

Б.Л. ВАН ДЕР ВАРДЕН

**МАТЕМАТИЧЕСКАЯ
СТАТИСТИКА**

И * Л

*Издательство
иностранной
литературы*

*

MATHEMATISCHE
STATISTIK

VON

Dr. B. L. VAN DER WAERDEN
Professor der Mathematik an der Universität
Zürich

SPRINGER — VERLAG
BERLIN - GÖTTINGEN - HEIDELBERG
1957

Б. Л. ван дер Варден

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Перевод с немецкого
Л. Н. БОЛЬШЕВА

Под редакцией
Н. В. СМЕРНОВА

ИЗДАТЕЛЬСТВО
ИНОСТРАННОЙ ЛИТЕРАТУРЫ
Москва, 1960

АННОТАЦИЯ

Эта книга знакомит читателя с основами математической статистики. Ее автору, известному ученому ван дер Вардену, удалось, не поступаясь математической строгостью, построить свое изложение таким образом, что для чтения книги не требуется знакомства ни с какими специальными разделами математики. Многочисленные примеры, иллюстрирующие применение математической статистики к разного рода научным и практическим задачам, представляют значительный интерес.

Книга принесет большую пользу как специалистам-прикладникам, использующим в своей работе методы математической статистики, так и научным работникам, аспирантам и студентам, специализирующимся в этой области.

ПРЕДИСЛОВИЕ К РУССКОМУ ПЕРЕВОДУ

По мере того как из года в год расширяется область приложения математической статистики, заметно возрастает спрос на различного типа руководства по этой дисциплине. Большинство из руководств ориентируется в первую очередь на потребности имеющих тех или иных специалистов-прикладников (экономистов, агрономов, биологов, медиков, физиков, инженеров-производственников и т. д.). В книгах такого типа статистические методы применяются в тех или иных конкретных случаях, а задача математического обоснования этих методов остается в стороне. Эта последняя задача рассматривается в сравнительно небольшом числе книг, рассчитанных главным образом на лиц, специализирующихся в области теории вероятностей и математической статистики (научных работников, студентов и аспирантов математических факультетов и т. п.).

Эти книги по своему объему, по детальности изложения и по уровню применяемого математического аппарата трудно доступны научным работникам и инженерам, занимающимся приложениями статистики. Между тем широкое проникновение математической статистики в различные науки и отрасли техники требует руководства особого типа, в котором рассматривались бы *общие постановки задач статистики* в форме, доступной для специалистов.

К такого рода книгам относится и выпускаемая в русском переводе «Математическая статистика» известного ученого Б. Л. ван дер Вардена (автора курса «Современная алгебра», получившего мировую известность).

Рассматривая свою книгу лишь как введение в современные статистические теории, автор ограничивает свою задачу изложением вполне сложившихся отделов математической статистики, оставляя в стороне менее разработанные, но, пожалуй, наиболее характерные для современного этапа развития науки проблемы последовательного анализа, теории решающих функций и статистику случайных процессов. В сжатом, но мастерском изложении автор наряду с теоретико-вероятностными основами статистики знакомит читателя с теорией статистической оценки параметров и проверки гипотез, простейшими схемами дисперсион-

ного анализа и теории корреляции. В трактовке этих классических задач и методов автору удалось найти во многих случаях новый нетрадиционный подход, значительно облегчающий их понимание и усвоение. Как интересные новинки следует отметить теорию несмещенных оценок (гл. VIII), развитую у нас в работах А. Н. Колмогорова и примененную им к некоторым практически важным задачам приемочного контроля; теорию некоторых новых непараметрических тестов (критерий Вилкоксона, X-тест и др.) (гл. XII), сопровождаемую поучительным исследованием их мощности, и, наконец, теорию ранговых коэффициентов корреляции (гл. XIII).

В книгу вошло много примеров из практики, частью составивших предмет консультаций, которые давались автором специалистам-прикладникам, частью заимствованных из специальных работ, главным образом по генетике, агрономии, медицине. Бегло и несколько поверхностно рассмотрены простейшие приложения математической статистики к изучению экономических явлений.

Переводчиком Л. Н. Большевым устранены некоторые неточности немецкого текста и в ряде случаев введены пояснения к примерам, слишком лаконично сформулированным автором.

Н. В. Смирнов

ПРЕДИСЛОВИЕ

Эта книга является результатом многолетней деятельности по практическому применению статистики. Со студенческих лет мне постоянно приходится иметь дело с экономистами, медиками, философами, биологами и инженерами, обращающимися ко мне со статистическими вопросами. Изучая литературу и размышляя, я знакомился все с лучшими и лучшими методами. Данная книга посвящается обоснованию этих методов и их применению к возможно наиболее поучительным примерам из области естествознания и социологии. Это позволит, как я надеюсь, уберечь читателя от тех или иных ложных путей, которыми я сам шел вначале. Примеры не являются искусственной конструкцией над теорией, а заимствованы из практики, поэтому некоторым примерам необходимо было предпослать обстоятельное объяснение.

Основные математические понятия я излагаю по возможности кратко и, смею надеяться, понятно. В некоторых случаях оказываются необходимыми длинные теоретические выводы, однако там, где это представляется возможным, более трудные доказательства будут заменяться ссылками на существующие хорошие учебники. Не имеет смысла вновь развивать математические теории, обстоятельно и ясно изложенные Колмогоровым, Каратеодори и Крамером.

Предполагается, что читателю известны элементы теории функций и теории интеграла Лебега. Это, конечно, не означает, что без такой подготовки читатель не сможет понять эту книгу: он должен будет лишь некоторые теоремы принять без доказательства или ограничиться изучением тех более элементарных разделов, в которых предполагаются известными только дифференциальное и интегральное исчисления и аналитическая геометрия (гл. I—IV и X—XII).

Эту книгу следует рассматривать лишь как введение в математическую статистику, не претендующее на полноту изложения. Многие важные разделы теории, такие, как последовательный анализ, теория решающих функций и вероятностные процессы, автор вынужден был целиком опустить. Этим теориям посвящены специальные труды выдающихся специалистов, например Wald A., *Sequential analysis* (Wiley), New York, 1947;

Wald A., *Statistical decision functions* (Wiley), New York, 1950;
Doob J. L., *Stochastic processes* (Wiley), New York, 1953¹.

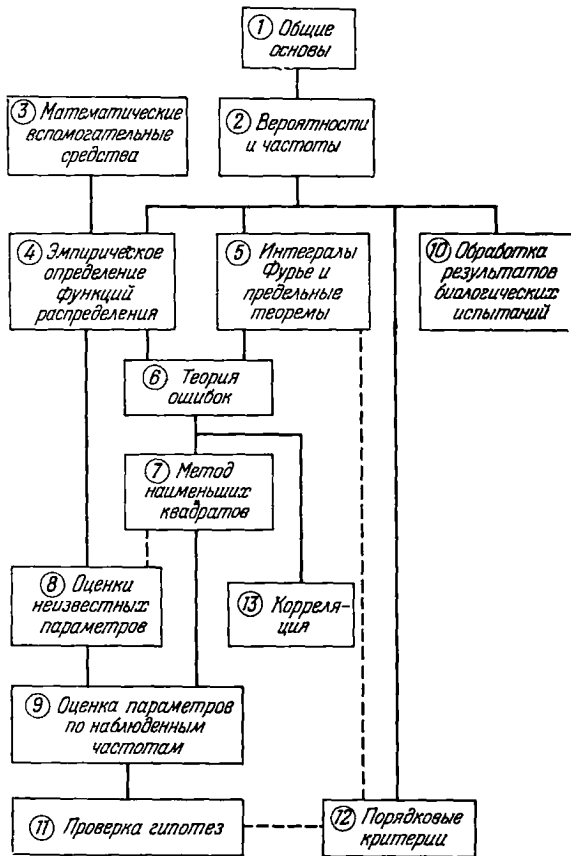
Дальнейшая литература указана во многих местах книги, причем ссылки делаются там, где это представляется удобным или полезным — непосредственно в тексте или в подстрочном примечании. Новая мода на литературные ссылки в конце главы или даже в конце книги неизбежно влечет за собой пугающую необходимость постоянного перелистывания страниц. Я не стремился к единообразию литературных ссылок и избегал чрезмерных сокращений.

Первая редакция этой книги возникла в 1945 г. Она служила пособием для курса теории ошибок и статистики в лаборатории нефтяной компании «Роял дач Шелл» в Амстердаме. Последующая редакция была критически прочитана доктором Е. Бачелетом (Базель). Ему, а также профессору Е. Л. Леману (Беркли) я очень благодарен за чрезвычайно ценные замечания. Я благодарен также г-ну Х. Р. Фишеру и г-ну Е. Нивергельту (Цюрих) за выполнение рисунков и участие в правке корректур.

Сентябрь 1956.

Б. Л. ван дер Варден

¹ Имеется русский перевод: Д у б Дж. Л., *Вероятностные процессы*, ИЛ, М., 1956. — *Прим. перев.*



Логическая структура книги.

Штриховые линии указывают, что для понимания соответствующих глав изучение предшествующих глав件лезно, но не является необходимым.

ВВЕДЕНИЕ

В старых трудах о количественных закономерностях в массовых совокупностях понятия частоты, среднего значения, дисперсии и т. д. развивались лишь для конечной совокупности. О понятии бесконечной совокупности не возникало и мысли. Английские и американские статистики, напротив, каждую статистическую совокупность воспринимают принципиально как случайную выборку из неограниченной совокупности возможностей. Частота события в этом понимании является лишь оценкой для *вероятности* события, а эмпирическое среднее (выборочное среднее) — оценкой для идеального *среднего значения*, или *математического ожидания*. С этой точки зрения узловым вопросом математической статистики является вопрос: *как далеко могут отклоняться величины, вычисленные по выборке, от соответствующих идеальных значений?*

Так сложилось к настоящему времени убеждение в необходимости построения математической статистики на основе теории вероятностей.

Теория вероятностей как точная математическая теория в надлежащем объеме впервые была построена А.Н. Колмогоровым. В дальнейшем мы будем опираться на аксиоматику Колмогорова, не заботясь о происхождении понятия вероятности. Эта теория, построенная чисто математически, оправдывается в приложениях так же хорошо, как Эвклидова геометрия или Ньютонова механика; этого вполне достаточно. Философское объяснение понятия вероятности, несомненно, интересно и важно, однако оно едва ли уместно в таком учебнике, как этот.

Логическая структура этой книги изображена в приведенной выше схеме.

Главы I—VI посвящены аксиоматике теории вероятностей по Колмогорову и ее различным статистическим приложениям, в том числе теории доверительных границ для неизвестной вероятности и доверительной зоны для неизвестной функции распределения, различным простым вариантам критерия χ^2 , гауссовой теории ошибок и критерию Стьюдента. В главах I, III и V излагаются вспомогательные математические средства; гл. II, IV и VI посвящены статистическим приложениям.

Центральную часть книги образуют два больших, взаимно связанных раздела: Теория оценок (гл. VII—IX) и Теория проверки гипотез (гл. XI, XII).

Отправным пунктом теории оценок является способ наименьших квадратов, развитый Гауссом. Гаусс указал два обоснования этого способа. Первое из них формулируется так: наимвероятнейшим значением неизвестного параметра является такое значение, для которого вероятность наблюдаемого события будет наибольшей. Второе обоснование, которому сам Гаусс отдавал предпочтение, исходит из того требования, что оценки должны иметь возможно меньшие средние ошибки.

Р. А. Фишер распространил оба этих обоснования на значительно более общие проблемы отыскания оценок. Требование наибольшей вероятности наблюдаемых значений приводит к оценкам наибольшего правдоподобия, а требование наименьшей средней ошибки — к понятию эффективных оценок. В широком классе случаев принцип максимального правдоподобия приводит действительно к эффективным оценкам. Уточнение этого понятия и точные доказательства по Фреше, Рао, Леману и Шеффе будут даны в гл. VIII, а применения к наблюдаемым частотам — в гл. IX.

Современная теория проверки гипотез берет свое начало от критерия χ^2 Пирсона и критерия t Стьюдента. Р. А. Фишер расширил область применения этих методов. Он ввел понятие «степеней свободы» и вскрыл взаимосвязь с теорией оценок, указав на то, что в критерий χ^2 следует пользоваться только эффективными оценками. Точные доказательства его утверждений даны Нейманом и Е. Пирсоном. Ими сформулированы также общие принципы, лежащие в основе современной теории критериев. Все это будет изложено и пояснено на примерах в гл. XI.

В теории ранговых критериев (гл. XII) также выявляется ценность этих принципов. Однако математические вспомогательные средства, необходимые для понимания этой главы, здесь много скромнее: они главным образом ограничиваются материалом гл. I и II и предельной теоремой из гл. V.

Обработке результатов биологических испытаний посвящена гл. X. И хотя здесь речь идет о задаче теории оценок в смысле гл. VIII, подготовительный материал также исчерпывается содержанием гл. I и II.

Заключительная гл. XIII посвящена коэффициентам корреляции и ранговой корреляции. Как видно из схемы логической структуры книги, здесь предполагается известным лишь содержание гл. I—VI.

ОБЩИЕ ОСНОВЫ

Изучение этой главы безусловно необходимо.

§ 1. Основные понятия теории вероятностей

А. ПРЕДВАРИТЕЛЬНОЕ РАЗЪЯСНЕНИЕ И ПРИМЕРЫ

В теории вероятностей рассматриваются *события*, наступление которых зависит от случая и *вероятности*, которые могут быть выражены числами.

Понятие вероятности является *статистическим* понятием. Для выяснения статистического смысла этого понятия следует многократно реализовать те условия, при которых может осуществляться определенное событие, и установить, с какой *частотой* это событие осуществляется. Если вероятность осуществления события равна p , то это означает, что в ряду из n таких повторений событие наступает в среднем np раз. Конечно, число наступлений события совершает колебания около среднего значения np , которые мы позже оценим точнее.

События обозначаются большими буквами A, B, \dots . Введем следующие обозначения.

AB (читается: A и B) — событие, которое осуществляется тогда и только тогда, когда A и B осуществляются одновременно.

\bar{A} (читается: отрицание A) — событие, которое осуществляется тогда и только тогда, когда не происходит событие A .

E — событие, которое осуществляется всегда.

$A \dot{+} B$ (читается: A или B) — событие, которое осуществляется тогда и только тогда, когда осуществляется либо A , либо B , либо оба эти события одновременно.

Если A и B несовместны, т. е. не могут осуществляться одновременно, то вместо $A \dot{+} B$ пишут $A + B$ (читается опять: A или B). Аналогично в случае любого конечного, а также и бесконечного числа попарно несовместных событий

$$\sum_1^n A_i = A_1 + \dots + A_n,$$

$$\sum_1^\infty A_i = A_1 + A_2 + \dots$$

Вероятность события A обозначается $P(A)$. Следующие примеры поясняют употребление этих понятий.

Пример 1. Игральная кость бросается три раза. Событиями являются все мыслимые результаты таких бросаний, как, например, $(6, 1, 1)$, или все комбинации таких результатов, связанные словом «или»: например, « $(6, 1, 1)$ или $(4, 5, 6)$ » является событием, а именно суммой событий $(6, 1, 1)$ и $(4, 5, 6)$. Вероятность выпадения 6 очков при первом бросании не обязательно равна $\frac{1}{6}$: кость может оказаться фальшивой и иметь случайные неправильности. Если она приближенно симметрична и однородна, то разумно предположить, что вероятность близка к $\frac{1}{6}$. В противном случае эту вероятность можно приближенно определить лишь с помощью большого числа бросаний кости, установив, сколько часто при этом выпадают 6 очков.

Пример 2. Производится стрельба по мишени. Идеализируя действительность, предположим, что пробонна в мишени является точкой и что мишень всегда поражается. Событием является попадание в какую-нибудь ограниченную часть мишени. Каждой частичной области мишени соответствует, таким образом, событие, в частности, всей мишени — событие E . Вероятность любого такого события тем больше, чем больше площадь соответствующей области, а также чем ближе к середине лежит эта область: целятся ведь в середину мишени. Попадание в отдельную точку также является событием, но вероятность этого события равна нулю, так как точка не имеет площади. Если события A и B соответствуют определенным областям мишени, то сумма $A + B$ соответствует объединению обеих областей, а произведение AB — их пересечению.

В. СОБЫТИЯ

Если желательно математически уточнить употребление формальных операций AB , \bar{A} , $A \dot{+} B$ и $A + B$, то имеются два пути: поле событий можно рассматривать либо как булевскую алгебру, либо как тело множеств.

В первом понимании «события» являются неопределяемыми объектами, и указанные операции должны лишь удовлетворять определенным аксиомам (см. С. Carathéodory, *Maß und Integral*, Basel, 1956).

Во втором понимании «события» являются подмножествами множества E , причем AB является пересечением, \bar{A} — дополнением, $A \dot{+} B$ — объединением.

Оба подхода эквивалентны, так как по известной теореме Стоуна¹ каждая булевская алгебра изоморфна некоторому телу множеств. Первый подход, может быть, естественнее (см. Карпос D. A., *Zur mathematischen Begründung der Wahrscheinlichkeitstheorie*, Sitzungsber. Bayer. Akad. München, 1948), но второй математически проще. Поэтому, следуя Колмогорову², мы будем все «события» трактовать как множества «элементарных событий».

¹ См. Stone M. H., *Trans. Amer. math. Soc.*, 40 (1936), 37, или Hemes H., *Einführung in die Verbandstheorie*, Springer-Verlag, 1955, § 20.

² См. Колмогоров А. Н., *Основные понятия теории вероятностей*, ОНТИ, М., 1936, а также Reichenbach H., *Wahrscheinlichkeitslehre*.

В этом понимании E является множеством всех элементарных событий, которые в данной ситуации можно рассматривать как возможные события.

В. ВЕРОЯТНОСТЬ

В основу теории вероятностей по Колмогорову можно положить следующие аксиомы¹:

1. Если A и B — события, то \bar{A} , AB и $A \dot{+} B$ тоже события.
2. Каждому событию A можно поставить в соответствие действительное число $P(A) \geq 0$.
3. E является событием с $P(E) = 1$.
4. Если A и B несовместны, то

$$P(A + B) = P(A) + P(B).$$

5. Если все события последовательности A_1, A_2, \dots не могут осуществиться одновременно, то

$$\lim_{n \rightarrow \infty} P(A_1 A_2 \dots A_n) = 0.$$

Из аксиом 3 и 4 следует, что

$$P(\bar{A}) = 1 - P(A) \quad (1)$$

и поэтому $P(A)$ не превосходит единицы:

$$0 \leq P(A) \leq 1. \quad (2)$$

Далее, если A_1, \dots, A_n взаимно несовместны, то имеет место теорема сложения:

$$P(A_1 + \dots + A_n) = P(A_1) + \dots + P(A_n). \quad (3)$$

Из аксиомы непрерывности 5 следует, что теорема сложения также справедлива для бесконечного количества событий; если

$$A = A_1 + A_2 + \dots$$

является событием, то

$$P(A) = P(A_1) + P(A_2) + \dots \quad (4)$$

Очень простые доказательства этих теорем имеются в цитированной монографии Колмогорова «Основные понятия теории вероятностей».

¹ Формулировка аксиом у автора несколько отличается от формулировки, данной в указанной выше работе А. Н. Колмогорова. — Прим. ред.

Г. УСЛОВНЫЕ ВЕРОЯТНОСТИ

Пусть $P(A) \neq 0$. Условная вероятность события B в предположении, что событие A произошло, определяется равенством

$$P(B|A) = \frac{P(AB)}{P(A)}. \quad (5)$$

Отсюда следует, что

$$P(AB) = P(B|A) P(A). \quad (6)$$

Последняя формула справедлива также и для $P(A) = 0$; множителем $P(B|A)$ в этом случае может быть любым числом.

В приложениях, как правильно заметил Рихтер¹, условная вероятность $P(B|A)$ почти никогда не вычисляется по определению (5), а делаются какие-либо предположения о $P(B|A)$, на основе которых по формуле (6) вычисляется $P(AB)$ (как в случае, изложенном в примере 3). Собственно говоря, следовало бы отказаться от определения условной вероятности $P(B|A)$ равенством (5) и включить ее в число неопределяемых понятий аксиоматики. В этом случае (6) можно бы было принять за аксиому.

Однако в этой книге мы не будем заниматься вопросами аксиоматики и, следуя Колмогорову, примем (5) в качестве определения.

Пример 3. Из урны, содержащей r белых и $N - r$ черных шаров, поочередно извлекаются 2 шара (без возвращения). Какова вероятность: а) при первом извлечении получить белый шар, б) при первом и втором извлечениях получить белые шары, в) при втором извлечении получить белый шар? При этом предполагается, что шары в урне хорошо перемешаны, так что вероятность извлечь определенный шар одинакова для всех шаров. Предполагается также, что условная вероятность извлечь во второй раз какой-либо шар, если при первом извлечении уже получен некоторый определенный шар, одинакова для всех оставшихся $N - 1$ шаров.

Обозначим A_j событие, которое заключается в том, что при первом извлечении появится шар с номером j , и B_k — событие, которое заключается в том, что при втором извлечении появится шар с номером k . Из указанных предположений следует, что

$$P(A_j) = \frac{1}{N}$$

и

$$P(B_k|A_j) = \frac{1}{N-1} \quad (j \neq k).$$

Согласно правилу умножения (6), вероятность появления при первом извлечении шара с номером j и при втором извлечении шара с номером k одинакова для всех пар (j, k) с $j \neq k$, а именно

$$P(A_j B_k) = \frac{1}{N} \cdot \frac{1}{N-1}.$$

¹ Richter H., Grundlegung der Wahrscheinlichkeitsrechnung, Math. Annalen, 125 и 126. См. также Finsler P., Elemente der Math., 2 (1947), 112.

Количество пар, у которых первый шар белый, равно $r(N-1)$. Следовательно, вероятность появления белого шара при первом извлечении равна

$$\frac{r(N-1)}{N(N-1)} = \frac{r}{N}.$$

Аналогично количество пар, у которых второй шар белый, равно $r(N-1)$, поэтому вероятность появления белого шара при втором извлечении равна r/N . Наконец, число пар (j, k) шаров белого цвета равно $r(r-1)$, следовательно, вероятность того, что оба раза появятся белые шары, равна

$$\frac{r(r-1)}{N(N-1)}.$$

Д. ФОРМУЛА ПОЛНОЙ ВЕРОЯТНОСТИ

Предположим, что при некотором опыте могут осуществиться взаимно исключающие друг друга события A_1, \dots, A_n , сумма которых является достоверным событием:

$$E = A_1 + A_2 + \dots + A_n.$$

Согласно (3) и (6), для любого такого разложения имеет место формула полной вероятности

$$P(B) = \sum_k P(A_k) P(B|A_k). \quad (7)$$

Е. НЕЗАВИСИМОСТЬ

Пара или более таких разложений

$$E = A_1 + A_2 + \dots + A_n, \quad E = B_1 + B_2 + \dots + B_m, \dots$$

называются *независимыми*, если для всех h, i, \dots, k справедливы равенства

$$P(A_h B_i \dots D_k) = P(A_h) P(B_i) \dots P(D_k). \quad (8)$$

События A, B, \dots, D в конечном числе называются *независимыми*, если независимы разложения $E = A + \bar{A}$, $E = B + \bar{B}, \dots$, $E = D + \bar{D}$.

В этом случае справедливы равенства

$$P(AB \dots D) = P(\bar{A}) P(B) \dots P(D),$$

$$P(\bar{A}B \dots D) = P(\bar{A}) P(B) \dots P(D) \text{ и т. д.}$$

В приложениях независимость обычно не определяется равенством (8), а постулируется. Два опыта полагают независимыми, если исход одного практически исключает какое-либо влияние на исход другого.

Ж. БЕСКОНЕЧНЫЕ СУММЫ

Бесконечная сумма $A_1 + A_2 + \dots$ несовместных событий не обязана являться событием и иметь вероятность. Однако методами лебеговской теории меры можно расширить тело «событий» до тела «измеримых множеств», определив, таким образом, для этих множеств A^* меру $P^*(A^*)$, что в расширенной области снова будут иметь место аксиомы 1—5 и для исходных событий A мера P^* будет совпадать с вероятностью P :

$$P^*(A) = P(A).$$

Если, кроме того, вероятность $P(A)$ зависит от неизвестного параметра ϑ , то рассматривают лишь такие множества A^* , которые измеримы при всех значениях ϑ .

Каждая сумма счетного числа множеств A^* является снова множеством типа A^* , причем имеет место *теорема о полной аддитивности*:

$$P^*(A_1^* + A_2^* + \dots) = P^*(A_1^*) + P^*(A_2^*) + \dots \quad (9)$$

Доказательство этой теоремы см., например, в книге С. Carathéodory, *Vorlesungen über reelle Funktionen* (1918), p. 237—258¹.

В дальнейшем, если это потребуется, мы будем предполагать указанное расширение тела событий выполненным, не отмечая звездочками различие между новыми множествами и их мерами, с одной стороны, и исходными событиями и вероятностями — с другой. Таким образом, впредь будет предполагаться, что сумма событий $A_1 + A_2 + \dots$ снова является событием и что для счетных сумм справедлива теорема сложения.

§ 2. Случайные величины. Функции распределения

А. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Случайной величиной, или *стохастической переменной*, называют величину, значение которой зависит от случая. Точнее: пусть x — действительная функция, определенная на множестве E , такая, что для каждого элементарного события ξ значение $x(\xi)$ является действительным числом. Пусть, далее, эта функция измерима в том смысле, что для любого действительного числа t множество тех ξ , для которых $x < t$, является измеримым множеством. Мы условились, однако, рассматривать каждое измеримое множество как событие (в расширенном смысле). Следовательно, предположение измеримости означает, что для любого t множество тех ξ , для которых $x < t$, представляет собой событие.

¹ О полной аддитивности меры см. также Халмош П., *Теория меры*, ИЛ, М., 1953. — *Прим. перев.*

Простейшим является случай, когда множество E представимо в виде конечной суммы частичных множеств

$$E = A_1 + A_2 + \dots + A_n,$$

причем на каждой части A_k функция \mathbf{x} принимает постоянное значение x_k . Если A_k являются событиями, то требование измеримости выполняется.

В примере 1 (§ 1) вообще число очков, выпавшее при трехкратном бросании игральной кости, является случайной величиной, принимающей конечное число значений (от 3 до 18). В примере 2 обе координаты \mathbf{x}, \mathbf{y} точки попадания — случайные величины.

Б. ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

Если \mathbf{x} — случайная величина, а t изменяется от $-\infty$ до $+\infty$, то вероятность события $\mathbf{x} < t$ представляет собой неубывающую, непрерывную слева функцию от t , которую мы, следуя Колмогорову¹, назовем *функцией распределения* $F(t)$ величины \mathbf{x} :

$$F(t) = \mathbf{P}(\mathbf{x} < t). \quad (1)$$

Функция $F(t)$ стремится к нулю при $t \rightarrow -\infty$ и к единице при $t \rightarrow +\infty$. Это легко выводится из аксиомы непрерывности 5.

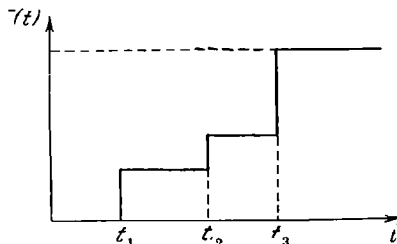


Рис. 1. График ступенчатой функции.

Если t стремится к a слева, то $F(t)$ стремится к $F(a)$, но если t стремится к a справа, то $F(t)$ стремится к $\mathbf{P}(\mathbf{x} \leq t)$. Разность этих пределов

$$\Delta F(a) = F(a + 0) - F(a - 0)$$

является вероятностью того, что \mathbf{x} в точности равна a . Далее для $a < b$ имеем

$$F(b) - F(a) = \mathbf{P}(a \leq \mathbf{x} < b). \quad (2)$$

¹ Другие авторы определяют $F(t)$ как вероятность события $\mathbf{x} \leq t$. В этом случае $F(t)$ непрерывна справа.

Для приложений особенно важны два случая. Если величина x принимает лишь конечное число значений t_1, t_2, \dots, t_n с вероятностями p_1, p_2, \dots, p_n , то $F(t)$ представляет собой *ступенчатую функцию*, график которой в точке $t = t_i$ имеет скачок, равный по величине p_i .

В. ПЛОТНОСТЬ ВЕРОЯТНОСТИ

В другом случае, когда $F(t)$ непрерывно дифференцируема,

$$F'(t) = f(t).$$

Согласно (2), тогда

$$P(a \leq x < b) = F(b) - F(a) = \int_a^b f(t) dt. \quad (3)$$

Из (3) предельным переходом при $a \rightarrow -\infty$ получаем

$$P(x < b) = F(b) = \int_{-\infty}^b f(t) dt.$$

откуда предельным переходом при $b \rightarrow \infty$ находим

$$\int_{-\infty}^{\infty} f(t) dt = 1. \quad (4)$$

Функция $f(t)$ называется *плотностью вероятности* величины x . Выражаясь популярно, но неточно, можно сказать, что $f(t)dt$ есть вероятность того, что x заключена в пределах между t и $t + dt$.

Г. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Известным примером плотности вероятности является гауссова *функция ошибок*

$$f(t) = \frac{1}{c\sqrt{2\pi}} e^{-\frac{(t-a)^2}{2c^2}}. \quad (5)$$

Рассматривая вместо x величину $(x - a)/c$, можно для этой величины получить плотность вероятности более простого вида

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}. \quad (6)$$

На рис. 2 сверху изображен график этой функции. В § 12 будет доказано, что

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2} dt = \sqrt{2\pi}, \quad (7)$$

поэтому функция (6) удовлетворяет условию (4). Соответствующая функция распределения имеет вид

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}\tau^2} d\tau \quad (8)$$

(см. рис. 2 снизу, а также табл. 1 в конце книги).

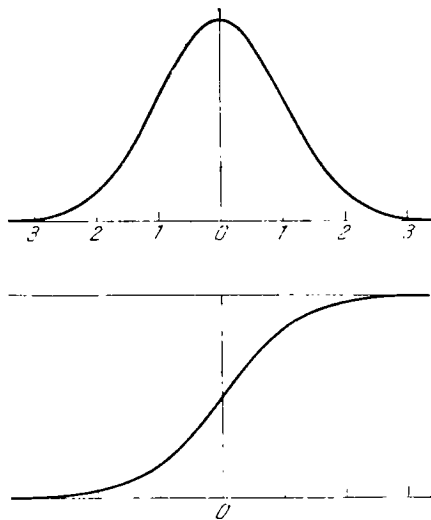


Рис. 2. График гауссовой функции ошибок.

Название «гауссова функция ошибок» основано на том, что, по Гауссу, плотность вероятности для случайных ошибок астрономических наблюдений выражается формулой (5). Однако имеется много других случайных величин, плотности вероятности которых точно или приближенно выражаются этой формулой. Поэтому функции (6) и (8) заслуживают более подробного изучения.

Для вычисления интеграла ошибок при не слишком больших значениях t плотность $f(t)$ разлагают в бесконечный ряд

$$f(t) = \frac{1}{\sqrt{2\pi}} \left(1 - \frac{t^2}{2} + \frac{t^4}{2! \cdot 2^2} - \frac{t^6}{3! \cdot 2^3} + \dots \right)$$

и интегрируют:

$$\Phi(t) = \frac{1}{2} + \int_0^t f(t) dt = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left(t - \frac{t^3}{2 \cdot 3} + \frac{t^5}{2! \cdot 2^2 \cdot 5} - \dots \right). \quad (9)$$

Для больших t имеется асимптотическое разложение, которое получается следующим образом. Имеем

$$1 - \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{1}{2}\tau^2} d\tau.$$

Если положить $\tau^2/2 = x$ и $t^2/2 = u$, то интеграл примет вид

$$\frac{1}{2\sqrt{\pi}} \int_u^\infty e^{-x} x^{-\frac{1}{2}} dx. \quad (10)$$

Интегрируя по частям, найдем

$$\int_u^\infty e^{-x} x^{-\frac{1}{2}} dx = e^{-u} u^{-\frac{1}{2}} - \frac{1}{2} \int_u^\infty e^{-x} x^{-\frac{3}{2}} dx = e^{-u} u^{-\frac{1}{2}} - R_1, \quad (11)$$

где $-R_1$ — отрицательный остаточный член. Для того чтобы оценить R_1 сверху, интегрируем по частям еще раз:

$$R_1 = \frac{1}{2} \int_u^\infty e^{-x} x^{-\frac{3}{2}} dx = \frac{1}{2} e^{-u} u^{-\frac{3}{2}} - \frac{1}{2} \cdot \frac{3}{2} \int_u^\infty e^{-x} x^{-\frac{5}{2}} dx < \frac{1}{2} e^{-u} u^{-\frac{3}{2}}.$$

Если (11) подставить в (10), то получится искомая асимптотическая формула

$$\Phi(t) = 1 - \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \left(\frac{1}{t} - S_1 \right), \quad (12)$$

$$0 < S_1 < \frac{1}{t^3}.$$

Если желательно сделать остаток величиной порядка t^{-5} , то нужно снова один раз проинтегрировать по частям и тогда получится

$$\Phi(t) = 1 - \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \left(\frac{1}{t} - \frac{1}{t^3} + S_2 \right), \quad (13)$$

$$0 < S_2 < 1 \cdot 3 \cdot \frac{1}{t^5}.$$

Путь дальнейших уточнений ясен. В качестве множителя при $\exp\left(-\frac{1}{2}t^2\right)$ будет получаться частная сумма знакочередующегося ряда

$$\frac{1}{t} - \frac{1}{t^3} + \frac{1 \cdot 3}{t^5} - \frac{1 \cdot 3 \cdot 5}{t^7} + \dots, \quad (14)$$

и остаточный член всегда будет меньше первого из отброшенных членов.

Асимптотический ряд (14) мало пригоден для численных расчетов со многими десятичными знаками, потому что, хотя вначале члены ряда и убывают, однако в дальнейшем они начинают возрастать. Однако ван Вейнгарден¹ указал преобразование, которое превращает (14) в сходящийся ряд, с успехом используемый в численных расчетах для средних и больших значений t .

В табл. 1 в конце книги табулирована функция $u = \Phi(t)$, в табл. 2 — обратная функция $t = \Psi(u)$. В этой книге указанные функции будут неизменно обозначаться буквами Φ и Ψ .

Точки перегиба кривой с уравнением $y = f(t)$ имеют абсциссы $t = \pm 1$, так как вторая производная

$$y'' = \frac{t^2 - 1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$$

в интервале $(-1, +1)$ отрицательна, а вне этого интервала неотрицательна. С ростом t функция $f(t)$ очень быстро убывает. На интервал от -2 до $+2$ приходится более 95% общей площади, расположенной между кривой и осью абсцисс, на интервал от -3 до $+3$ около 99,7%, а на интервал от -4 до $+4$ — только немногим менее 100%. Таким образом, для всех практических целей можно ограничиться интервалом от -4 до $+4$: вероятность того, что случайная величина $(x - a)/c$ примет значение вне этого интервала, практически равна нулю.

Все функции распределения, для которых плотности вероятности задаются формулой вида (5), называются *функциями нормального распределения*. Постоянная a указывает абсциссу точки максимума функции $f(t)$, а постоянная c — расстояние между проекциями точки максимума и точки перегиба на ось абсцисс. Таким образом, c является мерой «широты» кривой.

¹ Van Wijngaarden A., A transformation of formal series, Proc. Kon. Ned. Akad. Amsterdam, Section of Science, A 56 (1953), 537.

§ 3. Среднее значение и квадратичное отклонение

А. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

Среднее значение или математическое ожидание $\mathcal{E}x$ случайной величины x , принимающей лишь конечное число значений t_1, \dots, t_n , определяется как сумма этих значений, умноженных на их вероятности:

$$\hat{x} = \mathcal{E}x = \sum t_k p_k. \quad (1)$$

Из определения ясно, что среднее значение \hat{x} заключено между наименьшим и наибольшим возможными значениями случайной величины x . Среднее значение можно также определить как центр тяжести возможных значений t_k с весами p_k .

Если случайная величина x имеет плотность вероятности $f(t)$, то в определении среднего значения вместо суммы появляется интеграл:

$$\mathcal{E}x = \int_{-\infty}^{+\infty} t f(t) dt. \quad (2)$$

В случае произвольной функции распределения $F(t)$ интеграл (2) нужно заменить так называемым интегралом Стильтьеса

$$\mathcal{E}x = \int_{-\infty}^{+\infty} t dF(t). \quad (3)$$

Общее понятие интеграла Стильтьеса нам в дальнейшем не понадобится, так как мы вполне обойдемся специальными случаями (1) и (2); поэтому, точности ради, укажем лишь определение интеграла (3) по Фреше:

$$\mathcal{E}x = \lim_{h \rightarrow 0} \sum_{k=-\infty}^{+\infty} kh [F(kh + h) - F(kh)]. \quad (4)$$

Среднее значение является специальным случаем интеграла Лебега. А именно, если E_k представляет собой событие $kh \leq x < (k+1)h$, то интеграл Лебега от x в области B по мере $\mathbf{P}(A)$ определяется так:

$$\int_B x \mathbf{P}(dE) = \lim_{h \rightarrow 0} \sum_{k=-\infty}^{+\infty} kh \mathbf{P}(BE_k). \quad (5)$$

Если, в частности, положить $B = E$, то (5) перейдет в (4).

В теории интеграла Лебега доказывается, что сумма $x + y$ двух измеримых функций x и y также измерима и что интеграл

этой суммы по любому измеримому множеству B равен сумме интегралов слагаемых:

$$\int_B (\mathbf{x} + \mathbf{y}) \mathbf{P}(dE) = \int_B \mathbf{x} \mathbf{P}(dE) + \int_B \mathbf{y} \mathbf{P}(dE).$$

При этом предполагается, что оба интеграла в правой части равенства конечны. Доказательство можно найти в книге С. Carathéodory, «Vorlesungen über reelle Funktionen», или в новой книге того же автора «Maß und Integral».

Если, в частности, выбрать $B = \hat{E}$, то из последнего равенства будет следовать, что среднее значение суммы равно сумме средних значений слагаемых, если эти средние значения конечны:

$$\mathcal{E}(\mathbf{x} + \mathbf{y}) = \mathcal{E} \mathbf{x} + \mathcal{E} \mathbf{y} \quad (6)$$

и аналогично для произвольного числа слагаемых

$$\mathcal{E}(\mathbf{x}_1 + \dots + \mathbf{x}_n) = \mathcal{E} \mathbf{x}_1 + \dots + \mathcal{E} \mathbf{x}_n. \quad (7)$$

Если c — постоянная, то очевидно, что

$$\mathcal{E}(c\mathbf{x}) = c \mathcal{E} \mathbf{x}. \quad (8)$$

Для краткости мы иногда будем обозначать $\mathcal{E} \mathbf{x} = \hat{\mathbf{x}}$.

Б. НЕЗАВИСИМЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Пара или более случайных величин $\mathbf{x}, \mathbf{y}, \dots$ называются независимыми, если для любых действительных чисел t, u, \dots события

$$\mathbf{x} < t, \mathbf{y} < u, \dots$$

являются независимыми. Из независимости событий $\mathbf{x} < t, \mathbf{y} < u, \dots$ следует, что события

$$a \leq \mathbf{x} < b, \quad c \leq \mathbf{y} < d, \dots$$

также будут независимыми и поэтому, если речь идет, например, о двух независимых величинах \mathbf{x} и \mathbf{y} , справедливо равенство

$$\mathbf{P}(a \leq \mathbf{x} < b, c \leq \mathbf{y} < d) = \mathbf{P}(a \leq \mathbf{x} < b) \mathbf{P}(c \leq \mathbf{y} < d).$$

В этом случае среднее значение произведения равно произведению средних значений:

$$\mathcal{E}(\mathbf{x} \mathbf{y}) = \mathcal{E} \mathbf{x} \cdot \mathcal{E} \mathbf{y} \quad (9)$$

и аналогично для произвольного числа независимых случайных величин

$$\mathcal{E}(\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_n) = \hat{\mathbf{x}}_1 \hat{\mathbf{x}}_2 \dots \hat{\mathbf{x}}_n. \quad (10)$$

Доказательство см., например, в книге А. Н. Колмогорова «Основные понятия теории вероятностей», стр. 69, формула (6).

В КВАДРАТИЧНОЕ ОТКЛОНЕНИЕ И ДИСПЕРСИЯ

Квадратичное отклонение $\sigma = \sigma_{\mathbf{x}}$ случайной величины \mathbf{x} определяется как положительный квадратный корень из дисперсии

$$\sigma^2 = \mathfrak{E}(\mathbf{x} - \hat{\mathbf{x}})^2 = \mathfrak{E}(\mathbf{x}^2 - 2\hat{\mathbf{x}}\mathbf{x} + \hat{\mathbf{x}}^2). \quad (11)$$

Вычисляя правую часть (11) по формулам (7) и (8), получим

$$\sigma^2 = \mathfrak{E}(\mathbf{x}^2) - 2\hat{\mathbf{x}}^2 + \hat{\mathbf{x}}^2 = \mathfrak{E}(\mathbf{x}^2) - (\mathfrak{E}\mathbf{x})^2. \quad (12)$$

Если c — постоянная, то очевидно, что

$$\sigma_{c\mathbf{x}} = c\sigma_{\mathbf{x}}. \quad (13)$$

Для дисперсии суммы имеет место равенство

$$\begin{aligned} \sigma_{\mathbf{x}+\mathbf{y}}^2 &= \mathfrak{E}(\mathbf{x} - \hat{\mathbf{x}} + \mathbf{y} - \hat{\mathbf{y}})^2 = \\ &= \mathfrak{E}(\mathbf{x} - \hat{\mathbf{x}})^2 + 2\mathfrak{E}[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{y} - \hat{\mathbf{y}})] + \mathfrak{E}(\mathbf{y} - \hat{\mathbf{y}})^2. \end{aligned}$$

Если \mathbf{x} и \mathbf{y} независимы, то средний член обращается в нуль:

$$\mathfrak{E}[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{y} - \hat{\mathbf{y}})] = \mathfrak{E}(\mathbf{x} - \hat{\mathbf{x}})\mathfrak{E}(\mathbf{y} - \hat{\mathbf{y}}) = 0,$$

и мы получаем

$$\sigma_{\mathbf{x}+\mathbf{y}}^2 = \sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2. \quad (14)$$

Эта формула справедлива и для разности двух независимых случайных величин:

$$\sigma_{\mathbf{x}-\mathbf{y}}^2 = \sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2.$$

В случае произвольного количества попарно независимых величин дисперсия суммы $\mathbf{x} = \mathbf{x}_1 + \dots + \mathbf{x}_n$ равна сумме дисперсий слагаемых

$$\sigma^2 = \sigma_1^2 + \dots + \sigma_n^2. \quad (15)$$

Для вычисления средних значений и дисперсий полезна следующая формула, являющаяся обобщением формулы (2):

$$\mathfrak{E}g(\mathbf{x}) = \int_{-\infty}^{+\infty} g(t)f(t)dt. \quad (16)$$

В § 4 мы рассмотрим дальнейшее обобщение этой формулы и дадим ее доказательство. Теперь же мы рассмотрим пример применения этой формулы.

Пример 4. Случайная величина \mathbf{x} имеет гауссову плотность вероятности

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

Каковы среднее значение и квадратичное отклонение x^2 ? Согласно (2), среднее значение равно

$$\hat{x} = \int_{-\infty}^{+\infty} t f(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t e^{-\frac{1}{2} t^2} dt = 0.$$

Дисперсия σ^2 равна среднему значению случайной величины $(x - 0)^2 = x^2$. Если мы положим $g(t) = t^2$ и применим формулу (16), то получим

$$\sigma^2 = \mathcal{E}(x^2) = \int_{-\infty}^{+\infty} t^2 f(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{1}{2} t^2} dt.$$

Интегрируя по частям, найдем

$$\sigma^2 = \left[-\frac{1}{\sqrt{2\pi}} t e^{-\frac{1}{2} t^2} \right]_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} t^2} dt = 0 + 1 = 1,$$

следовательно, $\sigma = 1$. В общем случае

$$f(t) = \frac{1}{c\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(t-a)^2}{c^2}}.$$

Поступая так же, как и раньше, найдем, что $\mathcal{E}x = a$ и $\sigma = c$. Поэтому впредь мы всегда будем вместо c писать σ :

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(t-a)^2}{\sigma^2}}.$$

Г. НЕРАВЕНСТВО ЧЕБЫШЕВА

Практическое значение квадратичного отклонения σ заключается в том, что отклонения случайной величины x от ее среднего значения \hat{x} , намного превышающие квадратичное отклонение σ , являются маловероятными. Точный смысл этого утверждения формулируется с помощью *неравенства Чебышева*:

Если g — произвольное положительное действительное число, то вероятность события $|x - \hat{x}| > g\sigma$ меньше, чем g^{-2} .

Доказательство весьма просто и основано на определении математического ожидания величины $(x - \hat{x})^2$

$$\sigma^2 = \mathcal{E}(x - \hat{x})^2.$$

Для вычисления $\mathcal{E}(x - \hat{x})^2$, согласно определению (1), нужно возможные значения случайной величины

$$y = (x - \hat{x})^2$$

умножить на соответствующие вероятности и произведения сложить. Возможные значения распадутся на две группы. К пер-

вой группе принадлежат значения $\leq g^2\sigma^2$ и ко второй — значения $> g^2\sigma^2$. В соответствии с этим сумма произведений разобьется на две частичные суммы, первая из которых неотрицательна, так как неотрицательны все ее слагаемые. Вторая частичная сумма больше произведения $g^2\sigma^2 P$, где P — вероятность сбывания $(x - \hat{x})^2 > g^2\sigma^2$. Предполагается, что P не равно нулю (если $P = 0$, то все становится тривиальным). Таким образом, для всей суммы получаем неравенство

$$\sigma^2 > g^2\sigma^2 P,$$

откуда следует утверждение:

$$P < \frac{1}{g^2}.$$

Точно так же проводится доказательство и для случая, когда математическое ожидание определяется интегралом (3). Пусть $F(t)$ — функция распределения случайной величины y . Интеграл

$$\sigma^2 = \xi y = \int_{-\infty}^{\infty} t dF(t).$$

распадается на сумму двух интегралов, из которых первый берется по области $t \leq g^2\sigma^2$, а второй — по области $t > g^2\sigma^2$. Первое слагаемое этой суммы неотрицательно, а второе $> g^2\sigma^2 P$. Таким образом, $\sigma^2 > g^2\sigma^2 P$ и, следовательно, $P < g^{-2}$.

В большинстве случаев вероятность P оказывается значительно меньше, чем g^{-2} . Например, для нормального распределения вероятность события $|x - a| > 3\sigma$ равна 0,0027, т. е. она много меньше, чем $1/9$.

Частным случаем неравенства Чебышева для случая $\sigma = 0$ является следующее утверждение:

Если квадратичное отклонение равно нулю, то вероятность того, что x будет отличаться от постоянного значения \hat{x} , также равна нулю.

§ 4. Интегральные представления средних значений и вероятностей

А. Прямоугольники и открытые множества

Под *прямоугольником* в плоскости uOv мы понимаем множество точек, координаты которых удовлетворяют неравенствам

$$a \leq u < b, \quad c \leq v < d.$$

Любое открытое множество M в плоскости uOv представимо в виде суммы счетного числа таких прямоугольников. А именно если эту плоскость разбить на прямоугольники с помощью верти-

кальных и горизонтальных равноотстоящих прямых и выделить те прямоугольники, которые целиком принадлежат множеству M , затем каждый из оставшихся прямоугольников такими же прямыми разбить на четыре равных прямоугольника и снова выделить прямоугольники, целиком принадлежащие M и т. д., то каждая точка множества M в конце концов окажется в каком-либо из прямоугольников R_i , т. е.

$$M = R_1 + R_2 + \dots$$

Если x и y случайные величины, то $a \leq x < b$ и $c \leq y < d$ события, следовательно, одновременное выполнение этих неравенств также является событием. Наступление этого события происходит тогда и только тогда, когда точка X с координатами x, y принадлежит прямоугольнику R ($a \leq u < b, c \leq v < d$). Это справедливо для любого прямоугольника, поэтому принадлежность X множеству $M = R_1 + R_2 + \dots$ также является событием. При этом предполагается, что множество событий расширено таким образом, что бесконечная сумма событий $A_1 + A_2 + \dots$ также есть событие (§ 1Ж).

Если принадлежность точки X множеству M в плоскости uOv является событием и поэтому имеет вероятность, то множество M называется *измеримым*, а вероятность события, соответствующего множеству M , называется *мерой* этого множества. В силу аксиом теории вероятностей эта мера обладает обычными свойствами вполне аддитивных мер.

Согласно вышесказанному, каждое открытое множество измеримо. Отсюда следует, что любое замкнутое множество также измеримо, так как его дополнение открыто.

Пусть x и y — случайные величины и пусть $g(u, v)$ — непрерывная функция действительных переменных u и v . Тогда $g(x, y)$ снова является случайной величиной. Для доказательства этого нужно лишь показать, что при любом действительном t

$$g(x, y) < t$$

непрерывно является событием. После всего сказанного выше это очевидно, так как, в силу непрерывности функции g , множество тех точек (u, v) , для которых справедливо неравенство $g(u, v) < t$, является открытым множеством. Легко доказать, что то же самое остается в силе, если функция $g(u, v)$ *кусочно-непрерывна*; функция $g(u, v)$ называется *кусочно-непрерывной*, если плоскость E можно разбить на сумму конечного числа измеримых множеств $M_1 + \dots + M_n$, на каждом из которых $g(u, v)$ непрерывна. Впрочем, далее будут употребляться лишь простейшие случаи этой теоремы, в которых функция $g(u, v)$ или кусочно постоянна, или на каком-либо множестве M_1 непрерывна, а на дополнении $E - M_1$ равна нулю.

Б. ДВУМЕРНЫЕ ПЛОТНОСТИ ВЕРОЯТНОСТИ

Говорят, что пара случайных величин (\mathbf{x}, \mathbf{y}) обладает *плотностью вероятности* $f(u, v)$, если для любых действительных $a < b$ и $c < d$ вероятность события

$$a \leq \mathbf{x} < b \text{ и } c \leq \mathbf{y} < d$$

равна интегралу функции $f(u, v)$ по прямоугольнику

$$a \leq u < b, c \leq v < d$$

в плоскости uOv :

$$P(a \leq \mathbf{x} < b, c \leq \mathbf{y} < d) = \int_a^b \int_c^d f(u, v) du dv.$$

Интеграл в правой части должен быть обычным римановым интегралом; предполагается, что функция $f(u, v)$ интегрируема по Риману.

Теперь мы можем распространить формулу (16) из § 3 на функции двух переменных:

Теорема I. Если пара случайных величин (\mathbf{x}, \mathbf{y}) обладает плотностью вероятности $f(u, v)$, а функция $g(u, v)$ интегрируема по Риману, то

$$\mathcal{E} g(\mathbf{x}, \mathbf{y}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(u, v) f(u, v) du dv, \quad (1)$$

причем предполагается, что (риманов) интеграл справа сходится. Аналогичное утверждение справедливо для произвольного количества случайных величин $\mathbf{x}, \mathbf{y}, \dots$.

Доказательство. Сначала в плоскости uOv рассмотрим большой квадрат $-n \leq u < n, -n \leq v < n$ и предположим, что функция $g(u, v)$ вне этого квадрата равна нулю. Наш квадрат можно разбить на m^2 одинаковых частей, являющихся также квадратами. Если теперь на каждом частичном квадрате заменить функцию $g(u, v)$ ее верхней и нижней гранями, то получатся кусочно-постоянные нижняя функция $g_1(u, v)$ и верхняя функция $g_2(u, v)$. Для случайных величин $g_1(\mathbf{x}, \mathbf{y})$ и $g_2(\mathbf{x}, \mathbf{y})$, принимающих лишь конечное число значений, справедливость формулы (1) очевидна:

$$\mathcal{E} g_1(\mathbf{x}, \mathbf{y}) = \iint g_1(u, v) f(u, v) du dv, \quad (2)$$

$$\mathcal{E} g_2(\mathbf{x}, \mathbf{y}) = \iint g_2(u, v) f(u, v) du dv. \quad (3)$$

Левая часть (2) представляет собой сумму возможных значений $g_1(\mathbf{x}, \mathbf{y})$, умноженных на соответствующие вероятности, а правая часть -- сумму интегралов по частичным квадратам. Если в

каждом таком интеграле вынести постоянный множитель $g_1(u, v)$ за знак интеграла, то справа получим ту же сумму, что и слева. Доказательство справедливости формулы (3) будет аналогичным.

Для произвольной пары значений x, y случайных величин \mathbf{x}, \mathbf{y} имеем, очевидно,

$$g_1(x, y) \leq g(x, y) \leq g_2(x, y),$$

откуда следует

$$\mathcal{E} g_1(x, y) \leq \mathcal{E} g(x, y) \leq \mathcal{E} g_2(x, y).$$

Если теперь стороны частичных квадратов устремить к нулю, то правые части (2) и (3), между которыми заключено $\mathcal{E} g(x, y)$, будут стремиться к правой части (1). Отсюда следует справедливость формулы (1) для видоизмененных функций $g^{(n)}(u, v)$, принимающих нулевое значение вне большого квадрата.

Теперь остается совершить предельный переход при $n \rightarrow \infty$, который просто осуществляется применением теоремы из лебеговой теории интегрирования. Согласно этой теореме, интеграл по сумме множеств $A_1 + A_2 + \dots$ равен сумме интегралов по A_1, A_2, \dots . В нашем случае A_1 является событием, которое наступает, если точка (x, y) попадает в квадрат $-1 \leq u < 1, -1 \leq v < 1$. Аналогично, наступление события A_2 связано с попаданием этой точки в квадрат $-2 \leq u < 2, -2 \leq v < 2$ и т. д. Интеграл Лебега от случайной величины $g(x, y)$ по множеству $A_1 + A_2 + \dots + A_n$, согласно доказанному, равен интегралу Римана от произведения $g(u, v)f(u, v)$ по квадрату $-n \leq u < n, -n \leq v < n$. Отсюда непосредственным предельным переходом при $n \rightarrow \infty$ получаем формулу (1).

В. ВЫЧИСЛЕНИЕ ВЕРОЯТНОСТЕЙ ПОСРЕДСТВОМ ИНТЕГРИРОВАНИЯ

Важнейшее применение теоремы I заключается в следующем. Пусть G — произвольное открытое множество в плоскости uOv и пусть \mathbf{X} — случайная точка с координатами (x, y) . Спрашивается, какова вероятность события « \mathbf{X} принадлежит G »?

Для того чтобы осуществить возможность применения теоремы I, мы введем функцию

$$g(u, v) = \begin{cases} 1, & \text{если } (u, v) \text{ принадлежит } G, \\ 0, & \text{если } (u, v) \text{ не принадлежит } G. \end{cases}$$

Среднее значение случайной величины $g(x, y)$ равно вероятности того, что \mathbf{X} принадлежит G . Формула (1) в этом случае получает вид

$$P(\mathbf{X} \in G) = \int_G f(u, v) du dv. \quad (4)$$

Результат может быть без труда обобщен для произвольного количества случайных величин:

Теорема II. Если совокупность случайных величин x, y, \dots обладает плотностью вероятности $f(x, y, \dots)$, то вероятность попадания случайной точки \mathbf{X} с координатами x, y, \dots в произвольную область G равна интегралу от функции f по этой области:

$$P(\mathbf{X} \in G) = \int_G f(u, v, \dots) du dv \dots \quad (5)$$

г. ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ СУММЫ СЛУЧАЙНЫХ ВЕЛИЧИН

Применим теорему II к решению следующей задачи: Две независимые случайные величины x и y имеют плотности вероятности $f(u)$ и $g(v)$. Какова функция распределения суммы $x + y$?

В силу независимости x и y , вероятность того, что x лежит в пределах a и b , а y — в пределах c и d , равна произведению

$$\int_a^b f(u) du \int_c^d g(v) dv = \int_a^b \int_c^d f(u) g(v) du dv.$$

Следовательно, пара случайных величин (x, y) имеет плотность вероятности

$$f(u, v) = f(u) g(v).$$

Значение функции распределения $H(t)$ случайной величины $x + y$ в точке t равно вероятности того, что $x + y < t$:

$$H(t) = P(x + y < t).$$

По теореме II эта вероятность равна двойному интегралу

$$H(t) = \int \int_{u+v < t} f(u) g(v) du dv.$$

Этот интеграл можно записать в виде повторного интеграла:

$$H(t) = \int_{-\infty}^{+\infty} du \int_{-\infty}^{t-u} f(u) g(v) dv.$$

Если во внутреннем интеграле ввести новую переменную интегрирования $w = u + v$, то получим

$$H(t) = \int_{-\infty}^{+\infty} du \int_{-\infty}^t f(u) g(w - u) dw.$$

Меняя порядок интегрирования (для неотрицательных функций такая замена всегда возможна), найдем

$$H(t) := \int_{-\infty}^t \int_{-\infty}^{+\infty} f(u) g(w - u) du. \quad (6)$$

Эта функция распределения обладает плотностью вероятности

$$h(t) = \int_{-\infty}^{+\infty} f(u) g(t - u) du. \quad (7)$$

Таким образом, имеет место

Теорема III. Плотность вероятности суммы независимых случайных величин с плотностями $f(t)$ и $g(t)$ дается формулой (7).

Пример 5. Случайные величины x и y независимы и подчиняются нормальным распределениям. Какова функция распределения суммы $x + y$? Пусть плотности вероятности x и y имеют вид

$$f(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{t-a}{\sigma}\right)^2}, \quad (8)$$

$$g(t) = \frac{1}{\tau \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{t-b}{\tau}\right)^2}. \quad (9)$$

Без ограничения общности можно предполагать, что $a = b = 0$, так как заменой x и y на $x - a$ и $y - b$ соответственно мы всегда можем прийти к этому случаю. По теореме III сумма $z = x + y$ имеет плотность вероятности

$$\begin{aligned} h(t) &= \int_{-\infty}^{+\infty} f(u) g(t - u) du = \frac{1}{2\pi\sigma\tau} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} \left[\frac{u^2}{\sigma^2} + \frac{(t-u)^2}{\tau^2} \right]} du = \\ &= \frac{1}{2\pi\sigma\tau} \int_{-\infty}^{\infty} e^{-\frac{1}{2} (\alpha u^2 - 2\beta u + \gamma)} du, \end{aligned} \quad (10)$$

где

$$\alpha = \frac{1}{\sigma^2} + \frac{1}{\tau^2}, \quad \beta = \frac{t}{\tau^2}, \quad \gamma = \frac{t^2}{\tau^2}.$$

Вводя новую переменную интегрирования

$$v = u - \frac{\beta}{\alpha}, \quad (11)$$

получим

$$h(t) = \frac{1}{2\pi\sigma\tau} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} \alpha v^2 - \frac{1}{2} \delta} dv, \quad (12)$$

где

$$\delta = \frac{\alpha \gamma - \beta^2}{\alpha} = \frac{(\sigma^{-2} + \tau^{-2}) l^2 \tau^{-2} - l^2 \tau^{-4}}{\sigma^{-2} + \tau^{-2}} = \frac{l^2}{\sigma^2 + \tau^2},$$

или

$$\begin{aligned} h(t) &= \frac{1}{2\pi\sigma\tau} e^{-\frac{1}{2}\delta} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}av^2} dv = \frac{1}{2\pi\sigma\tau} e^{-\frac{1}{2}\delta} \sqrt{\frac{2\pi}{a}} = \\ &= \frac{1}{\sqrt{2\pi(\sigma^2 + \tau^2)}} e^{-\frac{1}{2} \frac{l^2}{\sigma^2 + \tau^2}}. \end{aligned} \quad (13)$$

Таким образом, z подчиняется нормальному распределению со средним значением 0 и дисперсией $\sigma^2 + \tau^2$. Следовательно, исходная сумма $x + y$ также распределена нормально со средним значением $a + b$ и дисперсией $\sigma^2 + \tau^2$.

Это же самое справедливо и для разности $x - y$ с той, конечно, поправкой, что среднее значение в этом случае будет равно $a - b$.

Повторным применением найденного результата устанавливается, что сумма произвольного количества независимых нормальных случайных величин снова распределена нормально.

ВЕРоятности и частоты

В этой главе основными являются первые три параграфа (§ 5, 6, 7). Последующие три параграфа (§ 8, 9, 10) посвящены приложениям к демографической статистике, медицине, биологии и физике. Тот, кто желает быстрее познакомиться с важными понятиями «эмпирического среднего» и «эмпирического квадратичного отклонения», может от § 7 сразу перейти к § 18.

§ 5. Биномиальное распределение

А. ФОРМУЛА БЕРНУЛЛИ

Предположим, что опыт, результат которого зависит от случая, многократно реализуется при одинаковых условиях. Например, ботаник путем скрещивания определенного родительского материала выводит потомство и классифицирует его по окраске цветов или по другим признакам. Каждый потомок является продуктом случая, а различные окраски цветов — возможными результатами опыта. Или хирург производит одну и ту же операцию над рядом пациентов и подсчитывает, сколько пациентов выздоравливает и сколько умирает.

Простоты ради предположим, что в каждом опыте имеется лишь две возможности A и \bar{A} , например, либо пациент умирает, либо не умирает. Смерть первого пациента является событием A_1 , смерть второго — событием A_2 и т. д. до A_n . Предполагается, что

1. Все опыты $E = A_1 + \bar{A}_1$, $E = A_2 + \bar{A}_2, \dots, E = A_n + \bar{A}_n$ независимы.

2. Все вероятности $P(A_1)$, $P(A_2), \dots, P(A_n)$ имеют одно и то же значение p .

Может показаться, что в приложениях к конкретным практическим задачам условие 2 выполняется редко. Пациенты данного хирурга имеют неодинаковую конституцию: у одних сердце лучше, у других хуже; большое значение имеют возраст и пол и т. д. Однако все это не мешает применимости теории, поскольку выбор пациентов не систематический, а случайный. Если наш хирург принимает пациентов независимо друг от друга и в том порядке, в каком они к нему поступают, то условие 2 можно считать выполненным. При этом не исключается, конечно, определен-

ный отбор для выявления тех пациентов, которым операция не нужна или для которых она слишком опасна. Недопустим лишь такой порядок операций, когда, например, сначала оперируют только мужчин, а затем — только женщин: мужчины и женщины должны следовать друг за другом в случайном порядке. Если $P(J)$ — вероятность того, что на операционном столе находится женщина, и $P(C|J)$ — вероятность смерти женщины, а $P(M)$ и $P(C|M)$ — соответствующие вероятности для мужчин, то по формуле полной вероятности (§ 1) вероятность смерти случайно выбранного пациента равна

$$P(C) = P(J) P(C|J) + P(M) P(C|M). \quad (1)$$

Таким образом, вероятность $P(C)$ остается постоянной, и поэтому условие 2 выполняется даже в том случае, когда условные вероятности смерти мужчин и женщин различны. Это заключение остается также справедливым, если пациентов классифицируют вместо пола по какому-нибудь другому признаку, и вероятности смерти для различных классов различны.

При фармакологических опытах, в которых подопытных животных подвергают действию определенного яда или какого-либо другого вещества, стремятся к тому, чтобы все животные находились в наиболее сходных условиях. Поэтому дозы вещества выбирают не равными, а пропорциональными весу животных. Однако для применимости общей теории это не является необходимым, так как если животные выбираются случайно и независимо (без учета их веса), то можно спокойно подвергать их воздействию одинаковых доз и, несмотря на это, считать, что вероятность реагирования одинакова для всех животных. Обосновать это можно точно так же, как в разобранным выше случае пациентов мужского и женского пола. Кроме того, следует отметить, что индивидуальное различие подопытных животных по их чувствительности к яду столь значительно различия в весе, что последним без колебаний можно пренебречь.

В теории вероятностей обычно полагают

$$P(A) = p \quad \text{и} \quad P(\bar{A}) = q \quad (p + q = 1).$$

Вероятность того, что в последовательности из n опытов k раз произойдет событие A и оставшиеся l раз — событие \bar{A} , согласно Бернулли, равна

$$W_k = \binom{n}{k} p^k q^l = \frac{n!}{k!l!} p^k q^l \quad (l = n - k). \quad (2)$$

Доказательство. Вероятность того, что в k определенных случаях наступит событие A , а в остальных случаях --- собы-

тие \bar{A} , равна $p^k q^l$. Количество различных комбинаций по k случаев из общего числа n равно $\binom{n}{k}$. Следовательно, вероятность того, что в каких-либо k случаях наступит событие A , а в остальных — событие \bar{A} , равна $\binom{n}{k} p^k q^l$.

Б. СРЕДНЕЕ ЗНАЧЕНИЕ И КВАДРАТИЧНОЕ ОТКЛОНЕНИЕ ЧИСЛА НАСТУПЛЕНИЙ СОБЫТИЯ

Если для каждого i определить случайную величину x_i , принимающую значения $+1$ или 0 , смотря по тому, осуществляется ли при i -м опыте событие A_i или нет, то сумма

$$x = x_1 + x_2 + \dots + x_n$$

будет случайной величиной, принимающей значение k , если событие A наступит k раз. Величины x_i независимы, так как опыты $E = A_i + \bar{A}_i$ предполагаются независимыми. Среднее значение величины x_i равно

$$\xi x_i = p \cdot 1 + q \cdot 0 = p.$$

Аналогично, средним значением величины x_i^2 является

$$\xi x_i^2 = p \cdot 1 + q \cdot 0 = p.$$

Следовательно, дисперсия величины x_i равна

$$\sigma_i^2 = \xi x_i^2 - (\xi x_i)^2 = p - p^2 = p(1 - p) = pq.$$

Среднее значение и дисперсия величины $x = x_1 + \dots + x_n$ равны

$$\xi x = np, \tag{3}$$

$$\sigma^2 = \sigma_1^2 + \dots + \sigma_n^2 = npq.$$

Поэтому квадратичное отклонение величины x имеет вид

$$\sigma = \sqrt{npq}. \tag{4}$$

В. ЗАКОН БОЛЬШИХ ЧИСЕЛ

Согласно неравенству Чебышева, вероятность того, что случайная величина $x - np$ примет значения, по абсолютной величине намного превышающие σ , чрезвычайно мала. *События, имеющие очень маленькую вероятность, следует рассматривать как почти невозможные; при однократной реализации условий, для которых эти события теоретически возможны, нельзя рассчитывать на их осуществление.* На этом принципе основано вообще

каждое практическое применение теории вероятностей. Поэтому те значения $k - np$, с которыми практически приходится считаться, являются величинами порядка не выше \sqrt{npq} . Множитель q можно отбросить и говорить: $k - np$ практически является величиной порядка \sqrt{np} . Преимущество последней формулировки особенно обнаруживается тогда, когда речь идет о редком событии и, следовательно, p близка к 0, а $q - к 1$.

Частота положительных исходов или частота наступления события A определяется отношением числа k к числу n :

$$h = \frac{k}{n}. \quad (5)$$

Если теперь мы построим случайную величину

$$h = \frac{x}{n},$$

значения которой равны $h = k/n$, то, согласно (3) и (4), математическое ожидание и квадратичное отклонение h будут даваться формулами

$$E h = p, \quad \sigma_h = \sqrt{\frac{pq}{n}}.$$

Постоянное подчеркивание различия между случайной величиной h и теми значениями h , которые эта величина может принимать, делает язык изложения излишне утомительным и поэтому в дальнейшем от него следует отказаться. Мы будем просто говорить о частоте h , молчаливо подразумевая при этом, что эта частота зависит от случая и, следовательно, обладает средним значением и квадратичным отклонением. Точно так же станем мы поступать и во всех подобных случаях. Буквы жирного шрифта будут встречаться в тексте все реже и реже и лишь там, где возникнет принципиальная необходимость вспомнить общие основы из гл. I.

Среднее значение h равно p . Следовательно, значения $|h - p|$, с которыми приходится встречаться при практических расчетах, являются величинами порядка не ниже, чем

$$\sqrt{\frac{pq}{n}}.$$

Так как $0 < pq \leq 1/4$, то можно также сказать: практически $|h - p|$ является величиной порядка не ниже, чем $n^{-1/2}$.

Смысл этого утверждения заключается в следующем: те значения $|h - p|$, которые велики сравнительно с $n^{-1/2}$, имеют все вместе лишь исчезающе малую вероятность.

Если количество опытов n возрастает, то $n^{-1/2}$ стремится к нулю. Таким образом, частота все более и более приближается к

значению вероятности p . На этом законе больших чисел основана принципиальная возможность статистического истолкования теории вероятностей.

§ 6. Как велико может быть отклонение частоты h от вероятности p ?

А. ПРИБЛИЖЕНИЕ БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ НОРМАЛЬНЫМ РАСПРЕДЕЛЕНИЕМ

Муавр и Лаплас исследовали биномиальное распределение при больших значениях n и аппроксимировали его некоторым значительно более удобным непрерывным распределением. Эти результаты хорошо известны, и поэтому мы ограничимся здесь лишь их кратким изложением¹. Сначала введем приближенное выражение вероятности W_k для больших np и nq , а именно

$$W_k \sim \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{z}{\sigma}\right)^2} \left[1 + \frac{1}{2} (p - q) \frac{z}{\sigma^2} - \frac{1}{6} (p - q) \frac{z^3}{\sigma^3} \right], \quad (1)$$

где $z = k - np$. При выводе этой формулы предполагается, что z являются величинами порядка $\sigma = \sqrt{npq}$. Однако формулу (1) можно применять, не задумываясь, и тогда, когда z велики по сравнению с σ , так как в этом случае правая и левая части (1) исчезающе малы². Рис. 3 показывает, насколько хорошим является приближение (1) даже при не очень больших n . Здесь для $n = 8$, $p = q = 1/2$ точные значения W_k изображены высотами прямоугольников, а приближенные значения — ординатами непрерывной кривой. Так как в этом случае $p - q = 0$, то непрерывная кривая является гауссовой кривой ошибок; если $p - q \neq 0$, то влияние дополнительного множителя в (1) сделает соответствующую кривую асимметричной. В пределе при $n \rightarrow \infty$ эта асимметрия исчезает.

Из приближенной формулы (1) следует, что функция распределения случайной величины x может быть достаточно хорошо

¹ Вывод этих результатов, основанный на формуле Стирлинга

$$n! = n^n e^{-n} \sqrt{2\pi n} \left(1 + \frac{\delta_n}{12n} \right) \quad (0 < \delta_n < 1),$$

можно найти в учебниках по теории вероятностей, из которых особого предпочтения заслуживает учебник: Марков Л. А., Исчисление вероятностей, 4-е изд., ГИЗ, 1924.

² Можно показать, что (1) остается справедливой, если $z^2/\sigma \rightarrow 0$ при $n \rightarrow \infty$. Если это условие не соблюдается, формулу (1) можно видоизменить так, чтобы и относительная погрешность приближения по-прежнему была бы мала при больших n (см. Феллер, Введение в теорию вероятностей и ее приложения, ИЛ, М., 1952, гл. VII, стр. 147). — Прим. ред.

аппроксимирована интегралом от гауссовой функции ошибок. Так как x может принимать лишь конечное число значений $k = 0, 1, \dots, n$, то в действительности функция распределения x является ступенчатой функцией¹ (рис. 4).

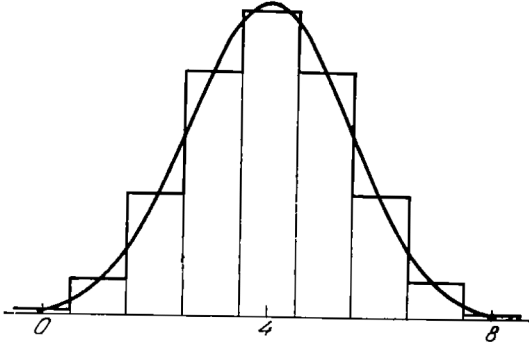


Рис. 3. Вероятности W_k при $n = 8$, $p = 1/2$.

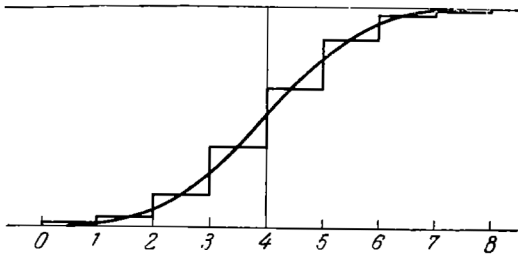


Рис. 4. График функции распределения случайной величины x при $n = 8$, $p = 1/2$.

Суммированием можно подсчитать общую вероятность тех значений k , для которых абсолютная величина разности $k - np$ не превышает произведения $g\sigma = g\sqrt{npq}$:

$$|k - np| \leq g\sqrt{npq}, \quad (2)$$

или, что то же самое,

$$|h - p| \leq g\sqrt{\frac{pq}{n}}. \quad (3)$$

¹ Имеются непрерывные кривые, которые даже при не очень больших n приближают биномиальное распределение еще лучше, чем гауссова кривая ошибок. См. Wise M. E., Proc. Kon. Ned. Akad. Amsterdam (section of sciences), A 57, 513.

Таким образом, устанавливается, что эта вероятность приближенно равна интегралу

$$2\Phi(g) - 1 = \frac{2}{\sqrt{2\pi}} \int_0^g e^{-\frac{1}{2}t^2} dt. \quad (4)$$

С ростом g функция $\Phi(g)$ столь быстро стремится к единице, что при $g = 2,58$ вероятность $2\Phi(g) - 1 = 0,99$, а при $g = 3$ равна даже $0,9973$. Это свойство словами выражается так: значения k — np , превышающие по абсолютной величине утроенное квадратичное отклонение, столь маловероятны, что при расчетах их едва ли следует принимать во внимание.

Этот результат справедлив¹ не только для очень больших, но, как показывают численные подсчеты, и для не очень больших значений n . Вообще в математической статистике часто оказывается, что 4 уже является большим числом. Если np и nq (или k и l) оба больше четырех, можно уверенно пользоваться сформулированным выше правилом 3σ .

Б. ОЦЕНКА КВАДРАТИЧНОГО ОТКЛОНЕНИЯ

Трудность применения правила 3σ заключается в том, что, хотя практически k , l и n (и, следовательно, частота h) бывают известны, однако p и q (а следовательно, и квадратичное отклонение σ) обычно остаются неизвестными. Для выхода из этого затруднения имеются различные пути.

Наиболее надежный из них заключается в замене pq наибольшим возможным значением этого произведения $1/4$. При этом $\sigma = \sqrt{npq}$ заменяется величиной $\sqrt{n}/2$. Этот прием особенно хорош тогда, когда наблюдаемая частота h близка к $1/2$.

Второй возможный путь заключается в замене p и q величинами h и $1 - h$. При этом $\sigma = \sqrt{npq}$ заменяется величиной

$$s = \sqrt{nh(1-h)} = \sqrt{\frac{kl}{n}}. \quad (5)$$

Для больших n это вполне допустимо, так как по закону больших чисел p близко к h ; здесь нужна лишь осторожность, когда k или l малы (скажем, меньше 4). А именно, если h близко к нулю, то результаты употребления $p(1-p)$ или $h(1-h)$ могут оказаться существенно различными. Иначе говоря, при малых k или l существует опасность, что определенное таким образом s

¹ Более точные приближения к биномиальному распределению были получены С. Н. Бернштейном [Изв. АН СССР, серия матем., 7 (1943), 3—16] и Феллером [Annales of Math. Statistic, 16 (1945), 319—329]. — Прим. ред.

окажется значительно меньше σ , поэтому, заменяя σ на s , мы получим заниженную оценку для квадратичного отклонения.

Особенно резко это заметно в крайнем случае при $k = 0$. Предположим, что некоторый хирург оперировал 90 пациентов и при этом не наблюдалось ни одного смертельного исхода. Статистика наблюдений здесь столь обширна, что о действительной смертности p можно высказать утверждение: величина p заведомо мала. Однако нет оснований полагать, что $p - h$ меньше утроенной оценки $s = \sqrt{(kl)/n}$ квадратичного отклонения σ : ведь в нашем случае $s = 0$, а вряд ли кто-нибудь решится утверждать, что смертность в точности равна нулю.

Один всегда возможный выход из этого затруднения будет указан в следующих параграфах¹. Здесь же мы ограничимся указанием небольшой поправки, которую при не слишком малых k и l целесообразно применить при вычислении s^2 для того, чтобы компенсировать возможное уменьшение значения квадратичного отклонения. Именно если сравнить $s^2 = (kl)/n$ с $\sigma^2 = npq$, то окажется, что среднее значение s^2 равно не σ^2 , а $\sigma^2(n-1)/n$. Путь вычислений таков:

$$\mathcal{E} s^2 = \mathcal{E} \frac{k(n-k)}{n} = \frac{\mathcal{E} kn - \mathcal{E} k^2}{n}.$$

Так как $\mathcal{E} k = np$ и

$$\mathcal{E} k^2 = (\mathcal{E} k)^2 + \sigma^2 = (np)^2 + npq,$$

то

$$\begin{aligned} \mathcal{E} s^2 &= \frac{pn^2 - (p^2n^2 + pqn)}{n} = np - np^2 - pq = \\ &= npq - pq = (n-1)pq = \frac{n-1}{n}\sigma^2. \end{aligned}$$

Следовательно, оценкой, среднее значение которой точно равняется σ^2 , будет не s^2 , а

$$s'^2 = \frac{n}{n-1} s^2 = \frac{kl}{n-1}. \quad (6)$$

Для того чтобы теперь получить исправленную оценку дисперсии частоты h , нужно s'^2 разделить на n^2 , так как $h = k/n$ и $\sigma_h^2 = \sigma^2/n^2$. Таким образом, оценка σ_h^2 имеет вид

$$s_h^2 = \frac{kl}{n^2(n-1)} = \frac{h(1-h)}{n-1}. \quad (7)$$

¹ Другой прием связан с отысканием такой функции от h , дисперсия которой почти не зависит от p . Подробнее об этом см. Рао С. Р., *Advanced Statistical Methods in Biometric Research*, Wiley, New York, 1952, p. 207—214.

Среднее значение s_n^2 равно

$$\mathbb{E} s_n^2 = \sigma_h^2 = \frac{\sigma^2}{n^2} = \frac{pq}{n}. \quad (8)$$

Если воспользоваться оценками h для p и s_n^2 для σ_h^2 , то можно всегда быть уверенным, что в среднем будут получаться правильные результаты: эти оценки не имеют *смещений*, т. е. лишены *систематических ошибок*.

Пример 6. С 1871 по 1900 г. в Швейцарии родились 1 359 671 мальчик и 1 285 086 девочек (см. Pölya, Handbuch der biol. Arbeitsmethoden, S. 742). Что можно сказать о величине вероятности рождения мальчика?

Частота рождения мальчика равна

$$h = \frac{k}{n} = \frac{1\,359\,671}{2\,644\,757} = 0,5141.$$

Число наблюдений очень велико, поэтому при расчетах можно, без сомнения, воспользоваться нормальным распределением. Квадратичное отклонение h равно

$$\sigma = \sqrt{\frac{pq}{n}} \approx \sqrt{\frac{1}{4n}} = 0,0003.$$

Если условиться, что возможные отклонения p от h не превышают 3σ , то мы должны будем сделать заключение, что, по-видимому, вероятность p лежит в пределах 0,5132 и 0,5150.

в. односторонние и двусторонние границы для h

Неравенство (3) указывает двусторонние границы для частоты h . Вероятность того, что h будет удовлетворять этому неравенству, приближенно равна $2\Phi(g) - 1$.

Если теперь мы положим $\Phi(g) = 1 - \beta$, то вероятность того, что неравенство (3) выполняется, будет равна

$$2\Phi(g) - 1 = 1 - 2\beta.$$

Следовательно, вероятность неравенства, противоположного (3), равна 2β . Число 2β можно сделать сколь угодно малым, если только выбрать g достаточно большим. Как уже упоминалось, для $g = 2,58$ будет $2\beta = 0,01$.

Если неравенство (3) не выполняется, то это означает, что либо h меньше нижней границы, либо h больше верхней границы, определяемых неравенством (3):

$$h - p > g \sqrt{\frac{pq}{n}} \quad \text{или} \quad h - p < -g \sqrt{\frac{pq}{n}}.$$

В обоих случаях вероятности почти одинаковы и, следовательно, приближенно равны β . Для того чтобы прийти к этому приближению, нужно в (1) пренебречь дополнительными членами

с z и z^2 ; следовательно, это приближение уже не такое хорошее, как (4). Однако если удовлетвориться этим грубым приближением, то можно сказать: с вероятностью $\sim 1 - \beta$ имеет место неравенство

$$h \leq p + g \sqrt{\frac{pq}{n}}$$

и точно так же с вероятностью $\sim 1 - \beta$ имеет место другое неравенство

$$h \geq p - g \sqrt{\frac{pq}{n}}.$$

Правые части этих неравенств указывают односторонние границы для h .

В табл. 3, в конце книги, указаны значения g , соответствующие различным доверительным уровням. В этой таблице доверительным уровнем односторонней границы называется величина β , а доверительным уровнем двусторонней границы — величина 2β . Величины β и g связаны соотношением

$$\Phi(g) = 1 - \beta. \quad (9)$$

§ 7. Доверительные границы для неизвестной вероятности

А. ПОСТАНОВКА ЗАДАЧИ

Пусть в условиях, описанных в предыдущем параграфе, наблюдается некоторая частота $h = k/n$. Какие границы можно указать для неизвестной вероятности p ?

Если требовать абсолютную надежность, то об этих границах нельзя сказать ничего более содержательного, чем то, что ими являются числа 0 и 1. Указание всяких других границ сопряжено всегда с риском совершить ошибку, вероятность которой называют доверительным уровнем. Допустимую вероятность ошибки, т. е. доверительный уровень двусторонних границ для p , мы будем снова обозначать 2β .

Выбор доверительного уровня в значительной степени зависит от той цели, которую мы перед собой ставим. Например, тарифы компаний по страхованию жизни должны быть рассчитаны таким образом, чтобы банкротство вследствие случайного повышения смертности было чрезвычайно маловероятным: здесь может оказаться неприемлемым даже уровень 0,01, так как он означает, что из ста таких страховых компаний в среднем одна обанкротится. С другой стороны, при статистических исследованиях в биологии и медицине имеется так много дополнительных источников ошибок (например, недостоверность теоретических пред-

положений, упрощающие допущения и т. д.), что дополнительная ошибка от применения статистики, соответствующая уровню 0,01, представляется сравнительно безобидной. Очень часто удовлетворяются даже величиной $2\beta = 0,05$.

Очень хорошие графические таблицы Коллера¹ рассчитаны для уровня $2\beta = 0,0027$, соответствующего утроенному квадратичному отклонению нормального распределения. Англичане в большинстве случаев пользуются уровнями $2\beta = 0,05$ или 0,01. В качестве доверительного уровня в дальнейшем мы, как правило, будем выбирать 0,01. Однако теоретические выводы останутся справедливыми для любого β .

Б. ПРИБЛИЖЕННОЕ РЕШЕНИЕ ЗАДАЧИ ПРИ БОЛЬШИХ n

Согласно формуле (3) § 6, при больших n с вероятностью $W = 1 - 2\beta$ выполняется неравенство

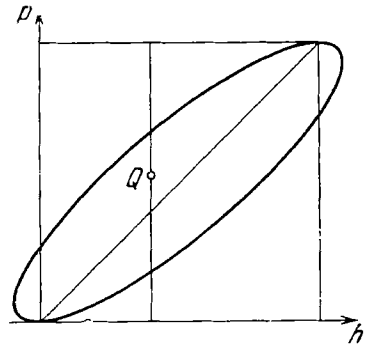
$$|h - p| \leq g \sqrt{\frac{pq}{n}}. \quad (1)$$

Вместо (1) можно также написать

$$(h - p)^2 \leq \frac{g^2}{n} p(1 - p). \quad (2)$$

Величину g , соответствующую заданному значению β , можно найти в табл. 3, в конце книги, где в последнем столбце указаны величины g^2 . Например, если выбрать $2\beta = 0,01$, то найдем, что $g = 2,58$ и $g^2 = 6,63$. Если же выбрать $2\beta = 0,05$, то будет $g = 1,96$ и $g^2 = 3,84$.

Если эмпирическую частоту h и вероятность p принять в качестве координат точки Q в плоскости hOp , то геометрическим местом точек, координаты которых удовлетворяют неравенству (2), будет замкнутая область с границей в виде эллипса, целиком расположенная между прямыми с уравнениями $h = 0$ и $h = 1$. Эллипс проходит через точки с координатами (0,0) и (1,1) и касается горизонтальных сторон единичного квадрата. Величины осей эллипса зависят от g и n : чем больше число опытов n , тем уже эллипс. Положение точки Q зависит от случая, так как координата h является случайной величиной. Вероятность попадания точки Q внутрь или на границу эллиптической области при любом



Р и с. 5. Доверительный эллипс.

¹ Koller S., Graphische Tafeln zur Beurteilung statistischer Zahlen, 2. Aufl., Dresden, 1943.

значении p приближенно равна $1 - 2\beta$. Иными словами (при $2\beta = 0,01$), утверждая, что точка Q не находится вне эллипса, мы в среднем ошибемся лишь в одном случае из ста.

На практике обычно p неизвестно, а h известно. Прямая с уравнением $h = \text{const}$ пересекает эллипс в двух точках, ординаты которых можно найти, решая квадратное уравнение

$$(h - p)^2 = \frac{g^2}{n} p(1 - p). \quad (3)$$

Отрезок, соединяющий эти точки, расположен внутри эллипса. Следовательно, если высказывается утверждение, что неизвестная вероятность p заключена между двумя корнями p_1 и p_2 квадратного уравнения (3), то вероятность ошибки равна 2β , так как это утверждение эквивалентно утверждению: Q не находится вне эллипса.

Это не означает, что в каждом отдельном случае при заданных h и n можно утверждать с вероятностью 0,99, что p лежит между p_1 и p_2 . В любом из этих случаев p имеет определенное, хотя и неизвестное значение, и высказывание: $p_1 \leq p \leq p_2$ — либо справедливо, либо ложно. Если оно справедливо, то ему соответствует вероятность 1, а если ложно, то — 0. Однако если имеется несколько последовательностей наблюдений, то, вычислив для каждой последовательности значения p_1 и p_2 , статистик может утверждать, что в любом случае $p_1 \leq p \leq p_2$. При этом он ошибется лишь в одном случае из ста. Следует подчеркнуть, что нельзя производить отбор опытов с целью получения какой-то определенной частоты: частоты h должны быть такими, какими они получаются по воле случая.

Решая квадратное уравнение (3), получим значения доверительных границ, между которыми предположительно заключено истинное значение p :

$$\left. \begin{aligned} p_1 &= \frac{hn + \frac{1}{2}g^2 - g\sqrt{h(1-h)n + \frac{1}{4}g^2}}{n + g^2} \\ p_2 &= \frac{hn + \frac{1}{2}g^2 + g\sqrt{h(1-h)n + \frac{1}{4}g^2}}{n + g^2} \end{aligned} \right\} \quad (4)$$

Вычисления по этим формулам довольно утомительны. Более практичен следующий графический прием.

Если в (3) положить

$$h - p = \frac{g}{n} x, \quad (5)$$

то уравнение эллипса (3) перейдет в уравнение окружности

$$x^2 = p(1 - p). \quad (6)$$

Если h известно, то в плоскости xOp уравнение (5) задает прямую, пересекающую окружность (6). Эта прямая проходит через точку с координатами $(0, h)$ и имеет угловой коэффициент $-g/\sqrt{n}$. Окружность (6) (диаметром около 10 см) можно заранее начертить на миллиметровой бумаге. Затем на оси Op (рис. 6) следует отло-

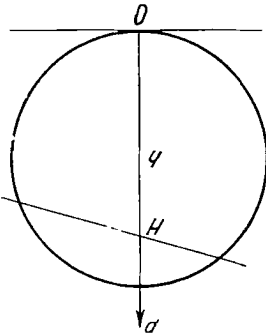


Рис. 6. Графическое построение доверительных границ.

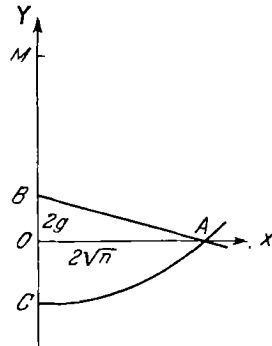


Рис. 7. Вспомогательный чертеж.

жить точку H с координатами $(0, h)$: через эту точку должна проходить прямая (5). Направление этой прямой можно определить с помощью вспомогательного чертежа (рис. 7) в плоскости XOY : на осях координат откладывают отрезки длины \sqrt{n} и g или, если желают иметь большую точность, отрезки $OA = 2\sqrt{n}$ и $OB = 2g$. Направление AB совпадает с направлением искомой прямой. Величину \sqrt{n} можно найти или по таблице квадратных корней или графически (если n не очень велико); для этого на оси OY в отрицательном направлении откладывается отрезок $OC = 2$ см, а в положительном направлении — отрезок $OM = (n - 1)$ см. Из точки M , как из центра, проводят окружность, проходящую через точку C . Эта окружность пересечет ось OX в точке A . Диаметр окружности равен $2MC = 2(n - 1)$ см, поэтому $(OA)^2 = OC(2MC - OC) = 4n$. Если теперь через точку H провести прямую, параллельную AB (см. рис. 6), то эта прямая пересечет нашу первую окружность в двух точках, ординаты которых равны p_1 и p_2 . Несредственно отсчитывая эти ординаты по миллиметровой бумаге, мы найдем искомые доверительные границы. Если воспользоваться лишь верхней (или лишь

нижней) доверительной границей, то доверительный уровень будет равен β .

Указанные здесь приближенные формулы становятся ненадежными, когда одно из математических ожиданий np или nq мало. Поэтому формулы (4) или эквивалентное им геометрическое построение я советую применять лишь тогда, когда оба результата наблюдений

$$k = hn \text{ и } n - k = (1 - h)n$$

по величине не менее четырех единиц.

Пример 7. С 1948 по 1952 г. в хирургической клинике Цюрихского университета было произведено 79 легочных операций с целью ликвидации расширения бронхов. Из 79 оперированных пациентов в течение послеоперационной недели умерли трое¹. Следовательно, наблюдаемая смертность равна

$$h \cdot 100 = \frac{3}{79} \cdot 100 = 3,8\%.$$

По формулам (4) или посредством конструкции, указанной на рис. 6, найдем следующие 5%-ные границы² для истинной смертности (в %):

$$p_1 = 1,3\%, \quad p_2 = 10,6\%.$$

Так как наблюдаемое количество смертельных исходов меньше четырех, то, осторожности ради, следует несколько расширить интервал между доверительными границами и сделать заключение: истинная смертность, по-видимому, более 1% и менее 11%.

Из этого примера видно, сколь мала точность определения вероятности по умеренному числу наблюдений.

В. ТОЧНОЕ РЕШЕНИЕ ЗАДАЧИ

Только что указанное приближенное решение задачи отыскания доверительных границ было основано на замене точного биномиального распределения (2) из § 5 некоторым непрерывным распределением (1) из § 6. Как показывает рис. 4, аппроксимирующая кривая иногда расположена ниже, а иногда выше точной ступенчатой линии. Вследствие этого точное значение доверительного уровня для приближенных доверительных границ в зависимости от истинного значения вероятности³ p иногда оказывается несколько больше, а иногда несколько меньше, чем 2β .

¹ Wegmann F., Die operative Behandlung der Bronchektasien, Diss. Zürich, 1955, Zusammenfassung, S. 39.

² 5%-ной границей называется такая доверительная граница, которой соответствует доверительный уровень 0,05. — Прим. перев.

³ В заметке van der Waerden B. L., Vertrauensgrenzen für unbekanntes Wahrscheinlichkeiten, Sitzungsber. sächs. Akad. Wiss., 91 (1939), 213, имеется график доверительного уровня как функции p , для предельного случая редких событий.

Однако, следуя Клопперу и Е. Пирсону¹, можно указать такие доверительные границы, для которых доверительный уровень $\leq 2\beta$, причем вероятность выхода за каждую из обеих доверительных границ не превышает β . Рассмотрим точную формулу биномиальных вероятностей

$$W_k(p) = \binom{n}{k} p^k q^{n-k} \quad (7)$$

Если p не слишком близка к 0 или к 1, то W_k с ростом k сначала монотонно возрастает, в некоторой точке k , близкой к np , она принимает наибольшее значение и затем монотонно убывает (см. рис. 3). Сумма всех W_k равна единице:

$$\sum_0^n W_k(p) = 1. \quad (8)$$

Односторонняя доверительная граница с доверительным уровнем $\leq \beta$ определяется следующим образом. Пусть K — целое число ($0 \leq K < n$). Образум из (8) частичную сумму от 0 до K :

$$S_K(p) = \sum_0^K W_k(p). \quad (9)$$

$S_K(p)$ равна вероятности того, что k примет какое-либо из значений от 0 до K .

Если S_K продифференцировать по p и затем привести подобные члены, то мы получим следующую отрицательную функцию:

$$\begin{aligned} \frac{d}{dp} S_K(p) &= -n q^{n-1} + \sum_1^K \left[k \binom{n}{k} p^{k-1} q^{n-k} - (n-k) \binom{n}{k} p^k q^{n-k-1} \right] = \\ &= -\frac{n!}{K!(n-K-1)!} p^K q^{n-K-1}. \end{aligned} \quad (10)$$

Попутно заметим, что из (10) следует интересное интегральное представление $S_K(p)$ в виде «неполной бета-функции»:

$$\begin{aligned} S_K(p) &= \int_p^1 \frac{n!}{K!(n-K-1)!} x^K (1-x)^{n-K-1} dx = \\ &= \int_0^{1-p} \frac{n!}{K!(n-K-1)!} (1-y)^K y^{n-K-1} dy. \end{aligned} \quad (11)$$

Однако теперь мы воспользуемся лишь тем обстоятельством, что, в силу (10), S_K является непрерывной убывающей функцией

¹ Clopper C. J. and Pearson E. S., *Biometrika*, 26 (1934), 404.

от p , которая при $p = 0$ принимает значение 1, а при $p = 1$ — значение 0. Отсюда следует, что S_K один и только один раз принимает любое промежуточное значение. Следовательно, для каждого $K < n$ можно определить p_K так, чтобы при $p = p_K$ функция $S_K(p)$ принимала значение β :

$$S_K(p_K) = \beta.$$

Клоппер и Пирсон сформулировали следующее правило: *Если в некотором эксперименте получена частота k/n , то p_k следует принять в качестве верхней доверительной границы, т. е. отбросить все те значения p , которые больше, чем p_k .* В этом случае можно утверждать, что вероятность ошибочно отбросить истинное значение p меньше β .

Доказательство. Если истинное p отброшено, то это значит, что $p > p_k$. Так как S_k является убывающей функцией p , то отсюда следует, что

$$S_k(p) < S_k(p_k) = \beta.$$

Пусть K — наибольшее из тех значений индексов r , для которых $S_r(p) < \beta$, тогда $k \leq K$, т. е. k совпадает с одним из значений $0, 1, \dots, K$. Но вероятность того, что k примет одно из значений от 0 до K , в точности равна $S_K(p)$ и, следовательно, она меньше β , что и требовалось доказать.

Таким образом, для $k < n$ верхняя доверительная граница p_k совпадает с решением уравнения

$$W_0(p) + W_1(p) + \dots + W_k(p) = \beta.$$

Аналогично если $k > 0$, то соответствующая нижняя доверительная граница является решением уравнения

$$W_k(p) + W_{k+1}(p) + \dots + W_n(p) = \beta.$$

При $k = 0$ нижней доверительной границей является, конечно, нуль и точно так же при $k = n$ верхней доверительной границей является единица.

Точные доверительные границы, определенные по Клопперу и Пирсону, значительно шире, чем приближенные границы p_1 и p_2 , описанные выше в разделе Б этого параграфа¹. Это связано

¹ Если $n \rightarrow \infty$, то можно показать, что отношение длин приближенного и точного доверительных интервалов стремится к единице, следовательно, эти интервалы асимптотически эквивалентны. При этом разность между приближенной и точной границами стремится к нулю. Для практического построения точных доверительных границ с доверительными уровнями $\beta = 0,005$ и $\beta = 0,025$ удобно пользоваться таблицами, имеющимися в книге Дунина-Барковского И. В. и Смирнова Н. В., Теория вероятностей и математическая статистика в технике (общая часть), ГИТТЛ, М., 1955. — *Прим. перев.*

с тем, что истинный доверительный уровень приближенных доверительных границ с изменением k колеблется около 2β , в то время как уровень значимости точных границ не превышает 2β (как правило, он бывает меньше 2β).

Какими же границами следует пользоваться, точными или приближенными? Мне представляется, что если условились считать допустимым риск ошибки, вероятность которого равна 2β , то, по-видимому, следует примириться и с колебаниями этой вероятности около 2β . Если оценивается не одна, а несколько вероятностей, и при этом каждый раз пользуются приближенными доверительными границами, то отклонения доверительных уровней вверх и вниз будут компенсировать друг друга и поэтому в среднем доверительные границы будут ошибочными лишь в $2\beta \cdot 100$ случаях из ста возможных. При очень малых значениях k (например, $k < 4$) для большей уверенности приближенную нижнюю доверительную границу следует несколько понизить.

§ 8. Проблема случайного отбора.

Выборочный метод

Из урны, содержащей K белых и L черных шаров ($K + L = N$), извлекается наугад n шаров (без возвращения). Какова вероятность того, что среди извлеченных шаров будет ровно k белых и l черных ($k + l = n$)?

Число возможных отборов n шаров из общего количества N равно $\binom{N}{n}$. Если шары хорошо перемешаны, то все эти отборы одинаково вероятны; следовательно, каждый из них имеет вероятность $1/\binom{N}{n}$. Число возможных отборов, содержащих k белых и l черных шаров, равно $\binom{K}{k}\binom{L}{l}$. Таким образом, искомая вероятность равна

$$W_k = \frac{\binom{K}{k}\binom{L}{l}}{\binom{N}{n}} = \frac{K!}{k!(K-k)!} \cdot \frac{L!}{l!(L-l)!} \cdot \frac{n!(N-n)!}{N!}. \quad (1)$$

Если теперь, так же как в § 5, ввести случайную величину x , значения которой совпадают с количеством извлеченных белых шаров, то окажется, что x равна сумме $x_1 + \dots + x_n$, где x_i зависит от цвета i -го извлеченного шара, а именно, если i -й шар белый, то $x_i = 1$, если же этот шар черный, то $x_i = 0$. Вероятность того, что x примет значение k , выражается формулой (1).

Распределение вероятностей, соответствующее (1), называют *гипергеометрическим распределением*.

Вычислим теперь среднее значение и дисперсию случайной величины \mathbf{x} . Согласно § 1 (пример 3), вероятность того, что $\mathbf{x}_i = 1$, равна K/N . Аналогично вероятность того, что $\mathbf{x}_i \mathbf{x}_j = 1$, равна либо $\frac{K(K-1)}{N(N-1)}$ (если $i \neq j$), либо K/N (если $i = j$).

Таким образом,

$$\mathcal{E}(\mathbf{x}_i^2) = \mathcal{E} \mathbf{x}_i = \frac{K}{N},$$

$$\mathcal{E}(\mathbf{x}_i \mathbf{x}_j) = \frac{K(K-1)}{N(N-1)} \quad (i \neq j).$$

Отсюда следует, что

$$\mathcal{E} \mathbf{x} = \sum \mathcal{E} \mathbf{x}_i = n \frac{K}{N},$$

$$\begin{aligned} \mathcal{E} \mathbf{x}^2 &= \mathcal{E}(\sum \mathbf{x}_i)^2 = \mathcal{E}(\sum \mathbf{x}_i^2 + 2 \sum_{j>i} \mathbf{x}_i \mathbf{x}_j) = \\ &= n \frac{K}{N} + n(n-1) \frac{K(K-1)}{N(N-1)}, \end{aligned}$$

$$\begin{aligned} \sigma^2 &= \mathcal{E} \mathbf{x}^2 - (\mathcal{E} \mathbf{x})^2 = n \frac{K}{N} + n(n-1) \frac{K(K-1)}{N(N-1)} - n^2 \left(\frac{K}{N}\right)^2 = \\ &= nK \cdot \frac{N(N-1) + N(n-1)(K-1) - n(N-1)K}{N^2(N-1)} = \\ &= \frac{nK(N-K)(N-n)}{N^2(N-1)} = \frac{KLn(N-n)}{N^2(N-1)}. \end{aligned} \quad (2)$$

Таким образом, значения k , с которыми практически приходится иметь дело, лежат вблизи от среднего значения nK/N , причем отклонение

$$k - n \frac{K}{N} = z$$

является величиной порядка

$$\sigma = \frac{1}{N} \sqrt{\frac{KLn(N-n)}{N-1}}. \quad (3)$$

С помощью формулы Стирлинга для $n!$, которую мы выведем в § 12, можно найти асимптотическую формулу для вероятности (1) при больших K , L , n и $N-n$. Опуская длинные выкладки, укажем лишь результат¹:

¹ Теорему о нормальном приближении для гипергеометрического, распределения см. в книге С. Н. Бернштейна, Теория вероятностей, изд. 4. ГТТИ, М., 1946. — *Прим. перее.*

$$W_k \sim \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \frac{z^2}{\sigma^2}} \left[1 + \frac{(K-L)(N-2n)}{N^2} \left(\frac{z}{2\sigma^2} - \frac{z^3}{6\sigma^4} \right) \right]. \quad (4)$$

Следовательно, как и в § 6, вероятность того, что k будет заключено в пределах $(nK/N) - g\sigma$ и $(nK/N) + g\sigma$, приближенно равна

$$\frac{2}{\sigma \sqrt{2\pi}} \int_0^g e^{-\frac{1}{2} t^2} dt = 2 \Phi(g) - 1. \quad (5)$$

Величину g по-прежнему можно выбрать таким образом, чтобы интеграл (5) принимал заданное значение $1 - 2\beta$ (см. табл. 3).

Равенство

$$\left| k - n \frac{K}{N} \right| \leq g\sigma, \quad (6)$$

вероятность которого выражается формулой (5), можно переписать так:

$$(kN - nK)^2 \leq g^2 N^2 \sigma^2$$

или, если воспользоваться для σ формулой (3),

$$\frac{(kN - nK)^2 (N-1)}{KLn(N-n)} \leq g^2. \quad (7)$$

Следовательно, вероятность неравенства (7) приближенно равна $1 - 2\beta$. Это приближение является равномерным в следующем смысле: для всякого $\varepsilon > 0$ существует такое $M(\varepsilon)$, что коль скоро все математические ожидания четырех случайных величин $k, l, K - k$ и $L - l$ будут больше M , то вероятность неравенства (7) будет отличаться от $1 - 2\beta$ менее чем на ε . Этим обстоятельством мы воспользуемся в следующем разделе.

Пример 8. Сущность *выборочного метода*, применяемого в статистике народонаселения и в экономической статистике, заключается в том, что статистическому обследованию подвергают лишь некоторую часть всего населения. Эта часть в том или ином смысле должна представлять все население (как говорят, выборка должна быть репрезентативной, т. е. представительной), например население больших городов, малых городов и деревень, северных и южных областей и т. д. должно встречаться в выборке примерно в тех же соотношениях, которые характерны для всего населения страны. Тогда частоты, вычисленные по выборке (например, смертность по различным причинам), будут являться приближенными значениями соответствующих частот для всего населения. Какие при этом следует ожидать отклонения выборочных частот h от соответствующих частот H для всего населения?

Если выборка из всего населения произведена случайно, то эта задача, очевидно, идентична нашей задаче с урной. Частотами, о которых шла речь, являются

$$h = \frac{k}{n} \quad \text{и} \quad H = \frac{K}{N}. \quad (8)$$

Среднее значение h равно H , квадратичное отклонение h дается формулой

$$\sigma_h = \frac{\sigma}{n} = \frac{1}{N} \sqrt{\frac{KL(N-n)}{n(N-1)}} = \sqrt{\frac{H(1-H)}{n} \cdot \frac{N-n}{N-1}}. \quad (9)$$

Следует ожидать, что отклонения $|h - H|$ в 99% всех случаев окажутся меньше, чем $2,58 \cdot \sigma_h$.

Обстоятельное изложение задач, связанных с выборочным методом, можно найти в недавно появившейся книге¹: *Schmetterer L., Einführung in die mathematische Statistik, Springer-Verlag, Wien, 1956, Kap. 2₁*.

§ 9. Сравнение двух вероятностей

А. ПОСТАНОВКА ЗАДАЧИ

Еще более важной задачей, чем задача оценки отдельной вероятности (особенно в медицине и биологии), является задача сравнения двух вероятностей. Например, если хирург испытал новый метод операции на ряде пациентов и при этом частота смертельных исходов оказалась меньше, чем при прежнем методе операции, то свидетельствует ли это об уменьшении смертности? Или, например, найдено новое лечебное средство против некоторой болезни; раньше из 400 пациентов умирали 40, т. е. 10%, а после применения нового средства из 50 пациентов умер лишь один, т. е. 2%. Свидетельствует ли это о действенности лекарства или же различие частот следует приписать влиянию случая?

Пусть найденные частоты равны

$$h_1 = \frac{k_1}{n_1}, \quad h_2 = \frac{k_2}{n_2}, \quad (1)$$

а соответствующие вероятности равны p_1 и p_2 . Предположим, что оказалось $h_1 > h_2$; как велика должна быть разность $h_1 - h_2$, чтобы с достаточной уверенностью можно было утверждать, что $p_1 > p_2$?

Как мы уже видели в § 5, случайная величина h_1 имеет среднее значение p_1 и квадратичное отклонение $\sigma_1 = \sqrt{p_1 q_1 / n_1}$. График функции распределения h_1 близок к гауссовой кривой с ошибкой с дополнительной асимметрией, влияние которой хорошо аппроксимируется дополнительными членами в формуле (1), § 6. Анало-

¹ См. также Дунин-Барковский И. В. и Смирнов Н. В., Теория вероятностей и математическая статистика в технике (общая часть), ГИТТЛ, М., 1955; Хальд А., Математическая статистика с техническими приложениями (перевод с англ.), ИЛ, М., 1956. — *Прим. перев.*

гичное обстоятельство справедливо и для h_2 . Следовательно, разность $h_1 - h_2$ имеет среднее значение $p_1 - p_2$ и дисперсию

$$\sigma^2 = \sigma_1^2 + \sigma_2^2. \quad (2)$$

В § 4 (пример 5) мы видели, что разность двух независимых нормальных случайных величин снова подчиняется нормальному распределению. Если же h_1 и h_2 распределены лишь приблизительно нормально, то $h_1 - h_2$ будет также иметь приблизительно нормальное распределение. Это приближение будет еще лучше, чем соответствующие приближения для распределений h_1 или h_2 , так как функция распределения разности $h_1 - h_2$ имеет меньшие скачки и меньшую асимметрию, чем функции распределения h_1 или h_2 .

Таким образом, мы, без сомнений, можем принять, что $h_1 - h_2$ имеет нормальное распределение со средним значением $p_1 - p_2$ и квадратичным отклонением σ . Но тогда следует, что значения $|(h_1 - h_2) - (p_1 - p_2)|$, превышающие g -кратное квадратичное отклонение, будут очень редки, причем g связано известным соотношением с заданным доверительным уровнем 2β (например, $g = 2,58$ при $2\beta = 0,01$). Следовательно, если $h_1 - h_2$ больше, чем $g\sigma$, то можно считать, что разность $p_1 - p_2$ положительна.

Но здесь опять возникает трудность, связанная с тем, что точное значение σ неизвестно. Имеются два пути для преодоления этой трудности. Идя по первому менее предпочтительному пути, в формуле (2) неизвестные величины σ_1^2 и σ_2^2 заменяют их приближенными значениями

$$s_1^2 = \frac{h_1(1-h_1)}{n_1-1} \quad \text{и} \quad s_2^2 = \frac{h_2(1-h_2)}{n_2-1}$$

и образуют сумму

$$s^2 = s_1^2 + s_2^2, \quad (3)$$

среднее значение которой равно σ^2 , однако в отдельных случаях s^2 может заметно отклоняться от σ^2 . Вместо условия $h_1 - h_2 > g\sigma$ теперь требуют, чтобы разность $h_1 - h_2$ была больше, чем gs . Если n_1 и n_2 велики, а частоты h_1 и h_2 не слишком близки к нулю или к единице, то при применении этого правила с $g = 2,58$ наши выводы будут ошибочными в среднем лишь в одном случае из ста.

В. КРИТЕРИЙ x^2

Второй путь связан с более простыми вычислениями, кроме того, он предпочтительнее и теоретически. Основной гипотезой, которую, может быть, следует отвергнуть и которая поэтому нуждается в проверке, является предположение, что различие h_1 и h_2 чисто случайное и что в действительности $p_1 = p_2$. Посту-

пая так, как поступал Сократ, когда он хотел диалектически опровергнуть утверждение своего собеседника, сначала допускают, что гипотеза $p_1 = p_2$ верна, и затем делают вывод: если эта гипотеза находится в противоречии с фактами, то ее следует отвергнуть.

В предположении, что $p_1 = p_2 = p$, имеем

$$\sigma_1^2 = \frac{pq}{n_1}, \quad \sigma_2^2 = \frac{pq}{n_2}$$

и

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 = \frac{pq(n_1 + n_2)}{n_1 n_2} = \frac{pqN}{n_1 n_2} = \frac{pq}{N} \cdot \frac{N^2}{n_1 n_2},$$

где $N = n_1 + n_2$. Теперь снова заменим неизвестную вероятность p соответствующей частотой. Для этой цели теперь имеется значительно больше материала, чем имелось раньше для отдельных p_1 и p_2 , так как мы теперь можем объединить вместе результаты всех $N = n_1 + n_2$ опытов, из которых $k_1 + k_2 = K$ дали положительный исход, а $l_1 + l_2 = L$ — отрицательный. Поэтому p и q заменяют частотами

$$H = \frac{K}{N} \quad \text{и} \quad 1 - H = \frac{L}{N} \quad (4)$$

и получают приближенное значение σ^2 (в силу указанных в § 6 Б соображений знаменатель N заменяют на $N - 1$):

$$s^2 = \frac{KL}{N-1} \cdot \frac{1}{n_1 n_2}. \quad (5)$$

Эта формула более надежна, чем выведенная ранее формула (3)¹. Если окажется, что

$$|h_1 - h_2| > gs,$$

или, что то же самое,

$$(h_1 - h_2)^2 > g^2 s^2,$$

или

$$\frac{(h_1 - h_2)^2 n_1 n_2 (N - 1)}{K L} > g^2,$$

то гипотезу $p_1 = p_2$ следует отвергнуть.

Положим

$$\chi^2 = \frac{(h_1 - h_2)^2 n_1 n_2 (N - 1)}{K L} \quad (6)$$

и назовем только что сформулированный критерий *критерием*

¹ Это верно, если верна основная гипотеза. В противном случае более надежной может оказаться формула (3), так как формула (5) будет в этом случае давать приближенные значения σ^2 со значительной систематической ошибкой. — *Прим. перев.*

χ^2 для сравнения двух вероятностей. Если h_1 и h_2 заменить их значениями, то вместо (6) можно написать

$$\chi^2 = \frac{(k_1 n_2 - k_2 n_1)^2 (N-1)}{K l_1 n_1 n_2}. \quad (7)$$

В. ОБОСНОВАНИЕ

Для обоснования этого критерия нужно доказать, что если гипотеза $p_1 = p_2$ верна, то вероятность неравенства

$$\chi^2 \leq g^2 \quad (8)$$

приближенно равна

$$1 - 2\beta = \frac{2}{\sqrt{2\pi}} \int_0^g e^{-\frac{1}{2}t^2} dt. \quad (9)$$

При этом будет предполагаться, что n_1 и n_2 являются большими числами и $p_1 = p_2 = p$ не слишком близки к нулю или к единице, так что математические ожидания pn_1 , pn_2 , qn_1 и qn_2 также являются большими числами. Тогда с большой вероятностью будут велики k_1 , k_2 , $l_1 = n_1 - k_1$ и $l_2 = n_2 - k_2$. Случаи, когда одно из этих четырех чисел мало, хотя и возможны, однако они не играют сколько-нибудь заметной роли при вычислении вероятности неравенства (8).

Пусть $P(K)$ — вероятность того, что $k_1 + k_2$ примет определенное целочисленное значение K , и пусть $P(\chi^2 \leq g^2 | K)$ — условная вероятность неравенства $\chi^2 \leq g^2$ в предположении, что $k_1 + k_2 = K$. Тогда по формуле полной вероятности (§ 1, формула (7))

$$P(\chi^2 \leq g^2) = \sum_K P(K) P(\chi^2 \leq g^2 | K). \quad (10)$$

Следовательно, если мы сможем доказать, что все условные вероятности $P(\chi^2 \leq g^2 | K)$ приближенно равны $1 - 2\beta$, то применение критерия χ^2 будет оправдано. Действительно, если каждый отдельный множитель в правой части (10) заключен в пределах $1 - 2\beta - \varepsilon$ и $1 - 2\beta + \varepsilon$, то левая часть (10) лежит в тех же пределах.

Вероятность, соответствующая паре значений (k_1, k_2) с $k_1 + k_2 = K$, равна

$$P(k_1, k_2) = \binom{n_1}{k_1} p^{k_1} q^{l_1} \binom{n_2}{k_2} p^{k_2} q^{l_2} = \binom{n_1}{k_1} \binom{n_2}{k_2} p^K q^L.$$

Вероятность $P(K)$ совпадает с вероятностью того, что из N независимых испытаний ровно K испытаний будут иметь положительный исход, следовательно,

$$P(K) = \binom{N}{K} p^K q^L.$$

Условная вероятность, соответствующая паре значений (k_1, k_2) при условии, что $k_1 + k_2 = K$, по определению условной вероятности (§ 1, формула (5)) равна отношению

$$P(k_1, k_2 | K) = \frac{P(k_1, k_2)}{P(K)} = \frac{\binom{n_1}{k_1} \binom{n_2}{k_2}}{\binom{N}{K}} = \frac{n_1! n_2! K! L!}{k_1! l_1! k_2! l_2! N!}.$$

Если теперь для упрощения обозначений положим

$$n_1 = n, \quad k_1 = k, \quad l_1 = l,$$

то получим

$$P(k, K - k | K) = \frac{n! (N - n)! K! L!}{k! l! (K - k)! (L - l)! N!}. \quad (11)$$

При делении множители p и q сократились, и найденный результат совпадает с формулой (1) § 8. Это означает, что *условная вероятность $P(k, K - k | K)$ в точности равна вероятности появления k белых и l черных шаров при n -кратном извлечении из урны, содержащей K белых и L черных шаров.*

Далее, если в (7) положить $k_2 = K - k$ и $n_2 = N - n$, то получим

$$\chi^2 = \frac{(kN - nK)^2 (N - 1)}{KLn(N - n)}. \quad (12)$$

Следовательно, неравенство $\chi^2 \leq g^2$ в точности совпадает с неравенством (7) § 8. Но, согласно § 8, вероятность этого неравенства приближенно равна $1 - 2\beta$, что и требовалось доказать.

Идея этого доказательства принадлежит М. Гипперту. Формула (7) с $N - 1$ в числителе впервые была указана Х. Шеллингом; ранее вместо $N - 1$ всегда писали N .

Г. ПРИМЕНЕНИЕ ОДНОСТОРОННЕГО И ДВУСТОРОННЕГО КРИТЕРИЕВ χ^2

На практике критерий χ^2 применяют не только для проверки гипотезы $p_1 = p_2$, но также и для выявления, какая из двух вероятностей больше: p_1 или p_2 ? Невольно напрашивается следующее правило (мы скоро убедимся в его справедливости). Если $\chi^2 > g^2$ и при этом $h_1 > h_2$, то следует считать, что $p_1 > p_2$. С другой стороны, если $\chi^2 > g^2$ и $h_1 < h_2$, то следует считать, что $p_1 < p_2$.

Если в действительности $p_1 = p_2$, то, как мы видели, вероятность ошибочного заключения $p_1 \neq p_2$ близка к 2β . А именно, в силу обоснования критерия χ^2 , вероятности неравенств $h_1 > h_2$ и $h_1 < h_2$ приближенно равны, поэтому с вероятностью, примерно равной β , может быть сделан ошибочный вывод $p_1 > p_2$ и почти с такой же вероятностью может возникнуть другой ошибочный вывод $p_1 < p_2$.

Но если $p_1 < p_2$, то вероятность одновременного осуществления двух событий $\chi^2 > g^2$ и $h_1 > h_2$ будет меньше β . Следовательно, в этом случае вероятность ошибочного заключения $p_1 > p_2$ будет также меньше β .

Аналогично если $p_1 > p_2$, то вероятность ошибочного вывода $p_1 < p_2$, полученного на основе критерия χ^2 , будет меньше, чем β .

Во всех трех случаях этого критерия вероятности ошибочных выводов не превышают 2β .

Если применяется односторонний вариант критерия χ^2 , то это означает, что заключение $p_1 > p_2$ (или соответственно $p_1 < p_2$) делается лишь в том случае, когда величина χ^2 достаточно велика и $h_1 > h_2$ (или, соответственно, $h_1 < h_2$); во всех остальных случаях от выводов воздерживаются. Практически очень часто, например, бывает интересно выяснить, действительно ли новое лекарство лучше старого? При этом не требуется ответа на вопрос, не будет ли действие нового лекарства одинаковым или худшим, чем действие старого? Доверительный уровень одностороннего критерия составляет лишь половину соответствующего уровня значимости двустороннего критерия.

Д. НАДЕЖНОСТЬ ПРИ НЕБОЛЬШИХ N

Критерий χ^2 можно уверенно применять и при небольших значениях N . На рис. 8, заимствованном из работы Гильдемайстера и автора этой книги, графически изображена зависимость истинного уровня критерия от p для некоторых типичных случаев (приближенный уровень значимости выбран равным $2\beta = 0,01$). Сплошные линии соответствуют случаю, когда в числителе (7) стоит N , а штриховые — случаю, когда N заменено величиной $N - 1$. Лишь в отдельных местах штриховые линии превышают 1%-ую границу, причем величина этого превышения мала и большинство кривых расположено ниже указанной границы.

