глава 4. Мера Лебега линейных точечных множеств	30
1. Длина интервала. — 2. Обобщение. — 3. Мера суммы интервалов. — 4. Внешняя и внутренняя меры ограниченного множества. — 5. Измеримые множества и лебегова мера. — 6. Класс измеримых множеств. — 7. Измеримые множества и борелевские множества.	
Глава 5. Интеграл Лебега от функций одной переменной	46
1. Интеграл от ограниченной функции по множеству конечной меры. — 2. В-измеримые функции. — 3. Свойства интеграла. — 4. Интеграл от неограниченной функции на множестве конечной меры. — 5. Интеграл на множестве бесконечной меры. — 6. Интеграл Лебега как аддитивная функция множества.	

Глава 6. Неотрицательные аддитивные функции множества в R_1	62
1. Обобщение меры Лебега и интеграла Лебега. — 2. Функции множества и функции точки. — 3. Построение функции множества. — 4. P -мера. — 5. Ограниченнные функции множества. — 6. Распределения. — 7. Последовательности распределений. — 8. Теорема сходимости.	
Глава 7. Интеграл Лебега — Стильтьеса от функции одного переменного	77
1. Интеграл от ограниченной функции по множеству конечной P -меры. — 2. Несограниченные функции и множества бесконечной P -меры. — 3. Интегралы Лебега — Стильтьеса с параметром. — 4. Интегралы Лебега — Стильтьеса относительно распределения. — 5. Интеграл Римана — Стильтьеса.	
Литература к главам 4—7	90

Главы 8—9

Теория меры и интегрирования в R_n

Глава 8. Мера Лебега и другие аддитивные функции множества в R_n	91
1. Мера Лебега в R_n . — 2. Неотрицательные аддитивные функции множества в R_n . — 3. Ограниченнные функции множества. — 4. Распределения. — 5. Последовательности распределений. — 6. Распределения в произведении пространств.	
Глава 9. Интеграл Лебега — Стильтьеса от функций n переменных	100
1. Интеграл Лебега — Стильтьеса. — 2. Интегралы Лебега — Стильтьеса относительно распределения. — 3. Теорема о повторном интегрировании. — 4. Интеграл Римана — Стильтьеса. — 5. Неравенство Шварца.	

Главы 10—12

Различные вопросы

Глава 10. Интегралы Фурье	105
1. Характеристическая функция распределения в R_1 . — 2. Некоторые вспомогательные функции. — 3. Теоремы единственности для характеристических функций в R_1 . — 4. Теорема непрерывности для характеристических функций в R_1 . — 5. Некоторые интегралы. — 6. Характеристическая функция распределения в R_n . — 7. Теорема непрерывности для характеристических функций в R_n .	
Глава 11. Матрицы, детерминанты и квадратичные формы	120
1. Матрицы. — 2. Векторы. — 3. Матричное обозначение линейных преобразований. — 4. Матричное обозначение для биллинейных и	

квадратичных форм. — 5. Детерминанты. — 6. Ранг. — 7. Присоединенная и обратная матрицы. — 8. Линейные уравнения. — 9. Ортогональные матрицы. Характеристические числа. — 10. Неотрицательные квадратичные формы. — 11. Разложение формы $\sum_1^6 x_i^2$. — 12. Некоторые интегральные формулы.

Глава 12. Различные дополнения	140
1. Символы O , o и ∞ . — 2. Формула Эйлера — Маклорена. — 3. Гамма-функция. — 4. Бета-функция. — 5. Формула Стирлинга. — 6. Ортогональные полиномы.	

ЧАСТЬ ВТОРАЯ

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Главы 13—14

Основания

Глава 13. Статистика и вероятность	157
1. Случайные эксперименты. — 2. Примеры. — 3. Статистическая устойчивость. — 4. Объект математической теории. — 5. Математическое понятие вероятности.	
Глава 14. Основные определения и аксиомы	173
1. Случайные величины (аксиомы 1—2). — 2. Составные величины (аксиома 3). — 3. Условные распределения. — 4. Независимые величины. — 5. Функции случайных величин. — 6. Заключение.	