Пример 9. С 1946 по 1951 г. в медицинской клинике Цюрихского университета для лечения последствий тромбоза—образования сгустков крови в кровеносных сосудах — 252 раза применялись антикоагулянты¹. Из 252 пациентов умерли 7, следовательно, смертность равнялась 2,8%.

¹ P u g a t s c h I., Zur Antikoagulantienbehandlung der Venenthrombosen in der inneren Medizin, Diss. Zürich, 1954.

С 1937 по 1942 г. антикоагулянты вообще не применялись. При подсчетах из всех случаев «консервативного» лечения тромбоза этих лет были исключены те, в которых лечение антикоагулянтами противопоказано. Оказалось, что из 205 оставшихся пациентов, подвергнутых консервативному лечению, умерли 37, т. е. 18,0%. Можно ли благотворность действия антикоагулянтов считать доказанной?

Вычисляя χ^2 по формулам (6) или (7), получим

$$\chi^2 = 30,2.$$

1%-ная граница равна 6,6, 0,1%-ная граница равна 10,8. Величина χ^2 значительно превышает обе эти границы. Следовательно, о случайности

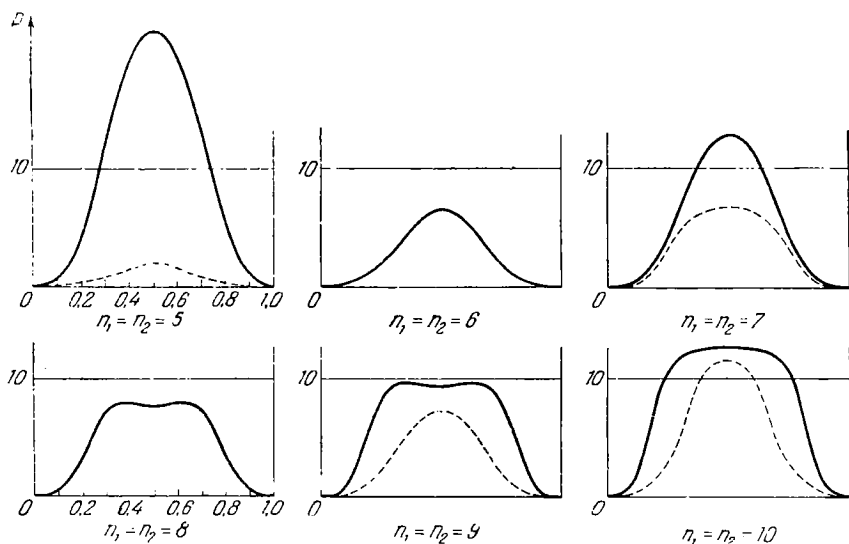


Рис. 8. Зависимость уровня значимости критерия χ^2 от p [из работы: *Gildemeister und van der Waerden, Ber. säch. Akad. Wiss., 95 (1943)*].

снижения смертности при антикоагулянтной терапии не может быть и речи.

С методической точки зрения можно бы было возразить против того, что оба ряда опытов относятся к различным периодам времени. Статистик-фанатик, возможно, стал бы часть пациентов лечить консервативно и одновременно другую часть — по новой терапии. Однако медик, стремящийся сделать все возможное для спасения жизни пациентов, в случае грозящего смертью тромбоза никогда так не поступит.

Если статистические данные о результатах лечения получены из двух рядов экспериментов, проводившихся не одновременно, то всегда следует ставить вопрос, не могут ли, помимо примененной терапии, с течением времени оказывать влияние и другие факторы (колебания эпидемиологических условий, гигиенических условий, условий питания и т. д.)? В случае тромбоза, конечно, следует заключить, что решающим фактором оказалась новая терапия.

Е. ТОЧНЫЙ КРИТЕРИЙ ФИШЕРА

Применяя тот же самый ход рассуждений, которым мы воспользовались в § 9 В для обоснования критерия χ^2 в предельном случае больших математических ожиданий, можно, как показал Р. А. Фишер, построить точный критерий, для которого односторонний уровень значимости всегда не превосходит β .

Конструкцию этого критерия можно уяснить на примере, заимствованном из работы Точера (*K. D. Tocher, Biometrika, vol. 37, 130*). Пусть наблюдались следующие числа:

$$\begin{array}{c|c|c} k_1 = 2 & l_1 = 5 & (n_1 = 7) \\ k_2 = 3 & l_2 = 2 & (n_2 = 5) \\ \hline (K = 5) & (L = 7) & (N = 12) \end{array}$$

Из этих чисел образуем так называемую «таблицу 2×2 » и рядом запишем все те таблицы 2×2 , для которых суммы по строчкам и столбцам имеют те же самые значения, что и в первой таблице, а величины k_1 строго меньше соответствующей наблюдаемой величины:

Наблюдаемая таблица	Дополнительные таблицы
$\begin{array}{c c c} 2 & 5 & 7 \\ 3 & 2 & 5 \\ \hline 5 & 7 & 12 \end{array}$	$\begin{array}{c c c c} 1 & 6 & 7 & 0 & 7 & 7 \\ 4 & 1 & 5 & 5 & 0 & 5 \\ \hline 5 & 7 & 12 & 5 & 7 & 12 \end{array}$

Затем вычислим и сложим условные вероятности, соответствующие всем этим таблицам, предполагая, что $k_1 + k_2 = K$.

$$P(k_1, k_2 | K) = \frac{n_1! n_2! K! L!}{k_1! l_1! k_2! l_2! N!} \quad (13)$$

В данном случае получим

$$P = 0,265 + 0,044 + 0,001 = 0,310.$$

Критерий гласит: Если сумма P не превосходит β , то гипотеза $p_1 = p_2$ отвергается в пользу конкурирующей гипотезы $p_1 < p_2$.

Например, если $\beta = 0,05$ и из опыта получены указанные выше числа (2, 5, 3, 2), то гипотеза $p_1 = p_2$ не отвергается. Однако если бы осуществился один из двух случаев, соответствующих дополнительным таблицам, то гипотезу следовало бы отвергнуть. Условная вероятность осуществления «дополнительных» случаев удовлетворяет неравенству

$$0,044 + 0,001 < 0,05,$$

следовательно, условная вероятность того, что гипотеза $p_1 = p_2$ будет ошибочно отвергнута, не превосходит 0,05.

Вообще, пусть A — событие, которое наступает тогда и только тогда, когда на основе указанного выше критерия гипотеза $p_1 = p_2$ отвергается в пользу конкурирующей гипотезы $p_1 < p_2$. Обозначим $P(A|K)$ условную вероятность этого события при заданных значениях сумм по столбцам $K = k_1 + k_2$ и $L = l_1 + l_2$ и в предположении, что гипотеза $p_1 = p_2$ справедлива. Тогда, в силу самой конструкции критерия, всегда имеет место неравенство

$$P(A|K) \leq \beta. \quad (14)$$

Безусловная вероятность события A равна

$$P(A) = \sum_K P(K) P(A|K), \quad (15)$$

где суммирование распространяется на все возможные значения K . Так как, согласно (14), $P(A|K) \leq \beta$ для всех K , то сумма (15) не превосходит β :

$$P(A) \leq \beta \sum P(K) = \beta.$$

Следовательно, вероятность ошибки одностороннего критерия Фишера всегда не превосходит β . В случае двустороннего критерия эта вероятность не превосходит, конечно, 2β .

В действительности при малых или не очень больших n_1 и n_2 вероятность ошибки критерия Фишера оказывается, как правило, существенно меньше 2β . Этот критерий излишне острожен и требует значительно больше вычислений, чем критерий χ^2 .

§ 10. Частота редких событий

А. ФОРМУЛА ПУАССОНА

Если в задаче Бернулли n велико, а вероятность p очень мала, так что np является небольшим числом, то формулы (2), (3) и (4) из § 5 и связанные с ними следствия остаются, конечно, справедливыми, однако указанное там асимптотическое разложение уже не будет иметь места. Для этого крайнего случая Пуассоном найдена другая асимптотическая формула, а именно

$$W_k \sim \frac{\lambda^k}{k!} e^{-\lambda}. \quad (1)$$

Здесь W_k — вероятность того, что в последовательности из n независимых испытаний редкое событие A осуществится ровно k раз, $\lambda = np$ — математическое ожидание k . Выражение (1) от n зависит неявно. Эта формула применима ко всем редким событиям, таким, как несчастные случаи, случаи разрушения атомного ядра и т. д.

Доказательство приближенной формулы (1) очень просто. Согласно точной формуле,

$$\begin{aligned} W_k &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{-k}. \end{aligned}$$

Первый множитель не зависит от n , второй при $n \rightarrow \infty$ стремится к $e^{-\lambda}$, а остальные множители стремятся к 1. Следовательно, предел W_k равен правой части (1).

До сих пор n было конечным числом (хотя, может быть, и очень большим), поэтому (1) являлась лишь приближенной формулой. Рассмотрим теперь идеализированный эксперимент, в котором число успехов может быть любым целым числом $k = 0, 1, 2, \dots$ и для которого формула (1) является не приближенной, а точной. Иными словами, мы рассматриваем случайную величину x , возможными значениями которой являются числа $k = 0, 1, 2, \dots$, которым соответствуют вероятности

$$W_k = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (2)$$

Ступенчатая функция распределения этой случайной величины x называется *функцией распределения Пуассона*:

$$F(t) = \mathbf{P}(x < t) = \sum_{k < t} \frac{\lambda^k}{k!} e^{-\lambda}.$$

Как и следовало ожидать, $F(t)$ при $t \rightarrow \infty$ стремится к единице.

Среднее значение x равно

$$\mathcal{E} x = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \quad (3)$$

Точно так же

$$\begin{aligned} \mathcal{E}(x^2) &= \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} (k-1+1) \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \\ &= \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} e^{-\lambda} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} = \\ &= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} = \lambda^2 + \lambda. \end{aligned}$$

Следовательно, дисперсия равна

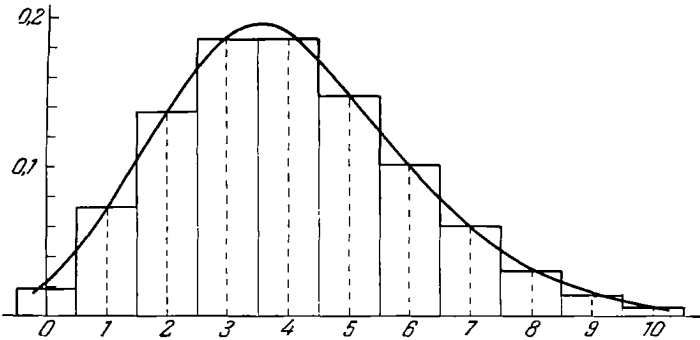
$$\sigma^2 = \mathcal{E}(x^2) - (\mathcal{E} x)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda,$$

поэтому $\sigma = \sqrt{\lambda}$. То же самое получим и из выведенной ранее формулы

$$\sigma = \sqrt{npq}$$

предельным переходом при $np \rightarrow \lambda$ и $q \rightarrow 1$.

Следовательно, значения случайной величины x , с которыми приходится иметь дело при практических расчетах, заключены между $\lambda - g\sqrt{\lambda}$ и $\lambda + g\sqrt{\lambda}$, при этом практически достаточно



Р и с. 9. Точное и асимптотическое распределение Пуассона ($\lambda = 4$).

выбрать значение g , не превышающее 3 или 4. Возможны ли дальнейшие уточнения?

Для больших λ и соответственно больших $k = \lambda + z$ существует асимптотическое разложение W_k :

$$W_k \sim \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{z^2}{2\lambda}} \left(1 - \frac{1}{2} \frac{z}{\lambda} + \frac{1}{6} \frac{z^3}{\lambda^2} \right). \quad (4)$$

Это разложение получается из (1) § 6 при $q = 1$. Рис. 9 показывает, что уже для $\lambda = 4$ соответствующее приближение оказывается очень хорошим. Число 4 снова выступает в качестве большого числа. Оказывается, что даже при не очень больших λ функцию распределения Пуассона (2) можно очень хорошо приблизить функцией нормального распределения с поправкой на асимметрию. Поэтому вероятность того, что x заключена в пределах $\lambda - g\sqrt{\lambda}$ и $\lambda + g\sqrt{\lambda}$, приближенно равна

$$2\Phi(g) - 1 = \sqrt{\frac{2}{\pi}} \int_0^g e^{-\frac{1}{2}t^2} dt.$$

Основываясь на этом, как и в § 7, по наблюдаемому значению

k можно построить доверительные границы для λ , решая квадратное уравнение $|k - \lambda| = \pm g\sqrt{\lambda}$ или

$$(k - \lambda)^2 = g^2 \lambda. \quad (5)$$

Значение g^2 снова следует взять из табл. 3. Обе доверительные границы являются решением (5):

$$\lambda_1 = k + \frac{1}{2} g^2 - g \sqrt{k + \frac{1}{4} g^2}, \quad \lambda_2 = k + \frac{1}{2} g^2 + g \sqrt{k + \frac{1}{4} g^2}. \quad (6)$$

Если используется лишь одна из этих двух границ, то соответствующий доверительный уровень уменьшится приблизительно вдвое против прежнего.

О вычислении точных доверительных границ см. *Garwood F., Biometrika, 28, (1936) 437.*

Пример 10. В течение 5 час. некоторым счетчиком космического излучения было зарегистрировано 80 космических частиц. Процесс регистрации, конечно, подвержен влиянию случая: если бы рядом с действовавшим был поставлен другой аналогичный счетчик, то вполне возможно, что он при той же интенсивности космического излучения зарегистрировал бы, скажем, 70 или 90 частиц. Целью наблюдений является не регистрация случайного числа частиц k , а оценка среднего значения λ числа частиц для всех аналогичных счетчиков в данной области в течение определенного отрезка времени. Среднее значение λ является мерой интенсивности космического излучения. Каковы границы, в которых может быть заключено λ ?

Так как совокупность космических частиц, попадающих в счетчик, составляет лишь крохотную долю совокупности всех частиц, пролетающих в данной части пространства, то можно применить формулу Пуассона для редких событий. Согласно этой формуле, квадратичное отклонение величины k равно $\sigma = \sqrt{\lambda}$. Так как k близко к λ , то $s = \sqrt{k} = \sqrt{80} \approx 9$ пригодно для оценки σ . Поэтому результат измерения в обычных обозначениях записывается так:

$$\begin{aligned} & \text{число частиц за 5 час.: } 80 \pm 9 \\ & \text{или число частиц за 1 час.: } 16 \pm 1,8. \end{aligned}$$

По правилу утроенного квадратичного отклонения отсюда следует, что среднее значение λ числа частиц за 5 час. предположительно заключено в границах

$$k - 3s = 80 - 27 = 53 \quad \text{и} \quad k + 3s = 80 + 27 = 107.$$

По более правильной формуле (6) с $g = 3$ находим несколько более высокие значения границ

$$\lambda_1 = 84,5 - 3\sqrt{82} = 57 \quad \text{и} \quad \lambda_2 = 84,5 + 3\sqrt{82} = 112.$$

Доверительный уровень, соответствующий величине $g = 3$, близок к 0,27%.

Б. СРАВНЕНИЕ ДВУХ ЧАСТОТ

При сравнении двух частот редких событий можно поступать подобно тому, как было указано в § 9. Пусть, например, в течение отрезков времени t_1 и t_2 произошло k_1 и k_2 успехов соответственно. Если средние количества успехов за единицу времени

$$m_1 = \frac{k_1}{t_1} \quad \text{и} \quad m_2 = \frac{k_2}{t_2} \quad (7)$$

различны, то возникает вопрос, можно ли это различие считать чисто случайным?

Предположим, что это различие было чисто случайным, т. е. что истинные средние значения количеств успехов за единицу времени были равны друг другу. Если мы это среднее значение обозначим буквой μ , то величины

$$\lambda_1 = \mu t_1 \quad \text{и} \quad \lambda_2 = \mu t_2$$

будут математическими ожиданиями количеств успехов x_1 и x_2 за t_1 и t_2 секунд соответственно. Квадратичные отклонения будут равны $\sqrt{\mu t_1}$ и $\sqrt{\mu t_2}$. Распределения x_1 и x_2 приближенно нормальны. Следовательно, разность

$$\frac{x_1}{t_1} - \frac{x_2}{t_2} \quad (8)$$

будет также иметь приближенно нормальное распределение с нулевым средним значением и дисперсией

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 = \frac{\mu}{t_1} + \frac{\mu}{t_2} = \frac{\mu}{t_1 t_2} (t_1 + t_2). \quad (9)$$

Если в результате эксперимента окажется, что абсолютная величина разности (8) превышает g -кратное квадратичное отклонение

$$\left| \frac{k_1}{t_1} - \frac{k_2}{t_2} \right| > g\sigma$$

или

$$\left(\frac{k_1}{t_1} - \frac{k_2}{t_2} \right)^2 > \frac{g^2 \mu}{t_1 t_2} (t_1 + t_2), \quad (10)$$

то гипотезу о чистой случайности различия частот следует отвергнуть.

Величина μ , однако, нам не известна. Как и в § 9, для преодоления этого затруднения постараемся заменить μ возможно наилучшей оценкой, использующей весь материал наблюдений. Так как за время $t_1 + t_2$ произошло $k_1 + k_2$ успехов, то в качестве оценки для μ выберем величину

$$m = \frac{k_1 + k_2}{t_1 + t_2}. \quad (11)$$

Если в неравенстве (10) заменить μ этой оценкой, то получится следующий практический критерий: *наблюдаемое различие коли-*

честв успехов за единицу времени следует считать случайным, коль скоро выполняется неравенство:

$$\left(\frac{k_1}{t_1} - \frac{k_2}{t_2}\right)^2 > \frac{g^2}{t_1 t_2} (k_1 + k_2). \quad (12)$$

Величина g^2 определяется по табл. 3 в конце книги.

Если применяется односторонний критерий, т. е. если заключение делается либо лишь в случае положительной разности, либо лишь в случае отрицательной разности в скобках левой части (12), то соответствующий уровень значимости составляет лишь половину прежнего.

Этот критерий можно записать так же, как критерий χ^2 , положив

$$\chi^2 = \left(\frac{k_1}{t_1} - \frac{k_2}{t_2}\right)^2 \frac{t_1 t_2}{k_1 + k_2} = \frac{(k_1 t_2 - k_2 t_1)^2}{t_1 t_2 (k_1 + k_2)}. \quad (13)$$

Обоснование критерия χ^2 достигается ровно так же, как это было сделано в § 9 В. Здесь мы можем отказаться от проведения обоснования, так как позднее мы снова вернемся к этому критерию и рассмотрим его с более общей точки зрения (§ 56 Ж).

МАТЕМАТИЧЕСКИЕ ВСПОМОГАТЕЛЬНЫЕ СРЕДСТВА

В первом чтении эту главу можно пропустить с тем, чтобы позднее, по мере надобности, обращаться к ней.

§ 11. Кратные интегралы.

Переход к полярным координатам

Мы будем называть *областью* любое открытое множество в пространстве действительных переменных x, y, \dots .

Как известно, двойной интеграл по плоской области G

$$I = \iint_G f(x, y) dx dy$$

можно вычислять двумя последовательными однократными интегрированиями

$$I = \int_c^d dy \int f(x, y) dx.$$

При этом внутреннее интегрирование по x производится по всем тем интервалам, которые образуются пересечением прямой $y = \text{const}$ с областью G . Что касается пределов интегрирования по y , то ими служат нижняя и верхняя грани координаты y для точек области G .

Точно так же $(m + n)$ -кратный интеграл

$$I = \int \dots \int_G f(x_1, \dots, x_m, y_1, \dots, y_n) dx_1 \dots dx_m dy_1 \dots dy_n$$

можно вычислить двумя последовательными интегрированиями: сначала по x_1, \dots, x_m , а затем по y_1, \dots, y_n :

$$I = \int \dots \int dy_1 \dots dy_n \int \dots \int f(x_1, \dots, x_m, y_1, \dots, y_n) dx_1 \dots dx_m. \quad (1)$$

При этом область внутреннего интегрирования задается множеством тех значений переменных x_1, \dots, x_m , для которых соответствующие точки с фиксированными y_1, \dots, y_n принадлежат области G . Иными словами, область внутреннего интегрирования определяется теми же неравенствами, что и область G , с той только разницей, что переменными в этих неравенствах

Таким же разложением преобразований и полной индукцией доказывается, что функциональный определитель преобразования (3) равен

$$\left| \frac{\partial(x_1, \dots, x_n)}{\partial(r, \varphi_1, \dots, \varphi_{n-1})} \right| = r^{n-1} \theta, \quad (5)$$

где θ зависит лишь от угловых переменных:

$$\theta = \sin^{n-2} \varphi_1 \sin^{n-3} \varphi_2 \dots \sin^2 \varphi_{n-3} \sin \varphi_{n-2}.$$

При доказательстве равенства (5) переход от $n - 1$ к n производится так:

$$\begin{aligned} \frac{\partial(x_1, \dots, x_n)}{\partial(r, \varphi_1, \dots, \varphi_{n-1})} &= \frac{\partial(x_1, x_2, \dots, x_n)}{\partial(x_1, r_1, \varphi_2, \dots, \varphi_{n-1})} \cdot \frac{\partial(x_1, r_1, \varphi_2, \dots, \varphi_{n-1})}{\partial(r, \varphi_1, \varphi_2, \dots, \varphi_{n-1})} = \\ &= \frac{\partial(x_2, x_3, \dots, x_n)}{\partial(r_1, \varphi_2, \dots, \varphi_{n-1})} \cdot \frac{\partial(x_1, r_1)}{\partial(r, \varphi_1)} = \\ &= r_1^{n-2} \sin^{n-3} \varphi_2 \dots \sin^2 \varphi_{n-3} \sin \varphi_{n-2} \cdot r = \\ &= (r \sin \varphi_1)^{n-2} \sin^{n-3} \varphi_2 \dots \sin^2 \varphi_{n-3} \sin \varphi_{n-2} \cdot r = r^{n-1} \theta. \end{aligned}$$

Таким образом, формула перехода к полярным координатам имеет вид

$$\int \dots \int f dx_1 \dots dx_n = \int \dots \int f r^{n-1} \theta d\varphi_1 \dots d\varphi_{n-1} dr. \quad (6)$$

Полагая в (6), для краткости, $\theta d\varphi_1 \dots d\varphi_{n-1} = d\Omega$, получим

$$\int \dots \int f dx_1 \dots dx_n = \int \dots \int f r^{n-1} dr d\Omega. \quad (7)$$

Если область G простирается в бесконечность или вблизи границы области G подынтегральная функция не ограничена, то *несобственный кратный интеграл* по G определяется как предел интегралов по последовательности ограниченных областей G_1, G_2, \dots , в каждой из которых подынтегральная функция ограничена (предполагается, что объединение G_1, G_2, \dots совпадает с G). Если существование этого предела не зависит от последовательности областей G_1, G_2, \dots и сам предел конечен, то соответствующий несобственный интеграл называют сходящимся. У тех интегралов, с которыми мы будем иметь дело в этой книге, подынтегральные функции при стремлении хотя бы одного из аргументов к бесконечности убывают столь быстро, что сходимость становится очевидной. В случае неотрицательных функций (например, в случае плотности вероятности) всегда имеется конечный или бесконечный предел, независимо от выбора последовательности областей G_1, G_2, \dots . Таким образом, здесь при исследовании сходимости можно ограничиться простейшими последовательностями областей.

Формула замены переменных (2) остается справедливой и для несобственных интегралов. Это же относится и к последовательному интегрированию (1), если только интегралы от обеих частей (1) сходятся.

§ 12. Бета- и гамма-функции

А. ГАММА-ФУНКЦИЯ

Эйлерова гамма-функция $\Gamma(z+1)$ для $z+1 > 0$ (или, в случае комплексного аргумента z , для $\operatorname{Re}(z+1) > 0$) задается формулой

$$\Gamma(z+1) = \int_0^{\infty} x^z e^{-x} dx. \quad (1)$$

Несобственный интеграл (1) определяется как предел собственного интеграла

$$\int_0^t x^z e^{-x} dx \quad (1a)$$

при $t \rightarrow \infty$, поэтому интеграл (1a) называют неполной гамма-функцией.

С помощью подстановок интегралу (1) можно придать другой вид: если положить $x = at$, то получим

$$\int_0^{\infty} t^z e^{-at} dt = a^{-(z+1)} \Gamma(z+1), \quad (2)$$

если же положить $x = \frac{1}{2} t^2$, то получим

$$\int_0^{\infty} t^{2z+1} e^{-\frac{1}{2} t^2} dt = 2^z \Gamma(z+1),$$

или, обозначив $2z+1 = n$ и заменив t на t/σ ,

$$\int_0^{\infty} t^n e^{-\frac{1}{2} \frac{t^2}{\sigma^2}} dt = 2^{\frac{n-1}{2}} \sigma^{n+1} \Gamma\left(\frac{n+1}{2}\right). \quad (3)$$

В частности,

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2} t^2} dt = 2 \int_0^{\infty} e^{-\frac{1}{2} t^2} dt = \sqrt{2} \Gamma\left(\frac{1}{2}\right). \quad (4)$$

Б. ФУНКЦИОНАЛЬНОЕ УРАВНЕНИЕ ГАММА-ФУНКЦИИ

Интегрированием по частям убеждаемся, что неопределенный интеграл (1а) удовлетворяет соотношению

$$\int x^z e^{-x} dx = -x^z e^{-x} + z \int x^{z-1} e^{-x} dx,$$

откуда после подстановки пределов интегрирования 0 и ∞ и в предположении, что $z > 0$, находим:

$$\Gamma(z + 1) = z \Gamma(z). \quad (5)$$

Очевидно, что $\Gamma(1) = 1$. Далее, из функционального уравнения (5) получаем

$$\Gamma(2) = 1 \cdot \Gamma(1) = 1,$$

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

и вообще для целочисленного n

$$\Gamma(n + 1) = n! \quad (6)$$

Чтобы вычислить $\Gamma(1/2)$, рассмотрим двойной интеграл по всей плоскости:

$$I = \iint e^{-\frac{1}{2}(x^2+y^2)} dx dy. \quad (7)$$

С одной стороны, интегрируя последовательно по x и y , согласно (4), получим

$$I = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} dx \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2} dy = 2 \left[\Gamma\left(\frac{1}{2}\right) \right]^2. \quad (8)$$

С другой стороны, в (7) можно ввести полярные координаты:

$$I = \iint e^{-\frac{1}{2}r^2} r dr d\varphi = \int_0^{2\pi} d\varphi \int_0^{\infty} e^{-\frac{1}{2}r^2} r dr = 2\pi \Gamma(1) = 2\pi. \quad (9)$$

Сравнивая (8) и (9), находим

$$\left[\Gamma\left(\frac{1}{2}\right) \right]^2 = \pi,$$

и так как $\Gamma(1/2) > 0$, то

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \quad (10)$$

Отсюда, пользуясь функциональным уравнением (5), можно определить $\Gamma(3/2)$, $\Gamma(5/2)$ и т. д.; например,

$$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2} \sqrt{\pi}. \quad (11)$$

В. ПЛОЩАДЬ ПОВЕРХНОСТИ МНОГОМЕРНОЙ СФЕРЫ

Если вместо (7) рассмотреть n -кратный интеграл по всему пространству

$$I = \int \dots \int e^{-\frac{1}{2}(x_1^2 + \dots + x_n^2)} dx_1 \dots dx_n, \quad (12)$$

то, с одной стороны, получим

$$I = \left[\int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} dx \right]^n = \left[\sqrt{2} \Gamma\left(\frac{1}{2}\right) \right]^n = (2\pi)^{\frac{n}{2}}, \quad (13)$$

а, с другой стороны, переходом к n -мерным полярным координатам (§ 11) найдем

$$I = \int_0^{\infty} e^{-\frac{1}{2}r^2} r^{n-1} dr \int d\Omega = 2^{\frac{n-2}{2}} \Gamma\left(\frac{n}{2}\right) \int d\Omega, \quad (14)$$

где интеграл $\int d\Omega$ распространяется на всю область изменения угловых переменных $\varphi_1, \dots, \varphi_{n-1}$. Сравнивая (13) и (14), получаем

$$\int d\Omega = \frac{2}{\Gamma\left(\frac{n}{2}\right)} \pi^{\frac{n}{2}}. \quad (15)$$

Например, если $n = 3$, то правая часть (15) равна известной со времен Архимеда площади поверхности шара единичного радиуса:

$$\int d\Omega = \frac{4}{\sqrt{\pi}} \pi^{\frac{3}{2}} = 4\pi.$$

Точно так же и (15) с геометрической точки зрения можно истолковать как площадь поверхности шара единичного радиуса в n -мерном пространстве.

Г. ФОРМУЛА СТИРЛИНГА

Выведем асимптотическую формулу для гамма-функции

$$\Gamma(\lambda + 1) = \int_0^{\infty} x^{\lambda} e^{-x} dx$$

при больших λ . Максимальное значение подинтегральной функции

$$f(x) = x^{\lambda} e^{-x}$$

достигается в точке $x = \lambda$. Для x , близких к λ , логарифм подинтегральной функции можно разложить в ряд

$$\begin{aligned} \ln f(x) &= \lambda \ln \lambda + \lambda \ln \frac{x}{\lambda} - x = \\ &= \lambda \ln \lambda + \lambda \left(\frac{x-\lambda}{\lambda} - \frac{(x-\lambda)^2}{2\lambda^2} + \dots \right) - x = \\ &= \lambda \ln \lambda - \lambda - \frac{(x-\lambda)^2}{2\lambda} + \dots, \end{aligned}$$

$$f(x) = \lambda^{\lambda} e^{-\lambda} e^{-\frac{1}{2\lambda}(x-\lambda)^2 + \dots}. \quad (16)$$

Если $|x - \lambda|$ меньше λ , то главный член ряда будет больше остальных членов, обозначенных от точием \dots , и поэтому ими можно пренебречь. Если же $x - \lambda$ является величиной того же порядка, что и λ , а λ велико сравнительно с единицей, то дополнительными членами также можно пренебречь, так как в этом случае как $f(x)$, так и правая часть (16) исчезающе малы. Следовательно, если отбросить дополнительные члены и проинтегрировать правую и левую части (16) от 0 до ∞ , то получим

$$\Gamma(\lambda + 1) \sim \lambda^{\lambda} e^{-\lambda} \int_0^{\infty} e^{-\frac{1}{2\lambda}(x-\lambda)^2} dx = \lambda^{\lambda - \frac{1}{2}} e^{-\lambda} \int_{-\sqrt{\lambda}}^{\infty} e^{-\frac{1}{2}t^2} dt.$$

Асимптотическое равенство \sim означает, что отношение обеих сторон стремится к единице при $\lambda \rightarrow \infty$. Справедливость асимптотического равенства не нарушится, если нижний предел интегрирования $-\sqrt{\lambda}$ заменить на $-\infty$. Воспользовавшись (4) и (10), получим *формулу Стирлинга*:

$$\Gamma(\lambda + 1) \sim \lambda^{\lambda + \frac{1}{2}} e^{-\lambda} \sqrt{2\pi}. \quad (17)$$

Если возьмем более точное разложение для $f(x)$, то найдем для гамма-функции более точное приближение¹:

¹ Точный вывод равенств (18) и (19) с оценкой остаточного члена и дальнейшее их уточнение можно найти в книге Крамера Г., Математические методы статистики, ИЛ, М., 1948, § 12.5. — *Прим. ред.*

$$\Gamma(\lambda + 1) = \lambda^{\lambda + \frac{1}{2}} e^{-\lambda} \sqrt{2\pi} \left(1 + \frac{1}{12\lambda} - R \right), \quad (18)$$

где остаточный член $-R$ отрицателен и является величиной порядка λ^{-2} . Последний множитель в (18) можно записать как $1 + \vartheta/12$, где $0 < \vartheta < 1$.

В частности, для целочисленных $\lambda = n$ из (17) следует асимптотическое равенство

$$n! \sim n^n e^{-n} \sqrt{2\pi n}. \quad (19)$$

Д. БЕТА-ФУНКЦИЯ

Эйлера бета-функция задается формулой

$$B(p + 1, q + 1) = \int_0^1 x^p (1 - x)^q dx. \quad (20)$$

Если оба параметра p и q по величине больше, чем -1 , то этот интеграл сходится. Подстановкой $u = ax$ получаем

$$\int_0^a u^p (a - u)^q du = a^{p+q+1} B(p + 1, q + 1). \quad (21)$$

Подстановкой $x = \sin^2 \varphi$ получаем

$$B(p + 1, q + 1) = 2 \int_0^{\frac{\pi}{2}} \sin^{2p+1} \varphi \cos^{2q+1} \varphi d\varphi. \quad (22)$$

Для того чтобы вычислить интеграл (20), рассмотрим двойной интеграл

$$I = \int_0^{\infty} \int_0^{\infty} e^{-\frac{1}{2}(x^2 + y^2)} x^{2q+1} y^{2p+1} dx dy.$$

С одной стороны, интегрируя последовательно по x и по y , согласно (3), получаем

$$\begin{aligned} I &= \int_0^{\infty} e^{-\frac{1}{2}x^2} x^{2q+1} dx \cdot \int_0^{\infty} e^{-\frac{1}{2}y^2} y^{2p+1} dy = \\ &= 2^q \Gamma(q + 1) 2^p \Gamma(p + 1) = 2^{p+q} \Gamma(p + 1) \Gamma(q + 1). \end{aligned} \quad (23)$$

С другой стороны, переходя к полярным координатам, находим

$$\begin{aligned} I &= \int_0^{\infty} e^{-\frac{1}{2}r^2} r^{2p+2q+3} dr \int_0^{\frac{\pi}{2}} \cos^{2q+1} \varphi \sin^{2p+1} \varphi d\varphi = \\ &= 2^{p+q+1} \Gamma(p+q+2) \frac{1}{2} B(p+1, q+1) = \\ &= 2^{p+q} \Gamma(p+q+2) B(p+1, q+1). \end{aligned} \quad (24)$$

Сравнение (23) и (24) показывает, что

$$\Gamma(p+1) \Gamma(q+1) = \Gamma(p+q+2) B(p+1, q+1),$$

таким образом,

$$B(p+1, q+1) = \frac{\Gamma(p+1) \Gamma(q+1)}{\Gamma(p+q+2)}. \quad (25)$$

Следующий интеграл можно свести к бета-функции:

$$K = \int_0^{\infty} (z^2 + a)^{-l} z^k dz \quad (k > -1, 2l - k > 1, a > 0). \quad (26)$$

А именно если положить $a(z^2 + a)^{-1} = y$ и, следовательно, $z^2 = a y^{-1} (1 - y)$, то этот интеграл преобразуется так:

$$K = \frac{1}{2} a^{\frac{k+1}{2} - l} \int_0^1 y^{l - \frac{k+3}{2}} (1 - y)^{\frac{k-1}{2}} dy = \frac{1}{2} a^{\frac{k+1}{2} - l} B\left(l - \frac{k+1}{2}, \frac{k+1}{2}\right)$$

или, согласно (25),

$$K = \frac{\Gamma\left(l - \frac{k+1}{2}\right) \Gamma\left(\frac{k+1}{2}\right)}{2\Gamma(l)} a^{\frac{k+1}{2} - l}. \quad (27)$$

В частности, для $k = 0$ имеем

$$\int_0^{\infty} (z^2 + a)^{-l} dz = \frac{\Gamma\left(l - \frac{1}{2}\right) \sqrt{\pi}}{2\Gamma(l)} a^{\frac{1}{2} - l}. \quad (28)$$

Таким образом, матрица обратного преобразования является транспонированной матрицей преобразования (1).

В силу (2), обратное преобразование (5) будет снова ортогональным, следовательно, и для транспонированной матрицы имеют место условия ортогональности:

$$\left. \begin{aligned} a_{i1}^2 + a_{i2}^2 + \dots + a_{in}^2 &= 1, \\ a_{i1} a_{j1} + a_{i2} a_{j2} + \dots + a_{in} a_{jn} &= 0. \end{aligned} \right\} \quad (6)$$

Точно так же доказывается обратное утверждение: условия (3) являются следствием условий (6).

Мы будем очень часто применять следующую теорему:

Любую начальную строку

$$y_1 = a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n,$$

коэффициенты которой удовлетворяют условию

$$a_{11}^2 + a_{12}^2 + \dots + a_{1n}^2 = 1,$$

можно дополнить до некоторого ортогонального преобразования (1).

Доказательство. Для коэффициентов второй строки имеются, согласно (6), одно линейное уравнение

$$a_{11} a_{21} + a_{12} a_{22} + \dots + a_{1n} a_{2n} = 0 \quad (7)$$

и одно квадратное

$$a_{21}^2 + a_{22}^2 + \dots + a_{2n}^2 = 1. \quad (8)$$

Линейное уравнение (7) заведомо имеет решение, отличное от нулевого. Умножением этого решения на подходящий множитель λ можно добиться, чтобы оно стало также и решением уравнения (8).

После того как первая и вторая строки найдены, для третьей строки мы получаем два линейных и одно квадратное уравнения. Так как оба линейных уравнения однородны, а число неизвестных n больше числа уравнений, то эти уравнения заведомо имеют решение, отличное от нулевого. Умножением этого решения на соответствующий множитель λ можно опять добиться того, чтобы найденное решение удовлетворяло соответствующему квадратному уравнению.

Продолжая этот процесс, в конце концов дойдем до последней строки. Здесь имеются $n - 1$ однородных линейных уравнений с n неизвестными и одно квадратное уравнение, которому можно удовлетворить подбором соответствующего множителя λ . Этим и завершается все доказательство.

§ 14. Квадратичные формы и их инварианты

А. ВЕКТОРЫ И ТЕНЗОРЫ

Вектором x называется упорядоченная совокупность из n чисел (x^1, \dots, x^n) . Если индексы указаны сверху, то x будем называть *верхним вектором*.

Линейная форма $L = \sum u_i x^i$ от переменных x^1, \dots, x^n определяется заданием *нижнего вектора* u с компонентами u_1, \dots, u_n . Аналогично квадратичная форма

$$G = g_{ik} x^i x^k \quad (g_{ik} = g_{ki})$$

определяется заданием некоторого (симметрического) *тензора* g_{ik} . В этом параграфе будет молчаливо предполагаться, что если один и тот же индекс встречается дважды (один раз наверху, а другой раз внизу), то по этому индексу производится суммирование. Квадратичная форма однозначно определяет билинейную форму от векторов x и y , полярную по отношению к этой квадратичной форме

$$G_{xy} = g_{ik} x^i y^k.$$

Если векторные компоненты x^i и y^i подвергаются некоторому обратимому линейному преобразованию

$$\left. \begin{aligned} x^i &= e_j^i x'^j, \\ y^i &= e_j^i y'^j, \end{aligned} \right\} \quad (1)$$

то u_i и g_{ik} должны так преобразоваться, чтобы формы $L = u_i x^i$ и $G_{xy} = g_{ik} x^i y^k$ остались инвариантными:

$$\begin{aligned} u_i x^i &= u_i e_j^i x'^j = u_j x'^j, \\ g_{ik} x^i y^k &= g_{ik} e_j^i e_r^k x'^j y'^r = g_{j'r'} x'^j y'^r. \end{aligned}$$

При этом, разумеется, вместе с G_{xy} останется инвариантсией также и $G_{xx} = G$. Таким образом, для нижних векторов и тензоров при преобразовании (1) имеют место соотношения

$$u_{j'} = u_i e_j^i, \quad (2)$$

$$g_{j'r'} = g_{ik} e_j^i e_r^k \quad (3)$$

Преобразование (2) называют *контраградиентным* (или *контравариантным*) преобразованием (1).

Если квадратичная форма G задана, то для каждого верхнего вектора y можно определить некоторый нижний вектор v :

$$v_i = g_{ik} y^k. \quad (4)$$

Билинейную форму $G_{xy} = g_{ik} x^i y^k$ можно теперь записать как $v_i x^i$. Так как G_{xy} при преобразовании (1) сохраняет инвариант-

ность, то $v_i x^i$ будет также инвариантной, т. е. v преобразуется в действительности как нижний вектор.

Б. ОБРАТНЫЕ МАТРИЦЫ

Если предположить, что форма G невырожденная, т. е. ее определитель g отличен от нуля, то систему уравнений (4) можно разрешить относительно y^k :

$$y^i = g^{ij} v_j. \quad (5)$$

Элементы матрицы (g^{ij}) равны соответствующим минорам матрицы (g_{ik}) , деленным на определитель g . Величины g^{ij} называют *элементами обратной матрицы*.

Если (4) подставить в (5), то получится тождество относительно y^i :

$$g^{ij} g_{jk} y^k = y^i.$$

В силу этого можно также написать

$$g^{ij} g_{jk} = \delta_k^i = \begin{cases} 1, & i = k, \\ 0, & i \neq k. \end{cases} \quad (6)$$

В (5) величины v_j произвольны. Если u — какой-либо второй нижний вектор, то форма $u_i y^i$ инвариантна, следовательно,

$$(uv) = g^{ij} u_i v_j \quad (7)$$

является инвариантом. Таким образом, для g^{ij} при преобразовании (1) имеют место соотношения, аналогичные (3).

Три инварианта

$$(xy) = g_{ik} x^i y^k, \quad (ux) = u_i x^i, \quad (uv) = g^{ij} u_i v_j$$

называют *скалярными произведениями*.

Как известно, любую квадратичную форму линейным преобразованием переменных можно представить как сумму и разность квадратов:

$$G = x_1'^2 + x_2'^2 + \dots + x_k'^2 - x_{k+1}'^2 - \dots - x_{k+l}'^2$$

(собственно говоря, индексы новых переменных x_i' следовало бы ставить сверху, однако от этого пришлось отказаться, так как квадраты символов с верхними индексами неудобны с полиграфической точки зрения).

Если $k = n$ и $l = 0$, то форма G принимает лишь положительные значения (случай, когда все переменные равны нулю, исключается) и называется *положительно определенной*; точно так же, если перед всеми квадратами стоит знак минус, то форма G назы-

вается отрицательно определенной. Таким образом, положительно определенную форму можно преобразовать в единичную форму

$$G = x_1'^2 + x_2'^2 + \dots + x_n'^2. \quad (8)$$

В случае такой единичной формы все скалярные произведения записываются особенно просто:

$$(xy) = \sum x'_i y'_i, \quad (ux) = \sum u'_i x'_i, \quad (uv) = \sum u'_i v'_i.$$

По теореме о произведении определителей из (3) следует, что

$$g' = |g_{j'l'}| = |g_{ik} e_{j'}^i| \cdot |e_{l'}^k| = |g_{ik}| \cdot |e_{j'}^i| \cdot |e_{l'}^k|$$

или

$$g' = g \Delta^2, \quad (9)$$

где Δ — определитель преобразования (1).

В частности, если преобразуемая форма является единичной формой, то $g' = 1$, следовательно,

$$\Delta = \pm \sqrt{g}.$$

В. ВЫЧИСЛЕНИЕ ОДНОГО ИНТЕГРАЛА

Вспользуемся этими алгебраическими вспомогательными средствами для вычисления интеграла

$$I = (2\pi)^{-\frac{n}{2}} g^{\frac{1}{2}} \int \dots \int e^{-\frac{1}{2}G} dx^1 dx^2 \dots dx^n, \quad (10)$$

B

где $G = g_{ik} x^i x^k$ — положительно определенная квадратичная форма и область интегрирования B задается двумя линейными неравенствами

$$(ux) > 0, \quad (vx) > 0. \quad (11)$$

Если введением новых переменных x'_1, \dots, x'_n преобразовать G в единичную форму (8), то (10) преобразуется в интеграл

$$I = (2\pi)^{-\frac{n}{2}} \int \dots \int e^{-\frac{1}{2}(x_1'^2 + \dots + x_n'^2)} dx'_1 \dots dx'_n, \quad (12)$$

B

где область B определяется неравенствами

$$(u'x') > 0, \quad (v'x') > 0.$$

Теперь с помощью ортогонального преобразования введем новые переменные y_1, \dots, y_n , полагая (см. § 13)

$$y_1 = \frac{(u'x')}{u} = \frac{u'_1 x'_1 + \dots + u'_n x'_n}{u}.$$

где

$$u = \sqrt{u_1^2 + \dots + u_n^2}.$$

При этом (vx) перейдет в некоторую линейную форму $(uy) = = w_1 y_1 + \dots + w_n y_n$. Наконец, вторым ортогональным преобразованием вместо y_2, \dots, y_n введем новые переменные z_2, \dots, z_n , полагая снова

$$z_2 = \frac{w_2 y_2 + \dots + w_n y_n}{w} \quad \text{и} \quad w = \sqrt{w_2^2 + \dots + w_n^2}.$$

Таким образом, мы получим

$$I = (2\pi)^{-\frac{n}{2}} \int \dots \int_B e^{-\frac{1}{2}(y_1^2 + z_2^2 + \dots + z_n^2)} dy_1 dz_2 \dots dz_n. \quad (13)$$

Формы $(ux) = (u'x')$ и $(vx) = (v'x')$ в новых переменных задаются формулами

$$\begin{aligned} (ux) &= uy_1, \\ (vx) &= w_1 y_1 + wz_2. \end{aligned}$$

Соответствующие скалярные произведения (в силу инвариантности скалярных произведений) равны:

$$\begin{aligned} (uu) &= (u'u') = u^2, \\ (uv) &= (u'v') = uw_1, \\ (vv) &= (v'v') = w_1^2 + w^2, \end{aligned}$$

а область интегрирования B задается неравенствами

$$uy_1 > 0, \quad u_1 y_1 + wz_2 > 0. \quad (14)$$

Теперь в (13) можно произвести интегрирование по z_2, \dots, z_n , а вместо остальных двух переменных y_1 и z_2 ввести полярные координаты:

$$y_1 = r \cos \varphi, \quad z_2 = r \sin \varphi.$$

Таким образом, получим

$$I = \frac{1}{2\pi} \int_0^\infty e^{-\frac{1}{2}r^2} r dr \int_\alpha^\beta d\varphi = \frac{\beta - \alpha}{2\pi}.$$

Пределы интегрирования по φ определяются так: каждое из неравенств (14) задает в плоскости $y_1 O z_2$ некоторую полуплоскость (рис. 10). Векторы с компонентами $(u, 0)$ и (u_1, w) являются внутренними нормальными к границам этих полуплоскостей, причем

косинус угла γ , заключенного между нормальями, дается формулой

$$\cos \gamma = \frac{uw_1}{\sqrt{u^2} \sqrt{w_1^2 + w^2}} = \frac{(uv)}{\sqrt{(uu)(vv)}}.$$

Отсюда следует, что область интегрирования сосредоточена внутри угла, образованного пересечением полуплоскостей, по величине равного $\pi - \gamma$. Следовательно,

$$\beta - \alpha = \pi - \gamma = \arccos \frac{-(uv)}{\sqrt{(uu)(vv)}}$$

и в свою очередь наш интеграл равен

$$I = \frac{1}{2\pi} \arccos \frac{-(uv)}{\sqrt{(uu)(vv)}}. \quad (15)$$

При этом скалярные произведения можно непосредственно вычислять по формуле

$$(uv) = g^{ik} u_i v_k, \quad (16)$$

не выполняя в действительности ни одного из трех линейных преобразований координат.

Если бы область B определялась тремя линейными неравенствами

$$(ux) > 0, \quad (vx) > 0, \quad (wx) > 0,$$

то вычисление интеграла по указанному методу свелось бы к вычислению площади поверхности сферического треугольника. Как известно, эта площадь пропорциональна сферическому избытку, который определяется как разность суммы углов сферического треугольника и π , следовательно,

$$I = \frac{1}{4\pi} \left[\arccos \frac{-(uv)}{\sqrt{(uu)(vv)}} + \arccos \frac{-(uw)}{\sqrt{(uu)(ww)}} + \arccos \frac{-(vw)}{\sqrt{(vv)(ww)}} - \pi \right].$$

При четырех неравенствах пришлось бы вычислять объем сферического тетраэдра, что не так просто.

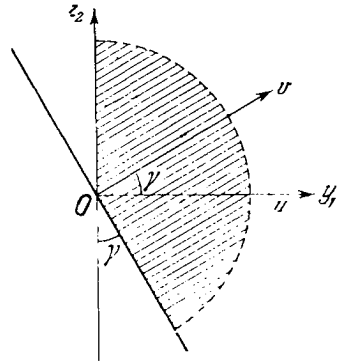


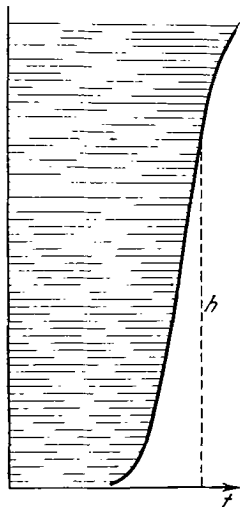
Рис. 10. Область интегрирования в плоскости $y_1 O z_2$

**ОЦЕНКИ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ,
СРЕДНИХ ЗНАЧЕНИЙ И ДИСПЕРСИЙ**

Важнейшими разделами этой главы являются § 15 и § 18.

§ 15. Кривая Кетле

Я до сих пор живо помню, как однажды, когда я был еще ребенком, мой отец привел меня на край города, где на берегу стояли ивы, и велел мне сорвать наугад сотню ивовых листочков. После отбора листьев с поврежденными кончиками у нас осталось 89 целых листиков. Вернувшись домой, мы расположили их в ряд по росту, как солдат. Затем мой отец через кончики листьев провел кривую и сказал: «Это и есть кривая Кетле. Глядя на нее, ты видишь, что посредственности всегда составляют подавляющее большинство и лишь немногие поднимаются выше или так и остаются внизу».



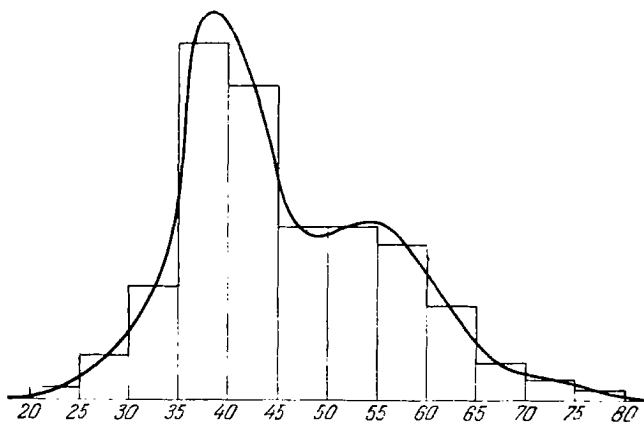
Р и с. 11. Кривая Кетле.

Если эту кривую расположить вертикально (рис. 11) и в качестве единицы масштаба на оси ординат выбрать отрезок, длина которого равна высоте всей фигуры, то ордината h , соответствующая абсциссе t , будет, очевидно, представлять собой частоту (или долю) тех ивовых листьев, длина которых меньше t . И так как частота h приближенно равна вероятности p , то наша кривая приближенно представляет $p = F(t)$ — функцию распределения длины листьев.

Измеренные длины ивовых листьев x_1, \dots, x_n образуют в совокупности то, что ныне называют выборкой. По выборке с помощью только что указанного приема можно эмпирически оценить функцию распределения $F(t)$. Определив приближенно $F(t)$, можно графическим дифференцированием оценить плотность вероятностей $f(t)$, однако результаты, как правило, бывают мало надежными.

Другой часто употребляемый способ оценки $f(t)$ и $F(t)$ основан на группировке наблюдаемых значений x . Интервал (t_0, t_r) , в котором заключены наблюдаемые значения x , произвольно вы-

в этом случае обусловлено перемешиванием друг с другом различных «чистых линий». Каждая «чистая линия» — потомство одного боба — подчиняется приблизительно нормальному распределению, в котором среднее значение при дальнейшей селекции либо совсем не меняется, либо меняется, но



Р и с. 12. Распределение веса бобов первого поколения, по Юханнсену.

очень незначительно. Средний вес 11 чистых линий имеет следующее распределение:

Границы весовых интервалов	35	40	45	50	55	60
Количество линий	4	2	0	2	3	

Смешением этих одиннадцати почти нормальных распределений и объясняется форма найденного эмпирического распределения.

§ 16. Оценки функций распределения

При первом изучении этой главы § 16 и 17 можно пропустить. Понятия из этих двух разделов будут использоваться значительно позднее.

На основе соображений, которые в предыдущем параграфе были наглядно объяснены на примере с ивовыми листьями, Колмогоровым была построена точная теория. Сначала, с помощью выборки x_1, \dots, x_n он определил *эмпирическую функцию распределения* $F_n(t)$, значение которой в произвольной точке t равно эмпирической частоте события $x_i < t$, т. е. равно количеству тех x_i , которые меньше t , деленному на n . График эмпирической функции

¹ Эта точка зрения не является общепринятой. — Прим. ред.

распределения — это не та гладкая кривая, которую с наивным воодушевлением строили Кетле и его ученики, а ступенчатая ломаная линия со скачками, по величине равными $\delta = 1/n$ во всех точках x_i (рис. 13)¹.

Спрашивается, насколько истинная функция распределения $y = F(t)$ может отличаться от эмпирической функции $y = F_n(t)$? Мы исследуем сначала положительные отклонения $F - F_n$, а затем — отрицательные. В практических приложениях F_n задана, а F неизвестна; однако при теоретических исследованиях мы можем $F(t)$ считать заданной, а $F_n(t)$ — зависящей от случая, так как наблюдаемые значения x_1, \dots, x_n являются случайными величинами. Пусть Δ — максимум разности $F - F_n$; требуется определить функцию распределения случайной величины Δ .

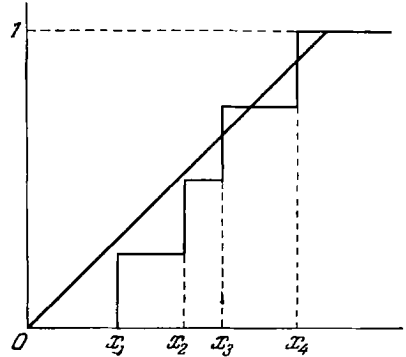


Рис. 13. Графики истинной и эмпирической функций распределения.

Относительно функции распределения $F(t)$ мы не делаем никаких предположений, кроме предположения о ее непрерывности. Так как непрерывное монотонное преобразование оси t не меняет разности $F - F_n$, то

вместо t и x в качестве новых переменных можно выбрать

$$t' = F(t) \text{ и } x' = F(x).$$

Это не изменит максимум Δ разности $F - F_n$. Если мы новые переменные снова обозначим t и x , то функция распределения будет иметь простой вид:

$$F(t) = t \quad (0 < t < 1). \tag{1}$$

Следовательно, график функции распределения представляет собой диагональ единичного квадрата. Событие, при котором x примет значение либо меньшее 0, либо большее 1, является невозможным; поэтому мы можем положить (рис. 13):

$$F(t) = \begin{cases} 0, & \text{если } t \leq 0, \\ 1, & \text{если } t \geq 1. \end{cases}$$

¹ Согласно определению $F_n(t)$, данному в тексте, в каждой точке x_i ($i = 1, 2, \dots, n$) $F_n(x_i - 0) = F_n(x_i) = \frac{i-1}{n}$ и $F_n(x_i + 0) = \frac{i}{n}$.

В полусегменте $x_{i-1} < t \leq x_i$ $F_n(t)$ сохраняет постоянное значение $\frac{i-1}{n}$; $F_n(t) = 0$ при $t \leq x_1$ и $F_n(t) = 1$ при $t > x_n$. — Прим. ред.

Плотность вероятности равна

$$f(t) = \begin{cases} 1, & \text{если } 0 < t < 1, \\ 0 & \text{в противном случае.} \end{cases}$$

Графически эта функция $y = f(t)$ изображается прямоугольником. Поэтому естественно также распределение называть *прямоугольным распределением*.

Мы хотим вычислить вероятность того, что Δ будет больше некоторой границы ε . Так как, по предположению, все x_1, \dots, x_n независимы и одинаково распределены с плотностью вероятности $f(t) = 1$, то, согласно § 4 (теорема II), искомая вероятность Q равна n -кратному интегралу

$$Q = \int_G \dots \int dx_1 dx_2 \dots dx_n \quad (2)$$

по области интегрирования G , определяемой неравенствами

$$0 < x_1 < 1, \dots, 0 < x_n < 1 \text{ и } \Delta > \varepsilon.$$

В дальнейшем для задания области интегрирования удобно ограничиться неравенствами

$$x_1 < x_2 < \dots < x_n.$$

Эти неравенства задают лишь часть всей области G . Однако перестановкой x_i ее можно перевести в любую другую аналогично определяемую часть области интегрирования (например, в $x_2 < x_1 < x_3 < \dots < x_n$). Такие перестановки не меняют ни ступенчатой функции F_n , ни максимума Δ . Все эти части области интегрирования имеют одинаковый объем и, следовательно, им соответствуют одинаковые вероятности. Граничной поверхности $x_i = x_k$ соответствует вероятность, равная нулю. Таким образом, искомая вероятность Q равна

$$Q = n! q, \quad (3)$$

где q — вероятность события

$$0 < x_1 < x_2 < \dots < x_n < 1, \Delta > \varepsilon. \quad (4)$$

В каждой точке x_k функция $F_n(t)$ совершает скачок от $(k-1)\delta$ к $k\delta$. Ясно, что разность $F - F_n$ достигает максимального значения Δ в одной из этих точек, причем если x_k — точка максимума, то $F_n(x_k) = (k-1)\delta$ и

$$\Delta = x_k - (k-1)\delta. \quad (5)$$

Следовательно, событие $\Delta > \varepsilon$ наступает тогда, когда хотя бы одна из разностей $x_k - (k-1)\delta$ окажется больше ε .

Пусть q_k -- вероятность того, что это событие произошло в точке с индексом k и не произошло в точках с меньшими индексами $j < k$. Так как случаи $k = 1, 2, \dots, n$ несовместны, то искомая вероятность q равна сумме q_k :

$$q = q_1 + q_2 + \dots + q_n. \quad (6)$$

Итак, нам осталось только вычислить q_k . Величина q_k представляет собой вероятность события

$$\left. \begin{aligned} 0 < x_1 < x_2 < \dots < x_n < 1, \\ x_k - (k-1)\delta > \varepsilon, \\ x_j - (j-1)\delta \leq \varepsilon \text{ для } j < k. \end{aligned} \right\} \quad (7)$$

Пусть сначала $k = 1$. Тогда мы имеем лишь неравенства

$$\left. \begin{aligned} x_1 < x_2 < \dots < x_n < 1, \\ x_1 > \varepsilon. \end{aligned} \right\} \quad (8)$$

Следовательно, все x_i лежат между ε и 1. Вероятность такого случая равна $(1 - \varepsilon)^n$. Так как всевозможные способы чередования индексов у x_i равновероятны, то вероятность события (8) равна

$$q_1 = \frac{1}{n!} (1 - \varepsilon)^n. \quad (9)$$

Если, далее, $k > 1$, то мы положим $k = h + 1$. Тогда (7) распадутся на такие неравенства, которые содержат лишь x_1, \dots, x_h :

$$\left. \begin{aligned} 0 < x_1 < \dots < x_h \\ x_j \leq \varepsilon + (j-1)\delta \text{ для } j = 1, \dots, h, \end{aligned} \right\} \quad (10)$$

и такие, которые содержат лишь x_{h+1}, \dots, x_n :

$$\left. \begin{aligned} x_{h+1} < x_{h+2} < \dots < x_n < 1, \\ x_{h+1} < \varepsilon + h\delta. \end{aligned} \right\} \quad (11)$$

Неравенство $x_h < x_{h+1}$, связывающее x_h и x_{h+1} , является следствием (10) и (11), и поэтому его можно отбросить.

События (10) и (11) относятся к неперекрывающимся интервалам и потому независимы. Следовательно, вероятность события (7) равна произведению вероятностей событий (10) и (11):

$$q_k = q_{h+1} = p_h r_h. \quad (12)$$

Вероятность r_h события (11) можно определить совсем просто. Метод тот же, что и при вычислении вероятности события (8); в результате получим

$$r_h = \frac{1}{(n-h)!} (1 - \varepsilon - h\delta)^{n-h}. \quad (13)$$

При этом предполагается, что $1 - \varepsilon - h\delta \geq 0$. Если это не так, то событие, соответствующее последнему неравенству в (11), является невозможным и поэтому $r_h = 0$.

Вероятность p_h события (10) равна

$$p_h = \int_0^\varepsilon dx_1 \int_{x_1}^{\varepsilon+\delta} dx_2 \int_{x_2}^{\varepsilon+2\delta} dx_3 \dots \int_{x_{h-1}}^{\varepsilon+(h-1)\delta} dx_h. \quad (14)$$

Для $h = 1$ находим без труда

$$p_1 = \varepsilon.$$

Если вычислить p_2 и p_3 , то возникает предположение, что

$$p_h = \frac{\varepsilon}{h!} (\varepsilon + h\delta)^{h-1}. \quad (15)$$

Это предположение можно легко проверить полной индукцией по h . Для $h = 1$ оно верно. Предположим, что оно верно для некоторого $h \geq 1$. Согласно (14), имеем

$$p_{h+1} = \int_0^\varepsilon dx_1 \int_{x_1}^{\varepsilon+\delta} dx_2 \int_{x_2}^{\varepsilon+2\delta} dx_3 \dots \int_{x_h}^{\varepsilon+h\delta} dx_{h+1}.$$

Если вместо x_2, x_3, \dots, x_{h+1} ввести новые переменные y_1, y_2, \dots, y_h по формулам

$$x_j = x_1 + y_{j-1},$$

то окажется, что

$$p_{h+1} = \int_0^\varepsilon R dx_1, \quad (16)$$

где

$$R = \int_0^{\varepsilon+\delta-x_1} dy_1 \int_{y_1}^{\varepsilon+2\delta-x_1} dy_2 \dots \int_{y_{h-1}}^{\varepsilon+h\delta-x_1} dy_h. \quad (17)$$

Если положить

$$\varepsilon + \delta - x_1 = \varepsilon',$$

то можно заметить, что интеграл R имеет точно такой же вид, как и p_h в (14), но только с заменой ε на ε' . Следовательно, согласно индуктивному предположению,

$$R = \frac{1}{h!} \varepsilon' (\varepsilon' + h\delta)^{h-1} = \frac{1}{h!} (\varepsilon + \delta - x_1) [\varepsilon + (h+1)\delta - x_1]^{h-1}. \quad (18)$$

Если (18) подставить в (16) и в качестве новой переменной выбрать $\varepsilon + (h + 1)\delta - x_1$, то интегрирование станет легко выполнимым. Результат

$$p_{h+1} = \frac{\varepsilon}{(h+1)!} [\varepsilon + (h+1)\delta]^h$$

имеет ту же самую форму, что и (15), но только с заменой h на $h + 1$. Этим завершается доказательство справедливости формулы (15).

Если (13) и (15) подставить в (12), то получим

$$q_k = \frac{\varepsilon}{h!(n-h)!} (\varepsilon + h\delta)^{h-1} (1 - \varepsilon - h\delta)^{n-h}. \quad (19)$$

Согласно (9), эта формула справедлива также и для $k = 1$. Наконец, из (3) и (6) получаем результат, впервые найденный Бернбаумом и Тинги¹

$$Q = \sum_{h=0}^H \binom{n}{h} \varepsilon (\varepsilon + h\delta)^{h-1} (1 - \varepsilon - h\delta)^{n-h}, \quad (20)$$

где

$$\delta = \frac{1}{n}. \quad (21)$$

Верхний предел суммирования H задается условием $1 - \varepsilon - h\delta \geq 0$. Следовательно, H является целой частью $n(1 - \varepsilon)$:

$$H = [n(1 - \varepsilon)]. \quad (22)$$

При больших n вычисление суммы (20) очень утомительно, поэтому для больших n лучше применять асимптотическую формулу, выведенную Н. В. Смирновым, согласно которой

$$Q \sim e^{-2n\varepsilon^2}. \quad (23)$$

Для каждого n и для заданного β (например, $\beta = 0,01$ или $\beta = 0,05$) можно найти такую границу ε , для которой $Q = \beta$. При малых n пользуются формулой (20), а при больших n — формулой (23). В табл. 4 для некоторых значений n и β указаны точные и асимптотические границы ε , по Бернбауму и Тинги. Из этой таблицы видно, что уже при $n = 50$ точные и асимптотические границы мало отличаются друг от друга. Так как асимптотические границы больше соответствующих точных границ (см. табл. 4), то при любом конечном n вероятность того, что Δ превзойдет асимптотическую границу, будет меньше β . Следовательно, за-

¹ Точная формула для распределения Δ при конечных n впервые дана Н. В. Смирновым в 1944 г. См. его статью «Приближение законов распределения случайных величин по эмпирическим данным», Усп. матем. наук, (10), (1944), 179—206. — *Прим. перев.*

мена точных границ асимптотическими изменяет доверительный интервал для Δ в сторону увеличения его надежности.

Найденная граница ε позволяет сформулировать *односторонний критерий* для проверки гипотезы, согласно которой функция распределения равна $F(t)$. А именно *если максимум Δ разности $F - F_n$ превосходит ε , то предположение, что функция распределения равна $F(t)$, должно быть отвергнуто*. Этот критерий мы будем называть Δ -критерием. Уровень значимости Δ -критерия равен β .

Если величины x и t заменить на $1 - x$ и $1 - t$ соответственно, то разность $F - F_n$ изменит знак, вследствие чего получится односторонний критерий с противоположной стороны: гипотетическую функцию распределения следует отвергнуть, если

$$\Delta' = \max (F_n - F) > \varepsilon.$$

Уровень значимости снова равен $\beta \sim e^{-2n\varepsilon^2}$.

Если этими двумя критериями воспользоваться одновременно, то получится *двусторонний критерий Колмогорова*. Согласно этому критерию, гипотетическая функция распределения отвергается, если максимум разности $|F - F_n|$ окажется больше ε . Уровень значимости такого критерия, очевидно, не превосходит 2β . При малых n вполне достаточно в качестве уровня значимости выбрать 2β : это лишь увеличит надежность критерия¹. При больших n приближенную формулу для уровня значимости можно уточнить, воспользовавшись асимптотической формулой Колмогорова:

$$2\beta \sim 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2 n \varepsilon^2}.$$

Этот ряд сходится очень быстро. Для практических целей можно ограничиться его первым членом

$$2e^{-2n\varepsilon^2},$$

который соответствует грубому правилу, сформулированному выше. В табл. 5 указаны значения ε для $2\beta = 0,01$ и $0,05$.

Практически критерием Колмогорова пользуются следующим образом. Сначала определяют эмпирическую функцию распределения $F_n(t)$. Затем, при заданном β , по табл. 5 находят ε и строят полосу, границами которой служат ступенчатые линии с

¹ Действительно, пусть $\mathbf{P}[\max (F - F_n) > \varepsilon] = \beta$, тогда, как показано выше, $\mathbf{P}[\min (F - F_n) < -\varepsilon] = \beta$. Если с помощью этого же ε построить двусторонний критерий, то соответствующий уровень значимости будет равен вероятности объединения событий: $\max (F - F_n) > \varepsilon$ и $\min (F - F_n) < -\varepsilon$. Так как эти события не являются несовместными, то вероятность их объединения меньше суммы их вероятностей, т. е. меньше 2β . — Прим. перев.

уравнениями $y = F_n(t) + \varepsilon$ и $y = F_n(t) - \varepsilon$. Предположительно, эта полоса целиком покрывает истинную кривую распределения с уравнением $y = F(t)$.

Л и т е р а т у р а к § 16

Kolmogoroff A., Determinazione empirica di una legge di distribuzione, *Giornale Istit. Ital. Attuari*, 4 (1933), 83.

Смирнов Н. В., Об уклонениях эмпирической кривой распределения, *Матем. сб.*, 6 (48), (1939), стр. 3.

Смирнов Н. В., Приближение законов распределения случайных величин по эмпирическим данным, *Усп. матем. наук*, (10), (1944), 179—206.

Feller W., On the Kolmogorov—Smirnov limit theorems for empirical distributions, *Ann. Math. Statist.*, 19 (1948), 177.

Birnbaum Z. W. and Tingey F. H., One-sided confidence contours for distribution functions, *Ann. Math. Statist.*, 22 (1951), 592.

Birnbaum Z. W., On the power of a one-sided test of fit for continuous probability functions, *Ann. Math. Statist.*, 24 (1953), 484.

§ 17. Порядковые статистики

Пусть снова x_1, \dots, x_n — выборка, состоящая из n независимых наблюдений случайной величины x с непрерывной функцией распределения $F(t)$.

Если все x_i расположены в порядке возрастания их величины и члены такой возрастающей последовательности обозначены $x^{(h)}$:

$$x^{(1)} < x^{(2)} < \dots < x^{(n)},$$

то каждый из $x^{(h)}$ называется *порядковой статистикой*, а соответствующая возрастающая последовательность — *вариационным рядом*.

Примеры порядковых статистик:

$x^{(1)}$ — наименьшее значение, $x^{(n)}$ — наибольшее значение из всех x_i .

Если n — нечетное число:

$$n = 2m - 1,$$

то $x^{(m)}$ называется *выборочной медианой*. Выборочная медиана является приближением *медианы* ζ , которая определяется как решение уравнения

$$F(\zeta) = \frac{1}{2}.$$

Для распределения с симметричной плотностью $y = f(t)$ и, в частности, для нормального распределения медиана ζ совпадает со средним значением \hat{x} ; поэтому выборочной медианой $x^{(m)} = Z$ можно в этом случае пользоваться в качестве удобной оценки для \hat{x} .

Аналогично если $n = 4r - 1$, то можно определить две *выборочные квартили*:

$$Z_1 = x^{(r)} \text{ и } Z_3 = x^{(3r)}.$$

Эти выборочные квартили и выборочная медиана разбивают вариационный ряд $x^{(1)}, \dots, x^{(n)}$ на четыре части, каждая из которых содержит по $r - 1$ величин. При больших n выборочные квартили близки к *квартилям* ζ_1 и ζ_3 соответствующего распределения, которые определяются как решения уравнений

$$F(\zeta_1) = \frac{1}{4} \text{ и } F(\zeta_3) = \frac{3}{4}.$$

Точно так же можно определить *выборочные секстили* Y_1 и Y_5 , которые служат приближенными значениями *секстилей* η_1 и η_5 , определяемых как решения уравнений

$$F(\eta_1) = \frac{1}{6} \text{ и } F(\eta_5) = \frac{5}{6}.$$

В случае нормального распределения η_1 и η_5 приближенно равны $\hat{x} - \sigma$ и $\hat{x} + \sigma$, так как если $\Phi(x)$ — функция нормального распределения с нулевым средним значением и единичной дисперсией, то

$$\Phi(-1) = 0,16\dots \text{ и } \Phi(+1) = 0,84\dots$$

Поэтому половина расстояния между выборочными секстилями, построенными по выборке из приближенно нормального распределения, может быть использована как удобная предварительная оценка квадратичного отклонения σ . Оценку для σ можно построить также и с помощью выборочных квартилей; к этому мы вернемся в § 20.

Для того чтобы можно было судить о точности этих оценок, нужно найти функции распределения $G^{(h)}(t)$ порядковых статистик $x^{(h)}$.

Значение $G^{(h)}(t)$ в точке t определяется как вероятность события $x^{(h)} < t$. Она равна вероятности того, что наименьшие h из всех n случайных величин x_i окажутся меньше t . Если W_k — вероятность события, которое заключается в том, что среди всех x_i имеется ровно k наблюдений, по величине меньших t , то, согласно § 5 (2),

$$W_k = \binom{n}{k} [F(t)]^k [1 - F(t)]^{n-k}. \quad (1)$$

Отсюда

$$G^{(h)}(t) = W_h + W_{h+1} + \dots + W_n. \quad (2)$$

Поставленная задача решена. Однако это решение несколько неудобно. Поэтому мы предположим, что $F(t)$ дифференцируема, обозначим

$$F'(t) = f(t)$$

и, дифференцируя (2) по t , вычислим плотность вероятностей

$$g^{(h)}(t) = W'_h + W'_{h+1} + \dots + W'_n.$$

При дифференцировании все члены, кроме первого, взаимно уничтожаются и остается лишь

$$g^{(h)}(t) = n \binom{n-1}{h-1} [F(t)]^{h-1} [1 - F(t)]^{n-h} f(t). \quad (3)$$

Произведение $F^{h-1} (1 - F)^{n-h}$ достигает наибольшего значения в точке

$$F = \frac{h-1}{n-1}.$$

Пусть этот случай осуществляется при $t = t_0$. Если обозначим

$$F_0 = F(t_0) = \frac{h-1}{n-1}, \quad f_0 = f(t_0), \quad F = F_0 + X,$$

то из (3) получим

$$g^{(h)}(t) = g^{(h)}(t_0) \left(1 + \frac{X}{F_0}\right)^{h-1} \left(1 - \frac{X}{1-F_0}\right)^{n-h} \frac{f}{f_0}. \quad (4)$$

Предположим теперь, что h и $n - h$ велики, и исследуем поведение трех сомножителей (4) в окрестности точки $t = t_0$.

Первый сомножитель

$$g^{(h)}(t_0) = \frac{n(n-1)!}{(h-1)!(n-h)!} \left(\frac{h-1}{n-1}\right)^{h-1} \left(\frac{n-h}{n-1}\right)^{n-h} f_0$$

можно аппроксимировать с помощью формулы Стирлинга

$$n! \sim n^n e^{-n} \sqrt{2\pi n}.$$

Получаем

$$g^{(h)}(t_0) \sim n \sqrt{\frac{n-1}{2\pi(h-1)(n-h)}} f_0 = \frac{\alpha}{\sqrt{2\pi}} f_0. \quad (5)$$

Второй сомножитель

$$Q = \left(1 + \frac{X}{F_0}\right)^{h-1} \left(1 - \frac{X}{1-F_0}\right)^{n-h}$$

прологарифмируем и разложим в ряд

$$\begin{aligned} \ln Q &= (h-1) \ln \left(1 + \frac{X}{F_0}\right) + (n-h) \ln \left(1 - \frac{X}{1-F_0}\right) = \\ &= (h-1) \left(\frac{X}{F_0} - \frac{X^2}{2F_0^2}\right) + (n-h) \left(-\frac{X}{1-F_0} - \frac{X^2}{2(1-F_0)^2}\right) + \dots = \\ &= -\frac{1}{2} X^2 \frac{(n-1)^2}{(h-1)(n-h)} + \dots \end{aligned}$$

Отброшенные члены являются величинами порядка не ниже, чем nX^3 , следовательно, при малых X остаток будет меньше главного члена, который является величиной порядка nX^2 . Если теперь в главном члене $(n-1)^3$ заменить на $n^2(n-1)$ (при больших n приближенная формула от этого не ухудшится), то получим асимптотическую формулу для второго сомножителя:

$$Q \sim e^{-\frac{1}{2} \alpha^2 X^2}. \quad (6)$$

Третий сомножитель (4) в окрестности точки $t = t_0$ можно заменить единицей. Поэтому, согласно (4), (5) и (6), получаем

$$g^{(h)}(t) \sim \frac{\alpha}{\sqrt{2\pi}} f_0 e^{-\frac{1}{2} \alpha^2 X^2}, \quad (7)$$

где

$$\alpha = n \sqrt{\frac{n-1}{(h-1)(n-h)}}. \quad (8)$$

Если произведение αX велико сравнительно с единицей, то правая часть (7) очень мала. Можно доказать, что в этом случае левая часть также мала. Тот факт, что интеграл от левой части мал¹, легко следует, впрочем, из закона больших чисел, а при практических приложениях только с таким интегралом и приходится иметь дело.

Теперь мы должны еще выразить X через t . Так как разность $t - t_0$ мала, а F дифференцируема, то разность $F - F_0$ приближенно равна $(t - t_0)f_0$. Следовательно, вместо (7) можно записать

$$g^{(h)}(t) \sim \frac{\alpha}{\sqrt{2\pi}} f_0 e^{-\frac{1}{2} \alpha^2 f_0^2 (t - t_0)^2}. \quad (9)$$

Формула (9) означает, что если h и $n-h$ велики, то порядковая статистика $x^{(h)}$ является асимптотически нормальной случайной величиной со средним значением t_0 и квадратичным отклонением

$$\sigma = \frac{1}{\alpha f_0} = \frac{1}{n} \sqrt{\frac{(h-1)(n-h)}{n-1}} \frac{1}{f_0}. \quad (10)$$

Для выборочной медианы Z отсюда находим

$$\sigma = \frac{1}{n} \sqrt{\frac{n-1}{2}} \frac{1}{f_0} \sim \frac{1}{2f_0 \sqrt{n}}.$$

¹ Здесь автор имеет в виду интеграл от $g^{(h)}(t)$ по такой области изменения t , где αX велико сравнительно с единицей. — *Прим. перев.*

В случае нормального распределения с нулевым средним значением и единичной дисперсией

$$t_0 = 0 \quad \text{и} \quad \frac{1}{f_0} = \sqrt{2\pi},$$

следовательно,

$$\sigma_z \sim \sqrt{\frac{\pi}{2n}}. \quad (11)$$

Для крайних порядковых статистик, т. е. для статистик, у которых малы либо h , либо $n - h$, асимптотическая оценка распределения более трудна. Этим оценкам посвящены важные труды Фишера и Типпета, Фреше, Мизеса и Гумбела¹. Читателю было бы полезно познакомиться с докладом Уилкса о порядковых статистиках [Wilks S. S., Bull. Amer. Math. Soc., 54 (1948)]².

Для пары порядковых статистик $x^{(h)}$, $x^{(j)}$ плотность вероятностей $f(u, v)$ может быть вычислена так же, как для одной порядковой статистики. Если, например, мы хотим найти совместное распределение наименьшего и наибольшего наблюдений в выборке, $x^{(1)}$ и $x^{(n)}$, то для этой цели можно воспользоваться тем обстоятельством, что вероятность одновременного осуществления событий $x^{(1)} > u$ и $x^{(n)} < v$ равна

$$[F(v) - F(u)]^n.$$

Отсюда дифференцированием по u и v получим искомую плотность вероятностей:

$$f(u, v) = n(n-1)[F(v) - F(u)]^{n-2} f(u)f(v). \quad (12)$$

Одной из важных функций от порядковых статистик, встречающейся во многих приложениях, является *размах*

$$W = x^{(n)} - x^{(1)}. \quad (13)$$

Функцию распределения размаха $H(t)$ можно получить интегрированием равенства (12):

$$H(t) = \mathbf{P}(W < t) = \int_{0 < v - u < t} \int f(u, v) \, du \, dv = \int_{-\infty}^{\infty} du \int_u^{u+t} f(u, v) \, dv.$$

Неопределенный интеграл от $f(u, v)$ по v равен, очевидно,

$$n[F(v) - F(u)]^{n-1} f(u).$$

¹ См. также Смирнов Н. В., Предельные распределения для членов вариационного ряда, Труды Математич. ин-та АН СССР, 25 (1949). — Прим. перев.

² Законченную теорию предельных распределений крайних членов вариационного ряда дал Б. В. Гнеденко [Ann. of Math. (44), (1943), 423—453]. — Прим. ред.

Если подставить пределы интегрирования u и $u + t$, то получим

$$\int_u^{u+t} f(u, v) dv = n[F(u+t) - F(u)]^{n-1} f(u),$$

следовательно,

$$H(t) = n \int_{-\infty}^{\infty} [F(u+t) - F(u)]^{n-1} f(u) du. \quad (14)$$

§ 18. Выборочное среднее значение и выборочная дисперсия

Среди постоянных, которыми характеризуется функция распределения, важнейшими являются *среднее значение* или *математическое ожидание*

$$\hat{x} = \mathcal{E} x = \int_{-\infty}^{\infty} t dF(t) \quad (1)$$

и дисперсия

$$\sigma^2 = \mathcal{E}(x - \hat{x})^2 = \int_{-\infty}^{\infty} (t - \hat{x})^2 dF(t). \quad (2)$$

Положительный квадратный корень из дисперсии называется *квадратичным отклонением* σ .

Мы предполагаем, что оба интеграла (1) и (2) сходятся. Как оценить по результатам наблюдений \hat{x} и σ^2 ?

Пусть имеется *выборка* (x_1, \dots, x_n) , т. е. имеются n наблюдаемых значений x_1, \dots, x_n случайной величины x . С точки зрения теории вероятностей x_1, \dots, x_n являются независимыми случайными величинами с одной и той же функцией распределения $F(t)$. Построим *арифметическое среднее* выборки

$$M = \frac{1}{n} (x_1 + \dots + x_n) = \frac{1}{n} x_1 + \dots + \frac{1}{n} x_n. \quad (3)$$

M называется *выборочным средним значением*. Часто вместо M пишут \bar{x} .

Величина M является случайной. Ее среднее значение равно

$$\mathcal{E} M = \hat{M} = \frac{\hat{x}}{n} + \dots + \frac{\hat{x}}{n} = \hat{x}, \quad (4)$$

а дисперсия, согласно § 3 (15), равна

$$\sigma_M^2 = \frac{\sigma^2}{n^2} + \dots + \frac{\sigma^2}{n^2} = n \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}. \quad (5)$$

Таким образом, дисперсия M при больших n оказывается значительно меньше дисперсии отдельного наблюдения x_i . Так как по неравенству Чебышева (§ 3 В) модуль разности $M - \hat{x}$ может являться с большой вероятностью лишь величинсой порядка

$$\sigma_M = \frac{\sigma}{\sqrt{n}},$$

то M является полезным приближенным значением для \hat{x} . Это приближение тем лучше, чем больше n .

Точно так же для σ^2 имеется приближенное значение

$$s_0^2 = \frac{1}{n} \sum (x_i - \hat{x})^2. \quad (6)$$

Математическое ожидание s_0^2 , очевидно, равно σ^2 .

Формулу (6) можно применять лишь тогда, когда \hat{x} точно известно. Этого, однако, в большинстве случаев не бывает. Поэтому здесь может помочь замена \hat{x} на M . Но тогда, как показывают расчеты, выражение (6) несколько уменьшается. А именно, для произвольного a имеет место тождество

$$\sum (x_i - M)^2 = \sum (x_i - a)^2 - n(M - a)^2, \quad (7)$$

которое доказывается так:

$$\begin{aligned} \sum (x_i - M)^2 &= \sum [(x_i - a) - (M - a)]^2 = \\ &= \sum (x_i - a)^2 - 2 \sum (x_i - a)(M - a) + n(M - a)^2 = \\ &= \sum (x_i - a)^2 - 2n(M - a)(M - a) + n(M - a)^2 = \\ &= \sum (x_i - a)^2 - n(M - a)^2. \end{aligned}$$

Положив в (7) $a = \hat{x}$, получим

$$\sum (x_i - M)^2 = \sum (x_i - \hat{x})^2 - n(M - \hat{x})^2. \quad (8)$$

Если в (6) \hat{x} заменить на M , то, как показывает тождество (8), правая часть (6), вообще говоря, уменьшится. Чтобы компенсировать это уменьшение, естественно сумму в левой части (8) разделить не на n , а на $n - 1$. Таким образом, получаем

$$s^2 = \frac{1}{n-1} \sum (x_i - M)^2. \quad (9)$$

Вычисляя математические ожидания правой и левой частей (8), найдем

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

или

$$\mathcal{E}(s^2) = \sigma^2. \quad (10)$$

Следовательно, знаменатель $n-1$ в (9) обеспечивает равенство математического ожидания s^2 величине σ^2 . s^2 называют *выборочной дисперсией*, а s — *выборочным квадратичным отклонением*¹.

Для упрощения вычислений s^2 можно использовать тождество (7). Целесообразно поступать так:

1. Сначала числовые значения x_i следует округлить таким образом, чтобы разность между наибольшим и наименьшим наблюдениями имела две значащие цифры. Практически это округление не окажет на s^2 никакого влияния.

2. Затем у всех округленных x_i можно отбросить запятую, т. е. помножить их на такую степень числа 10, чтобы все x_i стали целочисленными.

3. По округленным значениям x_i с помощью формулы (3) следует вычислить среднее M с одним запасным десятичным знаком.

4. Затем из интервала, в котором заключены все x_i , следует выбрать «круглое» число a и образовать разности $x_i - a$. Контроль: сумма этих разностей должна быть равна $n(M - a)$.

5. Разности $x_i - a$ не более чем двузначные целые числа, следовательно, их можно возвести в квадрат либо «в уме», либо с помощью очень короткой таблицы квадратов чисел.

6. После этого вычисляют $\sum(x_i - a)^2$, а затем, по формуле (7), $\sum(x_i - M)^2$ и, по формуле (9), s^2 (при вычислении s^2 результат округляется до целых чисел)².

7. Для контроля полезна формула (7) с $a = 0$:

$$\sum(x_i - M)^2 = \sum x_i^2 - nM^2. \quad (11)$$

Пример 12. Осадки за 90 лет в Ротемстеде (по книге R. A. Fisher, *Statistical Methods for Research Workers*, § 14). В первом столбце указано количество осадков x (в дюймах), во втором — соответствующее количество лет k , в течение которых наблюдалось данное количество осадков. В качестве приближенного значения для выборочного среднего выбрано $a = 28$. Третий столбец содержит разности $x - a$, четвертый — $k(x - a)$, пятый — $k(x - a)^2$.

¹ Эта терминология не является общепринятой. Чаще выборочной дисперсией называют величину $\sum(x_i - M)^2/n$, которая равна дисперсии соответствующей эмпирической функции распределения. — *Прим. перев.*

² В заключение s^2 нужно разделить на 10^{2p} , где p — степень, о которой говорилось в п. 2. — *Прим. перев.*

x	k	$x - a$	$k(x - a)$	$k(x - a)^2$
16	1	-12	-12	144
19	3	-9	-27	243
20	2	-8	-16	128
21	3	-7	-21	147
23	3	-5	-15	75
24	2	-4	-8	32
25	12	-3	-36	108
26	4	-2	-8	16
27	7	-1	-7	7
28	4	0	0	0
29	8	1	8	8
30	9	2	18	36
31	6	3	18	54
32	7	4	28	112
33	4	5	20	100
34	4	6	24	144
35	4	7	28	196
36	3	8	24	192
37	3	9	27	243
39	1	11	11	121
Сумма	90		56	2106

Поправка для выборочного среднего $56 : 90 = 0,62$
 Выборочное среднее $M = 28,62$

$$n(M - a)^2 = 35$$

$$\Sigma (x - M)^2 = 2106 - 35 = 2071$$

Выборочная дисперсия $s^2 = 2071 : 89 = 23$
 Выборочное квадратичное отклонение $s = 4,9$

§ 19. Поправки Шеппарда

Оба следующих раздела (§ 19 и 20) можно пропустить. Главная их цель — облегчить читателю знакомство с литературой, в которой зачастую идет речь о поправках Шеппарда, «вероятных ошибках» и т. п.

Наблюдаемые значения x_1, \dots, x_n часто округляют или группируют, т. е. объединяют в большие классы. Если класс с номером k содержит значения x_i , заключенные между $\tau_k - h/2$ и $\tau_k + h/2$, то τ_k называется *серединой класса*. Пусть n_k — количество наблюдаемых значений x в k -м интервале. Если в фор-

муле (3) § 18 все x_i заменить серединами соответствующих классов τ_k , то вместо M получим его приближенное значение

$$M' = \frac{1}{n} \sum n_k \tau_k, \quad (1)$$

и аналогично вместо s_0^2

$$s_0'^2 = \frac{1}{n} \sum n_k (\tau_k - \hat{x})^2. \quad (2)$$

M' может случайно оказаться несколько больше или несколько меньше, чем M , но разность $M' - M$ в среднем равна нулю; напротив, $s_0'^2$ в среднем несколько больше, чем s_0^2 . Для того чтобы в этом убедиться и определить поправку, которую нужно вычесть из $s_0'^2$, чтобы получить s_0^2 , мы предположим, что все интервалы имеют одинаковую длину h , а номера интервалов меняются от $-\infty$ до $+\infty$. Класс с номером 0 расположен между точками¹ $t - h/2$ и $t + h/2$, класс с номером k — между точками $t + kh - h/2$ и $t + kh + h/2$. Если, кроме того, мы выберем начало отсчета так, чтобы было $\hat{x} = 0$, то

$$M' = \frac{1}{n} \sum n_k (t + kh),$$

$$s_0'^2 = \frac{1}{n} \sum n_k (t + kh)^2.$$

Математические ожидания M' и $s_0'^2$ получаются из этих формул заменой n_k их математическими ожиданиями np_k . При этом

$$p_k = F\left(t + kh + \frac{1}{2}h\right) - F\left(t + kh - \frac{1}{2}h\right)$$

есть вероятность того, что x лежит между $t + kh - h/2$ и $t + kh + h/2$. Таким образом,

$$A(t) = \mathfrak{E}M' = \sum_{-\infty}^{+\infty} \left[F\left(t + kh + \frac{1}{2}h\right) - F\left(t + kh - \frac{1}{2}h\right) \right] (t + kh), \quad (3)$$

$$B(t) = \mathfrak{E}s_0'^2 = \sum_{-\infty}^{+\infty} \left[F\left(t + kh + \frac{1}{2}h\right) - F\left(t + kh - \frac{1}{2}h\right) \right] (t + kh)^2. \quad (4)$$

Оба выражения (3) и (4) являются периодическими функциями от t . А именно они переходят сами в себя при замене t на $t + h$.

В большинстве случаев эти периодические функции почти постоянны, т. е. они мало отличаются от постоянных составля-

¹ t является, таким образом, абсциссой середины нулевого класса. — Прим. ред.

ющих своих рядов Фурье и поэтому приближенно равны соответствующему интегральному среднему значению¹. Величиной интегрального среднего полезно интересоваться даже и в тех случаях, когда функции (3) и (4) значительно отличаются от постоянных. Действительно, выбор границ между классами $t + kh \pm \pm h/2$ довольно произволен и, в известном смысле, случаен, поэтому t можно рассматривать как случайную величину, которая равномерно распределена в интервале от 0 до h . В этом случае $A(t)$ будет также случайной величиной, среднее значение которой равно интегральному среднему

$$\hat{A} = \frac{1}{h} \int_0^h A(t) dt; \quad (5)$$

аналогично для $B(t)$

$$\hat{B} = \frac{1}{h} \int_0^h B(t) dt. \quad (6)$$

Если (3) подставим в (5), то получим

$$\begin{aligned} h\hat{A} &= \sum_{-\infty}^{+\infty} \int_0^h \left[F\left(t + kh + \frac{1}{2}h\right) - F\left(t + kh - \frac{1}{2}h\right) \right] (t + kh) dt = \\ &= \sum_{-\infty}^{+\infty} \int_{kh}^{(k+1)h} \left[F\left(t + \frac{1}{2}h\right) - F\left(t - \frac{1}{2}h\right) \right] t dt = \\ &= \int_{-\infty}^{+\infty} \left[F\left(t + \frac{1}{2}h\right) - F\left(t - \frac{1}{2}h\right) \right] t dt. \end{aligned} \quad (7)$$

Отсюда интегрированием по частям находим

$$\begin{aligned} h\hat{A} &= \int_{-\infty}^{+\infty} \frac{1}{2} t^2 d \left[F\left(t - \frac{1}{2}h\right) - F\left(t + \frac{1}{2}h\right) \right] = \\ &= \int_{-\infty}^{+\infty} \frac{1}{2} t^2 dF\left(t - \frac{1}{2}h\right) - \int_{-\infty}^{+\infty} \frac{1}{2} t^2 dF\left(t + \frac{1}{2}h\right) = \end{aligned}$$

¹ Так будет, если функция распределения удовлетворяет известным условиям. Строгий вывод поправок Шеппарда см. в книге Г. Крамера, Математические методы статистики, ИЛ, М., 1948, § 27.9. — *Прим. ред.*

$$\begin{aligned}
&= \int_{-\infty}^{+\infty} \frac{1}{2} \left(t + \frac{1}{2} h \right)^2 dF(t) - \int_{-\infty}^{+\infty} \frac{1}{2} \left(t - \frac{1}{2} h \right)^2 dF(t) = \\
&= \int_{-\infty}^{+\infty} \frac{1}{2} \left[\left(t + \frac{1}{2} h \right)^2 - \left(t - \frac{1}{2} h \right)^2 \right] dF(t) = \\
&= \int_{-\infty}^{+\infty} t h dF(t),
\end{aligned}$$

следовательно,

$$\hat{A} = \int_{-\infty}^{+\infty} t dF(t) = \hat{x} = 0,$$

и точно так же

$$\begin{aligned}
h\hat{B} &= \int_{-\infty}^{+\infty} \frac{1}{3} \left[\left(t + \frac{1}{2} h \right)^3 - \left(t - \frac{1}{2} h \right)^3 \right] dF(t) = \\
&= \int_{-\infty}^{+\infty} \left(t^2 h + \frac{1}{12} h^3 \right) dF(t) = \\
&= h \int_{-\infty}^{+\infty} t^2 dF(t) + \frac{1}{12} h^3
\end{aligned}$$

или

$$\hat{B} = \int_{-\infty}^{+\infty} t^2 dF(t) + \frac{1}{12} h^2 = \sigma^2 + \frac{1}{12} h^2.$$

Таким образом, математическое ожидание $s_0'^2$, осредненное по всем возможным значениям t , на $h^2/12$ больше математического ожидания s^2 , равного σ^2 . Для того чтобы для σ^2 получить несмещенную оценку, нужно из $s_0'^2$ вычесть поправку Шеннарда $h^2/12$.

Эту же самую поправку применяют также и при вычислении s^2 . Сначала вычисляют

$$s'^2 = \frac{1}{n-1} \sum n_k (\tau_k - M')^2$$

и из s'^2 вычитают $h^2/12$, при этом в среднем получается то же

самое s^2 , которое можно построить непосредственно по первоначальным значениям¹ x_i .

Пример 13. В § 18 говорилось, что при вычислении s^2 наблюдаемые значения x можно спокойно округлять таким образом, чтобы размах имел две значащие цифры. Теперь это утверждение должно быть подтверждено вычислением поправки Шеппарда. При этом мы предположим, что n не слишком велико и что случайная величина x распределена приблизительно нормально.

Пусть каждое x_i округлено ближайшим целым числом и пусть размах $W = x^{(n)} - x^{(1)}$ (т. е. разность между наибольшим и наименьшим наблюдениями) является двузначным числом

$$10 \leq W < 100.$$

Мы можем считать, что квадратичное отклонение σ больше чем 2. Так как если бы было $\sigma \leq 2$, то, с большой вероятностью, все x_i лежали бы в пределах $\hat{x} - 5$ и $\hat{x} + 5$ и, против предположения, размах был бы менее 10. С помощью формулы для функции распределения, выведенной в § 17, последнее заключение можно было бы сделать более точным, но для наших целей уже достаточна приближенная оценка.

Округление сводится к группировке по интервалам $(g - 1/2, g + 1/2)$, где g — целые числа. Математическое ожидание s^2 превосходит 4, а поправка Шеппарда равна лишь $1/12$. Следовательно, в среднем поправка Шеппарда составляет менее $1/48$ от s^2 , т. е. лишь 2%. Соответствующая величина поправки для s отличается от s в среднем менее чем на 1% и поэтому не имеет практического значения.

При очень больших n или в случае распределения, сильно отличающегося от нормального, эти отношения могут оказаться менее благоприятными. Поэтому при больших n округление, осторожности ради, не следует делать слишком грубо. Если размах W , вычисленный по округленным значениям, окажется меньше 20, то нужно округление произвести заново, сохранив дополнительно еще один десятичный знак. В этом случае W будет менее 200 и a всегда можно выбрать так, чтобы разности $x_i - a$ были двузначными числами, которые можно легко возводить в квадрат.

§ 20. Другие числовые характеристики распределения

Вместо среднего значения \hat{x} часто пользуются *медианой* ζ , которая для непрерывной функции распределения определяется как решение уравнения

$$F(\zeta) = \frac{1}{2}. \quad (1)$$

Для распределений с симметричной плотностью вероятности и, в частности, для нормального распределения медиана ζ равна среднему значению \hat{x} .

В качестве приближенного значения медианы ζ используют *выборочную медиану* Z . Если объем выборки является нечетным

¹ Если график плотности $F'(t) = f(t)$ не имеет соприкосновения высокого порядка с осью Ot на концах интервала, в котором сосредоточено распределение вероятностей, то лучше не применять поправок Шеппарда.
— Прим. ред.

числом $n = 2m - 1$, то Z равна среднему члену $x^{(m)}$ вариационного ряда $x^{(1)} < x^{(2)} < \dots < x^{(n)}$ (члены этого ряда являются элементами выборки x_1, x_2, \dots, x_n , расположенными в порядке возрастания их величины). Если же объем выборки равен четному числу $n = 2m$, то Z определяется как арифметическое среднее членов $x^{(m)}$ и $x^{(m+1)}$ вариационного ряда $Z = (x^{(m)} + x^{(m+1)})/2$.

Вычисление выборочной медианы осуществляется легче, чем вычисление выборочного среднего M . Однако в случае приближенно нормального распределения выборочное среднее M заслуживает большего доверия.

А именно если элементы выборки являются независимыми одинаково нормально распределенными случайными величинами, то, согласно § 17, выборочная медиана Z распределена приближенно нормально с дисперсией

$$\sigma_Z^2 = \frac{\pi}{2n} \sigma^2,$$

в то время как дисперсия выборочного среднего M равна

$$\sigma_M^2 = \frac{1}{n} \sigma^2.$$

Следовательно, дисперсия Z в $\pi/2$ раз больше дисперсии M .

Однако существуют функции распределений, для которых выборочная медиана Z точнее выборочного среднего M . В качестве примера рассмотрим функцию распределения Коши

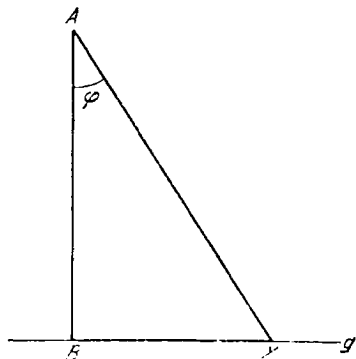


Рис. 14.

$$F(t) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arc} \operatorname{tg} t, \quad (2)$$

которой соответствует плотность вероятности

$$f(t) = \frac{1}{\pi(1+t^2)}. \quad (3)$$

Это распределение получается следующим образом. Пусть точка A находится на расстоянии, равном единице от фиксированной прямой g (рис. 14), и пусть через точку A проведена произвольная прямая. Если X — точка пересечения этой прямой с прямой g , B — проекция точки A на прямую g и φ — угол, образованный отрезками AX и AB , то расстояние от точки X до основания перпендикуляра B будет равно

$$x = \operatorname{tg} \varphi.$$

Если угол φ распределен равномерно в интервале между $-\pi/2$ и $\pi/2$ (т. е. φ является случайной величиной, плотность распре-

деления которой в указанном интервале постоянна и равна $1/\pi$), то $x = \operatorname{tg} \varphi$ имеет функцию распределения (2) и плотность вероятности (3).

Пусть x_1 и x_2 — независимые случайные величины с одинаковыми функциями распределения (2). Если по теореме III (§ 4) вычислить функцию распределения суммы $x_1 + x_2$, а затем — функцию распределения среднего

$$M_2 = \frac{1}{2} (x_1 + x_2),$$

то неожиданно окажется, что M_2 имеет плотность вероятности, в точности равную (3). Среднее из двух таких средних

$$M_4 = \frac{1}{4} (x_1 + x_2 + x_3 + x_4)$$

снова имеет ту же плотность и т. д. Следовательно, с помощью осреднения вообще нельзя добиться повышения точности¹.

Напротив, если по выборке x_1, \dots, x_n нечетного объема n определить выборочную медиану Z , то, согласно § 17, ее распределение будет приближенно нормальным со средним значением нуль и дисперсией

$$\sigma_Z^2 = \frac{\pi^2}{4n}.$$

Таким образом, с ростом n выборочная медиана становится все более и более точной оценкой для истинной медианы² $\zeta = 0$.

У распределения с плотностью вероятности (3) дисперсия и среднее значение не существуют, так как соответствующие интегралы расходятся.

Аналогично медиане ζ , обе *квартили* ζ_1 и ζ_3 определяются как решения уравнений

$$F(\zeta_1) = \frac{1}{4} \quad \text{и} \quad F(\zeta_3) = \frac{3}{4}.$$

¹ Для функции распределения (2) среднее значение \hat{x} не существует. Поэтому здесь идет речь об оценке медианы $\zeta = 0$. Так как M имеет то же распределение, что и каждый элемент выборки, то нет оснований считать M более точной оценкой для ζ , чем отдельное наблюдение x_i . — *Прим. перев.*

² В случае распределения (2) плотность вероятности выборочной медианы Z для выборки объема $n = 2m - 1$ задается формулой

$$f_Z(t) = \frac{(2m-1)!}{(m-1)!(m-1)!} \frac{1}{2^{2m-2}\pi} \left(1 - \frac{4}{\pi^2} \operatorname{arc} \operatorname{tg}^2 t\right)^{m-1} \frac{1}{1+t^2}.$$

Так как $|\operatorname{arc} \operatorname{tg} t| \sim \frac{\pi}{2} \left(1 - \frac{2}{\pi |t|}\right)$ при $t \rightarrow \pm \infty$, то $f_Z(t) = O(|t|^{-m-1})$. Следовательно, дисперсия Z существует, если $m \geq 3$, т. е. если $n \geq 5$. Аналогично математическое ожидание Z существует (и равно нулю), если $n \geq 3$. — *Прим. перев.*

Приближенными значениями для ζ_1 и ζ_3 служат *выборочные квартили* Z_1 и Z_3 , которые в общем случае определяются так. Пусть

$$q = \left\lfloor \frac{n}{4} \right\rfloor,$$

тогда Z_1 и Z_3 равны членам вариационного ряда с номерами $q + 1$ и $n - q$ соответственно: $Z_1 = x^{(q+1)}$ и $Z_3 = x^{(n-q)}$. Таким образом, найдутся q элементов выборки x_i , по величине меньших, чем Z_1 , и q элементов, по величине больших, чем Z_3 . Для $n = 4m - 1$ это определение совпадает с определением, указанным в § 17.

Квартилям родственна старомодная мера разброса, называемая *вероятным отклонением* w . В случае непрерывной функции распределения $F(t)$ вероятное отклонение w определяется условием, согласно которому случайная величина x с вероятностью $1/2$ должна принадлежать интервалу $(\zeta - w, \zeta + w)$:

$$F(\zeta + w) - F(\zeta - w) = \frac{1}{2}.$$

Если распределение симметрично, то $\zeta - w$ и $\zeta + w$ совпадают с квартилями:

$$\zeta_1 = \zeta - w, \quad \zeta_3 = \zeta + w.$$

Для нормального распределения

$$w = 0,6745 \sigma. \quad (4)$$

Вместо σ или s в старой литературе часто пользовались величиной w . Однако к настоящему времени от этого отказались. Дополнительный расчет оценки w' для w по формуле

$$w' = 0,6745 s$$

представляет собой излишнюю вычислительную операцию.

Наряду с σ и w третьей традиционной мерой разброса является *среднее отклонение* δ , которое определяется как математическое ожидание модуля $|x - \hat{x}|$:

$$\delta = \mathcal{E} |x - \hat{x}| = \int_{-\infty}^{\infty} |t - \hat{x}| dF(t). \quad (5)$$

Выборочным приближенным значением для δ является *выборочное среднее отклонение* d , которое определяется как арифметическое среднее абсолютных величин отклонений $x_i - M$:

$$d = \frac{1}{n} \sum |x_i - M|. \quad (6)$$

В случае нормального распределения отношение δ к σ равно некоторой постоянной. А именно если, путем изменения начала отсчета и изменения масштаба, случайную величину x нормировать так, чтобы после преобразования выполнялись условия $\hat{x} = 0$ и $\sigma = 1$, то

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

Поэтому для любого нормального распределения справедливо соотношение

$$\delta = \sigma \sqrt{\frac{2}{\pi}}. \quad (7)$$

Выборочное среднее отклонение d вычисляется несколько проще, чем выборочное квадратичное отклонение s , однако s является принципиально более важной и, как правило, более точной оценкой для σ , чем d .

Позднее мы увидим, что в случае нормального распределения s^2 является, в некотором определенном смысле, наилучшей оценкой для σ^2 .

ИНТЕГРАЛЫ ФУРЬЕ И ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ

§ 21. Характеристические функции

А. СРЕДНИЕ ЗНАЧЕНИЯ КОМПЛЕКСНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

До сих пор математическое ожидание было определено лишь для действительных случайных величин. Однако из двух действительных случайных величин x и y можно построить комплексную случайную величину

$$z = x + i y$$

и определить ее математическое ожидание формулой

$$\mathcal{E} z = \mathcal{E} x + i \mathcal{E} y.$$

Точно так же, в общем случае, определяется среднее значение вектора $v = (x_1, \dots, x_n)$:

$$\mathcal{E}(x_1, \dots, x_n) = (\mathcal{E} x_1, \dots, \mathcal{E} x_n).$$

Для двух произвольных случайных векторов из одного и того же векторного пространства или для двух комплексных случайных величин v и w справедливо равенство

$$\mathcal{E}(v + w) = \mathcal{E} v + \mathcal{E} w. \quad (1)$$

Две комплексные случайные величины $z = x + i y$ и $w = u + i v$ называются *независимыми*, если u и v независимы от x и y , т. е. если для любых a, b, c и d справедливо равенство

$$P(x < a, y < b, u < c, v < d) = P(x < a, y < b) P(u < c, v < d).$$

Если z и w независимы, то

$$\mathcal{E}(zw) = (\mathcal{E} z)(\mathcal{E} w). \quad (2)$$

Последнее равенство доказывается выделением действительной и мнимой частей zw с последующим использованием формулы (1).

Б. ХАРАКТЕРИСТИЧЕСКАЯ ФУНКЦИЯ

Пусть x — случайная величина с функцией распределения $F(u) = \mathbf{P}(x < u)$. *Характеристической функцией* случайной величины x называют математическое ожидание $\exp(itx)$:

$$\varphi(t) = \mathcal{E} e^{itx} = \int_{-\infty}^{\infty} e^{itu} dF(u). \quad (3)$$

Этот интеграл сходится для всех действительных t , так как абсолютные величины действительной и мнимой частей $\exp(itu)$ не превосходят единицы и $\int dF(u)$ сходится.

Если для $F(u)$ существует плотность вероятности $f(u)$, то характеристическая функция $\varphi(t)$ является обычным интегралом Фурье:

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itu} f(u) du. \quad (4)$$

С другой стороны, если x — дискретная случайная величина, принимающая значения x_1, x_2, \dots в конечном или счетном числе с вероятностями p_1, p_2, \dots , то

$$\varphi(t) = \sum p_k e^{itx_k}. \quad (5)$$

Легко видеть, что

$$\varphi(0) = 1$$

и

$$|\varphi(t)| \leq 1$$

для всех действительных t .

Если характеристическая функция случайной величины x равна $\varphi(t)$, то характеристическая функция $ax + b$ (a и b — постоянные) равна

$$\mathcal{E} e^{i(a x + b)} = e^{ib} \mathcal{E} e^{iatx} = e^{ib} \varphi(at).$$

В. НЕПРЕРЫВНОСТЬ ХАРАКТЕРИСТИЧЕСКОЙ ФУНКЦИИ

Одним из общих свойств интеграла Лебега является его непрерывная зависимость от подинтегральной функции, если она мажорируется функцией, интеграл Лебега от которой конечен. Соответствующая теорема гласит: Пусть $g_\nu(u)$ — последовательность интегрируемых функций (индекс ν можно заменить непрерывно меняющимся параметром t). Если для всех ν

$$|g_\nu(u)| \leq G(u), \quad (6)$$

интеграл $\int G(u) dF(u)$ конечен и для всех действительных u существует предел

$$\lim_{v \rightarrow \infty} g_v(u) = g(u), \quad (7)$$

то

$$\lim_{v \rightarrow \infty} \int_{-\infty}^{\infty} g_v(u) dF(u) = \int_{-\infty}^{\infty} g(u) dF(u). \quad (8)$$

Доказательство можно найти в теории интеграла Лебега.

Из этой теоремы непосредственно следует, что характеристическая функция $\varphi(t)$ непрерывна по t .

Г. МОМЕНТЫ

Моментом n -го порядка случайной величины \mathbf{x} называют математическое ожидание \mathbf{x}^n :

$$\alpha_n = \mathcal{E} \mathbf{x}^n = \int_{-\infty}^{\infty} u^n dF(u). \quad (9)$$

Обобщая это понятие, можно построить *моменты относительно точки c* :

$$\mathcal{E}(\mathbf{x} - c)^n = \int_{-\infty}^{\infty} (u - c)^n dF(u). \quad (10)$$

Особенно важное значение имеют моменты относительно математического ожидания $\hat{\mathbf{x}}$:

$$\mu_n = \mathcal{E}(\mathbf{x} - \hat{\mathbf{x}})^n = \int_{-\infty}^{\infty} (u - \hat{\mathbf{x}})^n dF(u). \quad (11)$$

Очевидно, что $\mu_0 = 1$ и $\mu_1 = 0$. μ_2 равен дисперсии:

$$\mu_2 = \sigma^2 = \mathcal{E}(\mathbf{x} - \hat{\mathbf{x}})^2.$$

Интегралы (9) сходятся лишь тогда, когда $F(u)$ при $u \rightarrow -\infty$ достаточно быстро стремится к нулю, а при $u \rightarrow +\infty$ достаточно быстро стремится к единице. Если для данного n интеграл (9) сходится, то говорят, что момент α_n *существует*. В этом случае

существуют моменты всех более низких порядков $\alpha_1, \dots, \alpha_{n-1}$, а также им соответствующие интегралы (10) и (11)¹.

Если α_n существует, то интеграл (3) можно n раз дифференцировать под знаком интеграла, причем

$$\varphi^{(n)}(t) = i^n \int_{-\infty}^{\infty} u^n e^{itu} dF(u). \quad (12)$$

Докажем справедливость этой формулы для первой производной; для остальных производных доказательство будет точно таким же. Рассмотрим отношение приращений

$$\frac{\varphi(t+h) - \varphi(t)}{h} = \int_{-\infty}^{\infty} e^{ihu} \frac{e^{ihu} - 1}{h} dF(u) \quad (13)$$

и устремим h к нулю. Предел дроби

$$\frac{e^{ihu} - 1}{h} = e^{\frac{1}{2}ihu} \frac{\sin \frac{1}{2}hu}{\frac{1}{2}hu} iu$$

равен iu и абсолютная величина дроби не превосходит $|u|$. Из теоремы непрерывности, сформулированной в разделе В этого параграфа, непосредственно вытекает, что

$$\varphi'(t) = i \int_{-\infty}^{\infty} u e^{itu} dF(u). \quad (14)$$

Так как α_n существует и подинтегральная функция в (12) непрерывна по t , то $\varphi^{(n)}(t)$ также непрерывна по t . Для $t = 0$ получаем

$$\varphi^{(n)}(0) = i^n \alpha_n. \quad (15)$$

¹ Пусть $0 < k < n$, тогда $|u - c|^k / |u|^n \rightarrow 0$ при $|u| \rightarrow \infty$, т. е. найдется такое U , что $|u - c|^k / |u|^n < 1$ для всех $|u| > U$. В силу очевидных неравенств

$$\begin{aligned} \int_{-\infty}^{\infty} |x - c|^k dF &= \int_{|u| < U} |u - c|^k dF + \int_{|u| > U} |u - c|^k dF \leq \\ &\leq \int_{|u| \leq U} |u - c|^k dF + \int_{|u| > U} |u|^n dF \leq \int_{|u| \leq U} |u - c|^k dF + \int_{-\infty}^{\infty} |u|^n dF. \end{aligned}$$

Так как оба последних интеграла конечны, то $\int |x - c|^k dF$ существует. Доказательство существования $\int x^k dF$ и $\int (x - \hat{x})^k dF$ получается заменой c на 0 или \hat{x} . — Прим. перев.

Следовательно, если моменты α_n существуют, то их можно найти посредством дифференцирования характеристической функции.

Если функция $\varphi(t)$ является аналитической в точке $t = 0$, то в некотором круге $|t| < r$ ее можно представить в виде ряда Тэйлора, коэффициенты которого определяются моментами:

$$\varphi(t) = \sum_0^{\infty} \frac{\alpha_n}{n!} (it)^n. \quad (16)$$

Как показал Крамер (Крамер Г., Математические методы статистики, ИЛ, М., 1948, стр. 199), для справедливости формулы (16) при $|t| < r$ достаточно, чтобы ряд $\sum \alpha_n r^n/n!$ сходиллся абсолютно.

д. ФОРМУЛЫ ОБРАЩЕНИЯ

Как известно, обращение интеграла Фурье (4) задается формулой

$$f(u) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T e^{-iu} \varphi(t) dt.$$

Если левую и правую части последнего равенства проинтегрировать от a до b , то получим

$$F(b) - F(a) = \lim_{T \rightarrow \infty} \frac{i}{2\pi} \int_{-T}^T \frac{e^{-itb} - e^{-ita}}{t} \varphi(t) dt. \quad (17)$$

Эту формулу можно записать и так:

$$F(u+h) - F(u-h) = \lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \frac{\sin ht}{t} e^{-iu} \varphi(t) dt. \quad (18)$$

Как показал Поль Леви, формулы (17) и (18) остаются справедливыми и в том случае, когда $F(u)$ не дифференцируема, а лишь непрерывна в точках a и b (соответственно в $u-h$ и $u+h$). Доказательство можно найти, например, в книге Г. Крамера «Математические методы статистики», стр. 109. Там же указана еще и другая формула обращения, а именно

$$\int_0^{\infty} [F(u+v) - F(u-v)] dv = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \cos ht}{t^2} e^{-iu} \varphi(t) dt. \quad (19)$$

Из формулы обращения следует, что

Функция распределения $F(u)$ однозначно определяется своей характеристической функцией $\varphi(t)$.

Однозначное определение $F(u)$ в ее точках непрерывности непосредственно вытекает из формулы (17). Если b — точка разрыва функции $F(u)$, то b можно представить в виде предела возрастающей последовательности точек непрерывности¹ b_ν , и так как $F(u)$ непрерывна слева в каждой точке, то справедливо равенство

$$F(b) = \lim_{\nu \rightarrow \infty} F(b_\nu). \quad (20)$$

Если, далее, a_ν — стремящаяся к $-\infty$ последовательность точек непрерывности, то

$$\lim_{\nu \rightarrow \infty} F(a_\nu) = 0. \quad (21)$$

Вычитая (21) из (20), получим, что

$$F(b) = \lim_{\nu \rightarrow \infty} [F(b_\nu) - F(a_\nu)]. \quad (22)$$

Из (22) следует однозначная определенность $F(b)$ для произвольного b .

Е. ХАРАКТЕРИСТИЧЕСКАЯ ФУНКЦИЯ СУММЫ

Пусть x и y — независимые случайные величины. Тогда, согласно (2),

$$\mathcal{E} e^{it(x+y)} = \mathcal{E} e^{itx} \mathcal{E} e^{ity},$$

или, словами: *характеристическая функция суммы независимых случайных величин равна произведению характеристических функций слагаемых.*

Это же самое справедливо, конечно, и для суммы $x_1 + \dots + x_n$ произвольного числа слагаемых.

Указанная теорема вместе с теоремой однозначности во многих случаях является очень удобным средством для отыскания функции распределения суммы независимых случайных величин. Это иллюстрируется ниже следующими примерами.

§ 22. Примеры

А. БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Пусть x_1, \dots, x_n — независимые случайные величины, каждая из которых может принимать значения 1 и 0 с вероятностями p и $q = 1 - p$ соответственно. В этом случае сумма $x_1 + \dots + x_n$

¹ Функция, монотонно возрастающая от 0 до 1, имеет не более чем счетное количество точек разрыва. Это можно доказать, например, используя то обстоятельство, что наша функция имеет лишь конечное число скачков ≥ 1 , конечное число скачков $\geq 1/2$, конечное число скачков $\geq 1/3$ и т. д. Следовательно, все скачки (а значит, и точки скачков) можно занумеровать в порядке их возрастания.

имеет биномиальное распределение: она принимает целочисленные значения k ($0 \leq k \leq n$) с вероятностями

$$W_k = \binom{n}{k} p^k q^{n-k}.$$

Согласно (5) § 21, характеристическая функция отдельного слагаемого x_j равна

$$\varphi(t) = pe^{it} + q. \quad (1)$$

Следовательно, характеристическая функция суммы задается формулой

$$[\varphi(t)]^n = (pe^{it} + q)^n. \quad (2)$$

Применяя к правой части (2) формулу бинома Ньютона, получим сумму, совпадающую с определением характеристической функции биномиального распределения

$$(pe^{it} + q)^n = \sum \binom{n}{k} p^k e^{ikt} q^{n-k} = \sum W_k e^{ikt}.$$

Б. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Если случайная величина x подчиняется нормальному распределению с единичной дисперсией и нулевым средним значением, т. е. если для x плотность вероятности задается формулой

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}, \quad (3)$$

то соответствующая характеристическая функция равна

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2 + itu} du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(u-it)^2 - \frac{1}{2}t^2} du. \quad (4)$$

Выбрав в качестве новой переменной интегрирования $w = u - it$, получим

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \int e^{-\frac{1}{2}w^2} dw, \quad (5)$$

где интегрирование производится в комплексной плоскости вдоль прямой, параллельной действительной оси. Если путь интегрирования перенести на действительную ось, то найдем, что

$$\varphi(t) = e^{-\frac{1}{2}t^2}. \quad (6)$$

Случайная величина σx также нормальна с дисперсией σ^2 и нулевым средним значением. Ее характеристическая функция равна

$$\mathcal{E} e^{it\sigma x} = \varphi(t\sigma) = e^{-\frac{1}{2} \sigma^2 t^2}. \quad (7)$$

Таким образом, характеристическая функция нормально распределенной случайной величины с нулевым средним значением лишь постоянным множителем отличается от гауссовой функции ошибок с дисперсией, равной обратной величине дисперсии исходного нормального распределения.

Если к случайной величине x прибавить постоянную a , то характеристическая функция $x + a$ будет равна произведению $\varphi(t)$ и $\exp(it a)$. Следовательно, для характеристической функции нормально распределенной случайной величины со средним значением a и квадратичным отклонением σ справедлива формула

$$\varphi(t) = e^{ita} e^{-\frac{1}{2} \sigma^2 t^2}. \quad (8)$$

Произведение двух функций этого вида опять будет функцией того же вида. Этим самым мы снова получили найденный ранее результат, но со значительно меньшим количеством выкладок:

Сумма двух независимых нормально распределенных случайных величин снова распределена нормально.