Главы 15—20

Случайные величины и распределения в R_1

Глава 15. Общие свойства	188
1. Функция распределения и плотность вероятности. — 2. Два простых типа распределений. — 3. Средние значения. — 4. Моменты. — 5. Характеристики расположения. — 6. Характеристики рассеяния. — 7. Теорема Чебышева. — 8. Характеристики асимметрии и эксцесса. — 9. Характеристические функции. — 10. Семи-инварианты. — 11. Независимые величины. — 12. Сложение независимых случайных величин.	
Глава 16. Различные дискретные распределения	217
1. Функция $\varepsilon(x)$. — 2. Биномиальное распределение. — 3. Теорема Бернулли. — 4. Теорема Муавра. — 5. Распределение Пуассона. — 6. Обобщенное биномиальное распределение Пуассона.	

Глава 17. Нормальное распределение	233
1. Нормальные функции. — 2. Нормальное распределение. — 3. Сложение независимых нормальных величин. — 4. Центральная предельная теорема. — 5. Дополнительные замечания к центральной предельной теореме. — 6. Ортогональное разложение, основанное на нормальном распределении. — 7. Асимптотическое разложение, основанное на нормальном распределении. — 8. Роль нормального распределения в статистике.	
Глава 18. Различные распределения, связанные с нормальным распределением	258
1. χ^2 -распределение. — 2. Распределение Стьюдента. — 3. z -распределение Фишера. — 4. Бета-распределение.	
Глава 19. Другие непрерывные распределения	270
1. Прямоугольное распределение. — 2. Распределения Коши и Лапласа. — 3. Усеченные распределения. — 4. Система Пирсона.	
Глава 20. Некоторые теоремы о сходимости	276
1. Сходимость распределений и случайных величин. — 2. Сходимость некоторых распределений к нормальному. — 3. Сходимость по вероятности. — 4. Теорема Чебышева. — 5. Теорема Хинчина. — 6. Теорема о сходимости.	
Упражнения к главам 15—20	283

Главы 21—24

Случайные величины и распределения в R_n

Глава 21. Случай двух измерений	288
1. Два простых типа распределений. — 2. Средние значения, моменты. — 3. Характеристические функции. — 4. Условные распределения. — 5. Регрессия, I. — 6. Регрессия, II. — 7. Коэффициент корреляции. — 8. Линейное преобразование случайных величин. — 9. Корреляционное отношение и средняя квадратическая связность. — 10. Эллипс рассеяния. — 11. Сложение независимых случайных величин. — 12. Нормальное распределение.	
Глава 22. Общие свойства распределений в R_n	321
1. Два простых типа распределений. Условные распределения. — 2. Замена переменных в непрерывном распределении. — 3. Средние значения, моменты. — 4. Характеристические функции. — 5. Ранг распределения. — 6. Линейное преобразование величин. — 7. Эллипсоид рассеяния.	

Глава 23. Регрессия и корреляция в R_n 332

1. Поверхности регрессии. — 2. Линейная средняя квадратическая регрессия. — 3. Остатки, остаточная дисперсия. — 4. Частная корреляция. — 5. Сводный коэффициент корреляции. — 6. Ортогональная средняя квадратическая регрессия.

Глава 24. Нормальное распределение 341

1. Характеристическая функция. — 2. Собственное нормальное распределение. — 3. Несобственное нормальное распределение. — 4. Линейное преобразование нормально распределенных величин. — 5. Распределение суммы квадратов. — 6. Условные распределения. — 7. Сложение независимых величин. Центральная предельная теорема.

Упражнения к главам 21—24 348

ЧАСТЬ ТРЕТЬЯ

СТАТИСТИЧЕСКИЕ ВЫВОДЫ

Главы 25—26

Общие понятия

Глава 25. Предварительные понятия, относящиеся к выбору 355

1. Вводные замечания. — 2. Простой случайный выбор. — 3. Распределение выборки. — 4. Выборочные значения как случайные величины. Выборочные распределения. — 5. Статистический аналог для распределения. — 6. Пристрастный выбор. Таблицы случайных чисел. — 7. Выбор без возвращения. Метод репрезентативной выборки.

Глава 26. Статистические выводы 365

1. Вводные замечания. — 2. Согласованность теории с данными опыта. Критерий значимости. — 3. Описание. — 4. Анализ. — 5. Предсказание.

Главы 27—29

Выборочные распределения

Глава 27. Характеристики выборочных распределений 375

1. Обозначения. — 2. Выборочное среднее значение \bar{x} . — 3. Моменты a_r . — 4. Дисперсия m_2 . — 5. Старшие центральные моменты и семи-инварианты. — 6. Несмещенные оценки. — 7. Функции от моментов. — 8. Характеристики многомерных распределений. — 9. Правки к группировке.

Глава 28. Асимптотические свойства выборочных распределений 398

1. Вводные замечания. — 2. Моменты. — 3. Центральные моменты. —
4. Функции от моментов. — 5. Квантили. — 6. Экстремальные зна-
чения и ширма выборки.

Глава 29. Точные выборочные распределения 413

1. Постановка проблемы. — 2. Лемма Фишера. Степени свободы. —
3. Совместное распределение величин \bar{x} и s^2 в выборках из кор-
мального распределения. — 4. Студентово отношение. — 5. Лем-
ма. — 6. Выбор из двумерного нормального распределения. —
7. Коэффициент корреляции. — 8. Коэффициенты регрессии. — 9. Вы-
бор из k -мерного нормального распределения. — 10. Обобщенная
дисперсия. — 11. Обобщенное студентово отношение. — 12. Коэф-
фициенты регрессии. — 13. Частные и сводные коэффициенты корре-
ляции.

Главы 30—31

Критерии значимости, I

Глава 30. Критерии согласия и аналогичные критерии 453

1. Критерий χ^2 в случае полностью определенного гипотетиче-
ского распределения. — 2. Примеры. — 3. Критерий χ^2 в слу-
чае, когда по выборке оцениваются некоторые параметры. — 4. При-
меры. — 5. Таблицы сопряженности признаков. — 6. χ^2 как критерий
однородности. — 7. Критерий для процента смертности. — 8. Даль-
нейшие критерии согласия.

Глава 31. Критерии значимости для параметров 489

1. Критерии, основанные на стандартных ошибках. — 2. Критерии,
основанные на точных распределениях. — 3. Примеры.

Главы 32—34

Теория оценок

Глава 32. Классификация оценок 513

1. Постановка проблемы. — 2. Две леммы. — 3. Минимум дисперсии
оценки. Эффективные оценки. — 4. Достаточные оценки. — 5. Асимп-
тотически-эффективные оценки. — 6. Случай двух неизвестных па-
раметров. — 7. Случай нескольких неизвестных параметров — 8. Обоб-
щение.

Глава 33. Методы нахождения оценок 540

1. Метод моментов. — 2. Метод максимума правдоподобия. —

3. Асимптотические свойства оценок максимального правдоподобия. —	
4. Метод минимума χ^2 .	
Глава 34. Доверительные области	550
1. Вводные замечания. — 2. Единственный неизвестный параметр. —	
3. Общий случай. — 4. Примеры.	
<i>Главы 35—37</i>	
Критерии значимости, II	
Глава 35. Общая теория проверки статистических гипотез	571
1. Выбор критерия значимости. — 2. Простые и сложные гипотезы. — 3. Критерий для простых гипотез. Наиболее мощные критерии. — 4. Несмешанные критерии. — 5. Критерии для сложных гипотез.	
Глава 36. Дисперсионный анализ	584
1. Изменчивость средних значений. — 2. Простая группировка величин. — 3. Обобщение. — 4. Случайные блоки. — 5. Латинские квадраты.	
Глава 37. Некоторые проблемы регрессии	597
1. Проблемы, содержащие неслучайные величины. — 2. Простая регрессия. — 3. Множественная регрессия. — 4. Дальнейшие проблемы регрессии.	
Таблицы I — II. Нормальное распределение	608—609
Таблица III. Распределение χ^2	610
Таблица IV. t-распределение	611
Цитированная литература	612
Предметный указатель	621
Дополнение ко второму изданию	625