В. РАСПРЕДЕЛЕНИЕ ПУАССОНА

Если x принимает значения $k = 0, 1, 2, \dots$ с вероятностями (§ 10)

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}, \quad (9)$$

то соответствующая характеристическая функция, согласно (5), равна

$$\varphi(t) = \sum_0^{\infty} p_k e^{itk} = e^{-\lambda} \sum_0^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it} - 1)}. \quad (10)$$

Произведение двух таких функций с параметрами λ_1 и λ_2 снова является функцией вида (10) с параметром $\lambda = \lambda_1 + \lambda_2$. Отсюда следует, что

Сумма двух независимых случайных величин x_1 и x_2 , подчиняющихся распределению Пуассона со средними значениями λ_1 и λ_2 , снова имеет распределение Пуассона со средним значением $\lambda_1 + \lambda_2$.

§ 23. Распределение χ^2

В связи с гауссовой теорией ошибок астроном Ф. Р. Хельмерт исследовал суммы квадратов нормально распределенных случайных величин и при этом пришел к функции распределения $G(u)$, которую позднее К. Пирсон назвал функцией распределения χ^2 . Для отрицательных u функция $G(u) = 0$, а для неотрицательных u

$$G(u) = \alpha \int_0^u y^{\lambda-1} e^{-\frac{1}{2}y} dy, \quad (1)$$

где $\lambda = f/2$ и f — натуральное число, которое, по Р. А. Фишеру, называется *числом степеней свободы*. Множитель α определяется так, чтобы выполнялось равенство $G(\infty) = 1$, т. е.

$$\alpha = \frac{1}{2^\lambda \Gamma(\lambda)}. \quad (2)$$

Соответствующая плотность вероятности задается формулой

$$g(u) = \alpha u^{\lambda-1} e^{-\frac{1}{2}u} \quad (u > 0). \quad (3)$$

В простейшем случае $\lambda = 1$ (две степени свободы) и плотность вероятностей является показательной функцией

$$g(u) = \frac{1}{2} e^{-\frac{1}{2}u} \quad (u > 0). \quad (4)$$

Случай $\lambda = 1/2$ (одна степень свободы) получается прямо из нормального распределения, согласно следующей теореме:

Если случайная величина x распределена нормально с нулевым средним значением и единичной дисперсией, то x^2 подчиняется распределению χ^2 с одной степенью свободы.

Для доказательства этой теоремы нужно лишь вычислить вероятность события $x^2 < u$. Она равна

$$G(u) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{u}}^{\sqrt{u}} e^{-\frac{1}{2}z^2} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{u}} e^{-\frac{1}{2}z^2} dz.$$

Если теперь в качестве новой переменной интегрирования выбрать $y = z^2$, то немедленно получим искомый результат

$$G(u) = \frac{1}{\sqrt{2\pi}} \int_0^u y^{-\frac{1}{2}} e^{-\frac{1}{2}y} dy. \quad (5)$$

Перейдем теперь к остальным случаям. Характеристическая функция распределения χ^2 имеет вид

$$\varphi(t) = \int_0^{\infty} \alpha u^{\lambda-1} e^{-\frac{1}{2}u + itu} du. \quad (6)$$

Выбирая в качестве новой переменной интегрирования $v = (\frac{1}{2} - it)u$, получим

$$\varphi(t) = 2^\lambda (1 - 2it)^{-\lambda} \alpha \int v^{\lambda-1} e^{-v} dv, \quad (7)$$

где интегрирование производится в комплексной плоскости v вдоль полупрямой, которая расположена в правой полуплоскости и уходит из нуля в бесконечность. Уравнение этой прямой

$$v = \left(\frac{1}{2} - it\right)u,$$

где t фиксировано, а u возрастает от 0 до ∞ . Этот путь интегрирования можно перевести в положительную часть действительной оси; значение интеграла от этого не изменится. Следовательно, интеграл в (7) равен $\Gamma(\lambda)$, поэтому

$$\varphi(t) = \frac{1}{(1 - 2it)^\lambda}. \quad (8)$$

Первый и второй моменты распределения χ^2 можно легко вычислить либо по определению момента

$$\alpha_n = \int_0^{\infty} u^n dG(u) = \int_0^{\infty} u^n g(u) du,$$

либо по формуле (15) предыдущего параграфа. Мы получим

$$\begin{aligned} \alpha_1 &= 2\lambda = f, \\ \alpha_2 &= 4\lambda(\lambda + 1) = f^2 + 2f. \end{aligned}$$

Отсюда находим среднее значение и дисперсию случайной величины y , подчиняющейся распределению χ^2 :

$$\mathcal{E} y = \alpha_1 = f, \quad (9)$$

$$\sigma^2 = \mathcal{E} y^2 - (\mathcal{E} y)^2 = \alpha_2 - \alpha_1^2 = 2f. \quad (10)$$

Пусть теперь y и z — независимые случайные величины, подчиняющиеся распределению χ^2 с степенями f и f' свободы соответственно. Согласно (8), характеристические функции y и z равны

$$(1 - 2it)^{-\frac{1}{2}f} \text{ и } (1 - 2it)^{-\frac{1}{2}f'}.$$

Их произведение снова является характеристической функцией того же самого вида. Отсюда следует, что

Если две независимые случайные величины y и z подчиняются распределению χ^2 с f и f' степенями свободы соответственно, то их сумма $y + z$ имеет распределение χ^2 с $f + f'$ степенями свободы.

Справедливость этой теоремы можно также проверить прямым вычислением интеграла

$$h(v) = \int_0^v g_1(u) g_2(v - u) du \quad (11)$$

по формуле (7) § 4 Б. Вычисление сведется к бета-функции. Таким образом, можно избежать применения характеристических функций; правда, при этом будет больше выкладок.

Само собой разумеется, что эта теорема остается справедливой и для сумм $y_1 + \dots + y_n$, где $n > 2$. В применении к сумме квадратов нормально распределенных случайных величин с нулевыми средними значениями и единичными дисперсиями эта теорема гласит:

Если x_1, x_2, \dots, x_n — независимые нормально распределенные случайные величины с нулевыми средними значениями и единичными дисперсиями, то сумма квадратов

$$\chi^2 = x_1^2 + x_2^2 + \dots + x_n^2 \quad (12)$$

подчиняется распределению χ^2 с n степенями свободы.

Таким образом, мы пришли к результату Хельмерта, сформулированному в начале этого раздела. Если дисперсия равна не единице, а σ^2 , то для того, чтобы получить распределение χ^2 , следует вместо (12) положить

$$\chi^2 = \frac{x_1^2 + \dots + x_n^2}{\sigma^2}. \quad (13)$$

§ 24. Предельные теоремы

А. ПРЕДЕЛЬНАЯ ТЕОРЕМА ЛЕВИ—КРАМЕРА

Из формул обращения характеристической функции (§ 21 Д) следует предельная теорема:

Если последовательность характеристических функций $\varphi_1(t)$, $\varphi_2(t)$, \dots при каждом t стремится к пределу $\varphi(t)$, непрерывному в точке $t = 0$, то $\varphi(t)$ является характеристической функцией некоторой функции распределения $F(u)$, причем последовательность функций распределения $F_1(u)$, $F_2(u)$, \dots сходится к $F(u)$ во всех тех точках u , где функция $F(u)$ непрерывна.

Доказательство этой теоремы имеется в книге Г. Крамера «Математические методы статистики», стр. 112.

Если $F(u)$ — функция распределения, и для всех тех точек u , где $F(u)$ непрерывна, имеет место соотношение $\lim_{n \rightarrow \infty} F_n(u) = F(u)$, то в дальнейшем мы будем кратко говорить: *последовательность F_n стремится к F* .

Б. ПРИМЕР ИСПОЛЬЗОВАНИЯ ПРЕДЕЛЬНОЙ ТЕОРЕМЫ:
БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Характеристической функцией биномиального распределения является

$$\varphi(t) = (pe^{it} + q)^n. \quad (1)$$

Среднее значение случайной величины $x = x_1 + \dots + x_n$ равно np , а квадратичное отклонение $\sigma = \sqrt{npq}$. Если

$$\frac{x - np}{\sigma} \quad (2)$$

ввести в качестве новой случайной величины, то соответствующая характеристическая функция будет задаваться формулой

$$\varphi_n(t) = e^{-\frac{itnp}{\sigma}} (pe^{\frac{it}{\sigma}} + q)^n. \quad (3)$$

Логарифм $\varphi_n(t)$ равен

$$\ln \varphi_n(t) = -\frac{itnp}{\sigma} + n \ln [1 + p(e^{\frac{it}{\sigma}} - 1)]. \quad (4)$$

Если t фиксировано и $n \rightarrow \infty$, то it/σ стремится к нулю; показательную функцию в круглых скобках можно разложить в степенной ряд

$$e^{\frac{it}{\sigma}} - 1 = \frac{it}{\sigma} - \frac{1}{2} \left(\frac{t}{\sigma}\right)^2 + \dots$$

Произведение $p [\exp(it/\sigma) - 1]$ для достаточно больших n мало сравнительно с единицей, следовательно, логарифм в правой части (4) можно также разложить в степенной ряд:

$$\ln [1 + p(e^{\frac{it}{\sigma}} - 1)] = \frac{itp}{\sigma} - \frac{1}{2} pq \left(\frac{t}{\sigma}\right)^2 + \dots$$

Подставляя этот результат в (4), получим

$$\ln \varphi_n(t) = -\frac{npq}{2} \left(\frac{t}{\sigma}\right)^2 + \dots = -\frac{1}{2} t^2 + \dots \quad (5)$$

Если теперь n устремим к бесконечности, то в пределе.

$$\lim_{n \rightarrow \infty} \varphi_n(t) = e^{-\frac{1}{2} t^2}. \quad (6)$$

Правая часть (6) является характеристической функцией нормального распределения. Отсюда следует, что функция распределения нормированной случайной величины (2) с нулевым средним значением и единичной дисперсией при $n \rightarrow \infty$ сходится к функции нормального распределения. Этот результат можно сформулировать и так: *случайная величина x асимптотически нормальна со средним значением np и квадратичным отклонением σ .*

Мы уже знакомы с этим результатом, однако только что изложенный вывод требует меньше вычислений. Точно таким же способом с помощью характеристической функции можно без труда получить дополнительный член порядка $1/\sqrt{n}$ (см. (1) § 6).

В. ЗАКОН БОЛЬШИХ ЧИСЕЛ

Закон больших чисел в § 5 был сформулирован так: если вероятность некоторого события равна p , то частота h наступления этого события в n независимых испытаниях для достаточно больших n с вероятностью, сколь угодно близкой к единице, будет произвольно мало отличаться от p . Это же самое можно сформулировать и так: *частота h сходится по вероятности к p при $n \rightarrow \infty$.* Или еще: *при $n \rightarrow \infty$ частота h является состоятельной оценкой для p .* Все эти высказывания означают одно и то же, а именно — что для всякого $\varepsilon > 0$ вероятность события $|h - p| < \varepsilon$ сколь угодно близка к единице, если n достаточно велико.

Случайная величина h была определена как отношение x/n , где

$$x = x_1 + \dots + x_n \quad (7)$$

является суммой независимых случайных величин, каждая из которых принимает значения 1 или 0 с вероятностями p и $q = 1 - p$ соответственно. Закон больших чисел можно обобщить, выбрав в качестве x_1, \dots, x_n какие-либо независимые случайные величины с одной и той же функцией распределения $F(u)$. При этом, согласно Хинчину, на $F(u)$ накладывается лишь требование существования конечного среднего значения

$$a = \int_{-\infty}^{\infty} u dF(u). \quad (8)$$

По Дюге, достаточно даже более слабое предположение о существовании конечной производной от характеристической функции x_1 в точке $t = 0$:

$$\varphi'(0) = ia. \quad (9)$$

Согласно Хинчину и Дюге, обобщенный закон больших чисел гласит:

Если x_1, \dots, x_n — независимые случайные величины с одной и той же функцией распределения и если выполняется условие (9), то арифметическое среднее

$$m = \frac{1}{n} (x_1 + \dots + x_n) \quad (10)$$

сходится по вероятности к a при $n \rightarrow \infty$.

Доказательство чрезвычайно просто. В некоторой окрестности точки $t = 0$ характеристическая функция $\varphi(t)$ близка к единице, следовательно, мы можем в этой окрестности положить

$$\varphi(t) = e^{\psi(t)}. \quad (11)$$

Характеристическая функция арифметического среднего m равна

$$\left[\varphi\left(\frac{t}{n}\right) \right]^n = e^{n\psi\left(\frac{t}{n}\right)} =: e^{t \frac{n}{t} \psi\left(\frac{t}{n}\right)}. \quad (12)$$

Так как в точке $t = 0$ функция $\varphi(t)$ дифференцируема, то дифференцируема также и $\psi(t)$, причем ее производная равна

$$\psi'(0) = \frac{\varphi'(0)}{\varphi(0)} = ia. \quad (13)$$

По определению производной $(n/t)\psi(t/n)$ стремится к $\psi'(0)$ при $n \rightarrow \infty$, следовательно, (12) при $n \rightarrow \infty$ имеет предел

$$e^{t\psi'(0)} = e^{iat}. \quad (14)$$

Но правая часть (14) является характеристической функцией величины a , которая с достоверностью принимает лишь одно значение a . Таким образом, характеристическая функция m при каждом t стремится к характеристической функции постоянной величины a . Функцией распределения a является такая функция $E(u)$, которая в точке a совершает скачок от 0 к 1 и в дальнейшем с ростом u остается равной единице. Согласно предельной теореме, функция распределения m стремится к этой функции $E(u)$ во всех тех точках u , где $E(u)$ непрерывна. Следовательно, функция распределения m стремится к нулю при $u < a$ и стремится к единице при $u > a$. Этот результат в точности совпадает с утверждением теоремы.

Только что доказанную теорему называют «слабым законом больших чисел». Наряду с этим законом существует еще «усиленный закон больших чисел», но в математической статистике он едва ли играет заметную роль. [См. Хишчин А. Я., Sur la loi des grands nombres, Comptes Rendus de l'Acad. des Sciences, Paris, 188 (1929), 477, а также Гнеденко Б. В. и Колмогоров А. И., Предельные распределения для сумм независимых случайных величин, ГИТТЛ, М., 1949, стр. 112 и 142.]

Г. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

Случайная величина x , функция распределения которой зависит от параметра n , называется *асимптотически нормальной*, если существуют два числа a и c (может быть, зависящие от n) такие, что функция распределения случайной величины

$$\frac{x-a}{c} \quad (15)$$

стремится к нормированной функции нормального распределения $\Phi(u)$ при $n \rightarrow \infty$. Согласно разделу А этого параграфа, необходимым и достаточным условием асимптотической нормальности x является сходимость характеристической функции случайной величины (15) к характеристической функции нормального распределения

$$e^{-\frac{1}{2}t^2} \quad (16)$$

при каждом t .

Во многих случаях a является средним значением, а c — квадратичным отклонением случайной величины x , однако может случиться, что квадратичное отклонение x не существует или даже среднее значение не существует и, несмотря на это, имеются числа a и c с упомянутыми свойствами.

В разделе Б мы видели, что при n опытах, в каждом из которых положительный исход наступает с одной и той же вероятностью p , общее число положительных исходов $x = x_1 + \dots + x_n$ асимптотически нормально. При этом каждое слагаемое x_j принимает лишь значения 1 и 0 с вероятностями p и $q = 1 - p$ соответственно.

Смысл содержания центральной предельной теоремы заключается в том, что при определенных условиях каждая сумма независимых случайных величин

$$x = x_1 + \dots + x_n \quad (17)$$

распределена асимптотически нормально.

Уже Лаплас и Гаусс предполагали справедливость этой теоремы и указали основания для своих догадок. Первое полное доказательство принадлежит А. М. Ляпунову (1901)¹. Поль Леви привлек для доказательства характеристические функции. Позднее Хинчин, Леви и Феллер доказали эту теорему при значительно более слабых предположениях. (По этому вопросу см. Levy P., *Théorie de l'addition des variables aléatoires*, Paris, 1954; Хинчин А. Я., *Предельные теоремы для сумм независимых слу-*

¹ Первое строгое доказательство центральной предельной теоремы дал с помощью метода моментов А. А. Марков (1898). При этом он следовал по пути, указанному П. Л. Чебышевым. — *Прим. ред.*

чайных величин, ГОНТИ, 1938; Гнеденко Б. В. и Колмогоров А. Н., Предельные распределения для сумм независимых случайных величин, ГИТТЛ, М., 1949.)

Определенные ограничения, которые в центральной предельной теореме накладываются на распределения слагаемых (17), нужны для того, чтобы, во-первых, исключить тот случай, когда величина отдельного слагаемого составляет слишком большую часть всей суммы (17), и, во-вторых, чтобы обеспечить достаточно быстрое стремление функций распределения x_j к нулю и к единице при $u \rightarrow -\infty$ и $u \rightarrow +\infty$ соответственно. Если, например, все отдельные слагаемые x_j подчиняются распределению Коши (§ 20), то сумма x имеет то же самое распределение и центральная предельная теорема не выполняется. Линдберг (*Math. Zeitschr.*, 15, 1922) указал довольно слабое условие, достаточное для выполнимости центральной предельной теоремы, однако условия Феллера (*Math. Zeitschr.*, 40 und 42) еще слабее, так как Феллер не требует даже конечности дисперсий¹.

Мы не будем здесь входить в эти тонкости и рассмотрим лишь случай, когда все x_j одинаково распределены и имеют конечное среднее значение и конечную дисперсию. Мы докажем, таким образом, теорему:

Если x_1, \dots, x_n — независимые одинаково распределенные случайные величины, имеющие среднее значение μ и квадратичное отклонение σ , то сумма (17) асимптотически нормальна со средним значением $n\mu$ и квадратичным отклонением $\sigma\sqrt{n}$.

Доказательство. Мы можем предположить, что $\mu = 0$. Пусть $\varphi(t)$ — характеристическая функция случайной величины x_1 , тогда $[\varphi(t)]^n$ — характеристическая функция случайной величины x .

Нам нужно доказать, что

$$\left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n \quad (18)$$

стремится к $\exp(-t^2/2)$ при $n \rightarrow \infty$.

Первая и вторая производные от $\varphi(z)$ в точке $z = 0$ равны соответственно $i\mu = 0$ и $i^2\sigma^2 = -\sigma^2$. Следовательно, для $\varphi(z)$ справедлива формула Тэйлора:

$$\varphi(z) = 1 - \frac{1}{2}\sigma^2 z^2 + R,$$

¹ См. монографию Б. В. Гнеденко и А. Н. Колмогорова, указанную выше, а также статью А. Н. Колмогорова, Некоторые работы последних лет в области предельных теорем теории вероятностей, Вест. Моск. ун-та, № 10 (1953), 29. — *Прим. перев.*

где остаточный член R при $z \rightarrow 0$ стремится к нулю быстрее, чем z^2 . Из последнего равенства получаем, что

$$\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + R, \quad (19)$$

причем R при $n \rightarrow \infty$ стремится к нулю быстрее, чем $1/n$. Логарифм характеристической функции равен

$$\ln \varphi\left(\frac{t}{\sigma\sqrt{n}}\right) = -\frac{t^2}{2n} + R', \quad (20)$$

где R' снова стремится к нулю быстрее, чем $1/n$. Если (20) умножить на n , то найдем логарифм (18). Устремляя затем n к бесконечности, получим в пределе $-t^2/2$, следовательно,

$$\lim_{n \rightarrow \infty} \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n = e^{-\frac{1}{2}t^2}, \quad (21)$$

что и требовалось доказать.

д. ПРИМЕР: РАСПРЕДЕЛЕНИЕ χ^2

Для суммы квадратов независимых одинаково нормально распределенных случайных величин с нулевым средним значением и единичной дисперсией

$$\chi^2 = x_1^2 + x_2^2 + \dots + x_n^2 \quad (22)$$

выполнены все предположения только что доказанной теоремы. Среднее значение x_1^2 равно единице, а квадратичное отклонение равно $\sqrt{2}$. Следовательно, сумма (22) асимптотически нормальна со средним значением n и квадратичным отклонением $\sqrt{2n}$ (ср. § 23). Так как квадратичное отклонение мало по сравнению со средним значением, то случайная величина $\sqrt{2}\chi^2$ распределена также асимптотически нормально. Нормальное приближение для распределения $\sqrt{2}\chi^2$ лучше, чем для самого распределения χ^2 (см. Fisher R. A., *Statistical Methods*, § 20). Среднее значение $\sqrt{2}\chi^2$ приближенно равно $\sqrt{2n} - 1$, а дисперсия близка к единице¹.

¹ По поводу асимптотической нормальности функций от асимптотически нормальных случайных величин см. Крамер Г., *Математические методы статистики*, ИЛ, М., 1948, стр. 401. Можно указать также функции от χ^2 , распределение которых существенно лучше приближается нормальным распределением, чем распределения $(\chi^2 - n)/\sqrt{2n}$ или $\sqrt{2}\chi^2 - \sqrt{2n} - 1$. Примером такой функции является функция $(\chi^2/n)^{1/3}$, распределенная асимптотически нормально со средним значением $1 - 2/(9n)$ и дисперсией $2/(9n)$ [см. Wilson E. B. and Hilferty M. M., *Nat. Acad. Sci. USA*, 17 (1931), 694]. — *Прим. перев.*

Е. ВТОРАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

Очень полезна также «вторая предельная теорема» Фреше и Шохата¹, которая гласит:

Если все $F_n(t)$ из последовательности функций распределения $\{F_n(t)\}$ обладают конечными моментами $\alpha_k(n)$ любого порядка и если $\alpha_k(n) \rightarrow \beta_k$ ($n \rightarrow \infty$) при каждом k , то β_k — моменты некоторой функции распределения $F(t)$. Если, кроме того, $F(t)$ своими моментами определяется однозначно, то при $n \rightarrow \infty$ последовательность $F_n(t)$ сходится к $F(t)$ в каждой точке непрерывности функции $F(t)$.

Доказательство этой теоремы можно найти в цитированном исследовании Фреше и Шохата или в книге М. G. Kendall, *Advanced Theory of Statistics*, vol. I, Griffin and Co, London, 1948, § 4.24.

Наиболее важен тот случай, когда β_k являются моментами функции нормального распределения $\Phi(t)$:

$$\beta_{2r-1} = 0, \quad \beta_{2r} = \frac{(2r)!}{2^r r!} \quad r = 1, 2, \dots \quad (23)$$

Функция нормального распределения всюду непрерывна и однозначно определяется своими моментами. Следовательно,

Если $\alpha_k(n)$ при $n \rightarrow \infty$ стремятся к моментам нормального распределения (23), то в каждой точке t последовательность $F_n(t)$ сходится к $\Phi(t)$.

Ж. ОДНА ЭЛЕМЕНТАРНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

Доказательства всех предыдущих предельных теорем были основаны на интегральном преобразовании Фурье. Однако следующая теорема будет совсем элементарной. Формулировка этой теоремы заимствована мною из книги Г. Крамера, «Математические методы статистики», ИЛ, М., 1948, стр. 281, 20. 6.

Пусть x_1, x_2, \dots — последовательность случайных величин с функциями распределения F_1, F_2, \dots . Предположим, что $F_n(u)$ стремится к функции распределения $F(u)$ при $n \rightarrow \infty$.

Пусть, далее, y_1, y_2, \dots другая последовательность случайных величин. Предположим, что y_n сходится по вероятности к некоторой постоянной c . Тогда функция распределения суммы

$$z_n = x_n + y_n \quad (24)$$

¹ Впервые эта теорема была доказана А. А. Марковым (1898) для случая сходимости к нормальному распределению (см. Марков А. А., Исчисление вероятностей, 4-е изд., ГИЗ, 1924, стр. 514). Доказательство формулируемой здесь общей теоремы дано в работе Fréchet M. and Shohat J., A Proof of the Generalized Second Limit Theorem, Trans. Amer. Math. Soc., 33 (1931), 533. — Прим. перев.

при $n \rightarrow \infty$ стремится к $F(u - c)$. Если $c > 0$, то с соответствующими очевидными изменениями это утверждение справедливо для произведения $x_n y_n$ и отношения x_n / y_n .

Важно отметить, что в этой теореме не требуется независимости входящих в нее случайных величин.

Мы проведем доказательство для сумм (24). В случае произведений или отношений доказательство остается аналогичным.

Пусть u — точка непрерывности функции $F(u - c)$. Тогда для каждого $\varepsilon > 0$ найдется такое $\delta = \delta(\varepsilon, u)$, что $F(u - c - \delta)$ и $F(u - c + \delta)$ будут отличаться от $F(u - c)$ менее чем на ε . Так как, согласно сделанному ранее замечанию (см. сноску в § 21 Д), множество точек разрыва функции F не более чем счетно, то мы можем так выбрать δ , чтобы точки $u - c + \delta$ и $u - c - \delta$ являлись точками непрерывности функции F .

Пусть $G_n(u)$ — вероятность события $z_n < u$. Нам нужно доказать, что $G_n(u)$ при $n \rightarrow \infty$ стремится к $F(u - c)$.

Если $x_n < u - c - \delta$ и $y_n \leq c + \delta$, то $z_n < u$. Отсюда следует, что если $x_n < u - c - \delta$, то или $z_n < u$ или $y_n > c + \delta$. Следовательно,

$$\mathbf{P}(x_n < u - c - \delta) \leq \mathbf{P}(z_n < u) + \mathbf{P}(y_n > c + \delta)$$

или

$$F_n(u - c - \delta) \leq G_n(u) + \mathbf{P}(y_n > c + \delta). \quad (25)$$

Так как y_n стремится по вероятности к c , то вероятность события $y_n > c + \delta$ для достаточно больших n будет меньше ε . Таким образом, из (25) получается, что

$$F_n(u - c - \delta) < G_n(u) + \varepsilon.$$

Поскольку F_n стремится к F , то для всех достаточно больших n справедливо неравенство

$$F(u - c - \delta) < G_n(u) + 2\varepsilon$$

и, далее,

$$F(u - c) < G_n(u) + 3\varepsilon. \quad (26)$$

Точно так же, поменяв ролями x_n и z_n , можно доказать, что

$$G_n(u) < F_n(u - c + \delta) + \varepsilon < F(u - c) + 3\varepsilon. \quad (27)$$

Из (26) и (27) следует равенство $\lim G_n(u) = F(u - c)$, чем и завершается доказательство предельной теоремы.

§ 25. Прямоугольное распределение. Ошибки округления

Случайная величина x называется равномерно распределенной между a и b , если ее плотность вероятности равна постоянной в интервале (a, b) и равна нулю вне этого интервала:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{если } a < x < b, \\ 0, & \text{если } x < a \text{ или } x > b. \end{cases} \quad (1)$$

Так как график функции $f(x)$ изображается в виде прямоугольника (рис. 15), то такое распределение называют *прямо-*



Р и с. 15. График плотности прямоугольного распределения.

угольным. Вопрос о том, определена ли $f(x)$ в конечных точках интервала a и b и если определена, то как именно, не имеет никакого значения. Функция прямоугольного распределения задается равенствами:

$$F(x) = \begin{cases} 0, & \text{если } x \leq a, \\ \frac{x-a}{b-a}, & \text{если } a < x \leq b, \\ 1, & \text{если } b < x. \end{cases} \quad (2)$$

Среднее значение x равно $(a+b)/2$, дисперсия $(b-a)^2/12$. Для большей определенности мы предположим, что $a = -1/2$ и $b = 1/2$. В этом случае x_1 равномерно распределена между $-1/2$ и $+1/2$ с нулевым средним значением и дисперсией $1/12$.

Такое распределение встречается тогда, когда результаты числовых расчетов округляются до целых чисел. Если точные результаты вычислений зависят от случая и изменяются в широких границах с плотностью вероятности, которая в интервале длины единица не сильно отличается от постоянной, то ошибки округления будут распределены приблизительно равномерно между $-1/2$ и $+1/2$.

Характеристическая функция прямоугольного распределения равна

$$\varphi(t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{itx} dx = \frac{2}{t} \sin \frac{t}{2}. \quad (3)$$

Рассмотрим теперь распределение суммы независимых одинаково равномерно распределенных в интервале $(-\frac{1}{2}, \frac{1}{2})$ случайных величин:

$$\mathbf{x} = \mathbf{x}_1 + \dots + \mathbf{x}_n. \quad (4)$$

Среднее значение суммы равно нулю, квадратичное отклонение равно $\sqrt{n}/12$. Если \mathbf{x} нормировать таким образом, чтобы квадратичное отклонение нормированной величины равнялось единице а затем n устремить к бесконечности, то характеристическая функция нормированной случайной величины $\mathbf{x} \sqrt{12/n}$

$$\varphi_n(t) = \left[\varphi \left(t \sqrt{\frac{12}{n}} \right) \right]^n \quad (5)$$

будет очень быстро приближаться к гауссовой функции $\exp(-t^2/2)$. Таким образом, нужно ожидать, что функция распределения \mathbf{x} очень быстро приближается к функции нормального распределения.

Это подтверждается расчетами. Если $f_n(x)$ — плотность вероятности суммы \mathbf{x} , то справедливы рекуррентные формулы:

$$f_1(x) = \begin{cases} 1, & \text{если } -\frac{1}{2} < x < \frac{1}{2}, \\ 0, & \text{если } x < -\frac{1}{2} \text{ или } x > \frac{1}{2}, \end{cases} \quad (6)$$

$$f_n(x) = \int_{x-\frac{1}{2}}^{x+\frac{1}{2}} f_{n-1}(u) du. \quad (7)$$

С помощью этих формул находим

$$f_2(x) = \begin{cases} x+1, & \text{если } -1 < x \leq 0, \\ (x+1) - 2x, & \text{если } 0 \leq x < 1, \end{cases} \quad (8)$$

$$f_3(x) = \begin{cases} \frac{1}{2} \left(x + \frac{3}{2} \right)^2, & \text{если } -\frac{3}{2} < x \leq -\frac{1}{2}, \\ \frac{1}{2} \left[\left(x + \frac{3}{2} \right)^2 - 3 \left(x + \frac{1}{2} \right)^2 \right], & \text{если } -\frac{1}{2} \leq x \leq \frac{1}{2}, \\ \frac{1}{2} \left[\left(x + \frac{3}{2} \right)^2 - 3 \left(x + \frac{1}{2} \right)^2 + 3 \left(x - \frac{1}{2} \right)^2 \right], & \text{если } \frac{1}{2} \leq x < \frac{3}{2}. \end{cases} \quad (9)$$

и т. д., вообще

$$f_n(x) = \frac{1}{(n-1)!} \left[\left(x + \frac{n}{2}\right)^{n-1} - \binom{n}{1} \left(x + \frac{n}{2} - 1\right)^{n-1} + \binom{n}{2} \left(x + \frac{n}{2} - 2\right)^{n-1} + \dots \right], \quad (10)$$

причем в сумму (10) входят все те слагаемые $x + n/2$, $x + n/2 - 1$, \dots , которые при данном x неотрицательны. Графики функций f_2 , f_3 и f_4 изображены на рис. 16, 17 и 18. Распределение с плот-

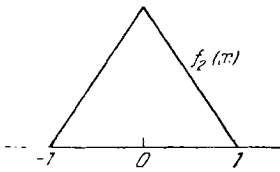


Рис. 16.

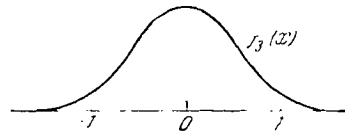


Рис. 17.

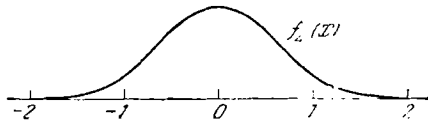


Рис. 18.

ностью $f_2(x)$ называют «треугольным распределением»: соответствующий график представляет собой равнобедренный треугольник, вершина которого имеет абсциссу $x = 0$, а основанием является отрезок $[-1, +1]$. График плотности $f_3(x)$ составлен из трех отрезков квадратичных парабол и очень похож на кривую Гаусса. Кривая, соответствующая $f_4(x)$, почти не отличима от кривой Гаусса.

Пример 14. «Основной аргумент» z в Сатурновых таблицах Хилла¹ представляет собой сумму 24 членов, каждый из которых получают посредством интерполяции из некоторой четырехзначной таблицы. Если отдельные члены были бы вычислены более точно, т. е. с 5 или 6 десятичными знаками, и лишь сумма округлена до четырех знаков, то насколько этот результат отличался бы от z , вычисленного с помощью четырехзначных таблиц?

Умножением всех слагаемых на 10^4 можно добиться того, чтобы табличные значения были целыми числами. Если между двумя табличными значениями y_n и y_{n+1} производится линейная интерполяция по формуле

$$y = y_n + x(y_{n+1} - y_n) = y_n(1 - x) + y_{n+1}x \quad (0 \leq x < 1) \quad (11)$$

¹ Astron. Papers Amer. Ephemeris, VIII (1898), 145—285.

и если y_n и y_{n+1} имеют ошибки округления u и v соответственно, то ошибка y равна

$$w = u(1 - x) + vx. \quad (12)$$

Предположим теперь, что u , v и x — независимые случайные величины, причем u и v равномерно распределены между $-1/2$ и $+1/2$, а x равномерно распределена между 0 и 1. В этом случае w имеет нулевое среднее значение и дисперсию

$$\sigma^2 = \mathcal{E} w^2 = \int_0^1 dx \int_{-\frac{1}{2}}^{\frac{1}{2}} du \int_{-\frac{1}{2}}^{\frac{1}{2}} dv [u(1-x) + vx]^2. \quad (13)$$

Если в (13) сначала произвести интегрирование по u и v , а затем — по x , то получим

$$\sigma^2 = \int_0^1 \left[\frac{1}{12} (1-x)^2 + \frac{1}{12} x^2 \right] dx = \frac{1}{18}. \quad (14)$$

Дисперсия ошибки линейной интерполяции в таблицах с двойным входом будет еще меньше. Однако так как у Хилла по таблицам с двойным входом получаются лишь три слагаемых из двадцати четырех, то точное вычисление соответствующих дисперсий не имеет никакого значения.

В некоторых таблицах y_n на больших промежутках сохраняют постоянные значения. В этом случае u и v уже не являются независимыми, и результат (14) не имеет места. Ошибка интерполяции оказывается приближенно равной ошибке округления табличных значений, поэтому здесь

$$\sigma^2 \sim \frac{1}{12}. \quad (15)$$

В большинстве таблиц линейная интерполяция недопустима, и нужно пользоваться квадратичной интерполяцией, например по формуле

$$y = y_n + (y_{n+1} - y_n) x + \frac{y_{n+1} - 2y_n + y_{n-1}}{2} x(x-1) \quad \left(-\frac{1}{2} < x \leq \frac{1}{2} \right).$$

Тогда с помощью вычислений, аналогичных указанным выше, получим несколько большую дисперсию, а именно

$$\sigma^2 = \frac{47}{384} = 0,12 \dots \quad (16)$$

Если квадратичная интерполяция применялась в 15 случаях, то, вычисляя дисперсию 15 слагаемых по формуле (16), а остальных девяти — по формуле (15), мы найдем для дисперсии суммы z оценку сверху. Это лишь увеличит надежность наших выводов. По центральной предельной теореме сумма 24 случайных величин распределена почти нормально с дисперсией 2,58. К этой дисперсии нужно еще добавить дисперсию ошибки округления точной суммы, равную $1/12$. Таким образом, разность между точным

значением z , округленным до целых единиц, и тем значением z , которое получается с помощью четырехзначных таблиц слагаемых, представляет собой практически нормально распределенную случайную величину с дисперсией

$$\sigma^2 = 2,67.$$

Возвращаясь снова к четырехзначным числам, получаем, что квадратичное отклонение σ этой разности округленно равно $1,6 \cdot 10^{-4}$. Следовательно, общая ошибка суммы, превышающая 4 единицы четвертого десятичного знака, будет встречаться лишь чрезвычайно редко.

Если бы мы сложили теоретически возможные максимальные ошибки всех отдельных слагаемых, то получили бы 14 единиц четвертого десятичного знака. Теоретически возможная ошибка суммы пропорциональна числу слагаемых m , а практическая ошибка пропорциональна \sqrt{m} .

ГАУССОВА ТЕОРИЯ ОШИБОК И КРИТЕРИЙ СТЬЮДЕНТА

В этой главе предполагается известным содержание гл. I—IV, а также центральная предельная теорема и понятие о распределении χ^2 из гл. V.

§ 26. Гауссова теория ошибок

А. РАВНОТОЧНЫЕ НАБЛЮДЕНИЯ

Повторные измерения физической величины, даже если эта величина в процессе измерений остается постоянной, не всегда дают одинаковые результаты: наблюдаемые значения x имеют некоторый разброс около среднего значения \hat{x} . Разность $x - \hat{x}$, по Гауссу, называется *случайной ошибкой* одного наблюдения.

Среднее значение \hat{x} не обязательно равняется истинному значению измеряемой величины, так как наблюдения могут содержать *систематическую ошибку*, которая возникает как следствие принятого метода измерений. При известных условиях влияние систематической ошибки можно понизить улучшением измерительной аппаратуры или добавлением к результатам измерений надлежащих поправок. Вопрос устранения систематических ошибок выходит за рамки статистической теории ошибок, которая посвящена лишь случайным ошибкам $x - \hat{x}$.

В теории ошибок всегда предполагается, что случайная величина x имеет конечные среднее значение \hat{x} и квадратичное отклонение σ . Иногда делается еще дополнительное предположение о законе распределения ошибок; однако сначала мы постараемся выяснить, как далеко можно продвинуть эту теорию без определенных предположений о распределении ошибок.

Следуя Гауссу, квадратичное отклонение σ называют *средней ошибкой* наблюдения. Осреднением многих наблюдений среднюю ошибку можно понизить. А именно, согласно § 18, выборочное среднее из n независимых наблюдений

$$M = \bar{x} = \frac{1}{n} \sum x_i \tag{1}$$

имеет то же среднее значение \hat{x} , что и отдельные наблюдения x_i , и средняя ошибка M в \sqrt{n} раз меньше средней ошибки x_i :

$$\sigma_M = \frac{\sigma}{\sqrt{n}}. \tag{2}$$

Для того чтобы можно было судить о точности M как оценки для \hat{x} , нужно найти приближенное значение для σ^2 . Как показано в § 18, в качестве приближенного значения принимают

$$s^2 = \frac{1}{n-1} \sum (x_i - M)^2. \quad (3)$$

Для дисперсии выборочного среднего M

$$\sigma_M^2 = \frac{\sigma^2}{n} \quad (4)$$

соответствующее приближенное значение равно

$$s_M^2 = \frac{s^2}{n}. \quad (5)$$

Если количество наблюдений n велико, то s_M является хорошим приближением для средней ошибки σ_M . Найденное выборочное среднее M и его выборочную среднюю ошибку s_M обычно объединяют в выражение $M \pm s_M$.

Пример 15. Повторные определения широты Капштадта в течение периода с 1892 по 1894 г. (Szuber, Wahrscheinlichkeitsrechnung, Beispiel LVI) дали 15 результатов, указанных в таблице.

Отбрасывая градусы и минуты, найдем выборочное среднее

$$M = \frac{48'',92}{15} = 3'',261$$

и в качестве округленного начала отсчета выберем $a = 3'',26$. При вычислении s^2 за новую единицу принята $0'',01$ (§ 18,2). Поправочный член $-n(M-a)^2$ в данном случае столь мал, что им можно пренебречь. Таким образом, находим

$$s^2 = \frac{3406}{14} = 243,$$

следовательно, $s = 16$;

$$s_M^2 = \frac{243}{15} = 16,$$

следовательно, $s_M = 4$.

Результат можно записать в виде символического равенства:

$$\varphi = -33^\circ 56' 3'',26 \pm 0'',04.$$

	x	$x - a$	$(x - a)^2$
1	$-33^\circ 56' 3'',48$	+22	484
2	$3'',50$	+24	576
3	$3'',50$	+24	576
4	$3'',32$	+6	36
5	$3'',09$	-17	289
6	$2'',98$	-28	784
7	$3'',07$	-19	361
8	$3'',28$	+2	4
9	$3'',27$	+1	1
10	$3'',20$	-6	36
11	$3'',30$	+4	16
12	$3'',25$	-1	1
13	$3'',11$	-15	225
14	$3'',30$	+4	16
15	$3'',27$	+1	1
	$48'',92$	+2	3406

Б. НЕРАВНОТОЧНЫЕ НАБЛЮДЕНИЯ

Если отдельные наблюдения имеют различную точность, то при образовании среднего их естественно снабдить различными весами и вместо (1) построить взвешенное выборочное среднее, которое мы снова обозначим буквой M :

$$M = g_1 x_1 + \dots + g_n x_n. \quad (6)$$

При этом сумма весов должна быть равна единице:

$$g_1 + \dots + g_n = 1. \quad (7)$$

Среднее значение и дисперсия M равны

$$\mathcal{E} M = g_1 \hat{x}_1 + \dots + g_n \hat{x}_n, \quad (8)$$

$$\sigma_M^2 = g_1^2 \sigma_1^2 + \dots + g_n^2 \sigma_n^2. \quad (9)$$

Если все x_i имеют одно и то же среднее значение \hat{x} , что мы и будем впредь всегда предполагать, то (8) превращается в равенство $\mathcal{E} M = \hat{x}$.

Спрашивается, каким образом нужно выбрать веса g_i , чтобы средняя ошибка σ_M была наименьшей?

Из (7) и (9) следует, что

$$\sigma_M^2 = g_1^2 \sigma_1^2 + \dots + g_{n-1}^2 \sigma_{n-1}^2 + (1 - g_1 - \dots - g_{n-1})^2 \sigma_n^2.$$

Как известно, этот квадратичный относительно g_1, \dots, g_{n-1} многочлен достигает в некоторой точке минимума, если все его частные производные по g_1, \dots, g_{n-1} в этой точке обращаются в нуль. Дифференцируя по g_1 , находим

$$2g_1 \sigma_1^2 - 2(1 - g_1 - \dots - g_{n-1}) \sigma_n^2 = 0$$

или

$$g_1 \sigma_1^2 = g_n \sigma_n^2.$$

Точно так же, дифференцируя по g_2 , найдем

$$g_2 \sigma_2^2 = g_n \sigma_n^2$$

и т. д. Это означает, что веса g_1, \dots, g_n должны быть обратно пропорциональны квадратам средних ошибок отдельных наблюдений.

При вычислениях дополнительное условие (7) удобно отбросить, заменив g величинами, пропорциональными весам. Тогда вместо (6) нужно написать

$$M = \frac{g_1 x_1 + \dots + g_n x_n}{g_1 + \dots + g_n} = \frac{\sum g x}{\sum g}. \quad (10)$$

g_i должны быть обратно пропорциональны дисперсиям σ_i^2 :

$$g_1 : g_2 : \dots : g_n = \sigma_1^{-2} : \sigma_2^{-2} : \dots : \sigma_n^{-2}. \quad (11)$$

Положим теперь

$$g_i \sigma_i^2 = \sigma^2 \quad (12)$$

и назовем σ «средней ошибкой наблюдения на единицу веса». В данном случае вместо (9) получим

$$\sigma_M^2 = \frac{\sum g_i^2 \sigma_i^2}{(\sum g_i)^2}$$

или, если воспользоваться (12),

$$\sigma_M^2 = \frac{\sigma^2}{\sum g} \quad (13)$$

По результатам наблюдений можно вычислить приближенное значение для σ^2 , а именно

$$s^2 = \frac{1}{n-1} \sum g_i (x_i - M)^2 \quad (14)$$

Оправданием формулы (14) служит то, что среднее значение s^2 равно σ^2 . Доказательство вполне аналогично приведенному ранее доказательству формулы (10) в § 18. Для упрощения вычислений мы предварительно выберем начало отсчета на оси Ox так, чтобы выполнялось условие $\hat{x} = 0$. В этом случае имеем

$$\begin{aligned} \mathcal{E} \sum g_i (x_i - M)^2 &= \mathcal{E} (\sum g_i x_i^2 - 2 \sum g_i x_i M + \sum g_i M^2) = \\ &= \sum g \mathcal{E} x_i^2 - \mathcal{E}(M^2) \sum g_i = \\ &= \sum g_i \sigma_i^2 - \sigma_M^2 \sum g_i = n \sigma^2 - \sigma^2 = (n-1) \sigma^2, \end{aligned} \quad (15)$$

следовательно, согласно (14),

$$\mathcal{E} s^2 = \sigma^2 \quad (16)$$

В силу (13), для дисперсии σ_M^2 взвешенного выборочного среднего M получаем приближенное значение

$$s_M^2 = \frac{s^2}{\sum g} \quad (17)$$

Формулу (17) можно применять лишь тогда, когда известно, что веса правильны, т. е. обратно пропорциональны дисперсиям σ_i^2 . Если веса заменены их приближенными оценками, то при выводах надо соблюдать осторожность.

При вычислении s^2 можно M опять заменить каким-либо близким числом a и затем из результата вычесть $(\sum g)(M - a)^2$:

$$s^2 = \frac{1}{n-1} [\sum g(x-a)^2 - (\sum g)(M-a)^2] \quad (18)$$

Пример 16. Для определения периода колебаний физического маятника фиксировались 20 последовательных моментов прохождения этого маятника в одном и том же направлении через положение равновесия. Пусть t_1, \dots, t_{20} — зафиксированные моменты времени. С помощью этих двадцати наблюдений можно построить десять независимых оценок для периода T , а именно:

$$T_1 = \frac{1}{19} (t_{20} - t_1),$$

$$T_2 = \frac{1}{17} (t_{19} - t_2),$$

$$T_{10} = (t_{11} - t_{10}).$$

Если все разности $(t_i - t_k)$ имеют одинаковую среднюю ошибку σ , то T_1, \dots, T_{10} имеют средние ошибки

$$\frac{\sigma}{19}, \quad \frac{\sigma}{17}, \quad \dots, \quad \frac{\sigma}{1}$$

соответственно.

Поэтому в качестве весов можно выбрать числа

$$g_1 = 19^2, \quad g_2 = 17^2, \dots, \quad g_{10} = 1^2.$$

Таким образом, взвешенное выборочное среднее равно

$$M = \frac{19^2 T_1 + 17^2 T_2 + \dots + 1^2 T_{10}}{19^2 + 17^2 + \dots + 1^2}.$$

Средняя ошибка σ_M взвешенного среднего M задается формулой (13):

$$\sigma_M^2 = \frac{\sigma^2}{19^2 + 17^2 + \dots + 1^2}.$$

Для оценки σ^2 можно воспользоваться формулой (14):

$$s^2 = \frac{1}{9} [19^2 (T_1 - M)^2 + 17^2 (T_2 - M)^2 + \dots + 1^2 (T_{10} - M)^2].$$

Однако эта оценка не очень точна, так как в нее входят лишь 10 разностей $t_{20} - t_1, t_{19} - t_2, \dots, t_{11} - t_{10}$. Если использовать все 20 наблюдений и вычислить квадраты их отклонений от «линии регрессии», которая, в некотором смысле, наилучшим образом сглаживает наблюдаемые точки, то с помощью суммы квадратов таких отклонений можно вывести лучшую оценку для средней ошибки отдельного наблюдения t_i , а значит, и лучшую оценку для средней ошибки разностей $t_i - t_k$. К этому мы вернемся в § 32 и 33.

V. ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ ВЫБОРОЧНОГО СРЕДНЕГО M

До сих пор теория не зависела ни от каких специальных предположений о функции распределения случайной величины x . Гаусс обычно предполагал, что рассматриваемые случайные величины распределены нормально. Для обоснования этого предположения Гаусс выдвинул «гипотезу элементарных ошибок», которая гласит, что общая ошибка наблюдения является суммой большого числа независимых малых ошибок, порождаемых раз-

личными причинами и обладающих малыми дисперсиями. Таким образом, справедлива центральная предельная теорема (§ 24 Г):

Сумма очень большого количества независимых случайных величин, дисперсия каждой из которых составляет лишь малую часть от дисперсии всей суммы, имеет приближенно нормальное распределение.

Если, следуя Гауссу, предположить, что все x_i подчиняются нормальному распределению, то их сумма, а значит, и выборочное среднее M , будут тоже иметь нормальное распределение. Поэтому имеет место правило:

Абсолютная величина разности $M - \hat{x}$ с вероятностью 0,95 меньше чем $1,96 \sigma_M$ и с вероятностью 0,99 меньше чем $2,58 \sigma_M$, где σ_M — средняя ошибка M .

При больших n указанные правила справедливы даже тогда, когда величины x_1, \dots, x_n имеют распределение, отличное от нормального. Это вызвано тем, что выборочное среднее M является суммой множества слагаемых, каждое из которых обладает лишь относительно малым квадратичным отклонением. Следовательно, согласно центральной предельной теореме, распределение M близко к нормальному распределению. Это тем более верно, если уже сами отдельные x_i имеют приближенно нормальное распределение.

Применение сформулированных выше правил требует знания средней ошибки σ_M . Можно ли, применяя эти правила, вместо истинной средней ошибки $\sigma_M = \sigma/\sqrt{n}$ воспользоваться ее выборочным приближенным значением $s_M = s/\sqrt{n}$? Для ответа на этот вопрос мы должны сначала исследовать величину отклонения s^2 от σ^2 или, иными словами, найти функцию распределения случайной величины s^2 .

§ 27. Распределение s^2

А. СВЯЗЬ С РАСПРЕДЕЛЕНИЕМ χ^2

Пусть x_1, \dots, x_n — независимые одинаково нормально распределенные случайные величины со средним значением $\hat{x} = a$ и квадратичным отклонением¹ σ , и пусть снова

¹ Если x_1, \dots, x_n имеют различные квадратичные отклонения $\sigma_1, \dots, \sigma_n$, то вместо x_i можно ввести новые величины

$$x'_i = \frac{x_i - a}{\sigma_i},$$

которые распределены одинаково нормально с нулевым средним значением и единичной дисперсией.

$$M = \frac{1}{n} \sum x_i, \quad (1)$$

$$s^2 = \frac{1}{n-1} \sum (x_i - M)^2. \quad (2)$$

Какова функция распределения случайной величины s^2 ? Вместо s^2 удобнее рассматривать величину

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{1}{\sigma^2} \sum (x_i - M)^2,$$

которая не изменяется при изменении масштаба на оси Ox . Заменой x на $(x - a)/\sigma$ мы всегда можем добиться, чтобы среднее значение и квадратичное отклонение удовлетворяли условиям: $\hat{x} = 0$ и $\sigma = 1$. В этом случае

$$\chi^2 = \sum (x_i - M)^2 = \sum x_i^2 - nM^2 = \sum x_i^2 - \frac{1}{n} (\sum x_i)^2, \quad (3)$$

и плотность вероятности для каждого x_i равна

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

Так как x_1, \dots, x_n независимы, то их совместная плотность вероятности равна произведению

$$f(t_1, \dots, t_n) = f(t_1) f(t_2) \dots f(t_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(t_1^2 + \dots + t_n^2)}.$$

Отсюда, по теореме II § 4, следует, что искомая функция распределения $G(u) = \mathbf{P}(\chi^2 < u)$ равна n -кратному интегралу¹

$$\begin{aligned} G(u) &= \int \dots \int_{\chi^2 < u} f(x_1, \dots, x_n) dx_1 \dots dx_n = \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} \int \dots \int_{\chi^2 < u} e^{-\frac{1}{2}(x_1^2 + \dots + x_n^2)} dx_1 \dots dx_n. \end{aligned} \quad (4)$$

¹ Обозначение переменных интегрирования и случайных величин одними и теми же буквами x_1, \dots, x_n логически не корректно, но удобно. Такие функции от x_i , как M, χ^2 и появляющиеся в следующем разделе y_1, \dots, y_n , носят двойкий смысл: как случайные величины и как функции от переменных интегрирования.

Если y_1, \dots, y_n считать случайными величинами, то результат (6) можно получить еще более простым путем. Плотность вероятности для y та же самая, что и для x :

$$(2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2} \sum y^2} = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2} \sum x^2},$$

так как $\sum x^2 = \sum y^2$. Следовательно, y_1, \dots, y_n — независимые одинаково нормально распределенные случайные величины с нулевым средним значением и единичной дисперсией. Случайная величина $\chi^2 = y_1^2 + \dots + y_n^2$ не зависит от y_1 , поэтому ее функция распределения может быть найдена интегрированием только по y_2, \dots, y_n . В этом и заключается результат (6).

Как было доказано в § 23, функция распределения суммы квадратов $y_2^2 + \dots + y_n^2$ является функцией распределения χ^2 с $f = n - 1$ степенями свободы:

$$G(u) = \alpha \int_0^u v^{\lambda-1} e^{-\frac{1}{2}v} dv, \quad \alpha = \frac{1}{2^\lambda \Gamma(\lambda)}, \quad \lambda = \frac{f}{2} = \frac{n-1}{2}. \quad (7)$$

Ранее этот результат был очень просто получен с помощью «характеристической функции». Однако интеграл (6) можно вычислить и независимо от предшествующих результатов, воспользовавшись преобразованием к полярным координатам (§ 11). Первая из полярных координат, обозначаемая обычно буквой r , в нашем случае называется χ , так как $\chi^2 = y_2^2 + \dots + y_n^2$. Таким образом, получаем

$$G(u) = (2\pi)^{-\lambda} \int \dots \int_{\chi^2 < u} e^{-\frac{1}{2}\chi^2} \chi^{n-2} d\chi d\Omega.$$

Так как условие $\chi^2 < u$ не зависит от угловых координат, то можно произвести интегрирование по $d\Omega$:

$$G(u) = (2\pi)^{-\lambda} \int_0^{\sqrt{u}} d\Omega \int e^{-\frac{1}{2}\chi^2} \chi^{n-2} d\chi.$$

В силу результатов § 12, а также в силу того, что $\lambda = (n - 1)/2$,

$$\int d\Omega = \frac{2}{\Gamma(\lambda)} \pi^\lambda,$$

следовательно,

$$2^\lambda \Gamma(\lambda) G(u) = 2 \int_0^{\sqrt{u}} e^{-\frac{1}{2}\chi^2} \chi^{n-2} d\chi.$$

Если теперь в качестве новой переменной интегрирования выбрать $v = \chi^2$, то получим в точности ту же формулу, что и (7).

Интеграл в правой части (7) является *неполной гамма-функцией*. Соответствующая плотность вероятностей равна

$$g(u) = \alpha u^{\lambda-1} e^{-\frac{1}{2}u} \quad \text{для } u > 0. \quad (8)$$

в. НЕЗАВИСИМОСТЬ M И χ^2

Тем же методом можно определить вероятность одновременного осуществления двух событий: $\chi^2 < u$ и $b \leq M < c$, где b и c — произвольные числа, удовлетворяющие условию $b < c$:

$$P = \int_{\chi^2 < u} \dots \int f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

$$b \leq M < c$$

Действительно, если ввести то же самое ортогональное преобразование координат, которое было указано выше, то получим произведение двух интегралов: в первом интеграле интегрирование производится по переменной y_1 в пределах от $b\sqrt{n}$ до $c\sqrt{n}$, а во втором — по области интегрирования $\chi^2 < u$, не зависящей от y_1 :

$$P = (2\pi)^{-\frac{1}{2}} \int_{b\sqrt{n}}^{c\sqrt{n}} e^{-\frac{1}{2}y_1^2} dy_1 (2\pi)^{-\lambda} \int_{\chi^2 < u} e^{-\frac{1}{2}x^2} dy_2 \dots dy_n =$$

$$= P(b \leq M < c) P(\chi^2 < u).$$

Таким образом, при любых u и $b < c$ вероятность одновременного осуществления двух событий $b \leq M < c$ и $0 \leq \chi^2 < u$ равна произведению вероятностей этих событий. Это и означает, что *случайные величины M и χ^2 независимы*.

Следовательно, совместная плотность вероятности пары случайных величин (M, χ^2) равна произведению плотностей вероятностей M и χ^2 . Плотность вероятности M нормальна с квадратичным отклонением σ_M , а плотность вероятности χ^2 задается формулой (8).

г. СРЕДНЕЕ ЗНАЧЕНИЕ И ДИСПЕРСИЯ χ^2

Среднее значение и дисперсия случайной величины $Q = \chi^2$ были уже указаны в § 23; их можно легко получить непосредственно по формуле (8):

$$\mathcal{E} Q = \alpha \int_0^{\infty} u^{\lambda-1} e^{-\frac{1}{2}u} du = \frac{1}{2^{\lambda} \Gamma(\lambda)} 2^{\lambda+1} \Gamma(\lambda+1) = 2\lambda = f = n-1,$$

$$\begin{aligned} \mathcal{E} Q^2 &= \alpha \int_0^{\infty} u^2 u^{\lambda-1} e^{-\frac{1}{2}u} du = \frac{1}{2^{\lambda} \Gamma(\lambda)} 2^{\lambda+2} \Gamma(\lambda+2) = \\ &= 4\lambda(\lambda+1) = f^2 + 2f, \\ \sigma_Q^2 &= \mathcal{E} Q^2 - (\mathcal{E} Q)^2 = 2f = 2(n-1). \end{aligned} \quad (9)$$

Ранее было установлено, что

$$\chi^2 = \frac{f s^2}{\sigma^2}. \quad (10)$$

Поэтому среднее значение s^2 равно σ^2 (этот результат был получен в § 26 (16)), а квадратичное отклонение s^2 равно

$$\sigma_{(s^*)} = \frac{\sigma^2}{f} \sqrt{2f} = \sigma^2 \sqrt{\frac{2}{n-1}}. \quad (11)$$

д. ДОВЕРИТЕЛЬНЫЕ ГРАНИЦЫ ДЛЯ s^*

Значения функции $G(u)$ можно определять с помощью существующих таблиц неполной гамма-функции¹. Эти таблицы позволяют найти такую границу K , для которой событие $\chi^2 < K$ имеет заданную вероятность. Если положим

$$G(K) = \mathbf{P}(\chi^2 < K) = 1 - \beta$$

и выберем, например, $\beta = 0,01$, то K будет являться верхней границей для χ^2 , причем χ^2 лишь изредка будет превышать границу K . Если же положим

$$G(K') = \mathbf{P}(\chi^2 < K') = \beta,$$

то получим нижнюю границу K' , причем лишь в редких случаях χ^2 будет меньше этой границы.

Случайные величины χ^2 и s^2/σ^2 связаны соотношением (10), следовательно, верхняя граница для s^2/σ^2 определяется верхней границей K для χ^2 , т. е. при заданном σ^2 величина $\sigma^2 K/f$ является верхней границей для s^2 , а при заданном s^2 величина $s^2 f/K$ является нижней границей для σ^2 . Точно так же из нижней границы K' получается верхняя граница $s^2 f/K'$ для σ^2 . Вероятность¹ нарушения каждого из этих неравенств в отдельности $s^2 f/K < \sigma^2$ и $s^2 f/K' > \sigma^2$, равна β .

¹ См., например, С л у ц к и й Е. Е., Таблицы для вычисления неполной Γ -функции и функции вероятностей χ^2 , изд. АН СССР, М., 1950. — *Прим. перев.*

Е. АДДИТИВНОСТЬ СЛУЧАЙНЫХ ВЕЛИЧИН χ^2

В § 23 была доказана теорема: *если две независимые случайные величины подчиняются распределению χ^2 с f и g степенями свободы, то сумма этих величин имеет распределение χ^2 с $f + g$ степенями свободы.* Эту теорему можно легко доказать непосредственно, не пользуясь характеристическими функциями. Случайные величины χ_1^2 и χ_2^2 имеют те же функции распределения, что и суммы

$$y_1^2 + \dots + y_f^2 \quad \text{и} \quad y_{f+1}^2 + \dots + y_{f+g}^2$$

независимых одинаково нормально распределенных случайных величин y_1, \dots, y_{f+g} с нулевым средним значением и единичной дисперсией. Следовательно, $\chi_1^2 + \chi_2^2$ имеет такую же функцию распределения, как и сумма

$$y_1^2 + \dots + y_{f+g}^2,$$

т. е. $\chi_1^2 + \chi_2^2$ подчиняется распределению χ^2 с $f + g$ степенями свободы.

§ 28. Критерий Стьюдента

Вернемся к вопросу, поставленному в конце § 26. Можно ли, применяя правило

$$\frac{|M - \hat{x}|}{\sigma_M} < 1,96, \text{ с вероятностью } 0,95,$$

или

$$\frac{|M - \hat{x}|}{\sigma_M} < 2,58, \text{ с вероятностью } 0,99,$$

заменить в знаменателе σ_M на s_M ?

Для ответа на этот вопрос нужно найти функцию распределения для отношения

$$t = \frac{M - \hat{x}}{s_M}. \quad (1)$$

Прежде всего упростим задачу.

Смещением начала отсчета на оси Ox можно добиться равенства $\hat{x} = 0$, и в этом случае

$$t = \frac{M}{s_M}. \quad (2)$$

При изменении масштаба на оси Ox отношение t остается неизменным, поэтому мы можем считать, что $\sigma = 1$. Если числитель

и знаменатель (2) умножить на \sqrt{n} , то, в обозначениях § 27, получим

$$t = \frac{M\sqrt{n}}{s_M\sqrt{n}} = \frac{y_1}{s}. \quad (3)$$

Умножая в (3) числитель и знаменатель на $\sqrt{f} = \sqrt{n-1}$ и принимая во внимание равенство $fs^2 = \chi^2$, найдем

$$t = \frac{y_1}{\chi} \sqrt{f}. \quad (4)$$

Таким образом, задача отыскания функции распределения $H(a)$ случайной величины t сводится к вычислению вероятности события

$$\frac{y_1}{\chi} \sqrt{f} < a. \quad (5)$$

Если мы положим

$$c = \frac{a}{\sqrt{f}}. \quad (6)$$

то неравенство (5) упростится:

$$y_1 < c\chi. \quad (7)$$

В § 27 В было доказано, что y_1 и χ^2 независимы, следовательно, плотность вероятностей пары случайных величин (y_1, χ^2) равна произведению плотностей вероятности y_1 и χ^2 :

$$f(y)g(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \alpha z^{\frac{1}{2}f-1} e^{-\frac{1}{2}z}, \quad (8)$$

$$\alpha = \frac{1}{2^{\frac{f}{2}} \Gamma\left(\frac{f}{2}\right)}.$$

Поэтому искомая вероятность имеет вид

$$\begin{aligned} H(a) &= \alpha' \iint_{y < c\sqrt{z}} e^{-\frac{1}{2}y^2} z^{\frac{1}{2}f-1} e^{-\frac{1}{2}z} dy dz = \\ &= \alpha' \int_0^{\infty} z^{\frac{1}{2}f-1} e^{-\frac{1}{2}z} dz \int_{-\infty}^{c\sqrt{z}} e^{-\frac{1}{2}y^2} dy \quad \left(\alpha' = \frac{\alpha}{\sqrt{2\pi}}\right). \end{aligned} \quad (9)$$

Для того чтобы при интегрировании по y верхний предел был постоянным, сделаем подстановку $y = x\sqrt{z}$:

$$H(a) = \alpha' \int_0^{\infty} z^{\frac{f-1}{2}} e^{-\frac{1}{2}z} dz \int_{-\infty}^c e^{-\frac{1}{2}x^2z} dx \quad (10)$$

и изменим порядок интегрирования, что допустимо в силу положительности подинтегральных функций:

$$H(a) = \alpha' \int_{-\infty}^c dx \int_0^{\infty} z^{\frac{f-1}{2}} e^{-\frac{1}{2}(1+x^2)z} dz. \quad (11)$$

Интегрируя по z , получим, согласно § 12 (2), гамма-функцию

$$\int_0^{\infty} z^{\frac{f-1}{2}} e^{-\frac{1}{2}(1+x^2)z} dz = \left(\frac{1+x^2}{2}\right)^{-\frac{f+1}{2}} \Gamma\left(\frac{f+1}{2}\right).$$

Поэтому

$$H(a) = \gamma \int_{-\infty}^c (1+x^2)^{-\frac{f+1}{2}} dx = \frac{\gamma}{\sqrt{f}} \int_{-\infty}^a \left(1 + \frac{t^2}{f}\right)^{-\frac{f+1}{2}} dt, \quad (12)$$

где

$$\gamma = \alpha \pi^{-\frac{1}{2}} 2^{\frac{f}{2}} \Gamma\left(\frac{f+1}{2}\right) = \pi^{-\frac{1}{2}} \Gamma\left(\frac{f+1}{2}\right) / \Gamma\left(\frac{f}{2}\right). \quad (13)$$

Формула (12) является решением поставленной задачи. Интеграл $H(a)$ можно вычислить в элементарных функциях. Плотность вероятности случайной величины t задается равенством

$$h(t) = \frac{\gamma}{\sqrt{f}} \left(1 + \frac{t^2}{f}\right)^{-\frac{f+1}{2}}. \quad (14)$$

График функции (14) имеет колоколообразную форму и похож на гауссову кривую ошибок. При $f \rightarrow \infty$ функция (14) стремится, очевидно, к плотности нормального распределения

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

Если $H(a)$ вычислена в явном виде как функция a , то можно найти, при каком a вероятность $H(a)$ принимает заданное значение $1 - \beta$, причем β можно выбрать, например, равным 0,025 или 0,005. Тогда вероятность того, что t превзойдет границу $a = t_{\beta}$, будет равна β . Так как $-t$ имеет то же самое распределение, что и t , то $-t$ сможет превзойти границу t_{β} также лишь с вероятностью β .

Абсолютная величина $|t|$ может превзойти границу t_β с вероятностью 2β .

Указанная выше постановка вопроса и ответ на этот вопрос в виде формулы (12) исходят от английского статистика Госсета, который под скромным псевдонимом «Стьюдент» опубликовал работу¹, совершившую переворот в статистике. Поэтому распределение случайной величины t называют *распределением Стьюдента*. В свою очередь правило, согласно которому гипотетическую величину среднего значения \hat{x} отвергают тогда, когда модуль отношения t превосходит границу t_β , называют *критерием Стьюдента*.

Граница t_β зависит от уровня значимости β и от числа степеней свободы $f = n - 1$. Границы t_β , табулированы в табл. 7 в конце книги.

При использовании одностороннего критерия предполагаемое среднее значение \hat{x} отвергают лишь тогда, когда t положительно и превосходит t_β , или лишь тогда, когда t отрицательно и не превосходит $-t_\beta$. При двустороннем критерии на знак t не обращают внимания, а лишь учитывают абсолютную величину $|t|$. Вероятность того, что правильное значение \hat{x} по критерию Стьюдента будет отвергнуто, в случае одностороннего критерия равна β , а в случае двустороннего критерия равна 2β . При этом предполагается, что все x_i — независимые одинаково нормально распределенные случайные величины. Если это не так, то β и 2β будут лишь приближенными значениями для уровней значимости соответствующих критериев.

§ 29. Сравнение двух средних значений

Во всех экспериментальных естественных науках большое значение имеет следующая постановка вопроса:

Пусть имеется g случайно выбранных значений x_1, \dots, x_g некоторой случайной величины x и h значений y_1, \dots, y_h другой случайной величины y . Обозначим \bar{x} и \bar{y} выборочные средние в первой и второй выборке соответственно. Предположим, что \bar{y} оказалось несколько меньше (или несколько больше), чем \bar{x} . Означает ли это, что истинное среднее значение \hat{y} , в силу тех или иных условий, также меньше (соответственно больше) другого истинного среднего значения \hat{x} , или различие выборочных средних носит чисто случайный характер? Иными словами: как велика должна быть разность $D = \bar{x} - \bar{y}$, чтобы ее можно было считать *значимой*, т. е. чтобы можно было утверждать, что $\hat{x} \neq \hat{y}$?

¹ Student, The probable error of a mean, Biometrika, 6 (1908), 1.

На этот вопрос гауссова теория ошибок дает следующий ответ: так как при больших g и h выборочные средние \bar{x} и \bar{y} имеют приближенно нормальные распределения с дисперсиями

$$\sigma_{\bar{x}}^2 = \frac{1}{g} \sigma_x^2 \quad \text{и} \quad \sigma_{\bar{y}}^2 = \frac{1}{h} \sigma_y^2$$

и приближенные значения этих дисперсий задаются формулами

$$s_{\bar{x}}^2 = \frac{1}{g} \cdot s_x^2 = \frac{\sum (x - \bar{x})^2}{g(g-1)}, \quad (1)$$

$$s_{\bar{y}}^2 = \frac{1}{h} \cdot s_y^2 = \frac{\sum (y - \bar{y})^2}{h(h-1)}, \quad (2)$$

то разность

$$D = \bar{x} - \bar{y}$$

имеет также приближенно нормальное распределение с дисперсией

$$\sigma_D^2 = \sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2,$$

приближенным значением которой является

$$s_D^2 = s_{\bar{x}}^2 + s_{\bar{y}}^2. \quad (3)$$

Если истинная разность $\hat{x} - \hat{y} = 0$, то отношение $q = D/\sigma_D$ распределено приближенно нормально с нулевым средним значением и единичной дисперсией. Следовательно, с вероятностью, близкой к $1 - 2\beta$, $|q|$ не превосходит границы q_β , которую можно найти по таблицам нормального распределения: например, при $2\beta = 0,01$ граница $q_\beta = 2,58$. Однако так как знаменатель σ_D неизвестен, то σ_D заменяют величиной s_D . Вследствие такой замены ненадежность границы усиливается, поэтому обычно q_β несколько завышают; например, в качестве границы для отношения D/s_D часто выбирают 3 или, если желают достичь еще большей надежности, 4. Если нормированная выборочная разность $|D|/s_D$ превышает эту границу, то гипотезу $\hat{x} = \hat{y}$ следует отвергнуть и считать, что $\hat{x} > \hat{y}$ или $\hat{x} < \hat{y}$, смотря по тому, положительна или отрицательна выборочная разность D .

Как видно, это грубое правило лишено вполне удовлетворительного обоснования. Что же в конце концов нужно выбирать в качестве границы для D/s_D : 2,58; 3 или 4, и каков уровень значимости этого критерия? Для очень больших g и h все ясно, так как в этом случае σ_D можно заменить на s_D без риска совершить сколько-нибудь значительную ошибку, но для малых или умеренно больших g и h хотелось бы уже знать точнее, сколь большой должна быть выбрана граница q_β для D/s_D , чтобы в случае $\hat{x} = \hat{y}$ вероятность события $|D|/s_D > q_\beta$ равнялась бы, скажем, 0,01?

К сожалению, на этот вопрос нельзя ответить совсем точно. Вероятность того, что отношение D/s_D превысит заданную границу, зависит (хотя и в незначительной степени) от неизвестного отношения истинных квадратичных отклонений σ_x и σ_y . Поэтому постановку вопроса несколько изменяют.

Согласно основной гипотезе, которую нужно проверить и, может быть, отвергнуть, различие между \bar{x} и \bar{y} является чисто случайным и истинные средние значения \hat{x} и \hat{y} равны друг другу. Но если различие между x и y чисто случайное, то можно предположить, что соответствующие средние ошибки σ_x и σ_y также равны друг другу. Таким образом, мы исходим из гипотезы, что x и y имеют не только одинаковые средние значения $\hat{x} = \hat{y} = \mu$, но также и одинаковые квадратичные отклонения σ . Нужно проверить, согласуется ли найденная выборочная разность $D = \bar{x} - \bar{y}$ с этой гипотезой?

Если истинные дисперсии σ_x^2 и σ_y^2 равны друг другу, то неразумно вычислять для них два различных приближенных значения s_x^2 и s_y^2 . В этом случае следует вычислить одно-единственное приближенное выражение для обеих дисперсий, а именно, взвешенное среднее s_x^2 и s_y^2 . Веса, которые следует приписать величинам s_x^2 и s_y^2 , согласно § 26 Б, должны быть обратно пропорциональны дисперсиям случайных величин s_x^2 и s_y^2 . Эти дисперсии, согласно § 27 (11), равны

$$\frac{2\sigma^4}{g-1} \quad \text{и} \quad \frac{2\sigma^4}{h-1}.$$

Следовательно, вес g^2 относится к весу h^2 , как $g-1$ к $h-1$. Поэтому взвешенное среднее равно

$$s^2 = \frac{(g-1)s_x^2 + (h-1)s_y^2}{(g-1) + (h-1)} = \frac{\sum (x - M_x)^2 + \sum (y - M_y)^2}{g + h - 2}. \quad (4)$$

С помощью этого s^2 образуем теперь

$$S_1^2 = \frac{1}{g} s^2, \quad S_2^2 = \frac{1}{h} s^2, \quad (5)$$

$$S^2 = S_1^2 + S_2^2 = \left(\frac{1}{g} + \frac{1}{h}\right) s^2. \quad (6)$$

В большинстве случаев величина S^2 лишь незначительно отличается от s^2 , вычисленной по формулам (1), (2) и (3). Если $g = h$ или если $s_x^2 = s_y^2$, то $S^2 = s^2$. Таким образом, результаты вычислений по формулам (1), (2), (3) и (4), (5), (6) практически будут одинаковыми.

Следуя Стьюденту и Фишеру¹, построим отношение

$$t = \frac{D}{S} \quad (7)$$

¹ Fisher R. A., Applications of „Student's” distribution, *Metron*, 5 (1926), 90.

и постараемся определить функцию распределения этого отношения в предположении, что справедлива гипотеза, согласно которой x_1, \dots, x_g и y_1, \dots, y_h — независимые одинаково нормально распределенные случайные величины со средним значением μ и квадратичным отклонением σ .

Так как при одинаковом изменении масштабов осей Ox и Oy отношение t не меняется, то мы можем считать, что $\mu = 0$ и $\sigma = 1$. Положим

$$\begin{aligned}\chi_1^2 &= (g-1) s_x^2 = \sum (x - \bar{x})^2, \\ \chi_2^2 &= (h-1) s_y^2 = \sum (y - \bar{y})^2.\end{aligned}$$

Тогда

$$(g+h-2) s^2 = \chi_1^2 + \chi_2^2. \quad (8)$$

Покажем, что случайные величины $\chi_1^2, \chi_2^2, \bar{x}$ и \bar{y} являются независимыми. Вероятность одновременного выполнения неравенств

$$\left. \begin{aligned}a_1 \leq \chi_1^2 < b_1, \quad a_3 \leq \bar{x} < b_3, \\ a_2 \leq \chi_2^2 < b_2, \quad a_4 \leq \bar{y} < b_4\end{aligned} \right\} \quad (9)$$

равна интегралу от совместной плотности вероятности системы случайных величин $(x_1, \dots, x_g, y_1, \dots, y_h)$:

$$\int \dots \int f(x_1) \dots f(x_g) f(y_1) \dots f(y_h) dx_1 \dots dx_g dy_1 \dots dy_h \quad (10)$$

по области (9). Этот интеграл немедленно распадается на два множителя, из которых первый является интегралом по x_1, \dots, x_g , а второй — интегралом по y_1, \dots, y_h . Поэтому вероятность одновременного осуществления событий (9) равна произведению вероятностей

$$\mathbf{P} \left\{ \begin{aligned} a_1 \leq \chi_1^2 < b_1 \\ a_3 \leq \bar{x} < b_3 \end{aligned} \right\} \cdot \mathbf{P} \left\{ \begin{aligned} a_2 \leq \chi_2^2 < b_2 \\ a_4 \leq \bar{y} < b_4 \end{aligned} \right\}. \quad (11)$$

Но в § 27 было доказано, что χ_1^2 и M_x независимы, следовательно, первый множитель в (11) снова распадается на два множителя

$$\mathbf{P}(a_1 \leq \chi_1^2 < b_1) \cdot \mathbf{P}(a_3 \leq \bar{x} < b_3).$$

То же самое справедливо и для второго сомножителя в формуле (11). Таким образом, мы в конце концов получаем произведение четырех вероятностей, соответствующих случайным величинам $\chi_1^2, \chi_2^2, \bar{x}$ и \bar{y} . Этим завершается доказательство независимости указанных величин.

Их совместная плотность вероятности $f(u_1, u_2, v_1, v_2)$ в силу только что доказанной теоремы представляет собой произведение

четырёх плотностей вероятности. Формулы этих плотностей были выведены ранее в § 27: χ_1^2 имеет плотность вероятности

$$g_1(u) = \alpha_1 u^{\frac{g-3}{2}} e^{-\frac{1}{2}u}$$

(распределение χ^2 с $g - 1$ степенями свободы), аналогично для χ_2^2

$$g_2(u) = \alpha_2 u^{\frac{h-3}{2}} e^{-\frac{1}{2}u},$$

в то время как \bar{x} и \bar{y} распределены нормально с нулевым средним и дисперсиями $1/g$ и $1/h$ соответственно. В силу независимости χ_1^2 , χ_2^2 , \bar{x} и \bar{y} , случайные величины

$$\chi^2 = (g + h - 2) s^2 = \chi_1^2 + \chi_2^2, \quad (12)$$

$$D = \bar{x} - \bar{y} \quad (13)$$

также независимы. Вероятность одновременного осуществления двух событий $c_1 \leq \chi^2 < d_1$ и $c_2 \leq D < d_2$ равна четырехкратному интегралу от плотности вероятности $f(u_1, u_2, v_1, v_2)$, который распадается на два множителя:

$$P(c_1 \leq \chi^2 < d_1) \cdot P(c_2 \leq D < d_2).$$

Согласно § 27 E, плотность вероятности для χ^2 снова является плотностью типа

$$g(u) = \alpha u^{-\frac{l-2}{2}} e^{-\frac{1}{2}u} \quad (14)$$

с $f = g + h - 2$ степенями свободы, в то время как D распределена нормально с нулевым средним значением и дисперсией

$$\sigma_D^2 = \sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2 = \frac{1}{g} + \frac{1}{h}. \quad (15)$$

Согласно (6) и (12),

$$S^2 = \left(\frac{1}{g} + \frac{1}{h}\right) s^2 = \left(\frac{1}{g} + \frac{1}{h}\right) \frac{\chi^2}{g+h-2} = \left(\frac{1}{g} + \frac{1}{h}\right) \frac{\chi^2}{f}. \quad (16)$$

Из (15) и (16) следует, что

$$S^2 = \sigma_D^2 \frac{\chi^2}{f}. \quad (17)$$

Отношение (7) можно теперь записать так:

$$t = \frac{D}{S} = \frac{D}{\sigma_D} \frac{\sqrt{f}}{\chi} = Y \frac{\sqrt{f}}{\chi}, \quad (18)$$

где

$$Y = \frac{D}{\sigma_D}. \quad (19)$$

Формула (18) полностью аналогична формуле (4) из § 28, согласно которой

$$t = y_1 \frac{\sqrt{f}}{x}. \quad (20)$$

Y и χ^2 так же, как ранее y_1 и χ^2 , являются независимыми случайными величинами, из которых первая распределена нормально с нулевым средним значением и единичной дисперсией, а вторая подчиняется распределению χ^2 с f степенями свободы. Следовательно, отношение t имеет то же самое распределение, что и раньше.

Таким образом, получается следующий критерий, который называют *критерием Стьюдента для проверки различия средних значений*:

Если абсолютная величина отношения

$$|t| = \frac{|D|}{S} = \frac{|\bar{x} - \bar{y}|}{S}$$

превышает границу t_{β} из табл. 7, то гипотезу $\hat{x} = \hat{y}$ следует отвергнуть и считать, что $\hat{x} > \hat{y}$ или $\hat{x} < \hat{y}$, смотря по тому, будет ли разность D положительной или отрицательной.

Если распределения отдельных наблюдений не слишком отклоняются от нормальных и соответствующие дисперсии равны друг другу, то вероятность ошибочно отвергнуть правильную гипотезу $\hat{x} = \hat{y}$ для указанного критерия равна 2β . Если же в действительности $\hat{x} > \hat{y}$ и дисперсии равны, то вероятность ошибочного вывода $\hat{x} < \hat{y}$ оказывается даже меньше, чем β . В случае одностороннего критерия вероятность ошибочно отвергнуть правильную гипотезу не превышает¹ β .

Если одно из выборочных значений сильно отклоняется от выборочного среднего остальных наблюдений, то возникает вопрос, можно ли это отклонение считать случайным или оно является следствием грубой ошибки в измерениях и поэтому соответствующее наблюдение нужно из выборки исключить? Для ответа на этот вопрос можно воспользоваться критерием Стьюдента, положив в формулах этого параграфа $h = 1$. Пусть $y_1 = x_{g+1}$ — отдельное наблюдение, которое подлежит проверке, а x_1, \dots, x_g — остальные наблюдения. Вероятность того, что на основе этого критерия наблюдение x_{g+1} будет ошибочно отвергнуто, равна 2β . Если этот же критерий применить последовательно ко всем

¹ В случае одностороннего критерия основной гипотезой является $\hat{x} \leq \hat{y}$ (или $\hat{x} \geq \hat{y}$). Поэтому здесь вероятность отвергнуть правильную гипотезу существенно зависит от разности $\hat{x} - \hat{y}$. Максимальное значение этой вероятности достигается при $\hat{x} = \hat{y}$ и равно β . — *Прим. перев.*

элементам выборки x_1, \dots, x_{g+1} , то общая вероятность того, что хотя бы один элемент выборки будет ошибочно отвергнут, не превосходит $2\beta(g+1) = 2\beta n$, где n — количество наблюдений. Если, к примеру, выбрать $2\beta = 0,01$, то при $n = 20$ вероятность ошибочно отвергнуть хотя бы одно наблюдение из двадцати почти равна 0,2. Исключение одного из элементов выборки практически не влияет на точность выводов: выборочное среднее оставшихся 19 наблюдений будет почти точно таким же, как и выборочное среднее всех двадцати наблюдений¹.

Пример 17 (по книге Kendall M. G., *Advanced Theory of Statistics*, vol. II, *Example 21.4*).

В одном классе из 20 детей были случайно отобраны 10, которым ежедневно стали выдавать апельсиновый сок. Остальные 10 человек ежедневно получали молоко. Через некоторое время было зафиксировано следующее увеличение веса детей (в фунтах):

Первая группа $4, 2\frac{1}{2}, 3\frac{1}{2}, 4, 1\frac{1}{2}, 1, 3\frac{1}{2}, 3, 2\frac{1}{2}, 3\frac{1}{2}$.

Вторая группа $1\frac{1}{2}, 3\frac{1}{2}, 2\frac{1}{2}, 3, 2\frac{1}{2}, 2, 2, 2\frac{1}{2}, 1\frac{1}{2}, 3$.

Среднее увеличение веса на одного ребенка в первой группе равно 2,9 фунта, а во второй группе — 2,4 фунта. Значимо ли это различие средних весов?

Находим

$$D = 2,9 - 2,4 = 0,5,$$

$$s^2 = \frac{13,3}{18} = 0,74,$$

$$S^2 = \left(\frac{1}{10} + \frac{1}{10}\right) s^2 = 0,148,$$

$$t = \frac{D}{S} = 1,30.$$

5%-граница для t с 20 — 2 = 18 степенями свободы равна 2,10. Следовательно, нет оснований для заключения, что различие выборочных средних вызвано различием истинных средних.

¹ Указанный вариант критерия Стьюдента обычно используют в процессе эксперимента для предсказания границ, в которых будет находиться какое-либо из очередных, еще не известных наблюдений; при этом предполагается, что известны все предшествующие наблюдения x_1, \dots, x_n или их часть. Для исключения наблюдений, содержащих грубые ошибки, критерий Стьюдента практически не применяется, так как для этой цели выгоднее пользоваться другими более совершенными критериями. См., например, Дунин-Барковский И. В. и Смирнов Н. В., *Теория вероятностей и математическая статистика в технике (общая часть)*, ГИТТЛ, М., 1955, гл. 8, § 4 и табл. XXIV. — *Прим. перев.*

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

§ 30. Выравнивание ошибок наблюдений

Запросы астрономии и геодезии привели Гаусса к следующей проблеме:

Пусть $\vartheta_1, \dots, \vartheta_r$ — истинные значения каких-либо неизвестных физических постоянных (например, значения элементов траектории некоторой планеты). Далее, пусть результатами наблюдений являются не сами $\vartheta_1, \dots, \vartheta_r$, а некоторые другие величины x_1, \dots, x_n (например, координаты положений планеты, наблюдаемые с земли в различные моменты времени). При этом предполагается, что истинные значения ξ_1, \dots, ξ_n величин x_1, \dots, x_n определенным образом зависят от $\vartheta_1, \dots, \vartheta_r$:

$$\xi_i = \varphi_i(\vartheta_1, \dots, \vartheta_r). \quad (1)$$

Какие значения параметров ϑ_i лучше всего согласуются с наблюдаемыми значениями x_1, \dots, x_n ?

Лежандр предлагал считать согласие «наилучшим» в том случае, когда сумма квадратов ошибок

$$Q = (x_1 - \xi_1)^2 + \dots + (x_n - \xi_n)^2 \quad (2)$$

минимальна. Теоретико-вероятностное обоснование этого подхода было дано Гауссом, который заметил, что если все наблюдения лишены систематических ошибок и имеют одинаковое квадратичное отклонение σ , то, согласно гауссовой теории ошибок, вероятность того, что результаты наблюдения x_i будут лежать между $t_i - \frac{1}{2} \delta t_i$ и $t_i + \frac{1}{2} \delta t_i$ ($i = 1, 2 \dots n$), при малых δt_i приближенно равна

$$\delta W = \sigma^{-n} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2} \frac{(t_1 - \xi_1)^2 + \dots + (t_n - \xi_n)^2}{\sigma^2}} \delta t_1 \dots \delta t_n. \quad (3)$$

При этом точки с координатами (ξ_1, \dots, ξ_n) принадлежат некоторому многообразию n -мерного пространства, определяемому равенствами (1). Для заданных t_i и δt_i вероятность δW достигает

наибольшего значения в той точке (ξ_1, \dots, ξ_n) этого многообразия, в которой квадратичная форма

$$(t_1 - \xi_1)^2 + \dots + (t_n - \xi_n)^2$$

будет минимальной. Если в эту форму вместо t_i подставить результаты наблюдений x_i , то получится форма (2). Следовательно, по Гауссу, «наилучшими значениями» $\vartheta_1, \dots, \vartheta_r$ будут те значения, которым соответствует наибольшая вероятность (3).

Впоследствии Гаусс дал другое обоснование принципа «наименьших квадратов», не зависящее от предположения нормальной распределенности x_1, \dots, x_n . Отыскание оценки некоторого параметра ϑ он сравнивал с азартной игрой, в которой нельзя выиграть, а можно лишь проиграть. Если T выбрано в качестве оценки параметра ϑ , то «проигрыш» от этого будет тем больше, чем больше абсолютная величина ошибки $T - \vartheta$. За меру проигрыша Гаусс принял квадрат $(T - \vartheta)^2$ и потребовал, чтобы оценка удовлетворяла двум условиям: во-первых, она не должна иметь систематической ошибки, т. е. $\xi T = \vartheta$, и, во-вторых, дисперсия оценки $\xi(T - \vartheta)^2$, являющаяся математическим ожиданием проигрыша, должна быть наименьшей. Затем он доказал, что эти условия минимизации приводят в точности к методу наименьших квадратов.

Если результаты наблюдений имеют различные квадратичные отклонения $\sigma_1, \dots, \sigma_n$, то вместо формы (2), естественно, появляется форма

$$\frac{(x_1 - \xi_1)^2}{\sigma_1^2} + \frac{(x_2 - \xi_2)^2}{\sigma_2^2} + \dots + \frac{(x_n - \xi_n)^2}{\sigma_n^2}, \quad (4)$$

или форма

$$Q = g_1 (x_1 - \xi_1)^2 + \dots + g_n (x_n - \xi_n)^2, \quad (5)$$

(если, как и в § 26 Б, ввести веса g_1, \dots, g_n , обратно пропорциональные дисперсиям $\sigma_1^2, \dots, \sigma_n^2$).

В применениях этого метода к биологическим и экономическим задачам разности $x_i - \xi_i$ являются не ошибками наблюдений, а отклонениями величин x_i от их математических ожиданий ξ_i . Величины x_i предполагаются независимыми и случайными. Их математические ожидания ξ_i , согласно (1), зависят от неизвестных параметров $\vartheta_1, \dots, \vartheta_r$. В качестве оценок для $\vartheta_1, \dots, \vartheta_r$ принимают такие значения этих параметров, для которых форма (5) достигает своего минимума.

При отыскании минимума полагают

$$\left. \begin{aligned} \vartheta_1 &= \vartheta_1^0 + u, \\ \vartheta_2 &= \vartheta_2^0 + v, \\ \dots \dots \dots \end{aligned} \right\} \quad (6)$$

где ξ_i^0 — предварительные приближенные значения для ϑ_i , а u, v, \dots — поправки. Предполагается, что при малых u, v, \dots функции (1) можно достаточно точно приблизить линейными функциями

$$\xi_i = \xi_i^0 + a_i u + b_i v + \dots \quad (7)$$

Величины ξ_i^0 являются приближенными значениями ξ_i соответствующими предварительным приближениям ϑ^0 :

$$\xi_i^0 = \varphi_i(\vartheta^0).$$

В качестве коэффициентов a_i, b_i, \dots линейных приближений (7) можно выбрать значения частных производных от функций (1) в точке $(\vartheta_1^0, \dots, \vartheta_r^0)$:

$$a_i = \left(\frac{\partial \xi_i}{\partial \vartheta_1}\right)^0, \quad b_i = \left(\frac{\partial \xi_i}{\partial \vartheta_2}\right)^0, \dots \quad (8)$$

Для упрощения вычислений мы будем предполагать, что начало координат в пространстве переменных x_1, \dots, x_n перенесено в точку $(\xi_1^0, \dots, \xi_n^0)$. Таким образом, вместо x_i мы вводим в качестве новых переменных наблюдаемые отклонения

$$l_i = x_i - \xi_i^0. \quad (9)$$

Их математические ожидания равны:

$$\lambda_i = \xi_i - \xi_i^0 = a_i u + b_i v + \dots \quad (10)$$

Форма Q теперь запишется так:

$$Q = \sum g_i (l_i - \lambda_i)^2 = \sum g_i (l_i - a_i u - b_i v - \dots)^2. \quad (11)$$

Координаты (u, v, \dots) точки минимума этой формы удовлетворяют системе уравнений, получающейся приравнованием нулю частных производных (11). После деления всех производных на 2 получаем

$$\left. \begin{aligned} \sum g_i a_i (a_i u + b_i v + \dots - l_i) &= 0, \\ \sum g_i b_i (a_i u + b_i v + \dots - l_i) &= 0, \\ \dots\dots\dots \end{aligned} \right\} \quad (12)$$

Если, следуя Гауссу, для краткости ввести обозначения

$$\sum g_i a_i^2 = [gaa], \quad \sum g_i a_i b_i = [gab], \dots,$$

то система уравнений (12) сведется в конце концов к системе *нормальных уравнений*

$$\left. \begin{aligned} [gaa]u + [gab]v + \dots &= [gal], \\ [gba]u + [gbb]v + \dots &= [gbl], \\ \dots\dots\dots \end{aligned} \right\} \quad (13)$$

Количество нормальных уравнений равно количеству неизвестных параметров $\vartheta_1, \dots, \vartheta_r$. Если все наблюдения равноточны, то можно считать, что все весовые множители g_i равны единице. В этом случае (13) запишется более просто:

$$\left. \begin{aligned} [aa]u + [ab]v + \dots &= [al], \\ [ba]u + [bb]v + \dots &= [bl], \\ \dots & \dots \end{aligned} \right\} \quad (14)$$

В способе записи нормальных уравнений я возможно ближе придерживался традиции, берущей свое начало от Гаусса. Введением матричных обозначений запись системы уравнений можно было несколько сократить, однако старомодный способ записи (14) очень удобен для приложений. Обычно g_i, a_i, b_i, \dots и l_i записывают столбцами и затем вычисляют коэффициенты $[aa]$ или $[gaa]$ и т. д.

Системы уравнений (13) или (14) всегда имеют решение, так как положительный квадратичный многочлен всегда достигает минимума. Однако решение не обязательно является однозначным. Может случиться, что нормальные уравнения однозначно разрешимы лишь для некоторых определенных линейных комбинаций параметров u, v, \dots , а относительно самих u, v, \dots однозначного решения нет¹. Следуя индийскому статистическому Рао², такие линейные комбинации параметров мы будем называть *допускающими оценку*.

Чтобы исследовать точнее, какие функции параметров допускают оценку, рассмотрим линейные формы

$$\lambda_i = a_i u + b_i v + \dots \quad (15)$$

Пусть среди них имеется, например, ровно p линейно независимых форм ($p \leq r$). Без ограничения общности можно считать, что линейно независимыми формами являются $\lambda_1, \dots, \lambda_p$ и что все остальные $\lambda_{p+1}, \dots, \lambda_n$ через них выражаются линейно, поэтому (11) можно записать в виде квадратичного многочлена, зависящего только от $\lambda_1, \dots, \lambda_p$. Квадратичная часть этого многочлена представляет собой сумму

$$\sum_{i=1}^p g_i \lambda_i^2 + \sum_{i=p+1}^n g_i \lambda_i^2 = \sum_{i=1}^p g_i (a_i u + b_i v + \dots)^2 \geq \sum_{i=1}^p g_i \lambda_i^2 \geq 0,$$

где λ_i при $i > p$ являются линейными комбинациями $\lambda_1, \dots, \lambda_p$. Эта сумма обращается в нуль тогда и только тогда, когда все

¹ Этот случай осуществляется тогда и только тогда, когда ранг матрицы системы (13) или (14) строго меньше количества уравнений и неизвестных r . — *Прим. перев.*

² Рао С. Р., *Advanced Statist. Methods in Biometric Research*, New York, 1952.

$\lambda_1, \dots, \lambda_p$ равны нулю, поэтому она положительно определена. Как известно, всякую положительно определенную квадратичную форму порядка p невырожденным линейным преобразованием переменных можно привести к сумме p квадратов:

$$\sum_{i=1}^p g_i \lambda_i^2 + \sum_{i=p+1}^n g_i \lambda_i^2 = (\mu^1)^2 + \dots + (\mu^p)^2,$$

где μ^j ($j = 1, 2, \dots, p$) — независимые линейные формы от $\lambda_1, \dots, \lambda_n$. Форма Q получается из этой квадратичной формы добавлением некоторой линейной функции от $\lambda_1, \dots, \lambda_p$, поэтому при данном линейном преобразовании Q будет иметь вид

$$Q = (\mu^1 - c^1)^2 + \dots + (\mu^p - c^p)^2 + (c^0)^2.$$

Эта форма достигает минимума $(c^0)^2$ в точке $\mu^1 = c^1, \dots, \mu^p = c^p$. Следовательно, μ^1, \dots, μ^p , а вместе с ними и λ_i допускают оценку. Отсюда заключаем, что.

Только те линейные формы параметров u, v, \dots допускают оценку, которые представимы в виде линейных комбинаций форм (15).

Если вместо прежних переменных u, v, \dots ввести новые переменные $\lambda_1, \dots, \lambda_p$, то относительно последних система нормальных уравнений будет однозначно разрешимой. В дальнейшем мы будем предполагать, что такая замена уже произведена, и поэтому система нормальных уравнений имеет единственное решение.

Простейшим способом решения систем (13) и (14) является совсем примитивный школьный способ, указанный еще Гауссом. По этому способу первое уравнение разрешают относительно u и результат подставляют во все остальные уравнения и т. д. При вычислениях целесообразно в левых частях (13) и (14) заменить символические коэффициенты их числовыми значениями, а свободные члены справа оставить неопределенными. В этом случае решение будет представлять собой линейные комбинации свободных членов:

$$\left. \begin{aligned} u &= h^{11} [gal] + h^{12} [gbl] + \dots, \\ v &= h^{21} [gal] + h^{22} [gbl] + \dots, \\ &\dots \dots \dots \end{aligned} \right\} \quad (16)$$

где h^{ik} — элементы матрицы, обратной матрице коэффициентов системы (13).

Вычислив u, v, \dots , находят ϑ по формулам (6) и λ — по формулам (10). Так как результаты этих вычислений представляют собой не истинные значения ϑ и λ , а лишь оценки этих параметров,

то мы обозначим их \tilde{v} и $\tilde{\lambda}$. Зная $\tilde{\lambda}$, можно получить оценки для ξ по формулам

$$\tilde{\xi}_i = \xi_i^0 + \tilde{\lambda}_i$$

и оценки поправок k_i для наблюдений¹

$$k_i = \tilde{\xi}_i - x_i = \tilde{\lambda}_i - l_i. \quad (17)$$

Если оценки \tilde{v} сильно отклоняются от начального приближения v^0 и если функции (1) не являются линейными, то вычисления нужно повторить еще раз, приняв за начальное приближение \tilde{v} вместо v^0 .

При практических расчетах контроль вычислений, безусловно, необходим. В качестве одного из способов контроля можно использовать то обстоятельство, что величины k_i , согласно (12), должны удовлетворять условиям:

$$\left. \begin{aligned} [gak] &= 0, \\ [gbk] &= 0, \\ \dots\dots\dots \end{aligned} \right\} \quad (18)$$

Другой способ контроля основан на вычислении минимума \tilde{Q} для формы Q , определяемой равенством (11). Значения λ_i , при которых форма Q достигает минимума, в точности равны $\tilde{\lambda}_i$, поэтому имеет место равенство

$$\tilde{Q} = \sum g_i(l_i - \tilde{\lambda}_i)^2 = \sum g_i k_i^2 = [gkk]. \quad (19)$$

С другой стороны, простое выражение для \tilde{Q} можно, согласно (18), получить так:

$$\begin{aligned} \tilde{Q} &= \sum g_i(\tilde{\lambda}_i - l_i)(\tilde{\lambda}_i - l_i) = \sum g_i(-l_i + a_i u + b_i v + \dots)k_i = \\ &= -[glk] + [gak]u + [gbk]v + \dots = -[glk]. \end{aligned}$$

Если теперь k_i снова заменить на $\tilde{\lambda}_i - l_i$, то найдем:

$$\tilde{Q} = [gl\tilde{\lambda}] - [gal]u - [gbl]v - \dots \quad (20)$$

Формула (20) служит для вычисления \tilde{Q} , а (19) — для контроля. Согласно (16), u, v, \dots представляют собой линейные функции от наблюдаемых отклонений $l_i = x_i - \xi_i^0$:

$$\left. \begin{aligned} u &= \alpha_1 l_1 + \dots + \alpha_n l_n, \\ v &= \beta_1 l_1 + \dots + \beta_n l_n, \\ \dots\dots\dots \end{aligned} \right\} \quad (21)$$

¹ У Гаусса оценки поправок обозначаются λ_i .

Коэффициенты α_i, β_i, \dots можно легко вычислить по формулам (16):

$$\alpha_i = g_i (h^{11}a_i + h^{12}b_i + \dots). \quad (22)$$

Для практических вычислений формулы (21) и (22) не имеют никакого значения, но они нам понадобятся в следующих параграфах для вычисления дисперсий.

Пример 18 (из книги Helmert F. R., Die Ausgleichsrechnung, Leipzig, 1872). При построении триангуляционной сети Швердом в районе Шпейера определялись углы между направлениями из основного пункта D' в пункты A, B, W, H и N . Шверд произвел многократные измерения, арифметические средние которых указаны ниже:

BA	(90 измерений)	$19^\circ 25' 59'', 42$
BW	(80 измерений)	$34^\circ 18' 43'', 61$
AW	(70 измерений)	$14^\circ 52' 44'', 33$
HW	(20 измерений)	$15^\circ 34' 58'', 80$
BH	(20 измерений)	$18^\circ 43' 45'', 60$
NA	(40 измерений)	$12^\circ 26' 24'', 65$
BN	(60 измерений)	$6^\circ 59' 34'', 51$
NH	(20 измерений)	$11^\circ 44' 11'', 60$

Многократностью опытов удалось в значительной мере компенсировать инструментальные ошибки. Следовательно, мы можем предположить, что наблюдения лишены систематических ошибок, и выбрать веса g пропорциональными числу измерений. За неизвестные ϑ_i примем четыре угла: BN, BH, BA и BW , через которые можно выразить все остальные. В качестве начальных приближений выберем измеренные значения этих четырех углов; таким образом, мы получим

$$\begin{aligned} \vartheta_1 = BN &= 6^\circ 59' 34'', 51 + u, \\ \vartheta_2 = BH &= 18^\circ 43' 45'', 60 + v, \\ \vartheta_3 = BA &= 19^\circ 25' 59'', 42 + u, \\ \vartheta_4 = BW &= 34^\circ 18' 43'', 61 + t. \end{aligned}$$

Восемь углов $\xi_1 = BA, \dots, \xi_8 = NH$, как уже говорилось выше, выражаются через эти неизвестные:

$$\begin{aligned} \xi_1 = BA &= 19^\circ 25' 59'', 42 + w, \\ \xi_2 = BW &= 34^\circ 18' 43'', 61 + t, \\ \xi_3 = AW &= 14^\circ 52' 44'', 19 - w + t, \\ &\dots \dots \dots \\ \xi_8 = NH &= 11^\circ 44' 11'', 09 - u + v. \end{aligned}$$

Коэффициенты этих выражений a_i, b_i, c_i, d_i , веса g_i и отклонения l_i указаны в следующей таблице:

g	a	b	c	d	l
9	0	0	+1	0	0
8	0	0	0	+1	0
7	0	0	-1	+1	+0,14
2	0	-1	0	+1	+0,79
2	0	+1	0	0	0
4	-1	0	+1	0	-0,26
6	+1	0	0	0	0
2	-1	+1	0	0	+0,51

Соответствующие нормальные уравнения имеют вид

$$\left. \begin{aligned} 12u - 2v - 4w &= -0,02, \\ -2u + 6v - 2t &= +0,56, \\ -4u + 20w - 7t &= -2,02, \\ -2v - 7w + 17t &= +2,56. \end{aligned} \right\} \quad (23)$$

Заменяя правые части уравнений (23) буквами A, B, C, D и затем решая эту систему относительно u, v, w, t , получим «неопределенное» решение вида

$$\left. \begin{aligned} u &= 0,00978 A + 0,00375 B + 0,00247 C + 0,00146 D, \\ v &= 0,00375 A + 0,01890 B + 0,00178 C + 0,00296 D, \\ w &= 0,00247 A + 0,00178 B + 0,00650 C + 0,00289 D, \\ t &= 0,00146 A + 0,00296 B + 0,00289 C + 0,00742 D. \end{aligned} \right\} \quad (24)$$

Коэффициенты правых частей (24) являются элементами h^{11}, \dots, h^{44} матрицы, обратной матрице системы (23). Если в (24) вместо A, B, C, D подставить свободные члены уравнений (23), то получим

$$u = -0,032, \quad v = -0,065, \quad w = -0,067, \quad t = +0,115.$$

При этом поправки k_i равны

$$\begin{aligned} k_1 &= -0,067, & k_5 &= -0,065, \\ k_2 &= +0,115, & k_6 &= +0,225, \\ k_3 &= +0,042, & k_7 &= -0,032, \\ k_4 &= -0,609, & k_8 &= -0,543. \end{aligned}$$

Теперь можно вычислить \tilde{Q} по формуле (19) или (20). Находим, что, согласно обеим формулам,

$$\tilde{Q} = 1,71.$$

§ 31. Средние значения и дисперсии оценок \tilde{y}

Оценки \tilde{y} , полученные по методу наименьших квадратов, являются линейными функциями результатов наблюдений x_k и, следовательно, случайными величинами. Вычислим их средние значения и дисперсии.

А. СРЕДНИЕ ЗНАЧЕНИЯ

Пусть система нормальных уравнений (13) § 30 имеет единственное решение u, v, \dots , которое мы обозначим теперь u^1, \dots, u^r , а сами нормальные уравнения запишем в виде

$$\sum h_{jk} u^k = r_j. \quad (1)$$

В этом случае решение (1) равно

$$u^i = \sum h^{ij} r_j, \quad (2)$$

где (h^{ij}) — матрица, обратная матрице (h_{ij}) :

$$\sum_j h^{ij} h_{jk} = \delta_k^i = \begin{cases} 1, & \text{если } i = k, \\ 0, & \text{если } i \neq k. \end{cases} \quad (3)$$

Соответствующие оценки \tilde{v} запишутся в виде равенств

$$\tilde{v}_k = v_k^0 + w^k \quad (k = 1, \dots, r). \quad (4)$$

Для упрощения вычислений средних значений \tilde{v}_k мы в качестве v_k^0 выберем истинные значения параметров v_k . При практических расчетах равенства $v_k^0 = v_k$, конечно, может и не быть, так как истинные значения неизвестны, но при теоретическом вычислении средних значений и дисперсий это допущение не окажет никакого влияния на результат¹. Таким образом, вместо (4) мы напишем

$$\tilde{v}_k = v_k + w^k. \quad (5)$$

Так как v_k^0 выбраны равными истинным значениям v_k , то соответствующие ξ_i^0 равны ξ_i — математическим ожиданиям x_i . Поэтому математическое ожидание разности

$$l_i = x_i - \xi_i^0 \quad (6)$$

равно нулю. Отсюда, согласно (21) § 30, следует, что w^k также имеют нулевые математические ожидания. Поэтому, в силу (5),

Математические ожидания оценок \tilde{v}_k равны истинным значениям параметров v_k .

Это же самое можно выразить и так:

Оценки \tilde{v}_k лишены систематических ошибок,

или

Оценки \tilde{v}_k являются несмещенными.

Б. ДИСПЕРСИИ

При вычислении дисперсий мы будем исходить из предположения, что все x_i являются независимыми случайными величинами с постоянными дисперсиями σ_i^2 , не зависящими от \tilde{v} . В § 30 веса g_i были выбраны обратно пропорциональными дисперсиям σ_i^2 , следовательно, мы можем положить

$$g_i \sigma_i^2 = \sigma^2. \quad (7)$$

¹ Это утверждение автора, безусловно, верно тогда, когда функции (1) § 30 линейны. В противном случае будет лишь приближенное равенство $\xi(\tilde{v}_k + w^k) \approx \xi(v_k + w^k)$, причем разность между левой и правой частями, вообще говоря, стремится к нулю при неограниченном возрастании числа наблюдений n . — *Прим. перев.*

Величина σ^2 равна дисперсии наблюдения, которому соответствует вес, равный единице, поэтому σ^2 называют «дисперсией на единицу веса».

Согласно (5), дисперсия \tilde{v}_k равна дисперсии w^k , для вычисления которой мы снова воспользуемся формулами (21) § 30. При $k = 1$ имеем

$$w^1 = u = \alpha_1 l_1 + \dots + \alpha_n l_n. \quad (8)$$

Так как все l_i — независимые случайные величины с дисперсиями σ_i^2 , то дисперсия u равна

$$\sigma_u^2 = \alpha_1^2 \sigma_1^2 + \dots + \alpha_n^2 \sigma_n^2. \quad (9)$$

Это же выражение, согласно (7), можно записать так:

$$\sigma_u^2 = \sum \alpha_i^2 \frac{\sigma_i^2}{g_i}$$

или, согласно (22) § 30,

$$\begin{aligned} \sigma_u^2 &= \sum g_i (h^{11} a_i + h^{12} b_i + \dots)^2 \sigma^2 = \\ &= (h^{11} h^{11} [gaa] + 2h^{11} h^{12} [gab] + h^{12} h^{12} [gbb] + \dots) \sigma^2. \end{aligned}$$

Величины $[gaa], \dots$ являются коэффициентами нормальных уравнений. Эти коэффициенты мы обозначили буквами h_{ij} , поэтому

$$\sigma_u^2 = \left(\sum_j \sum_k h^{1j} h_{jk} h^{1k} \right) \sigma^2 \quad (10)$$

или, в силу (3),

$$\sigma_u^2 = h^{11} \sigma^2. \quad (11)$$

Точно так же при $k = 2$ получим

$$\sigma_v^2 = h^{22} \sigma^2 \quad (12)$$

и т. д.

В. НАГЛЯДНОЕ ГЕОМЕТРИЧЕСКОЕ ИСТОЛКОВАНИЕ

Для геометрической иллюстрации метода наименьших квадратов мы рассмотрим случай, когда имеется лишь один неизвестный параметр ($r = 1$) и три равноточных результата наблюдений. В этом случае наблюдаемые значения x_1, x_2, x_3 можно истолковать как координаты *наблюдённой точки* X в пространстве переменных x_1, x_2, x_3 .

В качестве начала координат примем точку ξ^0 , которая соответствует начальному приближению, введённому в § 30. Предположением же, согласно которому ξ^0 совпадет с истинной точкой ξ , мы теперь пользоваться не будем.

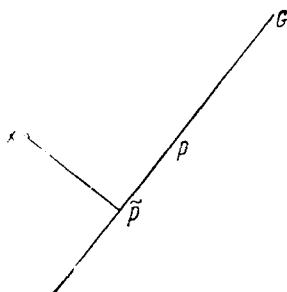
Уравнения (7) § 30 являются параметрическими уравнениями некоторой прямой линии. Так как в данном случае имеется лишь один параметр u , по предположению, все $\xi_i^0 = 0$, то уравнения прямой упрощаются:

$$\xi_i = a_i u \quad (i = 1, 2, 3). \quad (13)$$

Этой прямой G принадлежит «истинная точка» P , координаты которой равны математическим ожиданиям результатов наблюдений: $\xi_i = \hat{x}_i$. Наблюденная точка X расположена где-то вблизи от точки P (рис. 19). Форма

$$Q = (x_1 - \xi_1)^2 + (x_2 - \xi_2)^2 + (x_3 - \xi_3)^2 \quad (14)$$

представляет собой квадрат расстояния между наблюдаемой точкой X и прямой G . Отыскание минимума Q равносильно отысканию на прямой G такой точки \tilde{P} , которая менее всего удалена от X . Следовательно, \tilde{P} является основанием перпендикуляра, опущенного из X на прямую G .



Р и с. 19. Метод наименьших квадратов.

Формулы для вычисления координат основания перпендикуляра станут проще, если мы предварительно произведем ортогональное преобразование координат. В качестве одной из новых осей выберем прямую G , а две другие оси расположим перпендикулярно G . В новых координатах параметрическими уравнениями G будут являться:

$$\eta_1 = au, \quad \eta_2 = 0, \quad \eta_3 = 0 \quad (a^2 = a_1^2 + a_2^2 + a_3^2). \quad (15)$$

В общем случае (r параметров и n величин x_i) подпространство G задается параметрическими уравнениями

$$\xi_i = a_i u + b_i v + \dots \quad (16)$$

Ортогональное преобразование можно записать в виде равенств

$$x_i = \sum e_{ik} y_k. \quad (17)$$

Для того чтобы r новых координатных направлений лежали в подпространстве G , нужно, чтобы первые r столбцов матрицы (e_{ik}) являлись линейными комбинациями векторов (a_i) , (b_i) , \dots . Выберем в качестве первого столбца вектор (λa_k) , в качестве второго столбца — линейную комбинацию векторов $(\mu a_i + \nu b_i)$ и т. д., а затем определим коэффициенты λ , μ , ν , \dots из условия ортогональности.

Если вычислить математические ожидания правых и левых частей (17), то получим равенства

$$\xi_i = \sum e_{ik} \eta_k,$$

где η_1, \dots, η_r представляют собой линейные комбинации u, v, \dots , а $\eta_{r+1}, \dots, \eta_n$ все равны нулю, как это было в (15) для $r = 1$ и $n = 3$.

При ортогональном преобразовании форма Q остается инвариантной, поэтому в случае $n = 3$ имеем

$$Q = (y_1 - \eta_1)^2 + (y_2 - \eta_2)^2 + (y_3 - \eta_3)^2,$$

или, согласно (15),

$$Q = (y_1 - \eta_1)^2 + y_2^2 + y_3^2. \quad (18)$$

Эта форма достигает минимума \tilde{Q} при $\eta_1 = y_1$, следовательно,

$$\tilde{\eta}_1 = y_1, \tilde{\eta}_2 = \tilde{\eta}_3 = 0$$

и

$$\tilde{Q} = y_2^2 + y_3^2. \quad (19)$$

Равенство (18) является выражением «теоремы Пифагора»:

$$(PX)^2 = (P\tilde{P})^2 + (X\tilde{P})^2. \quad (20)$$

Левая часть равенства (20) представляет собой форму Q , первое слагаемое справа равно $(y_1 - \eta_1)^2$, а второе слагаемое — $\tilde{Q} = y_2^2 + y_3^2$.

В случае r параметров и n наблюдаемых величин получаем следующие обобщения формул (18) и (19):

$$Q = (y_1 - \eta_1)^2 + \dots + (y_r - \eta_r)^2 + y_{r+1}^2 + \dots + y_n^2, \quad (21)$$

$$\tilde{Q} = y_{r+1}^2 + \dots + y_n^2. \quad (22)$$

И в этом общем случае \tilde{Q} можно также истолковать как квадрат расстояния $X\tilde{P}$, но уже в n -мерном пространстве.

Случай неравноточных наблюдений подстановками

$$x_i = x'_i \sigma_i$$

сводится к случаю, который только что был рассмотрен и проиллюстрирован геометрически.

Г. ТЕОРЕМА ГАУССА

Гаусс дал второе обоснование метода наименьших квадратов, опирающееся на следующую теорему:

Среди всех несмещенных оценок параметра ϑ_1 , являющихся линейными функциями наблюдений x_i , наименьшей дисперсией обладает оценка $\tilde{\vartheta}$.

Очень короткое доказательство этой теоремы Гаусса указано в работе Plackett R. L., *Biometrika*, **36** (1949), 458. Здесь будет воспроизведено доказательство, основанное на ортогональном преобразовании (17). Мы ограничимся случаем $r = 1$, $n = 3$, а обобщение для произвольных r и n предоставим читателю.

Пусть T — некоторая оценка параметра ϑ_1 , являющаяся линейной функцией от x_1, x_2, x_3 , а значит, и линейной функцией от y_1, y_2, y_3 :

$$T = c_0 + c_1 y_1 + c_2 y_2 + c_3 y_3. \quad (23)$$

Когда мы говорим, что оценка T является несмещенной, то подразумеваем, что математическое ожидание T является функцией от ϑ_1 , тождественно равной ϑ_1 . Математические ожидания y_2 и y_3 равны нулю, а математическое ожидание y_1 равно η_1 , следовательно,

$$\mathcal{E} T = c_0 + c_1 \eta_1 = c_0 + c_1 a u. \quad (24)$$

По условию теоремы это выражение тождественно по u должно равняться

$$\vartheta_1 = \vartheta_1^0 + u, \quad (25)$$

поэтому

$$c_0 = \vartheta_1^0 \quad \text{и} \quad c_1 = \frac{1}{a}. \quad (26)$$

Далее, имеем

$$T - \mathcal{E} T = c_1 (y_1 - \eta_1) + c_2 y_2 + c_3 y_3. \quad (27)$$

Для того чтобы вычислить дисперсию T , нужно (27) возвести в квадрат и найти математическое ожидание:

$$\begin{aligned} \sigma_T^2 = & c_1^2 \mathcal{E} (y_1 - \eta_1)^2 + 2c_1 c_2 \mathcal{E} (y_1 - \eta_1) y_2 + 2c_1 c_3 \mathcal{E} (y_1 - \eta_1) y_3 + \\ & + c_2^2 \mathcal{E} y_2^2 + 2c_2 c_3 \mathcal{E} y_2 y_3 + c_3^2 \mathcal{E} y_3^2. \end{aligned} \quad (28)$$

Каждое из математических ожиданий в правой части (28) можно легко вычислить, например:

$$\begin{aligned} \mathcal{E} (y_1 - \eta_1)^2 &= \mathcal{E} \left[\sum e_{1i} (x_i - \xi_i) \right]^2 = \\ &= \sum_i \sum_j e_{1i} e_{1j} \mathcal{E} (x_i - \xi_i) (x_j - \xi_j) = \sum_i e_{1i}^2 \sigma^2 = \sigma^2. \end{aligned}$$

Если таким же образом воспользоваться свойствами ортогональности матрицы (e_{ik}) , то найдем, что математические ожидания квадратов и произведений в правой части (28) равны σ^2 и 0 соответственно. Таким образом, получаем

$$\sigma_T^2 = (c_1^2 + c_2^2 + c_3^2) \sigma^2, \quad (29)$$

где c_1 уже задано равенством (26). Следовательно, дисперсия (29) будет минимальной при

$$c_2 = c_3 = 0,$$

поэтому несмещенная линейная оценка с минимальной дисперсией задается формулой

$$T = \vartheta_1^0 + \frac{y_1}{a}.$$

Но в точности такая же оценка получается и по методу наименьших квадратов. Этим завершается доказательство теоремы.

Геометрический смысл этой теоремы заключается в следующем. Пусть через наблюдаемую точку проходит плоскость, параллельная некоторой фиксированной плоскости. Несмещенной линейной оценкой T является то значение параметра, которое соответствует точке пересечения фиксированной прямой G с плоскостью, проходящей через наблюдаемую точку. Надлежащим выбором фиксированной плоскости можно получить любую наперед заданную линейную несмещенную оценку. Если эта плоскость перпендикулярна прямой G , то получается оценка \tilde{b} с наименьшей дисперсией.

§ 32. Оценка дисперсии σ^2

Минимум квадратичной формы

$$Q = \sum (x_i - \xi_i)^2 \quad (1)$$

в предыдущих параграфах был обозначен символом \tilde{Q} . Ортогональным преобразованием (17) § 31 эта форма преобразуется в форму

$$Q = (y_1 - \eta_1)^2 + \dots + (y_r - \eta_r)^2 + y_{r+1}^2 + \dots + y_n^2, \quad (2)$$

минимум которой равен

$$\tilde{Q} = y_{r+1}^2 + \dots + y_n^2. \quad (3)$$

Этот минимум достигается в точке $\tilde{\eta}_1 = y_1, \dots, \tilde{\eta}_r = y_r$.

Математическое ожидание \tilde{Q} представляет собой сумму математических ожиданий квадратов y_{r+1}^2, \dots, y_n^2 . Эти математические ожидания вычисляются так же, как в § 31 Г. Таким образом, получаем

$$\mathbb{E} \tilde{Q} = (n - r) \sigma^2. \quad (4)$$

Отсюда следует, что

$$s^2 = \frac{\tilde{Q}}{n - r} \quad (5)$$

можно использовать в качестве оценки для σ^2 . Эта оценка является несмещенной.

Если дисперсии наблюдений σ_i^2 неодинаковы, то вместо (1) нужно рассматривать форму

$$Q = \sum g_i(x_i - \xi_i)^2. \quad (6)$$

Однако подстановками

$$x'_i = x_i \sqrt{g_i} \quad (7)$$

этот случай можно свести к предыдущему, и поэтому в качестве несмещенной оценки для дисперсии наблюдения с единичным весом снова получается выражение (5).

Ясно, что оценка (5) при малых $n - r$ очень неточна и только с увеличением $n - r$ ее точность повышается. Для уточнения этого высказывания нам нужно исследовать функцию распределения случайной величины \tilde{Q} . С этой целью мы предположим, что x_1, \dots, x_n — независимые нормально распределенные случайные величины с одинаковым квадратичным отклонением σ . Совместная плотность вероятности (x_1, \dots, x_n) задается формулой

$$f(x_1, \dots, x_n) = \sigma^{-n} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum (x_i - \xi_i)^2}. \quad (8)$$

Если опять вместо x_i посредством ортогонального преобразования ввести новые переменные y_i , то плотность вероятности (8) почти не изменится:

$$g(y_1, \dots, y_n) = \sigma^{-n} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^r (y_i - \eta_i)^2 - \frac{1}{2\sigma^2} \sum_{i=r+1}^n y_i^2}. \quad (9)$$

Следовательно, y_1, \dots, y_n — независимые нормально распределенные случайные величины со средними значениями

$$\eta_1, \dots, \eta_r, 0, \dots, 0$$

и квадратичным отклонением σ . Случайные величины $y_{r+1}/\sigma, \dots, y_n/\sigma$ имеют нулевые средние значения и единичные дисперсии. Поэтому сумма их квадратов

$$\chi^2 = \frac{\tilde{Q}}{\sigma^2} = \frac{y_{r+1}^2 + \dots + y_n^2}{\sigma^2} \quad (10)$$

подчиняется распределению χ^2 с $n - r$ степенями свободы.

Математическое ожидание χ^2 равно $n - r$, что согласуется с формулой (5). При больших $n - r$ случайная величина χ^2 распределена приблизительно нормально со средним значением $n - r$ и дисперсией $2(n - r)$. Следовательно, отношение

$$\frac{\chi^2}{n - r} = \frac{\tilde{Q}}{(n - r)\sigma^2} = \frac{s^2}{\sigma^2}$$

распределено приблизительно нормально с единичным средним значением и квадратичным отклонением $\sqrt{2/(n - r)}$. Поэтому при

больших $n - r$ величина s^2 является хорошей оценкой для σ^2 ; при малых $n - r$ эта оценка очень неточна. Доверительные границы для s^2/σ^2 можно получить с помощью таблиц для χ^2 (табл. 6).

При практическом вычислении \tilde{Q} следует пользоваться формулами (19) и (20) из § 30. Из \tilde{Q} , согласно (5), можно получить s^2 , а из s^2 по формулам (11) и (12) из § 31 можно получить приближенные значения для σ_u^2 , σ_v^2 и т. д.:

$$\left. \begin{aligned} s_u^2 &= h^{11}s^2, \\ s_v^2 &= h^{22}s^2, \\ &\dots\dots\dots \end{aligned} \right\} \quad (11)$$

Из (10) следует, что отношение

$$\frac{(n-r)s_u^2}{\sigma_u^2} = \frac{(n-r)s^2}{\sigma^2} = \frac{\tilde{Q}}{\sigma^2} = \chi^2$$

подчиняется распределению χ^2 с $n - r$ степенями свободы. Так как χ^2 не зависит от $y_1 = \tilde{\eta}_1, y_2 = \tilde{\eta}_2, \dots, y_r = \tilde{\eta}_r$, то χ^2 не зависит ни от u, v, \dots . Отсюда, как и в § 28, заключаем, что отношение

$$t = \frac{\tilde{v}_1 - \tilde{v}_1}{s_u} = \frac{u - \xi u}{s_u} = \frac{u - \xi u}{\sigma_u} \frac{\sigma_u}{s_u} = \frac{u - \xi u}{\sigma_u} \frac{\sqrt{n-r}}{\chi} \quad (12)$$

подчиняется t -распределению (распределению Стьюдента) с $n - r$ степенями свободы. Таким образом,

С целью построения доверительного интервала для какого-либо одного из неизвестных параметров $\tilde{v}_1, \tilde{v}_2, \dots$ следует воспользоваться соответствующей оценкой по методу наименьших квадратов $\tilde{v}_1, \tilde{v}_2, \dots$ и для отыскания границ отношения (12) применить таблицы распределения Стьюдента с $n - r$ степенями свободы (см. табл. 7).

Пример 19. В византийских солнечных таблицах¹ указаны моменты вступления солнца в каждую из 12 частей пояса зодиака²:

Весы: 23 сентября	12ч. 00 м. дня	Овен: 20 марта	5ч. 20 м. ночи
Скорпион: 23 октября	3ч. 30 м. дня	Телец: 21 апреля	11ч. 00 м. ночи
Стрелец: 21 ноября	10ч. 30 м. дня	Близнецы: 22 мая	1ч. 40 м. ночи
Козерог: 20 декабря	3ч. 20 м. ночи	Рак: 23 июня	6ч. 31 м. дня
Водолей: 19 января	2ч. 20 м. дня	Лев: 24 июля	3ч. 00 м. ночи
Рыбы: 18 февраля	2ч. 20 м. дня	Дева: 24 августа	0ч. 30 м. ночи

¹ Van der Waerden B. L., Eine byzantinische Sonnentafel, Sitzungsber. Bayer. Akad. München (math.-nat.), (1954), 159.

² Пояс зодиака — ряд созвездий, расположенных вдоль эклиптики — большого круга небесной сферы, по которому совершается видимое годичное движение солнца. Пояс зодиака подразделяется на 12 равных частей, получивших названия соответствующих созвездий. — *Прим. перев.*

Дневное время отсчитывается с 6 час. утра, а ночное — с 6 час. вечера. Если за начало отсчета времени принять момент вступления солнца в ту часть пояса зодиака, которая называется «Весы», то моменты вступления солнца в остальные 11 частей зодиака будут задаваться числами $t_i - t_{-6}$, указанными в третьем столбце таблицы, следующей ниже. В четвертом столбце указано среднее время вступления солнца в каждую из частей зодиака; это среднее время вычислено в предположении, что видимое вращение солнца вокруг земли совершается с постоянной угловой скоростью.

i	λ_i	$t_i - t_{-6}$	Среднее время	l_i
-6	-180	0	0	0
-5	-150	29 д. 15 ч. 30 м.	30 д. 10 ч. 30 м.	- 19 ч. 00 м.
-4	-120	58 д. 22 ч. 30 м.	60 д. 21 ч. 00 м.	- 46 ч. 30 м.
-3	- 90	88 д. 3 ч. 20 м.	91 д. 7 ч. 30 м.	- 76 ч. 10 м.
-2	- 60	117 д. 14 ч. 20 м.	121 д. 18 ч. 00 м.	- 99 ч. 40 м.
-1	- 30	147 д. 14 ч. 20 м.	152 д. 4 ч. 30 м.	-110 ч. 10 м.
0	0	178 д. 5 ч. 20 м.	182 д. 15 ч. 00 м.	-105 ч. 40 м.
1	30	209 д. 11 ч. 00 м.	213 д. 1 ч. 30 м.	- 86 ч. 30 м.
2	60	241 д. 1 ч. 40 м.	243 д. 12 ч. 00 м.	- 58 ч. 20 м.
3	90	272 д. 18 ч. 30 м.	273 д. 22 ч. 30 м.	- 28 ч. 00 м.
4	120	304 д. 3 ч. 00 м.	304 д. 9 ч. 00 м.	- 6 ч. 00 м.
5	150	335 д. 0 ч. 30 м.	334 д. 19 ч. 30 м.	+ 5 ч. 00 м.

Если из моментов, указанных в третьем столбце, вычесть соответствующее среднее время, то получим поправки l_i , величины которых даны в пятом столбце. При этом предполагается, что год состоит из 365 дней и 6 час. Такое предположение вполне допустимо, так как моменты, указанные в византийских таблицах, за очевидным исключением, округлены до чисел, кратных 10 мин. В первом столбце указаны номера границ между частями зодиака, а во втором столбце — угловые координаты λ_i этих границ. За начало отсчета углов выбрана граница между частями «Рыбы» и «Овен», причем направление отсчета совпадает с направлением движения солнца.

Определенная симметрия этих чисел (хотя и приближенная), а также другие родственные тексты наводят на мысль, что византийские таблицы, по-видимому, вычислены на основе теории эксцентриков. Согласно этой теории, солнце S движется с постоянной скоростью по некоторой окружности (по эксцентрику), эксцентрично расположенной относительно неподвижной земли E (рис. 20). Предполагая, что византийские таблицы вычислены по этой теории и что результаты, указанные в таблицах, содержат случайные ошибки, мы постараемся возможно более точно определить эксцентриситет e соответствующего эксцентрика и угловую координату α (на эклиптике) для апогея¹.

Примем радиус эксцентрика за единицу. Пусть λ — угловая координата солнца в момент его вступления в данную часть зодиака. Разность $x = \lambda - \alpha$ называют *истинной аномалией*. Угловое расстояние на эксцентрике между апогеем A и солнцем S называют *средней аномалией*; мы обозначим

¹ Эксцентриситет — отношение расстояния между землей и центром эксцентрика к радиусу эксцентрика, апогей — точка эксцентрика, максимально удаленная от земли. — *Прим. перев.*

ее $x + \omega$. Разность между истинной и средней аномалиями, равная $-\omega$, называется *уравнением центра*. В частности, если $\lambda = 0$, то $x = -\alpha$ и средняя аномалия равна $\omega_0 - \alpha$. По теореме синусов плоской тригонометрии ω , x и эксцентриситет e связаны соотношениями:

$$\sin \omega = e \sin x, \quad \sin \omega_0 = e \sin(-\alpha) \quad (13)$$

или

$$\omega = \arcsin(e \sin x), \quad \omega_0 = -\arcsin(e \sin \alpha). \quad (14)$$

Так как e мало, то правые части (14) можно разложить в степенные ряды, в каждом из которых мы ограничимся первыми двумя членами:

$$\left. \begin{aligned} \omega &= e \sin x + \frac{1}{6} e^3 \sin^3 x = e \sin x + \\ &+ \frac{1}{24} e^3 (3 \sin x - \sin 3x) = \\ &= \left(e + \frac{1}{8} e^3 \right) \sin x - \frac{1}{24} e^3 \sin 3x, \\ \omega_0 &= -\left(e + \frac{1}{8} e^3 \right) \sin \alpha + \frac{1}{24} e^3 \sin 3\alpha. \end{aligned} \right\} (15)$$

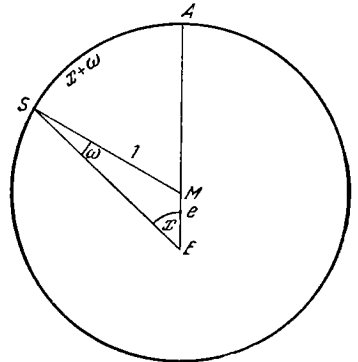


Рис. 20. Движение по эксцентрику.

Если в начальный момент τ_0 солнце имело угловую координату 0 (т. е. находилось в точке 0 эксцентрика), то в точку эксцентрика с угловой координатой λ , отсчитываемой по эклиптике, оно придет в момент времени

$$\tau = \tau_0 + \frac{T}{2\pi} (x + \omega - \omega_0 + \alpha) = \tau_0 + \frac{T}{2\pi} \lambda + \frac{T}{2\pi} (\omega - \omega_0),$$

где $T = 365,25$ дня — период обращения солнца. Очевидно, что величина $T\lambda/2\pi$ равна времени осредненного движения солнца из точки 0 в точку с координатой λ (осредненное движение — движение по эклиптике с постоянной скоростью).

Моменты времени τ_i отличаются от соответствующих моментов t_i византийских таблиц неизвестными ошибками k_i , возникшими в результате ошибок вычислений, опусок в тексте и округления. Таким образом, в каждой точке с координатой λ_i имеет место соотношение

$$t_i - \tau_0 - \frac{T}{2\pi} \lambda_i - \frac{T}{2\pi} (\omega_i - \omega_0) = k_i. \quad (16)$$

Если в (16) вместо ω_i и ω_0 подставить ранее найденные выражения (15), то получим уравнения

$$t_i - \tau_0 - \frac{T}{2\pi} \lambda_i - a (\sin x_i - \sin x_0) - b (\sin 3x_i - \sin 3x_0) = k_i, \quad (17)$$

где

$$a = \frac{T}{2\pi} \left(e + \frac{1}{8} e^3 \right), \quad b = -\frac{T}{2\pi} \frac{1}{24} e^3 = -ce^3$$

и c — известная постоянная. Если, кроме того, воспользоваться равенствами $x_i = \lambda_i - \alpha$, то система (17) преобразуется к виду

$$t_i - \tau_0 - \frac{T}{2\pi} \lambda_i - a[\sin(\lambda_i - \alpha) + \sin \alpha] - b[\sin 3(\lambda_i - \alpha) + \sin 3\alpha] = k_i. \quad (18)$$

Система (18) состоит из 12 уравнений с тремя неизвестными e , α и τ_0 . В качестве приближенных значений этих параметров выберем такие числа \tilde{e} , $\tilde{\alpha}$ и $\tilde{\tau}_0$, для которых сумма квадратов $k_1^2 + \dots + k_{12}^2$ принимает наименьшее значение.

Вычисления станут особенно удобными, если сначала пренебречь малыми слагаемыми, содержащими b . После того как найдено приближенное значение \tilde{e} , можно вычислить $\tilde{b} = -c \tilde{e}^3$ и найти второе приближение, считая b постоянным и равным \tilde{b} . При этом окажется, что второе приближение для a в точности равно первому приближению, так как при построении нормальных уравнений члены, содержащие $\sin 3\alpha$, взаимно уничтожаются. Таким образом, в (18) можно сразу отбросить члены с b и записать систему уравнений так:

$$t_i - \tau_0 - \frac{T}{2\pi} \lambda_i - a \sin \alpha - a \cos \alpha \sin \lambda_i - a \sin \alpha \cos \lambda_i = k_i. \quad (19)$$

Если ввести новые переменные

$$u = a \cos \alpha, \quad v = -a \sin \alpha, \quad w = \tau_0 + a \sin \alpha = \tau_0 - v,$$

то (19) преобразуется в систему 12 линейных уравнений

$$\left(t_i - \frac{T}{2\pi} \lambda_i \right) - u \sin \lambda_i - v \cos \lambda_i - w = k_i, \quad (20)$$

которой соответствует система нормальных уравнений

$$\left. \begin{aligned} [aa]u + [ab]v + [ac]w &= [ad], \\ [ba]u + [bb]v + [bc]w &= [bd], \\ [ca]u + [cb]v + [cc]w &= [cd]. \end{aligned} \right\} \quad (21)$$

Вычисление коэффициентов системы (21) не представляет труда:

$$[aa] = \sum \sin^2 \lambda_i = 6, \quad [ab] = \sum \sin \lambda_i \cos \lambda_i = 0, \quad [ad] = \sum \left(t_i - \frac{T}{2\pi} \lambda_i \right) \sin \lambda_i,$$

$$[bb] = \sum \cos^2 \lambda_i = 6, \quad [ac] = \sum \sin \lambda_i = 0, \quad [bd] = \sum \left(t_i - \frac{T}{2\pi} \lambda_i \right) \cos \lambda_i,$$

$$[cc] = \sum 1 = 12, \quad [bc] = \sum \cos \lambda_i = 0, \quad [cd] = \sum \left(t_i - \frac{T}{2\pi} \lambda_i \right),$$

и поэтому, в данном случае, нормальные уравнения можно записать совсем просто:

$$\left. \begin{aligned} 6u &= \sum \left(t_i - \frac{T}{2\pi} \lambda_i \right) \sin \lambda_i, \\ 6v &= \sum \left(t_i - \frac{T}{2\pi} \lambda_i \right) \cos \lambda_i, \\ 12w &= \sum \left(t_i - \frac{T}{2\pi} \lambda_i \right). \end{aligned} \right\}$$

В указанной выше таблице величины l_i равны $t_i - t_{-6} - T\lambda_i/2\pi - T/2$, поэтому, если воспользоваться очевидными равенствами

$$\sum \sin \lambda_i = \sum \cos \lambda_i = 0,$$

то последние три уравнения можно будет записать так:

$$12 \left[\dot{w} - \left(t_{-6} + \frac{T}{2} \right) \right] = \sum l_i \cdot \left. \begin{array}{l} 6u = \sum l_i \sin \lambda_i, \\ 6v = \sum l_i \cos \lambda_i, \end{array} \right\} \quad (22)$$

Если решения (22) \tilde{u} , \tilde{v} и \tilde{w} найдены, то $\tilde{\alpha}$, \tilde{a} и $\tilde{\tau}_0$ можно определить из уравнений

$$\left. \begin{array}{l} \tilde{a} \cos \tilde{\alpha} = \tilde{u}, \\ \tilde{a} \sin \tilde{\alpha} = -\tilde{v}, \\ \tilde{\tau}_0 = \tilde{w} + \tilde{v} \end{array} \right\} \quad (23)$$

и, наконец, \tilde{e} — из уравнения

$$\tilde{e} - \frac{1}{8} \tilde{e}^3 = \frac{2\pi}{T} \tilde{a}. \quad (24)$$

В данном случае

$$\sum l_i \sin \lambda_i = 142,87, \quad \sum l_i \cos \lambda_i = -316,61, \quad \sum l_i = -631,$$

поэтому, согласно (22), (23) и (24),

$$\tilde{e} = 0,04157, \quad \tilde{\alpha} = 65^\circ 40', \quad \tilde{\tau}_0 - t_{-6} = 178 \text{ дн. } 5 \text{ ч. } 39 \text{ м.}$$

Гиппарх и Птолемей — создатели теории эксцентриков — полагали

$$e = \frac{1}{24} = 0,04167, \quad \alpha = 65^\circ 30'.$$

Согласие следует признать отличным. Таким образом, нужно считать, что византийские таблицы вычислены на основе модели движения солнца, предложенной Гиппархом.

Если бы мы вычислили по формуле (5) выборочную среднюю ошибку v моментов вступления солнца в каждую отдельную часть зодиака при $n = 12$ и $r = 3$, то нашли бы, что v приблизительно равна 20 мин. Однако эта оценка очень неточна; так, знаменатель $n - r = 9$ не очень велик. Кроме того, ненадежность этой оценки обусловлена отсутствием уверенности в том, что в византийских таблицах ошибки отдельных значений являются независимыми. Если воспользоваться указанными выше точными значениями e и α и вычислить ошибки отдельных значений византийских таблиц, то окажется, что из двенадцати чисел шесть в точности соответствуют модели Гиппарха, а остальные шесть подвержены грубым ошибкам от 30 до 50 мин. При этом имеется два случая, когда моменты вступления солнца в две соседние части зодиака указаны с одинаковыми ошибками, соответственно равными 50 и 30 мин. По-видимому, в византийских таблицах отдельные значения не вычислялись независимо друг от друга.

Только что проведенные вычисления в конечном счете сводятся к гармоническому анализу периодической функции $U(\lambda)$, заданной в 12 точках. Гармонический анализ является очень полезным вспомогательным сред-

ством исследования таких астрономических таблиц, закон составления которых неизвестен. Во многих случаях этот метод вносит полную ясность в рассматривавшийся вопрос¹.

§ 33. Линии регрессии

Пусть y — случайная величина, распределение которой зависит от некоторой независимой переменной x . В экономической статистике x в большинстве случаев является временем, а y — величиной, имеющей статистическое истолкование. Примером такой величины может служить количество выплавленной стали, которое, хотя с течением времени и меняется определенным образом, но, с другой стороны, помимо времени, зависит и от многих других факторов. Обе величины x и y могут оказаться случайными с некоторой определенной зависимостью, как, например, количество браков в текущем году и число новорожденных в следующем году.

Пусть в результате некоторого опыта наблюдались n пар значений $(x_1, y_1), \dots, (x_n, y_n)$ величин x и y . Постараемся исследовать характер функциональной зависимости y от x и с этой целью предположим, например, что y и x связаны соотношением *линейной регрессии*

$$y = \vartheta_0 + \vartheta_1 x + u, \quad (1)$$

причем линия регрессии, уравнение которой имеет вид $y = \vartheta_0 + \vartheta_1 x$, должна проходить достаточно близко от наблюдаемых точек с координатами (x_i, y_i) , так что «случайные отклонения» u в некотором смысле являются наименьшими. Можно сделать также и другие предположения: например, можно считать, что y является многочленом второй степени (*квадратичная регрессия*), а соответствующая линия регрессии — параболой. Более общим является тот случай, когда y представляет собой многочлен более высокой степени («регрессия r -го порядка»)

$$y = \vartheta_0 + \vartheta_1 x + \dots + \vartheta_r x^r + u, \quad (2)$$

или, при периодических колебаниях, тригонометрический многочлен

$$y = \vartheta_0 + \vartheta_1 \cos \omega x + \vartheta_2 \sin \omega x + u.$$

Остатком u измеряется отклонение истинного графика функции $y(x)$ от соответствующей линии регрессии. Требование, согласно которому этот остаток должен быть по возможности малым, можно уточнить, воспользовавшись методом наименьших

¹ Van der Waerden B. L., Die Bewegung der Sonne nach griechischen und indischen Tafeln, Sitzungsber. Bayer. Akad. (math.-nat.) (1952), 219; Krishna Rao I. V. M., The Motion of the Moon in Tamil Astronomy, Centaurus, 4 (1956).

квадратов. Согласно этому методу, в качестве значений параметров v_0, v_1, \dots нужно выбрать такие числа, для которых величина квадратичной формы

$$Q = \sum u_i^2 \quad (3)$$

будет наименьшей. Вычисления выполняются точно так же, как в § 26. Например, в случае линейной регрессии из условия

$$\sum u_i^2 = \sum (y_i - v_0 - v_1 x_i)^2 = \text{minimum}$$

посредством дифференцирования получаем

$$\begin{aligned} & - \sum y_i + v_0 n + v_1 \sum x_i = 0, \\ & - \sum y_i x_i + v_0 \sum x_i + v_1 \sum x_i^2 = 0 \end{aligned}$$

или, в обозначениях Гаусса (см. § 30):

$$\begin{cases} n v_0 + [x] v_1 = [y], \\ [x] v_0 + [xx] v_1 = [xy]. \end{cases} \quad (4)$$

Для линии регрессии r -го порядка аналогичным образом получается система $(r + 1)$ нормальных уравнений

$$\begin{cases} n v_0 + [x] v_1 + \dots + [x^r] v_r = [y], \\ [x] v_0 + [x^2] v_1 + \dots + [x^{r+1}] v_r = [xy], \\ \dots \\ [x^r] v_0 + [x^{r+1}] v_1 + \dots + [x^{2r}] v_r = [x^r y]. \end{cases} \quad (5)$$

Отыскание уравнения регрессии посредством решения системы (5) сопряжено с большим количеством утомительных вычислений. Правая часть уравнения регрессии представляет собой линейную комбинацию многочленов $1, x, x^2, \dots, x^r$. Если эти многочлены предварительно *ортгонализованы*, то при отыскании уравнения регрессии количество вычислений существенно сокращается. Две функции $\varphi(x)$ и $\psi(x)$, определенные в точках x_1, \dots, x_n , называются ортогональными, если

$$\sum \varphi(x_i) \psi(x_i) = 0.$$

Если теперь нам заданы m функций $\varphi_1, \varphi_2, \dots, \varphi_m$, то их можно заменить системой ортогональных функций $\psi_1, \psi_2, \dots, \psi_m$, определяемых соотношениями¹

$$\begin{aligned} \psi_1 &= \varphi_1, \\ \psi_2 &= \varphi_2 - \alpha \psi_1, \\ \psi_3 &= \varphi_3 - \beta \psi_1 - \gamma \psi_2, \\ & \dots \end{aligned}$$

¹ Предполагается, что все m векторов $[\varphi_k(x_1), \dots, \varphi_k(x_n)]$ ($k = 1, 2, \dots, m$) линейно независимы. — Прим. перев.

Постоянная α определяется таким образом, чтобы ψ_1 и ψ_2 были ортогональны; постоянные β и γ определяются двумя условиями ортогональности: ψ_3 к ψ_1 и ψ_3 к ψ_2 и т. д.

Любую линейную комбинацию

$$\vartheta_1 \varphi_1 + \dots + \vartheta_m \varphi_m$$

можно записать как

$$\mu_1 \psi_1 + \dots + \mu_m \psi_m.$$

Если теперь для определения μ снова воспользоваться методом наименьших квадратов, то каждое нормальное уравнение будет содержать лишь какое-либо одно неизвестное μ_i , и поэтому решение можно выписать непосредственно, минуя утомительные вычисления, связанные с решением системы (5).

В случае линейной регрессии вычисления выполняются так: исходными функциями равенства (1) являются 1 и x , а соответствующими ортогональными функциями

$$\psi_0 = 1 \text{ и } \psi_1 = x - \bar{x},$$

где \bar{x} — арифметическое среднее всех x_i .

Дифференцированием устанавливаем, что сумма квадратов

$$\sum (y - \mu_0 \psi_0 - \mu_1 \psi_1)^2$$

достигает наименьшего значения при тех значениях μ_0 и μ_1 , которые удовлетворяют уравнениям

$$\left. \begin{aligned} - \sum y \psi_0 + \mu_0 \sum \psi_0^2 &= 0, \\ - \sum y \psi_1 + \mu_1 \sum \psi_1^2 &= 0 \end{aligned} \right\} \quad (6)$$

или

$$\left. \begin{aligned} \mu_0 n &= \sum y, \\ \mu_1 \sum (x - \bar{x})^2 &= \sum y (x - \bar{x}). \end{aligned} \right\} \quad (7)$$

Если, ради простоты обозначений, положим $m_0 = \tilde{\mu}_0$ и $m_1 = \tilde{\mu}_1$, то решения уравнений (7) можно записать в виде

$$m_0 = y = \frac{1}{n} \sum y, \quad (8)$$

$$m_1 = \frac{\sum (x - \bar{x}) y}{\sum (x - \bar{x})^2} = \frac{\sum (x - \bar{x}) (y - \bar{y})}{\sum (x - \bar{x})^2}. \quad (9)$$

Формула

$$y - \bar{y} = m_1 (x - \bar{x}) \quad (10)$$

называется уравнением эмпирической линии регрессии. Эта линия представляет собой прямую с угловым коэффициентом m_1 , который является случайной величиной и называется выборочным коэффи-

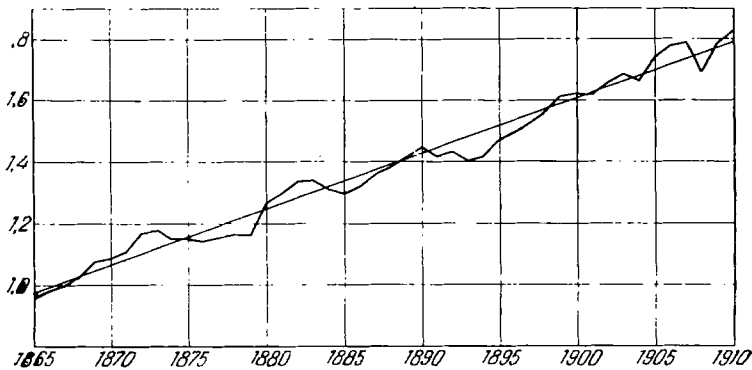
t	x	y	$t - a$	$y - b$	$(t - a)^2$	$(t - a)(y - b)$
1865	9,10	959	-25	-441	625	+ 11 025
1866	9,66	985	-24	-415	576	+ 9 960
1867	10,06	1003	-23	-397	529	+ 9 131
1868	10,71	1030	-22	-370	484	+ 8 140
1869	11,95	1077	-21	-323	441	+ 6 783
1870	12,26	1088	-20	-312	400	+ 6 240
1871	12,85	1109	-19	-291	361	+ 5 529
1872	14,84	1172	-18	-228	324	+ 4 104
1873	15,12	1180	-17	-220	289	+ 3 740
1874	13,92	1144	-16	-256	256	+ 4 096
1875	14,12	1150	-15	-250	225	+ 3 750
1876	13,96	1145	-14	-255	196	+ 3 570
1877	14,19	1152	-13	-248	169	+ 3 224
1878	14,54	1162	-12	-238	144	+ 2 856
1879	14,41	1159	-11	-241	121	+ 2 651
1880	18,58	1269	-10	-131	100	+ 1 310
1881	19,82	1297	- 9	-103	81	+ 927
1882	21,56	1334	- 8	- 66	64	+ 528
1883	21,76	1338	- 7	- 62	49	+ 434
1884	20,46	1311	- 6	- 89	36	+ 534
1885	19,84	1298	- 5	-102	25	+ 510
1886	20,81	1318	- 4	- 82	16	+ 328
1887	22,82	1358	- 3	- 42	9	+ 126
1888	24,03	1381	- 2	- 19	4	+ 38
1889	25,88	1413	- 1	+ 13	1	- 13
1890	27,87	1445	0	45	0	0
1891	26,17	1418	1	18	1	18
1892	26,92	1430	2	30	4	60
1893	25,26	1402	3	2	9	6
1894	26,03	1416	4	16	16	64
1895	29,37	1468	5	68	25	340
1896	31,29	1495	6	95	36	570
1897	33,46	1525	7	125	49	875
1898	36,46	1562	8	162	64	1 296
1899	40,87	1611	9	211	81	1 899
1900	41,35	1616	10	216	100	2 160
1901	41,14	1614	11	214	121	2 354
1902	44,73	1651	12	251	144	3 012
1903	46,82	1670	13	270	169	3 510
1904	46,22	1665	14	265	196	3 710
1905	54,79	1739	15	339	225	5 085
1906	59,66	1776	16	376	256	6 016
1907	61,30	1787	17	387	289	6 579
1908	48,80	1688	18	288	324	5 184
1909	60,60	1782	19	382	361	7 258
1910	66,20	1821	20	421	400	8 420
		63413	-115	-987	8395	147 937

центом регрессии. Если y_i — случайные величины, а значения x_i от случая не зависят (например, x_i — заданные моменты времени), то, пользуясь результатами § 30, можно определить среднее значение и дисперсию выборочного коэффициента регрессии m_1 .

В том случае, когда x является временем, регрессия носит название *тренд*¹.

Пример 20. Во втором столбце таблицы на стр. 178 указано количество чугуна, которое ежегодно выплавлялось во всем мире с 1865 по 1910 г.². Постараемся возможно наилучшим образом разложить изменение выплавки на тренд и конъюнктурные колебания.

В таблице используются следующие обозначения: t — номер года, x — количество выплавленного чугуна (в миллионах тонн), y — десятичный



Р и с. 21. Логарифм количества чугуна, которое ежегодно выплавлялось во всем мире с 1865 по 1910 г.

логарифм x , умноженный на 1000 ($y = 1000 \lg x$). С целью получения более удобных малых чисел из всех табличных значений t и y были вычтены $a = 1890$ и соответственно $b = 1400$.

Если предварительно по заданным числам t и x в плоскости tOx построить грубую кривую, то обнаружится, что с ростом t эта кривая поднимается вверх значительно быстрее, чем прямая линия или квадратная парабола. Следовательно, нет оснований считать регрессию линейной или квадратичной. Напротив, показательная функция оказывается хорошо согласующейся с табличными значениями. При этом колебания эмпирической кривой около показательной с ростом t усиливаются. Это наводит на мысль перейти от абсолютных чисел x к их логарифмам y и затем в плоскости tOy постараться найти такую прямую, которая наилучшим образом согласуется с табличными значениями (t, y).

¹ Если $y(x)$ — случайные величины, распределение которых зависит от времени x , то трендом называют такую функцию $\hat{y}(x)$, значения которой в каждой точке x равны среднему значению $\mathcal{E} y(x)$. — *Прим перев.*

² Cassel G., Theoret. Sozialökonomie, 3 Aufl., 587, Figur S. 532.
12*

Находим

$$\bar{t} = 1890 - \frac{115}{46} = 1890 - 2,5 = 1887,5,$$

$$m_0 = \bar{y} = 1400 - \frac{987}{46} = 1400 - 21 = 1379.$$

Точка с координатами (\bar{t}, m_0) принадлежит линии регрессии, которая является прямой с угловым коэффициентом

$$m_1 = \frac{\sum (t - \bar{t})(y - \bar{y})}{\sum (t - \bar{t})^2} = \frac{147\,937 - 2,5 \cdot 987}{8395 - 46 \cdot (2,5)^2} = \frac{145\,470}{8107,5} = 17,94.$$

Уравнение линии регрессии задается формулой

$$y = m_0 + m_1 (t - \bar{t}) = \bar{y} + m_1 (t - \bar{t}).$$

следовательно, в данном случае

$$y = 1379 + 17,94 (t - 1887,5).$$

Рисунок 21 показывает, что эмпирическая линия регрессии очень хорошо согласуется с основным характером роста эмпирической ломаной линии. Это приближение можно еще несколько улучшить, добавив к правой части уравнения регрессии квадратичный член $m_2 \psi_2$, где

$$\psi_2 = (t - \bar{t})^2 - \gamma.$$

Постоянная γ подбирается таким образом, чтобы функция ψ_2 была ортогональна постоянной $\psi_0 = 1$:

$$\sum_t \psi_0(t) \psi_2(t) = 0.$$

Это приводит к условию

$$\sum_t (t - \bar{t})^2 - 46\gamma = 0.$$

И так как $\sum (t - \bar{t})^2 = 8107,5$, то $\gamma = 176,25$.

Метод ортогонализации имеет то преимущество, что для отыскания квадратичного приближения не нужно заново пересчитывать уже вычисленные коэффициенты m_0 и m_1 линейного приближения. Достаточно лишь вычислить m_2 из третьего нормального уравнения и новый член $m_2 \psi_2$ прибавить к правой части линейного уравнения регрессии. Осуществление этого плана предоставляется читателю.

§ 34. Выяснение причин изменения экономических показателей

Если некоторый экономический показатель w зависит от величин x, y, \dots и, кроме того, подвержен влиянию других, не поддающихся учету факторов, то можно попытаться найти возможно более тесную зависимость между изменением w и изменением x, y, \dots , что открывает доступ к теоретическому расчету динамики показателя w .

Классическим примером такого рода исследований служит работа А. Хану¹ о циклических колебаниях цены на свиней. Высокая цена на свиней является для крестьян стимулом к усилению интенсивности свиноводства. Вследствие этого примерно через полтора года количество свиней на рынке увеличивается и цена на них падает. С этого момента начинается обратный процесс и т. д. Если никакие другие причины не нарушают течения такого процесса, то цена на свиней будет претерпевать колебания с периодом примерно в три года.

При изучении конъюнктуры причинная зависимость оказывается не такой простой, как в приведенном выше примере. И все же стоит попытаться исследовать, как далеко мы можем продвинуться в причинном объяснении явлений. При этом исходными данными являются те значения x, y, \dots и w , которые наблюдались в течение определенного ряда лет. Из каждого показания вычитают арифметическое среднее за период наблюдения (благодаря этому выборочные средние оказываются равными нулю), и подбором надлежащих функций времени (как правило, линейных) в полученных рядах наблюдений устраняют временной тренд. Таким образом, вместо исходных наблюдений величин x, y, \dots и w получают некоторые новые ряды значений, подверженных лишь периодическим и нерегулярным колебаниям. Далее делается предположение, что причинная зависимость колебаний w от колебаний x, y, \dots приближенно линейная, и поэтому, обозначив колебания теми же буквами, которыми ранее обозначались соответствующие величины, можно написать

$$w = \lambda x + \mu y + \dots + u, \tag{1}$$

причем u представляет собой остаток, возникающий вследствие невыявленных причин. Предполагается, что величина u столь мала, что не оказывает на w заметного влияния, и поэтому в качестве оценок для неизвестных коэффициентов выбирают такие значения λ, μ, \dots , для которых сумма квадратов значений u становится минимальной:

$$[uu] = \sum u_i^2 = \text{minimum}. \tag{2}$$

Если $[uu]$ продифференцировать по λ, μ, \dots и производные приравнять нулю, то, как в § 30, получим систему нормальных уравнений

$$\left. \begin{aligned} \lambda [xx] + \mu [xy] + \dots &= [xw], \\ \lambda [yx] + \mu [yy] + \dots &= [yw], \\ \dots &\dots \end{aligned} \right\} \tag{3}$$

¹ Хану А., Die Prognose der Schweinpreise, Sonderheft 18 der Vierteljahrhefte zur Konjunkturforschung, Berlin, 1930.

Решая систему (3), можно найти оценки для неизвестных коэффициентов λ, μ, \dots

Если предполагается, что экономический показатель x оказывает влияние на подлежащий объяснению показатель w с некоторым запаздыванием (как в предыдущем примере повышение цены на свиней вызывало, с запаздыванием на полтора года, повышение предложения на рынке) и если такое предположение является теоретически оправданным, то при вычислениях это запаздывание учитывается посредством смещения значений x во времени. С этой целью проще всего предварительно выяснить, при каких смещениях значений x по времени получается наибольшая корреляция между x и w . Таким образом, сначала вычисляют коэффициент корреляции между x_i и w_i , затем между x_{i-1} и w_i , между x_{i-2} и w_i и т. д. (эти смещения варьируются, конечно, в умеренных границах, соответствующих разумным теоретическим соображениям) и выбирают такое запаздывание по времени, для которого коэффициент корреляции получается наибольшим. Учитывая это наиболее возможное запаздывание, снова делают предположение (1). Имеется и другой путь, согласно которому неизвестное запаздывание считают дополнительным неизвестным параметром и вместе с остальными параметрами оценивают его методом наименьших квадратов. С этой целью, последовательно, для различных значений запаздывания составляют системы нормальных уравнений, решают их и вычисляют $[uu]$. В качестве оценки принимают такую величину запаздывания, которой соответствует наименьшая сумма квадратов $[uu]$.

Примеры использования этого метода можно найти в работе Tinbergen J., *Business Cycles in the United States*. Publ. Völkerbund, Genf, 1939. С появлением этого основополагающего труда к применению метода, изложенного выше, стали относиться более осторожно: сначала посредством какого-либо критерия «независимости» стремятся убедиться, не слишком ли сильна зависимость между «независимыми величинами» x, y, \dots . Не имея возможности углубляться здесь в эти тонкие методы эконометрики, сошлемся лишь на работы: Tintner G., *Econometrics*, New York and London, 1952; Klein L. R., *A textbook of Econometrics*, Evanston and New York, 1953; Hood W. C. and Koopmans T. C., *Studies in Econometric Method*, Cowles Monograph No. 14, New York (Wiley), 1953.

ОЦЕНКИ НЕИЗВЕСТНЫХ ПАРАМЕТРОВ

Эта глава распадается на четыре части. Первая часть (§ 35 и 36) посвящена методу наибольшего правдоподобия и поясняющим примерам. Эта часть в первую очередь предназначена для тех читателей, которые еще не знакомы с этим методом. Тот, кто при практическом применении метода наибольшего правдоподобия столкнется со сложными уравнениями, сможет в § 36 найти вспомогательные указания для отыскания корней.

Во второй части (§ 37—39) показано, что для оценок неизвестного параметра существует некоторая граница точности и никакая оценка не может иметь точность выше этой границы. В некоторых случаях метод наибольшего правдоподобия является наилучшим методом отыскания оценок, так как часто точность оценок наибольшего правдоподобия достигает границы точности. Вспомогательным средством этой второй части является неравенство Фреше.

Для пояснения постановки задачи вторая часть очень полезна, но логически она не представляется необходимой.

В третьей части (§ 40—44) развивается метод, с помощью которого можно достигнуть большего, чем с помощью упомянутого неравенства. Метод третьей части приводит к наиболее точным несмещенным оценкам даже в тех случаях, когда метод наибольшего правдоподобия перестает действовать.

Совсем короткая четвертая часть (§ 45) содержит обзор асимптотических свойств оценок наибольшего правдоподобия.

В этой главе в большинстве случаев предполагается, что результаты наблюдений, служащие отправным пунктом при построении оценки неизвестного параметра, являются непрерывными случайными величинами x_1, \dots, x_n . В следующей, гл. 9 будет рассмотрен случай, когда результаты наблюдений являются частотами. Однако примеры этого рода будут встречаться уже и в этой главе (примеры 21, 28 и 31).

§ 35. Метод наибольшего правдоподобия Р. А. Фишера

Как мы видели в § 30, основой гауссовского обоснования метода наименьших квадратов является принцип, согласно которому наилучшими значениями неизвестных параметров $\vartheta_1, \dots, \vartheta_r$ являются те значения, при которых результат наблюдений имеет наибольшую вероятность¹. Этот принцип Р. А. Фишер использовал в качестве основы общего метода, позволяющего оценить неизвестные параметры $\vartheta_1, \dots, \vartheta_r$ в том случае, когда результаты наблюдений являются случайными величинами с распределением, зависящим от $\vartheta_1, \dots, \vartheta_r$.

Наблюдаемые величины x_1, \dots, x_n могут быть дискретными или непрерывными. Введем функцию

$$g(t | \vartheta) = g(t_1, \dots, t_n | \vartheta_1, \dots, \vartheta_r)$$

и будем считать, что в дискретном случае $g(t | \vartheta)$ — вероятность того, что величины x_1, \dots, x_n соответственно равны t_1, \dots, t_n . Если же x_1, \dots, x_n — непрерывные случайные величины, то $g(t | \vartheta) = g(t_1, \dots, t_n | \vartheta_1, \dots, \vartheta_r)$ будем истолковывать как плотность совместного распределения величин x_1, \dots, x_n . Напомним, что в теории метода наименьших квадратов $g(t | \vartheta)$ представляет собой произведение гауссовых функций ошибок

$$g(t | \vartheta) = \sigma^{-n} (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum (t_i - \xi_i)^2}, \quad (1)$$

где «истинные значения» ξ_i являются заданными функциями от ϑ . Теперь мы откажемся от этого специального предположения и будем считать, что $g(t | \vartheta)$ — произвольная плотность вероятности, зависящая от ϑ .

Непосредственно наблюдаемые значения x_i Фишер подставил в $g(t | \vartheta)$ (т. е. положил $t_i = x_i$) и полученную функцию $g(x | \vartheta)$ от $\vartheta_1, \dots, \vartheta_r$ назвал *функцией правдоподобия*. Те значения параметров ϑ , для которых функция правдоподобия достигает максимума, называются *правдоподобными значениями* параметров ϑ . Согласно *методу наибольшего правдоподобия*, в качестве оценок для истинных значений параметров ϑ выбирают правдоподобные значения ϑ .

Логарифм функции $g(x | \vartheta)$ мы в дальнейшем будем обозначать² символами $L(x | \vartheta)$ или $L(\vartheta)$.

¹ Так как x_1, \dots, x_n — непрерывные случайные величины, то вероятность, о которой говорит автор, всегда равна нулю. На самом деле, максимальной должна быть не вероятность, а соответствующая плотность вероятности (см. (3) § 30). — *Прим. перев.*

² Функцию $L(x/\vartheta) = \ln g(x/\vartheta)$ называют логарифмической функцией, правдоподобия. — *Прим. перев.*

Функцию правдоподобия не следует смешивать с вероятностью. Хотя эта функция и определяется с помощью вероятностей (в дискретном случае) или плотности вероятности (в непрерывном случае), однако и вероятности, и плотность вероятности относятся не к неизвестным параметрам, а к результатам наблюдений. Параметры совсем не зависят от случая и поэтому не имеют плотности вероятности. Для каждого фиксированного результата наблюдений некоторые значения параметров могут оказаться более правдоподобными, так как при таких значениях вероятность получения наблюдаемых результатов имеет заметную величину, а другие — менее правдоподобными, потому что наблюдаемые результаты при таких предположениях являются весьма маловероятными.

Пусть $g(t | \vartheta)$ — плотность совместного распределения случайных величин x_1, \dots, x_n . Если вместо x_i ввести новые случайные величины x'_i по формулам $x_i = \varphi_i(x'_1, \dots, x'_n)$, то соответствующая плотность вероятности будет иметь вид (см. (2) § 11)

$$h(t' | \vartheta) = g(t | \vartheta) \left| \frac{\partial(t_1, \dots, t_n)}{\partial(t'_1, \dots, t'_n)} \right|,$$

где $t_i = \varphi_i(t'_1, \dots, t'_n)$. Таким образом, $h(t' | \vartheta)$ отличается от $g(t | \vartheta)$ множителем, зависящим лишь от t' , и, следовательно, при замене переменных t_i точка максимума функции правдоподобия остается неизменной. Точно так же в дискретном случае с целью упрощения функции $g(t | \vartheta)$ мы будем считать себя вправе умножать ее на некоторый множитель, зависящий лишь от t . Точка максимума g как функции ϑ от этого, очевидно, не изменится.

В качестве примеров использования метода наибольшего правдоподобия в первую очередь служат все примеры отыскания оценок методом наименьших квадратов (гл. VI).

Теперь мы укажем три новых примера, имеющих принципиальный интерес.

Пример 21. Оценка неизвестной вероятности.

Рассмотрим последовательность n независимых испытаний, в каждом из которых положительный исход наступает с постоянной и неизвестной вероятностью p . Пусть x — наблюдаемое число положительных исходов. Каково наиболее правдоподобное значение p ?

Согласно формуле Бернулли (§ 5 А), функция правдоподобия имеет вид

$$\binom{n}{x} p^x (1-p)^{n-x}$$

или, если отбросить число сочетаний, зависящее лишь от x ,

$$g(x|p) = p^x (1-p)^{n-x}. \quad (2)$$

При отыскании максимума вместо самой функции $g(x|p)$ можно воспользоваться ее логарифмом

$$L(p) = x \ln p + (n-x) \ln (1-p).$$

Дифференцированием по p получаем

$$L'(p) = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x-np}{p(1-p)}.$$

Производная $L'(p)$ обращается в нуль при $p = x/n$, причем при $p < x/n$ она положительна, а при $p > x/n$ — отрицательна. Поэтому $p = x/n$ является точкой максимума функции $L(p)$. Следовательно, правдоподобное значение p равно

$$\tilde{p} = h = \frac{x}{n}. \quad (3)$$

Оценка (3) является несмещенной: математическое ожидание h в точности равно истинному значению p . Далее, оценка (3) *состоятельная*, т. е. h стремится, по вероятности, к истинному значению p при $n \rightarrow \infty$. Это является следствием закона больших чисел (§ 5 и 33).

Пример 22. Случайная величина x имеет нормальную плотность вероятности

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}$$

с неизвестным средним значением μ и неизвестной дисперсией σ^2 . Пусть в результате наблюдений получены n независимых значений x_1, \dots, x_n случайной величины x . Каковы правдоподобные значения μ и σ^2 ?

Если плотность совместного распределения x_1, \dots, x_n умножить на несущественный множитель $(2\pi)^{n/2}$, то функция правдоподобия окажется равной

$$g(x_1, \dots, x_n | \mu, \sigma) = \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2},$$

а ее логарифм

$$L(\mu, \sigma) = -n \ln \sigma - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2. \quad (4)$$

Второй член справа является отрицательно определенным квадратичным многочленом относительно μ . Точку максимума $\tilde{\mu}$ этого многочлена находим дифференцированием (4) по μ :

$$\left. \begin{aligned} \sum (x_i - \tilde{\mu}) &= 0, \\ \tilde{\mu} &= \frac{1}{n} \sum x_i = \bar{x}. \end{aligned} \right\} \quad (5)$$

Следовательно, правдоподобное значение μ является арифметическим средним из наблюдаемых значений x . Гаусс нашел этот же результат методом наименьших квадратов.

Если $\tilde{\mu}$ подставить в (4) и затем (4) продифференцировать по σ , то получим

$$\frac{d}{d\sigma} L(\tilde{\mu}, \sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum (x_i - \bar{x})^2.$$

Эта производная обращается в нуль при

$$n\sigma^2 = \sum (x_i - \bar{x})^2.$$

Если $\sigma^2 < \sum (x_i - \bar{x})^2/n$, то производная положительна, если же $\sigma^2 > \sum (x_i - \bar{x})^2/n$, то она отрицательна. Следовательно, правдоподобным значением σ^2 является

$$\tilde{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2. \quad (6)$$

Ранее вместо (6) мы имели приближенное значение

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2, \quad (7)$$

где знаменатель $n-1$ был выбран для того, чтобы среднее значение s^2 в точности равнялось σ^2 . Очевидно, что математическое ожидание (6) несколько меньше математического ожидания (7). Следовательно, оценка наибольшего правдоподобия (6) является смещенной: ее математическое ожидание не равно истинному значению σ^2 .

В этом примере смещение оценки $\tilde{\sigma}^2$ мало: в пределе при $n \rightarrow \infty$ оно исчезает. Дисперсия оценки $\tilde{\sigma}^2$ стремится к нулю при $n \rightarrow \infty$. В силу этих двух свойств, по неравенству Чебышева (§ 3 В) эта оценка является *состоятельной*¹.

Пример 23. В этом примере метод наибольшего правдоподобия не приводит к состоятельной оценке.

В лаборатории измерялись n концентраций, каждая из которых определялась дважды. Точность измерений была одинаковой во всех опытах, а истинные значения n концентраций были различными. Предположим, что все $2n$ результатов измерений $x_1, y_1; \dots; x_n, y_n$ являются независимыми нормальными случайными величинами с совместной плотностью вероятности

$$g(x_i, y_i | \sigma, \mu_i) = \frac{1}{\sigma^2 n} \frac{1}{(2\pi)^n} e^{-\frac{1}{2\sigma^2} \sum [(x_i - \mu_i)^2 + (y_i - \mu_i)^2]}. \quad (8)$$

Неизвестными являются n средних значений μ_1, \dots, μ_n и дисперсия σ^2 . Правдоподобные значения для μ_i , конечно, снова равны арифметическим средним

$$\tilde{\mu}_i = \frac{1}{2} (x_i + y_i).$$

Если все $\tilde{\mu}_i$ подставить в (8), то получим

$$(2\pi)^n g(x_i, y_i | \sigma, \tilde{\mu}_i) = \frac{1}{\sigma^{2n}} e^{-\frac{1}{4\sigma^2} \sum (x_i - y_i)^2}.$$

Вычисляя логарифмическую производную последнего выражения, найдем, что функция правдоподобия достигает максимума в точке $\sigma^2 = \tilde{\sigma}^2$, где

$$\tilde{\sigma}^2 = \frac{1}{4n} \sum (x_i - y_i)^2. \quad (9)$$

¹ Если $\tilde{\vartheta}$ — смещенная оценка параметра ϑ , причем $\mathcal{E}(\tilde{\vartheta} - \mathcal{E}\tilde{\vartheta})^2 \rightarrow 0$ и $(\mathcal{E}\tilde{\vartheta} - \vartheta)^2 / \mathcal{E}(\tilde{\vartheta} - \mathcal{E}\tilde{\vartheta})^2 \rightarrow 0$ ($n \rightarrow \infty$), то $\tilde{\vartheta}$ называют асимптотически несмещенной оценкой для ϑ . В данном случае $\mathcal{E}\tilde{\sigma}^2 = (1-1/n)\sigma^2$ и $\mathcal{E}(\tilde{\sigma}^2 - \mathcal{E}\tilde{\sigma}^2)^2 = 2\sigma^4(n-1)/n^2$, поэтому $\tilde{\sigma}^2$ — асимптотически несмещенная оценка для σ^2 . — *Прим. перев.*

Математическое ожидание $\tilde{\sigma}^2$ равно

$$E(\tilde{\sigma}^2) = \frac{1}{2} \sigma^2$$

— вдвое меньше σ^2 . Следовательно, в этом случае метод наибольшего правдоподобия дает оценку для σ^2 с постоянным отрицательным смещением $-\sigma^2/2$.

В качестве несмещенной оценки следовало бы выбрать

$$s^2 = \frac{1}{2n} \sum (x_i - y_i)^2. \quad (10)$$

При всех i разность $x_i - y_i$ распределена нормально с нулевым средним значением и дисперсией $2\sigma^2$, поэтому математическое ожидание $\sum (x_i - y_i)^2$ равно $2n\sigma^2$.

Оценка (10) является состоятельной. Позднее мы докажем, что среди всех несмещенных оценок s^2 обладает наименьшей дисперсией.

Из этих примеров мы видим, что в некоторых случаях метод наибольшего правдоподобия дает хорошие несмещенные оценки, в других случаях — по крайней мере состоятельные оценки, но имеется и третий класс случаев, когда этот метод не приводит ни к каким хорошим результатам.

Возникает вопрос, в каких случаях метод наибольшего правдоподобия хорош и в каких случаях плох? На этот вопрос едва ли можно найти исчерпывающий ответ. Можно сказать лишь следующее. Если имеется много независимых наблюдений x_1, \dots, x_n и лишь один неизвестный параметр или лишь ограниченное число параметров $\vartheta_1, \dots, \vartheta_r$, а функции распределения удовлетворяют определенным условиям регулярности, то метод наибольшего правдоподобия оказывается хорошим и становится все лучше и лучше с возрастанием n . Но если n невелико или если r возрастает одновременно с n (как это было в нашем последнем примере), то на метод наибольшего правдоподобия полагаться нельзя. Для этих случаев имеются другие методы, позволяющие находить наилучшую несмещенную оценку. С одним из таких методов мы познакомимся в § 41.

Но прежде мы несколько полнее изучим метод наибольшего правдоподобия. При этом сначала будем предполагать, что имеется лишь *один* неизвестный параметр ϑ .

§ 36. Вычисление максимума

При практическом отыскании оценок методом наибольшего правдоподобия прежде всего требуется найти решение *уравнения правдоподобия*:

$$L'(x | \tilde{\vartheta}) = 0, \quad (1)$$

где $L'(x | \vartheta)$ — логарифмическая производная функции правдоподобия $g(x | \vartheta)$ по ϑ . Поэтому

$$L'(x | \vartheta) = \frac{g'(x | \vartheta)}{g(x | \vartheta)}.$$

Пусть $\tilde{\vartheta}$ — оценка наибольшего правдоподобия, удовлетворяющая условию (1), ϑ_0 — неизвестное истинное значение параметра ϑ , $\mathcal{E}_0 y$ — математическое ожидание случайной величины y с плотностью вероятности $g(t | \vartheta_0)$ и $\mathcal{E}_\vartheta y$ — математическое ожидание случайной величины y с плотностью вероятности $g(t | \vartheta)$. Штрих всегда будет обозначать дифференцирование по ϑ .

Сначала предположим, что x_1, \dots, x_n — независимые наблюдения, имеющие одинаковую плотность вероятности $f(x | \vartheta)$, зависящую от параметра ϑ . В этом случае

$$g(x | \vartheta) = f(x_1 | \vartheta) \dots f(x_n | \vartheta),$$

следовательно,

$$L'(x | \vartheta) = \sum \varphi(x_k | \vartheta), \quad (2)$$

где $\varphi = f'/f$ — логарифмическая производная функции f .

В отдельных случаях удается найти решение уравнения (1) элементарными средствами; с такими случаями мы познакомимся в § 35. В большинстве же случаев, однако, (1) представляет собой сложное алгебраическое или трансцендентное уравнение, которое можно решить методом последовательных приближений.

Простейший вариант этого метода заключается в следующем. Сначала выбирают какое-либо приближенное значение ϑ_1 и вычисляют $L'(x | \vartheta_1)$ как сумму логарифмических производных, относящихся к отдельным наблюдениям x_k :

$$L'(x | \vartheta_1) = \sum \varphi(x_k | \vartheta_1). \quad (3)$$

Далее, стараются найти улучшенное приближение

$$\vartheta_2 = \vartheta_1 + h \quad (4)$$

по методу Ньютона. С этой целью $L'(x | \vartheta_2)$ разлагают по формуле Тэйлора до членов первого порядка включительно:

$$L'(x | \vartheta_2) \sim L'(x | \vartheta_1) + h L''(x | \vartheta_1).$$

Приравнивая это выражение нулю, получают

$$h = \frac{L'(x | \vartheta_1)}{-L''(x | \vartheta_1)}, \quad (5)$$

где знаменатель вычисляется по формуле

$$-L''(x | \vartheta_1) = -\sum \varphi'(x_k | \vartheta_1). \quad (6)$$

Сумма в правой части (6) равна арифметическому среднему всех φ' , умноженному на n . Если арифметическое среднее заменить соответствующим математическим ожиданием, то можно добиться большого упрощения вычислений. Это математическое ожидание задается формулой

$$\mathcal{E}_0 \varphi'(x | \vartheta_1) = \int \varphi'(t | \vartheta_1) f(t | \vartheta_0) dt,$$

где интегрирование производится по всему множеству значений случайной величины x .

Значение ϑ_0 является неизвестным. Однако поскольку нас интересует лишь приближенное значение параметра, то ϑ_0 можно спокойно заменить величиной ϑ_1 . Таким образом, знаменатель — $L''(x | \vartheta_1)$ в (5) мы заменим величиной $n j(\vartheta_1)$, где

$$j(\vartheta) = - \int \varphi'(t | \vartheta) f(t | \vartheta) dt. \quad (7)$$

Следовательно, вместо (5) получаем

$$h_1 = \frac{L'(x | \vartheta_1)}{n j(\vartheta_1)}. \quad (8)$$

Функцию $j(\vartheta)$ в знаменателе (8) можно представить так:

$$j(\vartheta) = - \int \left(\frac{f'}{f} \right)' f dt = \int \left(\frac{f' f'}{f} - f'' \right) dt. \quad (9)$$

Но $f(t | \vartheta)$ — плотность вероятности, поэтому

$$\int f(t | \vartheta) dt = 1. \quad (10)$$

Продифференцируем дважды равенство (10), предполагая, что его левую часть можно дифференцировать под знаком интеграла. Мы получим

$$\int f'' dt = 0,$$

и поэтому (9) можно записать так:

$$j(\vartheta) = \int \left(\frac{f'}{f} \right)^2 f dt = \mathcal{E}_\vartheta \left(\frac{f'}{f} \right)^2. \quad (11)$$

Если $j(\vartheta)$ умножить на n , то получим выражение, которое Р. А. Фишер назвал *информацией, содержащейся в выборке*:

$$I(\vartheta) = n j(\vartheta) = n \mathcal{E}_\vartheta \left(\frac{f'}{f} \right)^2. \quad (12)$$

В качестве второго приближения для ϑ мы теперь имеем

$$\vartheta_2 = \vartheta_1 + h_1. \quad (13)$$

где

$$h_1 = \frac{L'(x|\vartheta_1)}{I(\vartheta_1)}, \quad (14)$$

причем $L'(x|\vartheta_1)$ и $I(\vartheta_1)$ вычисляются по формулам (2) и (12) соответственно.

Выражение $I(\vartheta)$ можно образовать и тогда, когда x_k имеют различные плотности вероятности f_k . В этом случае

$$I(\vartheta) = \sum_k \mathcal{E}_\vartheta \left(\frac{f'_k}{f_k} \right)^2.$$

Общее выражение для $I(\vartheta)$ имеет вид

$$I(\vartheta) = \mathcal{E}_\vartheta [L'(x|\vartheta)]^2 = \int [L'(t|\vartheta)]^2 g(t|\vartheta) dt, \quad (15)$$

где интегрирование производится по всему n -мерному пространству изменения t .

Если имеются два независимых ряда наблюдений, x_1, \dots, x_m и y_1, \dots, y_n , то информация I , содержащаяся в этих рядах, представляет собой сумму информации отдельных рядов:

$$I(\vartheta) = I_1(\vartheta) + I_2(\vartheta). \quad (16)$$

Функция $I(\vartheta)$ всегда неотрицательна. Если $g(x|\vartheta)$ не зависит от ϑ (в этом случае значения x не содержат никакой информации о ϑ), то $I(\vartheta) = 0$. Эти свойства функции $I(\vartheta)$ могут служить оправданием употребления слова «информация»².

Пример 24. На плоской фольге в неизвестной точке находится источник радиоактивного излучения, посылающий лучи равномерно по всем направлениям пространства. Если параллельно фольге поставить экран, то на этом экране можно наблюдать вспышки, вызываемые радиоактивным излучением. Каким образом по местам вспышек на экране можно определить положение источника излучения на фольге?

Выберем плоскость экрана в качестве координатной плоскости xOy и примем за единицу длины расстояние между фольгой и экраном. Параллельные плоскости экрана и фольги в этом случае имеют уравнения $z = 0$ и $z = 1$ соответственно. Пусть $(\vartheta, \eta, 1)$ — координаты источника излучения.

С целью упрощения задачи предположим, что нас интересует лишь координата ϑ источника и что в соответствии с этим измеряются лишь координаты x_1, \dots, x_n точек экрана, в которых наблюдаются вспышки. Поэтому весь процесс можно спроектировать на плоскость xOz (рис. 22).

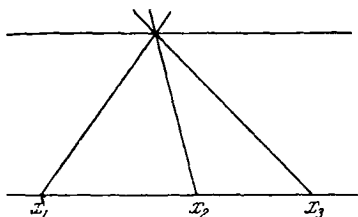
¹ Формулу (15) можно легко проверить, принимая во внимание равенство (2) и выражения n -мерной плотности $g(t|\vartheta)$. — *Прим. ред.*

² О более общем понятии информации и о связи этого понятия с информацией в смысле Фишера см. Колмогоров А. И., Теория передачи информации, изд. АН СССР, М., 1956; Шэннон К., Статистическая теория передачи электрических сигналов (в сборнике «Теория передачи электрических сигналов при наличии помех»), ИЛ, М., 1953. — *Прим. перев.*

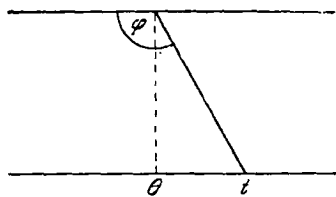
Функция распределения $F(t)$ координаты вспышки x_1 задается вероятностью того, что луч упадет на экран левее точки с координатой t (рис. 23). Все лучи, удовлетворяющие этому условию, заключены внутри угла, величина которого равна

$$\varphi = \frac{\pi}{2} + \text{arc tg } (t - \vartheta), \quad (17)$$

а все лучи, которые вообще могут попасть на экран, заключены внутри



Р и с. 22.



Р и с. 23.

угла, по величине равного π . Следовательно, в силу равномерности распределения лучей, искомая вероятность равна

$$F(t) = \frac{\varphi}{\pi} = \frac{1}{2} + \frac{1}{\pi} \text{arc tg } (t - \vartheta). \quad (18)$$

Таким образом, распределение координат вспышек является распределением Коши. Соответствующая плотность вероятности равна

$$f(t|\vartheta) = \frac{1}{\pi} \frac{1}{(t - \vartheta)^2 + 1}. \quad (19)$$

Функция правдоподобия имеет вид

$$g(x|\vartheta) = \pi^n f(x_1|\vartheta) \dots f(x_n|\vartheta) = \prod_1^n \frac{1}{(x_k - \vartheta)^2 + 1}. \quad (20)$$

Ее логарифм равен

$$L(x|\vartheta) = - \sum \ln [(x_k - \vartheta)^2 + 1]. \quad (21)$$

Дифференцированием находим, что координата ϑ точки максимума функции $L(x|\vartheta)$ должна удовлетворять условию

$$\sum \frac{2(x_k - \vartheta)}{(x_k - \vartheta)^2 + 1} = 0. \quad (22)$$

Само собой разумеется, что при $n = 1$ решением уравнения (22) будет

$$\vartheta = x_1.$$

При $n = 2$ получаем уравнение третьей степени

$$(x_1 - \vartheta) [(x_2 - \vartheta)^2 + 1] + (x_2 - \vartheta) [(x_1 - \vartheta)^2 + 1] = 0$$

или

$$(x_1 + x_2 - 2\vartheta) [(x_1 - \vartheta)(x_2 - \vartheta) + 1] = 0,$$

которое заведомо имеет решение

$$\vartheta_1 = \bar{x} = \frac{1}{2}(x_1 + x_2). \quad (23)$$

Два других решения удовлетворяют квадратному уравнению

$$\vartheta^2 - 2\vartheta \bar{x} + x_1 x_2 + 1 = 0,$$

которое можно записать так:

$$(\vartheta - \bar{x})^2 = \left(\frac{x_1 - x_2}{2}\right)^2 - 1. \quad (24)$$

Если расстояние между точками вспышек x_1 и x_2 меньше двух единиц, то уравнение (24) не имеет действительных корней и решение (23) является точкой максимума функции правдоподобия. Если это расстояние в точности равно двум, то все три корня уравнения правдоподобия оказываются равными друг другу и снова $\vartheta_1 = \bar{x}$ — точка максимума функции правдоподобия. Если же расстояние больше двух, то (23) является точкой минимума, а оба действительных решения уравнения (24) — точками максимумов, одна из которых лежит вблизи от x_1 , а другая — вблизи от x_2 . Метод наибольшего правдоподобия не дает указания, какое из этих двух решений следует выбрать в качестве оценки для ϑ . На практике обычно выбирают то из них, которое расположено ближе к середине фольги.

Для $n > 2$ уравнение (22) решают последовательными приближениями. В качестве первого приближения ϑ_1 выбирают, например, выборочную медиану Z (если n нечетно, то Z — координата средней точки вспышек, см. § 20). Тогда улучшенное приближение имеет вид $\vartheta_2 = \vartheta_1 + h_1$, где

$$h_1 = \frac{L'(x' \vartheta_1)}{I(\vartheta_1)}. \quad (25)$$

Числитель в (25) равен левой части (22) при $\vartheta = \vartheta_1$, а знаменатель представляет собой информацию $I(\vartheta_1) = n j(\vartheta_1)$, где $j(\vartheta)$ можно вычислить по формуле (11):

$$j(\vartheta) = \int_{-\infty}^{\infty} \left(\frac{f'}{f}\right)^2 f dt = \frac{4}{\pi} \int_{-\infty}^{\infty} \left[\frac{t - \vartheta}{(t - \vartheta)^2 + 1}\right]^2 \frac{1}{(t - \vartheta)^2 + 1} dt = \frac{4}{\pi} \int_{-\infty}^{\infty} \frac{u^2 du}{(u^2 + 1)^3} = \frac{1}{2}.$$

Поэтому информация

$$I(\vartheta) = n j(\vartheta) = \frac{n}{2} \quad (26)$$

не зависит от ϑ и, согласно (25),

$$h_1 = \frac{2}{n} L'(x | \vartheta_1) = \frac{4}{n} \sum \frac{x_k - \vartheta_1}{(x_k - \vartheta_1)^2 + 1}. \quad (27)$$

Аналогичным образом можно получить третье приближение. Вообще, если m -е приближение ϑ_m уже найдено, то $(m+1)$ -е приближение получим по формуле

$$\vartheta_{m+1} = \vartheta_m + \frac{4}{n} \sum_k \frac{x_k - \vartheta_m}{(x_k - \vartheta_m)^2 + 1}.$$

Эти последовательные приближения очень быстро сходятся к некоторому пределу $\hat{\vartheta}$, который и принимают в качестве оценки для ϑ . При больших n дисперсия оценки $\hat{\vartheta}$ асимптотически равна

$$\frac{1}{I(\hat{\vartheta})} = \frac{2}{n}. \quad (28)$$

Согласно § 20, дисперсия выборочной медианы дается асимптотической формулой

$$\frac{1}{4n [f(\hat{\vartheta}|\vartheta)]^2} = \frac{\pi^2}{4n}. \quad (29)$$

Сравнение (28) и (29) показывает, что оценка наибольшего правдоподобия лучше выборочной медианы. В свою очередь, арифметическое среднее \bar{x} много хуже выборочной медианы, так как дисперсия \bar{x} не существует, а функция распределения \bar{x} совпадает с функцией распределения отдельного наблюдения x_k .

§ 37. Неравенство Фреше

От хорошей оценки T неизвестного параметра ϑ требуется, чтобы ее значения были, по возможности, близки к истинному значению ϑ . Качество оценки определяется главным образом двумя ее характеристиками: математическим ожиданием $\hat{T} = \mathcal{E} T$ и дисперсией

$$\sigma_T^2 = \mathcal{E}(T - \hat{T})^2.$$

Так как математическое ожидание $\mathcal{E} T$ зависит от ϑ , то мы, вместо $\mathcal{E} T$, снова обозначим его $\mathcal{E}_\vartheta T$. От этого математического ожидания будем требовать, чтобы оно было равно ϑ или по крайней мере было близко к ϑ . Разность

$$\hat{T} - \vartheta = \mathcal{E}_\vartheta T - \vartheta = b(\vartheta) \quad (1)$$

называется *смещением* или *систематической ошибкой* оценки T . От дисперсии σ_T^2 будет требоваться, чтобы она была по возможности мала. Оценка с нулевым смещением и наименьшей дисперсией называется *наилучшей несмещенной оценкой*.

Можно легко указать оценки с нулевой дисперсией: для этого нужно лишь выбрать T равным произвольной постоянной T_0 , независимо от результатов наблюдений. Однако в этом случае, если T_0 сильно отличается от истинного значения ϑ , нам придется иметь дело с большим смещением $T_0 - \vartheta$. Таким образом, обнаруживается противоречие между смещением и дисперсией: обе эти величины нельзя сделать равными нулю (за исключением тривиальных случаев, когда ϑ заранее известно или когда ϑ с вероятностью единица определяется результатами наблюдений).

Эти предварительные выводы можно уточнить с помощью одного неравенства, которое при заданном смещении указывает нижнюю границу дисперсии оценки. Это неравенство было найдено, независимо друг от друга, Фреше, Рао и Крамером¹. В англо-американской литературе оно называется неравенством Крамера—Рао или, как недавно стали его называть, *неравенством информации*.

Если y и z — случайные величины, причем y^2 и z^2 имеют конечные средние значения, то справедливо *неравенство Шварца*²:

$$(\xi yz)^2 \leq (\xi y^2) (\xi z^2). \quad (2)$$

Доказательство справедливости (2) очень просто. Квадратичная форма

$$\xi(\lambda y + \mu z)^2 = \lambda^2 \xi y^2 + 2\lambda\mu \xi yz + \mu^2 \xi z^2 \quad (3)$$

может принимать лишь неотрицательные значения, следовательно, ее дискриминант неположителен:

$$(\xi yz)^2 - (\xi y^2) (\xi z^2) \leq 0. \quad (4)$$

Отсюда непосредственно следует (2). Легко можно убедиться, что неравенство (2) тривиальным образом справедливо и в том случае, когда хотя бы один из множителей в правой части (2) обращается в бесконечность.

Пусть теперь x_1, \dots, x_n — результаты наблюдений и пусть их совместная плотность вероятности³

$$g(x | \vartheta) = g(x_1, \dots, x_n | \vartheta) \quad (5)$$

зависит от единственного неизвестного параметра ϑ . Обозначим через $T = T(x)$ оценку этого параметра. Требуется вывести неравенство для дисперсий σ_T^2 .

Если в некоторой части пространства (x_1, \dots, x_n) функция $g(x | \vartheta)$ обращается в нуль, то при вычислении средних значений эту часть можно исключить из области интегрирования. Таким образом, интегрирование будет производиться лишь в той части

¹ Fréchet M., Rev. Intern. de Stat. (1943), 182; Rao C. R., Bull. Calcutta Math. Soc., 37, 81; Gramér H., Scandinavisk Aktuarie-tidskr., 29, 85, или Математические методы статистики, ИЛ, М., 1948, стр. 517. Далее, Wolfowitz J., Ann. of Math. Stat., 18, 215. О применениях см. Hodges and Lehmann, Proc. Second Berkeley Symposium on Math. Stat., Berkeley (1951), 13.

² Это неравенство опубликовано русским математиком В. Я. Буняковским в 1859 г. — на 25 лет раньше соответствующей публикации немецкого математика Г. А. Шварца, поэтому в советской литературе принято (2) называть неравенством Буняковского. — Прим. перев.

³ С этого момента мы пренебрегаем различием между наблюдаемыми случайными величинами x_1, \dots, x_n и независимыми переменными t_1, \dots, t_n . 13*

пространства, где $g(x | \vartheta) \neq 0$. Предположим, что эта часть не зависит от ϑ и что функция $g(x | \vartheta)$ дифференцируема по ϑ . Если производную по ϑ снова обозначим штрихом, то логарифмическая производная функции g будет равна

$$L'(x | \vartheta) = \frac{g'(x | \vartheta)}{g(x | \vartheta)},$$

где

$$L(x | \vartheta) = \ln g(x | \vartheta).$$

Далее, имеют место равенства

$$\vartheta + b(\vartheta) = \mathcal{E}_{\vartheta} T = \int T g(x | \vartheta) dx \quad (6)$$

и

$$1 = \int g(x | \vartheta) dx. \quad (7)$$

Предположим теперь, что производные от правых частей (6) и (7) можно вычислять дифференцированием под знаком интеграла. Выполняя это дифференцирование, получим

$$1 + b'(\vartheta) = \int T g' dx = \mathcal{E}_{\vartheta} \left(\frac{Tg'}{g} \right) = \mathcal{E}_{\vartheta} (T L'), \quad (8)$$

$$0 = \int g' dx = \mathcal{E}_{\vartheta} \left(\frac{g'}{g} \right) = \mathcal{E}_{\vartheta} L'. \quad (9)$$

Если (9) умножить на \hat{T} и вычесть из (8), то найдем, что

$$1 + b'(\vartheta) = \mathcal{E}_{\vartheta} [(T - \hat{T}) L']. \quad (10)$$

В правой части (10) находится математическое ожидание произведения случайных величин $T - \hat{T}$ и L' . Применяя к этому математическому ожиданию неравенство Шварца (2), получим

$$(1 + b')^2 \leq \sigma_T^2 \mathcal{E}_{\vartheta} (L')^2. \quad (11)$$

Если теперь предположить, что $\mathcal{E}_{\vartheta} (L')^2 \neq 0$ и обозначить $\mathcal{E}_{\vartheta} (L')^2 = I(\vartheta)$, то из (11) следует, что

$$\sigma_T^2 \geq \frac{[1 + b'(\vartheta)]^2}{I(\vartheta)}. \quad (12)$$

Это и есть неравенство Фреше (неравенство информации). Я еще раз сформулирую те предположения, при которых оно было выведено:

1. Часть пространства иксов, в которой $g(x | \vartheta) \neq 0$, не зависит от ϑ .

2. В формулах (6) и (7) допустимо дифференцирование под знаком интеграла.

3. Знаменатель в (12) не равен нулю.

Знаменатель в (12) представляет собой интеграл

$$I(\vartheta) = \mathcal{E}_{\vartheta} [L'(x | \vartheta)]^2 = \int (\ln g)' g' dx, \quad (13)$$

который мы уже ранее, следуя Р. А. Фишеру, назвали «информацией». Другое выражение для $I(\vartheta)$ получается интегрированием (13) по частям¹:

$$I(\vartheta) = -\mathcal{E}_{\vartheta} L''(x | \vartheta). \quad (14)$$

Если при всех ϑ из некоторой окрестности истинного значения параметра ϑ_0 смещение оценки T равно нулю, то числитель (12) при тех же значениях ϑ будет равен единице и мы получаем

$$\sigma_T^2 \geq \frac{1}{I(\vartheta)}. \quad (15)$$

Правая часть (15) не зависит от оценки T . Следовательно, существует нижняя граница для дисперсий несмещенных оценок и этой границей служит величина $1/I(\vartheta)$, обратная информации².

Неравенство Фреше и вытекающие из него следствия остаются справедливыми и в случае дискретных величин x_1, \dots, x_n . Нужно лишь во всех формулах заменить интегралы суммами. При этом предполагается, что суммы, соответствующие формулам (6) и (7), можно дифференцировать почленно. В случае конечных сумм это всегда допустимо.

§ 38. Достаточные оценки и наилучшие оценки

При каких условиях только что выведенные неравенства обращаются в равенства?

Неравенство Шварца (2) § 37, очевидно, обращается в равенство тогда и только тогда, когда форма (3) представляет собой

¹ Интеграл (13) многомерный и обычное интегрирование по частям к нему не применимо. Равенство (14) может быть получено следующим образом: из $L' = g'/g$ получаем $L'' = g''/g - (g'/g)^2$. Но $\mathcal{E}_{\vartheta}(g''/g) = \int g'' dx = 0$, (если в (7) можно дважды дифференцировать под знаком интеграла), поэтому

$$-\mathcal{E}_{\vartheta}[L''(x|\vartheta)] = \mathcal{E}_{\vartheta}(g'/g)^2 = \mathcal{E}_{\vartheta}[L'(x|\vartheta)]^2 = I(\vartheta),$$

откуда и следует (14). — *Прим. ред.*

² Нижняя граница $1/I(\vartheta)$ не обязательно является точной нижней гранью для дисперсий несмещенных оценок. Можно, например, показать, что если x_1, \dots, x_n независимы и нормальны со средним значением $a^{1/3}$ и единичной дисперсией, то нижняя грань дисперсий несмещенных оценок для a равна $(9a^4/n) + (18a^2/n^2) + (6/n^3)$, в то время как $1/I(a) = 9a^4/n$. — *Прим. перев.*

полный квадрат, или, иначе говоря, тогда и только тогда, когда существуют λ и μ не равные нулю одновременно, такие, что сумма $\lambda y + \mu z$ равна нулю с вероятностью единица. В применении к неравенству (12) это означает, что либо T принимает постоянное значение, $T = \hat{T}$ с вероятностью единица, либо с вероятностью единица имеет место равенство

$$L'(x | \vartheta) = K(T - \hat{T}), \quad (1)$$

где K не зависит от x .

В первом случае оценка T принимает постоянное значение T_0 , независимо от наблюдений, и поэтому этот случай мы можем не рассматривать. Если $T = T_0$, то $b(\vartheta) = T_0 - \vartheta$ очень сильно зависит от ϑ . Это случай крайнего «смещения» или, иными словами, случай предвзятого мнения, когда считают, что истинное значение ϑ нужно знать заранее, и вообще не заботятся ни о каких наблюдениях. При некоторых условиях такой подход может оказаться вполне разумным, а именно, тогда, когда предвзятое мнение хорошо обосновано и не опровергается наблюдениями сколько-нибудь убедительно. В этом случае и не возникает никакой проблемы отыскания «точнейшей оценки на основе наблюдений».

Остается случай (1). Интегрированием получаем

$$L(x | \vartheta) = \ln g(x | \vartheta) = A(\vartheta)T + B(\vartheta) + C(x),$$

следовательно,

$$g(x | \vartheta) = e^{A(\vartheta)T + B(\vartheta)} h(x), \quad (2)$$

где A и B зависят лишь от ϑ , а h зависит лишь от x .

Таким образом, неравенство (12) § 36 обращается в равенство тогда, когда выполняются два следующих условия:

а) *Функция правдоподобия $g(x | \vartheta)$ является произведением двух сомножителей*

$$g(x | \vartheta) = e(T | \vartheta) h(x), \quad (3)$$

из которых первый зависит лишь от ϑ и T , а второй — лишь от x .

б) *Первый множитель имеет вид*

$$e(T | \vartheta) = e^{AT+B}, \quad (4)$$

причем A и B зависят лишь от ϑ .

Если условие а) выполняется, то T называется *достаточной оценкой* параметра ϑ (или, по Р. А. Фишеру, *достаточной статистикой*).

Докажем теперь теорему:

Если выполняются условия 1, 2, 3 (§ 37), а также условия а) и б), то среди всех оценок с одинаковым смещением $b(\vartheta)$ наименьшей дисперсией обладает оценка T .

Доказательство. Из (3) и (4) следует, что

$$L'(x | \vartheta) = A'T + B'. \quad (5)$$

В свою очередь из (9) § 37 получаем

$$A'\xi T + B' = \xi(A'T + B') = \xi L' = 0,$$

следовательно,

$$B' = -A'\xi T = -A'\hat{T}'. \quad (6)$$

Если (6) подставить в (5), то убедимся, что

$$L'(x | \vartheta) = A'(T - \hat{T}'). \quad (7)$$

Так как L' пропорционально $T - \hat{T}'$, то неравенство Фреше обратится в равенство

$$\sigma_T^2 = \frac{[1 + b'(\vartheta)]^2}{I(\vartheta)}.$$

Для любой другой оценки справедлив лишь знак \geq . Следовательно, среди всех оценок со смещением $b(\vartheta)$ оценка T имеет наименьшую дисперсию σ_T^2 .

С целью выяснения, в какой мере эти результаты относятся к оценкам наибольшего правдоподобия, мы, помимо а) и б), сделаем еще одно предположение, а именно:

в) Оценка T является несмещенной.

Согласно предположению в),

$$b(\vartheta) = \hat{T}' - \vartheta = 0$$

или $\hat{T}' = \vartheta$. В соответствии с этим, в силу (7), получаем

$$L'(x | \vartheta) = A'(T - \vartheta). \quad (8)$$

Уравнение (10) § 37 в данном случае имеет вид

$$\xi_{\vartheta} [(T - \hat{T}') L'] = 1.$$

Если теперь L' заменить выражением (7), то получим

$$\xi_{\vartheta} [A'(T - \hat{T}')^2] = 1$$

или

$$A' \sigma_T^2 = 1. \quad (9)$$

Отсюда следует, что A' всегда положительно. Далее, из (7) находим

$$I = \xi [L'(x | \vartheta)]^2 = (A')^2 \xi (T - \hat{T}')^2 = (A')^2 \sigma_T^2,$$

поэтому, согласно (9),

$$I = A', \quad (10)$$

т. е. информация I равна производной по ϑ от коэффициента A в формуле (4).

Согласно (8), уравнение правдоподобия имеет вид

$$A'(T - \vartheta) = 0, \quad (11)$$

и так как A' всегда положительно, то это уравнение имеет единственное решение

$$\tilde{\vartheta} = T.$$

В силу (8), $L'(x | \vartheta)$ положительна при $\vartheta < T$ и отрицательна при $\vartheta > T$, поэтому $\vartheta = T$ — точка максимума функции L , а значит, и функции правдоподобия

$$g(x | \vartheta) = e^{L(x|\vartheta)}$$

Таким образом,

если предположения а), б) и в) справедливы, то оценка, найденная методом наибольшего правдоподобия, является несмещенной и наилучшей.

Если же условие в) не выполняется, то можно положить

$$\mathcal{E} T = \tau(\vartheta).$$

В этом случае T является несмещенной наилучшей оценкой для $\tau(\vartheta)$.

§ 39. Примеры

В некоторых важных случаях все условия 1, 2, 3 (§ 37) и а), б), в) (§ 38) оказываются выполненными. Простейшим случаем является следующий.

Пример 25. Оценка среднего значения в нормальном распределении.

Пусть результаты наблюдений x_1, \dots, x_n представляют собой независимые, одинаково нормально распределенные случайные величины с известным средним значением μ . При этом не имеет никакого значения, известно квадратичное отклонение σ или нет. Мы предположим для простоты, что $\sigma = 1$. Тогда функция правдоподобия будет иметь вид (см. § 35, пример 22)

$$g(x | \mu) = e^{-\frac{1}{2} \sum (x - \mu)^2}.$$

Эту формулу можно записать так:

$$g(x | \mu) = e^{-\frac{1}{2} \sum x^2 + \sum x\mu - \frac{n}{2} \mu^2}.$$

Если выборочное среднее обозначить буквой M :

$$M = \frac{1}{n} \sum x,$$

то $g(x|\mu)$ можно будет представить в виде произведения двух сомножителей

$$g(x|\mu) = e^{-n\left(\mu M - \frac{1}{2}\mu^2\right)} \cdot e^{-\frac{1}{2}\sum x^2}.$$

Первый множитель зависит лишь от M и μ , а второй — лишь от x . Следовательно, условие а) выполнено: M является достаточной оценкой для μ .

Проверка справедливости условий 1, 2, 3 (§ 37) не представляет труда. Выполнение условий б) и в) (§ 38) является очевидным. Таким образом, выборочное среднее M является несмещенной наилучшей оценкой для μ .

Пример 26. Оценка дисперсии в нормальном распределении с известным средним значением.

Если среднее значение μ известно, то смещением начала координат на оси Ox можно добиться, чтобы было $\mu = 0$. Тогда плотность совместного распределения x_1, \dots, x_n , с точностью до известного постоянного множителя, будет равна

$$g(x|\sigma) = \sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum x^2}. \quad (1)$$

Требуется найти оценку для $\vartheta = \sigma^2$. Если обозначим $\sum x^2 = ns^2$, то (1) можно записать так:

$$g(x|\sigma) = e^{-n \ln \sigma - \frac{n s^2}{2\sigma^2}}. \quad (2)$$

Эта функция имеет вид $\exp(A s^2 + B)$. Следовательно, оценка $T = s^2$ удовлетворяет условиям а) и б). Математическое ожидание s^2 равно σ^2 , поэтому условие в) также выполняется. Легко можно убедиться и в справедливости условий 1, 2, 3 (§ 37). Таким образом, $s^2 = \sum x^2/n$ является несмещенной, наилучшей оценкой для σ^2 .

Пример 27. Метод наименьших квадратов.

Пусть результаты наблюдений x_1, \dots, x_n представляют собой независимые, нормально распределенные случайные величины с известными квадратичными отклонениями $\sigma_1, \dots, \sigma_n$ соответственно. В § 26 предполагалось, что средние значения ξ_1, \dots, ξ_n наблюдений являются произвольными дифференцируемыми функциями неизвестных параметров $\vartheta_1, \vartheta_2, \dots$; затем эти функции аппроксимировались линейными функциями. Сейчас мы предположим, что все ξ_i являются линейными функциями единственного неизвестного параметра ϑ . Эту теорию можно обобщить на случай нескольких параметров, однако для нелинейных функций она будет справедлива лишь приближенно.

Посредством подстановок $x_i = \sigma_i x'_i$ общий случай можно свести к частному случаю, когда все x_i имеют единичную дисперсию, поэтому, не меняя обозначений, мы будем считать, что $\sigma_1 = \dots = \sigma_n = 1$. В этих предположениях совместная плотность вероятности, с точностью до постоянного множителя, равна

$$g(x|\vartheta) = e^{-\frac{1}{2}\sum (x_i - \xi_i)^2}. \quad (3)$$

Если в (3) заменить ξ_i соответствующими линейными функциями от ϑ

$$\xi_i = c_i + a_i \vartheta \quad (4)$$

(a_i и c_i предполагаются известными), то $g(x|\vartheta)$ примет вид

$$g(x|\vartheta) = e^{\frac{1}{2}(-k\vartheta^2 + 2l\vartheta - m)}, \quad (5)$$

где $k = \sum a_i^2 = \sum aa$ — постоянная величина, а $l = \sum (x_i - c_i) a_i$ и m — соответственно линейная и квадратичная функции от x . Если $k = 0$, то все a_i равны нулю и $g(x|\vartheta)$ вообще не зависит от ϑ ; в этом случае, по терминологии § 30, параметр ϑ не допускает оценки. Но если $k \neq 0$, то (5) можно записать так:

$$g(x|\vartheta) = e^{-\frac{1}{2}k(\vartheta - T)^2 - h(x)}, \quad (6)$$

где

$$T = \frac{\sum ax - \sum ac}{\sum aa}. \quad (7)$$

Ясно, что в (6) показатель степени будет максимальным при $\vartheta = T$. Этот показатель равен показателю степени в (3), который, согласно методу наименьших квадратов, также должен быть сделан максимальным. Следовательно, метод наименьших квадратов приводит к той же оценке T для параметра ϑ , что и метод наибольшего правдоподобия. Вычисляя среднее значение T , найдем, что $\hat{T} = \vartheta$, поэтому оценка T является несмещенной. Так как в данном случае условия а) и б) (§ 38), а также 1, 2, 3 (§ 37) выполнены, то мы получаем результат:

T — несмещенная, наилучшая оценка для ϑ .

Случайная величина T подчиняется нормальному распределению с плотностью вероятности

$$\sqrt{\frac{k}{2\pi}} e^{-\frac{k}{2}(T - \vartheta)^2}.$$

Дисперсия T равна

$$\sigma_T^2 = \frac{1}{k} = \frac{1}{\sum aa}. \quad (8)$$

Постоянная $k = \sum aa$ в точности совпадает с информацией I , так как в данном случае неравенство Фреше обращается в равенство.

Если вспомнить, что посредством преобразований $x_i = \sigma_i x'_i$ все дисперсии наблюдений были сделаны равными единице и что благодаря этому все веса также равны единице, то можно убедиться, что (8) согласуется с ранее полученным результатом (см. § 31 Б):

$$\sigma_u^2 = h^{11} \sigma^2 = \frac{\sigma^2}{[gaa]}. \quad (9)$$

Пример 28. Оценка вероятности.

Пусть вероятность p некоторого события неизвестна и пусть в n независимых опытах это событие наступило x раз. Какова наилучшая оценка для p ?

В данном случае x является дискретной случайной величиной, однако это не может служить причиной каких-либо затруднений. Рассмотрим функцию правдоподобия, вычисленную ранее в первом примере § 35 с точностью до множителя, зависящего лишь от x :

$$g(x|p) = p^x (1 - p)^{n-x}. \quad (10)$$

Эту функцию можно записать так:

$$g(x|p) = e^{x \ln p + (n-x) \ln(1-p)}. \quad (11)$$

Если в (10) положим $x = nh$, где h — частота,

$$h = \frac{x}{n},$$

то получим

$$g(x|p) = e^{hn \ln p + (1-h)n \ln(1-p)}. \quad (12)$$

Это выражение имеет в точности тот же вид, который требуется условиями а) и б) (§ 38). Математическое ожидание оценки h , как мы знаем, равно p . Условия 1, 2, 3 (§ 37) в данном случае также выполняются. Следовательно, частота h является несмещенной, наилучшей оценкой для вероятности p , т. е. среди всех несмещенных оценок h имеет наименьшую дисперсию.

До сих пор во всех рассмотренных случаях для доказательства минимальности дисперсий соответствующих несмещенных оценок мы пользовались тем обстоятельством, что неравенство Фреше обращалось в равенство. Но если условия а) и б) (§ 38) не выполняются, то это неравенство не может обратиться в равенство. Однако существуют другие методы, позволяющие находить несмещенные, наилучшие оценки. Эти методы были разработаны Рао и, при более общих предположениях, Леманном и Шеффе¹. Для подготовки к изложению этих методов нам сначала нужно ознакомиться с понятием условных математических ожиданий по Колмогорову.

§ 40. Условные математические ожидания

Случайные величины мы будем обозначать жирными буквами $\mathbf{t}, \mathbf{u}, \mathbf{v}, \mathbf{x}, \dots$. При этом предполагается, что $\mathbf{x}_1, \dots, \mathbf{x}_n$ — результаты наблюдений, а все остальные случайные величины являются функциями \mathbf{x}_k :

$$\mathbf{t} = T(\mathbf{x}); \mathbf{u} = U(\mathbf{x}); \dots$$

Пусть

$$t = T(x); u = U(x); \dots$$

— отдельные возможные значения этих функций.

Нужно дать определение *условного математического ожидания случайной величины \mathbf{u} для заданного значения t случайной величины \mathbf{t}* .

Если \mathbf{t} и \mathbf{u} принимают лишь конечное число значений, то это определение является очень простым. Пусть $P(t)$ — вероятность того, что \mathbf{t} примет значение, равное t . Если $P(t) \neq 0$, то для всех

¹ Lehmann E. L. and Scheffe H., Completeness, Similar Regions and Unbiased Estimation I, Sankhya (The Indian Journal of Stat.), 10 (1950), 305.

значений u_1, \dots, u_n случайной величины \mathbf{u} можно вычислить условные вероятности¹

$$P(u_k | t) = \frac{P(u_k, t)}{P(t)} = \frac{P\{\mathbf{u} = u_k \cap \mathbf{t} = t\}}{P\{\mathbf{t} = t\}} \quad (1)$$

и затем определить условное математическое ожидание $\mathcal{E}(\mathbf{u} | t)$ как сумму произведений u_k и соответствующих условных вероятностей:

$$\mathcal{E}(\mathbf{u} | t) = \sum u_k P(u_k | t). \quad (2)$$

Если (2) умножить на $P(t)$ и просуммировать по всем значениям t , принадлежащим некоторому множеству M , то, в силу (1), получим.²

$$\sum_{t \in M} \mathcal{E}(\mathbf{u} | t) P(t) = \sum_k u_k P\{\mathbf{u} = u_k \cap (\mathbf{t} \in M)\}. \quad (3)$$

Обратно, если равенство (3) справедливо для любого множества M , то (2) справедливо и для таких множеств, которые состоят лишь из одного значения t . Если в этом случае соответствующее равенство разделим на $P(t)$, то снова получим (2).

Предположение о том, что \mathbf{u} принимает лишь конечное число значений, не является существенным. Ведь конечную сумму в правой части (2) можно заменить бесконечным рядом или интегралом точно так же, как мы это делали в § 3 при определении обычного математического ожидания! Если $F(u | t)$ — функция условного распределения случайной величины \mathbf{u} (значение этой функции в любой точке u определяется как условная вероятность события $\mathbf{u} < u$ при условии, что $\mathbf{t} = t$), то вместо (2) можно написать

$$\mathcal{E}(\mathbf{u} | t) = \int_{-\infty}^{\infty} u dF(u | t), \quad (4)$$

а вместо (3) —

$$\sum_{t \in M} \mathcal{E}(\mathbf{u} | t) P(t) = \int_{M'} \mathbf{u} P(dE), \quad (5)$$

где M' — случайное событие, которое наступает тогда, когда $T(\mathbf{x}) \in M$. Правая часть (5) представляет собой интеграл Лебега

¹ Символом $(\mathbf{u} = u_k) \cap (\mathbf{t} = t)$ обозначено событие, являющееся пересечением событий $\mathbf{u} = u_k$ и $\mathbf{t} = t$, см. § 1. Поэтому $P(u_k, t) = P\{(\mathbf{u} = u_k) \cap (\mathbf{t} = t)\}$ — вероятность одновременного осуществления этих событий. — *Прим. перев.*

² Здесь используется символическое обозначение $t \in M$, которое следует читать так: « t принадлежит множеству M ». Случайное событие $t \in M$ наступает тогда и только тогда, когда случайная величина \mathbf{t} принимает значение, принадлежащее множеству M . — *Прим. перев.*

от функции \mathbf{u} , определенной на множестве M' , по мере $\mathbf{P}(A)$ (см. § 3 А).

Точно так же левую часть (5) можно понимать как интеграл Лебега по множеству M . Если функцию распределения случайной величины t обозначить $H(t)$, то (5) можно будет записать в виде равенства

$$\int_M \mathcal{E}(\mathbf{u} | t) dH(t) = \int_{M'} \mathbf{u} \mathbf{P}(dE). \quad (6)$$

До сих пор в качестве функции распределения случайной величины t мы рассматривали ступенчатую функцию со строго положительными скачками. Если t имеет непрерывную функцию распределения, то определения (1) и (2) оказываются неприменимыми, так как знаменатель в (1) будет равен нулю. Однако формула (6) всегда сохраняет смысл, и поэтому эту формулу, следуя Колмогорову, можно принять за определение условного математического ожидания $\mathcal{E}(\mathbf{u} | t)$. С помощью теоремы Никодима Колмогоров («Основные понятия теории вероятностей», V, § 4) доказал, что если $\mathcal{E}\mathbf{u}$ существует, то всегда найдется такая измеримая функция $f(t) = \mathcal{E}(\mathbf{u} | t)$, для которой справедливо равенство (6) при любом измеримом множестве M на оси t . Хотя функция $f(t) = \mathcal{E}(\mathbf{u} | t)$ равенством (6) определяется не однозначно, однако два решения $f_1(t)$ и $f_2(t)$ уравнения (6) могут отличаться друг от друга лишь на таком множестве точек оси t , которое имеет меру нуль.

Если $\mathbf{x}_1, \dots, \mathbf{x}_n$ имеют совместную плотность вероятности $g(\mathbf{x})$, то равенство (6) можно записать так:

$$\int_M \mathcal{E}(\mathbf{u} | t) dH(t) = \int_{M'} U(\mathbf{x}) g(\mathbf{x}) d\mathbf{x}. \quad (7)$$

Если t также имеет плотность вероятности $h(t)$, то интеграл Стильтьеса в левой части (7) можно заменить обычным интегралом:

$$\int_M \mathcal{E}(\mathbf{u} | t) h(t) dt = \int_{M'} U(\mathbf{x}) g(\mathbf{x}) d\mathbf{x}. \quad (8)$$

Равенство (8) справедливо для любых измеримых множеств M тогда и только тогда, когда оно справедливо для всех интервалов от $-\infty$ до b на оси t :

$$\int_{-\infty}^b \mathcal{E}(\mathbf{u} | t) h(t) dt = \int_{T(\mathbf{x}) < b} U(\mathbf{x}) g(\mathbf{x}) d\mathbf{x}. \quad (9)$$

Дифференцированием (9) по верхнему пределу b можно найти значения функции $f(t) = \mathcal{E}(\mathbf{u} | t)$ в тех точках t , где $f(t)$ непрерывна и $h(t) \neq 0$.

Покажем теперь, как в некоторых простейших случаях можно вычислить условное математическое ожидание.

Пусть, сначала, $t = x_1$. Плотность вероятности $h(t)$ можно найти интегрированием совместной плотности $g(t, x_2, \dots, x_n)$ по x_2, \dots, x_n :

$$h(t) = \int g(t, x_2, \dots, x_n) dx_2 \dots dx_n. \quad (10)$$

Если положим

$$\mathcal{E}(u | t) = \frac{\int U(t, x_2, \dots, x_n) g(t, x_2, \dots, x_n) dx_2 \dots dx_n}{\int g(t, x_2, \dots, x_n) dx_2 \dots dx_n}, \quad (11)$$

где интегрирование распространяется на все пространство переменных x_2, \dots, x_n , то сразу убеждаемся, что функция (11) удовлетворяет условию (9).

Далее, пусть $t = \sqrt{x_1^2 + \dots + x_n^2}$. Преобразованием к полярным координатам $r, \varphi_1, \dots, \varphi_{n-1}$ этот случай можно свести к предыдущему. Получаем

$$\mathcal{E}(u | r) = \frac{\int U(x) g(x) d\omega}{\int g(x) d\omega}, \quad (12)$$

где $d\omega$ — элемент поверхности шара единичного радиуса в n -мерном пространстве и интегрирование в числителе и знаменателе производится по поверхности сферы радиуса r в том же пространстве.

Условное математическое ожидание, коль скоро оно определено для всех точек оси t (с точностью до множества точек t меры нуль), обладает следующими свойствами:

1. $\mathcal{E}(u - v | t) = \mathcal{E}(u | t) - \mathcal{E}(v | t)$.
2. Если u равна постоянной c , то $\mathcal{E}(u | t) = c$.
3. Если $\mathcal{E}(u | t)$ равно нулю для всех t , то $\mathcal{E}u = 0$.
4. Если $v = \varphi(t)$, то $\mathcal{E}(uv | t) = \mathcal{E}(u | t) \varphi(t)$.

Первые три свойства непосредственно следуют из определения. Последнее свойство доказано Колмогоровым («Основные понятия...», стр. 50).

§ 41. Достаточные статистики

Вернемся к задаче отыскания наилучшей оценки неизвестного параметра ϑ . Мы снова будем предполагать, что совместная плотность распределения результатов наблюдений x_1, \dots, x_n имеет вид

$$g(x | \vartheta) = e(t | \vartheta) h(x), \quad (1)$$

где t является функцией x , не зависящей от ϑ :

$$t = T(x). \quad (2)$$

В прежних обозначениях случайная величина $t = T(x)$ могла оказаться достаточной оценкой для ϑ . Но так как t совсем не обязана быть оценкой для ϑ , то мы t назовем *достаточной статистикой*. Мы будем также говорить: статистика $t = T(x)$ *достаточна для ϑ* .

Условное математическое ожидание $\mathcal{E}(u | t)$ случайной величины $u = U(x)$, как и в § 40, определяется равенством

$$\int_{M'} U(x) g(x | \vartheta) dx = \int_M \mathcal{E}(u | t) dH(t), \quad (3)$$

где $H(t)$ — функция распределения случайной величины t . Докажем теперь теорему:

Если плотность вероятности $g(x | \vartheta)$ представима в виде (1), то функцию $\mathcal{E}(u | t)$ можно определить так, чтобы она не зависела от ϑ .

Сначала мы проведем доказательство в предположении, что существует хотя бы одно значение параметра ϑ_0 , такое, что $e(t | \vartheta_0) \neq 0$ для всех t . В этом случае, согласно (1), для произвольного ϑ имеем

$$g(x | \vartheta) = \frac{e(t | \vartheta)}{e(t | \vartheta_0)} e(t | \vartheta_0) h(x) = \frac{e(t | \vartheta)}{e(t | \vartheta_0)} g(x | \vartheta_0)$$

или, если дробь в правой части обозначить $Q(t)$,

$$g(x | \vartheta) = Q(t) g(x | \vartheta_0). \quad (4)$$

Пусть \mathcal{E} и \mathcal{E}_0 — символы математических ожиданий, а $H(t)$ и $H_0(t)$ — функции распределения, соответствующие значениям параметра ϑ и ϑ_0 . Согласно (3), имеем

$$\int_{M'} U(x) g(x | \vartheta_0) dx = \int_M \mathcal{E}_0(u | t) dH_0(t) \quad (5)$$

и

$$\int_{M'} U(x) g(x | \vartheta) dx = \int_M \mathcal{E}(u | t) dH(t) \quad (6)$$

или, в силу (4),

$$\int_{M'} U(x) Q(t) g(x | \vartheta_0) dx = \int_M \mathcal{E}(u | t) dH(t). \quad (7)$$

Случайная величина $Q(t)$ принимает значения $Q(t)$. Обозначим эту случайную величину буквой v :

$$v = Q(t) = Q[T(x)] = V(x).$$

Рассмотрим условное математическое ожидание произведения $\mathbf{uv} = U(\mathbf{x}) V(\mathbf{x})$ и воспользуемся свойством 4 (§ 40):

$$\mathcal{E}_0(\mathbf{uv} | t) = \mathcal{E}_0(\mathbf{u} | t) Q(t),$$

следовательно, по определению $\mathcal{E}_0(\mathbf{uv} | t)$,

$$\int_{M'} U(\mathbf{x}) V(\mathbf{x}) g(\mathbf{x} | \vartheta_0) d\mathbf{x} = \int_M \mathcal{E}_0(\mathbf{u} | t) Q(t) dH_0(t). \quad (8)$$

Так как $Q(t) = V(\mathbf{x})$, то левые части (7) и (8) совпадают, и поэтому

$$\int_M \mathcal{E}(\mathbf{u} | t) dH(t) = \int_M \mathcal{E}_0(\mathbf{u} | t) Q(t) dH_0(t). \quad (9)$$

В частности, при $\mathbf{u} = 1$, согласно (9), получаем

$$\int_M dH(t) = \int_M Q(t) dH_0(t), \quad (10)$$

где M — произвольное измеримое множество.

Из (10) следует, что для каждой кусочно постоянной функции $f(t)$ имеет место равенство

$$\int_M f(t) dH(t) = \int_M f(t) Q(t) dH_0(t). \quad (11)$$

Для доказательства (11) достаточно M разложить на такие подмножества, где функция $f(t)$ постоянна, и к каждому подмножеству применить формулу (10).

Каждую измеримую функцию можно так аппроксимировать кусочно постоянной функцией, чтобы их интегралы сколь угодно мало отличались друг от друга. Таким образом, формула (11) справедлива для всех тех измеримых функций, для которых левая часть (11) вообще имеет смысл. Если положим $f(t) = \mathcal{E}_0(\mathbf{u} | t)$, то получим

$$\int_M \mathcal{E}_0(\mathbf{u} | t) dH(t) = \int_M \mathcal{E}_0(\mathbf{u} | t) Q(t) dH_0(t). \quad (12)$$

Сравнивая (9) и (12), находим, что

$$\int_M \mathcal{E}(\mathbf{u} | t) dH(t) = \int_M \mathcal{E}_0(\mathbf{u} | t) dH(t). \quad (13)$$

Следовательно, в правой части (7) $\mathcal{E}(\mathbf{u} | t)$ можно заменить на $\mathcal{E}_0(\mathbf{u} | t)$ — равенство от этого не нарушится, т. е. при любом ϑ можно считать, что условное математическое ожидание $\mathcal{E}(\mathbf{u} | t)$ равно $\mathcal{E}_0(\mathbf{u} | t)$, что и требовалось доказать.

Если $e(t | \vartheta)$ обращается в нуль при некоторых значениях t , зависящих от ϑ , то доказательство этой теоремы становится несколько более трудным.

Мы предположим, что при всех ϑ функция $e(t | \vartheta)$ кусочно непрерывна по переменной t ; этого достаточно для всех приложений. Кроме того, предположим, что в точках разрыва $e(t | \vartheta) = 0$. Тогда, при каждом ϑ , множество тех точек оси t , где $e(t | \vartheta) \neq 0$, является открытым множеством.

Рассмотрим множество B_0 тех точек t , в которых $e(t | \vartheta) = 0$ при всех ϑ . В силу (1), при каждом ϑ вероятность события $t \in B_0$ равна нулю. На множестве B_0 условное математическое ожидание можно определить, например, равенством $\xi(\mathbf{u} | t) = 0$; это не повлияет на результат. Остается исследовать условное математическое ожидание на множестве C , которое представляет собой дополнение к множеству B_0 .

Для каждой точки t , принадлежащей множеству C , найдется хотя бы одно значение ϑ , такое, что $e(t | \vartheta) \neq 0$. Тогда существует также некоторая окрестность $B(t)$ точки t и во всех точках этой окрестности $e(t | \vartheta) \neq 0$. Таким образом, множество C покрывается открытыми множествами $B(t)$, поэтому из этих множеств можно выделить счетное покрытие¹ B_1, B_2, \dots множества C . Пусть, например, множеству B_1 соответствует значение параметра ϑ_1 , множеству B_2 соответствует ϑ_2, \dots и т. д., причем если $t \in B_i$, то $e(t | \vartheta_i) \neq 0$.

Из множества B_2 можно исключить те точки, которые уже принадлежат B_1 , точно так же из B_3 можно исключить все точки, которые принадлежат B_1 или B_2 и т. д. Видоизмененные таким способом множества B_1, B_2, \dots по-прежнему покрывают все множество C .

На основании ранее доказанных результатов при любом ϑ функцию $\xi(\mathbf{u} | t)$ можно видоизменить так, чтобы на B_1 она совпала с функцией $\xi_1(\mathbf{u} | t)$, соответствующей значению параметра $\vartheta = \vartheta_1$. Точно так же при любом ϑ можно $\xi(\mathbf{u} | t)$ видоизменить так, чтобы эта функция для $t \in B_2$ совпадала с $\xi_2(\mathbf{u} | t)$ и т. д. Таким образом, нам удалось определить функцию $\xi(\mathbf{u} | t)$ при любом t , независимо от ϑ . Для всех ϑ и M эта функция удовлетворяет условию (3). Действительно, каждое множество M можно разложить на счетное число подмножеств M_0, M_1, M_2, \dots , каждое из которых содержится в B_0, B_1, B_2, \dots соответственно, и так как (3) справедливо для каждой части, то (3) справедливо и для всего множества M .

Этим наше утверждение полностью доказано.

¹ См. Колмогоров А. Н. и Фомин С. В., Элементы теории функций и функционального анализа, Изд. Мос. ун-та (1954), 39. — Прим. перев.

§ 42. Применение теории условных математических ожиданий к задаче отыскания наилучших несмещенных оценок

А. УЛУЧШЕНИЕ ОЦЕНОК

Пусть снова x_1, \dots, x_n — результаты наблюдений, являющиеся случайными величинами с совместной плотностью вероятности $g(x | \vartheta)$, зависящей от неизвестного параметра ϑ . И пусть $t = T(x)$ — достаточная статистика, т. е.

$$g(x | \vartheta) = e(t | \vartheta) h(x). \quad (1)$$

Кроме того, пусть $u = U(x)$ — некоторая оценка параметра ϑ , обладающая конечным математическим ожиданием \hat{u} и конечной дисперсией σ_u^2 . При этом совершенно безразлично, выполняются ли эти предположения лишь для истинного значения параметра ϑ или они выполняются также и в некоторой окрестности истинного значения. Все следующие далее утверждения справедливы для тех значений ϑ , при которых математическое ожидание и дисперсия случайной величины u конечны.

Постараемся теперь найти *улучшенную* оценку v , зависящую только от достаточной статистики t :

$$v = F(t). \quad (2)$$

С этой целью потребуем, чтобы v принимала такие значения v , которые при каждом фиксированном значении t случайной величины t равны условному математическому ожиданию $\mathcal{E}(u | t)$:

$$v = F(t) = \mathcal{E}(u | t). \quad (3)$$

Колмогоров доказал¹, что $v = F(t)$ является случайной величиной. Согласно результатам § 41, функция $F(t) = \mathcal{E}(u | t)$ не зависит от ϑ , а зависит лишь от t .

Докажем теперь, что *математическое ожидание v равно математическому ожиданию u , а дисперсия v не превосходит дисперсии u .*

Доказательство целиком основано на свойствах 1—4 (§ 40). Из 2 и 4 прежде всего следует

$$\mathcal{E}(v | t) = \mathcal{E}(1 \cdot v | t) = \mathcal{E}(1 | t) F(t) = F(t). \quad (4)$$

Далее, из 1 получаем

$$\mathcal{E}(u - v | t) = \mathcal{E}(u | t) - \mathcal{E}(v | t) = F(t) - F(t) = 0. \quad (5)$$

Так как, согласно 3,

$$\mathcal{E}(u - v) = 0, \quad (6)$$

то $\mathcal{E}u = \mathcal{E}v$. Первое утверждение доказано.

¹ Колмогоров А. Н., Несмещенные оценки, Изв. АН СССР (сер. мат.), 14 (1950), 303. — *Прим. перев.*

Дисперсия u равна

$$\begin{aligned}\sigma_u^2 &= \mathcal{E}(u - \hat{u})^2 = \mathcal{E}(u - \hat{v})^2 = \mathcal{E}(u - v + v - \hat{v})^2 = \\ &= \mathcal{E}(u - v)^2 + 2\mathcal{E}(u - v)(v - \hat{v}) + \mathcal{E}(v - \hat{v})^2.\end{aligned}\quad (7)$$

Разность $v - \hat{v}$ зависит лишь от t . Если мы положим $v - \hat{v} = \varphi(t)$, то, согласно 4 и (6), получим

$$\mathcal{E}[(u - v)(v - \hat{v}) | t] = \mathcal{E}(u - v | t) \varphi(t) = 0, \quad (8)$$

следовательно, в силу 3.

$$\mathcal{E}(u - v)(v - \hat{v}) = 0. \quad (9)$$

Все это позволяет записать (7) более просто:

$$\sigma_u^2 = \mathcal{E}(u - v)^2 + \sigma_v^2. \quad (10)$$

Отсюда непосредственно следует, что

$$\sigma_u^2 \geq \sigma_v^2. \quad (11)$$

Второе утверждение доказано.

Если бы дисперсия σ_v^2 была бесконечной, то, согласно (10), дисперсия σ_u^2 также была бы бесконечной, что противоречит первоначальным предположениям. Таким образом, σ_v^2 конечна и не превосходит σ_u^2 .

Неравенство (11) обращается в равенство тогда и только тогда, когда разность $u - v$ принимает ненулевые значения лишь с вероятностью, равной нулю.

Предположение конечности дисперсии σ_u^2 может быть опущено. Действительно, если σ_u^2 бесконечна, то неравенство (11) становится тривиальным.

Таким образом, если существует достаточная статистика $t = T(x)$, то для каждой оценки $u = U(x)$ параметра ϑ существует улучшенная оценка $v = V(x)$, обладающая свойствами:

1. Смещение v равно смещению u , следовательно, если u — несмещенная оценка, то v — также несмещенная оценка.

2. Дисперсия v не превосходит дисперсии u .

3. Оценка v зависит лишь от достаточной статистики $t = T(x)$.

Теперь мы снова можем отказаться от жирного шрифта и обозначить результаты наблюдений и их значения через x_1, \dots, x_n , достаточную статистику — через $T = T(x)$ и оценки — через $U(x)$ и $V(x) = F(T)$.

Б. ИНТЕГРАЛЬНОЕ УРАВНЕНИЕ ДЛЯ НЕСМЕЩЕННЫХ ОЦЕНОК

На основании только что полученных результатов, при отыскании несмещенных наилучших оценок мы всегда можем ограничиться такими оценками $V = F(T)$, которые зависят только

от достаточной статистики. Предположим теперь, что T имеет плотность вероятности $q(t | \vartheta)$.

Несколько обобщая задачу, мы будем искать оценку не для самого параметра ϑ , а для некоторой его функции $\varphi(\vartheta)$. Условие, согласно которому искомая оценка должна быть несмещенной, непосредственно приводит к интегральному уравнению

$$\int q(t | \vartheta) F(t) dt = \varphi(\vartheta), \quad (12)$$

где интегрирование производится по всей области возможных значений t оценки T . Если F и F_1 — два решения уравнения (12), то их разность $D(t)$ должна удовлетворять интегральному уравнению

$$\int q(t | \vartheta) D(t) dt = 0. \quad (13)$$

Может оказаться, что однопараметрическое семейство $\{q(t|\vartheta)\}$ образует на оси t полную систему функций, так что никакая отличная от тождественного нуля функция $D(t)$ не может быть ортогональной ко всем функциям $q(t|\vartheta)$. В этом случае из (2) следует, что

$$D(t) = 0,$$

т. е. если решение (12) существует, то оно определяется однозначно. Этим самым доказана

Основная теорема. Если $T = T(x)$ — достаточная статистика для ϑ и семейство $\{q(t|\vartheta)\}$ образует полную систему функций, то любая несмещенная оценка для $\varphi(\vartheta)$, зависящая лишь от T , является наилучшей оценкой.

§ 43. Приложения

Метод отыскания несмещенных, наилучших оценок, изложенный в § 42, имеет много применений. Прежде всего заметим, что этим методом можно было бы воспользоваться во всех существующих примерах. Теперь мы укажем несколько новых примеров, из которых первые два заимствованы из книги С. Р. Рао, *Advanced Statistical Methods in Biometric Research*, New York, 1952.

Пример 29. Распределение χ^2 с множителем α .

Пусть имеется n независимых наблюдений x_1, \dots, x_n , каждое из которых подчиняется распределению χ^2 с неизвестным множителем α в показателе степени. Таким образом, плотность вероятности отдельного наблюдения задается формулой

$$f(x|\alpha) = c \alpha^p x^{p-1} e^{-\alpha x} \quad (x > 0), \quad (1)$$

где $c = 1/\Gamma(p)$. Если положим $\sum x = T(x) = T$, то плотность совместного распределения всех x_1, \dots, x_n будет иметь вид

$$g(x|\alpha) = c^n \alpha^{np} (x_1 \dots x_n)^{p-1} e^{-\alpha T}. \quad (2)$$

от достаточной статистики. Предположим теперь, что T имеет плотность вероятности $q(t | \vartheta)$.

Несколько обобщая задачу, мы будем искать оценку не для самого параметра ϑ , а для некоторой его функции $\varphi(\vartheta)$. Условие, согласно которому искомая оценка должна быть несмещенной, непосредственно приводит к интегральному уравнению

$$\int q(t | \vartheta) F(t) dt = \varphi(\vartheta), \quad (12)$$

где интегрирование производится по всей области возможных значений t оценки T . Если F и F_1 — два решения уравнения (12), то их разность $D(t)$ должна удовлетворять интегральному уравнению

$$\int q(t | \vartheta) D(t) dt = 0. \quad (13)$$

Может оказаться, что однопараметрическое семейство $\{q(t|\vartheta)\}$ образует на оси t полную систему функций, так что никакая отличная от тождественного нуля функция $D(t)$ не может быть ортогональной ко всем функциям $q(t | \vartheta)$. В этом случае из (2) следует, что

$$D(t) = 0,$$

т. е. если решение (12) существует, то оно определяется однозначно. Этим самым доказана

Основная теорема. Если $T = T(x)$ — достаточная статистика для ϑ и семейство $\{q(t | \vartheta)\}$ образует полную систему функций, то любая несмещенная оценка для $\varphi(\vartheta)$, зависящая лишь от T , является наилучшей оценкой.

§ 43. Приложения

Метод отыскания несмещенных, наилучших оценок, изложенный в § 42, имеет много применений. Прежде всего заметим, что этим методом можно было бы воспользоваться во всех предшествующих примерах. Теперь мы укажем несколько новых примеров, из которых первые два заимствованы из книги С. R. Rao, *Advanced Statistical Methods in Biometric Research*, New York, 1952.

Пример 29. Распределение χ^2 с множителем α .

Пусть имеется n независимых наблюдений x_1, \dots, x_n , каждое из которых подчиняется распределению χ^2 с неизвестным множителем α в показателе степени. Таким образом, плотность вероятности отдельного наблюдения задается формулой

$$f(x|\alpha) = c \alpha^p x^{p-1} e^{-\alpha x} \quad (x > 0), \quad (1)$$

где $c = 1/\Gamma(p)$. Если положим $\sum x := T(x) = T$, то плотность совместного распределения всех x_1, \dots, x_n будет иметь вид

$$g(x|\alpha) := c^n \alpha^{np} (x_1 \dots x_n)^{p-1} e^{-\alpha T}. \quad (2)$$

Формула (2) показывает, что T — достаточная статистика. Если от x_1, \dots, x_n перейдем к новым переменным T, y_1, \dots, y_n по формулам

$$x_i = Ty_i, \quad (3)$$

то все y_i будут связаны соотношением

$$\sum y_i = 1, \quad (4)$$

поэтому независимыми переменными можно считать лишь T и y_1, \dots, y_{n-1} . Если функцию (2) проинтегрировать по переменным y_1, \dots, y_{n-1} в области

$$y_1 > 0, \dots, y_n > 0, \quad \sum y_i = 1, \quad (5)$$

то получим плотность вероятности случайной величины T :

$$q(T|\alpha) = c' \alpha^{np} T^{np-1} e^{-\alpha T}. \quad (6)$$

Так как интеграл от $q(T, \alpha)$ в пределах от 0 до ∞ должен быть равен единице, то

$$c' = \frac{1}{\Gamma(np)}.$$

Среднее значение случайной величины $1/T$ равно

$$c' \alpha^{np} \int_0^{\infty} T^{np-2} e^{-\alpha T} dT = \frac{\alpha \Gamma(np-1)}{\Gamma(np)} = \frac{\alpha}{np-1}.$$

Следовательно,

$$F(T) = \frac{np-1}{T} \quad (7)$$

является несмещенной оценкой для α .

Если бы имелась другая несмещенная оценка, также зависящая только от T , то должно было бы существовать ненулевое решение $D(t)$ интегрального уравнения

$$\int_0^{\infty} D(t) t^{np-1} e^{-\alpha t} dt = 0. \quad (8)$$

Если введем новую переменную интегрирования $z = e^{-t}$ и положим $D(t) t^{np-1} = G(z)$, то из (8) будет следовать, что

$$\int_0^1 z^{\alpha-1} G(z) dz = 0 \quad \text{для } \alpha = 1, 2, 3, \dots \quad (9)$$

Но в интервале от 0 до 1 функции $1, z, z^2, \dots$ образуют полную систему функций¹. Таким образом, из (9) следует, что $G(z) = 0$, т. е. это интегральное уравнение имеет лишь нулевое решение. Следовательно, оценка (7) является наилучшей.

Оценка наибольшего правдоподобия

$$\tilde{\alpha} = \frac{np}{T}$$

¹ См., например, Курант Р. и Гильберт Д., Методы математической физики, т. 1, гл. 2, § 4, Гостехиздат, М., 1951.

Формула (11) показывает, что T является достаточной статистикой для ϑ . Согласно (11), математическое ожидание T равно

$$E T = n \vartheta^{-n} \int_0^{\vartheta} t^n dt = \frac{n}{n+1} \vartheta. \quad (12)$$

Поэтому

$$F = \frac{n+1}{n} T \quad (13)$$

является несмещенной оценкой для ϑ . Дисперсия этой оценки равна

$$\sigma^2 = \frac{\vartheta^2}{n(n+2)}.$$

Если бы существовала другая несмещенная и зависящая лишь от T оценка параметра ϑ , то существовало бы ненулевое решение $D(t)$ интегрального уравнения

$$\int_0^{\vartheta} D(t) n \vartheta^{-n} t^{n-1} dt = 0 \quad (14)$$

или

$$\int_0^{\vartheta} D(t) t^{n-1} dt = 0. \quad (15)$$

Но если (15) справедливо для всех ϑ , то должно выполняться равенство $D(t) = 0$. Следовательно, (13) является наилучшей оценкой.

Оценка наибольшего правдоподобия¹ $\tilde{\vartheta} = T$ при больших n незначительно отличается от F .

Следующий пример был мне любезно предоставлен Е. Л. Леманном. Этот пример особенно интересен тем, что в нем непосредственно используется метод улучшения несмещенной оценки, изложенный в § 42 А.

Пример 31. Поставщик поставляет некоторую продукцию, упакованную в ящики. Потребитель из каждого ящика берет случайную выборку, состоящую из n изделий, и проверяет качество каждого изделия. Если в выборке окажется три или больше дефектных изделий, то соответствующий ящик бракуется и возвращается поставщику, который в этом случае несет определенную потерю. Количества дефектных изделий, обнаруженных в каждом из ящиков, обязательно сообщаются поставщику. Так как все изделия изготавливаются в одинаковых условиях, то предполагается, что любое изделие может оказаться дефектным с постоянной вероятностью p . Пусть r — число ящиков, полученных потребителем, и пусть x_1, \dots, x_r — обнаруженные количества дефектных изделий в выборках объема n из

¹ В данном случае функция правдоподобия дается формулой

$$g(x|\vartheta) = \begin{cases} \vartheta^{-n}, & \text{если } \vartheta \geq T, \\ 0 & \text{— в противном случае.} \end{cases}$$

Для отыскания максимума $g(x|\vartheta)$ достаточно ограничиться лишь теми ϑ , которые удовлетворяют неравенству $\vartheta \geq T$. Так как при $\vartheta \geq T$ справедливо неравенство $\vartheta^{-n} \leq T^{-n}$, то $g(x|\vartheta) \leq T^{-n} = g(x|T)$. Следовательно, T — оценка наибольшего правдоподобия. — *Прим. перев.*

ящичков с номерами $1, \dots, r$ соответственно. Само собой разумеется, что несмещенной, наилучшей оценкой для p в данном случае будет

$$h = \frac{x_1 + \dots + x_r}{rn}.$$

Но поставщик желает знать также и несмещенную оценку для математического ожидания своих потерь. Ведь улучшение качества продукции (например, посредством более строгой предварительной инспекции) сопряжено с затратой средств, и на эту затрату поставщик пойдет лишь тогда, когда убедится в ее выгодности.

Вероятность того, что данный ящик не будет забракован потребителем, равна

$$\vartheta = q^n + n p q^{n-1} + \binom{n}{2} p^2 q^{n-2}.$$

Математическое ожидание потерь поставщика пропорционально $1 - \vartheta$. Следовательно, для отыскания несмещенной оценки потерь достаточно найти несмещенную оценку для ϑ .

Предположим, что x_1, \dots, x_r являются независимыми случайными величинами, подчиняющимися биномиальному распределению. Функция правдоподобия в данном случае задается формулой

$$P(x) = \binom{n}{x_1} \dots \binom{n}{x_r} p^{x_1} q^{n-x_1} \dots p^{x_r} q^{n-x_r}.$$

Если положим $x_1 + \dots + x_r = T$, то $P(x)$ будет иметь вид

$$P(x) = \binom{n}{x_1} \dots \binom{n}{x_r} p^T q^{nr-T}.$$

Из этой формулы следует, что T является достаточной статистикой для p , а значит, и для ϑ . Таким образом, если несмещенная, наилучшая оценка для ϑ существует, то можно считать, что она зависит лишь от T . Как найти такую функцию?

Рассмотрим случайную величину

$$U = \begin{cases} 1, & \text{если первый ящик принят,} \\ 0, & \text{если первый ящик забракован.} \end{cases}$$

Так как $E\{U\} = \vartheta$, то U — несмещенная оценка для ϑ , хотя и очень плохая. Вычислим теперь условное математическое ожидание U при условии, что $T = t$. Первый ящик может быть принят лишь в следующих трех взаимно исключающих друг друга случаях: когда в выборке нет дефектных изделий, когда в выборке одно дефектное изделие и, наконец, когда в выборке два дефектных изделия. Следовательно, условное математическое ожидание U будет равно сумме условных вероятностей указанных выше трех событий, в предположении, что $T = t$. Таким образом,

$$\begin{aligned} E(U|t) &= \frac{P\{(x_1 = 0) \cap (x_2 + \dots + x_r = t)\}}{P\{x_1 + \dots + x_r = t\}} + \\ &+ \frac{P\{(x_1 = 1) \cap (x_2 + \dots + x_r = t - 1)\}}{P\{x_1 + \dots + x_r = t\}} + \\ &+ \frac{P\{(x_1 = 2) \cap (x_2 + \dots + x_r = t - 2)\}}{P\{x_1 + \dots + x_r = t\}}. \end{aligned}$$

Все вероятности в числителях и в знаменателе содержат одинаковый множитель $p^t q^{nr-t}$, сокращая который, получим

$$\xi(U|t) = \frac{\binom{rn-n}{t} + \binom{n}{1} \binom{rn-n}{t-1} + \binom{n}{2} \binom{rn-n}{t-2}}{\binom{rn}{t}}.$$

Следовательно, улучшенная оценка для ϑ имеет вид

$$V = \binom{rn}{T}^{-1} \left[\binom{rn-n}{T} + \binom{n}{1} \binom{rn-n}{T-1} + \binom{n}{2} \binom{rn-n}{T-2} \right].$$

Для того чтобы показать, что V является наилучшей несмещенной оценкой, нам нужно еще убедиться в единственности несмещенной оценки для ϑ , зависящей лишь от T . Иначе говоря, мы должны проверить, будет ли уравнение

$$\sum_{t=0}^{nr} \binom{nr}{t} p^t q^{nr-t} F(t) = \vartheta$$

иметь единственное решение $F(t) = V(t)$. Если F и F_1 — два решения этого уравнения, то их разность $D(t)$ должна удовлетворять однородному уравнению

$$\sum \binom{nr}{t} p^t q^{nr-t} D(t) = 0.$$

Но многочлен от p на отрезке $0 \leq p \leq 1$ обращается тождественно в нуль тогда и только тогда, когда все его коэффициенты равны нулю. Следовательно, $D = 0$ — единственное решение однородного уравнения¹.

¹ В данном примере рассматривалась задача оценки потерь поставщика. Потребителю более важно знать другую величину, а именно, количество дефектных изделий, оставшихся в принятом ящике после контроля. Так как вероятность того, что выбранное наугад изделие окажется дефектным, равна p , то принятые дефектные изделия снова подчиняются биномиальному распределению. Следовательно, если контроль изделий сопряжен с их уничтожением, то изложенная процедура приемочного контроля не ведет к улучшению качества принятой продукции. Такая процедура целесообразна лишь в том случае, если для различных ящиков вероятности p различны и потребитель хочет гарантировать себя от приема ящиков с $p > p_0$, где p_0 — некоторый заранее установленный предел для доли брака.

О статистическом контроле качества продукции см. Д у н и н - Б а р к о в с к и й И. В. и С м и р н о в Н. В., Теория вероятностей и математическая статистика в технике (общая часть), ГИТТЛ, М., 1955, гл. VIII, § 4—5; К о л м о г о р о в А. Н., Несмещенные оценки, Изв. АН СССР, серия математич., 14 (1950), 303; К о л м о г о р о в А. Н., Статистический приемочный контроль при допустимом числе дефектных изделий, равно нулю, Изд. Ленингр. Общества научно-техн. пропаганды (1951); С и р а ж д и н о в С. Х., Одинарный статистический приемочный контроль, Труды ин-та Математики и механики, АН Узб. ССР, 15 (1955), 41. — Прим. перев.

§ 44. Оценка дисперсии нормального распределения

Рао, а также Леманн и Шеффе свою теорию наилучших оценок распространили на случай нескольких неизвестных параметров. Мы здесь не будем касаться этой общей теории, а ограничимся лишь одним примером, который особенно важен для приложений.

Пусть x_1, \dots, x_n — независимые и одинаково нормально распределённые случайные величины с неизвестным средним значением μ и неизвестной дисперсией ϑ . Следовательно, совместная плотность вероятности имеет вид

$$\begin{aligned} g(x_1, \dots, x_n | \mu, \vartheta) &= c \vartheta^{-\frac{n}{2}} e^{-\frac{1}{2\vartheta} \sum (x - \mu)^2} = \\ &= c \vartheta^{-\frac{n}{2}} e^{-\frac{1}{2\vartheta} (\sum x^2 - 2\mu \sum x + n\mu^2)}. \end{aligned} \quad (1)$$

Из формулы (1) непосредственно следует, что $\sum x$ и $\sum x^2$ — достаточные статистики для μ и ϑ . Как мы уже знаем, наилучшей оценкой для μ является

$$\bar{x} = \frac{1}{n} \sum x. \quad (2)$$

Постараемся найти наилучшую оценку для ϑ . При всех μ и ϑ эта оценка должна быть несмещенной.

Сначала ортогональным преобразованием введем новые координаты y_1, \dots, y_n , где

$$y_1 = \bar{x} \sqrt{n} = \frac{1}{\sqrt{n}} (x_1 + \dots + x_n). \quad (3)$$

В силу ортогональности преобразования,

$$\sum x^2 = \sum y^2,$$

поэтому

$$\begin{aligned} \sum x^2 - 2\mu \sum x + n\mu^2 &= \sum y^2 - 2\mu y_1 \sqrt{n} + n\mu^2 = \\ &= (y_1 - \mu \sqrt{n})^2 + y_2^2 + \dots + y_n^2. \end{aligned}$$

Таким образом, совместная плотность вероятности для y_1, \dots, y_n имеет вид

$$f(y_1, \dots, y_n | \mu, \vartheta) = c \vartheta^{-\frac{n}{2}} e^{-\frac{1}{2\vartheta} [(y_1 - \mu \sqrt{n})^2 + y_2^2 + \dots + y_n^2]}. \quad (4)$$

В пространстве переменных y_2, \dots, y_n введем полярные координаты $r, \varphi_1, \dots, \varphi_{n-2}$. Соответствующая плотность вероятности будет иметь вид

$$\begin{aligned} p(y_1, r, \varphi_1, \dots, \varphi_{n-2} | \mu, \vartheta) &= \\ &= c \vartheta^{-\frac{n}{2}} e^{-\frac{1}{2\vartheta} [(y_1 - \mu \sqrt{n})^2 + r^2]} \frac{r^{n-2}}{r} h(\varphi_1, \dots, \varphi_{n-2}). \end{aligned} \quad (5)$$

Вместо $\sum x^2$ и $\sum x$ в формуле (5) имеются величины r и y_1 , которые, конечно, также являются достаточными статистиками для μ и ϑ .

Как и в § 40, мы можем теперь для любой случайной величины u определить условное математическое ожидание $\mathcal{E}(u \mid y_1, r)$. Общая теория Колмогорова для этого даже и не потребуется: условное математическое ожидание можно определить посредством интегрирования по угловым координатам $\varphi_1, \dots, \varphi_{n-2}$.

С помощью этого условного математического ожидания можно, как в § 42 А, для любой оценки u параметра ϑ вывести улучшенную оценку v , которая будет иметь то же смещение, что и u , а дисперсия v не будет превосходить дисперсии u . Так как, кроме того, v должна зависеть лишь от y_1 и r , то оценку для ϑ мы сразу будем искать в виде функции от этих аргументов.

Такой функцией является

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1} = \frac{\sum x^2 - n\bar{x}^2}{n-1} = \frac{\sum y^2 - y_1^2}{n-1} = \frac{r^2}{n-1}. \quad (6)$$

Мы знаем, что s^2 является несмещенной оценкой для $\sigma^2 = \vartheta$. Если бы имелась какая-либо другая несмещенная оценка, зависящая лишь от y_1 и r , то интегральное уравнение

$$\iint D(y, r) e^{-\frac{1}{2\vartheta} [(y - \mu\sqrt{n})^2 + r^2]} r^{n-2} dr dy = 0 \quad (7)$$

имело бы ненулевое решение.

Полагая в (7)

$$\int_0^\infty D(y, r) e^{-\frac{1}{2\vartheta} r^2} r^{n-2} dr = F(y \mid \vartheta), \quad (8)$$

получим

$$\int_{-\infty}^\infty F(y \mid \vartheta) e^{-\frac{1}{2\vartheta} (y^2 - 2y\mu\sqrt{n} + \mu^2 n)} dy = 0$$

или, если обозначить $\alpha = \mu\sqrt{n}/\vartheta$ и постоянный множитель $\exp\{-\mu^2 n/2\vartheta\}$ вынести за знак интеграла,

$$\int_{-\infty}^\infty F(y \mid \vartheta) e^{-\frac{1}{2\vartheta} y^2} e^{\alpha y} dy = 0. \quad (9)$$

Последнее равенство должно быть справедливо при любых α и $\vartheta > 0$. Левая часть (9) является аналитической функцией α ,

которая определена¹ при всех комплексных α . Если эта функция даже на некотором маленьком отрезке действительной оси равна нулю, то она равна нулю тождественно при всех α . Поэтому заменим в (9) α на it и рассмотрим полученный интеграл, который можно истолковать как преобразование Фурье для функции $F(y | \vartheta) \exp\{-y^2/2\vartheta\}$. В силу (9), это преобразование равно нулю при всех t , следовательно, преобразуемая функция также равна нулю, т. е.

$$F(y | \vartheta) = 0. \quad (10)$$

Если (10) подставим в (8), то получим интегральное уравнение для $D(y, r)$:

$$\int_0^{\infty} D(y, r) e^{-\beta r^2} r^{n-2} dr = 0. \quad (11)$$

Выбрав r^2 в качестве новой переменной интегрирования, убеждаемся, что интегральное уравнение (11) имеет точно такой же вид, как и (8) § 43. Поэтому, как и тогда,

$$D(y, r) = 0.$$

Только что приведенное доказательство справедливо в классе функций $D(y, r)$, удовлетворяющих некоторым слабым условиям регулярности. Например, достаточно предположить, что интегралы (7) и (8) абсолютно сходятся для всех μ и ϑ из конечной области M :

$$\begin{aligned} a < \mu < b, \\ 0 < \vartheta < c, \end{aligned}$$

и что на каждом замкнутом подмножестве множества M интегралы (7) и (8) сходятся равномерно.

Точно тот же метод доказательства можно применить и к случаю, когда имеется несколько групп $x_1, \dots, x_m; y_1, \dots, y_n; \dots$ независимых нормально распределенных случайных величин с одинаковыми дисперсиями ϑ , но, может быть, разными средними значениями: μ — для первой группы, ν — для второй группы и т. д. Результат тот же самый:

$$s^2 = \frac{\sum (x - \bar{x})^2}{(m-1)} + \frac{\sum (y - \bar{y})^2}{(n-1)} + \dots \quad (12)$$

является несмещенной, наилучшей оценкой для ϑ .

¹ Доказательство. Сначала в (9) заменим пределы интегрирования $-\infty$ и $+\infty$ на $-M$ и $+M$ таким образом, чтобы для всех α , принадлежащих произвольному кругу $|\alpha| < R$, соответствующий интеграл отличался от (9) менее чем на ϵ . Затем $e^{i\alpha y}$ разложим в степенной ряд и проинтегрируем его почленно. Таким образом, мы получим разложение интеграла от $-M$ до $+M$ в степенной ряд по степеням α . Теперь M можно устремить к бесконечности: предел равномерно сходящейся последовательности регулярных функций является регулярной функцией в круге $|\alpha| < R$.

Если, в частности, каждая группа состоит лишь из двух случайных величин, то (12) превращается в формулу (10) § 35. Следовательно, пример 23 может служить иллюстрацией практического применения формулы (12).

§ 45. Асимптотические свойства

Все теоремы, с которыми мы до сих пор имели дело, были справедливы как для малых выборок, так и для больших, что особенно важно для приложений. Теперь, в заключение, мы хотим совсем кратко и без доказательств изложить важнейшие асимптотические свойства оценок в случае больших выборок.

А. СОСТОЯТЕЛЬНОСТЬ ОЦЕНОК НАИБОЛЬШЕГО ПРАВДОПОДОБИЯ

Вернемся к случаю одного параметра ϑ . Пусть x_1, \dots, x_n — независимые одинаково распределенные случайные величины с плотностью вероятности $f(x | \vartheta)$. Плотность их совместного распределения задается формулой

$$g(x | \vartheta) = f(x_1 | \vartheta) f(x_2 | \vartheta) \dots f(x_n | \vartheta). \quad (1)$$

Оценка T параметра ϑ называется *состоятельной*, если при $n \rightarrow \infty$ вероятность события $|T - \vartheta| < \varepsilon$ стремится к единице. При некоторых предположениях о регулярности функции $f(x | \vartheta)$ можно показать, что метод наибольшего правдоподобия приводит к состоятельной оценке параметра ϑ .

Простейшее из известных мне доказательств опирается на довольно слабые предположения регулярности; это доказательство принадлежит Вальду и Вольфовицу. Здесь мы это доказательство воспроизводить не будем, а отшлем читателя к оригинальной работе: A. Wald and J. Wolfowitz, *Annals of Mathematical Statistics*, 20 (1949), 595, 601.

Если при $n \rightarrow \infty$ вместе с увеличением n количество неизвестных параметров также возрастает, то теорема о состоятельности оценок наибольшего правдоподобия может и не быть справедливой. Пример этого нам уже встречался в § 35.

Б. АСИМПТОТИЧЕСКАЯ НОРМАЛЬНОСТЬ, АСИМПТОТИЧЕСКОЕ СРЕДНЕЕ ЗНАЧЕНИЕ И АСИМПТОТИЧЕСКАЯ ДИСПЕРСИЯ

Состоятельность оценки наибольшего правдоподобия $\hat{\vartheta}$ ранее была доказана Хотеллингем¹ и Дубсом² при более сильных ограничениях. Однако эти авторы, помимо состоятельности, доказали, что оценка $\hat{\vartheta}$ распределена асимптотически нормально со средним

¹ Hotelling H., *Trans. Amer. Math. Soc.*, 32 (1930), 847.

² Doob J. L., *Trans. Amer. Math. Soc.*, 36, 766; 39, 410.

значением ϑ и квадратичным отклонением c/\sqrt{n} . Это означает, что функция распределения случайной величины

$$U = (\tilde{\vartheta} - \vartheta) \sqrt{n} \quad (2)$$

при $n \rightarrow \infty$ стремится к функции нормального распределения с нулевым средним значением и квадратичным отклонением c . При этом c определяется равенством

$$\frac{1}{c^2} = \mathcal{E} \left(\frac{\partial \ln f}{\partial \vartheta} \right)^2. \quad (3)$$

В § 36 правую часть последнего равенства мы обозначили через $j(\vartheta)$. Если это выражение умножим на n , то получим введенную ранее «информацию» $I = I(\vartheta)$:

$$I = \frac{n}{c^2} = \mathcal{E} \left(\frac{\partial \ln g}{\partial \vartheta} \right)^2. \quad (4)$$

Следовательно, дисперсия асимптотического нормального распределения равна обратной величине информации:

$$\frac{c^2}{n} = \frac{1}{I}. \quad (5)$$

При определении понятий «асимптотическое среднее значение» и «асимптотическая дисперсия» нужно соблюдать большую осторожность, так как вполне может случиться, что при каждом n точное распределение оценки $\tilde{\vartheta}$ обладает бесконечной дисперсией. И тем не менее $\tilde{\vartheta}$ будет распределена асимптотически нормально с конечным средним значением ϑ и конечной дисперсией c^2/n . По этой причине нельзя сначала вычислять дисперсию и затем производить предельный переход при $n \rightarrow \infty$, а нужно сперва найти распределение случайной величины U , затем произвести предельный переход при $n \rightarrow \infty$ и, наконец, вычислить дисперсию. В дальнейшем выражения «асимптотическое среднее значение» и «асимптотическая дисперсия» всегда следует понимать именно в этом смысле.

Пусть T — оценка параметра ϑ . Если асимптотическое среднее значение разности $T - \vartheta$ (смысл этого понятия указан выше) мало сравнительно с $1/\sqrt{n}$, то оценка T называется *асимптотически несмещенной*; при этом асимптотическое распределение случайной величины

$$U = (T - \vartheta) \sqrt{n} \quad (6)$$

имеет нулевое среднее значение. Согласно упомянутым выше теоремам Хотеллинга и Дуба, оценка наибольшего правдоподобия $\tilde{\vartheta}$ является асимптотически несмещенной.

В. ЭФФЕКТИВНОСТЬ

Р. А. Фишер предполагал, что оценка \bar{v} асимптотически эффективна в том смысле, что она среди всех асимптотически несмещенных оценок обладает наименьшей асимптотической дисперсией. Однако более поздние исследования¹ показали, что это предположение соответствует действительности лишь тогда, когда множество допустимых оценок ограничено сильными условиями регулярности. Если же с оценкой наибольшего правдоподобия могут конкурировать произвольные асимптотически несмещенные оценки, то среди них можно найти так называемые «сверхэффективные» оценки, которые при некоторых значениях параметра имеют меньшую асимптотическую дисперсию, чем оценка наибольшего правдоподобия. Пример такой «сверхэффективной» оценки был указан Д. Л. Ходжесом. Пусть $f(x | \vartheta)$ — нормальная плотность вероятностей со средним значением ϑ и единичной дисперсией:

$$f(x | \vartheta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\vartheta)^2}. \quad (7)$$

Требуется найти оценку для ϑ по выборке, состоящей из n независимых наблюдений x_1, \dots, x_n . В качестве такой оценки можно принять выборочное среднее значение \bar{x} . Оценка \bar{x} является несмещенной, и ее дисперсия в данном случае равна $1/n$. Рассмотрим теперь другую оценку T :

$$T = \begin{cases} \bar{x}, & \text{если } |\bar{x}| \geq n^{-\frac{1}{4}} \\ \frac{1}{2}\bar{x}, & \text{если } |\bar{x}| < n^{-\frac{1}{4}}, \end{cases}$$

и докажем, что

1. T — асимптотически нормальна при любом ϑ .
2. T — асимптотически несмещенная оценка при любом ϑ .
3. Асимптотическая дисперсия T равна $1/n$, если $\vartheta \neq 0$, и равна $1/4n$, если $\vartheta = 0$.

Доказательство. Если $\vartheta \neq 0$, то при больших n вероятность события $|\bar{x}| < n^{-1/4}$ исчезающе мала, следовательно, практически всегда выполняется равенство $T = \bar{x}$, и поэтому асимптотическое распределение T совпадает с распределением \bar{x} . Напротив, если $\vartheta = 0$, то вероятность события $|\bar{x}| \geq n^{-1/4}$ будет исчезающе малой, и поэтому асимптотическое распределение T должно совпадать с распределением случайной величины $\bar{x}/2$.

Более обстоятельное изложение затронутых здесь вопросов можно найти в цитированной работе Ле Кама.

¹ Эти исследования подытожены в работе Le Cam L., On some asymptotic properties of Maximum Likelihood Estimates, Univ. of Calif. Publ. in Stat., 1, No 11 (1953), 277.

ОЦЕНКА ПАРАМЕТРОВ ПО НАБЛЮДЕННЫМ ЧАСТОТАМ

Постановка задачи в этой главе та же самая, что и в предыдущей, так как речь будет идти об оценке параметра ϑ по результатам наблюдений. При этом будет предполагаться, что все наблюдаемые величины являются частотами x/n , где n — число опытов, а x — число тех случаев, в которых осуществлялось данное событие. Каждая частота h представляет собой случайную величину, подчиняющуюся биномиальному распределению (§ 5). Это биномиальное распределение зависит от двух параметров: от известного объема выборки n и от вероятности p , с которой данное событие может осуществиться в каждом отдельном испытании. Если наблюдаются несколько частот h_i , то соответствующие вероятности p_i или неизвестны, или являются функциями неизвестных параметров ϑ . Задача заключается в отыскании оценок неизвестных параметров, а также в исследовании надежности этих оценок.

При этом будут предполагаться известными важнейшие результаты из гл. VII и VIII.

§ 46. Метод наибольшего правдоподобия

Для большей определенности мы предположим, что было произведено n независимых испытаний, в каждом из которых осуществлялся какой-либо один из трех взаимно исключающих друг друга случайных исходов. Пусть x_1 — число наступлений первого исхода, x_2 — число наступлений второго исхода и x_3 — число наступлений третьего исхода, тогда

$$x_1 + x_2 + x_3 = n.$$

Числам x_i соответствуют частоты $h_i = x_i/n$, удовлетворяющие соотношению

$$h_1 + h_2 + h_3 = 1.$$

Пусть p_1, p_2, p_3 — вероятности осуществления в отдельном испытании каждого из трех указанных исходов. Для этих вероят-

ностей, очевидно, должно иметь место равенство

$$p_1 + p_2 + p_3 = 1.$$

Математическое ожидание h_i равно p_i , поэтому математическое ожидание x_i равно np_i . Вычислим математические ожидания x_i^2 и $x_i x_k$ ($i \neq k$), которые нам понадобятся позднее.

Дисперсия x_i равна $np_i(1 - p_i)$, следовательно,

$$\begin{aligned} \mathcal{E} x_i^2 &= (\mathcal{E} x_i)^2 + np_i(1 - p_i) = n^2 p_i^2 + np_i - np_i^2 = \\ &= n(n - 1)p_i^2 + np_i. \end{aligned} \quad (1)$$

То же самое справедливо и для $x_i + x_k$ ($i \neq k$):

$$\mathcal{E}(x_i + x_k)^2 = n(n - 1)(p_i + p_k)^2 + n(p_i + p_k).$$

Если из обеих частей последнего равенства вычесть $\mathcal{E} x_i^2 + \mathcal{E} x_k^2$ и результаты разделить на 2, то получим

$$\mathcal{E} x_i x_k = n(n - 1)p_i p_k \quad (i \neq k). \quad (2)$$

После этой подготовки мы перейдем к основной задаче. Пусть вероятности p_1, p_2, p_3 являются функциями одного неизвестного параметра ϑ . Требуется найти *оценку* для этого параметра.

Вероятность того, что при n независимых испытаниях первый исход осуществится x_1 раз, второй — x_2 раз и третий — x_3 раз, равна

$$\frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}.$$

Согласно методу наибольшего правдоподобия, в качестве оценки для неизвестного параметра следует выбрать такое значение ϑ , при котором указанная вероятность будет наибольшей. Так как факториальный множитель не зависит от ϑ , то его можно отбросить и записать функцию правдоподобия так:

$$g(x | \vartheta) = p_1^{x_1} p_2^{x_2} p_3^{x_3}.$$

Кроме того, задачу отыскания максимума функции $g(x | \vartheta)$ можно заменить той же задачей для логарифма этой функции

$$L(x | \vartheta) = x_1 \ln p_1 + x_2 \ln p_2 + x_3 \ln p_3. \quad (3)$$

Если точка максимума находится внутри допустимого интервала изменений ϑ и если в этом интервале функции $p_i(\vartheta)$ дифференцируемы, то в точке максимума производная от L по ϑ равна нулю. Таким образом, получается *уравнение правдоподобия*

$$L'(x | \vartheta) = 0. \quad (4)$$

В простейших случаях уравнение правдоподобия удается решить непосредственно. В остальных же случаях приходится применять метод последовательных приближений, изложенный в § 36.

Если требуется определить лишь какую-либо одну вероятность p , то, как мы видели в § 35, метод наибольшего правдоподобия сразу приводит к оценке

$$\tilde{p} = h = \frac{x}{n},$$

которая является несмещенной и состоятельной. В § 39 (пример 28) было показано, что эта оценка является наилучшей.

Рассмотрим теперь два примера. В примере 32 метод наибольшего правдоподобия приводит к очень хорошей оценке. Пример же 33 показывает, что могут быть случаи, когда этот метод перестает действовать.

Пример 32. Рассмотрим один известный пример, который подробно обсуждается в гл. IX книги Fisher R. A., *Statistical Methods for Research Workers*, Edinburg—London, 8 ed., 1941.

У. А. Карвер изучал генетическое поведение двух наследственных признаков кукурузы. Для исследований были выбраны 4 «чистых» сорта: ($K_0 Z_0$) — крахмалистая кукуруза с зелеными листьями, ($K_0 B_0$) — крахмалистая с белыми листьями, ($C_0 Z_0$) — сахарная с зелеными листьями и ($C_0 B_0$) — сахарная с белыми листьями.

Для выяснения влияния признаков К и С на признаки З и Б было произведено внутрисортное скрещивание гибрида ($K_1 Z_1$) посредством самоопыления. В результате этого скрещивания было получено 3839 потомков со следующим распределением признаков¹:

Крахмалистая кукуруза (К)		Сахарная кукуруза (С)	
с зелеными листьями (КЗ)	с белыми листьями (КБ)	с зелеными листьями (СЗ)	с белыми листьями (СБ)
1997	906	904	32

¹ В этом примере автор исходит из закона Менделя, согласно которому при скрещивании одинаковых чистых сортов с вероятностью единица получается тот же сорт, а при скрещивании разных чистых сортов с вероятностью единица получается гибрид, причем некоторые признаки родителей свойственны и гибриду (доминирующие признаки), а другие признаки родителей у гибрида не проявляются (рецессивные признаки). В данном случае при скрещивании чистых сортов K_0 и C_0 крахмалистой и сахарной кукурузы получается крахмалистый гибрид K_1 , а при скрещивании чистых сортов Z_0 и B_0 — гибрид Z_1 с зелеными листьями. Поэтому признаки К и З называют доминирующими, а С и Б — рецессивными.

При скрещивании двух гибридов K_1 и K_1 (Z_1 и Z_1) получаются: с вероятностью $1/4$ — представитель чистого сорта $K_0(Z_0)$, с вероятностью $1/4$ — представитель чистого сорта $C_0(B_0)$ и с вероятностью $1/2$ — гибрид $K_1(Z_1)$. Следовательно, в первом поколении примерно $1/4$ потомства крахмалистого гибрида (зеленого гибрида) будет иметь рецессивный признак С (Б) и $3/4$ этого потомства будут иметь доминирующий признак К (З). — *Прим. перев.*

Общее отношение числа потомков с доминирующим признаком К к числу потомков с рецессивным признаком С, а также общее отношение числа потомков с доминирующим признаком З к числу потомков с рецессивным признаком Б очень близко к 3:1, что и должно быть по закону Менделя. Однако для сахарной кукурузы отношение числа зеленых индивидуумов к числу белых несколько не похоже на 3:1. Без труда можно было бы установить, что разность этих отношений значительно превышает те границы, внутри которых отклонение отношений друг от друга следовало бы признать случайным. Поэтому такие наследственные признаки, как качество плода (К, С) и цвет листьев (З, Б), не являются независимыми.

Рассмотрим теперь поведение наследственных признаков гибрида ($K_1 Z_1$) более подробно. Пусть $p/2$ — вероятность того, что женская гамета (половая клетка) этого гибрида будет иметь рецессивные признаки (С Б), и пусть $p'/2$ — вероятность того же события для мужской гаметы. Индивидуум с рецессивными признаками (С Б) может возникнуть лишь из женской и мужской гамет с теми же признаками, поэтому вероятность возникновения индивидуума (С Б) равна $pp'/4$. Пусть p_1, p_2, p_3, p_4 — вероятности возникновения индивидуумов с признаками (КЗ), (КБ), (СЗ) и (СБ) соответственно и пусть $pp' = \vartheta$, тогда, в силу закона Менделя, $p_1 + p_2 = 3/4$, $p_3 + p_4 = 1/4$, $p_1 + p_3 = 3/4$, $p_2 + p_4 = 1/4$, где $p_4 = \vartheta/4$. Таким образом,

$$p_1 = \frac{2 + \vartheta}{4}, \quad p_2 = p_3 = \frac{1 - \vartheta}{4}, \quad p_4 = \frac{\vartheta}{4}.$$

Эти вероятности задают, конечно, условное распределение признаков нового индивидуума при условии, что он действительно возник: мы рассматриваем лишь те случаи, когда встречаются гаметы разного пола.

Ясно, что произведение $pp' = \vartheta$ можно оценить по результатам наблюдений, указанным выше. Кроме того, если принять, что $p = p'$, то с помощью оценки для ϑ можно оценить и величину $p = \sqrt{\vartheta}$, называемую рекомбинационным отношением и равную удвоенной вероятности появления гаметы с рецессивными признаками.

Обозначим x_1, x_2, x_3, x_4 количества потомков с признаками (КЗ), (КБ), (СЗ) и (СБ) соответственно, тогда функция правдоподобия будет иметь вид

$$g(x|\vartheta) = p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4} = \left(\frac{2 + \vartheta}{4}\right)^{x_1} \left(\frac{1 - \vartheta}{4}\right)^{x_2 + x_3} \left(\frac{\vartheta}{4}\right)^{x_4}.$$

Если отбросить множитель $1/4$, то логарифмическая функция правдоподобия будет задаваться формулой

$$L(x|\vartheta) = x_1 \ln(2 + \vartheta) + (x_2 + x_3) \ln(1 - \vartheta) + x_4 \ln \vartheta.$$

Поэтому уравнение правдоподобия запишется так:

$$\frac{x_1}{2 + \vartheta} - \frac{x_2 + x_3}{1 - \vartheta} + \frac{x_4}{\vartheta} = 0$$

или

$$p\vartheta^2 - (x_1 - 2x_2 - 2x_3 - x_4)\vartheta - 2x_4 = 0. \quad (5)$$

Положительный корень этого уравнения является оценкой наибольшего правдоподобия $\hat{\vartheta}$.

Легко можно найти и другие оценки для параметра ϑ , в том числе даже и несмещенные. Например, очевидно, что

$$T_{\vartheta} = 4h_{\vartheta} = \frac{4x_4}{\vartheta}$$

и

$$T_1 = h_1 - h_2 - h_3 + h_4 = \frac{x_1 - x_2 - x_3 + x_4}{n}$$

представляют собой несмещенные оценки для ϑ . Но если интересоваться не только смещением, а и дисперсией, то окажется, что оценка наибольшего правдоподобия значительно лучше всех остальных. Смещение оценки $\tilde{\vartheta}$ является величиной порядка $1/n$, и ее дисперсия, при больших n , асимптотически равна той наименьшей дисперсии, которая вообще возможна, согласно неравенству Фреше, для несмещенных оценок. Хотя T_0 и T_1 представляют собой несмещенные оценки, однако их дисперсии значительно превосходят дисперсию $\tilde{\vartheta}$, причем эти оценки не являются асимптотически эффективными.

Фишер сравнивал друг с другом пять различных оценок T_1, \dots, T_5 , где T_1 совпадает с нашей оценкой T_1 и T_4 совпадает с оценкой $\tilde{\vartheta}$. T_5 является оценкой, полученной посредством минимизации

$$\chi^2 = \sum \frac{(x_i - np_i)^2}{np_i} = \text{minimum.}$$

Исследования Фишера показывают, что только последние три оценки, T_3, T_4 и T_5 , являются эффективными в том смысле, что их смещения являются величинами высшего порядка малости по сравнению с $1/\sqrt{n}$, а их дисперсии асимптотически равны той наименьшей дисперсии

$$\sigma_{\min}^2 = \frac{1}{I(\vartheta)} = \frac{c}{n},$$

которая возможна для несмещенных оценок, согласно неравенству Фреше.

Позднее мы еще вернемся к вопросу об эффективности оценок наибольшего правдоподобия.

Пример 33. Произведено n выстрелов из пушки по неподвижной точечной цели без перемены прицела, причем в k случаях наблюдался перелет, а в остальных l случаях — недолет ($k + l = n$). В дальнейшем, когда мы будем говорить «высота выстрела», то под этим будет подразумеваться высота точки попадания снаряда в вертикальную плоскость, проходящую через цель перпендикулярно к направлению стрельбы. Пусть все высоты выстрелов — независимые, одинаково нормально распределенные случайные величины с известным квадратичным отклонением и неизвестным математическим ожиданием. Какую поправку нужно внести в установку пушки, чтобы среднее значение высот выстрелов было возможно ближе к высоте цели?

Если высоты выстрелов отсчитывать от горизонтальной плоскости, проходящей через цель, а квадратичное отклонение положить равным единице, то плотность вероятности для отдельной высоты выстрела будет иметь вид

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

Среднее значение этого распределения равно μ , и поэтому искомая поправка равна $-\mu$. Вероятность того, что данный снаряд перелетит цель, задается формулой

$$\int_0^{\infty} g(x|\mu) dx = \Phi(\mu).$$

Вероятность того, что при n выстрелах k раз будет наблюдаться перелет, а остальные $n - k$ раз — не долет, равна

$$\binom{n}{k} [\Phi(\mu)]^k [1 - \Phi(\mu)]^{n-k}.$$

Логарифмическая функция правдоподобия

$$L(k|\mu) = k \ln \Phi(\mu) + (n - k) \ln [1 - \Phi(\mu)]$$

достигает максимума при таком значении μ , которое удовлетворяет уравнению

$$\Phi(\mu) = \frac{k}{n}.$$

Следовательно, правдоподобное значение $\tilde{\mu}$ выражается с помощью функции Ψ , обратной для функции Φ :

$$\tilde{\mu} = \Psi\left(\frac{k}{n}\right).$$

Если $k = 0$, то $\tilde{\mu} = -\infty$, если же $k = n$, то $\tilde{\mu} = +\infty$. И так как хотя бы одно из двух событий $k = 0$ и $k = n$ обязательно имеет положительную вероятность, то для случайной величины μ , строго говоря, не существует ни конечного среднего значения, ни тем более конечной дисперсии.

На практике этот недостаток можно, разумеется, легко исправить, если в обоих крайних случаях $k = 0$ и $k = n$ заменить оценку $\tilde{\mu}$ разумно выбранными конечными значениями; ведь положение цели, более или менее грубо, всегда бывает известно! Однако прежний результат остается в силе, так как в данном случае дословное применение метода наибольшего правдоподобия при любом n приводит к оценке с бесконечно большой дисперсией.

Тем не менее оценка $\tilde{\mu}$ является состоятельной¹: при $n \rightarrow \infty$ она сходится по вероятности к истинному значению μ .

§ 47. Состоятельность оценок наибольшего правдоподобия

При весьма общих предположениях оценка наибольшего правдоподобия является состоятельной. В § 45 мы ссылались на доказательство этого общего утверждения, принадлежащее Вальду и Вольфовицу. Теперь, когда в качестве результатов наблюдений рассматриваются частоты, мы постараемся исследовать этот вопрос несколько подробнее.

¹ Если $\varphi(\vartheta)$ — непрерывно дифференцируемая функция параметра ϑ , причем $\varphi'(\vartheta) \neq 0$, и если для ϑ существует асимптотически нормальная и асимптотически эффективная оценка $\tilde{\vartheta}$, то $\varphi(\tilde{\vartheta})$ будет асимптотически нормальной и асимптотически эффективной оценкой для $\varphi(\vartheta)$ (см. Neuman J., Contribution to the theory of the χ^2 , I. Berkeley Sumpos. on Math. Stat., Los Ang. (1949), 239). Так как k/n — наилучшая асимптотически нормальная оценка для $\Phi(\mu)$ и $\Phi'(\mu) \neq 0$, то, в силу этой теоремы, $\tilde{\mu} = \Psi(k/n)$ является не только состоятельной, но и, что более важно, асимптотически нормальной и асимптотически эффективной оценкой для μ . — Прим. перев.

Пусть снова наблюдаются три частоты:

$$h_i = \frac{x_i}{n} \quad (i = 1, 2, 3),$$

и пусть вероятности p_1, p_2, p_3 трех взаимно исключающих исходов являются функциями одного параметра ϑ .

Предположим, что между ϑ и (p_1, p_2, p_3) имеется непрерывное, взаимно однозначное соответствие, т. е. ϑ и ϑ^0 различны тогда и только тогда, когда различны соответствующие точки с координатами (p_1, p_2, p_3) и (p_1^0, p_2^0, p_3^0) , причем если $(p_1, p_2, p_3) \rightarrow (p_1^0, p_2^0, p_3^0)$, то $\vartheta \rightarrow \vartheta^0$.

Если не делать таких предположений, то по результатам наблюдений невозможно будет получить приближенное значение для ϑ , так как в данном случае результаты наблюдений — частоты — сами являются лишь приближенными значениями для вероятностей p_i .

Как и в § 46, функция правдоподобия задается формулой

$$g(x | \vartheta) = p_1^{x_1} p_2^{x_2} p_3^{x_3}.$$

Если эту формулу умножить на не зависящий от ϑ множитель¹

$$n^n x_1^{-x_1} x_2^{-x_2} x_3^{-x_3},$$

то получится функция

$$G(x | \vartheta) = \left(\frac{np_1}{x_1}\right)^{x_1} \left(\frac{np_2}{x_2}\right)^{x_2} \left(\frac{np_3}{x_3}\right)^{x_3}, \quad (1)$$

достигающая своего максимума в той же точке ϑ , что и $g(x | \vartheta)$. Логарифм G равен

$$L(x | \vartheta) = x_1 \ln \frac{np_1}{x_1} + x_2 \ln \frac{np_2}{x_2} + x_3 \ln \frac{np_3}{x_3}. \quad (2)$$

Формула (2) определяет $L(x | \vartheta)$ при всех допустимых значениях параметра ϑ и, в частности, при $\vartheta = \vartheta^*$, где ϑ^* — неизвестное истинное значение ϑ . Рассмотрим разности

$$z_i = x_i - np_i \quad (p_i > 0). \quad (3)$$

Если $\vartheta = \vartheta^*$ и n достаточно велико, то, согласно результатам § 5, с вероятностью, сколь угодно близкой к единице, случайные величины x_i будут отличаться от соответствующих математических ожиданий np_i не более, чем на величину порядка $\sqrt{np_i}$. Иначе говоря, для всяких $\delta > 0$ и $C > 0$ найдется такое $n_0 = n_0(\delta, C)$, что при любом $n > n_0$ абсолютные величины $|z_i|$

¹ Если $x_i = 0$, то условимся считать, что $x_i^{x_i} = 1$. — Прим. перев.

не будут превосходить $C\sqrt{np_i}$ с вероятностью $P \geq 1 - \delta$. При этом

$$\sum z = z_1 + z_2 + z_3 = 0. \quad (4)$$

Если в (2) все np_i выразить через x_i и z_i , то, в силу (4), получим

$$\begin{aligned} L(x | \vartheta) &= \sum x \ln \frac{x-z}{x} = \sum z + \sum x \ln \left(1 - \frac{z}{x}\right) = \\ &= \sum x \left[\frac{z}{x} + \ln \left(1 - \frac{z}{x}\right) \right] = \sum x \varphi \left(\frac{z}{x}\right). \end{aligned} \quad (5)$$

Эта формула, разумеется, справедлива не только для истинного значения ϑ^* , но и для всех допустимых значений параметра ϑ . Кроме того, она справедлива для произвольного количества частот и произвольного количества параметров $\vartheta_1, \dots, \vartheta_r$.

Функция

$$\varphi(t) = t + \ln(1-t),$$

входящая в (5), имеет максимум $\varphi(0) = 0$, так как ее производная

$$\varphi'(t) = 1 - \frac{1}{1-t}$$

обращается в нуль при $t = 0$, причем если $t < 0$, то $\varphi'(t) > 0$, если же $t > 0$, то $\varphi'(t) < 0$. Следовательно, в (5) все слагаемые неположительны.

Если $|t| < 1$, то $\varphi(t)$ можно разложить в степенной ряд:

$$\varphi(t) = -\frac{1}{2}t^2 - \frac{1}{3}t^3 - \dots \quad (6)$$

В частности, если $t = z/x$, где z является величиной порядка \sqrt{np} , то при $0 < p < 1$ имеем

$$|t| = \frac{|z|}{x} = \frac{|z|}{np+z} \leq \frac{|z|}{np-|z|},$$

следовательно, t — величина порядка $1/\sqrt{np}$ и, в силу (6), $-\varphi(t)$ — величина порядка $1/np$. Но если z — величина порядка \sqrt{np} , то $x = np + z$ — величина порядка np . Таким образом, все слагаемые (5) представляют собой величины порядка единицы. Иными словами, если $\vartheta = \vartheta^*$, то для всякого $\varepsilon > 0$ найдутся такие $n_1 = n_1(\varepsilon)$ и $g = g(\varepsilon)$, что при $n > n_1$ с вероятностью $P \geq 1 - \varepsilon$ будет выполняться неравенство

$$L(x | \vartheta) \geq -g. \quad (7)$$

Напротив, если $|z|$ велико сравнительно с \sqrt{x} , то $t = z/x$ велико сравнительно с $1/\sqrt{x}$, и поэтому $-\varphi(t)$ велико сравнительно с $1/x$. Таким образом, $-L(x | \vartheta)$ будет бесконечно большой

величиной при $n \rightarrow \infty$. Иначе говоря, если $\hat{v} \neq v^*$, то для всякого $\varepsilon > 0$ и всякого $g > 0$ найдется такое $n_2 = n_2(\varepsilon, g)$, что при $n > n_2$ с вероятностью $P \geq 1 - \varepsilon$ будет выполняться неравенство

$$L(x | \hat{v}) < -g. \quad (8)$$

Итак, мы установили, что для истинного значения параметра v^* (и, значит, для соответствующих истинных значений вероятностей p^*) с большой вероятностью будет выполняться неравенство (7). Пусть \tilde{v} — оценка наибольшего правдоподобия для v^* , т. е. такое значение параметра v , при котором функция $G(x|v)$ (а значит, и $L(x | v)$) достигает максимума, тогда $G(x | \tilde{v}) \geq G(x|v^*)$, и поэтому $L(x | \tilde{v})$ с большой вероятностью удовлетворяет неравенству (7), а соответствующие разности $z = x - n\tilde{p}$ не превосходят величин порядка \sqrt{x} . Таким образом, при $n \rightarrow \infty$ абсолютные величины $|\tilde{z}|$ со сколь угодно близкой к единице вероятностью, являются величинами порядка

$$\sqrt{x} \sim \sqrt{np^*}.$$

Так как $z = x - np^*$ с вероятностью, сколь угодно близкой к единице, являются величинами порядка $\sqrt{np^*}$, то разности $z - \tilde{z} = n(\tilde{p} - p^*)$ с вероятностью, сколь угодно близкой к единице, будут величинами того же порядка, т. е.

При достаточно большом n разности $\tilde{p} - p^$ с вероятностью, сколь угодно близкой к единице, будут являться величинами порядка $1/\sqrt{n}$.*

В силу непрерывности функции $\hat{v} = \hat{v}(p_1, p_2, p_3)$, заключаем, что $\tilde{v} = \hat{v}(\tilde{p}_1, \tilde{p}_2, \tilde{p}_3)$ сходится по вероятности к $v^* = \hat{v}(p_1^*, p_2^*, p_3^*)$ при $n \rightarrow \infty$, чем и завершается доказательство состоятельности оценки наибольшего правдоподобия¹.

Если $p_i = p_i(v)$ и $\hat{v} = \hat{v}(p_1, p_2, p_3)$ являются дифференцируемыми функциями, то имеет место следствие этой теоремы:

Разность $\tilde{v} - v^$ с вероятностью, сколь угодно близкой к единице, является величиной порядка $1/\sqrt{n}$.*

¹ В этой теореме предположение о непрерывном, взаимно однозначном соответствии между v и (p_1, p_2, p_3) является излишне сильным. Теорема останется справедливой, если потребовать лишь, чтобы параметр v являлся непрерывной функцией точки (p_1, p_2, p_3) и чтобы во всех n опытах значения вероятностей p_1, p_2, p_3 были такими же, как и в первом опыте. При этом доказательство будет точно таким же, как у автора. Указанная теорема имеет большое практическое значение, так как позволяет оценивать любые непрерывные функции от неизвестных вероятностей p_i . Асимптотические свойства таких оценок устанавливаются теоремой, указанной в сноске на стр. 229. — *Прим. перев.*

Если (6) подставим в (5), то получим очень полезное разложение функции $L(x | \vartheta)$ в ряд:

$$\begin{aligned} L(x | \vartheta) &= -\frac{1}{2} \sum \frac{z^2}{x} - \frac{1}{3} \sum \frac{z^3}{x^2} - \dots = \\ &= -\frac{1}{2} \sum \frac{(x-np)^2}{x} - \frac{1}{3} \sum \frac{(x-np)^3}{x^2} - \dots \end{aligned} \quad (9)$$

В большинстве случаев оказывается достаточным приближение

$$L(x | \vartheta) \sim -\frac{1}{2} \sum \frac{(x-np)^2}{x}. \quad (10)$$

§ 48. Наибольшее правдоподобие, минимум χ^2 и наименьшие квадраты

Только что указанное приближение функции правдоподобия очень удобно для быстрого и экономного вычисления первого приближения для $\tilde{\vartheta}$. При этом вместо максимума функции $L(x | \vartheta)$ нужно найти минимум квадратичной формы

$$\chi_x^2 = \sum \frac{(x-np)^2}{x}. \quad (1)$$

Форма χ_x^2 лишь незначительно отличается от известного выражения

$$\chi^2 = \sum \frac{(x-np)^2}{np}. \quad (2)$$

Для отыскания оценки параметра ϑ можно воспользоваться условием $\chi^2 = \text{minimum}$. Однако этот метод в большинстве случаев приводит к сложным вычислениям, так как для отыскания точки минимума приходится дифференцировать знаменатели np_i .

Но если в знаменателях все p заменить некоторыми подходящими оценками $p^{(0)}$, то дифференцирования знаменателей можно будет избежать, в этом случае будет идти речь о минимуме выражения

$$\chi_0^2 = \sum \frac{(x-np)^2}{np^{(0)}}. \quad (3)$$

Вычисление этих различных оценок мы исследуем несколько подробнее и докажем, что они отличаются друг от друга величинами порядка $1/n$. При этом мы рассмотрим лишь следующие три метода отыскания оценок:

А. Метод минимума χ_0^2 .

Б. Метод минимума χ_x^2 .

В. Метод наибольшего правдоподобия.

Для упрощения выкладок мы, далее, будем предполагать, что p_i являются линейными функциями параметров ϑ . Результаты

могут быть распространены на нелинейные дифференцируемые функции, так как в окрестности истинных значений параметров $\vartheta = \vartheta^*$ эти функции можно приблизить линейными. Поэтому функции p_i , встречающиеся в примерах этой главы, часто являются нелинейными, хотя теория излагается лишь для линейных функций¹.

А. МЕТОД МИНИМУМА χ_0^2

Квадратичная форма

$$\chi_0^2 = \sum \frac{(x - x')^2}{np^{(0)}} \quad (4)$$

в пространстве (x_1, \dots, x_m) определяет евклидову метрику: χ_0^2 — квадрат расстояния между точками $X(x_1, \dots, x_m)$ и $X'(x'_1, \dots, x'_m)$. Если положим

$$x'_i = np_i(\vartheta), \quad (i = 1, \dots, m),$$

где $p_i(\vartheta)$, как указано выше, — линейные функции параметров $\vartheta_1, \dots, \vartheta_r$, то множество всех точек X' с координатами (x'_1, \dots, x'_m) будет представлять собой линейное подпространство G . Условие $\chi_0^2 = \text{minimum}$ означает, что точка X' этого подпространства выбирается таким образом, чтобы расстояние между X' и наблюдаемой точкой X было наименьшим. Следовательно, из X нужно опустить перпендикуляр на линейное подпространство G ; основание этого перпендикуляра является точкой X' (рис. 24).

Тот же самый результат получается и с помощью вычислений. Если (3) продифференцировать по всем ϑ_α и производные приравнять нулю, то получим систему уравнений

$$\sum \frac{\left[\frac{x_i}{n} - p_i(\vartheta) \right] q_{i\alpha}}{p_i^{(0)}} = 0, \quad (5)$$

где

$$q_{i\alpha} = \frac{\partial p_i}{\partial \vartheta_\alpha}. \quad (6)$$

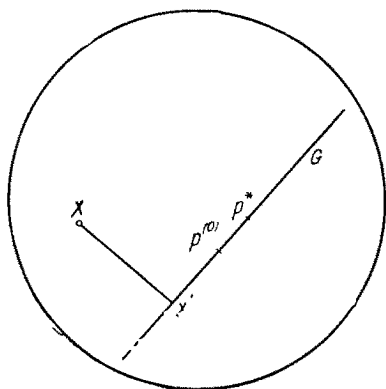


Рис. 24.

¹ Здесь имеются в виду лишь такие функции $p_i(\vartheta)$, которые являются линейными в некоторой ограниченной области изменения параметров ϑ . Поэтому определяемое далее множество G не является линейным подпространством (хотя бы потому, что оно ограничено). Выводы этого параграфа нельзя признавать строгими, однако они правильно отражают действительную связь рассматриваемых методов. обстоятельное и строгое изложение данного вопроса см., например, в книге Крамера Г., Математические методы статистики, ИЛ, М., 1948, гл. XXX. — *Прим. перев.*

Если $\vartheta_1, \vartheta_2, \dots, \vartheta_r$ — решение системы (5), то вектор, направленный из точки X с координатами x_i/n , в точку X' с координатами $p_i(\vartheta)$, будет перпендикулярен ко всем векторам, касающимся множества G в точке X' .

В этой формулировке наше утверждение справедливо даже и тогда, когда G не является линейным подпространством пространства всех X . Однако, по предположению, $p_i(\vartheta)$ — линейные функции параметров ϑ_α , поэтому производные $q_{i\alpha}$ равны постоянным и мы можем положить

$$p_i(\vartheta) = p_i(0) + \sum_{\beta} q_{i\beta} \vartheta_{\beta}. \quad (7)$$

Изменением начала отсчета в пространстве параметров можно добиться, чтобы при $\vartheta_{\beta} = 0$ соответствующие значения вероятностей p_i равнялись начальным приближениям $p_i^{(0)}$ в формуле (3), т. е. $p_i(0) = p_i^{(0)}$; это несколько упростит выкладки.

Если (7) подставим в (5), то получим r линейных уравнений с r неизвестными $\vartheta_1, \dots, \vartheta_r$:

$$\sum_{\beta} h_{\alpha\beta} \vartheta_{\beta} = \sum_i \frac{[x_i - np_i^{(0)}] q_{i\alpha}}{p_i^{(0)}}, \quad (8)$$

где

$$h_{\alpha\beta} = \sum_i \frac{nq_{i\alpha}q_{i\beta}}{p_i^{(0)}}. \quad (9)$$

Коэффициенты $h_{\alpha\beta}$ имеют тот же самый вид, что и суммы $[gaa]$, $[gab]$, ... в теории метода наименьших квадратов. Таким образом, задача отыскания минимума квадратичной формы χ_0^2 совпадает с задачей теории метода наименьших квадратов, причем в данном случае веса g_i наблюдаемых частот равняются $n/p_i^{(0)}$.

Попытаемся теперь выяснить, в какой степени выбор начального приближения $p_1^{(0)}, \dots, p_n^{(0)}$ влияет на точку X' .

В качестве координат точки X мы примем не x_i , а частоты $h_i = x_i/n$, и расстояние определим формулой

$$r^2 = \frac{1}{n} \chi_0^2 = \sum \frac{(h_i - h_i')^2}{p_i^{(0)}}. \quad (10)$$

Предположим, что точка X с координатами h_i и точка P_0 с координатами $p_i^{(0)}$ находится в окрестности «истинной точки» P^* с координатами p_i^* , причем радиус этой окрестности является величиной порядка $1/\sqrt{n}$, и, кроме того, координаты p_i всех точек P этой окрестности ограничены снизу некоторым положительным числом: $p_i \geq \delta > 0$.

Если теперь P_0 заменить точкой S_0 из той же окрестности, то все разности соответствующих координат $p_i^{(0)} - s_i^{(0)}$ будут величинами порядка $\varepsilon = 1/\sqrt{n}$. Так как точка X и линейное подпро-

странство G не зависят от выбора P_0 , то от замены P_0 на S_0 изменится лишь квадрат расстояния (10), причем $r^2(S_0)$ отличается от $r^2(P_0)$ на величину порядка ϵ . Направление перпендикуляра XX' также может измениться, однако угол между новым и старым направлениями является лишь величиной порядка ϵ . Но, по предположению, длина старого перпендикуляра представляет собой величину порядка ϵ , следовательно, координаты точки X' при замене P_0 на S_0 изменятся на величину порядка $\epsilon^2 = 1/n$.

Вообще, если разности координат $p_i^{(0)} - s_i^{(0)}$ являются величинами порядка η , то при замене P_0 на S_0 координаты точки X' изменятся на величину порядка $\epsilon\eta$. Или, точнее, если для всех i имеют место неравенства $|p_i^{(0)} - s_i^{(0)}| < \eta$, то абсолютные величины разностей координат точек $X'(P_0)$ и $X'(S_0)$ будут меньше, чем $c\epsilon\eta$, где $c > 0$ — постоянный числовой множитель.

Теоретическая оценка постоянной величины c не приносит значительной пользы. Такие оценки всегда оказываются слишком грубыми, так как они в большинстве случаев значительно превосходят те разности, которые встречаются на практике. Для практических целей достаточно установить, что точка X' лишь в очень незначительной степени зависит от выбора исходного приближения $p_i^{(0)}$

Б. МЕТОД МИНИМУМА χ_x^2

Если $p_i^{(0)}$ выбрать равными наблюдаемым частотам $h_i = x_i/n$, то χ_0^2 перейдет в χ_x^2 . Отсюда следует, что минимум χ_x^2 вычисляется точно так же, как и минимум χ_0^2 , а именно, по методу наименьших квадратов. При этом с вероятностью, сколь угодно близкой к единице, точка P_x минимума квадратичной формы χ_x^2 находится от ранее введенной точки X' на расстоянии порядка $1/n$.

В. МЕТОД НАИБОЛЬШЕГО ПРАВДОПОДОБИЯ

Если логарифмическая функция правдоподобия

$$L(x | \vartheta) = \sum x_i \ln p_i(\vartheta)$$

дифференцируема и достигает максимума внутри области допустимых значений параметров ϑ , то в точке максимума производные от L обращаются в нуль. Таким образом, координаты точки максимума должны удовлетворять уравнениям

$$\sum_i \frac{x_i q_{ia}}{p_i} = 0, \quad (11)$$

где $q_{ia} = \partial p_i / \partial \vartheta_a$. Так как сумма всех p_i равна единице, то суммы соответствующих производных равны нулю:

$$\sum_i q_{ia} = 0. \quad (12)$$

Если (12) умножим на n и результат вычтем из (11), то получим

$$\sum_i \frac{(x_i - np_i) q_{ia}}{p_i} = 0. \quad (13)$$

При этом p_i в числителях и знаменателях являются линейными функциями от ϑ . Решение системы (13) представляет собой оценку наибольшего правдоподобия $\hat{\vartheta}$, следовательно,

$$\sum_i \frac{[x_i - np_i(\hat{\vartheta})] q_{ia}}{p_i(\hat{\vartheta})} = 0. \quad (14)$$

Решение системы уравнений (14) можно очень легко найти методом последовательных приближений. Для этого сначала заменим ϑ в знаменателях каким-либо приближенным значением $\vartheta^{(0)}$. Соответствующая система уравнений

$$\sum_i \frac{[x_i - np_i(\vartheta)] q_{ia}}{p_i(\vartheta^{(0)})} = 0 \quad (15)$$

представляет собой систему (5) метода минимума χ_0^2 и, следовательно, ее можно легко решить как систему нормальных уравнений метода наименьших квадратов. Если решение $\vartheta^{(1)}$ системы (15) снова подставить в знаменатели (14), то тем же самым способом получим улучшенное приближение $\vartheta^{(2)}$ и т. д.

Последовательность приближений $\vartheta^{(k)}$ сходится, и предел этой последовательности не зависит от выбора исходного приближения $p_i^{(0)} = p_i(\vartheta^{(0)})$. Действительно, если $p_i^{(0)}$ и $s_i^{(0)}$ — два различных исходных приближения, отличающихся друг от друга лишь величинами порядка $\varepsilon = 1/\sqrt{n}$, то, согласно ранее доказанному, первые приближения $p_i^{(1)}$ и $s_i^{(1)}$ отличаются друг от друга лишь величинами порядка ε^2 , вторые приближения — лишь величинами порядка ε^3 и т. д. Если теперь в качестве $s_i^{(0)}$ выбрать решение уравнений правдоподобия, то будут выполняться равенства $s_i^{(0)} = s_i^{(1)} = s_i^{(2)} = \dots$, и так как для любых $p_i^{(0)}$ из ε -окрестностей $s_i^{(0)}$ справедливы неравенства $|p_i^{(k)} - s_i^{(k)}| = |p_i^{(k)} - s_i^{(0)}| < c\varepsilon^{k+1}$, то последовательности $p_i^{(0)}$, $p_i^{(1)}$, $p_i^{(2)}$, ... сходятся к решению $s_i^{(0)}$, независимо от выбора $p_i^{(0)}$ из ε -окрестностей $s_i^{(0)}$.

При доказательстве предполагалось, что решение уравнений наибольшего правдоподобия существует. Это предположение выполняется в том случае, когда все x_i строго положительны. Действительно, условия $\sum p_i = 1$, $p_i \geq 0$ определяют в пространстве всех (p_1, \dots, p_m) некоторую замкнутую ограниченную область. Часть линейного подпространства G , принадлежащая этой области, также является замкнутой и ограниченной. Функция правдоподобия

$$g(x | \vartheta) = \prod_i [p_i(\vartheta)]^{x_i}, \quad (16)$$

по предположению, непрерывна, следовательно, в замкнутой области она имеет максимум. Этот максимум не может достигаться на границе области, так как функция $g(x | \hat{v})$ там равна нулю.

Кроме того, из приведенного выше доказательства следует, что если точка с координатами (x_1, \dots, x_m) находится достаточно близко от подпространства G , то решение уравнений правдоподобия единственно. Действительно, если бы $p_i^{(0)}$ и $s_i^{(0)}$ были двумя различными решениями уравнений правдоподобия, то выполнялись бы равенства $p_i^{(0)} = p_i^{(1)} = p_i^{(2)} = \dots$ и $s_i^{(0)} = s_i^{(1)} = s_i^{(2)} = \dots$. Но первая последовательность, по доказанному, должна сходиться к $s_i^{(0)}$, что возможно лишь тогда, когда при всех i имеют место равенства $p_i^{(0)} = s_i^{(0)}$.

С небольшими изменениями это доказательство применимо и к случаю, когда p_i являются нелинейными функциями параметров, если только множество точек с координатами $p_1(\hat{v}), \dots, p_m(\hat{v})$ является замкнутым, или если оно не замкнуто, но все его несобственные предельные точки расположены на границе области $p_i \geq 0, \sum p_i = 1$. Если множество точек с координатами $p_1(\hat{v}), \dots, p_m(\hat{v})$ не замкнуто и некоторые его несобственные предельные точки расположены внутри области $p_i \geq 0, \sum p_i = 1$, то в доказательстве могут возникнуть осложнения.

Для приложений особенно важно то, что последовательные приближения хорошо сходятся. А именно: разности $p_i^{(k)} - s_i^{(0)}$ стремятся к нулю как степенная функция $\epsilon^{k+1} = n^{-(k+1)/2}$. Если нулевое приближение $p^{(0)}$ выбрано не слишком плохо, то спокойно можно довольствоваться первым приближением $p^{(1)}$: все дальнейшие приближения будут отличаться от $p^{(1)}$ лишь величинами порядка $\epsilon^2 = 1/n$. Так как неизбежные статистические отклонения \hat{v} от истинного значения v^* являются величинами порядка $\epsilon = 1/\sqrt{n}$, то бессмысленно добиваться дальнейшего повышения точности.

Результаты, установленные для методов А, Б и В, позволяют сделать вывод: *Оценки А и Б отличаются от оценки наибольшего правдоподобия В лишь величинами порядка $1/n$.*

Все эти соображения сохраняют силу и тогда, когда наблюдается не один, а несколько рядов частот, причем в каждом ряду сумма частот равна единице, например:

$$h_1 + h_2 = 1 \text{ или } x_1 + x_2 = n_1,$$

$$h_3 + h_4 = 1 \text{ или } x_3 + x_4 = n_2$$

и т. д. Выражения χ^2, χ_0^2, \dots останутся неизменными, с той только разницей, что каждую вероятность p_i нужно будет умножить на соответствующий множитель n_i , например:

$$\chi^2 = \sum \frac{(x_i - n_i p_i)^2}{n_i p_i}.$$

§ 49. Асимптотическое распределение χ^2 и $\tilde{\nu}$ при $n \rightarrow \infty$

Для простоты мы ограничимся случаем одного параметра $\tilde{\nu}$ и постараемся исследовать поведение функции распределения оценки $\tilde{\nu}$.

Оценка $\tilde{\nu}$, найденная методом наименьших квадратов, представляла собой линейную функцию результатов наблюдений x_i . Так как предполагалось, что все x_i распределены нормально, то и оценка $\tilde{\nu}$ имела нормальное распределение. В данном случае, однако, x_i являются дискретными случайными величинами, распределенными лишь приближенно нормально, а оценка $\tilde{\nu}$ представляет собой лишь приближенно линейную функцию результатов наблюдений x_i . Поэтому мы можем ожидать, что распределение $\tilde{\nu}$ будет лишь асимптотически нормальным при $n \rightarrow \infty$.

Вероятность того, что наблюдаемая точка X с координатами x_i будет принадлежать области B пространства иксов, равна сумме всех вероятностей, соответствующих отдельным точкам области B :

$$P(B) = \sum_{X \in B} P(X), \quad (1)$$

где

$$P(X) = \frac{n!}{x_1! \dots x_m!} p_1^{x_1} \dots p_m^{x_m}. \quad (2)$$

При этом $p_i = p_i(\tilde{\nu})$ представляют собой истинные вероятности p_1^*, \dots, p_m^* . В дальнейшем эти вероятности мы будем обозначать без звездочек.

Для больших n выражение (2) можно преобразовать с помощью формулы Стирлинга и получить

$$P(X) \sim x_1^{-\left(x_1 + \frac{1}{2}\right)} \dots x_m^{-\left(x_m + \frac{1}{2}\right)} n^{n + \frac{1}{2}} (2\pi)^{\frac{1-m}{2}} p_1^{x_1} \dots p_m^{x_m} \quad (3)$$

или

$$\gamma P(X) \sim \left(\frac{np_1}{x_1}\right)^{x_1 + \frac{1}{2}} \dots \left(\frac{np_m}{x_m}\right)^{x_m + \frac{1}{2}},$$

где

$$\gamma = [(2\pi n)^{m-1} p_1 \dots p_m]^{\frac{1}{2}}. \quad (4)$$

Логарифм $\gamma P(X)$ равен

$$\ln [\gamma P(X)] = \sum \left(x + \frac{1}{2}\right) \ln \frac{np}{x} + \dots, \quad (5)$$

причем остаток, обозначенный отточием, представляет собой величину порядка $1/n$. Положим снова

$$x_i = np_i + z_i, \quad (6)$$

тогда z_i с вероятностью, сколь угодно близкой к единице, будут являться величинами порядка $\sqrt{np_i}$. Поэтому

$$\begin{aligned} \ln[\gamma P(X)] &= - \sum \left(np + z + \frac{1}{2} \right) \ln \frac{np+z}{np} + \dots = \\ &= - \sum \left(np + z + \frac{1}{2} \right) \left(\frac{z}{np} - \frac{z^2}{2n^2p^2} + \frac{z^3}{3n^3p^3} \right) + \dots = \\ &= - \frac{1}{2} \sum \frac{z^2}{np} - \frac{1}{2} \sum \frac{z}{np} + \frac{1}{6} \sum \frac{z^3}{n^2p^2} + \dots, \end{aligned}$$

где остаток, обозначенный отточием, с вероятностью, сколь угодно близкой к единице, представляет собой величину порядка $1/n$. Если теперь обозначим

$$\sum \frac{z^2}{np} = \chi^2,$$

то получим асимптотическое равенство:

$$\begin{aligned} \gamma P(X) &\sim e^{-\frac{1}{2}\chi^2} \cdot e^{-\frac{1}{2} \sum \frac{z}{np} + \frac{1}{6} \sum \frac{z^3}{n^2p^2}} \sim \\ &\sim e^{-\frac{1}{2}\chi^2} \left(1 - \frac{1}{2} \sum \frac{z}{np} + \frac{1}{6} \sum \frac{z^3}{n^2p^2} \right). \quad (7) \end{aligned}$$

Последние два члена в круглых скобках являются величинами порядка $1/\sqrt{n}$. Эти члены мы рассмотрим несколько позднее, а пока ограничимся главным членом

$$\gamma P(X) \sim e^{-\frac{1}{2}\chi^2}, \quad (8)$$

где

$$\chi^2 = \sum \frac{(x - np)^2}{np}. \quad (9)$$

Формула (8) показывает, что вероятность, соответствующая точке $X(x_1, \dots, x_m)$, будет тем меньше, чем больше x_i отличаются от математических ожиданий np_i , так как с ростом абсолютных величин разностей $x_i - np_i$ функция χ^2 возрастает.

Как мы уже видели раньше, функция χ^2 определяет эвклидову метрику в пространстве переменных x_1, \dots, x_m : эта функция, с точностью до множителя n , представляет собой квадрат расстояния между переменной точкой X с координатами x_i/n и постоянной точкой P с координатами p_i . Согласно (8), чем дальше X удалена от P , тем меньше будет вероятность, соответствующая

точке X . В этой метрике поверхности с уравнениями $\chi = \text{const}$ являются сферами с центром в точке P . Если эти сферы (в случае $m = 3$) пересесть плоскостью $x_1 + x_2 + x_3 = n$, проходящей через точку P , то в пересечении получатся концентрические окружности с центром в точке P .

Так как с большой вероятностью разности $x - np$ являются лишь величинами порядка \sqrt{n} , то χ^2 с большой вероятностью является величиной порядка единицы, т. е. для всякого $\eta > 0$ найдутся такие R^2 и n_0 , что при всех $n > n_0$ вероятность события $\chi^2 < R^2$ будет больше, чем $1 - \eta$. Поэтому при выводе асимптотических формул для вероятностей всегда можно ограничиться такой окрестностью точки P , где $\chi^2 < R^2$.

В частности, мы хотим вычислить следующие вероятности:

- а) функцию распределения квадратичной формы χ^2 , т. е. вероятность события $\chi^2 < u$;
- б) функцию распределения оценки наибольшего правдоподобия $\tilde{\nu}$.

В указанной окрестности точки P введем новые координаты

$$y_i = \frac{x_i - np_i}{\sqrt{np_i}}, \quad (10)$$

где все y_i являются величинами порядка единицы. Тогда, в новых координатах, выражение для квадрата расстояния $\chi^2 = (PX)^2$ будет совсем простым:

$$\chi^2 = \sum y_i^2. \quad (11)$$

Следовательно, y_i являются обычными прямоугольными координатами в пространстве с метрикой (11).

Множество точек X с целочисленными координатами x_i , удовлетворяющими условию $\sum x_i = n$, определяет решетку в гиперплоскости $\sum y_i \sqrt{np_i} = 0$. Размерность этой гиперплоскости равна $m - 1$. В метрике (11) длины всех векторов базиса решетки являются величинами порядка $1/\sqrt{n}$, следовательно, объем каждой элементарной клетки представляет собой величину порядка $n^{-(m-1)/2}$. Вероятность того, что точка X принадлежит какой-либо области B , равна сумме вероятностей, соответствующих тем точкам решетки, которые принадлежат B .

Асимптотическая формула для этой вероятности легче всего выводится в случае а), когда область B определяется неравенством $\chi^2 < u$. В метрике (11) B является шаром. Величину суммы (1) по точкам решетки, расположенным внутри шара, оценивал в свое время К. Пирсон, который впервые использовал в математической статистике квадратичную форму χ^2 . Метод оценки суммы (1) заключается в следующем:

Сначала вероятности $\mathbf{P}(\bar{X})$ заменяют, по формуле (8), приближенными выражениями

$$\frac{1}{\gamma} e^{-\frac{1}{2}x^2} \quad (12)$$

(ошибка этого приближения является величиной порядка $1/\sqrt{n}$). Затем сумму по точкам решетки заменяют интегралом от (12) по области B , деленным на объем элементарной клетки. При этом ошибка возникает, главным образом, вблизи границы B . Порядок величины этой ошибки снова равен $1/\sqrt{n}$. Таким образом, в качестве приближенного значения для вероятности $\mathbf{P}(\chi^2 < u)$ получают интеграл

$$(2\pi)^{-\frac{m-1}{2}} \int \dots \int e^{-\frac{1}{2}x^2} dV_{m-1}, \quad (13)$$

где интегрирование производится по области $\chi^2 < u$. Как мы уже видели в § 27, вычисление этого интеграла приводит к известному распределению χ^2 с $m - 1$ степенями свободы. Более подробный вывод можно найти в работе Пирсона¹.

До сих пор в формуле (7) мы пренебрегали поправочными членами порядка $1/\sqrt{n}$. Но если учесть эти члены, то окажется, что они никак не влияют на результат. Действительно,

$$e^{-\frac{1}{2}x^2} \left(-\frac{1}{2} \sum \frac{z}{np} + \frac{1}{6} \sum \frac{z^3}{n^2 p^2} \right) \quad (14)$$

является нечетной функцией от z_1, \dots, z_m : при замене всех y_i на $-y_i$ эта функция лишь меняет свой знак. Интеграл от нечетной функции по симметричной области $\chi^2 < u$ равен нулю. Следовательно, асимметрия распределения точек \bar{X} почти не оказывает влияния на результат. Единственной ошибкой, которая при определенных условиях может стать величиной в точности порядка $1/\sqrt{n}$, является ошибка, возникающая вследствие замены суммирования интегрированием вблизи границы области $\chi^2 < u$.

б) Исследуем теперь распределение оценки наибольшего правдоподобия \tilde{v} . Сначала вместо \tilde{v} мы рассмотрим оценку \hat{v}' , полученную методом минимума χ_0^2 , где

$$\chi_0^2 = \sum \frac{[x_i - np_i(\hat{v}')]^2}{np_i}. \quad (15)$$

¹ Pearson K., On the criterion that a given system of deviations ... is such that it can be reasonably supposed to have arisen from random sampling, Philos. Mag., 50 (1900), 157. См. также Крамер Г., Математические методы статистики, ИЛ, М., 1948, гл. XXX.

В отличие от § 48, p_i в знаменателях формулы (15) являются истинными значениями соответствующих вероятностей. Хотя на практике p_i бывают неизвестны, однако при чисто теоретическом изучении функции распределения такая замена вполне уместна. Если $\tilde{\nu} = \tilde{\nu}'$ является для χ^2_0 точкой максимума, то, согласно результатам § 48, $\tilde{\nu}'$ и $\tilde{\nu}$ отличаются друг от друга величиной порядка $1/n$.

Оценка $\tilde{\nu}'$ получается по методу наименьших квадратов: из точки X опускается перпендикуляр на линейное подпространство G , определяемое параметрическим представлением $p_i(\tilde{\nu})$. Если основанием этого перпендикуляра является точка X' , то $\tilde{\nu}'$ — значение параметра $\tilde{\nu}$, соответствующее точке X' .

Уравнения для вычисления $\tilde{\nu}'$ в общем случае r неизвестных параметров $\tilde{\nu}_1, \dots, \tilde{\nu}_r$ были указаны в § 48. Если начало отсчета в пространстве параметров выбрать таким образом, чтобы $p_i(0)$ равнялись истинным вероятностям p_i , то, согласно (8) § 48, уравнения для $\tilde{\nu}'_1, \dots, \tilde{\nu}'_r$ будут иметь вид

$$\sum_{\beta} h_{\alpha\beta} \tilde{\nu}'_{\beta} = \sum_i \frac{(x_i - np_i) q_{i\alpha}}{p_i}, \quad (16)$$

где

$$h_{\alpha\beta} = \sum_i \frac{n q_{i\alpha} q_{i\beta}}{p_i}. \quad (17)$$

Из уравнений (16), во-первых, видно, что $\tilde{\nu}'_{\beta}$ являются линейными функциями от наблюдаемых частот $h_i = x_i/n$. Во-вторых, так как математические ожидания разностей $x_i - np_i$ равны нулю, то математические ожидания оценок $\tilde{\nu}'_{\beta}$ также равны нулю. Следовательно, оценки $\tilde{\nu}'_{\beta}$ являются несмещенными.

Далее, так как $x_i - np_i$ с вероятностью, сколь угодно близкой к единице, представляют собой величины порядка \sqrt{n} , а $h_{\alpha\beta}$ имеют порядок n , то $\tilde{\nu}'_{\alpha}$ с вероятностью, сколь угодно близкой к единице, являются величинами порядка $1/\sqrt{n}$.

Наконец, в силу того, что x_i распределены приближенно нормально, следует ожидать, что $\tilde{\nu}'_{\alpha}$ распределены также приближенно нормально. При доказательстве мы снова ограничимся случаем одного параметра $\tilde{\nu}$ и постараемся вывести асимптотическую формулу для вероятности события $\tilde{\nu}' < t/\sqrt{n}$ при $n \rightarrow \infty$.

Оценка $\tilde{\nu}'$ представляет собой значение параметра, соответствующее точке X' , которая является проекцией наблюдаемой точки X на прямую G (см. рис. 24). Пусть $P(\tilde{\nu})$ — точка с координатами $p_i(\tilde{\nu})$, принадлежащая прямой G . Рассмотрим точки $P_t = P(t/\sqrt{n})$ и $X' = P(\tilde{\nu}')$. Для всех $\tilde{\nu}'$, удовлетворяющих неравенству $\tilde{\nu}' < t/\sqrt{n}$, соответствующие точки X' расположены на прямой G по одну сторону от точки P_t . Поэтому $\tilde{\nu}' < t/\sqrt{n}$ тогда и только

тогда, когда наблюдаемые точки X расположены в полупространстве, ограниченном гиперплоскостью H_t , проходящей через точку P_t перпендикулярно к прямой G .

Выкладки станут проще, если предварительно с помощью линейного преобразования ввести новые прямоугольные координаты y_1, \dots, y_m таким образом, чтобы ось Oy_1 совпала с прямой G . Точка P_t тогда будет иметь координаты $y_1 = at, y_2 = \dots = y_m = 0$, а полупространство, ограниченное плоскостью H_t , будет определяться неравенством $y_1 < at$.

Искомая вероятность представляет собой сумму вероятностей $P(X)$ по всем точкам решетки, расположенным в полупространстве. $P(X)$ можно снова приблизить выражениями (8) или (12), а сумму заменить интегралом. Таким образом, получим интеграл вида

$$\begin{aligned} & (2\pi)^{-\frac{m-1}{2}} \int \dots \int e^{-\frac{1}{2}x^2} dV_{m-1} = \\ & = (2\pi)^{-\frac{m-1}{2}} \int \dots \int e^{-\frac{1}{2}(y_1^2 + \dots + y_m^2)} dV_{m-1}, \end{aligned} \quad (18)$$

где область интегрирования B представляет собой часть гиперплоскости $\sum x_i = n$, расположенную в полупространстве $y_1 < at$. При этом если R выбрано достаточно большим, то безразлично, производится ли интегрирование по всей полуплоскости или только по части полуплоскости, лежащей внутри шара $\chi^2 < R^2$. Если в (18) выполнить интегрирование по y_2, \dots, y_{m-1} , то вместо (18) останется интеграл

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{at} e^{-\frac{1}{2}y_1^2} dy_1 = \Phi(at). \quad (19)$$

Следовательно, распределение оценки \hat{v}' является асимптотически нормальным.

Для того чтобы нормальное распределение было полностью определено, нам нужно еще найти множитель a . Как видно из рис. 24, величина координаты $y_1 = at$ равна расстоянию PP_t в метрике (15), т. е.

$$a^2 = \frac{\chi^2}{i^2} = \frac{n}{i^2} \sum_i \frac{\left[p_i - p_i \left(\frac{t}{\sqrt{n}} \right) \right]^2}{p_i} = \sum_i \frac{q_i^2}{p_i} = \frac{h_{11}}{n}. \quad (20)$$

Асимптотическая формула для вероятности события $\hat{v}' < t/\sqrt{n}$ задается интегралом (19). Если положим $t/\sqrt{n} = t'$ и $a\sqrt{n} = a'$, то, в силу равенства $at = a't'$, асимптотическое распределение оценки \hat{v}' будет задаваться функцией

$$\Phi(at) = \Phi(a't') \sim P(\hat{v}' < t').$$

где

$$a' = a\sqrt{n} = \sqrt{h_{11}}. \quad (21)$$

Следовательно, асимптотическое квадратичное отклонение для \hat{v}' равно

$$\sigma_{\hat{v}'} = \frac{1}{a'} = \sqrt{\frac{1}{h_{11}}} = \sqrt{h^{11}}. \quad (22)$$

Точно так же, в случае r параметров $\hat{v}_1, \dots, \hat{v}_r$, квадратичное отклонение оценки \hat{v}_a равно

$$\sigma_a = \sqrt{h^{aa}}, \quad (23)$$

где (h^{ab}) — матрица, обратная матрице (h_{ab}) . Формулы (22) и (23) вполне аналогичны формулам § 31. Для доказательства справедливости формул (23) можно, например, с помощью обратной матрицы найти решение \hat{v}' системы (16), возвести \hat{v}'_a в квадраты и вычислить соответствующие средние значения. Так как $(\hat{v}'_a)^2$ представляют собой линейные комбинации случайных величин вида x_i^2 и $x_i x_j$, то при вычислении $E(\hat{v}'_a)^2$ можно будет воспользоваться ранее найденными точными формулами (1) и (2) § 46. При этом окажется, что (22) и (23) являются не только асимптотическими формулами при $n \rightarrow \infty$, но и точными. Попутно заметим, что, согласно (17), h_{ab} пропорциональны n , поэтому элементы обратной матрицы h^{ab} пропорциональны $1/n$. Отсюда следует, что квадратичные отклонения, вычисленные по формулам (23), являются величинами вида c/\sqrt{n} .

После того, как асимптотическое распределение \hat{v}' найдено, переход от \hat{v}' к \tilde{v} не представляет труда.

Если положим

$$\tilde{v} = \hat{v}' + \eta, \quad (24)$$

то, согласно результатам § 48, η с вероятностью, сколь угодно близкой к единице, будет являться величиной порядка $1/n$. Умножив правую и левую части (24) на \sqrt{n} , получаем

$$\tilde{v}\sqrt{n} = \hat{v}'\sqrt{n} + \eta\sqrt{n}. \quad (25)$$

Первое слагаемое представляет собой асимптотически нормальную случайную величину с нулевым средним значением и дисперсией, не зависящей от n , а второе слагаемое по вероятности стремится к нулю при $n \rightarrow \infty$. Следовательно, к сумме (25) применима элементарная предельная теорема из § 24 Ж. Таким образом, оценка наибольшего правдоподобия \tilde{v} распределена асимптотически нормально с тем же средним значением и с той же дисперсией, что и \hat{v}' .

Это же самое заключение справедливо для всех тех оценок, которые с вероятностью, сколь угодно близкой к единице, отличаются от ϑ' или $\tilde{\vartheta}$ величинами порядка $1/n$. Поэтому, например, оценка по методу минимума χ_x^2 и все оценки по методу минимума χ_0^2 являются асимптотически нормальными случайными величинами с тем же средним значением и с той же дисперсией, что и $\tilde{\vartheta}$.

§ 50. Асимптотическая эффективность

Мы снова ограничимся случаем одного параметра ϑ . Как мы видели, дисперсия асимптотического распределения оценки $\tilde{\vartheta}$ равна

$$\sigma^2 = \frac{1}{h_{11}} = h^{11} = \frac{c^2}{n}. \quad (1)$$

Это, однако, не означает, что при $n \rightarrow \infty$ дисперсия $\tilde{\vartheta}$ обязательно стремится к нулю. Как показывает пример 33 (§ 46), может даже случиться, что при всех n дисперсия $\tilde{\vartheta}$ будет бесконечной, следовательно, предел дисперсии будет также бесконечным. Формула (1) представляет собой не асимптотическую формулу для точной дисперсии, а формулу для *асимптотической дисперсии* оценки $\tilde{\vartheta}$ в смысле § 45 Б. Кроме того, $\tilde{\vartheta}$ является *асимптотически несмещенной оценкой* в смысле § 45 Б.

Сравним теперь асимптотическую дисперсию (1) с той наименьшей дисперсией, которая вообще возможна для несмещенных оценок, согласно неравенству Фреше. С этой целью предварительно вычислим «информацию» $I(\vartheta)$, которая, согласно § 37, определяется формулой

$$I(\vartheta) = \mathcal{E}[L'(x | \vartheta)]^2, \quad (2)$$

где

$$L(x | \vartheta) = \sum x_i \ln p_i(\vartheta).$$

Таким образом, если производные от p_i по ϑ обозначим q_i (ранее, в случае нескольких параметров, частные производные от p_i обозначались q_{ia}), то получим

$$L'(x | \vartheta) = \sum \frac{q_i}{p_i} x_i$$

и

$$I(\vartheta) = \mathcal{E} \left(\sum \frac{q_i}{p_i} x_i \right)^2 = \sum_i \sum_k \frac{q_i q_k}{p_i p_k} \mathcal{E}(x_i x_k). \quad (3)$$

Математические ожидания $x_i x_k$ были вычислены в § 46:

$$\mathcal{E}(x_i x_k) = n(n-1) p_i p_k, \quad \text{если } i \neq k, \quad (4)$$

$$\mathcal{E} x_i^2 = n(n-1) p_i^2 + n p_i. \quad (5)$$

Поэтому

$$I(\vartheta) = \sum_i \sum_k n(n-1) q_i q_k + \sum_i n \frac{q_i^2}{p_i}. \quad (6)$$

В формуле (6) двойная сумма равна нулю, так как $(\sum q_i)^2 = 0$, а второй член совпадает с h_{11} . Следовательно, как и в теории наименьших квадратов,

$$I(\vartheta) = h_{11}. \quad (7)$$

Согласно неравенству Фреше для дисперсий несмещенных оценок,

$$\sigma^2 \geq \frac{1}{I(\vartheta)} = \frac{1}{h_{11}} = h^{11} = \frac{c^2}{n}. \quad (8)$$

Оценка $\tilde{\vartheta}$ является асимптотически несмещенной и, в силу (1), ее асимптотическая дисперсия равна c^2/n . Следовательно, оценка $\tilde{\vartheta}$, в указанном смысле, асимптотически эффективна.

Это не означает, что среди всех асимптотически несмещенных оценок $\tilde{\vartheta}$ обладает наименьшей асимптотической дисперсией. Можно (подобно тому, как это делалось в § 45 Г) построить пример асимптотически несмещенной оценки, асимптотическая дисперсия которой при некоторых значениях ϑ будет меньше, чем $1/I(\vartheta)$. Минимальность дисперсии оценки $\tilde{\vartheta}$ можно доказать лишь в классе асимптотически несмещенных оценок, удовлетворяющих условиям регулярности. С этим доказательством мы должны теперь познакомиться ближе.

Если левую часть уравнения правдоподобия (14) § 48 разделить на n , то, отбрасывая у $q_{i\alpha}$ индекс α , получим

$$\sum_i \frac{[h_i - p_i(\tilde{\vartheta})] q_i}{p_i(\tilde{\vartheta})} = 0. \quad (9)$$

Это уравнение не содержит x_i и n в явном виде, а зависит лишь от частот h_i . Следовательно, оценка наибольшего правдоподобия $\tilde{\vartheta}$ является функцией одних только h_i . Более того, если точка с координатами h_i расположена вне области маловероятных отклонений от истинных значений p_1, \dots, p_m и h_i не слишком близки к нулю, то $\tilde{\vartheta}$ заведомо будет дифференцируемой¹ функцией частот h_i .

¹ Частные производные от $\tilde{\vartheta}$ по h_k имеют вид

$$\frac{\partial \tilde{\vartheta}}{\partial h_k} = \frac{q_k}{p_k(\tilde{\vartheta})} \left[\sum_i \frac{q_i^2}{p_i^2(\tilde{\vartheta})} h_i \right]^{-1}.$$

Сформулированные автором условия нужны главным образом для того, чтобы $p_k(\tilde{\vartheta})$ не обращались в нуль. — Прим. перев.

В качестве оценок, конкурирующих с \tilde{v} , мы рассмотрим лишь такие оценки T , которые являются дифференцируемыми функциями от h_i . Эти оценки мы будем называть регулярными. Таким образом, пусть T — асимптотически несмещенная, регулярная оценка. Мы хотим сравнить асимптотические дисперсии оценок T и \tilde{v} .

Как нам уже известно, точка (h_1, \dots, h_m) с большой вероятностью расположена в некоторой окрестности точки (p_1, \dots, p_m) , причем диаметр окрестности при $n \rightarrow \infty$ является бесконечно малой величиной. В такой окрестности каждую дифференцируемую функцию можно аппроксимировать линейной функцией. Полезным линейным приближением для \tilde{v} является ранее определенная оценка ϑ' ; пусть

$$T' = \sum c_i h_i \quad (10)$$

— линейная аппроксимация для T .

Асимптотическое распределение дифференцируемой функции T совпадает с асимптотическим распределением линейной функции T' так же, как совпадают асимптотические распределения \tilde{v} и ϑ' . Доказательство проводится тем же методом, что и раньше, причем асимптотическая формула для функции распределения T' выводится так же, как и для ϑ' . Мы снова должны просуммировать вероятности $P(X)$ по всем точкам некоторого полупространства. Суммирование заменяется интегрированием, а вероятность $P(X)$ — плотностью нормального распределения

$$C e^{-\frac{1}{2} \chi^2}, \quad (11)$$

где $\chi^2 = y_1^2 + \dots + y_m^2$. Все точки решетки, по которым нужно производить суммирование лежат в гиперплоскости $\sum h_i = 1$, следовательно, интегрирование распространяется лишь на эту гиперплоскость. Соответствующим ортогональным преобразованием величин y_i можно добиться, чтобы эта гиперплоскость имела уравнение $y_m = 0$, тогда плотность вероятности будет задаваться формулой

$$f(y_1, \dots, y_{m-1}) = C e^{-\frac{1}{2}(y_1^2 + \dots + y_{m-1}^2)}. \quad (12)$$

Эта формула справедлива тогда, когда в пространстве h_i в качестве начала прямоугольных координат y_1, \dots, y_m выбрана точка $(p_1(\vartheta), \dots, p_m(\vartheta))$. Если же началом координат является постоянная точка, не зависящая от ϑ , то (12) нужно заменить формулой

$$f(y_1, \dots, y_{m-1}) = C e^{-\frac{1}{2}[(y_1 - \hat{y}_1)^2 + \dots + (y_{m-1} - \hat{y}_{m-1})^2]}, \quad (13)$$

где \hat{y} равны математическим ожиданиям y .

Отсюда следует, что T' обладает асимптотически нормальным распределением. Асимптотическое среднее значение и асимптотическая дисперсия оценки T' равны, по определению, среднему значению и дисперсии асимптотического нормального распределения, т. е. равны среднему значению и дисперсии линейной функции

$$T' = \sum_{i=1}^m c_i h_i = b_0 + \sum_{k=1}^{m-1} b'_k y_k, \quad (14)$$

вычисленным в предположении, что y_1, \dots, y_{m-1} имеют плотность вероятности (13). Так как коэффициенты c_i не зависят от n , а среднее значение и дисперсию T' через эти коэффициенты можно выразить точно, то между асимптотической формулой для дисперсии и асимптотической дисперсией не будет никакого различия. Это же самое справедливо и для ϑ' .

Таким образом, мы имеем ту же самую ситуацию, что и в теории наименьших квадратов. y_1, \dots, y_{m-1} являются независимыми нормально распределенными случайными величинами с единичными дисперсиями и математическими ожиданиями \hat{y}_i , представляющими собой линейные функции одного параметра ϑ . Оценка ϑ' по методу наименьших квадратов является несмещенной и имеет минимальную дисперсию. Оценка T' также не имеет смещения, следовательно, ее дисперсия не меньше дисперсии оценки ϑ' . Равенство дисперсий будет осуществляться лишь тогда, когда соответствующие коэффициенты линейных функций T' и ϑ' равны. Следовательно:

Среди всех регулярных асимптотически несмещенных оценок T оценка наибольшего правдоподобия $\tilde{\vartheta}$ имеет наименьшую асимптотическую дисперсию. Если T и $\tilde{\vartheta}$ обладают равными асимптотическими дисперсиями и если в окрестности точки с координатами $h_i = p_i(\vartheta)$ обе эти функции разлагаются в ряд по степеням $h_i - p_i$, то по крайней мере линейные члены этих рядов совпадают; следовательно, T с вероятностью, сколь угодно близкой к единице, отличается от $\tilde{\vartheta}$ величиной высшего, чем $1/\sqrt{n}$, порядка малости.

Для того чтобы эту теорему можно было сформулировать короче, мы введем следующие определения:

Асимптотически несмещенная регулярная оценка T называется *асимптотически эффективной*, если среди всех оценок с теми же свойствами она обладает наименьшей асимптотической дисперсией. Две оценки, T_1 и T_2 , называются *асимптотически эквивалентными*, если их разность $D = T_1 - T_2$ с вероятностью, сколь угодно близкой к единице, является величиной высшего, чем $1/\sqrt{n}$, порядка малости, т. е. если $D\sqrt{n}$ стремится, по вероятности, к нулю.

Теперь указанную теорему можно сформулировать так:

Оценка наибольшего правдоподобия $\hat{\nu}$ является асимптотически эффективной, и каждая регулярная асимптотически эффективная оценка ей асимптотически эквивалентна. Асимптотическая дисперсия такой оценки: равна $1/I(\hat{\nu})$ — обратной величине информации.

При более общих предположениях и, в частности, без предположения линейности функций $p_i(\nu)$ эта теорема была доказана Нейманом¹.

Пример 34. Согласно общепринятой гипотезе Ф. Бернштейна², наличие у людей четырех групп крови: *O* (I группа), *A* (II группа), *B* (III группа) и *AB* (IV группа), вызывается тремя генами *A*, *B* и *O*, причем *A* и *B* доминируют над *O*. Если индивидуум имеет генную пару *OO*, то его кровь относится к группе *O*. Генные пары *AO* и *AA* приводят к группе *A*, а генные пары *BO* и *BB* — к группе *B*. Наконец, генная пара *AB* приводит к группе *AB*. Пусть

$$h_1 = \frac{x_1}{n}, \dots, h_4 = \frac{x_4}{n}$$

— известные частоты групп крови в выборке, состоящей из n индивидуумов, и пусть p, q и r ($p + q + r = 1$) — частоты генов *A, B* и *O* в крови населения. Требуется найти асимптотически эффективные оценки³ для p, q и r .

Мы предположим, что население хорошо перемешано, т. е. что оно не распадается на почти замкнутые группы с различными распределениями частот генов. При этом предположении вероятность образования генной пары *OO* равна r^2 , точно так же вероятность образования генной комбинации *AO* (или *OA*) равна pr и т. д. Следовательно, вероятности того, что выбранный наугад представитель населения будет иметь группу крови *O, A, B* или *AB*, равны соответственно:

$$\left. \begin{aligned} p_1 &= r^2, \\ p_2 &= 2pr + p^2 = (p+r)^2 - r^2, \\ p_3 &= 2qr + q^2 = (q+r)^2 - r^2, \\ p_4 &= 2pq. \end{aligned} \right\} \quad (15)$$

Эти уравнения можно разрешить относительно p и q :

$$\left. \begin{aligned} p &= 1 - (q+r) = 1 - \sqrt{p_1 + p_3}, \\ q &= 1 - (p+r) = 1 - \sqrt{p_1 + p_2}. \end{aligned} \right\} \quad (16)$$

¹ Neyman J., Contribution to the theory of the χ^2 -test, Berkeley Sympos. on Math. Stat., 1949, 239.

² Bernstein F., Z. f. induktive Abstammungs- und Vererbungslehre, 37 (1925), 236.

³ Общее определение совместно асимптотически эффективных оценок для нескольких неизвестных параметров см. в книге Крамер Г., Математические методы статистики, ИЛ, М., 1948. — *Прим. перев.*

С целью получения для p и q предварительных оценок можно в (16) вероятности p_1 , p_2 и p_3 заменить наблюдаемыми частотами h_1 , h_2 и h_3 . Таким образом, находим

$$\left. \begin{aligned} p_0 &= 1 - \sqrt{h_1 + h_3}, \\ q_0 &= 1 - \sqrt{h_1 + h_2}. \end{aligned} \right\} \quad (17)$$

Убедиться в том, что эти оценки не являются асимптотически эффективными, можно, например, так. Пусть h_1 , h_2 , h_3 — координаты в пространстве наблюдений, тогда частота h_4 будет являться функцией этих координат: $h_4 = 1 - h_1 - h_2 - h_3$. Все точки H с координатами h_1 , h_2 , h_3 , для которых оценки (17) остаются постоянными, расположены на прямой

$$\left. \begin{aligned} h_1 + h_3 &= (1 - p_0)^2, \\ h_1 + h_2 &= (1 - q_0)^2. \end{aligned} \right\} \quad (18)$$

Эта прямая пересекает поверхность, заданную параметрическими уравнениями (15), в точке P_0 с координатами $p_i(0)$, которым соответствуют значения параметров $p = p_0$, $q = q_0$. Если бы оценки (17) были асимптотически эффективными, то прямая (18) была бы перпендикулярна к этой поверхности (или, по крайней мере при больших n , приближенно перпендикулярна к ней) в смысле метрики, определяемой квадратичной формой (§ 48)

$$\chi_0^2 = \sum \frac{[h_i - p_i(0)]^2}{p_i(0)}. \quad (19)$$

Соответствующее условие ортогональности имеет вид

$$\sum \frac{u_i v_i}{p_i(0)} = 0, \quad (20)$$

где $u = (1, -1, -1, 1)$ — направляющий вектор прямой (18) и v — произвольный вектор, лежащий в касательной плоскости к поверхности (15). Два таких вектора, v и v' , получаются дифференцированием (15) по p и q . Если координаты этих векторов подставим в левую часть (20), то убедимся, что условие ортогональности не выполняется даже приближенно.

Асимптотически эффективные оценки можно получить с помощью отыскания максимума логарифмической функции правдоподобия

$$L(x|p, q) = x_1 \ln r^2 + x_2 \ln (2pr + p^2) + x_3 \ln (2qr + q^2) + x_4 \ln (2pq). \quad (21)$$

Дифференцирование L по p и q (при этом следует положить $r = 1 - p - q$) приводит к уравнениям

$$\frac{x_2 + x_4}{p} + \frac{x_2}{2r + p} = \frac{x_3 + x_4}{q} + \frac{x_3}{2r + q} = \frac{2x_1}{r} + \frac{2x_2}{2r + p} + \frac{2x_3}{2r + q}. \quad (22)$$

Для отыскания решения этих уравнений можно, например, p и q заменить новыми неизвестными u и v по формулам

$$\left. \begin{aligned} p &= p_0 + u, \\ q &= q_0 + v, \end{aligned} \right\} \quad (23)$$

и затем дроби в уравнениях (19) разложить в ряды по степеням u и v , удерживая лишь линейные члены. В результате возникнут два линейных уравнения, которые нужно будет разрешить относительно u и v .

Как показано в § 48, для отыскания асимптотически эффективных оценок вместо метода наибольшего правдоподобия можно также воспользоваться методами минимума χ_0^2 и χ_k^2 .

§ 51. Критерий χ^2

В § 49 мы вычислили асимптотическую формулу для функции распределения случайной величины

$$\chi^2 = \sum \frac{(x - np)^2}{np}, \quad (1)$$

в предположении, что p_i равны истинным значениям вероятностей:

$$p_i^* = p_i(\vartheta^*). \quad (2)$$

Однако на практике истинные p^* бывают неизвестны, и поэтому их заменяют оценками

$$\tilde{p}_i = p_i(\tilde{\vartheta}). \quad (3)$$

Выражение

$$\tilde{\chi}^2 = \sum \frac{(x - n\tilde{p})^2}{n\tilde{p}}, \quad (4)$$

построенное с помощью таких оценок \tilde{p}_i , оказывается, вообще говоря, меньшим, чем χ^2 , и имеет другую функцию распределения. Действительно, как мы увидим, случайная величина $\tilde{\chi}^2$ асимптотически имеет распределение χ^2 с $m - 1 - r$ степенями свободы, где r — число оцениваемых параметров $\vartheta_1, \dots, \vartheta_r$, в то время как, согласно § 49, случайная величина χ^2 , определенная формулой (1), асимптотически имеет распределение χ^2 с $m - 1$ степенями свободы.

Числители и знаменатели каждого слагаемого суммы (4) являются величинами порядка n . В знаменателях \tilde{p} можно заменить истинными значениями p^* , отличающимися от \tilde{p} величинами порядка $1/\sqrt{n}$: от этого функция распределения $\tilde{\chi}^2$ изменится лишь на бесконечно малую величину. В результате получим видоизмененное выражение

$$\chi_1^2 = \sum \frac{(x - n\tilde{p})^2}{np^*}, \quad (5)$$

функцию распределения которого можно определить несколько легче, чем функцию распределения $\tilde{\chi}^2$.

Теперь мы воспользуемся теорией, изложенной в § 48, причем в качестве $p^{(0)}$ выберем p^* и рассмотрим координаты

$$x'_i = n p_i(\vartheta') \quad (6)$$

гочки минимума квадратичной формы

$$\chi_0^2 = \sum \frac{(x - x')^2}{np^*}. \quad (7)$$

Как доказано в § 48, оценки

$$p'_i = p_i(\tilde{v}') = \frac{x_i}{n},$$

полученные методом минимума χ_0^2 , отличаются от соответствующих оценок наибольшего правдоподобия \tilde{p} величинами порядка $1/n$. Поэтому $n\tilde{p}$ в правой части (5) можно заменить на x' , тогда χ_1^2 перейдет в χ_0^2 и асимптотическое распределение не изменится.

Следовательно, нам осталось лишь найти асимптотическое распределение χ_0^2 . Вероятность события $\chi_0^2 < u$ снова равна сумме вероятностей $\mathbf{P}(X)$ по всем точкам X , принадлежащим области $\chi_0^2 < u$. Как и в § 49, эту сумму можно заменить интегралом. В результате получится искомая асимптотическая функция распределения

$$F(u) = (2\pi)^{-\frac{m-1}{2}} \int \dots \int e^{-\frac{1}{2}x^2} dV_{m-1}, \quad (8)$$

где интегрирование производится по области $\chi_0^2 < u$.

Введем теперь прямоугольные координаты y_1, \dots, y_m таким образом, чтобы выполнялось равенство, аналогичное (11) § 49:

$$\chi^2 = y_1^2 + \dots + y_m^2.$$

Ортогональным преобразованием этих координат можно добиться, чтобы гиперплоскость $\sum x_i = n$ в новых координатах имела уравнение $y_m = 0$; тогда переменными интегрирования будут y_1, \dots, y_{m-1} , и мы получаем

$$F(u) = (2\pi)^{-\frac{m-1}{2}} \int \dots \int e^{-\frac{1}{2}(y_1^2 + \dots + y_{m-1}^2)} dy_1 \dots dy_{m-1}. \quad (9)$$

Теперь мы можем еще раз ортогонально преобразовать координаты таким образом, чтобы новые оси Oy_1, \dots, Oy_r лежали в линейном пространстве G , определяемом параметрическими уравнениями $p_i = p_i(\tilde{v})$; остальные оси будут тогда перпендикулярны пространству G . Компоненты точки минимума формы (7) в этих новых координатах можно вычислить особенно легко. Действительно, точка X' принадлежит G , следовательно, из ее координат отличными от нуля могут быть лишь y'_1, \dots, y'_r ; остальные координаты $y'_{r+1}, \dots, y'_{m-1}$ равны нулю. В новых координатах форма χ_0^2 имеет вид

$$\chi_0^2 = (y_1 - y'_1)^2 + \dots + (y_r - y'_r)^2 + y_{r+1}^2 + \dots + y_{m-1}^2. \quad (10)$$

Она достигает своего минимума, когда все разности $y_1 - y'_1, \dots,$

$y_r - y'_r$ становятся равными нулю. В этом случае первые r слагаемых суммы (10) исчезают, и мы имеем

$$\chi_0^2 = y_{r+1}^2 + \dots + y_{m-1}^2. \quad (11)$$

Следовательно, в (9) область интегрирования является цилиндрической, так как левая часть неравенства $\chi_0^2 < u$ зависит лишь от y_{r+1}, \dots, y_{m-1} . Если в (9) произвести интегрирование по y_1, \dots, y_r , то получим

$$F(u) = (2\pi)^{-\frac{m-r-1}{2}} \int \dots \int e^{-\frac{1}{2}(y_{r+1}^2 + \dots + y_{m-1}^2)} dy_{r+1} \dots dy_{m-1}, \quad (12)$$

где область интегрирования задается неравенством

$$y_{r+1}^2 + \dots + y_{m-1}^2 < u. \quad (13)$$

Таким образом, $F(u)$ действительно является функцией распределения χ^2 с $m - 1 - r$ степенями свободы.

Общий критерий χ^2 можно теперь сформулировать так:

Если выражение χ^2 превосходит границу, указанную в табл. 6 для $m - 1 - r$ степеней свободы, то гипотезу, согласно которой истинные вероятности p^ имеют параметрическое представление $p(\tilde{v})$, следует отвергнуть.*

В символе $\tilde{\chi}^2$ волна введена только для того, чтобы ясно отличать $\tilde{\chi}^2$ от истинного χ^2 . В приложениях этим различием, как правило, пренебрегают.

Очень важным для приложений является вопрос, можно ли в правой части (4) вместо $\tilde{p} = p(\tilde{v})$ воспользоваться какой-либо другой оценкой $p(T)$?

Ответ гласит:

Если T — асимптотически эффективная оценка (§ 50, конец) и поэтому $|T - \tilde{v}|$ с вероятностью, сколь угодно близкой к единице, является величиной высшего, чем $1/\sqrt{n}$, порядка малости, то в правой части (4) \tilde{p} можно заменить на $p(T)$ и затем применить критерий χ^2 .

Доказательство очевидно. Если к \tilde{p} в числителях (4) прибавить некоторые дополнительные члены высшего, чем $1/\sqrt{n}$, порядка малости, то χ^2 изменится лишь на бесконечно малую величину, поэтому асимптотическое распределение останется неизменным.

Но если \tilde{p} заменить другими оценками, отличающимися от \tilde{p} величинами в точности порядка $1/\sqrt{n}$, то (4) может оказаться значительно больше χ^2 . Отсюда мы видим, как важно пользоваться лишь асимптотически эффективными оценками.

Пример 35. Если мы хотим проверить изложенную в примере 34 (§ 50) гипотезу Ф. Бернштейна о группах крови O , A , B и AB , то сначала можно найти оценки наибольшего правдоподобия для параметров p , q и $r = 1 - p - q$ и затем с помощью функций от этих оценок

$$\left. \begin{aligned} p_1 &= r^2, \\ p_2 &= 2pr + p^2, \\ p_3 &= 2qr + q^2, \\ p_4 &= 2pq \end{aligned} \right\} \quad (14)$$

вычислить величину χ^2 :

$$\chi^2 = \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_2 - np_2)^2}{np_2} + \frac{(x_3 - np_3)^2}{np_3} + \frac{(x_4 - np_4)^2}{np_4}. \quad (15)$$

Так как по результатам наблюдений оцениваются два параметра p и q , то число степеней свободы равно

$$f = 4 - 1 - 2 = 1.$$

Оценки параметров p и q очень сложны. Существует более простой способ проверки указанной гипотезы, который практически дает такой же результат, что и общий критерий χ^2 .

Прежде всего заметим, что уравнения (14) определяют в пространстве наблюдений некоторую поверхность F и что \tilde{P} представляет собой такую точку этой поверхности, которая находится на кратчайшем расстоянии от наблюдаемой точки H , в смысле метрики, определяемой выражением χ^2 . Следовательно, χ^2 — квадрат расстояния от H до поверхности F .

Вместо параметрического представления (14) поверхность F можно задать с помощью ее уравнения, которое имеет вид¹

$$\sqrt{p_1 + p_2} + \sqrt{p_1 + p_3} - \sqrt{p_1} - 1 = 0. \quad (16)$$

Если точка H не лежит на поверхности F , то расстояние от H до поверхности пропорционально величине

$$D = \sqrt{h_1 + h_2} + \sqrt{h_1 + h_3} - \sqrt{h_1} - 1. \quad (17)$$

Функцию D в окрестности «истинной точки» P , принадлежащей поверхности F , можно аппроксимировать линейной функцией координат h_1 , h_2 и h_3 .

Положив $h_i = p_i + u_i$, после небольших вычислений найдем, что

$$D \sim \frac{1}{2} \frac{u_1 + u_2}{\sqrt{p_1 + p_2}} + \frac{1}{2} \frac{u_1 + u_3}{\sqrt{p_1 + p_3}} - \frac{1}{2} \frac{u_1}{\sqrt{p_1}} = a_1 u_1 + a_2 u_2 + a_3 u_3. \quad (18)$$

Случайные величины $u_i = h_i - p_i$ распределены приближенно нормально с нулевым средним значением и дисперсиями

$$\mathcal{E} u_i^2 = np_i (1 - p_i). \quad (19)$$

Согласно (4) § 50, математические ожидания произведений $u_i u_k$ также известны:

$$\mathcal{E} u_i u_k = -np_i p_k. \quad (20)$$

¹ Bernstein F., Z. ind. Abstammungs- u. Vererbungslehre, 37, S. 245.

Таким образом, сумма $a_1 u_1 + a_2 u_2 + a_3 u_3$ (а вместе с ней и величина D) распределена приближенно нормально с нулевым средним значением и дисперсией

$$\begin{aligned} \sigma^2 = & a_1^2 \mathcal{E} u_1^2 + a_2^2 \mathcal{E} u_2^2 + a_3^2 \mathcal{E} u_3^2 + 2a_1 a_2 \mathcal{E} u_1 u_2 + \\ & + 2a_1 a_3 \mathcal{E} u_1 u_3 + 2a_2 a_3 \mathcal{E} u_2 u_3. \end{aligned} \quad (21)$$

Следовательно, случайная величина

$$\chi_D^2 = \frac{D^2}{\sigma^2} \quad (22)$$

асимптотически распределена, как χ^2 с одной степенью свободы. Выражение χ_D^2 приближенно равно выражению χ^2 , введенному ранее, и может быть использовано для проверки гипотезы Бернштейна.

При вычислении σ^2 вероятности p_i , входящие в (18), (19) и (20), можно заменить их приближенными значениями h_i . При тех больших значениях n , которые, как правило, встречаются при данного рода исследованиях, это приближение не может вызывать опасений, тем более что при небольших изменениях p_i величина χ^2 меняется не очень сильно.

ОБРАБОТКА РЕЗУЛЬТАТОВ БИОЛОГИЧЕСКИХ ИСПЫТАНИЙ

Эта глава посвящена обработке результатов биологических испытаний ядов и других веществ (*bio-assay*¹).

Если подопытные животные подвергаются действию различных доз некоторого яда, то для каждой дозы наблюдается определенная смертность. Кривая, изображающая зависимость смертности от дозы, называется *кривой эффекта*. Мы рассмотрим различные методы оценки кривых эффекта по результатам наблюдений. При этом будет предполагаться известным в основном лишь содержание гл. I и II.

§ 52. Кривая эффекта и логарифмическая кривая эффекта

Существуют препараты, действие которых может быть выявлено лишь тогда, когда некоторое количество подопытных животных подвергают действию различных доз испытываемого вещества и регистрируют, сколько животных реагирует определенным образом, например умирает. Каждой дозе соответствует определенная вероятность реагирования, которая может быть аппроксимирована эмпирической частотой. С возрастанием дозы, как правило, возрастает и вероятность p . Если эту вероятность как функцию дозы изобразить графически, то получится *кривая эффекта* данного препарата.

Очень часто в качестве абсциссы выбирают не саму дозу, а ее логарифм l . Соответствующая кривая в плоскости lOp называется *логарифмической кривой эффекта*. Одно из преимуществ применения логарифмов заключается в том, что кривые эффектов двух препаратов, отличающихся только различными концентрациями действующего вещества, получаются друг из друга параллельным сдвигом. Величина параллельного сдвига, очевидно, равна логарифму отношения концентраций.

Если имеются два препарата с похожими свойствами и нужно сравнить результаты их действия, то по большей части предполагают, что соответствующие логарифмические кривые эффек-

¹ *Bio-assay* (англ.) — биологическое испытание. — *Прим. перев.*

тов отличаются также лишь параллельным сдвигом. Только при этом предположении имеет смысл говорить об отношении эффективностей. Логарифм этого отношения опять-таки равен величине параллельного сдвига.

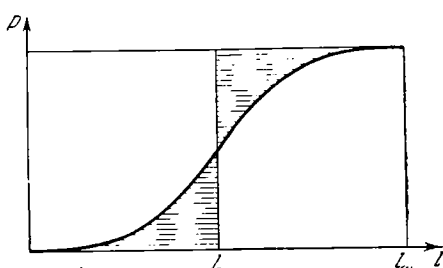


Рис. 25. Логарифмическая кривая эффекта.

Многие кривые эффекта возрастают от 0 до 1. Это означает, что на очень сильные дозы реагируют все подопытные животные. Однако бывают случаи, когда определенная доля подопытных животных нечувствительна к действующему веществу и не реагирует на самые большие дозы. Тогда кривая возрастает только до некоторого предельного значения $p_{\infty} < 1$. Для кривых этого типа обработка наблюдений более трудна.

Здесь мы ограничимся в основном лишь кривыми первого типа, которые возрастают от нуля до единицы.

Логарифмические кривые эффекта, наблюдаемые в природе, очень часто соответствуют графикам функций нормального распределения:

$$p = \Phi\left(\frac{l-L}{\sigma}\right), \quad (1)$$

где l — логарифм дозы, L — логарифм 50%-ной дозы¹, σ — квадратичное отклонение нормального распределения и, как всегда в этой книге,

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{1}{2}t^2} dt. \quad (2)$$

Существуют методы обработки наблюдений, целиком основанные на предположении такой «нормальности» кривой эффекта.

¹ 50%-ной дозой называют такую дозу, для которой $p = 0,5$. — Прим. перев.

Их называют «пробит-методами». Однако только в очень редких случаях это предположение проверено на достаточно обширном экспериментальном материале. По этой причине здесь сначала будут рассматриваться такие способы, которые не зависят от предположения нормальности кривой эффекта и используют лишь более слабое предположение о том, что с ростом дозы вероятность p возрастает от нуля до единицы.

В дальнейшем для простоты мы будем всегда называть вероятность p «смертностью», хотя излагаемые методы применимы не только к смертельным ядам, но и к любым биологическим препаратам. Мы будем также часто говорить «доза l », подразумевая под дозой ее логарифм l . Буквы l или L достаточно ясно указывают на логарифм.

§ 53. Метод площадей Берэнса и Кербера¹

Основой этого метода является понятие *средней смертельной дозы*. Если на рис. 25 провести вертикальную прямую таким образом, чтобы обе заштрихованные области, ограниченные кривой эффекта и указанной прямой, имели одинаковые площади, то абсцисса L точки пересечения этой прямой с осью Ol называется логарифмом средней смертельной дозы. Если кривая эффекта центрально-симметрична, то средняя смертельная доза равна 50%-ной дозе.

Смысл выражения «средняя смертельная доза» можно выяснить следующим образом. Предположим, что для каждого подопытного животного существует определенная наименьшая величина дозы, при которой животное заведомо умирает. Если подопытное животное выбирается случайно, то логарифм такой смертельной дозы является случайной величиной, в смысле § 2. Функция распределения $F(l)$ этой случайной величины l представляет собой вероятность смерти животного от дозы l . Таким образом, значение функции $F(l)$ в точке l совпадает со смертностью p от дозы l (под дозой мы всегда будем понимать логарифм дозы). Поэтому график функции $F(l)$ является логарифмической кривой эффекта. Согласно результатам § 3, среднее значение случайной величины l равно интегралу

$$E l = \int_{-\infty}^{\infty} l dF(l).$$

¹ Kärber E., Archiv exp. Path., 162 (1931), 480; Behrens und Kärber, Archiv exp. Path., 177 (1935), 637.

Если в этом интеграле в качестве новой переменной интегрирования выбрать $p = F(l)$, то получим

$$\mathfrak{E} l = \int_0^1 l(p) dp = L. \quad (1)$$

Следовательно, L является средним значением логарифмов индивидуальных смертельных доз подопытных животных.

Дисперсия σ^2 случайной величины l точно так же определяется интегралом

$$\sigma^2 = \mathfrak{E}(l - L)^2 = \int_{-\infty}^{\infty} (l - L)^2 dF(l) = \int_0^1 [l(p) - L]^2 dp. \quad (2)$$

Для определения дозы L можно воспользоваться также следующим методом. Пусть l_0 — столь малая доза, что соответствующая ей смертность p_0 практически равна нулю, и пусть доза l_ω так велика, что соответствующая смертность p_ω практически равна единице. Если на рис. 25 провести прямые с уравнениями $l = l_0$ и $l = l_\omega$, то площадь прямоугольника с высотой, равной единице, и основанием, лежащим между L и l_ω , будет приближенно¹ равна площади, расположенной между осью Ol и кривой эффекта:

$$l_\omega - L = \int_{l_0}^{l_\omega} p dl.$$

Это равенство можно разрешить относительно L :

$$L = l_\omega - \int_{l_0}^{l_\omega} p dl. \quad (3)$$

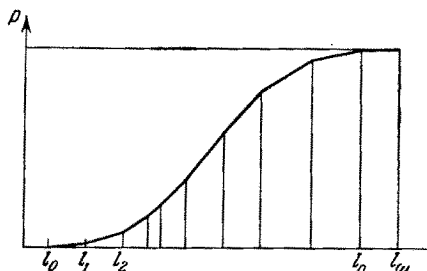


Рис. 26. Трапециoidalное приближение кривой эффекта.

Если интервал от l_0 до l_ω разделим на $(n + 1)$ частичных интервалов, граничными точками которых являются $l_0, l_1, \dots, l_{n+1} = l_\omega$, то интеграл в формуле (3) можно приближенно заменить суммой площадей трапеций (рис. 26); трапеция в любом интервале (l_i, l_{i+1}) получается заменой соответствующего отрезка кривой эффекта прямолинейным отрезком, проходящим через точки $(l_i, p(l_i))$ и $(l_{i+1}, p(l_{i+1}))$:

¹ Если $p(l_0) = 0$ и $p(l_\omega) = 1$, то, по определению средней смертельной дозы, формула (3) будет точной. — Прим. перев.

$$L \sim l_w - \frac{1}{2} [(p_0 + p_1)(l_1 - l_0) + (p_1 + p_2)(l_2 - l_1) + \dots + (p_n + p_w)(l_{n+1} - l_n)]. \quad (4)$$

Если сумму в правой части (4) расположить в порядке возрастания $p_0, p_1, \dots, p_n, p_w$ и учесть, что $p_0 = 0$ и $p_w = 1$, то получим

$$L \sim \frac{1}{2}(l_n + l_{n+1}) - \frac{1}{2}[p_1(l_2 - l_0) + p_2(l_3 - l_1) + \dots + p_n(l_{n+1} - l_{n-1})]. \quad (5)$$

Это выражение является исходным для эмпирического определения L по методу площадей.

Некоторое количество подопытных животных подвергают действию доз l_1, l_2, \dots, l_n и наблюдают частоты смертельных исходов h_1, h_2, \dots, h_n , соответствующие этим дозам. При этом количества животных для всех доз не обязательно должны быть одинаковыми. Нужно лишь, чтобы дозы *возрастали достаточно малыми скачками* [это обеспечит близость интеграла (3) и суммы (5)] и чтобы для *концевых доз l_1 и l_n частоты были близки соответственно нулю и единице* (это, практически, обеспечит уверенность, что для наименьшей дозы l_0 смертность почти равна нулю, а для наибольшей дозы $l_{n+1} = l_w$ смертность почти равна единице).

Следовательно, ряд доз должен достаточно далеко простираться в обе стороны. По большей части требуют, чтобы первая частота h_1 равнялась нулю, а последняя h_n равнялась единице, однако наблюдения можно считать удовлетворительными также и в том случае, когда, например, частоты h_1 и h_2 — обе близки к нулю. Зная весь набор частот и руководствуясь чутьем и опытом, экспериментатор может судить, будет ли p_0 почти нулем, а p_{n+1} — почти единицей.

Приближенное значение для L получают из формулы (5) заменой вероятностей p_1, \dots, p_n соответствующими частотами:

$$M = \frac{1}{2}(l_n + l_{n+1}) - \frac{1}{2}[h_1(l_2 - l_0) + h_2(l_3 - l_1) + \dots + h_n(l_{n+1} - l_{n-1})]. \quad (6)$$

На практике средняя смертельная доза вычисляется именно по этой формуле.

Количества животных, подвергающихся действию различных доз, не обязательно должны быть одинаковыми. Точность формулы (6) даже повысится, если эти количества для средних доз будут больше соответствующих количеств для доз на концах ряда. Излишнее употребление сильных доз, при которых почти все животные умирают, и слабых доз, при которых смертность

почти не наблюдается, является расточительством времени и опытного материала; на это указывал в свое время Гаддум¹.

Величина M зависит от случая; ее среднее значение равно L . В силу (6), M представляет собой сумму одного постоянного и n независимых случайных слагаемых. Так как постоянное слагаемое не оказывает влияния на дисперсию всей суммы, то, согласно (15) § 3, дисперсия M имеет вид

$$\sigma_M^2 = \sigma_1^2 + \dots + \sigma_n^2,$$

где, например, σ_1^2 — дисперсия члена $-\bar{h}_1(l_2 - l_0)/2$:

$$\sigma_1^2 = \frac{p_1(1-p_1)}{4N_1} (l_2 - l_0)^2.$$

Таким образом,

$$\sigma_M^2 = \sum_{k=1}^n \frac{p_k(1-p_k)}{4N_k} (l_{k+1} - l_{k-1})^2. \quad (7)$$

Но p_k неизвестны, поэтому, согласно известному правилу (§ 5, конец), p_k заменяют частотами \bar{h}_k , а числа N_k в знаменателях — величинами $N_k - 1$. В результате получают приближенное значение для дисперсии σ_M^2 :

$$s_M^2 = \sum_{k=1}^n \frac{\bar{h}_k(1-\bar{h}_k)}{4(N_k-1)} (l_{k+1} - l_{k-1})^2. \quad (8)$$

Среднее значение s_M^2 равно σ_M^2 , и для больших количеств животных s_M^2 не очень отличается от σ_M^2 . Следовательно, величина s_M позволяет судить о точности оценки M .

Пример 36. Исследуя влияние доз солянокислого гелеземина на смертность кроликов, Чен, Андерсон и Роббинс² наблюдали следующие частоты смертельных исходов (каждая доза применялась к десяти кроликам):

Дозы	6	7	8	9	10	11	12	13
Частоты смертельных исходов	0	0,1	0,3	0,6	0,8	0,5	0,8	1,0

При вычислениях по формуле (6) воспользуемся трехзначными десятичными логарифмами³:

$$l_1 = \lg 6 = 0,778 \text{ и т. д.}$$

Данные, указанные в таблице, позволяют надеяться, что при наименьшей дозе (примерно равной $l_0 = \lg 5$) смертность практически равна нулю,

¹ G a d d u m J. П., Med. Res. Council Rep. on Biol. Standards, 3 (1933), 27.

² Quarterly Journal Pharmacol., 11 (1938), 84.

³ Применение логарифмов не является обязательным, так как все формулы этого параграфа остаются справедливыми и для самих доз (а не только для их логарифмов). — Прим. перев.

а при наибольшей дозе (примерно равной $l_{n+1} = \lg 15$) смертность практически равна единице. Впрочем, выбор l_0 и l_{n+1} совершенно безразличен, так как в нашем случае $h_1 = 0$ и $h_n = 1$, поэтому в формулах (6) и (8) члены, содержащие l_0 и l_{n+1} , исчезают.

Согласно формулам (6) и (8), получаем¹

$$M = 0,957,$$

$$s_M = 0,016.$$

Следовательно, логарифм средней смертельной дозы равен

$$L = 0,957 \pm 0,016:$$

Этот результат показывает, что бессмысленно пользоваться более чем трехзначными логарифмами доз. Даже второй знак у M не заслуживает доверия, так как выборочное квадратичное отклонение превышает единицу второго десятичного знака.

С помощью этого же самого экспериментального материала, но другим методом, связанным с большим количеством вычислений, Блисс² нашел, что

$$L = 0,961 \pm 0,0166.$$

Отсюда видно, что метод площадей по точности почти не уступает более сложным методам вычисления оценок, основанным на предположении нормальности кривой эффекта. В заключительной части моей только что цитированной работы получен более точный результат, согласно которому средняя ошибка метода площадей, в случае одинаковых количеств животных, лишь на 1% превышает среднюю ошибку метода наибольшего правдоподобия, в предположении, что кривая эффекта является нормальной.

§ 54. Методы, основанные на предположении нормальности кривой эффекта

Если предположить, что логарифмическая кривая эффекта представляет собой график функции нормального распределения, то в нашем распоряжении окажется целый ряд методов обработки наблюдений.

А. ГРАФИЧЕСКИЙ МЕТОД

Основой графического метода является такое преобразование оси Op , при котором кривая эффекта превращается в прямую.

Абсциссу l , равную логарифму дозы, теперь удобнее обозначить буквой x . Тогда, согласно (1) § 52, нормальная кривая эффекта будет задаваться уравнением

$$p = \Phi \left(\frac{x - L}{\sigma} \right). \quad (1)$$

¹ Van der Waerden, *Archiv f. exp. Pathol.*, 195 (1940), 389.

² Bliss C. I., *Quarterly Journal Pharmacy and Pharmacol.*, 11 (1938), 202.

Введем теперь новую зависимую переменную y , связанную с p соотношением

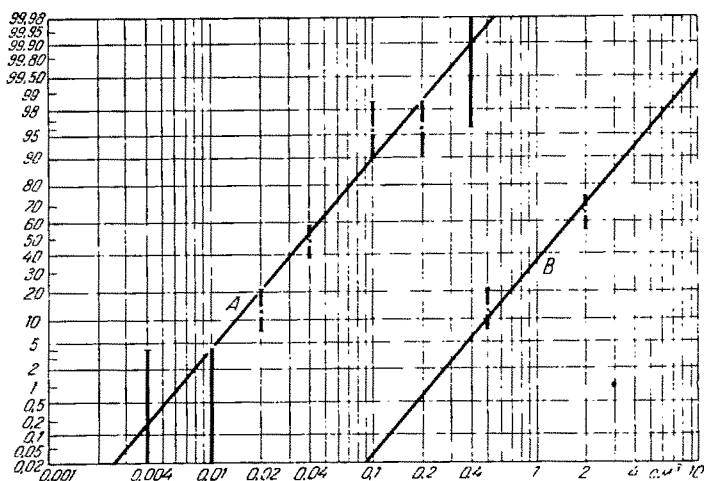
$$p = \Phi(y) \text{ или } y = \Psi(p). \quad (2)$$

В новых координатах x и y уравнение кривой эффекта будет иметь вид

$$y = \frac{x - L}{\sigma}. \quad (3)$$

Это уравнение задает прямую линию.

Иногда, для того чтобы не иметь дела с отрицательными числами, к значениям y прибавляют 5 и суммы $y + 5$ называют



Р и с. 27. Графическая оценка двух параллельных кривых эффекта методом пробитов.

пробитами. Однако с целью получения более простых формул мы будем пользоваться самими величинами y , а не $y + 5$.

Для того чтобы оценить L и σ , на оси Ox откладывают логарифмы применяемых доз, а на оси Oy пробиты $y = \Psi(h)$, соответствующие частотам h . Затем проводят такую прямую, которая возможно меньше отклоняется от полученных точек с координатами (x_i, y_i) . Существует бумага, похожая на миллиметровую бумагу и устроенная таким образом, чтобы точки (x_i, y_i) можно было наносить без предварительного вычисления логарифмов и пробитов (рис. 27).

Трудность заключается в том, что при $h = 0$ или $h = 1$ соответствующие пробиты принимают значения $-\infty$ или $+\infty$. Для того чтобы преодолеть это затруднение¹, сначала по наблю-

¹ Prigge R. und Schäfer W., Arch. exp. Path., 191 (1939), 303.

денным частотам h вычисляют доверительные границы p_1 и p_2 , причем в формулах из § 6 целесообразно положить $g = 1$: в этом случае границы не будут слишком широкими. Затем по этим p_1 и p_2 вычисляют соответствующие границы пробитов y_1 и y_2 , согласно формуле (2). В результате для каждой дозы получают отрезок, параллельный оси Oy и расположенный между точками с ординатами y_1 и y_2 . Если $h = 0$ или $h = 1$, то этот отрезок распространяется вниз или соответственно вверх до бесконечности. Предположительную прямую эффекта проводят таким образом, чтобы она пересекала все или по крайней мере большинство этих отрезков. Если это можно сделать многими способами, то прямую выбирают так, чтобы она проходила возможно ближе к точкам, соответствующим тем наблюдаемым частотам h_i , которые отличны от нуля и от единицы.

Этот метод иллюстрируется рис. 27, заимствованным из работы, цитированной в предыдущей сноске.

Б. НАИБОЛЬШЕЕ ПРАВДОПОДОБИЕ

Результаты графической оценки можно улучшить с помощью метода наибольшего правдоподобия¹. Этот метод, однако, требует очень большого количества вычислений, которые, по-моему, никогда не окупаются. Кажущаяся точность пробит-анализа является лишь иллюзорной, так как все выводы основываются на весьма недостоверной гипотезе о нормальности кривой эффекта. Если тем не менее обработку наблюдений все же рискуют строить на столь скользкой почве, то можно вполне довольствоваться грубыми графическими оценками. Для того чтобы получить для неизвестных параметров более надежные приближения, точность которых можно оценить, следует применить метод площадей, не зависящий от гипотезы нормальности кривой эффекта.

Лишь в одном случае методы получения асимптотически эффективных оценок оказываются, безусловно, необходимыми, а именно, тогда, когда нужно проверить гипотезу о том, что кривая эффекта является нормальной. В этом случае применяют критерий χ^2 (§ 51). Выражение χ^2 зависит от оценок \tilde{np} для математических ожиданий. Для того чтобы χ^2 не было слишком велико, нужно, чтобы эти оценки были *асимптотически эффективными*. Согласно § 50, такие асимптотически эффективные оценки можно получить методом наибольшего правдоподобия. Несколько более удобным является эффективный метод минимума χ_0^2 (§ 48) ($np^{(0)}$ в знаменателях χ_0^2 можно найти, например, с помощью графического метода).

¹ Bliss C. I., *Annals of Applied Biol.*, 22 (1935), 134, с дополнением Fisher R. A., 149. Этот метод обстоятельно изложен в книге: Finney D. J., *Probit Analysis*, Cambridge Univ. Press, 1947.

В. МЕТОД ДВУХ ТОЧЕК

Согласно этому методу, животных подвергают воздействию лишь двух доз испытываемого вещества, наблюдают соответствующие частоты h_1 и h_2 и на вероятностной бумаге (см. рис. 27) через точки с координатами $(x_1, \Psi(h_1))$ и $(x_2, \Psi(h_2))$ проводят прямую линию. Абсцисса точки пересечения этой прямой с осью Ox является оценкой для 50%-ной дозы.

Этот метод можно применять лишь тогда, когда количества подопытных животных велики, а параметры L и σ кривой эффекта (1) заранее приближенно известны. Для того чтобы, с большой вероятностью, обе точки были расположены по разные стороны от 50%-ной прямой $x = L$, нужно, чтобы первая доза была значительно меньше 50%-ной дозы, а вторая доза — значительно больше 50%-ной дозы. С другой стороны, нельзя допускать, чтобы дозы слишком отклонялись от 50%-ной дозы, так как если эмпирическая смертность будет равна 0 или 100%, то соответствующую точку нельзя будет нанести на вероятностную бумагу. Эти недостатки¹ усугубляются еще и тем, что изложенный метод существенно зависит от предположения нормальности кривой эффекта и что точность этого метода оценить очень трудно, а может быть, и вообще невозможно. Мне представляется более предпочтительной такая постановка эксперимента, при которой можно применять метод площадей.

Г. МЕТОД ОДНОЙ ТОЧКИ

Если приближенное значение для углового коэффициента прямой известно из других экспериментов, то наиболее точным методом получения оценки для L является метод одной точки, согласно которому ко всем подопытным животным применяется лишь одна доза, по возможности близкая к средней дозе L . В результате эксперимента получают точку (x, y) , через которую проводят прямую с известным угловым коэффициентом $1/\sigma$. В качестве оценки для средней дозы L принимают абсциссу точки пересечения прямой с осью Ox :

$$M = x - \sigma y. \quad (4)$$

На практике σ , конечно, заменяется оценкой s . Этот метод является надежным только тогда, когда примененная доза близка к средней смертельной дозе L . Следовательно, сначала с помощью предварительных опытов нужно определить приближенное значение для средней смертельной дозы (например, по методу площадей) и лишь затем применять метод одной точки. О вычислении

¹ По этому вопросу см. Behrens und Kärber, *Archiv exp. Path.*, 177 (1935), 637.

средней ошибки M см. мою ранее цитированную работу в журнале *Archiv exp. Path.*, 195 (1940).

Метод одной точки использует лишь среднюю, наиболее крутую часть кривой эффекта. Следовательно, он мало зависит от предположения нормальности кривой эффекта.

Д. ЛОГИСТИЧЕСКАЯ КРИВАЯ

Вместо нормальной кривой эффекта можно, следуя Берксону¹, принять за основу так называемую логистическую кривую, уравнение которой задается формулой

$$p = \frac{1}{e^{-z} + 1}, \quad (5)$$

где z — линейная функция от x . Если z возрастает от $-\infty$ до $+\infty$, то функция (5) так же, как $\Phi(z)$, возрастает от 0 до 1. Функция (5) очень мало отличается от функции нормального распределения. Кривая с уравнением (5) так же, как нормальная кривая, симметрична относительно точки с координатами $z = 0$, $p = 1/2$:

$$p(z) + p(-z) = 1.$$

Уравнение (5) можно легко разрешить относительно z :

$$z = -\ln\left(\frac{1}{p} - 1\right). \quad (6)$$

Величины z , определяемые формулой (6), называются *логитами*. Вычисления с помощью логитов выполняются легче вычислений с помощью указанных ранее пробитов $y = \Psi(p)$. Все методы пробит-анализа (графический метод, метод наибольшего правдоподобия, метод двух точек и метод одной точки) можно с одинаковым или еще большим успехом применять для логит-анализа.

§ 55. Методы «вверх и вниз»

А. МЕТОД ДИКСОНА И МУДА

Диксон и Муд² указали метод, который при меньшем количестве опытного материала позволяет получать столь же точные оценки, как и методы, изложенные выше. Согласно этому методу, сначала нужно выбрать исходную дозу (l — логарифм этой дозы) и подвергнуть ее воздействию одно животное. Затем, смотря по тому, реагирует ли животное определенным образом (например, умирает) или нет, логарифм уменьшается или соответственно

¹ Berkson J., Journal Amer. Statist. Assoc., 39, 41, 48.

² Dixon W. J. and Mood A. M., Journal Amer. Statist. Assoc., 43 (1948), 109.

увеличивается на d и новая доза применяется ко второму животному и т. д. — всегда вверх или вниз. Никакие другие дозы, кроме доз с логарифмами $l, l \pm d, l \pm 2d, \dots$, не применяются.

При таком методе большая часть экспериментов автоматически проводится с теми дозами, которым соответствует наиболее крутая часть кривой эффекта, так как если смертность близка к единице, то доза, с большой вероятностью, будет уменьшена, и точно так же если смертность близка к нулю, то доза будет увеличена. Это очень выгодно, так как приходится иметь дело как раз с той частью кривой эффекта, которая находится вблизи точки с ординатой $p = 1/2$ (50%-ная смертность). Чтобы метод действовал хорошо, разность d должна, по возможности, быть выбрана из интервала от $\sigma/2$ до 2σ . Таким образом, грубая оценка для σ должна быть известна заранее.

Для оценки неизвестных параметров можно бы было применить метод площадей. Однако Диксон и Муд воспользовались другим, чрезвычайно простым методом. Согласно этому методу, сначала надо подсчитать общее число «успехов» (т. е. число случаев, в которых животные реагировали положительно) и общее число «неудач». Пусть N — наименьшее из этих двух чисел и пусть соответствующее менее частое событие (успех или неудача) происходило при экспериментах с дозами l_0, l_1, l_2, \dots , причем для указанных доз это событие наступало n_0, n_1, n_2, \dots раз. Образуют суммы

$$A = \sum k n_k,$$

$$B = \sum k^2 n_k.$$

Тогда для L — логарифма 50%-ной дозы — можно указать следующую оценку:

$$M = l_0 + d \left(\frac{A}{N} \pm \frac{1}{2} \right). \quad (1)$$

Знак $+$ употребляется в том случае, когда успехов наблюдается больше, чем неудач, а знак $-$, когда успехов наблюдается меньше, чем неудач. В качестве оценки для σ используют выражение

$$s = 1,62 \left(\frac{NB - A^2}{N^2} + 0,03 \right). \quad (2)$$

Обоснование этих формул методом наибольшего правдоподобия можно найти в оригинальной работе. Броунли, Ходжес и Розенблатт (см. Brownlee, Hodges and Rosenblatt, J. Amer. Statist. Assoc., 48, 262) показали, что формулы (1) и (2) очень полезны даже при малом количестве опытов. Кроме того, эти авторы предложили вместо (1) и (2) несколько модификаций, предназначенных для того, чтобы сделать метод еще более эффективным.

При обосновании формул (1) и (2) кривая эффекта предполагалась нормальной, однако это предположение не очень суще-

ственно. Достаточно, чтобы часть кривой вблизи средней смертельной дозы приближенно представлялась некоторой нормальной кривой. Поведение ветвей кривой эффекта вблизи прямых $p = 0$ и $p = 1$ не имеет большого значения, так как метод устроен таким образом, что очень большие и очень малые дозы применяются чрезвычайно редко, и поэтому они не оказывают почти никакого влияния на среднее значение и дисперсию оценки M .

Недостатком метода является то обстоятельство, что последующий опыт можно проводить лишь тогда, когда установлен результат предыдущего опыта (успех или неудача). Этот недостаток можно частично уменьшить, поставив одновременно, например, четыре ряда опытов.

Б. МЕТОД СТОХАСТИЧЕСКОЙ АППРОКСИМАЦИИ

Согласно результатам Роббинса и Монро¹, метод «вверх и вниз» можно еще более улучшить, если дозы увеличивать или уменьшать не на постоянную величину d , а на некоторую переменную величину, стремящуюся к нулю при $n \rightarrow \infty$. Предварительно выбирают убывающую последовательность положительных чисел a_1, a_2, \dots и задают начальную дозу l_1 . Если на дозу l_n очередное животное реагирует, то в качестве следующей дозы выбирают

$$l_{n+1} = l_n - \frac{1}{2}a_n, \quad (3)$$

но если животное на дозу l_n не реагирует, то полагают

$$l_{n+1} = l_n + \frac{1}{2}a_n. \quad (4)$$

Можно также, при желании, проводить опыты одновременно с несколькими животными. Если h_n — частота успехов в n -м опыте, то в следующем опыте дозу выбирают равной

$$l_{n+1} = l_n + \left(\frac{1}{2} - h_n\right)a_n. \quad (5)$$

В качестве оценки для 50%-ной дозы принимают последнюю вычисленную дозу l_{N+1} , где N — число опытов. При некоторых ограничивающих предположениях Роббинс и Монро доказали, что при $N \rightarrow \infty$ оценка l_{N+1} сходится, по вероятности, к 50%-ной дозе, т. е. что для достаточно больших N с вероятностью, сколь угодно близкой к единице, имеет место неравенство

$$|l_{N+1} - L| < \varepsilon.$$

¹ Robbins H. and Monro S., A stochastic approximation method, Ann. Math. Stat., 22 (1951), 400.

Об асимптотическом распределении оценки l_{N+1} прежде всего см. работу Чжуна: Chung K. L., On a stochastic approximation method, *Ann. of Math. Stat.*, 25 (1954), 463.

Какой выбор коэффициентов a_n является наилучшим? Во всяком случае они должны стремиться к нулю, так как иначе, в силу формул (3) и (4), сходимость последовательности l_n к L будет невозможна. С другой стороны, если бы a_n столь быстро стремились к нулю, что ряд $\sum a_n$ был бы сходящимся, то с некоторого номера n успех или неудача в очередном опыте почти не оказывали бы влияния на величину последующих доз l_{n+1}, l_{n+2}, \dots , так как сумма всех поправочных членов $\pm \frac{1}{2} a_n \pm \frac{1}{2} a_{n+1} \pm \dots$ по абсолютной величине была бы меньше ε . Поэтому a_n выбирают таким образом, чтобы ряд $\sum a_n$ расходился.

Чжун рекомендует выбрать

$$a_n = \frac{c}{n^{1-\varepsilon}} \quad \left(0 < \varepsilon < \frac{1}{2}\right). \quad (6)$$

В противоположность этому, Роббинс и Моиро выбирают

$$a_n = \frac{c}{n}. \quad (7)$$

Коэффициенты (6) столь медленно стремятся к нулю, что это обеспечивает состоятельность оценки при довольно общих предположениях о виде кривой эффекта. При выборе последовательности (7) нужно соблюдать осторожность. Если a — угловый коэффициент касательной к кривой эффекта вблизи от 50%-ной дозы L , то постоянную c в формуле (7) нужно выбрать большей¹, чем $1/(2a)$.

Если вблизи 50%-ной дозы кривая эффекта приблизительно представима некоторой прямой линией, то, согласно результатам Чжуна, при определенных дополнительных предположениях, квадратичное отклонение оценки l_{N+1} будет асимптотически равно

$$\frac{c}{\sqrt{2ac-1}} \frac{\sigma_h}{\sqrt{N}}, \quad (8)$$

где σ_h — квадратичное отклонение для частоты h , соответствующей дозе, близкой к L . Если n' — количество животных в каждом опыте, то

$$\sigma_n^2 = \frac{pq}{n'} \sim \frac{1}{4n'}. \quad (9)$$

¹ Если кривая эффекта задается дифференцируемой функцией $p(x)$ и если $a = \sup_x p'(x)$, то c в формуле (7) должно удовлетворять неравенству $c > 1/(2a)$. — *Прим. перев.*

Выражение (8) достигает минимума при

$$c = \frac{1}{a}. \quad (10)$$

Но если угловой коэффициент a известен лишь приближенно, то c разумно выбрать несколько большим, чем $1/a$. На квадратичное отклонение это окажет лишь малое влияние, так как функция (8) вблизи своего минимума возрастает очень медленно. Во всех случаях, как уже говорилось, c нужно выбирать большим, чем $1/(2a)$. Если c стремится к $1/(2a)$, то выражение (8) становится очень большим.

ПРОВЕРКА ГИПОТЕЗ С ПОМОЩЬЮ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ

Статистические критерии являются важнейшей частью всех приложений. В этой главе мы сначала рассмотрим некоторые наиболее важные критерии и при этом вспомним те из них, которые были уже изложены ранее. Затем мы познакомимся с основными идеями общей теории Неймана и Пирсона.

Если даже читатель изучил и не все предшествующие главы этой книги, то, как я надеюсь, с помощью примеров он сможет уяснить те основные принципы, от которых зависит выбор критерия, соответствующего данным конкретным условиям. Необходимость знакомства с основными понятиями гл. I и II предполагается сама собой разумеющейся. При доказательствах в некоторых случаях будут, конечно, делаться ссылки на более поздние главы (а именно на гл. VIII и IX), а также на дополнительную литературу.

§ 56. Применения критерия χ^2

Общий критерий χ^2 , выведенный нами в § 51, включает в себя различные специальные случаи, часть которых была уже рассмотрена ранее. Напомним, что критерий χ^2 применяется тогда, когда по наблюдаемым частотам нужно проверить некоторую гипотезу относительно вероятностей.

А. ПРОВЕРКА ОДНОЙ ПРЕДПОЛАГАЕМОЙ ВЕРОЯТНОСТИ

Пусть некоторое событие в n независимых опытах наступило x_1 раз и не наступило x_2 раз ($x_1 + x_2 = n$), и пусть предполагается, что вероятность этого события равняется некоторому заданному числу p . Для проверки этой гипотезы полагают $q = 1 - p$ и образуют

$$\chi^2 = \frac{(x_1 - np)^2}{np} + \frac{(x_2 - nq)^2}{nq} \quad (1)$$

(np и nq — математические ожидания x_1 и x_2). Если χ^2 оказывается больше некоторой границы, выбранной по табл. 6, то пред-

полагаемую вероятность p отвергают. Так как два наблюдаемых количества, x_1 и x_2 , связаны одним линейным уравнением $x_1 + x_2 = n$, то число степеней свободы равно

$$f = 2 - 1 = 1.$$

Но

$$(x_1 - np) + (x_2 - nq) = 0,$$

следовательно,

$$(x_1 - np)^2 = (x_2 - nq)^2.$$

Последнее равенство позволяет записать (1) проще:

$$\chi^2 = \frac{(x_1 - np)^2 (p + q)}{npq} = \frac{(x_1 - np)^2}{npq}. \quad (2)$$

Это в точности то же самое выражение, которым мы пользовались раньше.

Б. НЕСКОЛЬКО ПРЕДПОЛАГАЕМЫХ ВЕРОЯТНОСТЕЙ

Пусть имеется выборка объема n из некоторой бесконечной совокупности, разбитой на m классов, и пусть x_1, \dots, x_m — наблюдаемые количества выборочных элементов, принадлежащих этим классам, причем $x_1 + \dots + x_m = n$. Нужно проверить гипотезу, согласно которой вероятности, соответствующие m классам, равны заданным числам p_1, \dots, p_m (например, в случае двух несцепленных генов, $\frac{1}{16}, \frac{3}{16}, \frac{3}{16}, \frac{9}{16}$). С этой целью вычисляют

$$\chi^2 = \sum \frac{(x_i - np_i)^2}{np_i} \quad (m - 1 \text{ степеней свободы}), \quad (3)$$

и если χ^2 превосходит границу, найденную по табл. 6, то гипотезу отвергают.

Строго говоря, распределение χ^2 имеет место лишь асимптотически при $n \rightarrow \infty$. Точное распределение случайной величины χ^2 является дискретным, так как при заданных n и p_i величина χ^2 может принимать лишь конечное число значений.

Для того чтобы проверить, насколько близки точное распределение и распределение χ^2 , я для случая

$$n = 10; p_1 = 0,5; p_2 = 0,3; p_3 = 0,2$$

вычислил точное распределение случайной величины

$$Y = 1 - e^{-\frac{1}{2}\chi^2} \quad (4)$$

и сравнил его с асимптотическим распределением. В этом случае

$$\chi^2 = \frac{(x_1 - 5)^2}{5} + \frac{(x_2 - 3)^2}{3} + \frac{(x_3 - 2)^2}{2}. \quad (5)$$

Имеется 66 возможных троек чисел (x_1, x_2, x_3) , удовлетворяющих условию $x_1 + x_2 + x_3 = 10$. Тройке (x_1, x_2, x_3) соответствует вероятность

$$\frac{10!}{x_1! x_2! x_3!} (0,5)^{x_1} (0,3)^{x_2} (0,2)^{x_3}. \quad (6)$$

Каждая тройка, согласно (4) и (5), приводит к определенному значению Y . Эти значения и их вероятности (6) определяют некоторую ступенчатую функцию, являющуюся функцией распределения случайной величины Y (рис. 28).

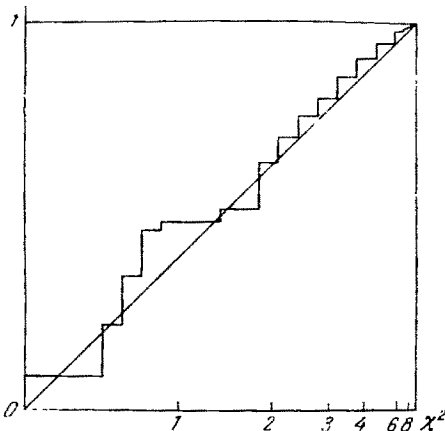


Рис. 28. Точное и асимптотическое распределение Y с двумя степенями свободы ($n = 10$, $p_1 = 0,5$, $p_2 = 0,3$, $p_3 = 0,2$).

Асимптотическая функция распределения χ^2 с двумя степенями свободы равна

$$G(u) = \int_0^u e^{-t} dt = 1 - e^{-u}.$$

Поэтому асимптотическая функция распределения случайной величины Y задается формулой

$$F(v) = \begin{cases} 0, & v < 0, \\ v, & 0 \leq v < 1, \\ 1, & v \geq 1 \end{cases} \quad (7)$$

и имеет своим графиком прямую, изображенную на рис. 28. Отклонение точной функции распределения Y от асимптотической мало, особенно в интервале от 0,95 до 1, который наиболее важен для приложений. Вероятность события $\chi^2 > 9,21$ при асимптотическом распределении должна быть равна 0,01, а в действительности она равна 0,0096. Вероятность события $\chi^2 > 5,99$, которая должна быть равной 0,05, в действительности равна 0,0502. В большинстве случаев истинный уровень значимости критерия χ^2 оказывается даже меньше уровня значимости, вычисленного по асимптотической формуле, поэтому примене-

ние асимптотического распределения, как правило, лишь увеличивает надежность критерия¹.

В литературе часто можно найти замечание, что асимптотическое распределение χ^2 можно применять лишь тогда, когда наблюдаемые x_i или их математические ожидания np_i не слишком малы. *Только что приведенный пример показывает, что математические ожидания np_i могут быть равны лишь двум или трем единицам и тем не менее асимптотическое распределение оказывается еще применимым.* Это же подтверждается и другими примерами. *Если имеется много классов, то математические ожидания могут быть даже равны единице.* Я однажды проводил вычисления в примере с десятью классами

$$np_1 = 1, np_2 = 1, \dots, np_{10} = 1 \quad (n = 10)$$

и нашел все еще удовлетворительное согласие с асимптотическим распределением χ^2 . Следовательно, чрезмерная осторожность является излишней.

В. СРАВНЕНИЕ ДВУХ ВЕРОЯТНОСТЕЙ

Пусть некоторое событие в n_1 опытах наступило x_1 раз и не наступило y_1 раз. И пусть в новых n_2 опытах это событие наступило x_2 раз и не наступило y_2 раз. Нужно проверить, изменилась ли вероятность события или нет? Все опыты предполагаются независимыми.

Гипотеза, которую мы хотим проверить, гласит: вероятность p в обоих случаях одинакова. Значения p мы не знаем. Для того чтобы вычислить χ^2 , мы должны p заменить некоторой оценкой, а именно, асимптотически эффективной оценкой, так как иначе значение χ^2 может получиться слишком большим (§ 51).

Для отыскания такой оценки мы воспользуемся методом наибольшего правдоподобия. Если вероятность осуществления события в каждом отдельном опыте равна p , то вероятность того, что в n_1 опытах это событие наступит x_1 раз, а в n_2 опытах наступит x_2 раз, равна

$$\frac{n_1!}{x_1! y_1!} \frac{n_2!}{x_2! y_2!} p^{x_1+x_2} q^{y_1+y_2}.$$

При вычислении максимума этого выражения на числовой множитель можно не обращать внимания. Максимум достигается в точке

$$\tilde{p} = \frac{x_1 + x_2}{n_1 + n_2}. \quad (8)$$

¹ Надежность критерия χ^2 увеличивается, так как реже будет ошибочно отвергаться испытываемая гипотеза, когда она имеет место в действительности и вместе с тем мы с большим основанием будем заключать о сущности (неслучайности) расхождений между наблюдаемым в опыте распределением и гипотетически допускаемым. — *Прим. ред.*

С помощью этого значения \tilde{p} и $\tilde{q} = 1 - \tilde{p}$ образуем теперь

$$\chi^2 = \frac{(x_1 - n_1\tilde{p})^2}{n_1\tilde{p}} + \frac{(y_1 - n_1\tilde{q})^2}{n_1\tilde{q}} + \frac{(x_2 - n_2\tilde{p})^2}{n_2\tilde{p}} + \frac{(y_2 - n_2\tilde{q})^2}{n_2\tilde{q}}, \quad (9)$$

или, что то же самое,

$$\chi^2 = \frac{(x_1 - n_1\tilde{p})^2}{n_1\tilde{p}\tilde{q}} + \frac{(x_2 - n_2\tilde{p})^2}{n_2\tilde{p}\tilde{q}}. \quad (10)$$

Но

$$(x_1 - n_1\tilde{p}) + (x_2 - n_2\tilde{p}) = 0,$$

следовательно, χ^2 можно записать короче:

$$\chi^2 = \frac{(x_1 - n_1\tilde{p})^2 (n_1 + n_2)}{n_1 n_2 \tilde{p} \tilde{q}}. \quad (11)$$

Ранее мы видели, что при малых n_1 и n_2 множитель $N = n_1 + n_2$ в числителе (11) целесообразно заменить на $N - 1$. При этом истинный уровень значимости будет лишь незначительно превышать заранее заданную величину (см. § 9). Таким образом, для сравнения двух вероятностей вместо χ^2 следует пользоваться статистикой

$$\chi_1^2 = \frac{(x_1 - n_1\tilde{p})^2 (n_1 + n_2 - 1)}{n_1 n_2 \tilde{p} \tilde{q}} = \frac{(x_1 n_2 - x_2 n_1)^2 (n_1 + n_2 - 1)}{n_1 n_2 (x_1 + x_2) (y_1 + y_2)}. \quad (12)$$

Число степеней свободы равно

$$f = 4 - 2 - 1 = 1,$$

так как наблюдались четыре количества x_1, y_1, x_2, y_2 , связанные двумя линейными уравнениями, и оценивался один неизвестный параметр p по формуле (8).

Г. ПРОВЕРКА НЕЗАВИСИМОСТИ ДВУХ ПРИЗНАКОВ

Пусть N объектов расклассифицированы по двум парам признаков, вследствие чего возникают четыре класса (так называемая *таблица сопряженности признаков* 2×2), и пусть $x_{11}, x_{12}, x_{21}, x_{22}$ — количества объектов в этих четырех классах. Нужно проверить, будут ли обе пары признаков независимыми? Пусть p_1 и p_2 — вероятности того, что в отдельном опыте объект будет обладать соответственно первым или вторым признаками, принадлежащими первой паре, и пусть q_1 и q_2 — вероятности признаков второй пары ($p_1 + p_2 = 1, q_1 + q_2 = 1$). Если признаки независимы, то четырем классам соответствуют вероятности $p_1 q_1, p_1 q_2, p_2 q_1, p_2 q_2$.

Значения p_i и q_k неизвестны. Если для их оценки снова воспользоваться методом наибольшего правдоподобия, то найдем

$$\tilde{p}_i = \frac{x_{i1} + x_{i2}}{N} \quad (i = 1, 2), \quad \tilde{q}_k = \frac{x_{1k} + x_{2k}}{N} \quad (k = 1, 2). \quad (13)$$

С помощью \tilde{p}_i и \tilde{q}_k можно образовать

$$\begin{aligned} \chi^2 = & \frac{(x_{11} - N\tilde{p}_1\tilde{q}_1)^2}{N\tilde{p}_1\tilde{q}_1} + \frac{(x_{12} - N\tilde{p}_1\tilde{q}_2)^2}{N\tilde{p}_1\tilde{q}_2} + \\ & + \frac{(x_{21} - N\tilde{p}_2\tilde{q}_1)^2}{N\tilde{p}_2\tilde{q}_1} + \frac{(x_{22} - N\tilde{p}_2\tilde{q}_2)^2}{N\tilde{p}_2\tilde{q}_2} \end{aligned} \quad (14)$$

Если в числителях и знаменателях $N\tilde{p}_i$ заменить величинами $n_i = x_{i1} + x_{i2}$, то получим то же самое выражение, что и (9), которое можно преобразовать в (10) или (11), или

$$\chi^2 = \frac{(x_{11}x_{22} - x_{12}x_{21})^2 N}{(x_{11} + x_{12})(x_{21} + x_{22})(x_{11} + x_{21})(x_{12} + x_{22})}. \quad (15)$$

Так как наблюдались четыре количества, связанные одним линейным соотношением

$$x_{11} + x_{12} + x_{21} + x_{22} = N,$$

и два неизвестных параметра p_1 и q_1 оценивались по результатам наблюдений с помощью формул (13), то число степеней свободы равно

$$f = 4 - 1 - 2 = 1.$$

Если множитель N в числителе (15) мал, то его снова целесообразно заменить величиной $N - 1$.

Среди английских статистиков в свое время велась большая дискуссия по вопросу о выборе числа степеней свободы. Если проверка независимости признаков производится на основе таблицы 2×2 , то чему должно равняться это число, 1 или 3? Первоначально Карл Пирсон вывел критерий χ^2 только для случая, когда вероятности p_i и q_k заданы. При этом предположении асимптотически для больших N имеет место распределение χ^2 с $4 - 1 = 3$ степенями свободы. Если истинные p_i и q_k заменить их приближенными значениями (13), то χ^2 может лишь уменьшиться. Поэтому истинное значение χ^2 не меньше приближенного значения (15). Следовательно, если приближенное значение (15) превосходит границу u , то и истинное значение χ^2 заведомо превосходит u . Пусть граница u выбрана по таблице с тремя степенями свободы, тогда вероятность того, что истинное значение χ^2 будет удовлетворять неравенству $\chi^2 \geq u$, равна β ($= 0,01$ или $0,05$). Таким образом, для того чтобы иметь уверенность, что уровень значимости критерия χ^2 не превосходит β , нужно

пользоваться таблицами распределения χ^2 с тремя степенями свободы — такой вывод делал Пирсон.

В противоположность этому, Фишер оправдывал свою точку зрения так: величина χ^2 , вычисленная по формуле (15) (назовем ее $\tilde{\chi}^2$), в большинстве случаев оказывается меньше истинного значения χ^2 . Следовательно, вероятность события $\tilde{\chi}^2 > u$ существенно меньше вероятности события $\chi^2 > u$. Таким образом, уровень значимости критерия χ^2 с тремя степенями свободы существенно меньше β , т. е. он излишне мал. Но если принять $f = 1$, то уровень значимости будет в точности равен β .

Так как Фишер не смог точно доказать справедливость своего взгляда и приводил лишь наводящие соображения, то Юль и Браунли с помощью обширных случайных экспериментов попытались установить, каково число степеней свободы у функции распределения $\tilde{\chi}^2$: $f = 3$ или $f = 1$? Опыты показали, что прав был Фишер, однако противники взгляда Фишера стали критиковать это экспериментальное доказательство. В конце концов спор был полностью решен Нейманом и Е. Пирсоном¹, которые с помощью математического доказательства установили, что интуиция Фишера была правильной.

д. СРАВНЕНИЕ НЕСКОЛЬКИХ ВЕРОЯТНОСТЕЙ

Пусть некоторое событие в первых n_1 опытах наступило x_1 раз и не наступило y_1 раз, во второй группе из n_2 опытов оно наступило x_2 раз и не наступило y_2 раз, и т. д. Нужно проверить, одинаковы вероятности этого события во всех группах опытов или нет?

Еще более общим является следующий случай. Пусть n_1 объектов разбиты по какому-то признаку на h классов, и пусть x_1, y_1, \dots, z_1 — количества объектов в этих классах. Следующим n_2 объектам аналогичным образом соответствуют количества x_2, y_2, \dots, z_2 и т. д. до x_k, y_k, \dots, z_k . Таким образом, в результате наблюдений получается hk чисел, которые можно расположить в прямоугольную таблицу

$$\begin{array}{cccc|c} x_1 & y_1 & \dots & z_1 & n_1 \\ x_2 & y_2 & \dots & z_2 & n_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_k & y_k & \dots & z_k & n_k \\ \hline \Sigma x & \Sigma y & \dots & \Sigma z & N \end{array}$$

Справа указаны суммы по строкам, снизу — по столбцам и, наконец, в правом нижнем углу указана общая сумма N .

¹ Neuman J. and Pearson E. S., On the use and interpretation of test criteria, *Biometrika*, 20 A, 175 и 263.

Нужно проверить, могут ли вероятности p, q, \dots, r , соответствующие h классам, быть одинаковыми для всех строк? Наилучшими оценками для p, q, \dots, r являются общие частоты классов

$$\tilde{p} = \frac{\sum x}{N}; \quad \tilde{q} = \frac{\sum y}{N}; \quad \dots; \quad \tilde{r} = \frac{\sum z}{N}. \quad (16)$$

С помощью этих оценок вычисляют оценки для математических ожиданий

$$n_i \tilde{p}, \quad n_i \tilde{q}, \quad \dots, \quad n_i \tilde{r}$$

и вычитают их из наблюдаемых количеств x_i, y_i, \dots, z_i . Полученные разности снова располагают в прямоугольную таблицу

$$\begin{array}{ccccccc} x_1 - n_1 \tilde{p} & y_1 - n_1 \tilde{q} & \dots & z_1 - n_1 \tilde{r} & & & \\ x_2 - n_2 \tilde{p} & y_2 - n_2 \tilde{q} & \dots & z_2 - n_2 \tilde{r} & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{array}$$

В этой таблице суммы по строкам и суммы по столбцам обязаны быть равными нулю. Этим свойством пользуются для контроля вычислений.

Если квадраты всех hk разностей разделить на соответствующие оценки для математических ожиданий и результаты сложить, то получим

$$\chi^2 = \sum \frac{(x_i - n_i \tilde{p})^2}{n_i \tilde{p}} + \sum \frac{(y_i - n_i \tilde{q})^2}{n_i \tilde{q}} + \dots \quad (17)$$

Число степеней свободы равно

$$f = hk - k - (h - 1) = (h - 1)(k - 1), \quad (18)$$

так как наблюдались hk количеств, связанных k линейными уравнениями

$$x_i + y_i + \dots + z_i = n_i;$$

кроме того, h параметров p, q, \dots, r оценивались по результатам наблюдений с помощью формул (16), и эти параметры удовлетворяют одному линейному уравнению

$$p + q + \dots + r = 1.$$

Следовательно, все вероятности определяются $h - 1$ независимыми параметрами, поэтому в (18) из $hk - k$ вычитается не h , а $h - 1$.

Е. РЕДКИЕ СОБЫТИЯ

Как уже ранее упоминалось, событие называют редким, если его вероятность p настолько мала, что во всех формулах $q = 1 - p$ можно заменить единицей. Тогда биномиальное распределение

перейдет в распределение Пуассона: вероятность того, что в n опытах данное событие осуществится x раз, будет равна¹

$$W_x = \frac{(np)^x}{x!} e^{-np} = \frac{\lambda^x}{x!} e^{-\lambda}. \quad (19)$$

Правая часть (19) зависит лишь от произведения $\lambda = np$, равного математическому ожиданию x , и не зависит от p и n в отдельности. Соответственно упрощается и формула для χ^2 . Слагаемым с np в знаменателе можно пренебречь, так как оно мало по сравнению со слагаемым, у которого в знаменателе np . В этом случае формула (1) будет иметь вид

$$\chi^2 = \frac{(x - np)^2}{np} = \frac{(x - \lambda)^2}{\lambda}. \quad (20)$$

Гипотезу о том, что λ принимает некоторое заданное значение, следует отвергнуть тогда, когда (20) превосходит границу для χ^2 с одной степенью свободы. Точно так же если имеются два независимых редких события, из которых первое наступило x раз, а второе — y раз, то гипотезу о том, что математические ожидания x и y равны соответственно λ и μ следует отвергнуть тогда, когда выражение

$$\chi^2 = \frac{(x - \lambda)^2}{\lambda} + \frac{(y - \mu)^2}{\mu} \quad (21)$$

превосходит некоторую границу, найденную по таблице функции распределения χ^2 с двумя степенями свободы.

Ж. СРАВНЕНИЕ ДВУХ РЕДКИХ СОБЫТИЙ

Эта задача была уже подробно изложена ранее (§ 10 Б). Теперь мы хотим лишь кратко показать, что критерий, найденный в § 10, можно непосредственно получить из общего критерия χ^2 .

Пусть за время t_1 первое редкое событие наблюдалось x_1 раз и за время t_2 другое редкое событие наблюдалось x_2 раз, и пусть математические ожидания x_1 и x_2 равны соответственно

$$\lambda_1 = \nu_1 t_1, \quad \lambda_2 = \nu_2 t_2.$$

Нужно проверить гипотезу, согласно которой $\nu_1 = \nu_2$. Если положим $\nu_1 = \nu_2 = \nu$, то

$$\lambda_1 = \nu t_1, \quad \lambda_2 = \nu t_2. \quad (22)$$

Для того чтобы можно было вычислить χ^2 , нужно оценить ν . Функция правдоподобия, в силу распределения Пуассона, имеет вид

$$(\nu t_1)^{x_1} e^{-\nu t_1} (\nu t_2)^{x_2} e^{-\nu t_2}.$$

¹ Это равенство является приближенным. Точный смысл формулы (19) указан в § 10. — *Прим. перев.*

Если отбросить множители, не зависящие от ϑ , и вычислить логарифм, то получим

$$L(x_1, x_2 | \vartheta) = (x_1 + x_2) \ln \vartheta - (t_1 + t_2) \vartheta. \quad (23)$$

Выражение (23) достигает максимума в точке

$$\tilde{\vartheta} = \frac{x_1 + x_2}{t_1 + t_2}. \quad (24)$$

Следовательно,

$$\chi^2 = \frac{(x_1 - \tilde{\vartheta} t_1)^2}{\tilde{\vartheta} t_1} + \frac{(x_2 - \tilde{\vartheta} t_2)^2}{\tilde{\vartheta} t_2}. \quad (25)$$

Так как наблюдались два количества x_1 и x_2 и один параметр ϑ оценивался по формуле (24), то число степеней свободы равно

$$f = 2 - 1 = 1. \quad (26)$$

3. ПРОВЕРКА НОРМАЛЬНОСТИ РАСПРЕДЕЛЕНИЯ

Пусть результатами наблюдений являются n независимых случайных величин z_1, \dots, z_n . Нужно проверить гипотезу, согласно которой все z_i распределены одинаково нормально.

С этой целью можно вычислить эмпирическую функцию распределения и применить критерий Колмогорова (§ 16). Ранее было уже отмечено, что «хвосты» распределения, т. е. очень большие и очень малые значения z , учитываются этим критерием относительно слабо. А как раз поведение «хвостов» может, при определенных условиях, оказаться решающим для суждения об отклонении от нормальности!

Применение критерия Колмогорова затрудняется еще и тем, что математическое ожидание и дисперсия нормального распределения, как правило, бывают неизвестны.

Хорошим методом, несколько сильнее учитывающим поведение «хвостов», является метод моментов. Мы здесь дадим лишь краткий обзор этого метода. Обоснование можно найти в книге Крамера (Крамер Г., Математические методы статистики, ИЛ, М., 1948, гл. 27.1 — 28.4 и 29.3).

Центральные выборочные моменты определяются формулами

$$m_k = \frac{1}{n} \sum (z - \bar{z})^k \quad (k = 1, 2, \dots).$$

По определению, первый момент m_1 равен нулю. С помощью m_2 , m_3 и m_4 вычисляются *асимметрия* и *эксцесс*, которые равны соответственно

$$g_1 = \frac{m_3}{\sqrt{m_2^3}}, \quad g_2 = \frac{m_4}{m_2^2} - 3.$$

При больших n все m_k , а также g_1 и g_2 распределены асимптотически нормально. Эти случайные величины можно использовать в качестве оце-

нок для истинных моментов μ_k , а также для асимметрии и эксцесса истинного распределения

$$\gamma_1 = \frac{\mu_3}{\sqrt{\mu_2^3}}, \quad \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3.$$

В случае нормального распределения γ_1 и γ_2 равны нулю.

При конечных n целесообразно заменить g_1 и g_2 величинами

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} g_1 \quad \text{и} \quad G_2 = \frac{n-1}{(n-2)(n-3)} [(n+1)g_2 + 6].$$

Если истинное распределение является нормальным, то математические ожидания G_1 и G_2 в точности равны нулю. Их дисперсии задаются формулами

$$\sigma_1^2 = \frac{6n(n-1)}{(n-2)(n+1)(n+3)}, \quad \sigma_2^2 = \frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}.$$

Следовательно, с помощью статистик G_1/σ_1 или G_2/σ_2 можно построить критерий для проверки нормальности истинного распределения. Обе статистики асимптотически нормальны с нулевым средним значением и единичной дисперсией.

Метод χ^2 можно применять не только при нормальном, но также и при других распределениях. Согласно этому методу, интервал изменения z разбивают на r частей, границами которых служат точки t_1, \dots, t_{r-1} , и подсчитывают количество z_j в каждом частичном интервале. Пусть эти количества равны x_1, \dots, x_r .

Для того чтобы можно было вычислить χ^2 , нужно знать математические ожидания np_i , а для этого нужно в свою очередь найти оценки m и s для среднего значения и квадратичного отклонения истинного нормального распределения. Зная эти оценки, можно положить

$$p_i = \Phi\left(\frac{t_i - m}{s}\right) - \Phi\left(\frac{t_{i-1} - m}{s}\right). \quad (27)$$

Если мы хотим применить теорию из § 51, то в качестве m и s мы должны выбрать асимптотически эффективные оценки, которые зависят лишь от x_1, \dots, x_n . За первое приближение можно принять известные оценки

$$m_0 = \frac{1}{n} \sum z, \quad (28)$$

$$s_0^2 = \frac{1}{n-1} \sum (z - m_0)^2. \quad (29)$$

Однако (28) и (29) не удовлетворяют указанному выше условию,

согласно которому оценки должны зависеть лишь от x_i . С помощью m_0 и s_0 образуем

$$p_{i0} = \Phi\left(\frac{t_i - m_0}{s_0}\right) - \Phi\left(\frac{t_{i-1} - m_0}{s_0}\right). \quad (30)$$

Для определения оценок m и s теперь можно воспользоваться методом наименьших квадратов и потребовать, чтобы выражение

$$\chi_0^2 = \sum \frac{(x_i - np_i)^2}{np_{i0}}. \quad (31)$$

было минимальным. Функции p_i в (31) целесообразно заменить линейными функциями

$$p_i = p_{i0} + (m - m_0) q_i + (s - s_0) r_i, \quad (32)$$

где q_i и r_i — значения частных производных от (27) в точке (m_0, s_0) :

$$q_i = \frac{\partial p_i}{\partial m}(m_0, s_0); \quad r_i = \frac{\partial p_i}{\partial s}(m_0, s_0). \quad (33)$$

Как известно, метод наименьших квадратов в этом случае приводит к двум линейным относительно $m - m_0$ и $s - s_0$ уравнениям, решая которые можно определить m и s .

Этот способ вычислений достаточно сложен. Спрашивается, нет ли более простого приближения?

Крамер рекомендует вычислять m и s^2 по группированным значениям z и затем для s^2 использовать поправку Шеппарда. При этом все z из интервала (t_{i-1}, t_i) нужно считать сконцентрированными в средней точке этого интервала $(t_{i-1} + t_i)/2$. С помощью таких модифицированных значений z и нужно вычислять среднее m и квадратичное отклонение s . Для того чтобы можно было применить поправку Шеппарда, нужно, чтобы все интервалы имели одинаковую длину h . Оценки m и s , найденные по этому методу, зависят лишь от x_i . Вопрос о том, являются ли они асимптотически эффективными, насколько мне известно, еще не исследовался.

Если имеется очень много классов и середины соседних классов расположены очень близко друг от друга, то отличие разных оценок для среднего значения и квадратичного отклонения настолько мало, что не возникает вопроса о том, какие оценки принимать за основу.

При грубых расчетах, когда количество интервалов мало, в качестве оценок рекомендуется использовать m_0 и s_0 и применить критерий с $r - 1$ степенями свободы. Строго говоря, распределение χ^2 с $r - 1$ степенями свободы имело бы место лишь в том случае, если бы выражение

$$\chi^2 = \sum \frac{(x_i - np_i)^2}{np_i} \quad (34)$$

было вычислено с помощью истинных значений $p_i = p_i(\mu, \sigma)$. Эти истинные значения нам не известны, однако нам известны наилучшие приближения m_0 и s_0 для μ и σ . Величина $\chi^2(m_0, s_0)$, построенная с помощью этих приближений, будет, как правило, несколько меньше истинной величины χ^2 , но не настолько, чтобы число степеней свободы можно было считать равным $r - 3$. Если при вычислении границы для χ^2 воспользоваться $r - 1$ степенями свободы, то это, во всяком случае, увеличит надежность критерия¹.

Вопрос о наилучшем выборе количества классов r и граничных точек между классами t_1, \dots, t_{r-1} исследовался, в частности, Манном и Вальдсом², которые хотя и не решили этот вопрос окончательно, однако дали ряд полезных указаний. Согласно этим исследованиям, при $n = 200$ или 400, или 1000, классы нужно выбрать таким образом, чтобы в каждый класс попадало примерно 12 (соответственно 20 или 30) наблюдений. Если воспользоваться этой рекомендацией, то размеры классов окажутся значительно меньше обычно употребляемых; соответственно возрастет и вычислительная работа.

И. ПРОВЕРКА НОРМАЛЬНОСТИ КРИВОЙ ЭФФЕКТА

Пусть имеется r групп подопытных животных, и пусть n_1, n_2, \dots, n_r — количества животных в этих группах. Если животные подвергались действию некоторого вещества, причем логарифмы доз в соответствующих группах были равны l_1, l_2, \dots, l_r , и если в результате опытов стали известны количества животных x_1, x_2, \dots, x_r , реагировавших на эти дозы, то с помощью методов из § 54 можно попытаться подобрать такую нормальную кривую эффекта, которая в том или ином смысле соответствует

¹ Рекомендации автора относительно критерия $\chi^2(m_0, s_0)$ для проверки нормальности при небольшом числе классовых промежутков слишком неопределенны. Как показали Чернов и Леманн (Chernoff H. and Lehmann E. L., The use of maximum likelihood estimates in χ^2 tests for goodness of fit, *Ann. Math. Stat.*, 25, № 3 (1954), 579—586), случайная величина $\chi^2(m_0, s_0)$ асимптотически распределена как сумма $y_1^2 + y_2^2 + \dots + y_{r-3}^2 + \lambda_2 y_{r-2}^2 + \lambda_2 y_{r-1}^2$, где y_i — независимые и нормально распределенные величины, $\sum (y_i) = 0$, $D(y_i) = 1$; числа λ_1 и λ_2 лежат между нулем и единицей и зависят от параметров проверяемого закона и принятого способа подразделения. Если для данного уровня значимости α мы зададим критическую границу, исходя из закона распределения с $r - 3$ степенями свободы (как это часто делается на практике), то такой выбор может привести к серьезному преуменьшению вероятности ошибок первого рода. Но и выбор $r - 1$ степеней свободы, который рекомендуется автором, может привести к заметной потере мощности критерия. — *Прим. ред.*

² Очень хороший сводный отчет об этих исследованиях дал Кочрен (см. Cochran W. G., The χ^2 test of goodness of fit, *Ann. Math. Stat.*, 23, 315).

наблюденым частотам $h_i = x_i/n_i$. Для того чтобы проверить согласие между этой кривой эффекта и результатами наблюдений, нужно вычислить соответствующие вероятности p_1, \dots, p_r , а также их дополнения $q_i = 1 - p_i$ и образовать

$$\chi^2 = \sum \frac{(x_i - n_i p_i)^2}{n_i p_i} + \sum \frac{(y_i - n_i q_i)^2}{n_i q_i}, \quad (35)$$

где $y_i = n_i - x_i$ — количество животных из группы с номером i , которые не реагировали на дозу l_i . Так как снова

$$(x_i - n_i p_i) + (y_i - n_i q_i) = 0,$$

то χ^2 можно записать короче:

$$\chi^2 = \sum \frac{(x_i - n_i p_i)^2}{n_i p_i q_i}. \quad (36)$$

При этом постоянные L и s , определяющие положение и наклон кривой эффекта, нужно вычислять с помощью асимптотически эффективных методов, например по методам пробит-анализа или по методу минимума χ^2 (§ 51). Графическая оценка прямой эффекта в этом случае не является достаточной, так как величина χ^2 может получиться чрезмерно большой.

Число степеней свободы равно

$$f = 2r - r - 2 = r - 2,$$

так как наблюдается $2r$ количеств x_1, \dots, x_r и y_1, \dots, y_r , связанных r линейными уравнениями

$$x_i + y_i = n_i,$$

и два параметра L и s оцениваются по результатам наблюдений.

Если для одного и того же действующего вещества имеется несколько эмпирических кривых эффекта, то для каждой такой кривой можно вычислить χ_k^2 и результаты сложить. Сумма χ_1^2 (с f_1 степенями свободы) и χ_2^2 (с f_2 степенями свободы), в силу § 23, подчиняется распределению χ^2 с $f_1 + f_2$ степенями свободы.

Чем больше количество слагаемых, из которых складывается общая величина χ^2 , тем надежнее можно полагаться на асимптотическое распределение χ^2 ; это следует из центральной предельной теоремы (§ 24 Г).

Если найденные χ^2 , а также их сумма не превышают границ, за которыми нормальность заведомо отвергается, то тем не менее к гипотезе нормальности нужно относиться скептически. Только тогда, когда на обширном экспериментальном материале удастся установить, что величины χ^2 все время колеблются около своих средних значений f (f — число степеней свободы) и, следовательно, сумма всех χ^2 близка к сумме всех f , только тогда к гипотезе нормальности кривой эффекта можно относиться с несколько большим доверием.

К. НАСКОЛЬКО ВЕЛИКИ ДОЛЖНЫ БЫТЬ МАТЕМАТИЧЕСКИЕ ОЖИДАНИЯ np , ЧТОБЫ БЫЛО ПРИМЕНИМО РАСПРЕДЕЛЕНИЕ χ^2 ?

В литературе можно часто встретить замечания такого рода: применение распределения χ^2 допустимо лишь тогда, когда математические ожидания np по величине не менее 5 или 10. По-видимому, эти замечания диктуются лишь взглядами их авторов. Кочрен, а вместе с ним все те, кто исследовал этот вопрос точнее, пришли к более оптимистическому заключению¹.

Символом X^2 Кочрен обозначает ту дискретную случайную величину, которая применяется в критерии χ^2 :

$$X^2 = \sum \frac{(x - np)^2}{np}.$$

Непрерывную случайную величину, подчиняющуюся распределению χ^2 с тем же числом степеней свободы f , что и X^2 , он обозначает χ^2 и, точно так же как мы это делали в § 56 Б, сравнивает распределения X^2 и χ^2 . При этом особое внимание уделяется сравнению указанных распределений в той области изменения u , где вероятность P события $X^2 > u$ заключена между 0,01 и 0,05. Согласие оказывается достаточно хорошим, особенно если число степеней свободы не слишком мало. Если оно больше 6, то одно из математических ожиданий np может снижаться даже до $\frac{1}{2}$ и при этом согласие будет вполне удовлетворительным. При 60 степенях свободы или более и при малых математических ожиданиях точное значение вероятности P оказывается существенно меньше приближенного значения, вычисленного с помощью распределения χ^2 , так как дисперсия χ^2 больше дисперсии X^2 . Таким образом, применение распределения χ^2 лишь увеличивает надежность критерия. Если вместо распределения χ^2 воспользоваться нормальным распределением с точной дисперсией, вычисленной Холдейном², то приближение станет еще более лучшим.

Как показывает пример в § 56 Б, при двух степенях свободы математическое ожидание может снижаться до 2 единиц.

Только при одной степени свободы нужно соблюдать осторожность и при этом либо требовать, чтобы математические ожидания не были меньше 4, либо, что еще лучше, умножить X^2 на $(N - 1)/N$, где N — общее число наблюдений (см. § 9).

Л. ПРИМЕРЫ КРИТЕРИЯ χ^2

Пример 37. Тройные гибриды примулы (т. е. гибриды по трем наследственным признакам) скрещивались с представителями чистой линии,

¹ Cochran W. G., The χ^2 test, Ann. Math. Stat., 23, 328.

² Haldane J. B. S., Biometrika, 29, 133 и 31, 346.

у которых все три признака были рецессивными¹. В качестве наследственных признаков рассматривались:

Ch — oh: Китайская примула — звездная примула

G — g: Зеленый пестик — красный пестик

W — w: Белый венчик — желтый венчик

В 12 семействах² потомства были получены следующие распределения восьми возможных типов:

Тип	Номер семейства												Сумма
	107	110	119	121	122	127	129	131	132	133	135	178	
ChGW	12	17	9	10	24	9	3	16	20	9	11	10	150
ChGw	20	16	10	7	23	3	6	24	18	2	13	12	154
ChgW	14	10	6	8	19	5	5	23	18	10	7	12	137
Chgw	13	13	9	8	9	6	3	12	18	1	9	12	113
chGW	5	5	16	2	30	3	8	21	19	4	9	12	134
chGw	12	6	14	3	16	5	7	13	14	4	13	10	117
chGW	7	3	18	2	11	5	4	14	23	4	6	13	110
chgw	10	8	10	4	23	5	4	22	23	7	8	16	140
Сумма	93	78	92	44	155	41	40	145	153	41	76	97	1055
$\chi^2 =$	12,6	19,2	10,1	12,4	18,1	4,9	4,8	9,2	3,2	14,2	5,0	2,0	115,7

Если три наследственных признака не являются сцепленными и если факторы летальности и неуживчивости не играют никакой роли, то для каждого типа следует ожидать частоту, близкую к $1/8$. Общая сумма квадратов отклонений от математических ожиданий для всех семейств равна $\chi^2 = 115,7$. В каждом семействе имеется семь степеней свободы, всего, следовательно, 84. 5%-ная граница в случае 84 степеней свободы равна 106,4, следовательно, χ^2 превосходит эту границу. Кроме того, в трех семействах соответствующие величины χ^2 превосходят 5%-ную границу для семи степеней свободы, равную 14,1. Семейство № 110 превосходит даже 1%-ную границу 18,5. Таким образом, наблюдаемые частоты значительно отклоняются от закона Менделя.

Для того чтобы исследовать, какие из наследственных признаков ведут себя нерегулярно и имеются ли сцепленные признаки, мы, следуя Фишеру, разложим общую величину χ^2 на составные части, соответствующие отдельным наследственным признакам и парам признаков. Тогда будет видно, какие составные части особенно велики.

¹ Gregory, de Winton and Bateson, *Genetics of Primula Sinensis*, *J. of Genetics*, 13 (1923), 236. Статистический анализ излагается по книге Fisher R. A., *Statist. Methods for Research Workers*, 11 ed., Ex. 15, p. 101. (Терминология и основные понятия объяснены в примере 32, § 46, где шла речь о двойных гибридах. — *Прим. перев.*)

² Семейства 54, 55, 58 и 59 из этой таблицы исключены, так как числа, указанные Фишером, не совпадают с теми данными, которые указаны в журнале — *J. of Genetics*, 13.

Если x_1, \dots, x_8 — наблюдаемые количества в одном семействе ($\sum x = n$), то для этого семейства

$$\chi^2 = \sum \frac{\left(x - \frac{1}{8}n\right)^2}{\frac{1}{8}n} = \frac{8}{n} \left(\sum x^2 - \frac{n^2}{8}\right).$$

С помощью ортогонального преобразования вместо x_1, \dots, x_8 введем новые переменные y_1, \dots, y_8 таким образом, чтобы величина $y_1 \sqrt{8}$ равнялась разности между числом потомков вида Ch и числом потомков вида ch:

$$y_1 \sqrt{8} = z_1 = x_1 + x_2 + x_3 + x_4 - x_5 - x_6 - x_7 - x_8. \quad (\text{Ch})$$

Точно так же определяются y_2, y_3 , соответствующие наследственным признакам G — g и W — w:

$$y_2 \sqrt{8} = z_2 = x_1 + x_2 - x_3 - x_4 + x_5 + x_6 - x_7 - x_8, \quad (\text{G})$$

$$y_3 \sqrt{8} = z_3 = x_1 - x_2 + x_3 - x_4 + x_5 - x_6 + x_7 - x_8. \quad (\text{W})$$

Следующая переменная y_4 соответствует парному признаку (GW, gw) — (gw, Gw):

$$y_4 \sqrt{8} = z_4 = x_1 - x_2 - x_3 + x_4 + x_5 - x_6 - x_7 + x_8. \quad (\text{GW})$$

Если наследственные признаки G — g и W — w не связаны, то математическое ожидание z_4 должно быть равно нулю.

Аналогично определяются z_5 и z_6 :

$$y_5 \sqrt{8} = z_5 = x_1 - x_2 + x_3 - x_4 - x_5 + x_6 - x_7 + x_8, \quad (\text{ChW})$$

$$y_6 \sqrt{8} = z_6 = x_1 + x_2 - x_3 - x_4 - x_5 - x_6 + x_7 + x_8. \quad (\text{ChG})$$

Для того чтобы ортогональное преобразование было полным, нужно ввести еще две переменные:

$$y_7 \sqrt{8} = z_7 = x_1 - x_2 - x_3 + x_4 - x_5 + x_6 + x_7 - x_8,$$

$$y_8 \sqrt{8} = z_8 = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8.$$

Величина z_7 не имеет сколько-нибудь простого биологического истолкования; $z_8 = n$ — количество растений в семействе.

Практически, для того чтобы избежать деления на $\sqrt{8}$, все вычисления проводят, конечно, с величинами z , а не с y . Величина χ^2 выражается через z_i следующим образом:

$$\chi^2 = \frac{8}{n} \left(\sum y^2 - \frac{n^2}{8}\right) = \frac{1}{n} \cdot \left(\sum z^2 - z_8^2\right) = \frac{1}{n} z_1^2 + \dots + \frac{1}{n} z_7^2.$$

Эта формула представляет собой разложение χ^2 на составные части, о которых говорилось выше. Каждая случайная величина z_k распределена приближенно нормально с нулевым средним и дисперсией n . Следовательно, каждое слагаемое z_k^2/n распределено приближенно, как χ^2 с одной степенью свободы. С помощью вычислений получаем для этих слагаемых следующие значения.

Семейство	(Ch)	(G)	(W)	(GW)	(ChW)	(ChG)	(z ₇)	Сумма
107	6,72	0,27	3,11	1,82	0,10	0,27	0,27	12,56
110	14,82	1,28	0,82	0,82	0,20	1,28	0	19,22
119	6,26	0,39	0,39	0,17	2,13	0,04	0,70	10,08
121	11,00	0	0	0,36	0,82	0,09	0,09	12,36
122	0,16	6,20	1,09	1,86	0,52	0,32	7,90	18,05
127	0,61	0,02	0,22	0,61	1,20	0,22	1,98	4,86
129	0,90	1,60	0	0,40	0,10	0,90	0,90	4,80
131	0,17	0,06	0,06	0,06	0,06	0,34	8,45	9,20
132	0,16	0,79	0,32	0,32	0,06	1,47	0,06	3,18
133	0,22	0,22	4,12	0,02	8,80	0,22	0,61	14,21
135	0,21	3,37	1,32	0,05	0,05	0	0,05	5,05
178	0,26	0,84	0,09	0,09	0,01	0,26	0,50	2,05
Сумма	41,49	15,04	11,54	6,58	14,05	5,41	21,51	115,62

1%-ная граница для отдельного значения равна 6,6 (одна степень свободы), для суммы значений в отдельном столбце — 26,2 (12 степеней свободы). Соответствующие 5%-ные границы равны 3,8 и 21,0. Числа, превосходящие 5%-ную границу, выделены в таблице жирным шрифтом. Три наименьших жирных числа — 6,26 в столбце (Ch), 6,20 в столбце (G) и 4,12 в столбце (W) — еще ни о чем не говорят, так как даже в том случае, когда всё в порядке, среди 84 чисел в среднем 4 будут превосходить 5%-ную границу. Все остальные случаи превышения 5%-ной границы сосредоточены в столбцах (Ch), (Ch W) и (z₇). В этих столбцах 1%-ная граница превышает шесть раз, а сумма чисел в столбце (Ch) превосходит даже 0,1%-ную границу 32,9. Следовательно, признак Ch — ch заведомо ведет себя ненормальным образом, и большая часть отклонений падает именно на этот признак. Вполне возможно, что признак Ch — ch связан с фактором летальности или неуживчивости.

Сцепления между какими-либо двумя из трех признаков — Ch — ch, G — g и W — w, по-видимому, не имеется, так как суммы в столбцах (GW), (Ch W) и (Ch G) не особенно велики.

Пример 38 (из книги Г. Крамера, Математические методы статистики, ИЛ, М., 1948, стр. 477). Юхансен измерял толщину 12 000 бобов. Результаты измерений были разбиты на 16 классов. К первому классу были отнесены бобы с толщиной менее 7 мм, ко второму — бобы с толщиной от 7 до 7,25 мм и т. д. (верхняя граница каждого последующего класса всегда на 0,25 мм больше верхней границы предыдущего класса). Количества бобов x_1, x_2, \dots, x_{16} в этих классах указаны во втором столбце таблицы, следующей ниже. Для того чтобы проверить, подчиняется ли толщина бобов нормальному распределению, сначала по группированным данным были вычислены m и s с шепардовской поправкой, причем оба конечных интервала (до 7,00 и свыше 10,50) были разбиты на частичные интервалы длины 0,25. В результате вычислений получились оценки

$$m = 8,512, \quad s = 0,6163.$$

По этим m и s была построена функция нормального распределения, с помощью которой были найдены оценки np_i для математических ожиданий (столбец 3). Разности $x_i - np_i$ указаны в столбце 4. Оказалось, что вели-

Классы	x	np	$x - np$
До 7,00	32	68	- 36
7,00— 7,25	103	132	- 29
7,25— 7,50	239	310	- 71
7,50— 7,75	624	617	+ 7
7,75— 8,00	1 187	1 046	+141
8,00— 8,25	1 650	1 506	+144
8,25— 8,50	1 883	1 842	+ 41
8,50— 8,75	1 930	1 920	+ 10
8,75— 9,00	1 638	1 698	- 60
9,00— 9,25	1 130	1 277	-147
9,25— 9,50	737	817	- 80
9,50— 9,75	427	444	- 17
9,75—10,00	221	205	+ 16
10,00—10,25	110	81	+ 29
10,25—10,50	57	27	+ 30
Свыше 10,50	32	10	+ 22
Сумма	12 000	12 000	0

чина χ^2 равна 196,5, в то время как 0,1%-ная граница для 13 степеней свободы равна 34,5. Следовательно, с большой уверенностью можно утверждать, что распределение толщины бобов не является нормальным. Разности $x - np$ показывают, что это распределение обладает значительной асимметрией: очень толстых бобов больше, а очень тонких бобов меньше, чем полагалось бы при нормальном распределении.

§ 57. Критерий, основанный на дисперсионном отношении (критерий F)

Пусть s_1^2 и s_2^2 — две независимые оценки для дисперсий σ_1^2 и σ_2^2 соответственно. Как проверить гипотезу $\sigma_1 = \sigma_2$?

Если s_1^2 и s_2^2 получены с помощью известной формулы по n_1 и n_2 наблюдениям соответственно и если отдельные наблюдения независимы и распределены нормально, то случайная величина

$$\chi_1^2 = \frac{(n_1 - 1) s_1^2}{\sigma_1^2} \quad (1)$$

подчиняется распределению χ^2 с $f_1 = n_1 - 1$ степенями свободы, а случайная величина

$$\chi_2^2 = \frac{(n_2 - 1) s_2^2}{\sigma_2^2} \quad (2)$$

подчиняется распределению χ^2 с $f_2 = n_2 - 1$ степенями свободы. Если $\sigma_1 = \sigma_2$, то

$$\frac{\chi_1^2}{\chi_2^2} = \frac{f_1 s_1^2}{f_2 s_2^2}. \quad (3)$$

Для проверки гипотезы $\sigma_1 = \sigma_2$ вычисляют отношение выборочных дисперсий (дисперсионное отношение)

$$F = \frac{s_1^2}{s_2^2}. \quad (4)$$

Если отношение F превосходит границу F_β , то гипотеза $\sigma_1 = \sigma_2$ отвергается. Это правило¹ называют критерием F . Граница F_β выбирается так, чтобы в том случае, когда гипотеза $\sigma_1 = \sigma_2$ верна, вероятность события $F > F_\beta$ в точности равнялась заданному уровню значимости β .

Для того чтобы вычислить F_β , мы должны найти функцию распределения случайной величины F при условии, что гипотеза $\sigma_1 = \sigma_2$ верна. Эта функция будет известна, если мы найдем функцию распределения $H(w)$ отнесения

$$\frac{\chi_1^2}{\chi_2^2} = \frac{f_1 s_1^2}{f_2 s_2^2} = \frac{f_1}{f_2} F. \quad (5)$$

Плотность вероятности для χ_1^2 задается формулой

$$g_1(t) = \alpha_1 t^{\frac{1}{2} f_1 - 1} e^{-\frac{1}{2} t}, \quad \text{где} \quad \alpha_1 = \frac{1}{\Gamma\left(\frac{f_1}{2}\right) 2^{\frac{f_1}{2}}}.$$

Для плотности χ_2^2 формула будет аналогичной. Следовательно, вероятность того, что отношение (5) окажется меньше w , равна

$$H(w) = \alpha_1 \alpha_2 \int \int t^{\frac{1}{2} f_1 - 1} e^{-\frac{1}{2} t} u^{\frac{1}{2} f_2 - 1} e^{-\frac{1}{2} u} dt du, \quad (6)$$

где интегрирование производится по области

$$t > 0, \quad u > 0, \quad \frac{t}{u} < w.$$

Двойной интеграл (6) можно представить в виде двух последовательных интегралов:

$$H(w) = \alpha_1 \alpha_2 \int_0^{uw} du \int_0^1 t^{\frac{1}{2} f_1 - 1} u^{\frac{1}{2} f_2 - 1} e^{-\frac{1}{2} t - \frac{1}{2} u} dt. \quad (7)$$

¹ Р. А. Фишер строил критерий с помощью статистики $z = (\ln F)/2$.

Если ввести новую переменную интегрирования $y = t/u$ и положить $f_1 + f_2 = f$, то получим

$$H(w) = \alpha_1 \alpha_2 \int_0^{\infty} du \int_0^w u^{\frac{1}{2}f-1} y^{\frac{1}{2}f_1-1} e^{-\frac{1}{2}uy - \frac{1}{2}u} dy \quad (8)$$

или, меняя порядок интегрирования,

$$H(w) = \alpha_1 \alpha_2 \int_0^w y^{\frac{1}{2}f_1-1} dy \int_0^{\infty} u^{\frac{1}{2}f-1} e^{-\frac{y+1}{2}u} du. \quad (9)$$

Внутренний интеграл представляет собой гамма-функцию:

$$\int_0^{\infty} u^{\frac{1}{2}f-1} e^{-\frac{y+1}{2}u} du = \left(\frac{2}{y+1}\right)^{\frac{f}{2}} \Gamma\left(\frac{f}{2}\right). \quad (10)$$

Таким образом,

$$H(w) = C \int_0^w y^{\frac{1}{2}f_1-1} (y+1)^{-\frac{1}{2}f} dy, \quad (11)$$

где

$$C = \frac{\Gamma\left(\frac{f}{2}\right)}{\Gamma\left(\frac{f_1}{2}\right) \Gamma\left(\frac{f_2}{2}\right)}. \quad (12)$$

Интеграл (11) является неполной бета-функцией, и при целочисленных f_1 и f его, очевидно, можно вычислить элементарно. Таким образом, функция распределения $H(w)$ известна.

Подстановкой

$$w = \frac{f_1}{f_2} w' \quad (13)$$

получим из $H(w)$ функцию распределения $G(w')$ для отношения F . Искомая граница $w' = F_{\beta}$ определяется как решение уравнения

$$G(w') = 1 - \beta. \quad (14)$$

Эта граница, помимо β , зависит также от f_1 и f_2 . Границы F_{β} для $\beta = 0,05$ и $0,01$ указаны в табл. 8А и 8Б.

Пример 39. В США в 30 лабораториях производился анализ газов. В каждой лаборатории делалось несколько анализов (в большинстве слу-

чаев 10), отдельные результаты которых были опубликованы М. Шеффердом¹.

Если по этим результатам попытаться вычислить дисперсию внутри лабораторий², то возникнет затруднение, связанное с тем, что внутри отдельных лабораторий выборочные дисперсии s^2 сильно отличаются друг от друга: имеются более хорошие и менее хорошие лаборатории. Если желательно вычислить среднюю выборочную дисперсию s^2 для хороших и средних лабораторий с тем, чтобы потом ее можно было принять за норму качества анализа для всех лабораторий, то совсем плохие лаборатории нужно из этого осреднения исключить. Однако те s^2 , которые лишь случайно оказались несколько больше всех остальных, исключать не следует, так как в противном случае среднее будет иметь отрицательную систематическую ошибку.

Для исключения больших s^2 можно применить критерий F . В качестве примера рассмотрим определение количества метана так называемым «методом сжигания А», который применялся большинством лабораторий.

В некоторых лабораториях анализ проводился двумя различными исследователями. При этом оказалось, что различие результатов обоих исследователей, как правило, несколько больше различия результатов, полученных отдельным исследователем. Поэтому для оценки дисперсии нужно выделить результаты каждого исследователя из общей массы наблюдений и с их помощью вычислить выборочные дисперсии по формуле

$$s^2 = \frac{Q}{n-1}, \text{ где } Q = \sum (x - \bar{x})^2,$$

где x — процентное содержание метана.

Полученные результаты указаны ниже в порядке возрастания выборочной дисперсии s^2 :

№	Q	n-1	s ²	№	Q	n-1	s ²
1	0,0	1	= 0,00	16	3,6	5	= 0,72
2	0,6	9	= 0,07	17	6,7	9	= 0,74
3	0,8	9	= 0,09	18	7,3	9	= 0,81
4	0,4	4	= 0,10	19	4,0	4	= 1,00
5	1,3	9	= 0,14	20	8,3	8	= 1,04
6	0,6	4	= 0,15	21	9,1	8	= 1,14
7	0,6	4	= 0,15	22	18,1	15	= 1,20
8	1,5	9	= 0,17	23	5,9	4	= 1,47
9	1,8	9	= 0,20	24	13,8	9	= 1,53
10	2,2	10	= 0,22	25	10,2	6	= 1,70
11	0,9	4	= 0,23	26	21,1	9	= 2,34
12	3,5	9	= 0,39	27	28,5	9	= 3,17
13	3,8	9	= 0,42	28	29,9	9	= 3,32
14			= 0,60	29	16,9	4	= 4,23
15			= 0,65	30	43,4	9	= 4,82
				31	15,2	2	= 7,60

Как видно, результаты исследователя 31 имеют значительно большую дисперсию, чем результаты всех остальных исследователей. Для того

¹ Shepherd M., J. Res. Nat. Bureau of Standards, 38 (1947), 19.

² Т. е. дисперсию результатов анализа в отдельной лаборатории — Прим. перев.

чтобы проверить, не является ли это случайностью, мы разделим выборочную дисперсию $s_1^2 = 7,60$ на среднюю дисперсию всех остальных наблюдений, вычисленную по формуле

$$s_2^2 = \frac{\sum Q}{\sum (n-1)}.$$

В результате получим $s_2^2 = 239,9/215 = 1,116$ и

$$F_{31} = \frac{s_1^2}{s_2^2} = \frac{7,60}{1,116} = 6,81.$$

Точно таким же образом можно вычислить F_1, F_2, \dots, F_{30} исследователю j при сравнении со всеми остальными ставится в соответствие отношение F_j ($j = 1, \dots, 31$). Наибольшим из этих F_j является F_{31} .

Если отвлечься от остальных результатов, то окажется, что F_{31} превосходит 5%-ную границу 3,04. Однако, хотя вероятность того, что определенное отдельное F_j превзойдет F_{β} , и равна 0,05, тем не менее для F_{31} мы не имеем права делать такое заключение: значение F_{31} является наибольшим из всех F_j , и событие, при котором хотя бы одно F_j превзойдет F_{β} , отнюдь не маловероятно. Если, в действительности, все σ^2 равны друг другу, то вероятность того, что все F_j будут меньше F_{β} , равна¹ $(0,95)^{31} = 0,20$, следовательно, вероятность того, что хотя бы одно F_j будет больше F_{β} , равна

$$1 - 0,20 = 0,80.$$

То, что F_{31} превосходит 1%-ную границу, также еще почти ни о чем не говорит, так как вероятность случайного осуществления этого события

$$1 - (0,99)^{31} = 0,27$$

все еще велика. Если бы F_{31} превзошло 0,1%-ную границу, то наличие грубых ошибок у 31-го исследователя было бы доказано, но так как эта граница равна 7,15, то F_{31} ее не превышает.

Более благоприятное положение складывается при проверке F_{30} , так как F_{30} имеет большее число степеней свободы, чем F_{31} . Находим

$$F_{30} = \frac{4,82}{1,066} = 4,53.$$

Эта величина значительно превосходит 0,1%-ную границу, равную 3,26. Вероятность того, что это событие произошло случайно, равна² лишь $1 - (0,999)^{31} = 0,03$. Следовательно, гипотезу о том, что результаты исследователя 30 имеют такую же дисперсию σ^2 , как и результаты остальных исследователей, следует отвергнуть.

Исключив исследователя 30, можно снова сравнить точность исследователя 31 со всеми остальными (с 1 по 29). В данном случае находим

$$F'_{31} = \frac{7,60}{1,00} = 7,60.$$

¹ Это утверждение было бы справедливо, если бы все F_j были независимы. На самом деле они являются зависимыми, так как связаны соотношением $\sum [v_i F_i / (1 + v_i F_i)] = 1$, где $v_i = (n_i - 1) / \sum_{j \neq i} (n_j - 1)$. — Прим. перев.

² Здесь автор снова допускает ошибку, так как вероятность события $F_{30} > 3,26$ не равна 0,03 (см. предыдущую сноску). — Прим. перев.

Эта величина превосходит 0,1%-ную границу, равную 7,15, поэтому исследователя 31 также надо исключить.

После исключения исследователей 30 и 31 можно продолжать проверку с помощью этого же метода. В результате последовательно будут исключены исследователи 28, 27, 29 и 26. Вероятность ошибочного исключения во всех случаях менее 0,03, следовательно, вероятность того, что хотя бы одно исключение из шести было ошибочным, не превосходит 0,18.

На первый взгляд заключение с вероятностью ошибки 0,18 может показаться неосторожным. Однако более точная проверка показывает, что большинство заключений остаются справедливыми также и при 0,05%-ной границе. В результате оказывается, что выборочное среднее значение содержания метана у исследователей 29 и 30 слишком велико, а у исследователя 31 слишком мало. Следовательно, исключение номеров 29, 30 и 31 не было ошибочным. При исключении трех остальных исследователей вероятность ошибки каждый раз оказывается меньше 0,014, а общая вероятность ошибки — меньше 0,042. Если ошибку, вероятность которой равна 0,05, считать допустимой, то исключение шести исследователей: 31, 30, 29, 28, 27 и 26, следует признать оправданным¹.

Средняя выборочная дисперсия на одного исследователя, вычисленная по оставшимся данным (с 1 по 25), равна

$$s^2 = \frac{\sum Q}{\sum(n-1)} = \frac{110,1}{175} = 0,63.$$

Величина s^2 представляет собой несмещенную оценку дисперсии σ^2 одного измерения и не отражает колебания выборочных дисперсий для отдельных исследователей, поэтому s^2 можно назвать *дисперсией воспроизводимости*². Величина s равна 0,8 (% метана).

Следует отметить, что даже в том случае, когда исключены показания исследователей, допускающих наибольшие отклонения, результаты измерений, произведенных в различных лабораториях, могут отличаться друг от друга на величину, почти вдвое превосходящую³ s . Для квадратичного

¹ Для проверки гипотезы об однородности ряда дисперсий в данном случае следовало бы применить специальные критерии (см. Халд А., Математическая статистика с техническими приложениями, ИЛ, 1956, гл. XI, § 6). — *Прим. ред.*

² Оценка s^2 представляет собой среднее взвешенное выборочных дисперсий для отдельных исследователей: $s^2 = [\sum (n_i - 1) s_i^2] / \sum (n_i - 1)$. Если гипотеза о равенстве дисперсий верна, то $\sum (n_i - 1) s_i^2 / \sigma^2$ подчиняется распределению χ^2 с $f = \sum (n_i - 1)$ степенями свободы. Это позволяет построить доверительный интервал для σ . Например, в данном случае 95%-ный интервал задается неравенством $0,72 < \sigma < 0,90$. — *Прим. перев.*

³ Пусть x_1 и x_2 — результаты измерений в двух различных лабораториях. По предположению, x_1 и x_2 — независимые нормальные величины с дисперсией σ^2 , поэтому разность $x_1 - x_2$ распределена нормально с нулевым средним и дисперсией $2\sigma^2$. Таким образом, $P(|x_1 - x_2| > 1,5s) = 2[1 - \Phi(1,5s/\sigma\sqrt{2})]$. Согласно предыдущей сноске, с вероятностью 0,975 имеет место неравенство $s/\sigma < 1,1$, поэтому $P(|x_1 - x_2| > 1,5s) > 2[1 - \Phi(1,5 \cdot 1,1/\sqrt{2})] = 2[1 - \Phi(1,17)] = 0,24$. Полученный результат позволяет с вероятностью 0,975 утверждать, что не менее чем в четверти всех случаев результаты измерений в двух различных лабораториях будут отличаться друг от друга более чем на 1,5s. — *Прим. перев.*

отклонения результатов различных лабораторий получилась оценка $S = 1,4$.

Сравнение с более точными методами измерений показывает, что систематическая ошибка «метода сжигания» ужасающе велика и достигает 2,6. Следовательно, метод сжигания A не очень надежен.

§ 58. Дисперсионный анализ

А. ДИСПЕРСИЯ ВНУТРИ КЛАССОВ И МЕЖДУ КЛАССАМИ

Пусть $x_1, \dots, x_n, y_1, \dots, y_n$ и z_1, \dots, z_n — независимые, нормально распределенные случайные величины, причем все x_i наблюдаются в одинаковых экспериментальных условиях, и поэтому можно предположить, что x_i имеют одинаковые средние значения и одинаковые дисперсии. Точно такие же предположения мы будем делать относительно y_j и z_k . Далее, предположим, что все x_i, y_j и z_k имеют одинаковую дисперсию σ^2 (эту гипотезу можно проверить с помощью критерия F). Возникает вопрос, одинаковы ли средние значения у x_i, y_j и z_k ?

Для ответа на этот вопрос можно использовать метод дисперсионного анализа. Идея этого метода основана на том, что если x, y и z имеют одинаковое среднее значение, то общая сумма квадратов

$$Q = \sum (x - M)^2 + \sum (y - M)^2 + \sum (z - M)^2, \quad (1)$$

где M — общее выборочное среднее значение всех x_i, y_j и z_k , распадается на две составные части, из которых первая связана с оценкой дисперсии *внутри трех классов*, а вторая — с оценкой дисперсии *между классами*. Эти две составные части сравниваются затем друг с другом посредством критерия F .

Математическим вспомогательным средством, позволяющим получить такое разложение, является ортогональное преобразование. Введем вместо x_1, \dots, x_n новые переменные u_1, \dots, u_n , из которых первое пропорционально арифметическому среднему x :

$$u_1 = \frac{x_1 + \dots + x_n}{\sqrt{n}} = \bar{x} \sqrt{n}, \quad (2)$$

а остальные переменные u_2, \dots, u_n определяются таким образом, чтобы преобразование было ортогональным. Тогда, в силу ортогональности,

$$\sum x^2 = \sum u^2 = u_1^2 + \dots + u_n^2, \quad (3)$$

следовательно,

$$u_2^2 + \dots + u_n^2 = \sum x^2 - u_1^2 = \sum x^2 - n\bar{x}^2 = \sum (x - \bar{x})^2. \quad (4)$$

Таким образом, данная часть суммы квадратов соответствует дисперсии в классе всех x .

Точно так же, заменив y_1, \dots, y_n и z_1, \dots, z_n новыми переменными v_1, \dots, v_n и w_1, \dots, w_n , получим

$$v_2^2 + \dots + v_n^2 = \sum (y - \bar{y})^2. \quad (5)$$

$$w_2^2 + \dots + w_n^2 = \sum (z - \bar{z})^2. \quad (6)$$

Складывая (4), (5) и (6), найдем

$$\begin{aligned} Q_2 &= (u_2^2 + \dots + u_n^2) + (v_2^2 + \dots + v_n^2) + (w_2^2 + \dots + w_n^2) = \\ &= \sum (x - \bar{x})^2 + \sum (y - \bar{y})^2 + \sum (z - \bar{z})^2. \end{aligned} \quad (7)$$

Сумму квадратов Q_2 можно использовать для оценки дисперсии внутри классов. Как всегда, такой оценкой является

$$s_2^2 = \frac{Q_2}{3(n-1)}. \quad (8)$$

Если три класса состоят из различных количеств наблюдений n_1, n_2, n_3 , то вместо (8) получится формула

$$s_2^2 = \frac{Q_2}{(n_1-1) + (n_2-1) + (n_3-1)}$$

или, короче,

$$s_2^2 = \frac{Q_2}{N-3}, \quad (9)$$

где $N = n_1 + n_2 + n_3$.

Для того чтобы найти «дисперсию между классами», u_1, v_1, w_1 подвергаются новому ортогональному преобразованию, при котором они переходят в u', v', w' , где u' пропорционально общему выборочному среднему M всех x_i, y_j и z_k :

$$u' = M \sqrt{N} = \frac{\sum x + \sum y + \sum z}{\sqrt{N}} = \frac{u_1 \sqrt{n_1} + v_1 \sqrt{n_2} + w_1 \sqrt{n_3}}{\sqrt{N}}. \quad (10)$$

В силу результатов § 13, такое преобразование возможно, так как сумма квадратов коэффициентов в правой части (10) равна единице:

$$\frac{n_1}{N} + \frac{n_2}{N} + \frac{n_3}{N} = 1.$$

Так как преобразование ортогонально, то

$$u_1^2 + v_1^2 + w_1^2 = u'^2 + v'^2 + w'^2. \quad (11)$$

Поэтому, положив $v'^2 + w'^2 = Q_1$, получим

$$Q_1 = u_1^2 + v_1^2 + w_1^2 - u'^2 = u_1^2 + v_1^2 + w_1^2 - NM^2. \quad (12)$$

Если Q_1 сложить с ранее полученной суммой

$$Q_2 = (u_2^2 + \dots + u_n^2) + (v_2^2 + \dots + v_n^2) + (u_2^2 + \dots + u_n^2),$$

то найдем, что

$$\begin{aligned} Q_1 + Q_2 &= \sum u_i^2 + \sum v_j^2 + \sum u_k^2 - NM^2 = \\ &= \sum x_i^2 + \sum y_j^2 + \sum z_k^2 - NM^2 = \\ &= \sum (x - M)^2 + \sum (y - M)^2 + \sum (z - M)^2. \end{aligned} \quad (13)$$

Правую часть (13) мы уже раньше обозначили буквой Q . Таким образом, разложение

$$Q = Q_1 + Q_2, \quad (14)$$

о котором говорилось выше, найдено.

Вторая составная часть Q_2 определяет оценку (9) для дисперсии внутри классов. Первая составная часть

$$Q_1 = v^2 + w^2 \quad (15)$$

зависит лишь от разностей между выборочными средними значениями \bar{x} , \bar{y} и \bar{z} внутри классов. А именно

$$\begin{aligned} Q_1 &= u_1^2 + v_1^2 + u_1^2 - NM^2 = \\ &= n_1 \bar{x}^2 + n_2 \bar{y}^2 + n_3 \bar{z}^2 - NM^2, \end{aligned} \quad (16)$$

где M — взвешенное среднее, составленное из \bar{x} , \bar{y} и \bar{z} с весами n_1 , n_2 и n_3 :

$$M = \frac{n_1 \bar{x} + n_2 \bar{y} + n_3 \bar{z}}{N}. \quad (17)$$

Поэтому вместо (16) можно записать

$$Q_1 = n_1 (\bar{x} - M)^2 + n_2 (\bar{y} - M)^2 + n_3 (\bar{z} - M)^2. \quad (18)$$

В теории наименьших квадратов к такому же выражению (18) приходят тогда, когда \bar{x} , \bar{y} и \bar{z} представляют собой неравноточные оценки для неизвестного параметра ϑ . Так как \bar{x} , \bar{y} и \bar{z} вычисляются по n_1 , n_2 и n_3 наблюдениям соответственно, то их нужно снабдить весами n_1 , n_2 и n_3 и вычислить взвешенное среднее (17). Согласно теории наименьших квадратов, «выборочная дисперсия одного наблюдения на единицу веса» равна

$$s_1^2 = \frac{n_1 (\bar{x} - M)^2 + n_2 (\bar{y} - M)^2 + n_3 (\bar{z} - M)^2}{3 - 1} = \frac{Q_1}{r - 1}.$$

Знаменатель равен числу классов, уменьшенному на единицу; следовательно, в случае r классов, в знаменателе должно стоять $r - 1$.

Если предположить, что все три класса x , y и z имеют не только равные дисперсии, но также и равные средние значения, то, согласно теории наименьших квадратов,

$$s_1^2 = \frac{Q_1}{r-1} \quad (19)$$

будет являться несмещенной оценкой для общей дисперсии σ^2 .

Независимо от теории наименьших квадратов мы снова покажем, что оценка (19) является несмещенной, т. е. покажем, что среднее значение Q_1 равно $(r-1)\sigma^2$ (в нашем случае это среднее значение равно $2\sigma^2$).

Пусть \bar{v} — общее среднее значение для x , y и z . Введением новых переменных $x - \bar{v}$, $y - \bar{v}$ и $z - \bar{v}$ можно добиться, чтобы общее среднее значение стало равным нулю. В этом случае средние значения x_i^2 , y_j^2 и z_k^2 будут равны σ^2 , а средние значения всех остальных произведений $x_i x_j$, $x_i y_j$ и т. д. будут равны нулю. Так как при ортогональном преобразовании указанные свойства средних значений для квадратов и произведений сохраняются, то средние значения u_i^2 , v_j^2 , w_k^2 и u'^2 , v'^2 , w'^2 также равны σ^2 . Следовательно, среднее значение (15) равно $2\sigma^2$, что и требовалось доказать.

С помощью (7) и (8) точно так же можно показать, что среднее значение s_2^2 равно σ^2 . Этот результат, разумеется, остается справедливым и в том случае, когда математические ожидания x , y и z различны, так как замена переменных

$$x'_i = x_i - a, \quad y'_j = y_j - b, \quad z'_k = z_k - c$$

не оказывает влияния на s_2^2 .

Наиболее удобными формулами для вычисления Q_1 и Q_2 являются (18) и (7). Для контроля можно воспользоваться (13) или формулой

$$Q_1 + Q_2 = Q = \sum(x - a)^2 + \sum(y - a)^2 + \sum(z - a)^2 - N(M - a)^2, \quad (20)$$

где a — произвольное число.

Если Q_1 и Q_2 уже вычислены, то для получения оценок s_1^2 и s_2^2 нужно лишь Q_1 и Q_2 разделить на соответствующее «число степеней свободы»:

$$s_1^2 = \frac{Q_1}{r-1}, \quad s_2^2 = \frac{Q_2}{N-r}. \quad (21)$$

Б. КРИТЕРИЙ F

Если величина s_1^2 меньше или лишь немногим больше величины s_2^2 , то нет оснований считать средние значения в классах различными. Однако если s_1^2 значительно превосходит s_2^2 , то возникает

подозрение, что эти истинные средние значения различны. Чтобы исследовать, насколько это подозрение является обоснованным, составим отношение

$$F = \frac{s_1^2}{s_2^2} = \frac{(N-r)Q_1}{(r-1)Q_2} \quad (22)$$

и применим критерий F .

Для того чтобы получить точную формулу для функции распределения отношения F , нужно, конечно, воспользоваться предположением, согласно которому все x_i , y_j и z_k независимы и распределены одинаково нормально. В этом случае u' , v' , w' и $u_2, v_2, w_2, \dots, u_n, v_n, w_n$ будут также независимыми нормально распределенными случайными величинами с одинаковой дисперсией σ^2 , так как плотность вероятности

$$C e^{-\frac{1}{2\sigma^2} [\sum (x-\bar{x})^2 + \sum (y-\bar{y})^2 + \sum (z-\bar{z})^2]}$$

при ортогональном преобразовании переменных переходит в плотность вероятности того же вида:

$$C e^{-\frac{1}{2\sigma^2} [(u'-\bar{u}')^2 + v'^2 + w'^2 + u_2^2 + v_2^2 + \dots + w_n^2]} \quad (23)$$

Следовательно,

$$\frac{Q_1}{\sigma^2} = \frac{v'^2 + w'^2}{\sigma^2} \quad (24)$$

подчиняется распределению χ^2 с

$$f_1 = 2 \quad (\text{в общем случае } f_1 = r - 1) \quad (25)$$

степенями свободы, и точно так же

$$\frac{Q_2}{\sigma^2} = \frac{u_2^2 + v_2^2 + w_2^2 + \dots + u_n^2 + v_n^2 + w_n^2}{\sigma^2} \quad (26)$$

подчиняется распределению χ^2 с

$$f_2 = N - 3 \quad (\text{в общем случае } f_2 = N - r)$$

степенями свободы. Оба отношения (24) и (26) независимы, так как плотность вероятности (23) представляет собой произведение двух сомножителей, из которых первый зависит лишь от u' , v' , w' , а второй — лишь от u_2, v_2, \dots, w_n . Таким образом,

Если все x_i, y_j и z_k независимы и распределены одинаково нормально со средним значением \bar{x} и дисперсией σ^2 , то отношение (22) подчиняется распределению F . Следовательно, для проверки гипотезы о равенстве средних можно применить критерий F . Число степеней

свободы в числителе и знаменателе равно $r - 1$ и $N - r$ соответственно.

В приложениях целесообразно составлять таблицу следующего вида:

	Сумма квадратов	Число степеней свободы	Оценка дисперсии
Дисперсия между классами	Q_1	$f_1 = r - 1$	$Q_1 : f_1 = s_1^2$
Дисперсия внутри классов	Q_2	$f_2 = N - r$	$Q_2 : f_2 = s_2^2$
Дисперсия по всем наблюдениям	Q	$N - 1$	$Q : (N - 1) = s^2$

В. РАСПРЕДЕЛЕНИЯ, ОТЛИЧНЫЕ ОТ НОРМАЛЬНОГО

Можно ли применять критерий F в том случае, когда распределение x , y и z отлично от нормального? При исследовании этого вопроса мы предположим, что в каждом классе количество величин n не слишком мало, например, $n \geq 4$. При этом предположении $f_1 = r - 1$ будет существенно меньше, чем $f_2 = (n - 1)r$. Отсюда следует, что относительная ошибка оценки s_1^2 подвержена значительно большим случайным колебаниям, чем относительная ошибка оценки s_2^2 . Поэтому функция распределения отношения F зависит главным образом от распределения числителя. Как видно из формул (18) и (19), числитель отношения F зависит лишь от \bar{x} , \bar{y} и \bar{z} , которые представляют собой выборочные средние значения, вычисленные по выборкам объема $n \geq 4$. Согласно центральной предельной теореме, такие средние значения имеют приблизительно нормальное распределение даже в том случае, когда распределения отдельных элементов этих выборок сильно отклоняются от нормального распределения. Если теперь по \bar{x} , \bar{y} , \bar{z} построить две линейные комбинации v' и w' и сложить их квадраты, то последствия отклонения распределений \bar{x} , \bar{y} , \bar{z} от нормального еще более сгладятся. Образование отношения (22) лишь незначительно ухудшит это приближение.

При малых n обстановка будет несколько менее благоприятной. Однако даже в случае $n = 2$ образование сумм и разностей

$$u_1 = \frac{x_1 + x_2}{\sqrt{2}}, \quad u_2 = \frac{x_1 - x_2}{\sqrt{2}}$$

с последующим сложением их квадратов ($u_1^2 + v_1^2 + u_1^2 + u_2^2 + v_2^2 + u_2^2$) в значительной степени выравнивает отклонения от

нормальности. Конкретные числовые примеры демонстрируют это свойство значительно лучше, чем рассуждения на основе общей теории.

Подводя итог, мы вправе сказать, что критерий F можно применять даже в том случае, когда не известно, подчиняются результаты наблюдений нормальному распределению или нет; при этом ошибка будет небольшой¹.

Пример 40 (по книге Fisher R. A., Stat. Meth., 11 ed. Ex. 38). Для экспериментального определения точности подсчета бактерий в почве некоторый участок земли был разбит на четыре равные части, с каждой из которых был произведен посев бактерий на семи пластинках. В результате на 28 пластинках были получены следующие количества колоний:

Пласти- нка	Проба почвы			
	I	II	III	IV
1	72	74	78	69
2	69	72	74	67
3	63	70	70	66
4	59	69	58	64
5	59	66	58	62
6	53	58	56	58
7	51	52	56	54
Сумма	426	461	450	440
Среднее	60,9	65,9	64,3	62,9

Для того чтобы проверить, являются ли отклонения средних в четырех выборках чисто случайными или нет, был применен метод дисперсионного анализа, который дал следующие результаты:

	Сумма квадратов	Число степе- ней свободы	Оценка дисперсии
Между классами	$Q_1 = 95$	$f_1 = 3$	$s_1^2 = 32$
Внутри классов	$Q_2 = 1446$	$f_2 = 24$	$s_2^2 = 60$
Все наблюдения	$Q = 1541$	$N - 1 = 27$	$s^2 = 57$

Выборочная дисперсия внутри классов здесь больше выборочной дисперсии между классами, и уже поэтому различие выборочных средних

¹ Для того, чтобы это утверждение было верно, нужно, чтобы выборочные средние \bar{x} , \bar{y} , \bar{z} были распределены приблизительно нормально, а для этого в свою очередь достаточно, чтобы у одинаково распределенных результатов наблюдений существовала дисперсия. — *Прим. перев*

незначимо: вычислять $F = 32/60$ совсем не нужно. Все 28 пластинок можно рассматривать как выборку из однородной совокупности. Общее выборочное среднее равно 63,5, а наилучшая оценка для дисперсии

$$s^2 = \frac{1541}{27} = 57,1.$$

Если бы отдельные результаты наблюдений подчинялись распределению Пуассона с математическим ожиданием 63,5, то истинная дисперсия равнялась бы $\sigma^2 = 63,5$ (§ 10 А). Для того чтобы проверить, насколько это предположение согласуется с результатами наблюдений, вычислим отношение

$$\chi^2 = \frac{27s^2}{\sigma^2} = \frac{1541}{63,5} = 24,3.$$

Если бы результаты наблюдений были распределены нормально, то случайная величина χ^2 подчинялась бы распределению χ^2 с 27 степенями свободы. Распределение Пуассона с большим математическим ожиданием 63,5 совсем незначительно отклоняется от соответствующего нормального распределения. Следовательно, в силу формул (9) и (10) из § 23, нужно ожидать, что χ^2 будет иметь среднее значение 27 и квадратичное отклонение $\sqrt{54} = 7,4$:

$$\chi^2 = 27 \pm 7,4.$$

Таким образом, отклонение найденного значения 24,3 от математического ожидания 27 следует признать случайным, и поэтому правдоподобное предположение о том, что результаты наблюдений подчиняются распределению Пуассона, экспериментом не опровергается. Наилучшая оценка для среднего значения распределения Пуассона в данном случае принимает значение $\bar{v} = 63,5$. Дисперсия распределения Пуассона равна среднему значению.

Г. СВЯЗЬ С КРИТЕРИЕМ t

Если имеется лишь два ряда наблюдений, то

$$\begin{aligned} s_1^2 &= Q_1 = n_1 (\bar{x} - M)^2 + n_2 (\bar{y} - M)^2 = \\ &= n_1 \left[\bar{x} - \frac{n_1 \bar{x} + n_2 \bar{y}}{n_1 + n_2} \right]^2 + n_2 \left[\bar{y} - \frac{n_1 \bar{x} + n_2 \bar{y}}{n_1 + n_2} \right]^2 = \\ &= \frac{n_1 n_2}{n_1 + n_2} (\bar{x} - \bar{y})^2 \end{aligned}$$

и

$$s_2^2 = \frac{Q_2}{N-2} = \frac{\sum (x - \bar{x})^2 + \sum (y - \bar{y})^2}{n_1 + n_2 - 2},$$

следовательно,

$$F = \frac{s_1^2}{s_2^2} = \frac{(\bar{x} - \bar{y})^2}{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) s_2^2}. \quad (27)$$

Правая часть (27) представляет собой квадрат статистики t , используемой в критерии Стьюдента. Это означает, что *при двух классах критерий F совпадает с двусторонним критерием t .*

Следовательно, если n_1 и n_2 не слишком малы (например, оба ≥ 4), то двусторонний критерий t можно применять даже в том случае, когда x и y имеют распределение, отличное от нормального.

Д. КОРРЕЛЯЦИЯ ВНУТРИ КЛАССОВ

Если каждый класс содержит лишь два наблюдения, то возникает ряд, состоящий из r пар (x, x') . Для вычисления Q_1 и Q_2 образуем из каждой пары среднее и разность

$$\bar{x} = \frac{x + x'}{2} \quad \text{и} \quad d = x - x'.$$

Если M — общее среднее всех \bar{x} , то, согласно (18),

$$Q_1 = 2 \sum (\bar{x} - M)^2 \quad (28)$$

и, согласно (7),

$$Q_2 = \frac{1}{2} \sum d^2. \quad (29)$$

Для контроля служит формула

$$Q_1 + Q_2 = Q = \sum [(x - M)^2 + (x' - M)^2]. \quad (30)$$

Далее, согласно (21),

$$s_1^2 = \frac{Q_1}{r-1}, \quad s_2^2 = \frac{Q_2}{r} \quad (31)$$

и, наконец, как всегда,

$$s^2 = \frac{Q}{N-1} = \frac{Q_1 + Q_2}{2r-1}. \quad (32)$$

Выборочный коэффициент корреляции внутри классов определяется формулой

$$r^* = \frac{2 \sum (x - M)(x' - M)}{\sum [(x - M)^2 + (x' - M)^2]}. \quad (33)$$

Выражение (33) устроено аналогично выборочному коэффициенту корреляции¹ (§ 66 Б). После небольших вычислений находим, что

$$r^* = \frac{Q_1 - Q_2}{Q_1 + Q_2}. \quad (34)$$

¹ Выборочный коэффициент корреляции целесообразно вычислять по формуле (33) в том случае, когда неизвестны дисперсии величин x и x' равны друг другу. — *Прим. перев.*

Если выборочная дисперсия внутри классов равна нулю, то коэффициент r^* равен $+1$. Если же выборочная дисперсия между классами равна нулю, то $r^* = -1$, однако практически этого никогда не наблюдается. Если обе выборочные дисперсии приближенно равны друг другу, то величина r^* близка к нулю.

Пример 41. Хедорн, Бертани и Галлера¹ разрезали на две симметричные части имагинальные диски, из которых развиваются мужские половые железы мухи-дрозофилы (*Drosophila melanogaster*), и полученные половинки пересаживали представителю того же вида. В результате из обеих частей возникали семенники приблизительно нормальных размеров. Однако от случая к случаю их размеры колебались в очень широких пределах. Нужно проверить, не вызывались ли эти колебания тем, что величины обих пересаженных половинок зачастую были неравными? Если это так, то нужно ожидать, что наряду с особенно большими семенниками внутри пар должна быть больше дисперсии между парами. Однако вычисления показывают, что, напротив, дисперсия внутри пар меньше, чем между парами. В таблице на стр. 306 указаны длины половых клеток x , x' , причем наибольшая из обих длин всегда предшествует наименьшей.

Находим

$$M = \frac{1}{36} \sum \bar{x} = 313,5,$$

$$Q_1 = 2 \sum (\bar{x} - M)^2 = 63\,222,$$

$$Q_2 = \frac{1}{2} \sum d^2 = 23\,006.$$

Выборочный коэффициент корреляции внутри пар равен

$$r^* = \frac{Q_1 - Q_2}{Q_1 + Q_2} = \frac{40\,216}{86\,228} = 0,47.$$

Выборочные дисперсии указаны в таблице:

	Сумма квадратов	Число степеней свободы	Оценка дисперсии
Между парами	$Q_1 = 63\,222$	$f_1 = 35$	$s_1^2 = 1806$
Внутри пар	$Q_2 = 23\,006$	$f_2 = 36$	$s_2^2 = 639$
	$Q = 86\,228$	$f = 71$	$s^2 = 1214$

Наконец,

$$F = \frac{1806}{639} = 2,83.$$

¹ Hädorn E., Bertani G. und Gallera J., Regulationsfähigkeit und Feldorganisation der männlichen Genital—Imaginalseibe von *Drosophila*, Wilhelm Roux' Archiv für Entwicklungsmechanik der Organismen, 2114 (1949), 31.

x	x'	\bar{x}	$(\bar{x} - M)^2$	d	d^2
394	328	361	2 256	66	4 356
382	344	363	2 450	38	1 444
375	328	351,5	1 444	47	2 209
369	319	344	930	50	2 500
369	319	344	930	50	2 500
369	293	331	306	76	5 776
363	350	356,5	1 849	13	169
357	325	341	756	32	1 024
357	300	328,5	225	57	3 249
356	331	343,5	900	25	625
353	347	350	1 332	6	36
350	297	323,5	100	53	2 809
347	325	336	506	22	484
344	313	328,5	225	31	961
335	319	327	182	16	256
331	319	325	132	12	144
328	269	298,5	225	59	3 481
325	300	312,5	1	25	625
322	300	311	6	22	484
319	313	316	6	6	36
319	306	312,5	1	13	169
319	303	311	6	16	256
316	303	309,5	16	13	169
313	272	292,5	441	41	1 681
313	234	273,5	1 600	79	6 241
309	300	304,5	81	9	81
309	294	301,5	144	15	225
306	300	303	110	6	36
297	287	292	462	10	100
297	253	275	1 482	44	1 936
294	281	287,5	676	13	169
287	253	270	1 892	34	1 156
281	272	276,5	1 369	9	81
281	269	275	1 482	12	144
275	263	269	1 980	12	144
250	234	242	5 112	16	256
11 811	10 763	11 287	31 611	1048	46 012

Следовательно, различие между классами значительно больше, чем внутри классов. 1%-ная граница для F равна 2,21. Таким образом, нет оснований утверждать, что колебания длин вызываются различием размеров обеих пересаженных половин.

Е. ДАЛЬНЕЙШИЕ ПРИЛОЖЕНИЯ

Метод дисперсионного анализа применим и к более сложным случаям. Например, может оказаться, что n наблюдаемых величин x_{ik} разбиваются на классы не только по строкам, но и по столбцам, и при этом требуется узнать, сколь существенно различие не только между строками, но также и между столбцами. Однако для того, чтобы в подобных случаях можно было применить критерий F , нужно сделать дополнительные предположения о распределении x_{ik} . Так, например, в нашем случае нужно предположить, что случайные величины x_{ik} представимы в виде сумм

$$x_{ik} = a_i + b_k + z_{ik},$$

где z_{ik} — независимые, нормально распределенные случайные величины с нулевым средним значением и одинаковыми дисперсиями.

Дальнейшее изложение намеченных здесь основных идей можно найти в соответствующей литературе, например, Fisher R. A., *The Design of Experiments* (Oliver and Boyd, 1935) или Kempthorne O., *The Design and Analysis of Experiments* (John Wiley and Sons, 1952).

§ 59. Общие принципы. Наиболее мощные критерии

А. ОСНОВНЫЕ ПОНЯТИЯ

С общей точки зрения вопросы проверки статистических гипотез были рассмотрены Нейманом и Е. Пирсоном¹. В этом параграфе мы постараемся изложить основные идеи их исследования.

Возможные результаты эксперимента можно изобразить в виде точек X некоторого пространства E . При этом безразлично, заполняют ли возможные точки все пространство или они могут попадать лишь в отдельные точки, принадлежащие E (например, в целочисленные точки). Эксперимент используется для проверки некоторой гипотезы H .

Пусть $PB = P(B|H)$ — вероятность, соответствующая произвольной измеримой области B из пространства E , вычисленная в предположении, что гипотеза H верна. Эту вероятность можно

¹ В первую очередь см. Neuman J. and Pearson E. S., *Phil. Trans. Roy. Soc. London, A* 231 (1932), 332.

получить суммированием вероятностей, соответствующих отдельным точкам из B или интегрированием плотности вероятности по области B .

Основой всех критериев является принцип, согласно которому гипотезу H отвергают тогда, когда наблюдаемая точка принадлежит некоторой определенной *критической области* V .

Область V выбирается таким образом, чтобы вероятность $P(V|H)$ была мала. Иными словами, должно выполняться неравенство

$$P(V|H) \leq \beta,$$

где β — заданная допустимая вероятность ошибки (например, 0,05 или 0,01).

Спрашивается, чем же нужно руководствоваться при выборе области V ? Ведь по заданной вероятности β область V определяется неоднозначно! Для ответа на этот вопрос введем понятия ошибок первого и второго рода.

Ошибка первого рода возникает тогда, когда отвергается правильная гипотеза. Как уже сказано, вероятность ошибки первого рода не превышает β . Если бы принималась во внимание лишь ошибка первого рода, то выбор V был бы в значительной степени произвольным. Может даже возникнуть мысль, что в качестве V удобно выбрать пустое множество: ведь в этом случае вероятность ошибки первого рода была бы равна нулю! Почему же так не поступают? Да потому, что имеется возможность совершить ошибку второго рода.

Ошибка второго рода возникает тогда, когда не отвергается ложная гипотеза. Если гипотеза H является ложной, то, вообще говоря, легко может случиться, что наблюдаемая точка не попадет в область V и, следовательно, гипотеза H не будет отвергнута. Однако этого нужно, по возможности, избегать. Целью эксперимента является решение вопроса, правильна или ложна гипотеза H , поэтому критерий должен быть устроен так, чтобы гипотеза H , по возможности, не отвергалась в том случае, когда она правильна, и чтобы она отвергалась, когда она ложна. Таким образом, следует стремиться к тому, чтобы *вероятность ошибки второго рода была возможно меньше.*

Но возникает новая трудность, связанная с тем, что вероятность ошибки второго рода нельзя указать заранее. Эта вероятность зависит от того, какая гипотеза H' является правильной, вместо ложной гипотезы H .

Сначала мы предположим, что имеется лишь одна альтернативная гипотеза H' . *Мощностью* P' некоторого определенного критерия относительно гипотезы H' называют вероятность отвергнуть гипотезу H , когда верна гипотеза H' :

$$P' = P'V = P(V|H').$$

Если гипотеза H' является правильной, то вероятность не отвергнуть H (т. е. вероятность ошибки второго рода) будет равна $1 - P'$.

Эта вероятность должна быть возможно меньшей, следовательно, мощность P' нужно сделать возможно большей. Если среди всех критериев, удовлетворяющих условию $PV \leq \beta$, данный критерий имеет наибольшую мощность P' , то он называется *наиболее мощным* критерием относительно альтернативной гипотезы H' .

Возникает следующая задача. Пусть заданы две функции множеств $PB = P(B|H)$ и $P'B = P(B|H')$, удовлетворяющие аксиомам теории вероятностей и определенные на всех измеримых множествах из пространства E . Требуется найти такую область V , для которой $P'V$ будет наибольшей при условии, что

$$PV \leq \beta. \quad (1)$$

Решения этой задачи в непрерывном и дискретном случаях требуют отдельного изложения.

Б. СЛУЧАЙ НЕПРЕРЫВНОГО РАСПРЕДЕЛЕНИЯ

Пусть E — пространство непрерывно меняющихся переменных x_1, \dots, x_n или часть этого пространства. Предположим, что обе функции множеств P и P' определяются непрерывными плотностями

$$f(X) = f(x_1, \dots, x_n) \text{ и } g(X) = g(x_1, \dots, x_n).$$

Если в некоторой части T пространства E функция f равна нулю, то мы можем, не нарушая условия (1), объединить T с V . От такого увеличения области V мощность $P'V$ может лишь увеличиться. Следовательно, мы можем рассматривать лишь дополнение множества T , равное разности $E - T$.

На множестве $E - T$ функция f не равна нулю, поэтому

$$U = U(X) = \frac{g}{f} \quad (2)$$

является непрерывной функцией от X . При любом положительном v событию $U < v$ в поле P соответствует определенная вероятность

$$G(v) = P(U < v) = P(g < v f). \quad (3)$$

Мы предположим сначала, что $1 - \beta$ принадлежит множеству значений функции распределения $G(v)$, т. е. найдется такое положительное v , что

$$P(g < v f) = 1 - \beta,$$

следовательно,

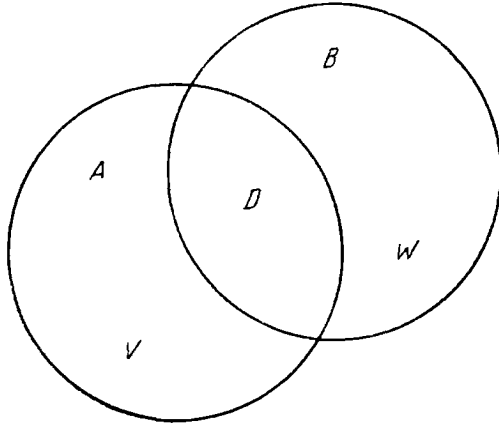
$$P(g \geq v f) = \beta. \quad (4)$$

В таком случае область V , определяемая неравенством

$$g \geq vf \text{ или } \frac{g}{f} \geq v,$$

является решением нашей экстремальной задачи.

Действительно, эта область, в силу (4), удовлетворяет условию (1). Если W — любая другая область, также удовлетворяющая условию (1), то можно показать, что $P'W \leq P'V$.



Р и с. 29.

Пусть D — пересечение областей V и W (рис. 29) и пусть

$$V = D + A, \quad W = D + B.$$

Таким образом, A представляет собой ту часть области V , которая не пересекается с W , а B — часть области W , которая не пересекается с V . Так как

$$P'V = P'D + P'A = \beta \text{ и } P'W = P'D + P'B \leq \beta,$$

то

$$P'A \geq P'B,$$

или, что то же самое,

$$\int_A f dX \geq \int_B f dX. \quad (5)$$

Но A принадлежит V , поэтому для всех точек X множества A справедливо неравенство $g \geq vf$. Следовательно,

$$P'V = P'D + P'A = P'D + \int_A g dX \geq P'D + \int_A vf dX.$$

Отсюда, в силу (5), получаем

$$\mathbf{P}' V \geq \mathbf{P}' D + \int_B v f dX.$$

Так как B не пересекается с V , то для всех точек X множества B справедливо неравенство $g < v f$, поэтому

$$\mathbf{P}' V \geq \mathbf{P}' D + \int_B g dX = \mathbf{P}' D + \mathbf{P}' B = \mathbf{P}' W.$$

Таким образом, область V является решением нашей экстремальной задачи.

Если функция $1 - G(t)$ не принимает значения β (т. е. $1 - G(t)$ совершает скачок от значения $< \beta$ к значению $> \beta$), то V строится следующим образом: сначала берут всю область $g > v f$ и затем добавляют к ней такую часть множества $g = v f$, чтобы общая вероятность $\mathbf{P}' V$ равнялась β . В остальном эта часть может быть выбрана произвольно. Доказательство останется тем же самым. Этот случай едва ли встречается в приложениях, и в дальнейшем мы его рассматривать не будем.

Плотность вероятности $f(X)$ называется *функцией правдоподобия* гипотезы H , и точно так же $g(X)$ называется функцией правдоподобия гипотезы H' . Поэтому отношение (2) называют *отношением правдоподобия*.

Только что найденный критерий, наиболее чувствительный к альтернативе H' , можно сформулировать так:

Гипотезу H следует отвергнуть тогда, когда отношение правдоподобия (2) будет не меньше v . При этом критическое значение v определяется таким образом, чтобы вероятность ошибки первого рода равнялась β , т. е. $\mathbf{P}(U \geq v) = \beta$.

Этот критерий называется *критерием отношения правдоподобия*. Он является наиболее мощным относительно гипотезы H' , и поэтому его целесообразно применять во всех тех случаях, когда имеется большая уверенность, что гипотеза H' может быть верна.

В. СЛУЧАЙ ДИСКРЕТНОГО РАСПРЕДЕЛЕНИЯ

Если пространство E состоит из конечного или счетного количества дискретных точек X , причем вероятность, соответствующая любому точечному множеству, равна сумме вероятностей для отдельных точек, то можно действовать точно так же, как мы действовали в непрерывном случае. Вместо плотностей вероятности $f(X)$ и $g(X)$ здесь появляются вероятности $\mathbf{P}X$ и $\mathbf{P}'X$, соответствующие отдельным точкам. Эти вероятности мы снова обозначим $f(X)$ и $g(X)$.

Если, в случае справедливости гипотезы H , некоторые точки пространства E имеют нулевые вероятности, то эти точки всегда можно присоединить к критической области V .

В остальных точках $f(X) \neq 0$, следовательно, можно построить случайную величину

$$U = \frac{g(X)}{f(X)} = \frac{P' X}{P X}.$$

Предположим сначала, что $1 - \beta$ принадлежит множеству значений функции распределения $G(v)$ случайной величины U , иными словами, найдется такое v , что

$$P(g \geq vf) = \beta.$$

Тогда, как и выше, в качестве V можно взять область, определяемую неравенством $g \geq vf$ с очевидной заменой интегрирования суммированием. Пусть W — любая другая область, удовлетворяющая условию $P(W) \leq \beta$, и пусть D — пересечение V и W . Если

$$V = D + A, \quad W = D + B,$$

то, как и в непрерывном случае,

$$P(A) \geq P(B)$$

или

$$\sum_A f(X) \geq \sum_B f(X).$$

В A имеет место неравенство $g \geq vf$, а в B — противоположное неравенство $g < vf$, поэтому

$$\begin{aligned} P' V &= P' D + P' A = P' D + \sum_A g(X) \geq P' D + \sum_A vf(X) \geq \\ &\geq P' D + \sum_B vf(X) \geq P' D + \sum_B g(X) = P' D + P' B = P' W. \end{aligned}$$

Таким образом, область V является решением экстремальной задачи, сформулированной выше.

Если функция $1 - G(v)$ не принимает значения β , а совершает скачок от значения $< \beta$ к значению $> \beta$, то для построения V сначала берут всю область $g > vf$ и, если это возможно, добавляют к ней столько точек X , удовлетворяющих условию $g = vf$, чтобы общая вероятность $P(V)$ равнялась β . Доказательство осуществляется точно так же, как и в непрерывном случае.

Если невозможно подобрать такие точки X , для которых $P(V) = \beta$, то к множеству $g > vf$ добавляют столько точек X , удовлетворяющих условию $g = vf$, чтобы вероятность $P(V)$ была по возможности близка к β (но не превосходила β). Пусть, например, $P V = \beta - \varepsilon$. Если к V прибавить еще одну точку X , удовлетворяющую условию $g(X) = vf(X)$, то вероятность ошибки

первого рода увеличится: $P(V + X) = \beta + \delta$, т. е. область $V + X$ — слишком велика. Следовательно, точку X нужно расщепить на две точки X_1 и X_2 и приписать им вероятности $P X_1 = \varepsilon$ и $P X_2 = \delta$. Затем точку X_1 следует включить в область V .

Для осуществления этого расщепления применяют следующий искусственный прием, аналогичный азартной игре с вероятностью выигрыша

$$p = \frac{\varepsilon}{\varepsilon + \delta}.$$

Если в результате эксперимента получена точка X , подлежащая расщеплению, то производится упомянутая выше «игра», причем в случае «выигрыша» гипотеза H отвергается, а в случае «проигрыша» не отвергается. Конечно, «игра» должна производиться независимо от результата опыта.

Докажем, что такой критерий будет являться решением нашей экстремальной задачи. Точке X соответствует вероятность $\varepsilon + \delta$. Событие X_1 осуществляется тогда и только тогда, когда наступает событие X и одновременно осуществляется «выигрыш». Следовательно, вероятность события X_1 равна

$$(\varepsilon + \delta)p = \varepsilon.$$

Поэтому

$$(\beta - \varepsilon) + \varepsilon = \beta$$

является вероятностью события $V + X_1$, т. е. $V + X_1$ представляет собой решение экстремальной задачи.

Указанная игра совсем не связана с выяснением вопроса о справедливости или ложности гипотезы H . Поэтому, практически, едва ли следует пользоваться такой игрой, а нужно в качестве критической области просто выбирать множество V (без X_1). И хотя в этом случае вероятность ошибки второго рода несколько возрастет, зато, однако, уменьшится вероятность ошибки первого рода, а именно она станет равной $\beta - \varepsilon$ (ранее она равнялась β). Если в качестве допустимой вероятности ошибки первого рода вместо β принять $\beta - \varepsilon$, то V будет решением новой экстремальной задачи, т. е. критерий, соответствующий критической области V , будет наиболее мощным с «уровнем значимости»¹ $\beta - \varepsilon$.

Г. ПРИМЕРЫ

Пример 42. Пусть E — пространство переменных x_1, \dots, x_n и пусть, согласно гипотезе H , случайные величины x_1, \dots, x_n независимы и одинаково нормально распределены с нулевым средним значением и единичной диспер-

¹ Уровень значимости критерия равен вероятности ошибки первого рода, т. е. этот уровень совпадает с вероятностью отвергнуть гипотезу H , когда она верна. Критерии проверки гипотез иногда также называют критериями значимости. — *Прим. перев.*

сией. Тогда совместная плотность вероятности будет задаваться формулой

$$f(X) = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}(x_1^2 + x_2^2 + \dots + x_n^2)}.$$

Далее, пусть, согласно конкурирующей гипотезе H' , случайные величины x_1, \dots, x_n предполагаются независимыми и одинаково нормально распределенными со средним значением $a > 0$ и единичной дисперсией:

$$g(X) = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}[(x_1 - a)^2 + \dots + (x_n - a)^2]} \quad (a > 0).$$

Отношение правдоподобия равно

$$U = \frac{g}{f} = e^{a \sum x - \frac{1}{2} na^2}.$$

U представляет собой монотонно возрастающую функцию от аргумента

$$\bar{x} = \frac{1}{n} \sum x.$$

Таким образом, гипотезу H следует отвергнуть тогда, когда выборочное среднее \bar{x} превосходит некоторое критическое значение c . Это критическое значение c определяется так, чтобы, в случае справедливости гипотезы H , вероятность события $\bar{x} > c$ равнялась β . Согласно гипотезе H , среднее \bar{x} распределено нормально с нулевым математическим ожиданием и дисперсией $1/n$. Поэтому

$$c = \frac{\sigma}{\sqrt{n}} \Psi(1 - \beta), \quad (6)$$

где Ψ — функция, обратная функции нормального распределения Φ . В нашем случае $\sigma = 1$, однако формула (6) справедлива при любом $\sigma > 0$.

Критерий, полученный в этом примере, замечателен тем, что он не зависит от a , если только a является положительным. Следовательно, односторонний критерий, отвергающий все значения $\bar{x} > c$, является *равномерно наиболее мощным* относительно всех гипотез H' с $a > 0$. Если бы мы в качестве конкурирующей гипотезы H' рассмотрели нормальное распределение с отрицательным a , то нужно бы было отвергнуть все значения $\bar{x} < -c$.

Пример 43. Пусть некоторое событие, согласно гипотезе H , имеет вероятность p и пусть, согласно альтернативной гипотезе H' , оно имеет большую вероятность p' . Предположим, что в n независимых опытах это событие осуществилось x раз. При каких x гипотезу H нужно отвергнуть?

Согласно гипотезе H , вероятность x -кратного осуществления события равна

$$f(x) = \binom{n}{x} p^x (1 - p)^{n-x}.$$

Согласно гипотезе H' , эта вероятность равна

$$g(x) = \binom{n}{x} p'^x (1 - p')^{n-x}.$$

Отношение правдоподобия задается формулой

$$u = \frac{g}{f} = \left(\frac{p'}{p}\right)^x \left(\frac{1-p'}{1-p}\right)^{n-x}.$$

Так как u является возрастающей функцией от x , то мы должны отвергнуть значения $x > c$. При этом граница c определяется таким образом, чтобы сумма вероятностей, соответствующих отброшенным значениям x , была наибольшей и не превосходила β :

$$\binom{n}{c+1} p^{c+1} q^{n-c-1} + \binom{n}{c+2} p^{c+2} q^{n-c-2} + \dots + q^n \leq \beta. \quad (7)$$

Левая часть неравенства (7) является возрастающей функцией от p , так как ее производная

$$(c+1) \binom{n}{c+1} p^c q^{n-c-1}$$

в интервале $0 < p < 1$ всегда положительна. Левая часть (7) при $p = 0$ равна нулю и при $p = 1$ равна единице, поэтому существует одно и только одно значение p_β , при котором левая часть (7) в точности равна β . Для $p \leq p_\beta$ неравенство (7) выполняется, а для $p > p_\beta$ не выполняется. Следовательно, этим критерием можно воспользоваться для проверки тех гипотез H , для которых $p \leq p_\beta$. Конкурирующей гипотезой H' в этом случае является $p > p_\beta$. Если $\bar{x} > c$, то гипотезы $p \leq p_\beta$ отвергаются, а $p > p_\beta$ не отвергаются.

Граница p_β в точности совпадает с односторонней доверительной границей для p , по Клопперу и Пирсону (§ 7). Таким образом, ранее изложенная теория доверительных границ подчиняется общим принципам проверки статистических гипотез.

§ 60. Сложные гипотезы

Простой гипотезой называется такая гипотеза, которая каждому событию из пространства E ставит в соответствие определенную вероятность. Если же вероятности зависят, кроме того, еще и от параметров, то мы имеем дело со *сложной гипотезой*. Сложная гипотеза состоит из простых гипотез, отвечающих фиксированным значениям параметров. Определение можно сформулировать и так: *сложная гипотеза есть множество простых гипотез*.

Если хотя бы проверить простую гипотезу H и если альтернатива H' также является простой, то, как мы видели в § 59, для проверки H всегда существует критерий, являющийся наиболее мощным относительно альтернативы H' . Но если H' является сложной, то могут иметь место два случая: либо существует *равномерно наиболее мощный критерий*, относительно всех простых гипотез, содержащихся в H' , либо такого критерия не существует.

Пример 42 (§ 59) может служить иллюстрацией обоих случаев. В этом примере гипотеза H является простой и гласит, что все

x_i независимы и распределены нормально с нулевым средним значением и единичной дисперсией. Альтернативная гипотеза H' зависит от параметра a и поэтому является сложной; согласно этой гипотезе, x_i независимы и распределены нормально со средним значением a и единичной дисперсией. Если допустимы лишь положительные значения a , то существует равномерно наиболее мощный критерий: гипотеза H отвергается тогда, когда \bar{x} превышает c/\sqrt{n} . Но если допустимы также и отрицательные значения a , то равномерно наиболее мощного критерия не существует. Критерий, отвергающий большие значения \bar{x} , теряет свою мощность при отрицательных a , и критерий, отвергающий малые значения \bar{x} , не является наиболее мощным при положительных a .

Для того чтобы в подобных случаях хорошие критерии можно было отличать от менее хороших, вводится понятие *несмещенности*. Критерий для проверки простой гипотезы H называется *несмещенным*, если вероятность отвергнуть H , когда она верна, не превосходит вероятность отвергнуть H , когда верна одна из гипотез H' , т. е.

$$P(V|H) \leq P(V|H') \text{ для всех } H'. \quad (1)$$

Другими словами, вероятность отвергнуть гипотезу H , когда она правильна, не должна превосходить вероятность отвергнуть H , когда H ложна. Вполне разумное требование!

Если в примере 42 расширить гипотезу H' и считать допустимыми все положительные и отрицательные средние значения a , то односторонние критерии, которые отвергают гипотезу H , коль скоро $\bar{x} > c/\sqrt{n}$ или коль скоро $\bar{x} < -c/\sqrt{n}$, не будут несмещенными. Для того чтобы получить несмещенный критерий, нужно воспользоваться абсолютной величиной $|\bar{x}|$ и определить критическую область с помощью неравенства $|\bar{x}| > c'/\sqrt{n}$. Если c' выбрать таким образом, чтобы вероятность отвергнуть гипотезу H , когда она правильна, в точности равнялась β , то этот критерий будет *несмещенным, наиболее мощным критерием* относительно всех альтернатив H' . Доказательство¹ можно найти в работе: Neyman J. and Pearson E. S., On the problem of the most efficient tests of statistical hypotheses, Philos. Trans. Royal Soc., London, A 231 (1933).

Если гипотеза H также является сложной, то и задача становится сложнее. Пусть, например, согласно гипотезе H , случайные величины x_1, \dots, x_n предполагаются независимыми и одинаково нормально распределенными с нулевым средним зна-

¹ Это доказательство приводится в книге Крамера Г., Математические методы статистики, ИЛ, М., 1948, стр. 581. — *Прим. переа.*

чением и произвольным (незаданным) квадратичным отклонением σ . Если гипотеза H верна, то совместная плотность вероятности для x_1, \dots, x_n задается формулой

$$f(x | \sigma) = (2\pi\sigma)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}(x_1^2 + \dots + x_n^2)} \quad (2)$$

Пусть V — критическая область. Если наблюдаемая точка X принадлежит V , то гипотезу H следует отвергнуть. При этом вероятность ошибки первого рода

$$P(V_1 | \sigma) = \int_V f(x | \sigma) dV, \quad (3)$$

вообще говоря, зависит от σ . Если $P(V | \sigma) \leq \beta$ для всех $\sigma > 0$, то говорят, что *уровень значимости критерия или области V не превосходит β* . Если же для всех σ

$$P(V | \sigma) = \beta,$$

то говорят, что *критерий (или область V) имеет уровень значимости, в точности равный β* . В этом случае Нейман и Пирсон называют V областью, *подобной пространству выборок*¹.

Нейман, Шэффе и Леманн² разработали общие методы, позволяющие находить области V с уровнем значимости, в точности равным β . Сущность этих методов мы поясним с помощью только что сформулированного примера. Доказательства можно найти в соответствующей литературе.

Формула плотности вероятности (2) непосредственно показывает, что

$$Q = x_1^2 + \dots + x_n^2 \quad (4)$$

является достаточной оценкой для $n\sigma^2$. Плотность распределения случайной величины Q равна

$$g(u | \sigma) = C \sigma^{-n} u^{\frac{n}{2}-1} e^{-\frac{1}{2\sigma^2}u}, \quad \text{где } C = \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}}. \quad (5)$$

Однопараметрическое семейство функций $g(u | \sigma)$ образует *ограниченно-полную систему*, в смысле Леманна и Шеффе; это

¹ Пространство выборок E удовлетворяет равенству $P(E|\sigma) = 1$ при всех σ . — *Прим. перев.*

² См. прежде всего Lehmann and Scheffé, *Completeness, Similar Regions and Unbiased Estimation*, *Sankhya*, 10 (1950), 305 и 15 (1956), 219. Там же указана дальнейшая литература.

означает, что если ограниченная интегрируемая функция $\varphi(t)$ удовлетворяет интегральному уравнению

$$\int_0^{\infty} \varphi(u) g(u | \sigma) du = 0 \text{ при всех } \sigma > 0, \quad (6)$$

то $\varphi(t) = 0$. Эта полнота становится тотчас же ясной, если, отбросив множитель $C\sigma^{-n}$, записать интегральное уравнение (6) так:

$$\int_0^{\infty} u^{\frac{n}{2}-1} \varphi(u) e^{-\lambda u} du = 0 \text{ при всех } \lambda > 0. \quad (7)$$

Леманн и Шеффе доказали, что в том случае, когда плотности вероятности образуют ограниченно-полную систему, все области V , имеющие уровень значимости, в точности равны β , можно найти по методу Неймана. Согласно этому методу, для каждого отдельного значения u достаточной статистики Q ищут такую область V_u , для которой условная вероятность, вычисленная при условии $Q = u$, принимает значение β . Если объединение V всех областей V_u измеримо, то оно имеет уровень значимости, в точности равный β .

В нашем случае V_u представляет собой область на сфере

$$x_1^2 + \dots + x_n^2 = u. \quad (8)$$

Условная плотность распределения случайных величин x_1, \dots, x_n на этой сфере задается отношением

$$\frac{f(x | \sigma)}{\int f(x | \sigma) d\omega_{n-1}}, \quad (9)$$

причем интегрирование в знаменателе производится по сфере (8). Область V_u на этой сфере нужно выбрать таким образом, чтобы интеграл от функции (9) по V_u в точности равнялся β . Так как функция $f(x | \sigma)$ на всей сфере равна некоторой постоянной величине, то в числителе и знаменателе (9) множители $f(x | \sigma)$ сократятся и интеграл будет просто равен отношению площади V_u к площади всей сферы. Таким образом, площадь V_u должна быть равна площади сферы, умноженной на β . В остальной области V_u можно выбирать произвольно (лишь бы они были не слишком дикими, с тем чтобы их объединение V оставалось измеримым).

Выбор областей V_u для построения наилучшего критерия в значительной мере зависит от рассматриваемых альтернативных гипотез H' . В данном случае в качестве альтернативы H' мы примем сложную гипотезу, согласно которой x_1, \dots, x_n — независи-

мые, одинаково нормально распределенные случайные величины с *положительным* средним значением a и произвольным квадратичным отклонением σ . Соответствующая плотность вероятности задается формулой

$$f_a(x | \sigma) = (2\pi\sigma)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} [(x_1 - a)^2 + \dots + (x_n - a)^2]} \quad (10)$$

Определение наиболее мощного критерия, соответствующего этой альтернативной гипотезе H' , не представляет труда. Сначала зафиксируем значение σ и определим область V_u таким образом, чтобы соответствующий критерий был наиболее мощным относительно отдельной гипотезы H'_σ . Так как интеграл от функции (9) не зависит от σ , то можно считать, что в (9) и (10) величины σ одинаковы. Тогда метод, изложенный в § 59, сам собой приведет нас к критерию отношения правдоподобия¹

$$\frac{f(x | \sigma)}{f_a(x | \sigma)} \geq v, \quad (11)$$

т. е., в нашем случае,

$$e^{\frac{a}{\sigma^2} (x_1 + \dots + x_n) - \frac{na^2}{2\sigma^2}} \geq v.$$

Таким образом, гипотезу H следует отвергнуть, коль скоро выборочное среднее

$$\bar{x} = \frac{1}{n} (x_1 + \dots + x_n)$$

превышает некоторое критическое значение w . Величина w выбирается таким образом, чтобы плоскость с уравнением $\bar{x} = w$ разбивала сферу (8) на две части и чтобы площадь той части, где $\bar{x} > w$, равнялась площади всей сферы, умноженной на β . Полученный критерий в точности совпадает с односторонним критерием t .

Таким образом, среди всех критериев, имеющих уровень значимости, в точности равный β , односторонний критерий t

¹ Так как мы хотим найти область V_u , расположенную на сфере (8), то мы должны воспользоваться условными плотностями (9) и

$$f_a(x | \sigma) \int f_a(x | \sigma) d\omega_{n-1},$$

где интегрирование производится по сфере (8). Так как интегралы в знаменателях условных плотностей зависят лишь от σ , то отношение правдоподобия можно записать в виде (11), где, вообще говоря, $v = v(\sigma)$. Если гипотеза H верна, то Q — достаточная статистика, поэтому условное распределение \bar{x} , при условии $Q = u$, не зависит от σ . — *Прим. перев.*

является равномерно наиболее мощным¹ относительно всех альтернатив H' с $a > 0$.

Пусть x_1, \dots, x_m и y_1, \dots, y_n — независимые, нормально распределенные случайные величины с одинаковыми неизвестными дисперсиями и пусть $\mathcal{E}x_1 = \dots = \mathcal{E}x_m = \mu$ и $\mathcal{E}y_1 = \dots = \mathcal{E}y_n = \nu$. Тем же самым методом можно доказать, что среди всех критериев с точным уровнем значимости β , предназначенных для проверки гипотезы $\mu = \nu$, односторонний критерий t является равномерно наиболее мощным относительно всех альтернатив H' с $\mu > \nu$.

Естественно поставить вопрос, не является ли критерий Стьюдента равномерно наиболее мощным также и среди всех критериев с уровнем значимости, не превосходящим β ? К сожалению, ответ на этот вопрос получается отрицательным².

¹ Можно показать, что среди всех несмещенных критериев с точным уровнем значимости 2β , предназначенных для проверки гипотезы $a = 0$, двусторонний критерий t является несмещенным, равномерно наиболее мощным относительно всех альтернатив $a \neq 0$ (см. Крамер Г., Математические методы статистики, ИЛ, М., 1948, стр. 583). — *Прим. перев.*

² Lehmann E. L. and Stein C., Most powerful tests of composite hypotheses I, *Ann. of Math. Stat.*, 19 (1948), 495.

ПОРЯДКОВЫЕ КРИТЕРИИ

Порядковыми критериями называют такие критерии, в которых используются не сами значения наблюдаемых величин, а лишь их упорядоченность, т. е. соотношения $x < y$ и $x > y$ (между двумя измеренными величинами). Такие критерии не зависят от функций распределения случайных величин x и y , и поэтому их называют *не зависящими от распределения или непараметрическими*.

Изучение теории порядковых критериев не требует больших предварительных познаний. Будет предполагаться известным лишь содержание гл. I и II.

§ 61. Критерий знаков

А. ОСНОВНОЙ ПРИНЦИП

Если 10 подопытных животных подвергались некоторому воздействию и во всех 10 случаях наблюдалось повышение кровяного давления, то, руководствуясь лишь одним чутьем, можно утверждать, что такой результат не является случайным! Это произвольное заключение можно обосновать следующим образом. Если бы наблюдаемые изменения кровяного давления колебались чисто случайно, то с большой вероятностью примерно половина всех разностей состояла бы из положительных величин, а другая половина — из отрицательных. Для каждого отдельного животного вероятность положительной разности — вероятность повышения кровяного давления — равнялась бы $1/2$. Следовательно, вероятность того, что все разности будут положительными, равнялась бы $(1/2)^{10} = 1/1024$. Столь маловероятный исход можно не принимать в расчет и поэтому следует заключить, что указанное выше утверждение соответствует действительности.

Этот совсем простой вывод заключения можно превратить в точный порядковый критерий с произвольно заданным уровнем значимости β .

Пусть в результате наблюдений получены n разностей $x_i - y_i$ ($i = 1, 2, \dots, n$), из которых k положительны и $n - k$ отрицательны. Возможность равенства $x_i = y_i$ мы пока исключаем. Гипотеза H , которую нужно проверить, утверждает, что при каждом i оба результата наблюдений x_i и y_i — независимые, одинаково

распределенные случайные величины. Согласно этой гипотезе, вероятность того, что разность $x_i - y_i$ окажется положительной, в точности равна вероятности, что эта разность будет отрицательной. Так как вероятность события $x_i = y_i$ равна нулю, то вероятности для положительных и отрицательных разностей равны $1/2$. Это и является тем следствием гипотезы H , которое нужно проверить с помощью критерия знаков.

Можно также положить $z_i = x_i - y_i$; разности z_1, \dots, z_n являются независимыми случайными величинами. Тогда подлежащая проверке гипотеза H утверждает, что для каждого i вероятности событий $z_i > 0$ и $z_i < 0$ равны друг другу:

$$P(z_i > 0) = P(z_i < 0). \quad (1)$$

Критерий знаков можно использовать для проверки гипотезы (1) также и в том случае, когда z не являются разностями.

Если вероятность события $z_i = 0$ равна нулю, то из (1) следует, что

$$P(z_i > 0) = \frac{1}{2}. \quad (2)$$

При этом предположении вероятность того, что среди всех z_1, \dots, z_n положительных величин окажется больше, чем m , равна

$$\left[\binom{n}{m+1} + \binom{n}{m+2} + \dots + \binom{n}{n} \right] \left(\frac{1}{2} \right)^n. \quad (3)$$

Если m — наименьшее число, для которого выражение (3) еще не превосходит β , то критерий знаков можно сформулировать так:

Гипотезу H следует отвергнуть, коль скоро k (число положительных z_i) окажется больше, чем m .

Уровень значимости этого критерия, т. е. вероятность отвергнуть гипотезу H , когда она правильна, очевидно, не превосходит β . Ведь критерий именно так и строился!

Сформулированное правило представляет собой *односторонний критерий знаков*. *Двусторонний* критерий отвергает гипотезу H не только тогда, когда количество k (число положительных z_i) превышает границу m , но также и тогда, когда количество $n - k$ (число отрицательных z_i) превышает эту же границу. Если граница m осталась той же самой, что и в одностороннем критерии, то уровень значимости двустороннего критерия вдвое больше уровня одностороннего критерия, следовательно, он не превосходит 2β .

В таблице 9 указаны границы m для $n \leq 50$, соответствующие обычным уровням значимости, а именно:

Двусторонний критерий: $2\beta = 0,05; 0,02; 0,01;$

Односторонний критерий: $\beta = 0,025; 0,01; 0,005.$

Б. СВЯЗИ

Спрашивается, как нужно поступать в том случае, когда имеются «связи», т. е. когда некоторые разности $x_i - y_i = z_i$ обращаются в нуль? Можно, например, половину связей считать положительными, а другую половину — отрицательными. Можно также для каждой связи бросать монету, и если выпадет герб, то считать разность z_i положительной. Однако лучше всего связи просто отбросить¹. Пусть k — количество положительных разностей z_i , l — количество отрицательных разностей и пусть $n = k + l$. Если к этим n разностям применить критерий значимости, то можно гарантировать, что уровень значимости не превзойдет β (в случае двустороннего критерия — 2β).

Доказательство. Предположим, что гипотеза (1) является правильной. Пусть N — общее число наблюдений ($N \geq n$) и пусть p_n — вероятность того, что n разностей будут отличными от нуля. Сумма всех p_n , разумеется, равна единице:

$$\sum_0^N p_n = 1. \quad (4)$$

Если количество отличных от нуля разностей равно n , то условная вероятность события $k > m$ (m — граница, соответствующая числу n) не превышает β . Обозначим эту условную вероятность P_n . Тогда

$$P_n \leq \beta. \quad (5)$$

Безусловная вероятность отвергнуть гипотезу H , в силу формулы полной вероятности, удовлетворяет неравенству

$$P = \sum p_n P_n \leq \sum p_n \beta = \beta. \quad (6)$$

Теорема доказана.

Пример 44. В опытах Фрица-Ниггли² яйца мухи-дрозофилы подвергались воздействию мягкого и жесткого излучения ($18 \cdot 10^4$ и $31 \cdot 10^6$ электрон-вольт). Из частот смертности в различных группах яиц, подвер-

¹ H e m e l g y k J., A theorem on the sign test when ties are present, Proc. Kon. Ned. Akad. section of sciences, A 55, 322.

² F r i t z - N i g g l i H., Vergleichende Analyse der Strahlenschädigung von Drosophila-Eiern, Fortschr. auf dem Geb. d. Röntgenstrahlen, 83 (1955), 178.

гавшихся воздействию одинаковых доз излучения, были сначала составлены средние. Затем, во всех случаях, вычислялись разности средних d для мягкого и жесткого излучений. К разностям d применялся критерий Стьюдента. При этом d имели следующие знаки (для яиц различного возраста) Возраст яиц (в часах):

1	+	+	+	+	-	+	+	-	(8 случаев)
1 ³ / ₄	+	+	+	-	-	-	+		(7 случаев)
3	+	+	+	+	+				(5 случаев)
4	+	+	+	+	+				(5 случаев)
5 ¹ / ₂	+	+							(2 случая)
7	+	+	+	+					(4 случая)

В возрастных группах от 1 до 3 час. критерий Стьюдента обнаружил превышение 5%-ной границы лишь в одном-единственном случае (а именно, у яиц возраста 1³/₄ часа); случаев же превышения 1%-ной границы вообще не оказалось. Напротив, в возрастных группах от 4 до 7 час. этот критерий в 7 случаях из 11 выявил превышение 5%-ной границы и в 5 случаях — превышение 1%-ной границы. Следовательно, практически можно считать установленным (по крайней мере для яиц старшего возраста), что при равных дозах мягкое излучение имеет более сильные летальные свойства, чем жесткое излучение.

Применение критерия Стьюдента было связано с большим количеством вычислений и, кроме того, потребовало предположения нормальности распределения. Поэтому возникает вопрос, нельзя ли сделать вывод о свойствах излучения с помощью одних только знаков + и —?

Если мы объединим возрастные группы от 4 до 7 час., то в 11 случаях из 11 будем иметь знак +. В качестве двусторонних 1%-ных границ табл. 9 указывает 1 и 10. Так как 11 находится вне этих границ, то с большой уверенностью можно утверждать, что летальные свойства мягкого излучения более эффективны, чем жесткого. Вероятность ошибочности этого вывода не превышает 0,01.

В возрастных группах от 1 до 3 час. знак + наблюдался в 15 случаях из 20. Двусторонние 5%-ные границы равны 6 и 14. Так как 15 находится вне этих границ, то эффективность мягкого излучения можно также считать установленной, степень уверенности здесь, конечно, меньше, чем в предыдущем случае.

Таким образом, критерий знаков позволяет установить почти без вычислений, что мягкие лучи на яйца старшего поколения действуют, практически достоверно, сильнее, а на яйца младшего возраста, вероятно, сильнее, чем жесткие лучи.

В. СИММЕТРИЯ РАСПРЕДЕЛЕНИЯ

Распределение случайной величины z называется *симметричным относительно нуля*, если при всех u имеет место равенство

$$P(z > u) = P(z < -u). \quad (7)$$

Если распределение задается плотностью вероятности $g(u)$, то равенство (7) означает, что $g(u)$ является четной функцией

$$g(u) = g(-u). \quad (8)$$

В частности, из (7) следует (1). Таким образом, для проверки симметрии распределения можно воспользоваться критерием знаков. Другие критерии симметрии изложены в работе: Hemelryk J., A family of parameterfree tests for symmetry, Proc. Kon. Ned. Akad. (section of sciences), 53, 945, 1186.

г. доверительные границы для медианы

Если рассматривать лишь *непрерывные* функции распределения $F(u)$, то (1) будет эквивалентно равенству

$$P(z < 0) = P(z > 0) = \frac{1}{2}. \quad (9)$$

В силу условия (9), (истинная) медиана равна нулю (§ 17). Поэтому критерий знаков можно использовать для проверки гипотезы о том, что распределение имеет нулевую медиану.

Если нужно проверить гипотезу, согласно которой медиана равна ζ , то можно в качестве новых случайных величин ввести $z_i - \zeta$ и затем воспользоваться критерием знаков. В силу одностороннего критерия, предположительное значение медианы ζ следует отвергнуть, если выборка (z_1, \dots, z_n) содержит более, чем m , положительных разностей $z_i - \zeta$.

Это же правило можно сформулировать еще и по-другому. Пусть z_1, \dots, z_n расположены в порядке их возрастания: $z^{(1)} \leq \dots \leq z^{(n)}$. Случайные величины $z^{(i)}$ представляют собой *порядковые статистики* (§ 17). Рассмотрим порядковую статистику с номером $n - m$ (т. е. рассмотрим $z^{(n-m)}$). Если $\zeta < z^{(n-m)}$, то количество положительных разностей $z^{(i)} - \zeta$ будет больше m , следовательно, все предполагаемые значения медианы ζ , удовлетворяющие неравенству $\zeta < z^{(n-m)}$, должны быть отвергнуты. Согласно двустороннему критерию, следует также отвергнуть все предполагаемые значения медианы $\zeta > z^{(m+1)}$. Таким образом, $z^{(n-m)}$ и $z^{(m+1)}$ являются *двусторонними доверительными границами для медианы* ζ . Соответствующий доверительный интервал имеет вид

$$z^{(n-m)} \leq \zeta \leq z^{(m+1)}. \quad (10)$$

Заключение (10) справедливо с вероятностью $\geq 1 - 2\beta$. Как легко убедиться, полученный результат сохраняет силу также и для распределений, не являющихся непрерывными.

§ 62. Задача двух выборок

а. постановка задачи

Пусть результатами наблюдений являются $n = g + h$ независимых случайных величин:

$$x_1, \dots, x_g; y_1, \dots, y_h,$$

и пусть все x_i наблюдаются в одинаковых экспериментальных условиях, т. е. можно предположить, что все они имеют одинако-

вые функции распределения. Такое же предположение мы будем делать и относительно y . Допустим, что наблюдается некоторое различие эмпирических распределений x и y ; например, все x могут оказаться больше, чем y , или область рассеяния x может быть шире области рассеяния y . Спрашивается, является ли различие эмпирических распределений следствием различия истинных распределений или же оно чисто случайное?

Нулевая гипотеза H_0 , подлежащая проверке, утверждает, что все x и y имеют одинаковые функции распределения и, значит, наблюдаемое различие эмпирических распределений является чисто случайным. Однако при этом мы не должны делать никаких специальных предположений о функции распределения x и y .

Два критерия, о которых мы уже говорили раньше, а именно критерий Стьюдента и критерий отношения дисперсий, основаны на предположении нормальности распределений x и y ; поэтому указанные критерии с самого начала нужно исключить из рассмотрения. И хотя оба критерия, с определенной степенью приближения, применимы к распределениям, отличным от нормального, однако в данном случае они оказываются непригодными, так как наша задача заключается в отыскании точных критериев, использующих лишь порядковые соотношения $x < y$ и $x > y$. Будет показано, что при некоторых условиях эти порядковые критерии являются даже более мощными, чем критерий Стьюдента, т. е. что существуют случаи, когда указанные критерии приводят к правильному решению, а критерий Стьюдента — нет (иными словами, критерий Стьюдента в этих случаях ложную гипотезу H_0 не отвергает).

Согласно гипотезе H_0 , все x_i и y_k распределены одинаково. Предположим, что их функция распределения $F(t)$ является непрерывной. Отсюда следует, что такие события, как $x_i = x_j$ или $x_i = y_k$, все имеют вероятность, равную нулю.

На практике это предположение непрерывности, строго говоря, никогда не выполняется, так как все результаты измерений являются округленными числами. В приложениях довольно часто оказывается, например, что некоторые x_i и y_k равны друг другу. Наличие таких «связей» влечет за собой небольшие затруднения в применении порядковых критериев. Способы преодоления этих затруднений мы изложим позднее.

Преобразование

$$t' = F(t)$$

переводит x_i и y_k в новые случайные величины x'_i и y'_k , подчиняющиеся «прямоугольному» распределению с функцией распределения

$$F'(t') = t' \quad (0 \leq t' \leq 1).$$

Упорядоченность величин x' и y' остается той же самой, что и для величин x и y . Следовательно, для порядковых критериев совершенно безразлично, оперируем ли мы с x и y или с x' и y' . Поэтому во всех тех случаях, когда это облегчает вычисление вероятностей, мы можем предположить, что x и y подчиняются прямоугольному распределению. При желании мы можем взять за основу и любое другое непрерывное распределение, например нормальное распределение с нулевым средним значением и единичной дисперсией.

Согласно гипотезе H_0 , все перестановки $n = g + h$ случайных величин $x_1, \dots, x_g, y_1, \dots, y_h$ равновероятны. Таких перестановок имеется $n!$, следовательно, каждой из них соответствует вероятность $1/n!$.

Построение критерия для проверки гипотезы H_0 эквивалентно указанию критической области V , включающей в себя некоторые из $n!$ перестановок. Если наблюдаемое расположение принадлежит области V , то гипотезу H_0 следует отвергнуть. Для того чтобы уровень значимости этого критерия не превосходил β , нужно, чтобы область V содержала не более чем $\beta n!$ перестановок.

Б. КРИТЕРИЙ Н. В. СМИРНОВА

Критерий Смирнова аналогичен критерию Колмогорова (§ 16). В критерии Колмогорова сравнивались эмпирическая и предполагаемая теоретическая функции распределения. В критерии Смирнова сравниваются две эмпирические функции распределения.

Пусть $F_g(t)$ — эмпирическая функция распределения, построенная по выборке x_1, \dots, x_g . Если $k(t)$ — количество тех x_i , которые удовлетворяют неравенству $x_i < t$, то

$$F_g(t) = \frac{k(t)}{g}.$$

Точно так же пусть $G_h(t)$ — эмпирическая функция распределения, построенная по выборке y_1, \dots, y_h , и пусть D — верхняя грань разности $|F_g - G_h|$. Согласно критерию Смирнова, гипотезу H_0 следует отвергнуть, если $D > D_\beta$. При этом D_β определяется так, чтобы вероятность события $D > D_\beta$, когда гипотеза H_0 верна, не превосходила β .

Смирнов доказал¹, что вероятность события

$$D > \lambda \sqrt{\frac{1}{g} + \frac{1}{h}}$$

¹ Смирнов Н. В., Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках, Бюлл. МГУ, 2, вып. 2 (1939), 1.

при больших n асимптотически равна сумме бесконечного ряда

$$2e^{-2\lambda^2} - 2e^{-2^2 \cdot 2\lambda^2} + 2e^{-3^2 \cdot 2\lambda^2} - \dots \quad (1)$$

Следовательно, если λ определить таким образом, чтобы сумма этого ряда равнялась 2β , то при больших n можно будет положить

$$D_\beta = \lambda \sqrt{\frac{1}{g} + \frac{1}{h}}. \quad (2)$$

Ряд (1) сходится очень быстро, и для практических целей его сумму можно заменить первым членом: это лишь увеличит надежность критерия. В результате получаем очень полезное приближение

$$\lambda \sim \sqrt{-\frac{1}{2} \ln \beta}. \quad (3)$$

Для того чтобы найти хорошее приближение для D_β , нужно лишь (3) подставить в (2).

Особым преимуществом критерия Смирнова является то, что этот критерий со сколь угодно большой вероятностью позволяет обнаружить любое отклонение между функциями распределения x и y , если только n достаточно велико. Таким образом, критерий Смирнова следует применять тогда, когда нужно проверить полное согласие функций распределения $F(t)$ и $G(t)$ случайных величин x и y во всем интервале изменения t и когда для этой проверки в нашем распоряжении имеется очень обширный материал наблюдений.

Но если речь идет лишь о том, чтобы установить, не будет ли x в среднем больше, чем y , то следует применять более мощные критерии, которые даже при небольших n могут привести к решению поставленного вопроса. Такого рода критериями являются критерий Вилкоксона и критерий X , к изложению которых мы теперь и переходим.

§ 63. Критерий Вилкоксона

А. ФОРМУЛИРОВКА КРИТЕРИЯ

Пусть наблюдаемые x_i и y_k расположены в порядке возрастания их величины. Если отбросить индексы, то получим последовательность, состоящую из букв x и y , например,

$$y y x y x y x x. \quad (1)$$

Если в этой последовательности x появляется позднее некоторого y , то говорят, что имеется одна *инверсия*. Например, последовательность (1) содержит 15 инверсий, так как первый x образует

с двумя предшествующими y две инверсии, второй x образует три инверсии и оба последних x — по пять инверсий.

Согласно критерию Вилкоксона, нулевая гипотеза отвергается, коль скоро количество инверсий U превосходит границу U_β . Граница U_β выбирается таким образом, чтобы, в случае если нулевая гипотеза верна, количество перестановок с числом инверсий $U > U_\beta$ не превышало $\beta n!$. Указанное правило представляет собой односторонний критерий.

При малых g и h граница U_β определяется непосредственным подсчетом последовательностей с наибольшими количествами инверсий. Для облегчения этого подсчета можно у x и y так же, как в (1), отбросить все индексы. В этом случае количество всех возможных последовательностей будет равно не $n!$, а

$$\binom{n}{g} = \frac{n!}{g! h!}.$$

Подсчет начинают с последовательности

$$y y \dots y y x x \dots x x, \quad (2)$$

которая имеет gh инверсий. Затем записывают последовательность

$$y y \dots y x y x \dots x x \quad (3)$$

с $gh - 1$ инверсиями и т. д. — до тех пор, пока не наберется больше чем $\beta \binom{n}{g}$ последовательностей. Количество инверсий в последней из полученных последовательностей и принимают в качестве U_β . Проиллюстрируем этот метод следующим примером:

$$g = h = 5, \quad \beta = 0,025$$

$$\beta \binom{n}{g} = \frac{252}{40} = 6,3.$$

1. $y y y y y x x x x x$
2. $y y y y x y x x x x$
3. $y y y x y y x x x x$
4. $y y y y x x y x x x$
5. $y y x y y y x x x x$
6. $y y y x y x y x x x$
7. $y y y y x x x y x x$

Последняя из выписанных последовательностей имеет 22 инверсии, следовательно, $U_{0,025} = 22$.

В данном случае существуют лишь 4 последовательности с большим чем 22 количеством инверсий, а именно последовательности с 1 по 4. Таким образом, уровень значимости соответствующий

щего критерия равен $4/252 = 0,016$, т. е. он значительно ниже допустимого уровня 0,025.

Это же обстоятельство имеет место и в других примерах. В большинстве случаев существует целый ряд последовательностей с одинаковым числом инверсий. В нашем примере последовательности 5, 6 и 7 имеют по 22 инверсии. Если бы все эти три последовательности были включены в критическую область, то уровень значимости соответствующего критерия превышал бы β . Однако если ни одну из этих трех последовательностей не включать в критическую область, то критерий окажется излишне слабым.

При больших g и h вычисление точной границы U_β очень утомительно. Но мы увидим, что в этом случае распределение случайной величины U можно аппроксимировать нормальным распределением.

При двустороннем варианте критерия Вилкоксона нулевая гипотеза отвергается не только тогда, когда количество инверсий превосходит границу U_β , но также и тогда, когда эту же границу превосходит количество $gh - U$ обратных инверсий. В этом случае уровень значимости критерия удваивается.

Вместо подсчета инверсий можно x_i и y_k пронумеровать в порядке возрастания их величины. Если при этом наименьший из всех x_i имеет порядковый номер r_1 , то количество предшествующих y_k равно $r_1 - 1$, и поэтому наименьшему x соответствуют точно $r_1 - 1$ инверсий. Если следующий за наименьшим x имеет порядковый номер r_2 , то ему соответствуют $r_2 - 2$ инверсий и т. д. Таким образом, в итоге получаем

$$\begin{aligned} U &= (r_1 - 1) + (r_2 - 2) + \dots + (r_g - g) = \\ &= \sum r_i - \frac{1}{2} g(g + 1) \end{aligned} \quad (4)$$

инверсий. Следовательно, для построения критерия Вилкоксона вместо U можно воспользоваться статистикой $\sum r_i$, представляющей собой сумму порядковых номеров случайных величин x_i .

Мы постараемся теперь исследовать распределение U несколько точнее. При этом мы сначала будем предполагать, что нулевая гипотеза верна. Напомним, что, согласно этой гипотезе, все x_i и y_k независимы и имеют одинаковые (непрерывные) функции распределения.

Б. СРЕДНЕЕ ЗНАЧЕНИЕ И ДИСПЕРСИЯ U

Для каждой пары наблюдений x_i, y_k определим функцию z_{ik} , принимающую лишь значения 0 или 1, а именно

$$z_{ik} = \begin{cases} 1, & \text{если } x_i > y_k, \\ 0, & \text{если } x_i \leq y_k. \end{cases}$$

Тогда, очевидно,

$$U = \sum z_{ik}. \quad (5)$$

Если нулевая гипотеза верна, то значения 0 и 1 для всех случайных величин z_{ik} являются равновероятными. Следовательно, среднее значение z_{ik} равно $1/2$. Из (5) тотчас же получаем среднее значение для U :

$$\hat{U} = \frac{1}{2} gh. \quad (6)$$

Вместо (5) мы можем теперь записать

$$U - \hat{U} = \sum \left(z_{ik} - \frac{1}{2} \right). \quad (7)$$

Для того чтобы вычислить дисперсию σ^2 случайной величины U , мы возведем (7) в квадрат и найдем среднее значение:

$$\sigma^2 = \mathfrak{E}(U - \hat{U})^2 = \sum \mathfrak{E} \left(z_{ik} - \frac{1}{2} \right) \left(z_{jl} - \frac{1}{2} \right). \quad (8)$$

Слагаемые с $i \neq j$ и $k \neq l$ равны нулю, так как в данном случае z_{ik} и z_{jl} независимы и их средние значения равны $1/2$. Слагаемые с $i = j$ и $k = l$ все равны $1/4$. Произведения $(z_{ik} - 1/2)(z_{jl} - 1/2)$ при $i = j$ и $k \neq l$ равны $-1/4$, если x_i расположен между y_k и y_l и равны $+1/4$ — в противном случае. Таким образом, среднее значение такого произведения равно

$$-\frac{1}{4} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{12}.$$

То же самое справедливо и при $k = l$ и $i \neq j$. Окончательно, в силу (8), получаем

$$\sigma^2 = \frac{1}{4} gh + \frac{1}{12} gh(h-1) + \frac{1}{12} gh(g-1) = \frac{gh}{12} (g+h+1). \quad (9)$$

В. АСИМПТОТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ U ПРИ $g \rightarrow \infty$ И $h \rightarrow \infty$

Манн и Уитни¹ вычислили не только среднее значение и дисперсию U , но также и нашли для больших g и h асимптотические формулы центральных моментов высших порядков. Моменты нечетного порядка равны нулю, так как распределение U сим-

¹ Mann H. B. and Whitney D. R., On a test whether one of two random variables is stochastically larger than the other, *Annals of Math. Stat.*, 18 (1947), 50.

метрично относительно среднего значения $g h \sqrt{2}$. Для моментов четного порядка имеет место формула

$$\mathfrak{E}(u^{2r}) = 1 \cdot 3 \cdot 5 \dots (2r-1) (gh)^r (g+h+1)^r \frac{1}{(12)^r} + R, \quad (10)$$

где $u = U - gh/2$ и R стремится к бесконечности (при $g \rightarrow \infty$ и $h \rightarrow \infty$) медленнее, чем главный член формулы (10). Если u^{2r} разделить на

$$\sigma^{2r} = (\mathfrak{E} u^2)^r = (gh)^r (g+h+1)^r \frac{1}{(12)^r},$$

затем вычислить математическое ожидание и устремить g и h к бесконечности, то, в силу (10), получим

$$\lim \mathfrak{E} \left(\frac{u}{\sigma} \right)^{2r} = 1 \cdot 3 \cdot 5 \dots (2r-1). \quad (11)$$

Согласно «второй предельной теореме» (§ 24 E), отсюда следует, что случайная величина u/σ при $g \rightarrow \infty$ и $h \rightarrow \infty$ распределена асимптотически нормально с нулевым средним значением и единичной дисперсией, или

Если g и h стремятся к бесконечности, то U распределена асимптотически нормально со средним значением $gh/2$ и дисперсией σ^2 .

Метод моментов, примененный здесь для доказательства асимптотической нормальности, можно использовать и во многих других случаях; например, с помощью этого метода можно доказать асимптотическую нормальность случайной величины U даже тогда, когда распределения x и y различны¹.

Г. АСИМПТОТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ U ПРИ $h \rightarrow \infty$

Если к бесконечности стремится только h , а g остается постоянным, то для отыскания асимптотического распределения U нужно применить другой метод. Основная идея этого метода станет особенно ясной, если мы сначала предположим, что $g = 2$.

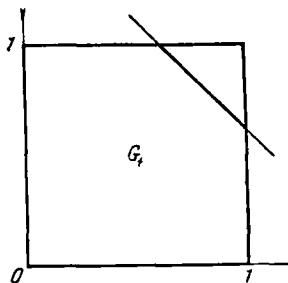
Пусть x_1, x_2 и y_1, \dots, y_h — независимые случайные величины и пусть $F(t) = t$ ($0 \leq t \leq 1$) — их общая функция распределения, т. е. все x_i и y_k распределены одинаково равномерно в интервале $(0,1)$. Далее, пусть u_1 и u_2 — количества инверсий для x_1 и x_2 соответственно. Общее количество инверсий равно $U = u_1 + u_2$.

¹ См. Lehmann E. L., Consistency and unbiasedness of nonparametric tests, Ann. of Math. Stat. 22, 167, Theorem 3.2 (здесь же указана литература), а также Hoeffding W., A combinatorial central limit theorem, Ann. of Math. Stat., 22, 558.

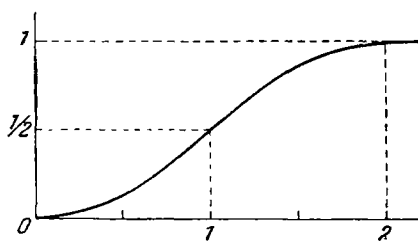
Сначала мы зафиксируем x_1 и x_2 . Если x_1 — постоянная величина, то вероятность события $y < x_1$ равна $F(x_1) = x_1$. Частота этого события задается отношением

$$v_1 = \frac{u_1}{h}, \quad (12)$$

так как из h величин y_1, \dots, y_h ровно u_1 оказались меньше x_1 . Если h велико, то частота события, с большой вероятностью, близка к вероятности этого события, следовательно, частота v_1 приближенно равна x_1 .



Р и с. 30.



Р и с. 31.

По той же самой причине частота v_2 близка к x_2 . Следовательно, отношение

$$v_1 + v_2 = \frac{u_1 + u_2}{h} = \frac{U}{h} \quad (13)$$

с большой вероятностью близко к $x_1 + x_2$.

Задача заключается в вычислении функции распределения случайной величины U , т. е. в вычислении вероятности события $U < u$. Вместо неравенства $U < u$ можно также записать

$$v_1 + v_2 = \frac{U}{h} < \frac{u}{h} = t. \quad (14)$$

Таким образом, мы должны вычислить вероятность события $v_1 + v_2 < t$.

Так как случайная величина $v_1 + v_2$, с большой вероятностью, близка к $x_1 + x_2$, то мы сначала вычислим вероятность события $x_1 + x_2 < t$. Случайные величины x_1 и x_2 независимы и распределены одинаково равномерно в интервале $(0, 1)$, поэтому совместная плотность вероятности для пары (x_1, x_2) внутри квадрата $0 < x_1 < 1$, $0 < x_2 < 1$ равна единице. Таким образом, вероятность события $x_1 + x_2 < t$ равна площади области G_t , определяемой неравенствами

$$0 < x_1 < 1, \quad 0 < x_2 < 1, \quad x_1 + x_2 < t.$$

Область G_t изображена на рис. 30. Она представляет собой часть единичного квадрата, лежащую под прямой с уравнением $x_1 + x_2 = t$. Площадь области G_t равна

$$H(t) = \begin{cases} 0, & \text{если } t \leq 0, \\ \frac{1}{2} t^2, & \text{если } 0 \leq t \leq 1, \\ 1 - \frac{1}{2} (2 - t)^2, & \text{если } 1 \leq t \leq 2, \\ 1, & \text{если } t \geq 2. \end{cases} \quad (15)$$

График функции $H(t)$ изображен на рис. 31. В интервале от 0

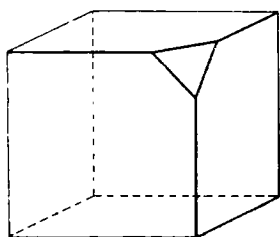


Рис. 32.

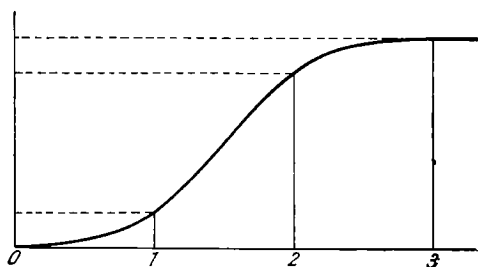


Рис. 33.

до 2 этот график состоит из двух дуг квадратных парабол. График соответствующей плотности вероятности указан в § 25, рис. 16.

Точно так же в случае $g = 3$ вероятность события $x_1 + x_2 + x_3 < t$ оказывается равной объему той части пространства, которая возникает в результате пересечения единичного куба плоскостью с уравнением $x_1 + x_2 + x_3 = t$ (рис. 32).

Вычисления показывают, что

$$H(t) = \begin{cases} 0, & \text{если } t \leq 0, \\ \frac{1}{6} t^3, & \text{если } 0 \leq t \leq 1, \\ \frac{1}{6} t^3 - \frac{1}{2} (t - 1)^3, & \text{если } 1 \leq t \leq 2, \\ 1 - \frac{1}{6} (3 - t)^3, & \text{если } 2 \leq t \leq 3, \\ 1, & \text{если } t \geq 3. \end{cases} \quad (16)$$

График функции $H(t)$ изображен на рис. 33, а график соответствующей плотности вероятности указан в § 25, рис. 17.

Уже в случае $g = 2$ и 3 графики функций (15) и (16) похожи на кривые нормального распределения. При $g = 4$ график функции

$H(t)$ почти совпадает с нормальной кривой; с увеличением g согласие станет еще лучшим.

Для перехода от $x_1 + x_2$ к $v_1 + v_2$ нам потребуется следующая лемма:

При любом целом положительном g и при любом $\varepsilon > 0$ функция $H(t)$ удовлетворяет условию

$$H(t + \varepsilon) - H(t) \leq \varepsilon. \quad (17)$$

Доказательство. Левая часть (17) представляет собой g -кратный интеграл

$$I = \int \dots \int dx_1 \dots dx_g, \quad (18)$$

где интегрирование производится по области, определяемой неравенствами

$$t \leq x_1 + x_2 + \dots + x_g < t + \varepsilon, \quad (19)$$

$$0 < x_i < 1 \quad (i = 1, 2, \dots, g). \quad (20)$$

Если сперва зафиксировать x_1, \dots, x_{g-1} и произвести интегрирование по x_g , то длина интервала интегрирования не будет превышать ε , так как неравенства (19) определяют интервал длины ε и, вследствие условия (20), этот интервал может лишь уменьшиться. Интегрированием по x_1, \dots, x_{g-1} в единичном кубе, принадлежащем $(g-1)$ -мерному пространству, убеждаемся, что интеграл (18) не превосходит ε . Лемма доказана.

Так как функция распределения $H(t)$ случайной величины $x_1 + x_2$ известна, то с помощью этой леммы мы можем оценить функцию распределения $v_1 + v_2$ сверху и снизу.

Пусть задано $\varepsilon > 0$. Покажем, что для достаточно больших h вероятность события $v_1 + v_2 < t$ отличается от $H(t)$ не более чем на 2ε .

Как мы уже знаем, разность $(x_1 + x_2) - (v_1 + v_2)$ при $h \rightarrow \infty$ по вероятности стремится к нулю. Отсюда следует, что для всех достаточно больших h вероятность события $(x_1 + x_2) - (v_1 + v_2) > \varepsilon$ будет меньше, чем ε . Если $v_1 + v_2 < t$, то либо $x_1 + x_2 < t + \varepsilon$, либо $(x_1 + x_2) - (v_1 + v_2) > \varepsilon$. Таким образом, событие $v_1 + v_2 < t$ содержится в объединении событий $x_1 + x_2 < t + \varepsilon$ и $(x_1 + x_2) - (v_1 + v_2) > \varepsilon$. Отсюда следует, что

$$\begin{aligned} \mathbf{P}(v_1 + v_2 < t) &\leq \mathbf{P}(x_1 + x_2 < t + \varepsilon) + \varepsilon = \\ &= H(t + \varepsilon) + \varepsilon \leq H(t) + 2\varepsilon. \end{aligned}$$

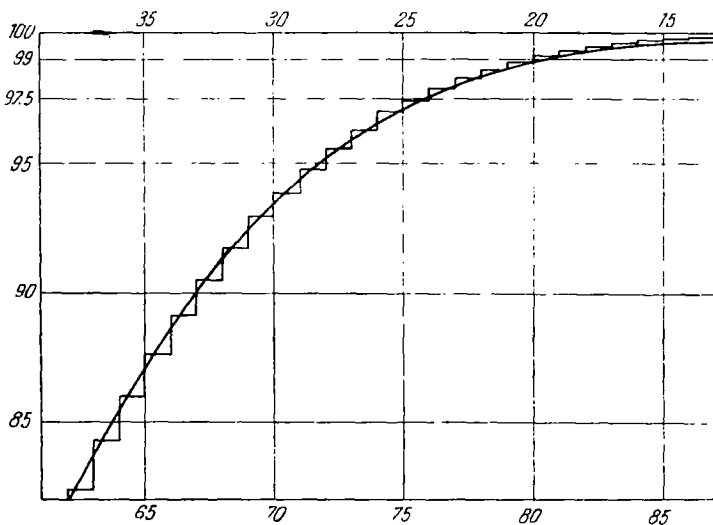
Точно так же находим

$$\begin{aligned} \mathbf{P}(v_1 + v_2 < t) &\geq \mathbf{P}(x_1 + x_2 < t - \varepsilon) - \varepsilon = \\ &= H(t - \varepsilon) - \varepsilon \geq H(t) - 2\varepsilon. \end{aligned}$$

Следовательно, как и утверждалось, $P(v_1 + v_2 < t)$ отличается от $H(t)$ не более чем на 2ε .

Этот же самый результат можно установить не только в случае $g = 2$, но и при любом g , а именно

При постоянном g и $h \rightarrow \infty$ случайная величина U асимптотически распределена, как сумма g независимых случайных величин, распределенных одинаково равномерно в интервале $(0, h)$.



Р и с. 34. Критерий Вилкоксона. Точная и асимптотическая кривые распределения U .

Следовательно, среднее значение этой суммы, равное $gh/2$, совпадает с точным средним значением случайной величины U . Дисперсия суммы равна

$$\frac{1}{12} gh^2,$$

в то время как дисперсия U задается формулой

$$\sigma_U^2 = \frac{1}{12} gh(g + h + 1).$$

При $g \geq 4$ распределение суммы можно достаточно точно аппроксимировать нормальным распределением. Поэтому нормальное приближение, найденное Манном и Уитни для больших g и h , применимо также для умеренных g и больших h , коль скоро $g > 3$.

Числовые примеры свидетельствуют о том, что h не обязательно должно быть очень большим. На рис. 34 изображены отрезок

точной кривой распределения (ломаная линия) при $g = h = 10$ и отрезок соответствующей кривой нормального распределения. Согласие вполне удовлетворительное, особенно в интервале между 95 и 90%, который наиболее важен для практических приложений. В интервале между 99 и 100% ломаная линия расположена над нормальной кривой, поэтому, в случае уровней значимости $\beta \leq 0,01$, применение нормального приближения увеличивает надежность критерия.

Изложенные результаты можно сформулировать так: При $g > 3$ и $g + h \geq 20$ нормальное приближение для критерия Вилкоксона оказывается достаточно точным. Если же значения g и h не удовлетворяют указанным неравенствам, то следует воспользоваться точным распределением.

д. таблицы для малых g и h

В таблице 10, в конце книги, указаны вероятности $p(u)$, соответствующие распределению статистики U критерия Вилкоксона, для всех g и h , удовлетворяющих условию

$$g \leq h \leq 10;$$

$p(u)$ определяется как вероятность события $U \leq u$, когда нулевая гипотеза верна. Если в результате эксперимента оказалось, что число инверсий равно u , и если $p(u) \leq \beta$, то это означает, что, согласно критерию Вилкоксона с заданным уровнем β , нулевую гипотезу следует отвергнуть. При двустороннем критерии нужно u заменить на $gh - u$ и применить то же самое правило. В случае одностороннего критерия истинный уровень значимости не превосходит β , а в случае двустороннего критерия он не превосходит 2β .

Таблица 10 была вычислена с помощью таблиц Ван дер Варта, опубликованных Математическим центром в Амстердаме (1952, отчет, стр. 32).

§ 64. Мощность критерия Вилкоксона

Как и в § 60, под мощностью критерия для проверки гипотезы H_0 относительно альтернативной гипотезы H' мы понимаем вероятность отвергнуть H_0 , когда H' правильна. В нашем случае, согласно нулевой гипотезе H_0 , все x_i и y_k независимы и имеют одинаковые функции распределения $F(t)$. В качестве альтернативы мы воспользуемся теперь гипотезой H' , которая утверждает, что все x_i и y_k независимы, причем все x_i имеют функцию распределения $F(t)$, а все y_k имеют функцию распределения $G(t)$, отличную от $F(t)$.

А. СРЕДНЕЕ ЗНАЧЕНИЕ И ДИСПЕРСИЯ U В СЛУЧАЕ
СПРАВЕДЛИВОСТИ ГИПОТЕЗЫ H'

Положим снова

$$U = \sum z_{ik}, \quad (1)$$

где $z_{ik} = 1$, если $x_i > y_k$, и $z_{ik} = 0$ — в противном случае, и определим среднее значение и дисперсию U с помощью того же метода, которым мы пользовались в § 63 Б. В результате получим¹

$$E U = gh p, \quad (2)$$

$$\sigma_U^2 = gh[(g-1)r^2 + (h-1)s^2 + pq], \quad (3)$$

где

$$p = \int G(t) dF(t), \quad (4)$$

$$q = 1 - p = \int F(t) dG(t), \quad (5)$$

$$r^2 = \int [F(t) - q]^2 dG(t), \quad (6)$$

$$s^2 = \int [G(t) - p]^2 dF(t), \quad (7)$$

причем во всех формулах интегрирование производится от $-\infty$ до $+\infty$.

Предположим, например, что x_i распределены нормально со средним значением $\mu > 0$ и единичной дисперсией и что y_k распределены нормально с нулевым средним значением и единичной дисперсией.

В этом случае

$$F(t) = \Phi(t - \mu), \quad G(t) = \Phi(t),$$

$$\begin{aligned} p &= \int_{-\infty}^{\infty} \Phi(t) d\Phi(t - \mu) = \int_{-\infty}^{\infty} \Phi(x + \mu) d\Phi(x) = \\ &= \frac{1}{2\pi} \int_{-x}^{\infty} e^{-\frac{1}{2}x^2} dx \int_{-\infty}^{x+\mu} e^{-\frac{1}{2}y^2} dy = \frac{1}{2\pi} \iint e^{-\frac{1}{2}(x^2 + y^2)} dx dy, \end{aligned}$$

где область интегрирования определяется неравенством $y < x + \mu$. Для того чтобы вычислить этот интеграл, мы введем новые переменные t и u с помощью ортогонального преобразования:

$$\begin{aligned} x + y &= t\sqrt{2}, \\ -x + y &= u\sqrt{2}. \end{aligned}$$

¹ См. van Dantzig D., Consistency and power of Wilcoxon's test, Proc. Kon. Ned. Akad., Amsterdam (Section of Sciences), A 54, 1.

Таким образом, получаем

$$p = \frac{1}{2\pi} \int \int e^{-\frac{1}{2}(t^2 + u^2)} dt du,$$

где область интегрирования определяется неравенством $u \sqrt{2} < \mu$. Если сначала проинтегрировать по t , а затем по u , то найдем

$$p = \Phi\left(\frac{\mu}{\sqrt{2}}\right). \quad (8)$$

Тогда

$$r^2 = \int [\Phi(t - \mu) - q]^2 d\Phi(t), \quad (9)$$

$$s^2 = \int [\Phi(t + \mu) - p]^2 d\Phi(t). \quad (10)$$

Подстановкой $t = -t'$ можно (9) перевести в (10). Следовательно, $r^2 = s^2$. Далее, если μ заменить на $-\mu$, то p перейдет в q и s^2 — в r^2 , т. е. s^2 перейдет само в себя. Поэтому s^2 является четной функцией от μ . Производить дальнейшие преобразования интеграла s^2 не имеет смысла, так как при этом не удастся получить простое выражение, зависящее от известных функций. Поэтому мы удовлетворимся тем замечанием, что p , r^2 и s^2 являются ограниченными функциями от μ (r^2 и s^2 даже стремятся к нулю при $\mu \rightarrow +\infty$ или $-\infty$) и что эти функции при малых μ можно разложить в ряды по степеням μ , начинающиеся с членов

$$p = \frac{1}{2} + \frac{\mu}{2\sqrt{\pi}} + \dots, \quad (11)$$

$$r^2 = s^2 = \frac{1}{12} + \dots \quad (12)$$

(выписаны лишь члены, содержащие μ^0 и μ^1).

Наша цель заключается в вычислении мощности одностороннего критерия Вилкоксона, т. е. в вычислении вероятности события $U > U_\beta$ как функции от μ . Обозначим эту функцию $P(\mu)$. Она нам, в частности, потребуется для сравнения мощностей критерия Вилкоксона и критерия Стьюдента. Ведь, как нам известно из § 60, если все x и y независимы и распределены нормально с одинаковыми дисперсиями, то среди всех односторонних критериев с точным уровнем значимости β , предназначенных для сравнения средних значений x и y , односторонний критерий Стьюдента является равномерно наиболее мощным. Так как речь идет об односторонних критериях, то, для определенности, мы снова предположим, что $\mu > 0$.

Пусть, например, $g \leq h$. Рассмотрим сначала два случая.

Первый случай: g и h — величины одинакового порядка.

Второй случай: h велико сравнительно с g , и g безгранично возрастает.

В обоих случаях мы будем рассматривать лишь такие значения μ , которые являются величинами порядка $1/\sqrt{g}$. Так как если μ велико сравнительно с $1/\sqrt{g}$, то искомая функция мощности критерия Вилкоксона и функция мощности критерия Стьюдента будут очень близки к единице, поэтому полное сравнение этих функций (при всех $\mu > 0$) не интересно. Доказывается это так.

Положим $U - gh/2 = V$ и $U_\beta - gh/2 = V_\beta$. Среднее значение V равно $(p - 1/2)gh$, следовательно, оно является величиной порядка μgh . Таким образом, если μ велико сравнительно с $1/\sqrt{g}$, то среднее значение V велико сравнительно с $h\sqrt{g}$. Согласно (3), квадратичное отклонение V является величиной порядка $h\sqrt{g}$ и граница V_β также является¹ величиной порядка $h\sqrt{g}$. Следовательно, вероятность события $V > V_\beta$ близка к единице. Так как при любом μ критерий Вилкоксона не мощнее критерия Стьюдента, то мощность критерия Стьюдента в обоих указанных случаях также будет близка к единице.

Итак, мы теперь предположим, что μ является величиной порядка $1/\sqrt{g}$. В этом случае в формулах (11) и (12) можно пренебречь членами порядка μ^2 или $1/g$ и ограничиться выписанными членами. Если (11) и (12) подставить в (2) и (3), то получим

$$\mathcal{E} U \sim \frac{1}{2} gh \left(1 + \frac{\mu}{2\sqrt{\pi}} \right), \quad (13)$$

$$\sigma_U^2 \sim \frac{1}{12} gh(g + h). \quad (14)$$

Б. ПЕРВЫЙ СЛУЧАЙ: g и h — величины
одинакового порядка

Для определенности мы предположим, что g и h стремятся к бесконечности таким образом, что отношение g/h остается постоянным. В этом случае Леманн² доказал, что случайная величина U распределена асимптотически нормально со средним значением ghp и квадратичным отклонением σ_U . Следовательно, вероятность $P(\mu)$ события $U > U_\beta$ асимптотически равна вероятности события $w > U_\beta$, где w — случайная величина, распределенная нормально со средним значением ghp и квадратичным отклонением σ_U :

$$P(\mu) \sim \Phi \left(\frac{ghp - U_\beta}{\sigma_U} \right). \quad (15)$$

¹ В обоих случаях $g \rightarrow \infty$ и $h \rightarrow \infty$, поэтому, согласно теореме, сформулированной в § 63 В, $U_\beta = O(\sigma_U) = O(h\sqrt{g})$. — Прим. перев.

² См. сноску на стр. 332.

Если (11) и (14) подставить в правую часть (15), то получим

$$P(\mu) \sim \Phi(b\mu - c), \quad (16)$$

где

$$b = \sqrt{\frac{3}{\pi} \frac{g\bar{h}}{g + \bar{h}}} \quad (17)$$

и постоянная c зависит от U_β . Напомним, что граница U_β была определена таким образом, чтобы вероятность события $U > U_\beta$ при $\mu = 0$ не превышала β и при $g \rightarrow \infty$ и $\bar{h} \rightarrow \infty$ стремилась к β . Поэтому вероятность (16) при $\mu = 0$ должна равняться β :

$$\Phi(-c) = \beta \quad (18)$$

Асимптотическая формула для функции мощности определяется формулами (16), (17) и (18).

В. СРАВНЕНИЕ С КРИТЕРИЕМ СТЬЮДЕНТА

В критерии Стьюдента для сравнения двух средних значений (§ 29) используется статистика

$$t = \frac{D}{S} = \frac{\bar{x} - \bar{y}}{S}, \quad (19)$$

где

$$\bar{x} = \frac{1}{g} (x_1 + \dots + x_g), \quad (20)$$

$$\bar{y} = \frac{1}{\bar{h}} (y_1 + \dots + y_{\bar{h}}), \quad (21)$$

$$S^2 = \left(\frac{1}{g} + \frac{1}{\bar{h}}\right) \frac{\sum (x - \bar{x})^2 + \sum (y - \bar{y})^2}{g + \bar{h} - 2}. \quad (22)$$

В формуле (19) среднее значение числителя D равно μ , в то время как среднее значение S^2 задается выражением

$$\sigma_D^2 = \left(\frac{1}{g} + \frac{1}{\bar{h}}\right) \sigma^2. \quad (23)$$

Числитель D представляет собой нормально распределенную случайную величину, у которой квадратичное отклонение является величиной того же порядка, что и μ . Так как квадратичное отклонение знаменателя S является величиной более высокого порядка малости, чем квадратичное отклонение числителя, то отсюда следует, что отношение t имеет асимптотически ту же самую функцию распределения, что и отношение D/σ_D . Таким образом, функция распределения t асимптотически равна функции распределения отношения

$$t' = \frac{\bar{x} - \bar{y}}{\sigma_D}.$$

Но случайная величина t' распределена нормально со средним

значением μ/σ_D и единичной дисперсией. Следовательно, вероятность события $t' > c$ равна

$$\Phi\left(\frac{\mu}{\sigma_D} - c\right). \quad (24)$$

При $\mu = 0$ снова должно иметь место равенство $\Phi(-c) = \beta$, следовательно, постоянная c принимает то же самое значение, что и раньше. Если в формулу (24) подставить (23), то получим асимптотическую формулу для функции мощности критерия Стьюдента

$$P'(\mu) \sim \Phi(b'\mu - c), \quad (25)$$

где

$$b' = \sqrt{\frac{gh}{g+h}}. \quad (26)$$

Сравнение (16) с (25) показывает, что асимптотически функции мощности $P(\mu)$ и $P'(\mu)$ при больших g и h отличаются друг от друга лишь множителем $\sqrt{3/\pi}$ в коэффициенте при μ . Иными словами, если критерий Вилкоксона применяется к выборкам объема g и h , а критерий Стьюдента — к выборкам объема g' и h' и если

$$g' \sim \frac{3}{\pi} g \text{ и } h' \sim \frac{3}{\pi} h, \quad (27)$$

то функции мощности $P(\mu)$ и $P'(\mu)$ будут асимптотически равны друг другу. Указанное свойство можно выразить и так: *асимптотическая эффективность критерия Вилкоксона равна $3/\pi$* . Это означает, что для критерия Стьюдента, который является наиболее мощным, объемы выборок g и h нужно уменьшить в $3/\pi$ раз, чтобы получить такую же функцию мощности, какую имеет критерий Вилкоксона, построенный по выборкам объема g и h .

Так как отношение $3/\pi$ приближенно равно $21/22$, то можно также сказать, что критерий Вилкоксона, примененный к 22 наблюдениям, приблизительно равноценен по мощности критерию Стьюдента, примененному к 21 наблюдению. Вследствие этого потеря мощности при переходе к критерию Вилкоксона оказывается очень малой.

Большим преимуществом критерия Вилкоксона является, конечно, возможность его применения в случае распределений, отличных от нормального. К тому же он требует значительно меньших вычислений, чем критерий Стьюдента. При больших g и h потеря мощности будет совсем незначительной и полностью окупается этими двумя преимуществами.

Г. ВТОРОЙ СЛУЧАЙ: h велико сравнительно с g

Пусть теперь h велико сравнительно с g . Если мы сначала g зафиксируем, то можно будет применить методы из § 63 Г. Сперва предположим, что $g = 2$ и что x_1 и x_2 — фиксированные числа.

Вероятность события $y < x_1$ равна $G(x_1)$, а частота этого события, как и (12) § 63, задается отношением

$$v_1 = \frac{u_1}{h}.$$

Так как частота события по вероятности стремится к вероятности события, то величина v_1 близка к $G(x_1)$ и точно так же v_2 близка к $G(x_2)$, следовательно, отношение

$$v_1 + v_2 = \frac{u_1 + u_2}{h} = \frac{U}{h}$$

близко к

$$G(x_1) + G(x_2).$$

Поэтому вероятность события $U < u$ или

$$\frac{U}{h} < \frac{u}{h} = t \quad (28)$$

асимптотически равна вероятности события

$$G(x_1) + G(x_2) < t. \quad (29)$$

Таким образом, функция распределения случайной величины U/h асимптотически равна функции распределения суммы двух (в общем случае — суммы g) независимых случайных величин, каждая из которых распределена так же, как $G(x_1)$, где x_1 — случайная величина с функцией распределения $F(t)$. Функция распределения случайной величины $G(x_1)$ равна вероятности события $G(x_1) < t$ или события $x_1 < G^{-1}(t)$, где G^{-1} — функция, обратная функции G . Таким образом, функция распределения для $G(x_1)$ задается формулой

$$K(t) = F[G^{-1}(t)]. \quad (30)$$

Если мы снова предположим, что x_i и y_k независимы и подчиняются нормальному распределению с единичной дисперсией и математическим ожиданием μ (для x) и 0 (для y), то

$$F(t) = \Phi(t - \mu), \quad G(t) = \Phi(t),$$

следовательно,

$$K(t) = \Phi[\Psi(t) - \mu],$$

где Ψ — функция, обратная функции нормального распределения Φ .

Согласно нулевой гипотезе, $F = G$ (или $\mu = 0$), поэтому $K(t) = t$, т. е. $G(x_1)$ подчиняется равномерному распределению в интервале $(0,1)$. При малых μ , а также и во всех тех случаях, когда F не сильно отличается от G , отклонение $K(t)$ от функции равномерного распределения будет небольшим. Во всяком случае, величина $G(x_1)$ заключена между нулем и единицей, поэтому ее распре-

деление ограничено и, значит, существуют моменты всех порядков.

Рассмотрим теперь распределение суммы $G(x_1) + G(x_2)$ или — в общем случае — распределение суммы

$$G(x_1) + \dots + G(x_g).$$

Согласно центральной предельной теореме (§ 24 Г), при больших g эта сумма распределена приблизительно нормально. При этом g не обязательно должно быть очень большим: уже при умеренных значениях g приближение оказывается очень хорошим. В том случае, когда $G(x_i)$ подчиняются равномерному распределению, аппроксимация становится отличной для всех g , начиная с $g = 4$. Если $K(t)$ несколько отклоняется от функции равномерного распределения, то качество приближения будет лишь немного хуже. Таким образом, если $g \geq 4$, то распределение отношения

$$v_1 + \dots + v_g = \frac{u_1 + \dots + u_g}{h} = \frac{U}{h}$$

мало отличается от нормального распределения.

Для того чтобы утверждение об асимптотическом распределении было теоретически правильным, нужно g устремить к бесконечности. При этом безразлично, остается ли отношение h/g ограниченным или оно стремится к бесконечности, так как в обоих случаях U распределено асимптотически нормально. Если μ не слишком велико, то для практических целей нормальное приближение оказывается вполне удовлетворительным уже при g и $h \geq 4$ и $g + h \geq 20$.

д. ДРУГИЕ СЛУЧАИ

При малых g (например, $g = 2$) и больших h можно применять этот же метод; нельзя только распределение суммы $G(x_1) + G(x_2)$ заменять нормальным распределением, а нужно воспользоваться точным распределением, для вычисления которого следует применить теорему III, § 4 Г. Если все x_i и y_k независимы и нормально распределены с единичной дисперсией и средними значениями μ (для x) и 0 (для y) и если $g = 2$ и $h \rightarrow \infty$, то при $\beta = 0,05$, $\mu = 1,5$ функции мощности критериев Вилкоксона и Стьюдента имеют значения¹ соответственно

$$P(\mu) = 0,64, \quad P'(\mu) = 0,68. \quad (31)$$

При больших g и h функции $P(\mu)$ и $P'(\mu)$ можно вычислять по формулам (16) и (25), где b и b' определяются формулами (17) и

¹ Van der Waerden B. L., Proc. Kon. Ned. Acad. Amsterdam. Series A, 55 (1952), 456.

(26), а c определяется равенством (18). Если положим снова $\beta = 0,05$ и выберем $b\mu = 2,03$, то получим, что $b'\mu = 2,08$ и

$$P(\mu) = 0,64, \quad P'(\mu) = 0,67, \quad (32)$$

т. е. почти тот же самый результат, что и (31).

Другим случаем, в котором легко осуществляются числовые оценки¹, является $g = h = 3$ и $\beta = 0,05$. В этом случае по критерию Вилкоксона нулевую гипотезу H_0 отвергают лишь тогда, когда x_1, x_2, x_3 и y_1, y_2, y_3 образуют последовательность

$$y y u x x x.$$

Уровень значимости при этом точно равен $\beta = 1/20$. При $\mu = 2$ функции мощности критериев Вилкоксона и Стьюдента принимают значения

$$P(\mu) = 0,62, \quad P'(\mu) = 0,65, \quad (33)$$

следовательно, разность этих значений столь же мала, как и в предыдущих случаях.

Однако при малых g и h критерий Вилкоксона обладает одним недостатком, вследствие которого мощность этого критерия в отдельных случаях существенно снижается. А именно, тогда, когда несколько перестановок имеют одинаковое число инверсий. Об этом говорилось в § 63 А и там же был указан соответствующий пример. Количество таких примеров можно увеличивать безгранично.

Пусть, например, $g = 4$, $h = 6$ и $\beta = 0,05$. Критерий Вилкоксона отвергает нулевую гипотезу в следующих случаях, имеющих 21 инверсию или больше:

1. $y y u y u y u x x x x$
2. $y y u y u x u x x x$
3. $y y u y x u y x x x$
4. $y y u y u x x u x x$
5. $y y u x u y u x x x$
6. $y y u y x u x u x x$
7. $y y u y u y x x u x$.

Так как заданный уровень значимости равен 0,05, то следовало бы отвергнуть $0,05 \cdot 210$ сочетаний, т. е. 10 сочетаний. Однако если к выписанным семи сочетаниям добавить все сочетания с 20 инверсиями, то получим 12 сочетаний; это количество слишком велико. Таким образом, критерий Вилкоксона с $\beta = 1/20$ не мощ-

¹ См. стр. 452 только что цитированной заметки.

нее того же критерия с $\beta = 1/30$, в то время как критерий Стьюдента с уровнем значимости $1/20$, конечно, значительно мощнее того же критерия с уровнем значимости $1/30$.

Точно так же можно показать, что в данном случае ($g = 4$, $h = 6$) критерий Вилкоксона с $\beta = 0,025$ не мощнее того же критерия с $\beta = 0,02$ или двусторонний критерий Вилкоксона с заданным уровнем значимости $0,05$ не мощнее двустороннего критерия с заданным уровнем значимости $0,04$ и т. д.

Мощность критерия можно было бы увеличить, если в сомнительных случаях вытаскивать карту из специально подобранной колоды и отвергать гипотезу H_0 в том случае, когда извлеченная карта окажется черной масти. Однако лучше воспользоваться более мощным критерием, а именно, критерием X , к изложению которого мы теперь и переходим.

§ 65. Критерий X

А. Эвристический вывод

Рассмотрим снова случай $g = 2$ и $h \rightarrow \infty$. Таким образом, пусть $x_1, x_2, y_1, \dots, y_h$ — результаты наблюдений и пусть u_1 — количество величин y_k , меньших чем x_1 , u_2 — количество величин y_k , меньших чем x_2 , v_1 — частота события $y < x_1$ и v_2 — частота события $y < x_2$, т. е.

$$v_1 = \frac{u_1}{h}, \quad v_2 = \frac{u_2}{h}. \quad (1)$$

Количество инверсий равно

$$U = u_1 + u_2. \quad (2)$$

Согласно критерию Вилкоксона, нулевая гипотеза отвергается тогда, когда $u_1 + u_2 > U_\beta$ или

$$v_1 + v_2 > b, \quad \text{где } b = \frac{U_\beta}{h}. \quad (3)$$

Вероятность этого события асимптотически равна вероятности события

$$G(x_1) + G(x_2) > b \quad (4)$$

(см. § 64 Г).

Сначала мы предположим, что y подчиняется нормальному распределению с нулевым средним значением и единичной дисперсией, т. е.

$$G(t) = \Phi(t). \quad (5)$$

Тогда вместо (4) можно написать

$$\Phi(x_1) + \Phi(x_2) > b. \quad (6)$$

Если x_1 и x_2 распределены нормально со средним значением $\mu \geq 0$ и единичной дисперсией, то равномерно наиболее мощный критерий для проверки нулевой гипотезы $\mu = 0$ отвергает эту гипотезу тогда, когда

$$x_1 + x_2 > c, \quad (7)$$

где постоянная c выбирается таким образом, чтобы вероятность события (7) в случае, если гипотеза $\mu = 0$ верна, точно равнялась β . Это приводит к условию

$$\Phi\left(\frac{c}{\sqrt{2}}\right) = 1 - \beta \quad (8)$$

или

$$c = \sqrt{2} \Psi(1 - \beta). \quad (9)$$

Критерий (7) имеет асимптотически ту же самую мощность, что и критерий Стьюдента. В этом можно убедиться непосредственно, вычислив мощность критерия (7) и сравнив ее с асимптотической оценкой мощности критерия Стьюдента, вычисленной в § 64 В.

Таким образом, различие асимптотических мощностей критериев Вилкоксона и Стьюдента возникает вследствие того, что в левой части (6) стоит $\Phi(x_1) + \Phi(x_2)$, тогда как левая часть (7) равна $x_1 + x_2$. Статистика $x_1 + x_2$ позволяет получить несколько лучший критерий.

Подстановка $v_i = G(x_i) = \Phi(x_i)$ переводит левую часть (3) в левую часть (6). Однако неравенство (3) можно легко видоизменить таким образом, чтобы в результате той же подстановки получалась левая часть (7). Для этого нужно лишь v_i формально заменить величинами $\Psi(v_i)$, где Ψ — функция, обратная функции Φ .

В результате получается видоизмененный критерий, согласно которому нулевая гипотеза отвергается тогда, когда сумма

$$S = \Psi(v_1) + \Psi(v_2) = \Psi\left(\frac{u_1}{h}\right) + \Psi\left(\frac{u_2}{h}\right) \quad (10)$$

превосходит надлежащим образом выбранную границу c . Если в (10) положить

$$v_i = \Phi(x_i), \quad (11)$$

то сумма S перейдет в $x_1 + x_2$ и получится критерий (7).

При произвольном g вместо (10) нужно рассматривать сумму

$$S = \Psi\left(\frac{u_1}{h}\right) + \Psi\left(\frac{u_2}{h}\right) + \dots + \Psi\left(\frac{u_g}{h}\right). \quad (12)$$

Однако соответствующий критерий будет иметь один недостаток, связанный с тем, что слагаемые суммы (12) могут обращаться

в $-\infty$ (при $u_i = 0$) или в $+\infty$ (при $u_i = h$); в этих случаях вычисление суммы невозможно. Для того чтобы преодолеть это затруднение, все x_i (а вместе с ними и u_i) располагают в порядке возрастания их величины, затем u_i заменяют числами

$$r_1 = u_1 + 1, \quad r_2 = u_2 + 2, \dots, \quad r_g = u_g + g, \quad (13)$$

и, наконец, в знаменатели вместо h подставляют

$$n + 1 = g + h + 1.$$

Таким образом, возникает окончательное выражение

$$X = \Psi\left(\frac{r_1}{n+1}\right) + \Psi\left(\frac{r_2}{n+1}\right) + \dots + \Psi\left(\frac{r_g}{n+1}\right). \quad (14)$$

Числа r_1, \dots, r_g , определенные равенствами (13), равны порядковым номерам x_1, \dots, x_g в общем вариационном ряду, составленном по объединенной выборке $x_1, \dots, x_g, y_1, \dots, y_h$. Порядковые номера r_i могут принимать лишь значения от 1 до $n = g + h$, поэтому слагаемые (14) никогда не обращаются в $\pm\infty$.

Если отвлечься от крайнего случая, когда в (12) некоторые u_i близки к 0 или к h (для очень больших h этот случай все равно является очень маловероятным), то окажется, что асимптотически при $h \rightarrow \infty$ сумма (14) ведет себя так же, как сумма (12). Изложенный выше эвристический вывод, который сперва привел нас к сумме S и критерию $S > c$, приводит, таким образом, к более удобной для приложений сумме X и к следующему критерию.

Б. КРИТЕРИЙ X

Пусть $n = g + h$ случайных величин x_1, \dots, x_g и y_1, \dots, y_h расположены в порядке их возрастания, и пусть r_i (или просто r) — порядковый номер x_i и s_k (или просто s) — порядковый номер y_k . Образует сумму

$$X = \sum_r \Psi\left(\frac{r}{n+1}\right), \quad (15)$$

$$Y = \sum_s \Psi\left(\frac{s}{n+1}\right). \quad (16)$$

В интервале $0 < t < 1$ функция $\Psi(t)$ принимает лишь конечные значения и удовлетворяет условию

$$\Psi(1-t) = -\Psi(t), \quad (17)$$

поэтому сумма $X + Y$ всегда равна нулю:

$$X + Y = \Psi\left(\frac{1}{n+1}\right) + \Psi\left(\frac{2}{n+1}\right) + \dots + \Psi\left(\frac{n}{n+1}\right) = 0. \quad (18)$$

При перемене ролей x и y величина X переходит в $Y = -X$. Если затем изменить еще и порядок следования (т. е. расположить все x и y не в порядке их возрастания, а в порядке убывания), то $-X$ снова перейдет в X .

Границу X_β следует определить таким образом, чтобы вероятность события

$$X > X_\beta, \quad (19)$$

вычисленная в предположении, что все $n!$ перестановок из $x_1, \dots, x_g, y_1, \dots, y_h$ являются равновероятными, была наибольшей и при этом не превосходила β . Указанное предположение является следствием *нулевой гипотезы* H_0 , которая утверждает, что все x_i и y_k независимы и обладают одинаковыми функциями распределения $F(x)$. Сначала эту функцию мы будем считать непрерывной и поэтому возможность осуществления таких событий, как $x_i = y_k$, можно не принимать в расчет.

Согласно *одностороннему критерию* X , нулевая гипотеза отвергается, коль скоро сумма X превосходит границу X_β . Этот критерий применяется тогда, когда интересуются, не будут ли x_i , вообще говоря, больше, чем y_k ? Уровень значимости этого критерия не превосходит β .

Согласно *двустороннему критерию* X , нулевая гипотеза отвергается тогда, когда X или Y превосходят границу X_β . Если говорится, что $X > X_\beta$, то считают, что x в среднем больше, чем y . В противоположность этому если $Y > X_\beta$, то полагают, что y в среднем больше, чем x . Уровень значимости двустороннего критерия не превосходит 2β .

В. ВЫЧИСЛЕНИЕ X_β

При малых g и h границу X_β можно вычислить точно с помощью непосредственного перечисления равновозможных сочетаний. В качестве примера снова рассмотрим случай $g = 4$, $h = 6$ и получим

$$\beta = \frac{1}{40} = 0,025, \quad \text{следовательно,} \quad 2\beta = \frac{1}{20} = 0,05.$$

Количество различных сочетаний вида $x y y \dots x$ равно $\binom{10}{4} = 210$. Сороковая часть этого количества¹ равна 5. Таким образом,

¹ Как и в случае критерия Вилкоксона, здесь идет речь о целой части $\left[\beta \binom{n}{g} \right]$. — *Прим. перев.*

мы должны выписать 5 таких сочетаний, для которых соответствующие значения X являются наибольшими.

Сначала составим таблицу значений Ψ , округленных до двух десятичных знаков:

$$\begin{array}{ll} \Psi \left(\begin{array}{c} 1 \\ 11 \end{array} \right) = -1,34 & \Psi \left(\begin{array}{c} 6 \\ 11 \end{array} \right) = 0,11 \\ \Psi \left(\begin{array}{c} 2 \\ 11 \end{array} \right) = -0,91 & \Psi \left(\begin{array}{c} 7 \\ 11 \end{array} \right) = 0,35 \\ \Psi \left(\begin{array}{c} 3 \\ 11 \end{array} \right) = -0,60 & \Psi \left(\begin{array}{c} 8 \\ 11 \end{array} \right) = 0,60 \\ \Psi \left(\begin{array}{c} 4 \\ 11 \end{array} \right) = -0,35 & \Psi \left(\begin{array}{c} 9 \\ 11 \end{array} \right) = 0,91 \\ \Psi \left(\begin{array}{c} 5 \\ 11 \end{array} \right) = -0,11 & \Psi \left(\begin{array}{c} 10 \\ 11 \end{array} \right) = 1,34. \end{array}$$

С помощью этой таблицы для каждого такого сочетания, как $уууууууууууууууу$, можно теперь вычислить значение X по формуле (15). Шестью сочетаниями с наибольшими значениями X являются

1. $уууууууууууууууу$ $X = 3,31$
2. $уууууууууууууууу$ $X = 2,96$
3. $уууууууууууууууу$ $X = 2,74$
4. $уууууууууууууууу$ $X = 2,71$
5. $уууууууууууууууу$ $X = 2,50$
6. $уууууууууууууууу$ $X = 2,49$

Если мы положим $X_p = 2,49$, то лишь для пяти сочетаний значения X будут превосходить X_p . При этом предполагается, что в практических применениях критерия X значения X вычисляются лишь с двумя десятичными знаками и что нулевая гипотеза отвергается только тогда, когда вычисленное значение X оказывается строго больше, чем X_p .

Для контроля целесообразно наряду с X вычислять также и Y . Сумма $X + Y$ должна быть точно равна нулю (даже при округленных значениях Ψ).

Указанное здесь перечисление всех возможных случаев практически удобно лишь при $g + h \leq 20$. При больших значениях g и h приходится переходить к асимптотическим оценкам.

Г. СРЕДНЕЕ ЗНАЧЕНИЕ И ДИСПЕРСИЯ X

Обозначим значения Ψ буквами a_1, \dots, a_n :

$$a_i = \Psi\left(\frac{i}{n+1}\right). \quad (20)$$

Согласно (15), статистика X представляет собой сумму

$$X = \sum a_r = a_{r_1} + a_{r_2} + \dots + a_{r_g}, \quad (21)$$

где r_1, \dots, r_g — некоторая выборка объема g , извлеченная из совокупности n возможных индексов $i = 1, 2, \dots, n$. В случае справедливости нулевой гипотезы все такие выборки равновероятны.

Каждое отдельное слагаемое a_r суммы (21) принимает значения a_1, \dots, a_n с одинаковыми вероятностями. Следовательно, математическое ожидание каждого a_r равно нулю, а поэтому

$$\mathcal{E} X = 0. \quad (22)$$

Для того чтобы вычислить дисперсию случайной величины X, определим сначала среднее значение для a_i^2 . Так как a_i^2 принимает значения a_1^2, \dots, a_n^2 с одинаковыми вероятностями, то

$$\mathcal{E} a_i^2 = \frac{1}{n} (a_1^2 + \dots + a_n^2) =: Q. \quad (23)$$

Определим теперь среднее значение произведения $a_{r_1} a_{r_2}$. Это произведение принимает значения вида $a_i a_k$ ($i \neq k$) с одинаковыми вероятностями, поэтому

$$\begin{aligned} \mathcal{E}(a_{r_1} a_{r_2}) &= \frac{1}{n(n-1)} \sum_{i \neq k} a_i a_k = \frac{1}{n(n-1)} [(\sum a_i)(\sum a_i) - \sum a_i^2] = \\ &= -\frac{1}{n(n-1)} \sum a_i^2 = -\frac{Q}{n-1}. \end{aligned} \quad (24)$$

Если (21) возвести в квадрат и вычислить среднее значение, то получим

$$\mathcal{E} X^2 = g \mathcal{E} a_i^2 + g(g-1) \mathcal{E} a_{r_1} a_{r_2} = gQ - \frac{g(g-1)}{n-1} Q =: \frac{g(n-g)}{n-1} Q$$

или

$$\sigma_X^2 = \frac{gh}{n-1} Q. \quad (25)$$

При этом Q, согласно (23), определяется формулой

$$Q = \frac{1}{n} \sum_{i=1}^n \left[\Psi\left(\frac{i}{n+1}\right) \right]^2.$$

Среднее значение и дисперсия случайной величины X задаются формулами (22) и (25). Величины Q табулированы в табл. 12.

Д. АСИМПТОТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ X

Пусть x_1, \dots, x_g и y_1, \dots, y_h — независимые, одинаково распределенные случайные величины, и пусть сначала h велико сравнительно с g . При этом безразлично, велико g или нет. В таком случае справедлива теорема: *случайная величина X распределена асимптотически нормально с нулевым средним значением и квадратичным отклонением σ_X* . Доказательство не представляет трудности; его можно найти в моей работе о критерии X , опубликованной в журнале *Math. Ann.*, 126 (1953), 94.

Согласно только что сформулированной теореме для каждого $\epsilon > 0$ найдется также M , что при всех $h/g > M$ функция распределения X будет отличаться от соответствующей функции нормального распределения менее чем на ϵ . Это же самое справедливо и для $g/h > M$. Таким образом, мы должны еще рассмотреть лишь случай, когда оба отношения g/h и h/g не превосходят M и при этом $n = g + h$ безгранично возрастает. В этом случае теорема об асимптотической нормальности также справедлива, но доказываемая она много тяжелее. В реферате моей только что цитированной работы (реферативный журнал *Math. Reviews*, 15, 46, референт G. E. Noether) отмечается, что доказательство можно провести с помощью одной теоремы Вальда и Вольфовица, которую в свою очередь можно доказать методом моментов, указанным в § 63 В. Полностью это доказательство провел Стокер в своей Амстердамской диссертации: [см. *Stoker, D. J. Oor 'n klas van toetsings-groothede vir die probleem van twee steekproewe* (1955)].

Поэтому при $n \rightarrow \infty$ случайная величина X распределена асимптотически нормально, независимо от того, стремятся в отдельности g и h к бесконечности или нет. Таким образом, статистика X и число инверсий U асимптотически ведут себя различно.

На основе этой теоремы были вычислены таблицы¹ для критерия X . При этом для малых n (т. е. для малых g и h) граница X_p определялась точно, путем перечисления возможных случаев. Для больших n использовалось асимптотическое нормальное распределение. При этом особое внимание уделялось членам a_1 и a_n , которые могут встретиться среди слагаемых суммы (21). Это позволило существенно улучшить приближение.

¹ Van der Waerden B. L. und Nievergelt E., *Tafeln zum Vergleich zweier Stichproben mittels X-Test und Zeichentest*, Springer-Verlag, 1956.

Е. СЛУЧАЙ, КОГДА НЕКОТОРЫЕ x И y МОГУТ
БЫТЬ РАВНЫМИ

До сих пор мы предполагали, что x и y обладают непрерывными функциями распределения и отсюда следовало, что возможность осуществления события $x_i = y_k$ можно не принимать в расчет. Однако на практике x_i и y_k всегда представляются округленными числами и, следовательно, имеют дискретное распределение; поэтому вполне возможен случай, когда $x_i = y_k$. Спрашивается, как в таком случае нужно определять порядковые номера r_i и s_k , которые используются при вычислении X и Y по формулам (15) и (16)? Такой же вопрос возникает также и в случае критерия Вилкоксона.

Были предложены различные методы. Например, для того чтобы решить, какую из двух равных величин x_i и y_k считать большей, можно бросать монету. Можно также условиться приписывать средний порядковый номер $r + 1/2$ тем равным величинам $x_i = y_k$, которые, в случае их неравенства, должны были бы иметь порядковые номера r и $r + 1$. Однако наилучшим оказывается следующий метод.

Мы рассмотрим сейчас наиболее общий случай, когда имеется $c = a + b$ равных величин x_1, \dots, x_a и y_1, \dots, y_b , занимающих места с порядковыми номерами $r, r + 1, \dots, r + c - 1$. Расположим величины $x_1, \dots, x_a, y_1, \dots, y_b$ на имеющихся в нашем распоряжении c местах всеми $c!$ возможными способами, для каждой такой перестановки вычислим X и из всех полученных значений X образуем арифметическое среднее.

При практических расчетах этот метод можно упростить. Следует суммировать не $c!$ слагаемых, а лишь c . А именно, нужно по имеющимся в нашем распоряжении порядковым номерам построить сумму

$$S_c = \Psi\left(\frac{r}{n+1}\right) + \Psi\left(\frac{r+1}{n+1}\right) + \dots + \Psi\left(\frac{r+c-1}{n+1}\right) \quad (26)$$

и к суммам X и Y , вычисленным по остальным порядковым номерам, добавить дроби, соответственно равные

$$\frac{a}{a+b} S_c \quad \text{и} \quad \frac{b}{a+b} S_c. \quad (27)$$

Если в другом месте имеется еще $c' = a' + b'$ равных друг другу величин x_i и y_k , то вычисляют аналогичное выражение и т. д. Суммированием всех таких выражений¹ получают X и соответственно Y .

¹ Этот прием формально применим и в случае $c = 1$, т. е. тогда, когда два соседних члена вариационного ряда не равны друг другу. Поэтому X и Y можно считать суммами выражений вида (27). — *Прим. перев.*

Эта модификация оказывает лишь небольшое влияние на функцию распределения X . Квадратичное отклонение X (а вместе с ним, вероятно, и истинный уровень значимости критерия) несколько уменьшается. Следовательно, если граница X_β остается неизменной, то указанная модификация лишь увеличивает надежность критерия.

Ж. СРАВНЕНИЕ С КРИТЕРИЕМ СТЬЮДЕНТА

Предположим теперь, что все x_i и y_k независимы и распределены нормально с единичной дисперсией и средними значениями $\mu \geq 0$ (для x) и 0 (для y). Далее, предположим, что g фиксировано и h стремится к бесконечности. В этих предположениях мы хотим найти асимптотическую оценку для мощности критерия X и сравнить ее с мощностью критерия Стьюдента.

Функция мощности $P(\mu)$ критерия X определяется как вероятность события

$$\Psi\left(\frac{r_1}{n+1}\right) + \Psi\left(\frac{r_2}{n+1}\right) + \dots + \Psi\left(\frac{r_g}{n+1}\right) > X_\beta. \quad (28)$$

Так как все перестановки величин x_1, \dots, x_g являются равновероятными, то мы можем считать, что $x_1 < x_2 < \dots < x_g$. В этом случае формулы (13) снова оказываются справедливыми, и вместо (28) можно записать

$$\Psi\left(\frac{u_1+1}{n+1}\right) + \Psi\left(\frac{u_2+1}{n+1}\right) + \dots + \Psi\left(\frac{u_g+g}{n+1}\right) > X_\beta. \quad (29)$$

Если в этом неравенстве, согласно (1), положить

$$u_i = hv_i$$

и $n = g + h$, то получим

$$\Psi\left(\frac{hv_1+1}{h+g+1}\right) + \Psi\left(\frac{hv_2+2}{h+g+1}\right) + \dots + \Psi\left(\frac{hv_g+g}{h+g+1}\right) > X_\beta. \quad (30)$$

Теперь мы так же, как в § 64 Г, заменим частоты v_i близкими к ним вероятностями. Тогда выражение

$$\Psi\left(\frac{hv_i+i}{h+g+1}\right)$$

перейдет в

$$\Psi\left[\frac{h\Phi(x_i)+i}{h+g+1}\right].$$

Если, кроме того, в числителе и знаменателе пренебречь теми членами, которые малы сравнительно с h , то получим

$$\Psi[\Phi(x_i)] = x_i$$

и (30) перейдет в

$$x_1 + x_2 + \dots + x_g > X_\beta. \quad (31)$$

Наконец, если при $h \rightarrow \infty$ X_β заменить соответствующим асимптотическим выражением, то (31) перейдет в неравенство

$$x_1 + x_2 + \dots + x_g > \sqrt{g} \Psi(1 - \beta). \quad (32)$$

Этот результат является обобщением критерия (7) на случай произвольного g . В том, что мы вернулись к этому критерию, нет ничего удивительного, так как в разделе А этого параграфа мы вывели критерий X , исходя из неравенства (7). В данном же случае мы шли тем же путем, но только в обратном направлении.

Только что полученный результат показывает, что асимптотически при $h \rightarrow \infty$ критерий X имеет ту же самую функцию мощности, что и критерий (32). Как мы уже видели, среди всех критериев для проверки нулевой гипотезы, обладающих точным уровнем значимости β , критерий (32) является равномерно наиболее мощным.

Легко можно вычислить функцию мощности критерия (32). Находим

$$P'(\mu) = \Phi(b'\mu - c), \quad (33)$$

где

$$b' = \sqrt{g} \quad \text{и} \quad c = \Psi(1 - \beta). \quad (34)$$

Следовательно, асимптотически функция мощности критерия X задается формулой (33). В § 64 В мы видели, что функция мощности критерия Стьюдента также задается асимптотической формулой (33). Таким образом, *при постоянном g и при $h \rightarrow \infty$ критерий X имеет такую же мощность, как и критерий Стьюдента.*

В данном случае я хотел лишь изложить основные идеи и наметить путь доказательства. Точный вывод можно найти в работе, цитированной выше (*Math. Ann.*, 126, § 5, 103).

Я подозреваю, что этот же результат останется справедливым и тогда, когда оба параметра g и h стремятся к бесконечности.

3. РАСПРЕДЕЛЕНИЯ, ОТЛИЧНЫЕ ОТ НОРМАЛЬНОГО

Критерий Стьюдента предназначен для проверки нулевой гипотезы в предположении, что в обеих выборках случайные величины распределены нормально с одинаковыми дисперсиями. Большим преимуществом порядковых критериев является их полная независимость от предположения нормальности. При этом, независимо от выбора непрерывной функции распределения $F(x)$, уровень значимости таких критериев всегда не превосходит β .

Что касается критерия Стьюдента, то его истинный уровень значимости может превышать β , если только распределения x и y отличны от нормального.

Однако при соответствующих предположениях о функции $F(x)$ и при больших g и h превышение заданного уровня значимости критерия Стьюдента оказывается не очень значительным. А именно, при достаточно большом g выборочное среднее \bar{x} независимых случайных величин x_1, \dots, x_g распределено приблизительно нормально, точно так же распределено и выборочное среднее \bar{y} при достаточно большом h , следовательно, разность $D = \bar{x} - \bar{y}$ распределена приблизительно нормально: При больших $n = g + h$ знаменатель S стьюдентовского отношения можно приблизительно заменить истинным квадратичным отклонением σ_D разности D . Таким образом, отношение D/S распределено приблизительно нормально с квадратичным отклонением, стремящимся к единице при $g + h \rightarrow \infty$. Поэтому если все x и y имеют одинаковые и не слишком дикое функции распределения $F(t)$ и если g и h велики, то истинный уровень значимости критерия Стьюдента будет приблизительно равен заданному значению β .

Однако при распределениях, отличных от нормального, критерий Стьюдента, в противоположность порядковым критериям, обладает другим недостатком, а именно незначительной мощностью. В § 6 моей уже упоминавшейся работы (*Math. Ann.*, 126, 106) я рассматривал случай, когда функции распределения F и G случайных величин x и y устроены таким образом, что при некотором однозначном преобразовании случайных величин

$$x' = \tau(x), \quad y' = \tau(y) \quad (35)$$

оба распределения переходят в нормальные распределения с равными дисперсиями, но различными средними значениями. Мощность критерия X точно так же, как и мощность любого другого порядкового критерия, при таком преобразовании остается, конечно, неизменной. Что касается функции мощности критерия Стьюдента, то она вследствие преобразования (35) может значительно уменьшиться. В частности, такое уменьшение происходит тогда, когда благодаря преобразованию (35) квадратичные отклонения σ_x и σ_y увеличиваются сильнее, чем разность средних значений $\hat{x} - \hat{y}$.

В одной из следующих заметок (*Proc. Kon. Akad. Amsterdam*, A 56, 311) я рассмотрел другой случай, когда x_1, \dots, x_4 распределены равномерно между нулем и единицей, а y_1, \dots, y_6 распределены равномерно между нулем и $1 + \mu$. При этом оказалось, что при $\mu \rightarrow \infty$ функции мощности порядковых критериев стремятся к единице, а функция мощности критерия Стьюдента к единице не стремится.

Мне и на практике приходилось иметь дело со случаями, в которых гипотеза о нормальном распределении x и y с равными дисперсиями заведомо не имела места и в которых критерий X нулевую гипотезу отвергал, в то время как критерий Стьюдента (с тем же уровнем значимости) не позволял сделать такой же вывод.

Пример 25. На одном промышленном предприятии измерялось время простоев, подверженное сильному рассеянию. Числовые значения y , к сожалению, забыл; поэтому мы воспользуемся теми данными, которые указаны в упоминавшихся ранее таблицах ван дер Вардена и Нивергельта:

$$x_1 = 11, \quad x_2 = 34, \quad x_3 = 13, \quad x_4 = 18.$$

После реорганизации производства время простоев сократилось и рассеяние уменьшилось, например:

$$y_1 = 8, \quad y_2 = 10, \quad y_3 = 7, \quad y_4 = 6.$$

В данном случае возможность применения критерия Стьюдента является весьма сомнительной, так как результаты наблюдений показывают, что распределения едва ли являются нормальными и что дисперсии не равны друг другу; кроме того, g и h не очень велики. Если тем не менее все-таки применить критерий Стьюдента (двусторонний критерий с уровнем значимости 0,05), то нулевая гипотеза не будет отвергнута: отношение t в данном случае принимает значение 2,1, а соответствующая граница (по табл. 7) равна 2,4.

Критерий Вилкоксона нулевую гипотезу немедленно отвергает. Так как все x больше любого из y , то количество инверсий равно 16. Для того чтобы применить табл. 10, нужно x и y поменять местами; в этом случае количество инверсий станет равным нулю. В столбце (4; 4) при $u = 0$ находим, что вероятность события $U = 0$ равна 0,0143. Так как эта вероятность меньше чем 0,025, то, согласно одностороннему критерию с $\beta = 0,025$ или согласно двустороннему критерию с $2\beta = 0,05$, нулевую гипотезу следует отвергнуть.

Для применения критерия X нужно все x и y расположить в порядке возрастания их величины (при этом x_i будут иметь порядковые номера 5, 8, 6 и 7) и вычислить

$$\begin{aligned} X &= \Psi\left(\frac{5}{9}\right) + \Psi\left(\frac{8}{9}\right) + \Psi\left(\frac{6}{9}\right) + \Psi\left(\frac{7}{9}\right) = \\ &= 0,14 + 1,22 + 0,43 + 0,76 = 2,55. \end{aligned}$$

Для вычисления X можно воспользоваться табл. 2, в конце этой книги, или более удобной табл. 2 ван дер Вардена—Нивергельта. При $n = 8$ и $g - h = 0$ двусторонняя 5%-ная граница равна 2,40 (табл. 11). Следовательно, по критерию X нулевую гипотезу нужно также отвергнуть.

КОРРЕЛЯЦИЯ

В этой главе будет предполагаться известным лишь содержание первых шести глав.

§ 66. Ковариация и коэффициент корреляции

А. ИСТИННЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Если x и y — две зависимые случайные величины, то дисперсия суммы $\lambda x + y$, помимо дисперсий слагаемых λx и y , содержит еще член, линейный относительно λ :

$$\begin{aligned} \mathfrak{E}(\lambda x + y - \lambda \hat{x} - \hat{y})^2 &= \lambda^2 \mathfrak{E}(x - \hat{x})^2 + \\ &+ 2\lambda \mathfrak{E}(x - \hat{x})(y - \hat{y}) + \mathfrak{E}(y - \hat{y})^2. \end{aligned} \quad (1)$$

Коэффициент при 2λ в правой части (1) называется *ковариацией* случайных величин x и y . Если ковариацию разделить на произведение квадратичных отклонений $\sigma_x \sigma_y$, которые предполагаются отличными от нуля, то получится *истинный коэффициент корреляции* ρ :

$$\rho = \frac{\mathfrak{E}(x - \hat{x})(y - \hat{y})}{\sigma_x \sigma_y}. \quad (2)$$

С помощью (2) формулу (1) можно записать так:

$$\sigma_{\lambda x + y}^2 = \lambda^2 \sigma_x^2 + 2\lambda \rho \sigma_x \sigma_y + \sigma_y^2. \quad (3)$$

Коэффициент корреляции тесно связан с *коэффициентом регрессии* γ , который определяется следующим образом. Положим

$$y = \gamma x + z \quad (4)$$

и определим γ так, чтобы дисперсия z была наименьшей. Дисперсию разности $z = y - \gamma x$ можно вычислить по формуле (3), заменив λ на $-\gamma$:

$$\sigma_z^2 = \gamma^2 \sigma_x^2 - 2\gamma \rho \sigma_x \sigma_y + \sigma_y^2. \quad (5)$$

Правая часть (5) представляет собой многочлен относительно γ , достигающий минимума в точке

$$\gamma = \rho \frac{\sigma_y}{\sigma_x}. \quad (6)$$

Формула (6) связывает коэффициент регрессии γ с коэффициентом корреляции ρ .

Значение минимума многочлена (5) равно

$$\sigma_z^2 = \rho^2 \sigma_y^2 - 2\rho^2 \sigma_y^2 + \sigma_y^2 = (1 - \rho^2) \sigma_y^2. \quad (7)$$

Из (7) непосредственно следует, что

$$1 - \rho^2 \geq 0.$$

Таким образом, значение коэффициента корреляции заключено в пределах -1 и $+1$.

Если ρ принимает одно из крайних значений, равных ± 1 , то, согласно (7), $\sigma_z = 0$. В силу последней теоремы из § 3, это возможно лишь тогда, когда z с вероятностью единица является постоянной величиной, т. е. когда y с вероятностью единица представляет собой линейную функцию от x :

$$y = \gamma x + \alpha. \quad (8)$$

Коэффициент корреляции ρ является мерой зависимости (мерой линейной зависимости) между x и y . В случае независимости этих величин $\rho = 0$. Если же x и y связаны точной линейной зависимостью (8), то $\rho = \pm 1$; при этом знак ρ , в силу (6), всегда равен знаку коэффициента регрессии γ .

Смысл коэффициента корреляции можно выяснить с помощью анализа дисперсии величины y . Из формулы (4) видно, что y является суммой двух случайных величин γx и z , из которых первая (γx) пропорциональна x , а вторая (z) с x некоррелирована, так как ковариация x и z равна нулю. Таким образом, дисперсия y представляет собой сумму дисперсий γx и z :

$$\sigma_y^2 = \gamma^2 \sigma_x^2 + \sigma_z^2. \quad (9)$$

Если в эту формулу вместо γ и σ_z^2 подставить (6) и (7), то первое слагаемое правой части (9) будет равно $\rho^2 \sigma_y^2$, а вторым слагаемым будет $(1 - \rho^2) \sigma_y^2$. Как и следовало ожидать, сумма этих слагаемых равна σ_y^2 . Таким образом, ρ^2 показывает, какая часть дисперсии случайной величины y приходится на долю слагаемого γx в формуле (4).

Б. ВЫБОРОЧНЫЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Если в результате наблюдений получены n пар значений $(x_1, y_1), \dots, (x_n, y_n)$ и если предполагается, что пары (x_i, y_i) являются независимыми двумерными случайными величинами с одинаковым двумерным распределением, то в качестве оценки для дисперсии $\sigma_{\lambda x + y}^2$ можно применить *выборочную дисперсию*

$$\begin{aligned} s_{\lambda x + y}^2 &= \frac{1}{n-1} \sum (\lambda x_i + y_i - \lambda \bar{x} - \bar{y})^2 = \\ &= \lambda^2 \frac{\sum (x - \bar{x})^2}{n-1} + 2\lambda \frac{\sum (x - \bar{x})(y - \bar{y})}{n-1} + \frac{\sum (y - \bar{y})^2}{n-1}. \end{aligned} \quad (10)$$

Поэтому в качестве оценки для ковариации $\mathcal{E}(x - \hat{x})(y - \hat{y})$ естественно воспользоваться *выборочной ковариацией*

$$\frac{1}{n-1} \sum (x - \bar{x})(y - \bar{y}), \quad (11)$$

где, как всегда, \bar{x} и \bar{y} являются выборочными средними:

$$\bar{x} = \frac{1}{n} \sum x, \quad \bar{y} = \frac{1}{n} \sum y. \quad (12)$$

Так как (10) представляет собой несмещенную оценку, то оценка (11) также является несмещенной.

Для того чтобы получить оценку для ρ , разделим (11) на $s_x s_y$. Такая оценка называется *выборочным коэффициентом корреляции*

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n-1)s_x s_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}. \quad (13)$$

Можно показать, что r обладает теми же свойствами, какими обладает указанный выше истинный коэффициент корреляции ρ . Положим $\lambda = -c$ и выберем c таким образом, чтобы значение многочлена (10) стало наименьшим. Точка минимума и минимальное значение многочлена задаются формулами

$$c = r \frac{s_x}{s_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} \quad (14)$$

и

$$s_{y-cx}^2 = (1 - r^2)s_y^2. \quad (15)$$

Так как выражение (15) всегда неотрицательно, то значение r всегда заключено между -1 и $+1$. Если $r = \pm 1$, то, в силу (15), все $y_i - cx_i$ принимают одинаковое значение a , т. е. все наблюдаемые точки с координатами (x_i, y_i) лежат на прямой с уравнением

$$y = cx + a.$$

Если же эти точки не лежат на одной прямой, то через точку с координатами (\bar{x}, \bar{y}) можно провести прямую, угловой коэффициент которой задается равенством (14). Эта прямая представляет собой *эмпирическую линию регрессии*, о которой говорилось выше, в § 33,

$$y - \bar{y} = c(x - \bar{x}). \quad (16)$$

В § 33 эта линия определялась таким образом, чтобы сумма квадратов отклонений точек (x_i, y_i) от прямой была наименьшей (отклонения измеряются по направлению Oy). Угловым коэффициентом c этой прямой называется *выборочным коэффициентом регрессии*. Связь между выборочным коэффициентом регрессии и выборочным коэффициентом корреляции выражается формулой (14).

Числитель (13) можно вычислять различными способами, контролируемыми друг друга:

$$\begin{aligned} \sum (x - \bar{x})(y - \bar{y}) &= \sum (x - \bar{x})y = \sum x(y - \bar{y}) = \\ &= \sum xy - n\bar{x}\bar{y} = \\ &= \sum (x - a)(y - b) - n(\bar{x} - a)(\bar{y} - b). \end{aligned}$$

Как уже отмечалось ранее, то же самое справедливо и для знаменателя.

Пример 46. Теммис¹ исследовал различные виды цветочной пыльцы и нашел связь между величиной пылинки и количеством пор для выхода пыльцевых трубок. В качестве примера мы рассмотрим результаты исследований пыльцы шаровидной фуксии (*Fuchsia Globosa*). Эта пыльца может иметь от 0 до 4 пор, расположенных в экваториальной плоскости пылинки. Для измерения диаметров пылинок были выделены 5 групп (по 10 пылинок в каждой группе) с количеством пор 0, 1, 2, 3 и 4. Результаты измерений округлялись до числа, кратного 5 микронам. Количества пылинок указаны в корреляционной таблице.

Диаметр	Количество пор				
	$x = 0$	1	2	3	4
$y = 10$	3				
15	7	3			
20		6			
25		1			
30			4		
35			5		
40			1	3	
45				4	
50				3	3
55					4
60					3

¹ Таммес P. M. L., On the origin of number and arrangement of the places of exit on the surface of pollen-grains. Diss., Groningen, 1930.

Очень красивый и редко встречающийся случай линейной регрессии! Находим

$$\begin{aligned}\bar{x} &= 2, & \sum(x - \bar{x})^2 &= 100, \\ \bar{y} &= 33,2, & \sum(y - \bar{y})^2 &= 12\,588, \\ & & \sum(x - \bar{x})(y - \bar{y}) &= 1090.\end{aligned}$$

Выборочный коэффициент регрессии равен

$$c = \frac{1090}{100} = 10,9.$$

Уравнение эмпирической линии регрессии имеет вид

$$y - \bar{y} = c(x - \bar{x})$$

или

$$y = 10,9x + 11,4.$$

Собственно говоря, выборочный коэффициент корреляции имеет смысл вычислять лишь тогда, когда пары (x, y) являются независимыми, т. е. получаются чисто случайно. Поэтому общее количество пылинок для каждого фиксированного x должно быть случайной величиной с частотой приближенно равной вероятности данного значения x . В нашем же случае для всех x выбирается по 10 пылинок. Если, несмотря на это, все-таки вычислить r по формуле (13), то корреляция окажется очень высокой:

$$r = \frac{1090}{\sqrt{100 \cdot 12\,588}} = 0,97.$$

§ 67. Коэффициент корреляции как признак зависимости

Так как r является оценкой для ρ и так как $\rho = 0$, для независимых x и y , то в том случае, когда r значительно отличается от нуля, можно сделать вывод, что $\rho \neq 0$ и поэтому случайные величины x и y зависимы¹.

Для того чтобы знать, при каких r можно уверенно делать указанный вывод, мы должны уметь отвечать на следующий вопрос: насколько может отклоняться от нуля выборочный коэффициент корреляции r , если, в действительности, случайные величины x и y независимы и поэтому $\rho = 0$?

¹ Практически использование коэффициента корреляции в качестве меры зависимости оправдано лишь тогда, когда предполагается, что случайные величины x и y распределены нормально. В общем случае коэффициент ρ как мера зависимости может оказаться неудовлетворительным. Например, если x принимает значения $1/n$, 1 и n с вероятностями $2n/(n-1)^2$, $1 - 4n/(n-1)^2$ и $2n/(n-1)^2$ соответственно и если $y = 1/x$, то $\rho_{xy} = -4n/(n-1)^2$. Таким образом, при $n \rightarrow \infty$ коэффициент корреляции стремится к нулю, хотя x и y связаны функциональной зависимостью. О других мерах зависимости, лишенных недостатков коэффициента корреляции, см. Дунин-Барковский И. В. и Смирнов Н. В., Теория вероятностей и математическая статистика в технике (общая часть), ГИТТЛ, М., 1955, гл. VII. — Прим. перев.

Предположим, что x и y независимы и распределены нормально. Заменой x на $a(x - \hat{x})$ и y на $b(y - \hat{y})$ можно добиться, чтобы обе величины имели нулевое среднее значение и единичную дисперсию. Следовательно, мы можем считать, что совместная плотность распределения случайных величин x и y задается формулой

$$f(x, y) = f(x) f(y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \tag{1}$$

Так как отдельные пары $(x_1, y_1), \dots, (x_n, y_n)$ предполагаются независимыми друг от друга, то совместная плотность вероятности всей системы $(x_1, y_1, \dots, x_n, y_n)$ равна произведению

$$f(x_1, y_1) \dots f(x_n, y_n) = \frac{1}{(2\pi)^n} e^{-\frac{1}{2} \sum x_i^2 - \frac{1}{2} \sum y_i^2} \tag{2}$$

Спрашивается, какова функция распределения r ?

Мы исследуем сейчас несколько более общий вопрос, а именно каково совместное распределение пяти случайных величин: $\bar{x}, \bar{y}, s_x^2, s_y^2$ и r , т. е. какова вероятность того, что все эти величины будут лежать в заданных границах?

Прежде всего с помощью ортогонального преобразования можно легко выделить распределение \bar{x} и \bar{y} . Для этого случайные величины x_1, \dots, x_n ортогонально преобразуем в u_1, \dots, u_n таким образом, чтобы u_1 была пропорциональна выборочному среднему \bar{x} :

$$\left. \begin{aligned} u_1 &= \frac{x_1}{\sqrt{n}} + \frac{x_2}{\sqrt{n}} + \dots + \frac{x_n}{\sqrt{n}} = \bar{x} \sqrt{n}, \\ u_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n, \\ &\dots \dots \dots \end{aligned} \right\} \tag{3}$$

К y_1, \dots, y_n применим ортогональное преобразование с теми же самыми коэффициентами:

$$\left. \begin{aligned} v_1 &= \frac{y_1}{\sqrt{n}} + \frac{y_2}{\sqrt{n}} + \dots + \frac{y_n}{\sqrt{n}} = \bar{y} \sqrt{n}, \\ v_2 &= a_{21}y_1 + a_{22}y_2 + \dots + a_{2n}y_n, \\ &\dots \dots \dots \end{aligned} \right\} \tag{4}$$

Тогда $\sum x_i^2 = \sum u_i^2$ и $\sum y_i^2 = \sum v_i^2$. И так как суммы $x_i + y_i$ подвергаются этому же преобразованию, например

$$u_2 + v_2 = a_{21}(x_1 + y_1) + a_{22}(x_2 + y_2) + \dots + a_{2n}(x_n + y_n),$$

то отсюда следует, что

$$\sum (x_i + y_i)^2 = \sum (u_i + v_i)^2.$$

Если из обеих частей последнего равенства вычесть $\sum x^2 = \sum u^2$, $\sum y^2 = \sum v^2$ и результат разделить на 2, то получим

$$\sum xy = \sum uv. \quad (5)$$

Таким образом,

$$\begin{aligned} r^2 &= \frac{(\sum xy - n\bar{x}\bar{y})^2}{(\sum x^2 - n\bar{x}^2)(\sum y^2 - n\bar{y}^2)} = \frac{(\sum uv - u_1v_1)^2}{(\sum u^2 - u_1^2)(\sum v^2 - v_1^2)} = \\ &= \frac{(u_2v_2 + \dots + u_nv_n)^2}{(u_2^2 + \dots + u_n^2)(v_2^2 + \dots + v_n^2)} \end{aligned} \quad (6)$$

и

$$(n-1)s_x^2 = \sum x^2 - n\bar{x}^2 = \sum u^2 - u_1^2 = u_2^2 + \dots + u_n^2, \quad (7)$$

$$(n-1)s_y^2 = \sum y^2 - n\bar{y}^2 = \sum v^2 - v_1^2 = v_2^2 + \dots + v_n^2. \quad (8)$$

Так как случайные величины u_1 и v_1 не зависят друг от друга и от остальных $u_2, v_2, \dots, u_n, v_n$, то \bar{x} и \bar{y} не зависят друг от друга и от s_x^2, s_y^2 и r .

Разумеется, \bar{x} и \bar{y} распределены нормально с дисперсией $1/n$. Поэтому нам осталось исследовать лишь распределение случайных величин s_x^2, s_y^2 и r , являющихся функциями от $u_2, v_2, \dots, u_n, v_n$. Эти функции задаются формулами (6), (7) и (8). Постараемся вычислить вероятность события

$$s_x^2 < a, \quad s_y^2 < b, \quad r < c. \quad (9)$$

Эта вероятность равна интегралу

$$W = \frac{1}{(2\pi)^{n-1}} \int_G \dots \int e^{-\frac{1}{2}\sum_1^nu^2 - \frac{1}{2}\sum_1^nv^2} du_2 dv_2 \dots du_n dv_n, \quad (10)$$

где область интегрирования G определяется неравенствами (9).

В формуле (10) можно сначала интегрировать по v_2, \dots, v_n , а затем — по u_2, \dots, u_n . При внутреннем интегрировании по v_2, \dots, v_n величины u следует рассматривать как постоянные. Введем ортогональное преобразование переменных v , считая u фиксированными:

$$\left. \begin{aligned} w_2 &= b_{22}v_2 + \dots + b_{2n}v_n, \\ \dots & \\ w_n &= b_{n2}v_2 + \dots + b_{nn}v_n, \end{aligned} \right\} \quad (11)$$

где коэффициенты первой строки определяются формулами

$$b_{2i} = \frac{u_i}{\sqrt{u_2^2 + \dots + u_n^2}} \quad (i = 2, \dots, n). \quad (12)$$

Сумма квадратов коэффициентов b_{2i} равна единице, следовательно, в силу § 13, коэффициенты остальных строк можно определить таким образом, чтобы все преобразование было ортогональным. Тогда внутренний интеграл запишется так:

$$I = \int \dots \int e^{-\frac{1}{2} \sum_1^n v_i^2} dv_2 \dots dv_n = \int \dots \int e^{-\frac{1}{2} \sum_1^n w_i^2} dw_2 \dots dw_n. \quad (13)$$

Согласно (12),

$$w_2 = \sum_2^n b_{2i} v_i = \frac{u_2 v_2 + \dots + u_n v_n}{\sqrt{u_2^2 + \dots + u_n^2}},$$

следовательно,

$$r = \frac{w_2}{\sqrt{v_2^2 + \dots + v_n^2}} = \frac{w_2}{\sqrt{w_2^2 + \dots + w_n^2}}. \quad (14)$$

Если положим

$$u_3^2 + \dots + u_n^2 = \zeta^2, \quad (15)$$

то, в силу (8) и (14), r и s_y^2 окажутся зависящими лишь от w_2 и ζ^2 :

$$(n-1) s_y^2 = u_2^2 + u_3^2 + \dots + u_n^2 = u_2^2 + \zeta^2, \quad (16)$$

$$r = \frac{w_2}{\sqrt{w_2^2 + \zeta^2}}. \quad (17)$$

Для того чтобы вычислить интеграл (13), мы вместо w_3, \dots, w_n введем полярные координаты $\zeta, \varphi_1, \dots, \varphi_{n-3}$. Так как неравенства (9), определяющие область интегрирования, не зависят от угловых переменных, то можно сразу же проинтегрировать по $\varphi_1, \dots, \varphi_{n-3}$ и, положив $w_2 = w$, записать внутренний интеграл (13) в виде

$$I = C \iint e^{-\frac{1}{2} w^2 - \frac{1}{2} \zeta^2} \zeta^{n-3} dw d\zeta, \quad (18)$$

где область интегрирования определяется неравенствами

$$\left. \begin{aligned} \zeta &\geq 0, \\ w^2 + \zeta^2 &= (n-1) s_y^2 < (n-1) b, \\ \frac{w}{\sqrt{w^2 + \zeta^2}} &= r < c. \end{aligned} \right\} \quad (19)$$

Интеграл I не зависит от u_i , поэтому в формуле (10) множитель I можно вынести за знак интеграла. Тогда получим

$$W = \frac{1}{(2\pi)^{n-1}} I \int \dots \int e^{-\frac{1}{2} \sum_1^n u_i^2} du_2 \dots du_n, \quad (20)$$

где область интегрирования задается неравенством

$$\sum_2^n u_i^2 = (n-1) s_x^2 < (n-1) a. \quad (21)$$

Если в этом интеграле также ввести полярные координаты χ , $\varphi'_1, \dots, \varphi'_{n-2}$, то снова можно будет проинтегрировать по угловым переменным. В результате получим

$$\begin{aligned} W &= C' I \int e^{-\frac{1}{2}x^2} \chi^{n-2} d\chi \dots \\ &= \alpha \iiint e^{-\frac{1}{2}x^2} \chi^{n-2} d\chi e^{-\frac{1}{2}\zeta^2} \zeta^{n-3} d\zeta e^{-\frac{1}{2}w^2} dw, \end{aligned} \quad (22)$$

где область интегрирования задается неравенствами

$$\left. \begin{aligned} \chi^2 &< (n-1) a, & \chi &\geq 0, \\ w^2 + \zeta^2 &< (n-1) b, & \zeta &\geq 0, \\ \frac{w}{\sqrt{w^2 + \zeta^2}} &< c. \end{aligned} \right\} \quad (23)$$

Само собой разумеется, что этот результат справедлив не только для области специального вида (9), но также и для произвольной области G в пространстве переменных s_x^2, s_y^2, r . Пусть G' — преобразованная область, расположенная в пространстве новых переменных χ, ζ, w , которые определяются равенствами

$$\chi^2 = (n-1) \frac{s_x^2}{\sigma_x^2}, \quad \chi \geq 0, \quad (24)$$

$$w^2 + \zeta^2 = (n-1) \frac{s_y^2}{\sigma_y^2}, \quad z \geq 0, \quad (25)$$

$$w = \frac{r}{\sqrt{w^2 + \zeta^2}} = \sqrt{n-1} \frac{rs_y}{\sigma_y}. \quad (26)$$

Знаменатели σ_x^2, σ_y^2 и σ_y добавлены для того, чтобы формулы были справедливы и тогда, когда σ_x и σ_y не равны единице. В этом случае области G соответствует вероятность

$$P G = W = \alpha \iiint e^{-\frac{1}{2}x^2} \chi^{n-2} d\chi \cdot e^{-\frac{1}{2}\zeta^2} \zeta^{n-3} d\zeta \cdot e^{-\frac{1}{2}w^2} dw, \quad (27)$$

где интегрирование распространяется на преобразованную область G' . Постоянную α , конечно, следует определить таким образом, чтобы интеграл по всему пространству $\chi \geq 0, \zeta \geq 0$ равнялся единице.

Результат можно сформулировать так:

$\chi^2 = u$, $\zeta^2 = v$ и w являются независимыми случайными величинами с плотностями вероятности

$$f(u) = \alpha_1 e^{-\frac{1}{2}u} u^{\frac{n-3}{2}}, \quad \alpha_1 = \frac{1}{\Gamma\left(\frac{n-1}{2}\right) 2^{\frac{n-1}{2}}}; \quad (28)$$

$$g(v) = \alpha_2 e^{-\frac{1}{2}v} v^{\frac{n-4}{2}}, \quad \alpha_2 = \frac{1}{\Gamma\left(\frac{n-2}{2}\right) 2^{\frac{n-2}{2}}}; \quad (29)$$

$$h(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2}. \quad (30)$$

Поэтому χ^2 и ζ^2 подчиняются распределению χ^2 с $n - 1$ и $n - 2$ степенями свободы соответственно. Случайная величина w распределена нормально с нулевым средним значением и единичной дисперсией.

Теперь легко можно вывести распределение выборочного коэффициента корреляции r . Сначала равенство (26) разрешим относительно ζ :

$$\zeta = \frac{w}{r} \sqrt{1 - r^2},$$

и затем построим отношение

$$t = \frac{w}{\zeta} \sqrt{n - 2} = \frac{r}{\sqrt{1 - r^2}} \sqrt{n - 2}. \quad (31)$$

Так как w имеет нормальное распределение и ζ^2 подчиняется распределению χ^2 с $n - 2$ степенями свободы, то, согласно § 28, случайная величина t подчиняется распределению Стьюдента с $n - 2$ степенями свободы. Зная границы для t , соответствующие обычным уровням значимости (5, 2 и 1%), и пользуясь формулой (31), немедленно получаем границы для выборочного коэффициента корреляции r . Эти границы табулированы в табл. 13, в конце книги.

Таблица 13 применяется следующим образом: если в результате практических вычислений получается такое значение r , абсолютная величина которого превосходит границу r_p из табл. 13, то случайные величины x и y считают зависимыми.

Относительно уровня значимости этого критерия можно сказать следующее.

Если x и y зависимы и если, согласно этому критерию, независимость отвергается, то в данном случае никакой ошибки не возникает¹.

Может быть и другой случай, когда x и y независимы и распределены приближенно нормально. В этом случае истинный уровень значимости критерия приближенно равен 2β , так как именно с таким уровнем строился точный критерий для проверки независимости двух нормальных случайных величин.

Имеется и третий случай, когда x и y независимы, но их распределения существенно отличаются от нормального. В этом случае истинный уровень значимости критерия может оказаться несколько больше, чем 2β . Однако если x и y обладают конечными дисперсиями и n достаточно велико, то отклонение истинного уровня значимости от 2β не будет значительным. Это происходит потому, что случайные колебания r определяются главным образом случайными колебаниями числителя

$$\sum (x - \bar{x})(y - \bar{y}). \quad (32)$$

Если n велико, то \bar{x} и \bar{y} можно приближенно заменить истинными средними значениями \hat{x} и \hat{y} . Тогда вместо (32) получим

$$\sum (x - \hat{x})(y - \hat{y}).$$

Это выражение представляет собой сумму большого количества независимых слагаемых, квадратичное отклонение каждого из которых равно $\sigma_x \sigma_y$. Следовательно, согласно центральной предельной теореме (§ 24 Г), эта сумма распределена приближенно нормально с нулевым средним значением и дисперсией $n\sigma_x^2\sigma_y^2$. Знаменателем r является $(n-1)s_x s_y$, следовательно, этот знаменатель приближенно равен $n\sigma_x \sigma_y$. Поэтому при $n \rightarrow \infty$ случайная величина r распределена асимптотически нормально с нулевым средним значением и дисперсией $1/n$, безотносительно к тому, распределены x и y нормально или нет. Таким образом, при $n \rightarrow \infty$ истинный уровень значимости стремится к 2β . Для конечных, не слишком малых n отклонение от 2β не очень значительно.

Пример 47 (из книги: Fisher R. A., *Statistical Methods for Research Workers*, 11th ed., Еж. 27). В восточной Англии в течение 20 лет (с 1885 по 1904 г.)

¹ Строго говоря, этот критерий предназначен для проверки независимости случайных величин x и y , распределенных нормально. В остальных случаях этот критерий употребляется для проверки гипотезы $\rho = 0$. При таком истолковании мощность критерия оказывается вполне удовлетворительной. Если же этот критерий применяется для проверки независимости случайных величин с распределениями, отличными от нормального, то, как показывает пример в предыдущей сноске, мощность критерия может оказаться очень малой. Автор не затрагивает вопроса оценки мощности этого критерия, а рассматривает лишь вероятность ошибки первого рода. — *Прим. перев.*

велись наблюдения за урожаем пшеницы и количеством осадков осенью. В результате оказалось, что урожай пшеницы и осенние осадки связаны отрицательной корреляцией, равной $-0,63$. По таблице находим, что $0,56$ является 1%-ной границей для r . Следовательно, взаимосвязь между осадками и урожаем пшеницы можно считать доказанной.

§ 68. Частные коэффициенты корреляции

А. ПОНЯТИЕ ЧАСТНОЙ КОРРЕЛЯЦИИ¹

В некоторых случаях корреляция между двумя величинами x и y целиком или частично вызывается тем, что существует значительная корреляция между x и y , с одной стороны, и третьей случайной величиной z — с другой. Можно попытаться исключить зависимость от z , заменив x и y такими величинами

$$x' = x - \lambda z \quad \text{и} \quad y' = y - \mu z,$$

которые некоррелированы с z (слово «некоррелированы» означает, что коэффициент корреляции равен нулю). Возникает вопрос, какова будет при этом оставшаяся корреляция между x и y ?

Сначала мы займемся изучением этого вопроса для истинного коэффициента корреляции ρ_{xy} . Если $\hat{x} = \hat{y} = \hat{z} = 0$, то

$$\rho_{xy} = \frac{\xi_{xy}}{\sigma_x \sigma_y}. \quad (1)$$

Множители λ и μ нужно выбрать так, чтобы имели место равенства

$$\begin{aligned} \xi_{x'z} &= \xi(xz - \lambda z^2) = 0, \\ \xi_{y'z} &= \xi(yz - \mu z^2) = 0. \end{aligned}$$

Отсюда находим

$$\lambda = \frac{\xi_{xz}}{\xi_{z^2}} = \frac{\rho_{xz} \sigma_x \sigma_z}{\sigma_z^2} = \rho_{xz} \frac{\sigma_x}{\sigma_z}, \quad (2)$$

$$\mu = \frac{\xi_{yz}}{\xi_{z^2}} = \frac{\rho_{yz} \sigma_y \sigma_z}{\sigma_z^2} = \rho_{yz} \frac{\sigma_y}{\sigma_z}. \quad (3)$$

«Частный коэффициент корреляции» $\rho_{xy \cdot z}$ определяется формулой

$$\rho_{xy \cdot z} = \rho_{x'y'} = \frac{\xi_{(x-\lambda z)(y-\mu z)}}{\sigma_{x-\lambda z} \sigma_{y-\mu z}}. \quad (4)$$

¹ Автор пользуется здесь терминами «очищенные» коэффициенты корреляции и «очищение» (*bereinigte Korrelationskoeffizienten, Bereinigung*), которые хорошо соответствуют логическому замыслу теории частной корреляции. — *Прим. ред.*

Числитель (4) можно записать так:

$$\begin{aligned} \mathfrak{E} xy - \lambda \mathfrak{E} zy - \mu \mathfrak{E} xz + \lambda\mu \mathfrak{E} zz &= \\ &= \varrho_{xy} \sigma_x \sigma_y - \lambda \varrho_{yz} \sigma_y \sigma_z - \mu \varrho_{xz} \sigma_x \sigma_z + \lambda\mu \sigma_z^2. \end{aligned}$$

Если в этом равенстве λ и μ заменить их значениями (2) и (3), то получим

$$\mathfrak{E}(x - \lambda z)(y - \mu z) = (\varrho_{xy} - \varrho_{xz}\varrho_{yz})\sigma_x\sigma_y. \quad (5)$$

Точно так же вычисляются

$$\sigma_{x-\lambda z}^2 = \mathfrak{E}(x - \lambda z)^2 = (1 - \varrho_{xz}^2) \sigma_x^2, \quad (6)$$

$$\sigma_{y-\mu z}^2 = \mathfrak{E}(y - \mu z)^2 = (1 - \varrho_{yz}^2) \sigma_y^2. \quad (7)$$

Если все эти выражения подставить в (4), то найдем

$$\varrho_{xy|z} = \frac{\varrho_{xy} - \varrho_{xz}\varrho_{yz}}{\sqrt{1 - \varrho_{xz}^2} \sqrt{1 - \varrho_{yz}^2}}. \quad (8)$$

Для того чтобы получить оценку для $\varrho_{xy|z}$, заменим в формуле (8) истинные коэффициенты корреляции ϱ выборочными r . В результате получим

$$r_{xy|z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{1 - r_{xz}^2} \sqrt{1 - r_{yz}^2}}. \quad (9)$$

Формулу (9) можно было бы вывести так же, как была выведена формула (8). Для этого нужно бы было определить, какие линейные комбинации $x'' = x - az$ и $y'' = y - bz$ имеют с z выборочный коэффициент корреляции, равный нулю. В этом случае выборочный коэффициент корреляции между x'' и y'' задавался бы формулой, в точности совпадающей с (9).

Отсюда следует, что $r_{xy|z}$ не изменится, если x заменить на $x - \lambda z$, а y — на $y - \mu z$, где λ и μ — произвольные числа. Действительно, выборочные коэффициенты корреляции между z и линейными комбинациями $x'' = x - az$ и $y'' = y - bz$ равны нулю, а поэтому при замене x на $x - \lambda z$ и y на $y - \mu z$ линейные комбинации x'' и y'' остаются неизменными. Таким образом, при изучении свойств случайной величины $r_{xy|z}$ мы всегда можем заменить x и y величинами x' и y' , некоррелированными с z , т. е. мы можем заранее предположить, что $\varrho_{xz} = \varrho_{yz} = 0$.

Б. ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ $r_{xy|z}$

Возникает тот же вопрос, что и в предыдущем параграфе: сколь велик должен быть частный коэффициент корреляции $r_{xy|z}$, чтобы можно было утверждать, что между $x - \lambda z$ и $y - \mu z$ действительно существует зависимость? Или, иными словами,

какие значения $r_{xy|z}$ следует считать чисто случайными, если $x - \lambda z$ и $y - \mu z$, в действительности, независимы?

Согласно сделанному выше замечанию, мы можем при этом предположить, что $\lambda = \mu = 0$, следовательно, $\rho_{xz} = \rho_{yz} = 0$. Мы пойдем еще дальше и предположим, что x, y, z — независимые, нормально распределенные случайные величины. Если их дисперсии равны единице, то совместная плотность вероятности будет задаваться формулой

$$f(x, y, z) = (2\pi)^{-\frac{3}{2}} e^{-\frac{1}{2}(x^2 + y^2 + z^2)} \quad (10)$$

При этих предположениях мы должны найти функцию распределения $H(c)$ случайной величины $r_{xy|z}$. Эта функция определяется кратным интегралом

$$H(c) = P(r_{xy|z} < c) = \iiint \dots \iiint f(x_1, y_1, z_1) \dots f(x_n, y_n, z_n) dx_1 dy_1 \dots dz_n, \quad (11)$$

где область интегрирования задается неравенством $r_{xy|z} < c$.

Как и в § 66, с помощью ортогонального преобразования можно выделить распределение выборочных средних \bar{x} , \bar{y} и \bar{z} . Для этого мы вместо x_i, y_j и z_k введем u_i, v_j и w_k таким образом, чтобы u_1, v_1 и w_1 были пропорциональны \bar{x}, \bar{y} и \bar{z} соответственно. Коэффициенты ортогонального преобразования для v_j и w_k выберем те же самые, что и для u_i . Если положим, для краткости,

$$[uv] = u_2 v_2 + \dots + u_n v_n, \quad (12)$$

а $[uw]$ и $[vw]$ определим аналогично, то выборочные коэффициенты корреляции будут задаваться формулами

$$r_{xy} = \frac{[uv]}{\sqrt{[uu][vv]}}, \quad r_{xz} = \frac{[uw]}{\sqrt{[uu][ww]}}; \quad r_{yz} = \frac{[vw]}{\sqrt{[vv][ww]}}. \quad (13)$$

Преобразованный интеграл выглядит точно так же, как первоначальный интеграл (11), но только с заменой x, y, z на u, v, w . Так как определение области интегрирования $r_{xy|z} < c$ не зависит от u_1, v_1, w_1 , то можно произвести интегрирование по этим переменным и получить

$$H(c) = (2\pi)^{-\frac{3n-3}{2}} \iiint \dots \iiint e^{-\frac{1}{2}([uu] + [vv] + [ww])} du_2 dv_2 \dots dw_n, \quad (14)$$

где область интегрирования задается неравенством $r_{xy|z} < c$.

В формуле (14) мы сначала произведем интегрирование по $u_2, \dots, u_n, v_2, \dots, v_n$, а затем по w_2, \dots, w_n . В качестве внутреннего интеграла получим

$$I = \iint \dots \iint e^{-\frac{1}{2}[uu] - \frac{1}{2}[vv]} du_2 \dots du_n dv_2 \dots dv_n. \quad (15)$$

Как и в § 67, введем теперь ортогональное преобразование переменных u и v , считая w постоянным:

$$u'_i = \sum a_{ik} u_k, \quad v'_i = \sum a_{ik} v_k. \quad (16)$$

При этом коэффициенты выбираются таким образом, чтобы, в частности, имели место равенства

$$u'_2 = \frac{[uu]}{\sqrt{[ww]}}, \quad v'_2 = \frac{[vw]}{\sqrt{[ww]}} \quad (17)$$

и, следовательно,

$$r_{xz} = \frac{u'_2}{\sqrt{[uu]}} = \frac{u'_2}{\sqrt{[u'u']}}, \quad r_{yz} = \frac{v'_2}{\sqrt{[vv]}} = \frac{v'_2}{\sqrt{[v'v']}},$$

$$r_{xy} = \frac{[uv]}{\sqrt{[uu][vv]}} = \frac{[u'v']}{\sqrt{[u'u'][v'v']}}.$$

Тогда (15) перейдет в интеграл

$$I = \iint \dots \iint e^{-\frac{1}{2}[u'u'] - \frac{1}{2}[v'v']} du'_2 \dots du'_n dv'_2 \dots dv'_n, \quad (18)$$

где w — постоянная величина, и область интегрирования, как и прежде, задается неравенством $r_{xy} |z| < c$. При этом

$$r_{xy} |z| = \frac{[u'v'] - u'_2 v'_2}{\sqrt{[u'u'] - u'^2_2} \cdot \sqrt{[v'v'] - v'^2_2}} =$$

$$= \frac{u'_3 v'_3 + \dots + u'_n v'_n}{\sqrt{u'^2_3 + \dots + u'^2_n} \cdot \sqrt{v'^2_3 + \dots + v'^2_n}}. \quad (19)$$

Так как определение области интегрирования не зависит от u'_2 и v'_2 , то, интегрируя по этим переменным, получим

$$I = 2\pi \int \dots \int_{r_{xy} |z| < c} e^{-\frac{1}{2}(u'^2_3 + \dots + u'^2_n + v'^2_3 + \dots + v'^2_n)} du'_3 \dots du'_n dv'_3 \dots dv'_n. \quad (20)$$

Формула (20) показывает, что интеграл I не зависит от w_1, \dots, w_n , следовательно, множитель I можно вынести за знак

интеграла (14) и затем проинтегрировать по w_2, \dots, w_n . Таким образом, заменив w_3, \dots на u_3, \dots , получаем

$$H(c) = (2\pi)^{-n+2} \int \dots \int e^{-\frac{1}{2}(u_3^2 + \dots + u_n^2 + v_3^2 + \dots + v_n^2)} du_3 \dots dv_n, \quad (21)$$

где область интегрирования определяется неравенством $r < c$, а r , в силу (19), задается формулой

$$r = \frac{u_3 v_3 + \dots + u_n v_n}{\sqrt{u_3^2 + \dots + u_n^2} \cdot \sqrt{v_3^2 + \dots + v_n^2}}. \quad (22)$$

Сравним теперь интеграл (21) с интегралом (10), вычисленным в § 67. Ранее имелось $2(n-1)$ переменных интегрирования u_2, \dots, u_n и v_2, \dots, v_n , теперь же таких переменных имеется $2(n-2)$: u_3, \dots, u_n и v_3, \dots, v_n . Область интегрирования для интеграла (10) § 67 определялась неравенствами

$$s_x^2 < a, \quad s_y^2 < b, \quad r < c.$$

Но если a и b устремить к бесконечности, то область интегрирования будет задаваться одним неравенством $r < c$. Таким образом, интеграл (21) является частным случаем интеграла (10) § 67, поэтому справедлива теорема:

Функция распределения выборочного частного коэффициента корреляции r_{xy} равна функции распределения обычного выборочного коэффициента корреляции r , вычисленного по двум рядам независимых, нормально распределенных случайных величин x_1, \dots, x_{n-1} и y_1, \dots, y_{n-1} , с той лишь разницей, что количество переменных в каждом ряду равно не n , а $n-1$. Еще раз напомним, что эта теорема справедлива лишь в том случае, когда справедлива гипотеза, согласно которой $x - \lambda z$, $y - \mu z$ и z являются независимыми, нормально распределенными случайными величинами. При этом наиболее важным является требование независимости; что же касается требования нормальной распределенности, то оно менее существенно,

Согласно этой теореме, для проверки зависимости случайных величин $x - \lambda z$ и $y - \mu z$ можно воспользоваться табл. 13, считая число наблюдений равным $n-1$ вместо n .

В. ГЕОМЕТРИЧЕСКОЕ ИСТОЛКОВАНИЕ

Результаты, полученные в этом и в предыдущих параграфах, можно вывести геометрически.

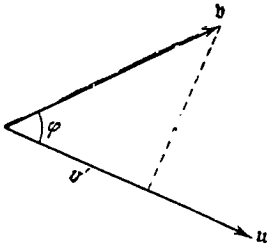
Пусть, например, $n=4$ и пусть (u_2, u_3, u_4) , (v_2, v_3, v_4) и (w_2, w_3, w_4) — компоненты трех векторов u , v и w , расположенных в трехмерном пространстве. В этом случае $[uv]$ представляет собой квадрат длины вектора u , $[uv]$ — скалярное произведение

\mathbf{u} и \mathbf{v} , и r_{xy} — косинус угла φ между \mathbf{u} и \mathbf{v} . Плотность совместного распределения случайных величин u_i и v_i

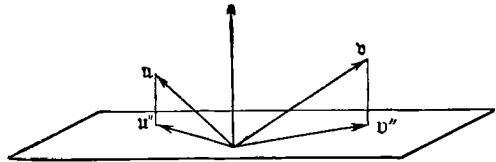
$$(2\pi)^{-n+1} e^{-\frac{1}{2}[\mathbf{uu}] - \frac{1}{2}[\mathbf{vv}]}$$

показывает, что все шесть векторных компонент u_i и v_j взаимно независимы и распределены нормально. Этот же вывод останется справедливым и в том случае, если к \mathbf{u}_i и \mathbf{v}_j добавить еще w_k .

Указанный закон распределения инвариантен относительно ортогональных преобразований. Таким образом, если одну из новых осей координат направить вдоль вектора \mathbf{u} , а две другие



Р и с. 35.



Р и с. 36.

ортогональные оси расположить в плоскости, перпендикулярной \mathbf{u} , то в новых координатах одна компонента v' вектора \mathbf{v} будет параллельна вектору \mathbf{u} , а две другие компоненты будут перпендикулярны \mathbf{u} . При этом все компоненты независимы и подчиняются нормальному распределению. Компонента v' , параллельная \mathbf{u} , распределена нормально с единичной дисперсией, а сумма квадратов компонент, перпендикулярных \mathbf{u} , подчиняется распределению χ^2 с двумя степенями свободы. Обозначим эту сумму квадратов v''^2 . Отношение $v' / |v''|$ равно котангенсу угла φ (рис. 35), следовательно,

$$\frac{v'}{|v''|} = \frac{\cos \varphi}{\sin \varphi} = \frac{r}{\sqrt{1-r^2}}.$$

Этот результат объясняет, почему случайная величина

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}$$

подчиняется распределению Стьюдента с $n-2$ степенями свободы.

Если к векторам \mathbf{u} и \mathbf{v} добавить еще вектор \mathbf{w} (рис. 36), то для вычисления выборочного частного коэффициента корреляции $r_{xy|z}$ нужно будет сначала \mathbf{u} и \mathbf{v} заменить векторами $\mathbf{u}'' = \mathbf{u} - a\mathbf{w}$ и $\mathbf{v}'' = \mathbf{v} - b\mathbf{w}$, перпендикулярными \mathbf{w} . Выберем новую систему

координат таким образом, чтобы одна из новых осей была направлена вдоль вектора \mathbf{w} , а две другие оси были перпендикулярны \mathbf{w} . Компоненты векторов \mathbf{u} и \mathbf{v} , перпендикулярные \mathbf{w} , определяются векторами \mathbf{u}'' и \mathbf{v}'' ; косинус угла между \mathbf{u}'' и \mathbf{v}'' равен выборочному частному коэффициенту корреляции $r_{xy|z}$. Так как эти векторы расположены в плоскости, перпендикулярной \mathbf{w} , то их размерность равна не $n - 1$, а $n - 2$; кроме того, \mathbf{u}'' и \mathbf{v}'' подчиняются нормальному распределению. Отсюда ясно, что $r_{xy|z}$ имеет ту же самую функцию распределения, что и обычный выборочный коэффициент корреляции, но только с заменой n на $n - 1$.

§ 69. Распределение выборочного коэффициента корреляции зависимых случайных величин

А. ДВУМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

До сих пор мы всегда предполагали, что, в действительности, x и y являются независимыми случайными величинами и что r лишь случайно отличается от нуля. Для зависимых x и y все становится значительно сложнее. Для того чтобы о распределении r можно было сказать что-то определенное, нужно сделать некоторые предположения о распределении x и y .

Наиболее простым является предположение, согласно которому величина x распределена нормально, а y равна $\lambda x + z$, где z также распределена нормально и не зависит от x . Если дисперсия случайной величины x равна $1/g$, а дисперсия z равна $1/h$, то плотность совместного распределения пары (x, z) задается формулой

$$\frac{\sqrt{gh}}{2\pi} e^{-\frac{1}{2}(gx^2 + hz^2)}, \tag{1}$$

следовательно, плотность совместного распределения пары (x, y) равна

$$\begin{aligned} f(x, y) &= \frac{\sqrt{gh}}{2\pi} e^{-\frac{1}{2}[gx^2 + h(y - \lambda x)^2]} = \\ &= \frac{\sqrt{gh}}{2\pi} e^{-\frac{1}{2}[(g + h\lambda^2)x^2 - 2\lambda hxy + hy^2]} \end{aligned} \tag{2}$$

Эту же плотность можно записать так:

$$f(x, y) = C e^{-\frac{1}{2}(ax^2 - 2bxy + by^2)} \tag{3}$$

Средние значения x и y выбраны равными нулю только для удобства. Если желательно рассматривать общий случай, то нужно x и y заменить разностями $x - \hat{x}$ и $y - \hat{y}$.

Функцию вида (3) с отрицательной квадратичной формой в показателе степени называют *плотностью двумерного нормального распределения*. Такие распределения очень часто приближенно осуществляются в биологии, а именно, тогда, когда на основе неселектированного материала изучается наследственность определенных размеров тела.

Каждую плотность вида (3) можно представить выражением (2), т. е. y можно представить как сумму двух случайных величин, из которых первая (λx) пропорциональна x , а вторая (z) от x не зависит. Так как x и y совершенно равноправны, то можно, конечно, наоборот, представить x в виде суммы линейного члена μy и случайной величины z' , не зависящей от y . Выбор одного из этих представлений зависит от той конкретной задачи, которую нужно решить.

Изменением масштабов на осях Ox и Oy всегда можно добиться, чтобы выполнялись равенства $a = h = 1$. Тогда $\lambda = b$ и $g = 1 - b^2$. В этом случае плотность распределения (3) задается простой формулой

$$f(x, y) = \frac{1}{2\pi} \sqrt{1 - b^2} e^{-\frac{1}{2}(x^2 - 2bxy + y^2)}, \quad (4)$$

которая в дальнейшем для нас будет являться основной.

Дисперсия x была равна

$$\sigma_x^2 = \frac{1}{g} = \frac{1}{1 - b^2}. \quad (5)$$

Дисперсия y , конечно, имеет то же самое значение

$$\sigma_y^2 = \frac{1}{1 - b^2}. \quad (6)$$

Ковариация x и y вычисляется так:

$$\begin{aligned} \mathcal{E} xy &= \mathcal{E} x(\lambda x + z) = \lambda \mathcal{E} x^2 + \mathcal{E} xz = \\ &= b \mathcal{E} x^2 + 0 = \frac{b}{1 - b^2}. \end{aligned} \quad (7)$$

Поэтому истинный коэффициент корреляции равен

$$\rho = \frac{\mathcal{E} xy}{\sigma_x \sigma_y} = b. \quad (8)$$

Таким образом, положив в (4) $b = \rho$, мы могли бы написать

$$f(x, y) = \frac{1}{2\pi} \sqrt{1 - \rho^2} e^{-\frac{1}{2}(x^2 - 2\rho xy + y^2)} \quad (9)$$

Б. АСИМПТОТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ r ПРИ БОЛЬШИХ n

Пусть $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ — независимые двумерные случайные величины, каждая из которых подчиняется распределению с плотностью (9). Тогда совместная плотность распределения всей системы $(x_1, y_1, \dots, x_n, y_n)$ задается формулой

$$f(x_1, y_1) \dots f(x_n, y_n) = \frac{1}{(2\pi)^n} (1 - \rho^2)^{\frac{n}{2}} e^{-\frac{1}{2} \sum (x_k^2 - 2\rho x_k y_k + y_k^2)} \quad (10)$$

При этих предположениях и при $n \rightarrow \infty$ выборочный коэффициент корреляции r распределен асимптотически нормально. Среднее значение r асимптотически равно ρ , а его квадратичное отклонение асимптотически равно

$$\sigma = \frac{1 - \rho^2}{\sqrt{n-1}}. \quad (11)$$

Доказательство будет очень простым. Сначала мы так же, как в § 67, с помощью ортогонального преобразования заменим x_k и y_k новыми переменными u_k и v_k таким образом, чтобы имели место равенства $u_1 = \bar{x} \sqrt{n}$ и $v_1 = \bar{y} \sqrt{n}$. Тогда r будет задаваться формулой

$$r = \frac{u_2 v_2 + \dots + u_n v_n}{\sqrt{u_2^2 + \dots + u_n^2} \cdot \sqrt{v_2^2 + \dots + v_n^2}}. \quad (12)$$

Плотность распределения u_k и v_k имеет тот же вид, что и плотность (10):

$$f(u_1, v_1) \dots f(u_n, v_n) = \frac{1}{(2\pi)^n} (1 - \rho^2)^{\frac{n}{2}} e^{-\frac{1}{2} \sum (u_k^2 - 2\rho u_k v_k + v_k^2)}. \quad (13)$$

Средние значения u_k^2 , v_k^2 и $u_k \cdot v_k$ равны

$$\mathcal{E} u_k^2 = \sigma_x^2 = \frac{1}{g},$$

$$\mathcal{E} v_k^2 = \sigma_y^2 = \frac{1}{g},$$

$$\mathcal{E} u_k v_k = \rho \sigma_x \sigma_y = \frac{\rho}{g}.$$

Следовательно, мы можем положить

$$\left. \begin{aligned} u_k^2 &= \frac{1 + p_k}{g}, \\ v_k^2 &= \frac{1 + q_k}{g}, \\ u_k v_k &= \frac{\rho + r_k}{g}. \end{aligned} \right\} \quad (14)$$

Если (14) подставить в (12), то получим

$$r = \frac{m\rho + \sum r_k}{\sqrt{m + \sum p_k} \cdot \sqrt{m + \sum q_k}}, \quad (15)$$

где $m = n - 1$ и $\sum p_k$, $\sum q_k$, $\sum r_k$ представляют собой суммы m случайных величин, каждая из которых имеет среднее значение, равное нулю, и квадратичное отклонение порядка единицы. Таким образом, с вероятностью, сколь угодно близкой к единице, эти суммы при $m \rightarrow \infty$ являются величинами порядка \sqrt{m} , т. е. все они меньше m . Поэтому (15) можно разложить в ряд по степеням $\sum p_k/m$, $\sum q_k/m$ и $\sum r_k/m$:

$$\begin{aligned} r &= \frac{\rho + \frac{\sum r_k}{m}}{\sqrt{1 + \frac{\sum p_k}{m}} \cdot \sqrt{1 + \frac{\sum q_k}{m}}} = \\ &= \rho + \frac{\sum r_k}{m} - \frac{1}{2} \rho \frac{\sum p_k}{m} - \frac{1}{2} \rho \frac{\sum q_k}{m} + \dots \end{aligned}$$

В результате получаем асимптотическую формулу

$$r - \rho \sim \sum \frac{r_k - \frac{1}{2} \rho p_k - \frac{1}{2} \rho q_k}{n-1}. \quad (16)$$

Правая часть представляет собой сумму большого числа независимых случайных величин, дисперсии которых ограничены. Согласно центральной предельной теореме (§ 24 Г), эта сумма распределена асимптотически нормально. Среднее значение правой части (16) равно нулю, а дисперсия равна сумме дисперсий отдельных слагаемых. Вычислив эти дисперсии, найдем, что квадратичное отклонение выражается формулой (11).

При практическом применении этих результатов возникают различные трудности. Так как значение ρ заранее не известно, то нельзя вычислить σ . Часто в правой части (11) ρ заменяют на r и, таким образом, получают следующую оценку для σ :

$$s = \frac{1 - r^2}{\sqrt{n-1}}. \quad (17)$$

Однако эта оценка является надежной лишь тогда, когда n очень велико и ρ^2 не близко к единице. Если значение n является лишь умеренно большим и если ρ^2 близко к единице (признаком этого является близость к единице значения r^2), то s может значительно отличаться от σ . Кроме того, математическое ожидание r отклоняется от ρ , т. е. оценка r имеет смещение. В этом можно

очень легко убедиться, вычислив несколько следующих членов приближения (16). В результате найдем, что

$$\mathcal{E} r = \rho - \frac{\rho}{2} \frac{1-\rho^2}{n-1} + \dots \quad (18)$$

Ко всему тому нужно еще добавить, что для умеренных значений n точное распределение r может значительно отклоняться от нормального распределения, особенно тогда, когда значение ρ^2 близко к единице. Таким образом, при вычислении доверительных границ для ρ по заданному r нельзя непосредственно применять нормальное распределение, а нужно воспользоваться точной функцией распределения r или по меньшей мере улучшенным приближением для этой функции.

в. ТОЧНОЕ РАСПРЕДЕЛЕНИЕ s_x^2, s_y^2 И r

Плотность совместного распределения трех случайных величин s_x^2, s_y^2 и r можно вычислить следующим образом. Введем сначала для оценок дисперсий и ковариации обозначения:

$$s_{xx} = s_x^2, \quad s_{yy} = s_y^2, \quad s_{xy} = r s_x s_y. \quad (19)$$

Пусть снова $y = \lambda x + z = \rho x + z$, где z — случайная величина, не зависящая от x . Построим оценки дисперсий и ковариации, аналогичные (19):

$$s_{xx} = s_x^2, \quad s_{xz} = r' s_x s_z, \quad s_{zz} = s_z^2, \quad (20)$$

где r' — выборочный коэффициент корреляции между x и z .

Величины (19) можно легко выразить через величины (20):

$$\begin{aligned} s_{xx} &= s_{xx}, \\ s_{xy} &= \rho s_{xx} + s_{xz}, \\ s_{yy} &= \rho^2 s_{xx} + 2\rho s_{xz} + s_{zz}. \end{aligned} \quad (21)$$

Вместо трех величин (20) можно ввести независимые случайные величины χ, w, ζ , указанные в § 67. Согласно (22) § 67, плотность совместного распределения χ, w и ζ задается формулой

$$f_0(\chi, w, \zeta) = \alpha e^{-\frac{1}{2}\chi^2 - \frac{1}{2}w^2 - \frac{1}{2}\zeta^2} \chi^{n-2} \zeta^{n-3}, \quad (22)$$

где α — постоянный нормирующий множитель.

Случайные величины (20) выражаются через χ, w, ζ следующим образом:

$$\left. \begin{aligned} (n-1) s_{xx} &= \sigma_x^2 \chi^2, \\ (n-1) s_{xz} &= \sigma_x \sigma_z \chi, \\ (n-1) s_{zz} &= \sigma_z^2 (w^2 + \zeta^2). \end{aligned} \right\} \quad (23)$$

При этом, согласно (5), $\sigma_x^2 = 1/g = 1/(1 - \rho^2)$ и $\sigma_y^2 = 1$.

С помощью плотности (22) и формул (23) можно сначала найти плотность распределения случайных величин s_{xx} , s_{xz} и s_{zz} , затем с помощью формул (21) вычислить плотность для s_{xx} , s_{xy} и s_{yy} и, наконец, воспользовавшись формулами (19), найти плотность совместного распределения s_x , s_y и r . В результате окажется, что плотность совместного распределения трех случайных величин s_x , s_y и r имеет вид

$$f(s_x, s_y, r) = \quad (24)$$

$$= \frac{1}{\pi \Gamma(n-2)} (1 - \rho^2)^{\frac{n-1}{2}} e^{-\frac{n-1}{1-\rho^2} (s_x^2 - 2\rho r s_x s_y + s_y^2)} s_x^{n-2} s_y^{n-2} (1 - r^2)^{\frac{n-4}{2}}.$$

Этот результат принадлежит Фишеру [Fisher R. A., *Biometrika*, 10 (1915), 507].

Г. ВСПОМОГАТЕЛЬНОЕ ПРЕОБРАЗОВАНИЕ ФИШЕРА

Интегрированием (24) по s_x и s_y получаем плотность вероятности для r :

$$f(r) = \frac{1}{\pi \Gamma(n-2)} (1 - \rho^2)^{\frac{n-1}{2}} (1 - r^2)^{\frac{n-4}{2}} \frac{d^{n-2}}{d(\rho r)^{n-2}} \frac{\arcsin \rho r}{\sqrt{1 - \rho^2 r^2}}. \quad (25)$$

О вычислении этой плотности см. Kendall M. G., *Advanced Theory of Statistics*, I, 14.14, или только что цитированную работу Фишера, опубликованную в журнале *Biometrika*, 10.

Фишер указал очень практичное преобразование случайной величины r , с помощью которого распределение с плотностью (25) приближенно преобразуется в нормальное распределение. Это преобразование задается формулой

$$z = \frac{1}{2} \ln \frac{1+r}{1-r}. \quad (26)$$

Распределение случайной величины z очень хорошо аппроксимируется нормальным распределением с не зависящей от ρ дисперсией

$$\sigma_z = \frac{1}{n-3} \quad (27)$$

и средним значением

$$\xi z = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)}. \quad (28)$$

Поправочный член в правой части (28) всегда мал сравнительно с квадратичным отклонением σ_z и поэтому таким сла-

гаемым можно пренебречь. Оно оказывается существенным лишь тогда, когда вычисляется много выборочных коэффициентов корреляции и из соответствующих значений z образуется среднее.

Насколько сильно улучшает нормальное приближение переход от r к z , можно видеть на рис. 37 и 38, заимствованных из книги: Fisher R. A., *Mathematical Methods for Research Workers* (11th ed., Fig. 8, p. 200).

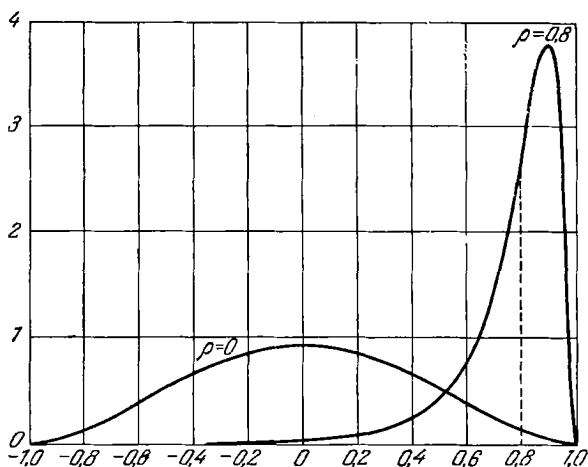


Рис. 37. Плотности распределения r при $\rho = 0$ и $\rho = 0,8$.

С помощью преобразования z можно решать следующие задачи.

1. Проверить, согласуется ли выборочный коэффициент корреляции r с предполагаемым значением теоретического коэффициента корреляции ρ ?

2. Найти доверительные границы для ρ по наблюдаемому значению r .

3. Проверить, соответствуют два выборочных коэффициента корреляции r_1 и r_2 одинаковым значениям ρ или нет?

4. Предполагается, что нескольким выборочным коэффициентам корреляции r_1, r_2, \dots соответствуют одинаковые значения ρ . Требуется найти наилучшую оценку для ρ .

Пример 48. По выборке, состоящей из 25 пар наблюдений, было найдено, что выборочный коэффициент корреляции равен 0,60. В каких доверительных границах находится истинный коэффициент корреляции ρ , если предполагается, что наблюдаемые пары независимы и распределены одинаково нормально?

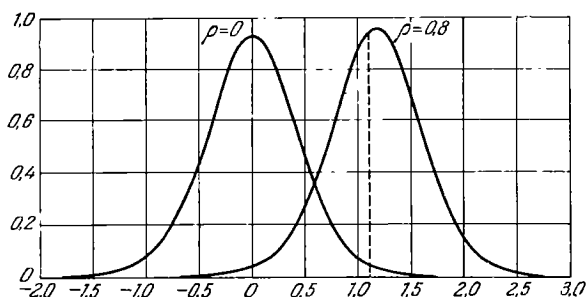
По формуле (26) находим $z = 0,693$. Согласно (27), квадратичное отклонение z равно

$$\sigma_z = \frac{1}{\sqrt{22}} = 0,2132.$$

Следовательно, доверительными границами для z с уровнем значимости 0,05 будут

$$z_1 = 0,693 - 1,96 \sigma_z = 0,275,$$

$$z_2 = 0,693 + 1,96 \sigma_z = 1,111.$$



Р и с. 38. Плотности распределения z при $\rho = 0$ и $\rho = 0,8$.

Решая уравнение (26), находим r как функцию от z :

$$r = \frac{e^{2z} - 1}{e^{2z} + 1}.$$

Таким образом, доверительные границы для ρ с уровнем значимости 0,05 равны

$$r_1 = 0,268 \text{ и } r_2 = 0,804.$$

Пример 49. По выборке, состоящей из 20 пар наблюдений, было найдено, что $r_1 = 0,6$. В другой выборке, состоящей из 25 пар наблюдений, выборочный коэффициент корреляции оказался равным $r_2 = 0,8$. Значимо ли различие r_1 и r_2 ?

Находим

$$z_1 = 0,693, z_2 = 1,099, d = z_2 - z_1 = 0,406.$$

Дисперсии z_1 , z_2 и d равны

$$\sigma_1^2 = \frac{1}{17} = 0,0588, \quad \sigma_2^2 = \frac{1}{22} = 0,0455, \quad \sigma_d^2 = \sigma_1^2 + \sigma_2^2 = 0,1043.$$

Зная d и σ_d , найдем их отношение:

$$\frac{d}{\sigma_d} = \frac{0,406}{0,323} = 1,26.$$

Так как 5%-ная граница для $|d/\sigma_d|$ равна 1,96, то различие r_1 и r_2 следует считать незначимым.

§ 70. Коэффициент ранговой корреляции R , по СпирменуА. ОПРЕДЕЛЕНИЕ R

Отдельные индивидуумы могут обладать таким признаком, который хотя и не поддается точной количественной оценке, однако позволяет сравнивать индивидуумы друг с другом. В результате всю совокупность индивидуумов удастся упорядочить, приписав каждому из них порядковый номер. Такие признаки мы будем называть *качественными признаками*. Примерами качественных признаков являются успеваемость школьников по определенному предмету, музыкальность, цвет волос. Если хотят проверить зависимость двух таких качественных признаков, то рассматривают выборку из n независимых индивидуумов и каждому индивидууму приписывают два порядковых номера, соответствующие двум данным признакам. Из этих порядковых номеров можно построить *коэффициент ранговой корреляции*.

Как обычно, мы будем приписывать порядковые номера в соответствии с убыванием качества: первый номер приписывается наилучшему индивидууму в данном классе и т. д.

Пусть n индивидуумам по двум сравниваемым признакам приписаны порядковые номера $1, 2, \dots, n$. Сначала, для того чтобы арифметическое среднее равнялось нулю, мы из этих номеров вычтем $(n + 1)/2$, а затем все результаты удвоим и обозначим их ξ (для первого качественного признака) и η (для второго качественного признака); ξ и η выражаются целыми числами. Такой порядковый номер индивидуума (ξ или η) равен $k - l$, если по данному признаку этот индивидуум превосходит l других индивидуумов и при этом его самого превосходят k индивидуумов ($k + l = n - 1$).

Сумма квадратов порядковых номеров ξ или η равна

$$Q = \sum \xi^2 = \sum \eta^2 = (n - 1)^2 + (n - 3)^2 + \dots + (-n + 1)^2 = \\ = \frac{n(n-1)(n+1)}{3}.$$

Коэффициент ранговой корреляции R , по Спирмену, определяется формулой

$$R = \frac{\sum \xi \eta}{Q}. \quad (1)$$

Спирмен применял его для психологических исследований¹. Обычно этот коэффициент обозначается буквой ρ , однако у нас эта буква имеет другое значение.

¹ Spearman C., The proof and measurement of association between two things, Amer. J. Psychol., 15 (1904), 88.

Крайними значениями R снова являются $+1$ и -1 , причем значение $+1$ достигается тогда, когда оба ряда порядковых номеров полностью совпадают ($\xi - \eta = 0$), а значение -1 достигается тогда, когда оба ряда полностью противоположны друг другу ($\xi + \eta = 0$).

Для вычисления R удобен следующий способ. Применим обычную нумерацию от 1 до n и для каждого индивидуума вычислим разность d порядковых номеров по обоим признакам. Тогда

$$R = 1 - 2 \frac{\sum d^2}{Q} = 1 - \frac{6 \sum d^2}{n(n-1)(n+1)}. \quad (2)$$

Два качественных признака называются независимыми, если при любом n порядковые номера ξ и η являются независимыми случайными величинами. Таким образом, в случае независимости признаков среднее значение каждого слагаемого в числителе (1) равно нулю, а поэтому $E R = 0$. Следовательно, если R значительно отличается от нуля, то можно сделать вывод, что признаки являются зависимыми. Для того чтобы уточнить смысл слов «значительно отличается», мы должны будем исследовать, сколь велико может оказаться чисто случайное отклонение R от нуля, когда оба признака независимы. Иными словами, мы должны исследовать функцию распределения R в случае двух независимых качественных признаков.

Б. РАСПРЕДЕЛЕНИЕ R В СЛУЧАЕ НЕЗАВИСИМЫХ ПРИЗНАКОВ

Если два качественных признака являются независимыми, то последовательность порядковых номеров η не зависит от последовательности ξ . Последовательности η представляют собой перестановки всех ξ , и для заданной последовательности ξ все такие перестановки равновероятны. Каждой перестановке соответствует определенное значение R . Таким образом, независимо от распределения случайных величин ξ и η искомая функция распределения представляет собой вполне определенную ступенчатую функцию, которую при малых n можно легко вычислить, если только иметь необходимое терпение. Например, в случае $n = 5$ нужно будет вычислить значения R для всех 120 перестановок η . В действительности, такие расчеты проводились до $n = 8$; результаты изложены в книге: Kendall M. G., Rank Correlation Methods (London, 1948), Appendix Table 2. При $n > 8$ расчеты становятся очень утомительными.

В случае независимости признаков среднее значение R равно нулю. Вычислим теперь дисперсию σ^2 . Из формулы (1) следует, что

$$Q^2 R^2 = (\sum \xi_i \eta_i) (\sum \xi_k \eta_k) = \sum_i \sum_k \xi_i \xi_k \eta_i \eta_k,$$

$$Q^2 \sigma^2 = Q^2 \varepsilon R^2 = \sum_i \sum_k \varepsilon \xi_i \xi_k \cdot \varepsilon \eta_i \eta_k.$$

Эта сумма содержит n слагаемых с индексами $i = k$, и так как все индексы равноправны, то все эти слагаемые имеют одинаковые значения. Точно так же оказываются равными друг другу и остальные слагаемые с индексами $i \neq k$; таких слагаемых имеется $n(n-1)$. Поэтому

$$Q^2 \sigma^2 = n \varepsilon \xi_1^2 \cdot \varepsilon \eta_1^2 + n(n-1) \varepsilon \xi_1 \xi_2 \cdot \varepsilon \eta_1 \eta_2. \quad (3)$$

В силу равенства $\sum \xi_i = 0$, имеем

$$0 = \varepsilon (\xi_1 \sum \xi_i) = \varepsilon \xi_1^2 + (n-1) \varepsilon \xi_1 \xi_2,$$

следовательно,

$$\varepsilon \xi_1 \xi_2 = -\frac{1}{n-1} \varepsilon \xi_1^2, \quad (4)$$

и точно так же

$$\varepsilon \eta_1 \eta_2 = -\frac{1}{n-1} \varepsilon \eta_1^2. \quad (5)$$

Далее,

$$\sum \xi_i^2 = Q. \quad (6)$$

Если от обеих частей (6) вычислить средние значения, то получим

$$n \varepsilon \xi_1^2 = Q.$$

То же самое, конечно, справедливо и для η_1 . Таким образом,

$$\varepsilon \xi_1^2 = \varepsilon \eta_1^2 = \frac{Q}{n}, \quad (7)$$

и далее, согласно (4) и (5),

$$\varepsilon \xi_1 \xi_2 = \varepsilon \eta_1 \eta_2 = -\frac{Q}{n(n-1)}. \quad (8)$$

Если (7) и (8) подставить в (3), то получим

$$Q^2 \sigma^2 = \frac{Q^2}{n} + \frac{Q^2}{n(n-1)} = \frac{Q^2}{n-1},$$

или, после сокращения Q^2 ,

$$\sigma^2 = \frac{1}{n-1}. \quad (9)$$

Тем же самым методом можно вычислить и четвертый момент R , т. е. математическое ожидание R^4 . Находим

$$\mathfrak{E} R^4 = \frac{3}{(n-1)^2} \frac{25n^3 - 38n^2 - 35n + 72}{25n(n^2 - 1)}. \quad (10)$$

В. СРАВНЕНИЕ С НОРМАЛЬНЫМ РАСПРЕДЕЛЕНИЕМ

Если в (10) последние два члена — $35n + 72$ заменить на $-25n + 38$ (от этого правая часть (10) изменится лишь очень незначительно), то можно будет числитель и знаменатель сократить на $n^2 - 1$. В результате получим

$$\mathfrak{E} R^4 \sim \frac{3}{(n-1)^2} \left(1 - \frac{38}{25n}\right). \quad (11)$$

Сравнение правой части (11) с четвертым моментом нормального распределения, обладающего нулевым средним значением и дисперсией $1/(n-1)$,

$$\mu_4 = \frac{3}{(n-1)^2}, \quad (12)$$

показывает, что выражения (10) и (12) асимптотически равны и что (10) несколько меньше, чем (12). Так как распределение больших отклонений R в значительной мере определяется средним значением R^4 , то отсюда следует, что значения R , сильно отклоняющиеся от нуля, возникают с несколько меньшей вероятностью, чем это было бы при нормальном распределении с той же дисперсией.

Если вычислить высшие моменты $\mathfrak{E} R^{2k}$, то окажется, что все они, после умножения на $(n-1)^k$, при $n \rightarrow \infty$ стремятся к соответствующим моментам нормального распределения¹ с нулевым средним значением и единичной дисперсией, а именно, к

$$\frac{(2k)!}{2^k k!}. \quad (13)$$

Согласно «второй предельной теореме» (§ 24 E), отсюда следует, что функция распределения $R\sqrt{n-1}$ при $n \rightarrow \infty$ стремится к функции нормального распределения с нулевым средним значением и единичной дисперсией, т. е. *случайная величина R распределена асимптотически нормально с нулевым средним значением и дисперсией $\sigma^2 = 1/(n-1)$.*

Если при проверке независимости действовать так, как если бы R была нормально распределенной случайной величиной, то надежность выводов лишь увеличится. Действительно, как мы только что видели, большим отклонениям R от нуля соответствует,

¹ См. Kendall M. G., Rank Correlation Methods, p. 61.

на самом деле, несколько меньшая вероятность, чем при нормальном распределении. Отсюда получаем следующий простой критерий зависимости:

Если значение коэффициента ранговой корреляции R (или, в случае двустороннего критерия, значение $|R|$) окажется больше, чем

$$R_p = \frac{\psi(1-\beta)}{\sqrt{n-1}}, \quad (14)$$

то гипотезу независимости качественных признаков следует отвергнуть.

Истинный уровень значимости одностороннего критерия $< \beta$, двустороннего критерия $< 2\beta$.

Г. СРАВНЕНИЕ С РАСПРЕДЕЛЕНИЕМ СТЬЮДЕНТА

Кендалл заметил, что распределение с плотностью вероятности

$$f(R) = \frac{1}{B\left(\frac{1}{2}, \frac{n}{2} - 1\right)} (1 - R^2)^{\frac{n}{2} - 2} \quad (15)$$

аппроксимирует распределение случайной величины R несколько лучше, чем нормальное распределение. Функция $B(p, q)$ представляет собой бета-функцию, определенную в § 12 Д. Второй и четвертый моменты распределения с плотностью (15) задаются формулами

$$\mu'_2 = \frac{1}{n-1},$$

$$\mu'_4 = \frac{3}{n^2 - 1}.$$

Следовательно, μ'_2 в точности равен дисперсии коэффициента ранговой корреляции. Четвертый момент μ'_4 можно записать так:

$$\mu'_4 = \frac{3}{(n-1)^2} \left(1 - \frac{2}{n+1}\right). \quad (16)$$

Если (16) сравнить с (11), то окажется, что μ'_4 несколько меньше (11). Таким образом, это приближение не увеличивает, а уменьшает надежность критерия.

Если вместо случайной величины R , подчиняющейся распределению с плотностью (15), ввести новую случайную величину

$$t = R \sqrt{\frac{n-2}{1-R^2}}, \quad (17)$$

то точным распределением t будет являться распределение Стьюдента с $n - 2$ степенями свободы. Практически это означает, что по найденному значению коэффициента ранговой корреляции R можно вычислить t по формуле (17) и затем применить критерий Стьюдента. Правда, при этом истинный уровень значимости будет несколько больше, чем β или 2β . Более простым и надежным является применение границы (14), основанной на нормальном приближении.

Выводы, полученные нами с помощью исследования вторых и четвертых моментов, можно непосредственно проверить в случае $n = 8$. Для границ получаем следующие значения:

Уровни значимости	$2\beta = 0,01$	$2\beta = 0,02$	$2\beta = 0,05$
Точные границы	0,86 (0,007)	0,82 (0,015)	0,72 (0,046)
Нормальное приближение	0,97 (0,0004)	0,88 (0,007)	0,74 (0,036)
Стьюдентовское приближение	0,83 (0,011)	0,79 (0,022)	0,71 (0,058)

Соответствующие этим границам истинные уровни значимости указаны в скобках. Из таблицы видно, что истинные уровни значимости границ, полученных с помощью стьюдентовского приближения, оказываются систематически слишком большими. С другой стороны, эти же уровни для границ, полученных с помощью нормального приближения, слишком малы. Можно бы было, например, в качестве границы выбрать арифметическое среднее, составленное из нормального и стьюдентовского приближений; вероятно, в этом случае надежность критерия должна всегда повышаться¹.

При очень больших n безразлично, каким приближением пользоваться — нормальным или стьюдентовским.

¹ С помощью небольшого усложнения формулы (14) можно нормальное приближение существенно улучшить. А именно, нужно воспользоваться уточненной асимптотической формулой

$$R_{\beta} = \frac{\Psi(1 - \beta)}{\sqrt{n-1}} \left\{ 1 - \frac{0,19}{n-1} [\Psi^2(1 - \beta) - 3] \right\}.$$

Значения R_{β} , вычисленные по этой формуле при $n = 8$ и $2\beta = 0,01; 0,02; 0,05$, равны соответственно 0,87; 0,82 и 0,72. Таким образом, указанное приближение значительно лучше нормального и стьюдентовского. Вычисления с большим количеством знаков показывают, что приближенные границы больше точных, т. е. улучшенное приближение лишь увеличивает надежность критерия. — *Прим. перев.*

Д. СЛУЧАЙ ЗАВИСИМЫХ ПРИЗНАКОВ

Мы хотим теперь исследовать, каково соотношение между истинным коэффициентом корреляции ρ и коэффициентом ранговой корреляции R в случае зависимых признаков?

Предположим, что в основе двух качественных признаков лежат две нормальные случайные величины x и y , плотность распределения которых задается формулой

$$f(x, y) = \frac{1}{2\pi} \sqrt{1 - \rho^2} e^{-\frac{1}{2}(x^2 - 2\rho xy + y^2)}. \quad (18)$$

Для n независимых пар (x_i, y_i) плотность совместного распределения равна

$$f(x_1, y_1) \dots f(x_n, y_n) = \frac{1}{(2\pi)^n} (1 - \rho^2)^{\frac{n}{2}} e^{-\frac{1}{2} \sum (x_i^2 - 2\rho x_i y_i + y_i^2)} \quad (19)$$

Согласно (1),

$$QR = \sum \xi_i \eta_i, \quad (20)$$

где ξ_i — разность между количеством тех x_k , которые больше x_i , и количеством тех x_k , которые меньше x_i .

Пусть x_{ik} и y_{ik} — случайные величины, которые при всех i и k определяются равенствами

$$x_{ik} = \begin{cases} +1, & \text{если } x_i < x_k, \\ 0, & \text{если } x_i = x_k, \\ -1, & \text{если } x_i > x_k, \end{cases} \quad y_{ik} = \begin{cases} +1, & \text{если } y_i < y_k, \\ 0, & \text{если } y_i = y_k, \\ -1, & \text{если } y_i > y_k. \end{cases}$$

Тогда

$$\xi_i = \sum_k x_{ik}, \quad \eta_i = \sum_k y_{ik}.$$

Если эти суммы подставим в (20), то получим

$$QR = \sum_i \sum_k \sum_l x_{ik} y_{il}. \quad (21)$$

Вычислим теперь математические ожидания от обеих частей равенства (21):

$$Q\hat{R} = \sum_i \sum_k \sum_l \mathfrak{E}(x_{ik} y_{il}). \quad (22)$$

Все слагаемые этой суммы, у которых $i = k$ или $i = l$, равны нулю, поэтому сумма (22) содержит $n(n-1)(n-2)$ равных друг другу слагаемых с $k \neq l$ и $n(n-1)$ одинаковых слагаемых с $k = l$. Таким образом,

$$Q\hat{R} = n(n-1)(n-2) \mathfrak{E}(x_{12} y_{13}) + n(n-1) \mathfrak{E}(x_{12} y_{12}),$$

ли, после деления на $Q = n(n-1)(n+1)/3$,

$$\hat{R} = 3 \frac{n-2}{n+1} \mathfrak{E}(x_{12} y_{13}) + \frac{3}{n+1} \mathfrak{E}(x_{12} y_{12}). \quad (23)$$

Нам нужно теперь вычислить средние значения произведений $x_{12} y_{12}$ и $x_{12} y_{13}$. Случайная величина $x_{12} y_{12}$ принимает значение 1,

если $x_1 < x_2$ и $y_1 < y_2$ или если $x_1 > x_2$ и $y_1 > y_2$; в противном случае, когда $x_1 < x_2$ и $y_1 > y_2$ или когда $x_1 > x_2$ и $y_1 < y_2$, она принимает значение -1 . Вероятность одновременного осуществления двух событий, $x_1 < x_2$ и $y_1 < y_2$, равна интегралу от плотности $f(x_1, y_1)f(x_2, y_2)$ по области, заданной неравенствами $x_1 < x_2$, $y_1 < y_2$:

$$W_1 = \frac{1}{(2\pi)^2} (1 - \rho^2) \times \\ \times \iiint\limits_{\substack{x_1 < x_2 \\ y_1 < y_2}} e^{-\frac{1}{2}(x_1^2 - 2\rho x_1 y_1 + y_1^2)} e^{-\frac{1}{2}(x_2^2 - 2\rho x_2 y_2 + y_2^2)} dx_1 dy_1 dx_2 dy_2.$$

Этот интеграл имеет точно такой же вид, как и интеграл, вычисленный в общем виде в § 14 В:

$$I = (2\pi)^{-\frac{n}{2}} \sqrt{g} \int \dots \int e^{-\frac{1}{2}G} dx^1 \dots dx^n, \quad (24)$$

(ux) > 0
(vx) > 0

где $n = 4$,

$$G = \sum g_{ik} x^i x^k = x_1^2 - 2\rho x_1 y_1 + y_1^2 + x_2^2 - 2\rho x_2 y_2 + y_2^2 \quad (25)$$

и

$$g = (1 - \rho^2)^2.$$

Для того чтобы перейти к обозначениям из § 14 В, нужно положить

$$\begin{aligned} x^1 &= x_1, & x^2 &= y_1, & x^3 &= x_2, & x^4 &= y_2, \\ u_1 &= -1, & u_2 &= 0, & u_3 &= +1, & u_4 &= 0, \\ v_1 &= 0, & v_2 &= -1, & v_3 &= 0, & v_4 &= +1. \end{aligned}$$

Квадратичной формой, контраградиентной форме G , является

$$\sum g^{ik} u_i u_k = \frac{u_1^2 + 2\rho u_1 u_2 + u_2^2}{1 - \rho^2} + \frac{u_3^2 + 2\rho u_3 u_4 + u_4^2}{1 - \rho^2}. \quad (26)$$

Отсюда получаем значения инвариантов:

$$\begin{aligned} (uu) &= \sum g^{ik} u_i u_k = \frac{1}{1 - \rho^2} + \frac{1}{1 - \rho^2} = \frac{2}{1 - \rho^2}, \\ (uv) &= \sum g^{ik} u_i v_k = \frac{\rho}{1 - \rho^2} + \frac{\rho}{1 - \rho^2} = \frac{2\rho}{1 - \rho^2}, \\ (vv) &= \sum g^{ik} v_i v_k = \frac{1}{1 - \rho^2} + \frac{1}{1 - \rho^2} = \frac{2}{1 - \rho^2}. \end{aligned}$$

Поэтому интеграл $W_1 = I$ равен

$$W_1 = \frac{1}{2\pi} \arccos \frac{-(uv)}{\sqrt{(uu)} \sqrt{(vv)}} = \frac{1}{2\pi} \arccos(-\rho). \quad (27)$$

Точно такое же значение имеет и вероятность W_2 одновременного осуществления событий $x_1 > x_2$ и $y_1 > y_2$. Вероятность W_3 одновременного осуществления событий $x_1 > x_2$ и $y_1 < y_2$ получается умножением v_i на (-1) :

$$W_3 = \frac{1}{2\pi} \arccos \rho. \quad (28)$$

Такое же значение имеет и вероятность W_4 одновременного осуществления событий $x_1 < x_2$ и $y_1 > y_2$. Поэтому среднее значение $x_{12} y_{12}$ равно

$$\begin{aligned} E x_{12} y_{12} &= (W_1 + W_2) \cdot 1 + (W_3 + W_4) (-1) = 2W_1 - 2W_3 = \\ &= \frac{1}{\pi} \arccos (-\rho) - \frac{1}{\pi} \arccos \rho = \\ &= \frac{1}{\pi} \left(\frac{\pi}{2} + \arcsin \rho \right) - \frac{1}{\pi} \left(\frac{\pi}{2} - \arcsin \rho \right) = \frac{2}{\pi} \arcsin \rho. \quad (29) \end{aligned}$$

Аналогично вычисляется среднее значение случайной величины $x_{12} y_{13}$. Она принимает значение 1, если $x_1 < x_2$ и $y_1 < y_3$ или если $x_1 > x_2$ и $y_1 > y_3$. Вероятность одновременного осуществления первых двух событий равна

$$W_5 = \frac{1}{(2\pi)^3} (1 - \rho^2)^{\frac{3}{2}} \iiint_{\substack{x_1 < x_2 \\ y_1 < y_3}} e^{-\frac{1}{2} G} dx_1 dy_1 dx_2 dy_2 dx_3 dy_3,$$

где

$$G = (x_1^2 - 2\rho x_1 y_1 + y_1^2) + (x_2^2 - 2\rho x_2 y_2 + y_2^2) + (x_3^2 - 2\rho x_3 y_3 + y_3^2),$$

$$\sum g^{ik} u_i u_k = \frac{u_1^2 + 2\rho u_1 u_2 + u_2^2}{1 - \rho^2} + \frac{u_3^2 + 2\rho u_3 u_4 + u_4^2}{1 - \rho^2} + \frac{u_5^2 + 2\rho u_5 u_6 + u_6^2}{1 - \rho^2}.$$

Инварианты имеют значения

$$(uu) = \frac{2}{1 - \rho^2}, \quad (uv) = \frac{\rho}{1 - \rho^2}, \quad (vv) = \frac{2}{1 - \rho^2}.$$

Следовательно,

$$W_5 = \frac{1}{2\pi} \arccos \sqrt{\frac{(uv)}{(uu)} \frac{(uv)}{(vv)}} = \frac{1}{2\pi} \arccos \frac{\rho}{2}. \quad (30)$$

Точно такое же значение имеет вероятность W_6 одновременного осуществления событий $x_1 > x_2$ и $y_1 > y_3$. Вероятности W_7 и W_8 остальных двух возможных случаев равны

$$W_7 = W_8 = \frac{1}{2\pi} \arccos \frac{\rho}{2}.$$

Поэтому

$$\mathcal{E}(x_{12} y_{13}) = (W_5 + W_6) \cdot 1 + (W_7 + W_8) (-1) = \frac{2}{\pi} \arcsin \frac{\varrho}{2}. \quad (31)$$

Если (29) и (31) подставить в (23), то получим

$$\hat{R} = \frac{6}{\pi} \frac{n-2}{n+1} \arcsin \frac{\varrho}{2} + \frac{6}{\pi(n+1)} \arcsin \varrho. \quad (32)$$

При больших n из (32) следует, что

$$\hat{R} \sim \frac{6}{\pi} \arcsin \frac{\varrho}{2}. \quad (33)$$

При не очень больших n значение \hat{R} несколько меньше, чем правая часть (33), так как, в силу неравенств

$$2 \arcsin \frac{\varrho}{2} < \arcsin \varrho < 3 \arcsin \frac{\varrho}{2},$$

имеем

$$\frac{6}{\pi} \frac{n}{n+1} \arcsin \frac{\varrho}{2} < \hat{R} < \frac{6}{\pi} \arcsin \frac{\varrho}{2}.$$

Различие между $n/(n+1)$ и 1 совсем незначительно, поэтому приближение (33) можно применять и для умеренно больших значений n . Если (33) разрешить относительно ϱ , то получим

$$\varrho \sim 2 \sin \frac{\pi}{6} \hat{R}. \quad (34)$$

Это означает, что при больших n выражение $2 \sin (\pi R/6)$ можно использовать в качестве оценки для истинного коэффициента корреляции ϱ .

Все это справедливо лишь в предположении, что совместное распределение x и y является нормальным. Если такое предположение не выполняется, то R всегда можно истолковать как оценку для «истинного коэффициента ранговой корреляции» \mathcal{R} , который определяется так. Пусть $F(x)$ и $G(y)$ — непрерывные функции распределения случайных величин x и y . Положим¹

$$\xi = F(x), \quad \eta = G(y).$$

Тогда случайные величины ξ и η будут распределены равномерно между нулем и единицей, поэтому их дисперсии равны

$$\sigma_\xi^2 = \sigma_\eta^2 = \frac{1}{12}.$$

¹ См. Kendall M. G., Rank correlation methods 9.7; 10.6. Величины ξ и η Кендалл называет „рангами“.

Истинный коэффициент ранговой корреляции R определяется как истинный коэффициент корреляции случайных величин ξ и η :

$$R = 12 \mathcal{E}\left(\xi - \frac{1}{2}\right)\left(\eta - \frac{1}{2}\right). \quad (35)$$

Если совместное распределение x и y является нормальным, то между ρ и R существует соотношение, аналогичное (34),

$$\rho = 2 \sin \frac{\pi}{6} R. \quad (36)$$

Это соотношение было найдено Карлом Пирсоном (см. Draper's Company Research Memoirs, Biometric Series IV, Cambridge 1907, 13).

Пример 50 (Pearson Karl, *Biometrika*, 13, 304). На экзаменах, которым подвергались 27 кандидатов на должность в службе связи, была принята следующая система оценок: по арифметике — от 1 до 300 баллов, по остальным четырем предметам (орфография, чистописание, география и сочинение на английском языке) — от 1 до 200 баллов. Оценки, полученные каждым кандидатом, складывались и всем кандидатам приписывались порядковые номера в соответствии с убыванием сумм оценок. В следующей таблице на первом месте указаны порядковые номера и общие суммы оценок по всем предметам, а на втором месте — оценки по арифметике и соответствующие им порядковые номера.

Пирсон нашел, что в данном случае коэффициент ранговой корреляции имеет значение

$$R = 0,8834.$$

Все предметы		Арифметика		Все предметы		Арифметика	
номер	оценка	номер	оценка	номер	оценка	номер	оценка
1	907	1	230	15	580	13	131
2	764	9	158	16	561	15	128
3	748	2	228	17	560	18	116
4	746	10	154	18	532	22	82
5	724	8	162	19	529	16	125
6	718	5	182	20	526	17	122
7	710	14	129	21	515	19	114
8	703	7	164	22	484	21	93
9	677	3	187	23	463	25	61
10	665	4	186	24	444	26	38
11	645	11	151	25	386	27	37
12	643	6	167	26	369	23	63
13	634	20	103	27	288	24	62
14	628	12	146				

Отсюда, в силу формулы (34), получается оценка для истинного коэффициента корреляции, а именно

$$r' = 2 \sin \frac{\pi}{6} R = 0,893.$$

Выборочный коэффициент корреляции r , вычисленный непосредственно по оценкам, равен

$$r = 0,896.$$

Пирсон справедливо замечает: «Согласие между r и r' в этом случае является превосходным».

§ 71. Коэффициент ранговой корреляции T , по Кендаллу

В родственной связи с R находится коэффициент ранговой корреляции τ , который для наших целей предпочтительнее обозначить буквой T . Этот коэффициент был введен Грейнером и Эшером и заново открыт Кендаллом. обстоятельное исследование свойств T можно найти в уже неоднократно цитированной книге: Kendall M. G., Rank Correlation Methods. В этом параграфе мы затронем лишь некоторые основные вопросы.

А. ОПРЕДЕЛЕНИЕ T

Пусть снова имеется n индивидуумов, упорядоченных по двум качественным признакам. Для каждой пары индивидуумов (i, k) мы определим функцию, принимающую значения: $+1$, если порядковые номера одного индивидуума превосходят соответствующие порядковые номера другого индивидуума, и -1 в противном случае. В обозначениях § 70 Д эта функция для пары индивидуумов (i, k) равна произведению $x_{ik}y_{ik}$. Сумма S таких произведений для всех пар,

$$S = \sum x_{ik}y_{ik}, \quad (1)$$

по абсолютной величине не превосходит

$$\binom{n}{2} = \frac{n(n-1)}{2}.$$

Следовательно, если положить

$$T = \frac{2S}{n(n-1)}, \quad (2)$$

то значения T будут принадлежать отрезку, расположенному между -1 и $+1$. При этом $T = +1$ тогда и только тогда, когда обе последовательности порядковых номеров совпадают (разность порядковых номеров каждого индивидуума равна нулю), и $T = -1$ тогда и только тогда, когда обе последовательности противо-

положны друг другу (сумма порядковых номеров каждого индивидуума равна $n + 1$).

Если номера первой последовательности расположить в возрастающем порядке от 1 до n и под каждым из них написать номер из второй последовательности

$$\xi_1 = 1, \quad \xi_2 = 2, \quad \dots, \quad \xi_n = n,$$

$$\eta_1 \quad , \quad \eta_2 \quad , \quad \dots, \quad \eta_n \quad ,$$

то S можно будет вычислить следующим образом: подсчитаем количество тех η_k , которые стоят правее η_1 и величина которых превосходит η_1 , затем подсчитаем, сколько имеется номеров η_k , больших η_2 и расположенных правее η_2 , и т. д. Пусть P — сумма всех этих количеств, тогда S представляет собой сумму P величин, равных $+1$, и $\binom{n}{2} - P$ величин, равных -1 , т. е.

$$S = 2P - \frac{1}{2} n(n - 1), \tag{3}$$

$$T = \frac{4P}{n(n-1)} - 1. \tag{4}$$

Б. РАСПРЕДЕЛЕНИЕ T

Если случайные величины x и y независимы, то математическое ожидание T , очевидно, равно нулю. Если же x и y зависимы и распределены нормально с коэффициентом корреляции ρ , то математическое ожидание S принимает значение

$$\mathcal{E} S = \mathcal{E} \sum x_{ik} y_{ik} = \frac{n(n-1)}{2} \mathcal{E} x_{12} y_{12}$$

т. е., согласно (29) § 70,

$$\mathcal{E} S = \frac{n(n-1)}{2} \frac{2}{\pi} \arcsin \rho.$$

Таким образом,

$$\mathcal{E} T = \frac{2}{\pi} \arcsin \rho. \tag{5}$$

Поэтому если совместное распределение x и y является нормальным, то величину

$$r'' = \sin \frac{\pi T}{2} \tag{6}$$

можно использовать в качестве оценки для ρ .

Вернемся к случаю независимых x и y и с помощью (1) вычислим дисперсию S :

$$\sigma_S^2 = \mathcal{E} S^2 = \mathcal{E} (\sum x_{ij} y_{ij})^2.$$

Вычисления полностью приводятся в § 5.6 книги Кендалла. Поэтому мы укажем лишь результат:

$$\sigma_S^2 = \frac{n(n-1)(2n+5)}{18}. \quad (7)$$

Отсюда, в силу (2), следует, что

$$\sigma_T^2 = \frac{2(2n+5)}{9n(n-1)}. \quad (8)$$

Если таким же образом вычислить и высшие моменты ξT^4 , $\xi T^6, \dots$ (моменты нечетного порядка равны нулю, так как значения T и $-T$ всегда имеют равные вероятности), то окажется, что они при $n \rightarrow \infty$ асимптотически равны соответствующим моментам нормального распределения с дисперсией σ_T^2 :

$$\mu_{2r} = \frac{(2r)!}{2^r r!} \sigma_T^{2r}. \quad (9)$$

Отсюда следует, что случайная величина T распределена асимптотически нормально с нулевым средним значением и дисперсией (8).

Асимптотическая нормальность остается справедливой даже и в том случае, когда x и y являются зависимыми величинами с произвольной функцией распределения, если только абсолютная величина математического ожидания ξT не слишком близка к единице. Доказательство см. в книге: Kendall M. G., Rank correlation methods 5.21.

В случае зависимых случайных величин теорема об асимптотической нормальности практически не очень полезна, так как при умеренно больших значениях n распределение T может обладать значительной асимметрией и, кроме того, дисперсия T неизвестна. Однако, в случае независимых x и y , дисперсия T известна [она задается формулой (8)], и уже при $n = 8$ нормальное приближение оказывается очень хорошим. Следовательно, T можно с успехом использовать в качестве статистики критерия для проверки независимости признаков. Точное распределение T известно для $n \leq 10$ (Kendall M. G., Appendix Table 1); при $n > 10$ можно воспользоваться нормальным приближением, т. е. гипотезу независимости следует отвергнуть тогда, когда T (или, в случае двустороннего критерия, $|T|$) превосходит границу

$$T_\beta = \sigma_T \Psi(1 - \beta). \quad (10)$$

Уровень значимости такого одностороннего критерия приближенно равен β , а уровень двустороннего критерия приближенно равен 2β .

В. СРАВНЕНИЕ R И T

Если мы предположим, что x и y распределены нормально, и разложим математические ожидания \hat{R} и \hat{T} в степенные ряды по ρ , то получим

$$\hat{R} = \frac{3}{\pi} \left(\frac{n}{n+1} \rho + \frac{1}{24} \frac{n+6}{n+1} \rho^3 + \dots \right), \quad (11)$$

$$\hat{T} = \frac{2}{\pi} \left(\rho + \frac{1}{6} \rho^3 + \dots \right). \quad (12)$$

Отсюда следует, что при больших n и малых ρ математические ожидания отнесутся приблизительно как 3:2. С другой стороны, если при $\rho = 0$ сравнить квадратичные отклонения, а именно

$$\sigma_R = \frac{1}{\sqrt{n-1}}, \quad (13)$$

$$\sigma_T = \sqrt{\frac{2(2n+5)}{9n(n-1)}} = \frac{2}{3} \sqrt{\frac{1 + \frac{5}{2n}}{n-1}}, \quad (14)$$

то при больших n снова найдем то же отношение 3:2. Отсюда можно, предположительно, сделать вывод, что случайные величины R и T отнесутся друг к другу приблизительно как 3:2, если только их абсолютные величины не слишком близки к единице. Этот вывод был подтвержден результатами Дэниелса, согласно которым коэффициент корреляции случайных величин R и T при $n \rightarrow \infty$ стремится к единице, причем он близок к единице уже при не слишком больших значениях n (см. Kendall M. G., Rank correlation methods 5.14).

Какая статистика является наилучшей для построения критерия независимости признаков, R или T ? Односторонний критерий R отвергает гипотезу независимости, если $R > R_\beta$; в случае критерия T тот же вывод делается при $T > T_\beta$. Нам нужно исследовать, какой из этих двух критериев имеет наибольшую мощность (в смысле § 59). Искл мощностью критерия в данном случае мы понимаем вероятность отвергнуть гипотезу независимости, когда случайные величины x и y действительно являются зависимыми.

Для ответа на этот вопрос мы должны будем сначала сделать некоторое предположение о распределении x и y . А именно, мы предположим, что совместное распределение x и y является нормальным. Тогда, в силу § 69, плотность этого распределения можно задать формулой

$$f(x, y) = \frac{1}{2\pi} \sqrt{1 - \rho^2} e^{-\frac{1}{2}(x^2 - 2\rho xy + y^2)}. \quad (15)$$

В этом случае мощности критериев R и T будут являться функциями от ϱ .

Предположим сперва, что n настолько велико, что обе случайные величины R и T распределены почти нормально. Тогда границы R_β и T_β можно будет определить с помощью нормального распределения:

$$R_\beta = \sigma_R \Psi(1 - \beta) = \frac{\Psi(1 - \beta)}{\sqrt{n-1}}, \quad (16)$$

$$T_\beta = \sigma_T \Psi(1 - \beta) = \frac{2}{3} \sqrt{1 - \frac{5}{2n}} \Psi(1 - \beta). \quad (17)$$

Разумеется, значения σ_R и σ_T в формулах (16) и (17) вычислены при $\varrho = 0$. Если ϱ отлично от нуля, то значения σ_R и σ_T изменятся, однако это изменение является лишь величиной порядка ϱ^2 и поэтому им сначала можно пренебречь. Таким образом, в первом приближении вычисление функций мощности для критериев R и T мы будем производить с помощью нормальных распределений, математические ожидания которых задаются формулами (11) и (12), а квадратичные отклонения — формулами (13) и (14). Тогда мощность критерия R , или вероятность события $R > R_\beta$, будет равна

$$M_R(\varrho) = 1 - \Phi\left(\frac{R_\beta - \hat{R}}{\sigma_R}\right) = \Phi\left(\frac{\hat{R} - R_\beta}{\sigma_R}\right), \quad (18)$$

и точно так же мощность критерия T будет равна

$$M_T(\varrho) = \Phi\left(\frac{\hat{T} - T_\beta}{\sigma_T}\right). \quad (19)$$

Если ϱ велико сравнительно с $1/\sqrt{n-1}$, то \hat{R} будет значительно больше σ_R , а \hat{T} — значительно больше σ_T . Таким образом, в этом случае выражения (18) и (19) окажутся практически равными единице. Поэтому можно предположить, что ϱ^2 является величиной порядка $1/n$. Если теперь в формулах (18) и (19) пренебречь всеми членами порядка малости $1/n$ или выше, то для $M_R(\varrho)$ и $M_T(\varrho)$ получим одну и ту же функцию $M(\varrho)$, а именно

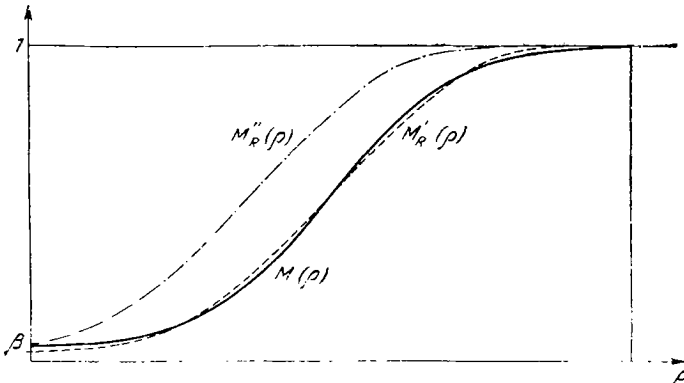
$$M(\varrho) = \Phi\left[\frac{3}{\pi} \varrho \sqrt{n-1} - \Psi(1 - \beta)\right]. \quad (20)$$

Следовательно, в первом приближении оба критерия имеют одну и ту же функцию мощности. График этой функции изображен сплошной линией на рис. 39 и обозначен $M(\varrho)$.

Это приближение мы теперь постараемся последовательно уточнить. Если сначала воспользоваться нормальными приближе-

ниями (18) и (19), но на этот раз в разложениях (11) и (12) учитывать и члены порядка ϱ^3 , а в формуле (17) — член $5/(2n)$, то результат будет практически тем же самым, что и раньше: графики функций $M_R(\varrho)$ и $M_T(\varrho)$ получаются почти совпадающими друг с другом.

Как мы видели, нормальное приближение для распределения T является хорошим, а для распределения R — нет. Истинная функция распределения R при больших положительных значениях R быстрее стремится к единице, а при больших отрицательных значениях R быстрее стремится к нулю, чем соответствующая



Р и с. 39. Графики функций мощности критериев R и T . Сплошная линия соответствует первому приближению для функций мощности R и T ; улучшенное приближение для T указано штриховой линией, а для R — штрихпунктирной линией.

функция нормального распределения. Если учесть это обстоятельство, но границу R_p оставить пока неизменной, то в качестве графика $M_R(\varrho)$ получается кривая, изображенная на рис. 39 штриховой линией и обозначенная $M'_R(\varrho)$. Сперва при малых ϱ эта кривая проходит под сплошной линией, затем она идет над кривой $M(\varrho)$; в точке с ординатой $M = 1/2$ кривая M'_R переходит вниз, и, наконец, при больших ϱ она снова оказывается выше кривой M . Рисунок является лишь качественно правильным: различие между кривыми несколько преувеличено.

При больших n нужно также еще учитывать и асимметрию распределений R и T . Влияние асимметрии приводит к уменьшению функций мощности $M_R(\varrho)$ и $M_T(\varrho)$, особенно при значениях ϱ , близких по абсолютной величине к единице. При этом M_R уменьшается несколько сильнее, чем M_T . Однако это уменьшение не очень значительно; на рисунке оно не указано.

Но мы должны учесть еще одну последнюю поправку, которая является решающей, а именно, поправку к границе R_β . Как мы видели, границы для T , вычисленные с помощью нормального приближения, являются приблизительно правильными, так как случайная величина T распределена приблизительно нормально. Однако для R соответствующие границы слишком велики. При $n = 8$ с помощью нормального приближения мы в качестве 0,5%-ной границы получили $R_\beta = 0,97$, в то время как точная граница равнялась 0,88. В результате мы оказываемся перед выбором: либо, с целью увеличения надежности критерия, оставить границу R_β неизменной (тогда, мощность критерия также не изменится, но истинный уровень значимости будет значительно меньше β), либо уменьшить R_β таким образом, чтобы истинный уровень значимости по-прежнему не превосходил β . В последнем случае штриховая линия сместится влево на значительное расстояние и в результате получится большая функция мощности $M_R''(\rho)$, график которой изображен на рисунке штрих-пунктирной линией. Поэтому

Критерий R по сравнению с критерием T имеет приблизительно равную мощность и меньший истинный уровень значимости или одинаковый истинный уровень значимости и большую мощность.

К этому следует добавить, что вычисление R требует меньшей вычислительной работы, чем вычисление T . Таким образом, оказывается, что старый коэффициент ранговой корреляции R , по Спирмену, теоретически и практически предпочтительнее своего более молодого конкурента T .

Т А Б Л И Ц Ы

Таблица 1

Функция нормального распределения $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{1}{2}x^2} dx$

t	0	1	2	3	4	5	6	7	8	9
-0,0	,5000	,4960	,4920	,4880	,4840	,4801	,4761	,4721	,4681	,4641
-0,1	,4602	,4562	,4522	,4483	,4443	,4404	,4364	,4325	,4286	,4247
-0,2	,4207	,4168	,4129	,4090	,4052	,4013	,3974	,3936	,3897	,3859
-0,3	,3821	,3783	,3745	,3707	,3669	,3632	,3594	,3557	,3520	,3483
-0,4	,3446	,3409	,3372	,3336	,3300	,3264	,3228	,3192	,3156	,3121
-0,5	,3085	,3050	,3015	,2981	,2946	,2912	,2877	,2843	,2810	,2776
-0,6	,2743	,2709	,2676	,2643	,2611	,2578	,2546	,2514	,2483	,2451
-0,7	,2420	,2389	,2358	,2327	,2297	,2266	,2236	,2206	,2177	,2148
-0,8	,2119	,2090	,2061	,2033	,2005	,1977	,1949	,1922	,1894	,1867
-0,9	,1841	,1814	,1788	,1762	,1736	,1711	,1685	,1660	,1635	,1611
-1,0	,1587	,1562	,1539	,1515	,1492	,1469	,1446	,1423	,1401	,1379
-1,1	,1357	,1335	,1314	,1292	,1271	,1251	,1230	,1210	,1190	,1170
-1,2	,1151	,1131	,1112	,1093	,1075	,1056	,1038	,1020	,1003	,0985
-1,3	,0968	,0951	,0934	,0918	,0901	,0885	,0869	,0853	,0838	,0823
-1,4	,0808	,0793	,0778	,0764	,0749	,0735	,0721	,0708	,0694	,0681
-1,5	,0668	,0655	,0643	,0630	,0618	,0606	,0594	,0582	,0571	,0559
-1,6	,0548	,0537	,0526	,0516	,0505	,0495	,0485	,0475	,0465	,0455
-1,7	,0446	,0436	,0427	,0418	,0409	,0401	,0392	,0384	,0375	,0367
-1,8	,0359	,0351	,0344	,0336	,0329	,0322	,0314	,0307	,0301	,0294
-1,9	,0288	,0281	,0274	,0268	,0262	,0256	,0250	,0244	,0239	,0233
-2,0	,0228	,0222	,0217	,0212	,0207	,0202	,0197	,0192	,0188	,0183
-2,1	,0179	,0174	,0170	,0166	,0162	,0158	,0154	,0150	,0146	,0143
-2,2	,0139	,0136	,0132	,0129	,0125	,0122	,0119	,0116	,0113	,0110
-2,3	,0107	,0104	,0102	,0099	,0096	,0094	,0091	,0089	,0087	,0084
-2,4	,0082	,0080	,0078	,0075	,0073	,0071	,0069	,0068	,0066	,0064
-2,5	,0062	,0060	,0059	,0057	,0055	,0054	,0052	,0051	,0049	,0048
-2,6	,0047	,0045	,0044	,0043	,0041	,0040	,0039	,0038	,0037	,0036
-2,7	,0035	,0034	,0033	,0032	,0031	,0030	,0029	,0028	,0027	,0026
-2,8	,0026	,0025	,0024	,0023	,0023	,0022	,0021	,0021	,0020	,0019
-2,9	,0019	,0018	,0018	,0017	,0016	,0016	,0015	,0015	,0014	,0014
t = -3,0	-3,1	-3,2	-3,3	-3,4	-3,5	-3,6	-3,7	-3,8	-3,9	
$\Phi(t) =$,0013	,0010	,0007	,0005	,0003	,0002	,0002	,0001	,0001	,0000

Таблица 1 (продолжение)

t	0	1	2	3	4	5	6	7	8	9
0,0	,5000	,5040	,5080	,5120	,5160	,5199	,5239	,5279	,5319	,5359
0,1	,5398	,5438	,5478	,5517	,5557	,5596	,5636	,5675	,5714	,5753
0,2	,5793	,5832	,5871	,5910	,5948	,5987	,6026	,6064	,6103	,6141
0,3	,6179	,6217	,6255	,6293	,6331	,6368	,6406	,6443	,6480	,6517
0,4	,6554	,6591	,6628	,6664	,6700	,6736	,6772	,6808	,6844	,6879
0,5	,6915	,6950	,6985	,7019	,7054	,7088	,7123	,7157	,7190	,7224
0,6	,7257	,7291	,7324	,7357	,7389	,7422	,7454	,7486	,7517	,7549
0,7	,7580	,7611	,7642	,7673	,7703	,7734	,7764	,7794	,7823	,7852
0,8	,7881	,7910	,7939	,7967	,7995	,8023	,8051	,8078	,8106	,8133
0,9	,8159	,8186	,8212	,8238	,8264	,8289	,8315	,8340	,8365	,8389
1,0	,8413	,8438	,8461	,8485	,8508	,8531	,8554	,8577	,8599	,8621
1,1	,8643	,8665	,8686	,8708	,8729	,8749	,8770	,8790	,8810	,8830
1,2	,8849	,8869	,8888	,8907	,8925	,8944	,8962	,8980	,8997	,9015
1,3	,9032	,9049	,9066	,9082	,9099	,9115	,9131	,9147	,9162	,9177
1,4	,9192	,9207	,9222	,9236	,9251	,9265	,9279	,9292	,9306	,9319
1,5	,9332	,9345	,9357	,9370	,9382	,9394	,9406	,9418	,9429	,9441
1,6	,9452	,9463	,9474	,9484	,9495	,9505	,9515	,9525	,9535	,9545
1,7	,9554	,9564	,9573	,9582	,9591	,9599	,9608	,9616	,9625	,9633
1,8	,9641	,9649	,9656	,9664	,9671	,9678	,9686	,9693	,9699	,9706
1,9	,9713	,9719	,9726	,9732	,9738	,9744	,9750	,9756	,9761	,9767
2,0	,9772	,9778	,9783	,9788	,9793	,9798	,9803	,9808	,9812	,9817
2,1	,9821	,9826	,9830	,9834	,9838	,9842	,9846	,9850	,9854	,9857
2,2	,9861	,9864	,9868	,9871	,9875	,9878	,9881	,9884	,9887	,9890
2,3	,9893	,9896	,9898	,9901	,9904	,9906	,9909	,9911	,9913	,9916
2,4	,9918	,9920	,9922	,9925	,9927	,9929	,9931	,9932	,9934	,9936
2,5	,9938	,9940	,9941	,9943	,9945	,9946	,9948	,9949	,9951	,9952
2,6	,9953	,9955	,9956	,9957	,9959	,9960	,9961	,9962	,9963	,9964
2,7	,9965	,9966	,9967	,9968	,9969	,9970	,9971	,9972	,9973	,9974
2,8	,9974	,9975	,9976	,9977	,9977	,9978	,9979	,9979	,9980	,9981
2,9	,9981	,9982	,9982	,9983	,9984	,9984	,9985	,9985	,9986	,9986

$t = 3,0$	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9
$\Phi(t) = ,9987$,9990	,9993	,9995	,9997	,9998	,9998	,9999	,9999	1,0000

Таблица 2

Обратная функция $\Psi(x)$

$x \rightarrow$ ↓	0	1	2	3	4	5	6	7	8	9
0,00	$-\infty$	-3,09	-2,88	-2,75	-2,65	-2,58	-2,51	-2,46	-2,41	-2,37
0,01	-2,33	-2,29	-2,26	-2,23	-2,20	-2,17	-2,14	-2,12	-2,10	-2,07
0,02	-2,05	-2,03	-2,01	-2,00	-1,98	-1,96	-1,94	-1,93	-1,91	-1,90
0,03	-1,88	-1,87	-1,85	-1,84	-1,83	-1,81	-1,80	-1,79	-1,77	-1,76
0,04	-1,75	-1,74	-1,73	-1,72	-1,71	-1,70	-1,68	-1,67	-1,66	-1,65
0,05	-1,64	-1,64	-1,63	-1,62	-1,61	-1,60	-1,59	-1,58	-1,57	-1,56
0,06	-1,55	-1,55	-1,54	-1,53	-1,52	-1,51	-1,51	-1,50	-1,49	-1,48
0,07	-1,48	-1,47	-1,46	-1,45	-1,45	-1,44	-1,43	-1,43	-1,42	-1,41
0,08	-1,41	-1,40	-1,39	-1,39	-1,38	-1,37	-1,37	-1,36	-1,35	-1,35
0,09	-1,34	-1,33	-1,33	-1,32	-1,32	-1,31	-1,30	-1,30	-1,29	-1,29
0,10	-1,28	-1,28	-1,27	-1,26	-1,26	-1,25	-1,25	-1,24	-1,24	-1,23
0,11	-1,23	-1,22	-1,22	-1,21	-1,21	-1,20	-1,20	-1,19	-1,19	-1,18
0,12	-1,18	-1,17	-1,17	-1,16	-1,16	-1,15	-1,15	-1,14	-1,14	-1,13
0,13	-1,13	-1,12	-1,12	-1,11	-1,11	-1,10	-1,10	-1,09	-1,09	-1,09
0,14	-1,08	-1,08	-1,07	-1,07	-1,06	-1,06	-1,05	-1,05	-1,05	-1,04
0,15	-1,04	-1,03	-1,03	-1,02	-1,02	-1,02	-1,01	-1,01	-1,00	-1,00
0,16	-0,99	-0,99	-0,99	-0,98	-0,98	-0,97	-0,97	-0,97	-0,96	-0,96
0,17	-0,95	-0,95	-0,95	-0,94	-0,94	-0,93	-0,93	-0,93	-0,92	-0,92
0,18	-0,92	-0,91	-0,91	-0,90	-0,90	-0,90	-0,89	-0,89	-0,89	-0,88
0,19	-0,88	-0,87	-0,87	-0,87	-0,86	-0,86	-0,86	-0,85	-0,85	-0,85
0,20	-0,84	-0,84	-0,83	-0,83	-0,83	-0,82	-0,82	-0,82	-0,81	-0,81
0,21	-0,81	-0,80	-0,80	-0,80	-0,79	-0,79	-0,79	-0,78	-0,78	-0,78
0,22	-0,77	-0,77	-0,77	-0,76	-0,76	-0,76	-0,75	-0,75	-0,75	-0,74
0,23	-0,74	-0,74	-0,73	-0,73	-0,73	-0,72	-0,72	-0,72	-0,71	-0,71
0,24	-0,71	-0,70	-0,70	-0,70	-0,69	-0,69	-0,69	-0,68	-0,68	-0,68
0,25	-0,67	-0,67	-0,67	-0,67	-0,66	-0,66	-0,66	-0,65	-0,65	-0,65
0,26	-0,64	-0,64	-0,64	-0,63	-0,63	-0,63	-0,63	-0,62	-0,62	-0,62
0,27	-0,61	-0,61	-0,61	-0,60	-0,60	-0,60	-0,59	-0,59	-0,59	-0,59
0,28	-0,58	-0,58	-0,58	-0,57	-0,57	-0,57	-0,57	-0,56	-0,56	-0,56
0,29	-0,55	-0,55	-0,55	-0,54	-0,54	-0,54	-0,54	-0,53	-0,53	-0,53
0,30	-0,52	-0,52	-0,52	-0,52	-0,51	-0,51	-0,51	-0,50	-0,50	-0,50
0,31	-0,50	-0,49	-0,49	-0,49	-0,48	-0,48	-0,48	-0,48	-0,47	-0,47

Таблица 2 (продолжение)

$x \rightarrow$ ↓	0	1	2	3	4	5	6	7	8	9
0,32	-0,47	-0,46	-0,46	-0,46	-0,46	-0,45	-0,45	-0,45	-0,45	-0,44
0,33	-0,44	-0,44	-0,43	-0,43	-0,43	-0,43	-0,42	-0,42	-0,42	-0,42
0,34	-0,41	-0,41	-0,41	-0,40	-0,40	-0,40	-0,40	-0,39	-0,39	-0,39
0,35	-0,39	-0,38	-0,38	-0,38	-0,37	-0,37	-0,37	-0,37	-0,36	-0,36
0,36	-0,36	-0,36	-0,35	-0,35	-0,35	-0,35	-0,34	-0,34	-0,34	-0,33
0,37	-0,33	-0,33	-0,33	-0,32	-0,32	-0,32	-0,32	-0,31	-0,31	-0,31
0,38	-0,31	-0,30	-0,30	-0,30	-0,30	-0,29	-0,29	-0,29	-0,28	-0,28
0,39	-0,28	-0,28	-0,27	-0,27	-0,27	-0,27	-0,26	-0,26	-0,26	-0,26
0,40	-0,25	-0,25	-0,25	-0,25	-0,24	-0,24	-0,24	-0,24	-0,23	-0,23
0,41	-0,23	-0,23	-0,22	-0,22	-0,22	-0,21	-0,21	-0,21	-0,21	-0,20
0,42	-0,20	-0,20	-0,20	-0,19	-0,19	-0,19	-0,19	-0,18	-0,18	-0,18
0,43	-0,18	-0,17	-0,17	-0,17	-0,17	-0,16	-0,16	-0,16	-0,16	-0,15
0,44	-0,15	-0,15	-0,15	-0,14	-0,14	-0,14	-0,14	-0,13	-0,13	-0,13
0,45	-0,13	-0,12	-0,12	-0,12	-0,12	-0,11	-0,11	-0,11	-0,11	-0,10
0,46	-0,10	-0,10	-0,10	-0,09	-0,09	-0,09	-0,09	-0,08	-0,08	-0,08
0,47	-0,08	-0,07	-0,07	-0,07	-0,07	-0,06	-0,06	-0,06	-0,06	-0,05
0,48	-0,05	-0,05	-0,05	-0,04	-0,04	-0,04	-0,04	-0,03	-0,03	-0,03
0,49	-0,03	-0,02	-0,02	-0,02	-0,02	-0,01	-0,01	-0,01	-0,01	-0,00
0,50	0,00	0,00	0,01	0,01	0,01	0,01	0,02	0,02	0,02	0,02
0,51	0,03	0,03	0,03	0,03	0,04	0,04	0,04	0,04	0,05	0,05
0,52	0,05	0,05	0,06	0,06	0,06	0,06	0,07	0,07	0,07	0,07
0,53	0,08	0,08	0,08	0,08	0,09	0,09	0,09	0,09	0,10	0,10
0,54	0,10	0,10	0,11	0,11	0,11	0,11	0,12	0,12	0,12	0,12
0,55	0,13	0,13	0,13	0,13	0,14	0,14	0,14	0,14	0,15	0,15
0,56	0,15	0,15	0,16	0,16	0,16	0,16	0,17	0,17	0,17	0,17
0,57	0,18	0,18	0,18	0,18	0,19	0,19	0,19	0,19	0,20	0,20
0,58	0,20	0,20	0,21	0,21	0,21	0,21	0,22	0,22	0,22	0,23
0,59	0,23	0,23	0,23	0,24	0,24	0,24	0,24	0,25	0,25	0,25
0,60	0,25	0,26	0,26	0,26	0,26	0,27	0,27	0,27	0,27	0,28
0,61	0,28	0,28	0,28	0,29	0,29	0,29	0,30	0,30	0,30	0,30
0,62	0,31	0,31	0,31	0,31	0,32	0,32	0,32	0,32	0,33	0,33
0,63	0,33	0,33	0,34	0,34	0,34	0,35	0,35	0,35	0,35	0,36
0,64	0,36	0,36	0,36	0,37	0,37	0,37	0,37	0,38	0,38	0,38
0,65	0,39	0,39	0,39	0,39	0,40	0,40	0,40	0,40	0,41	0,41

Таблица 2 (продолжение)

$X \rightarrow$ ↓	0	1	2	3	4	5	6	7	8	9
0,66	0,41	0,42	0,42	0,42	0,42	0,43	0,43	0,43	0,43	0,44
0,67	0,44	0,44	0,45	0,45	0,45	0,45	0,46	0,46	0,46	0,46
0,68	0,47	0,47	0,47	0,48	0,48	0,48	0,48	0,49	0,49	0,49
0,69	0,50	0,50	0,50	0,50	0,51	0,51	0,51	0,52	0,52	0,52
0,70	0,52	0,53	0,53	0,53	0,54	0,54	0,54	0,54	0,55	0,55
0,71	0,55	0,56	0,56	0,56	0,57	0,57	0,57	0,57	0,58	0,58
0,72	0,58	0,59	0,59	0,59	0,59	0,60	0,60	0,60	0,61	0,61
0,73	0,61	0,62	0,62	0,62	0,63	0,63	0,63	0,63	0,64	0,64
0,74	0,64	0,65	0,65	0,65	0,66	0,66	0,66	0,67	0,67	0,67
0,75	0,67	0,68	0,68	0,68	0,69	0,69	0,69	0,70	0,70	0,70
0,76	0,71	0,71	0,71	0,72	0,72	0,72	0,73	0,73	0,73	0,74
0,77	0,74	0,74	0,75	0,75	0,75	0,76	0,76	0,76	0,77	0,77
0,78	0,77	0,78	0,78	0,78	0,79	0,79	0,79	0,80	0,80	0,80
0,79	0,81	0,81	0,81	0,82	0,82	0,82	0,83	0,83	0,83	0,84
0,80	0,84	0,85	0,85	0,85	0,86	0,86	0,86	0,87	0,87	0,87
0,81	0,88	0,88	0,89	0,89	0,89	0,90	0,90	0,90	0,91	0,91
0,82	0,92	0,92	0,92	0,93	0,93	0,93	0,94	0,94	0,95	0,95
0,83	0,95	0,96	0,96	0,97	0,97	0,97	0,98	0,98	0,99	0,99
0,84	0,99	1,00	1,00	1,01	1,01	1,02	1,02	1,02	1,03	1,03
0,85	1,04	1,04	1,05	1,05	1,05	1,06	1,06	1,07	1,07	1,08
0,86	1,08	1,09	1,09	1,09	1,10	1,10	1,11	1,11	1,12	1,12
0,87	1,13	1,13	1,14	1,14	1,15	1,15	1,16	1,16	1,17	1,17
0,88	1,18	1,18	1,19	1,19	1,20	1,20	1,21	1,21	1,22	1,22
0,89	1,23	1,23	1,24	1,24	1,25	1,25	1,26	1,26	1,27	1,28
0,90	1,28	1,29	1,29	1,30	1,30	1,31	1,32	1,32	1,33	1,33
0,91	1,34	1,35	1,35	1,36	1,37	1,37	1,38	1,39	1,39	1,40
0,92	1,41	1,41	1,42	1,43	1,43	1,44	1,45	1,45	1,46	1,47
0,93	1,48	1,48	1,49	1,50	1,51	1,51	1,52	1,53	1,54	1,55
0,94	1,55	1,56	1,57	1,58	1,59	1,60	1,61	1,62	1,63	1,64
0,95	1,64	1,65	1,66	1,67	1,68	1,70	1,71	1,72	1,73	1,74
0,96	1,75	1,76	1,77	1,79	1,80	1,81	1,83	1,84	1,85	1,87
0,97	1,88	1,90	1,91	1,93	1,94	1,96	1,98	2,00	2,01	2,03
0,98	2,05	2,07	2,10	2,12	2,14	2,17	2,20	2,23	2,26	2,29
0,99	2,33	2,37	2,41	2,46	2,51	2,58	2,65	2,75	2,88	3,09

Таблица 3

Доверительные границы g в случае нормального распределения и доверительные границы g^2 в случае распределения χ^2 с одной степенью свободы

Уровень значимости		g	g^2
одно-сторонний	дву-сторонний	нормальное распределение	распределение χ^2
5%	10%	1,64	2,71
2,5%	5%	1,96	3,84
1%	2%	2,33	5,02
0,5%	1%	2,58	6,63
0,1%	0,2%	3,09	9,55
0,05%	0,1%	3,29	10,83

Таблица 4

Критерий Н. В. Смирнова. Точные и асимптотические односторонние границы для верхней грани разности истинной и эмпирической функций распределения

n	Уровень значимости 5%			Уровень значимости 1%		
	точная граница	асимптотическая граница	отношение	точная граница	асимптотическая граница	отношение
5	0,5094	0,5473	1,074	0,6271	0,6786	1,082
8	0,4096	0,4327	1,056	0,5065	0,5365	1,059
10	0,3687	0,3870	1,050	0,4566	0,4799	1,051
20	0,2647	0,2737	1,034	0,3285	0,3393	1,033
40	0,1891	0,1935	1,023	0,2350	0,2399	1,021
50	0,1696	0,1731	1,021	0,2107	0,2146	1,019

При $n > 50$ следует применять асимптотическую границу

$$\bar{\varepsilon}_\beta = \sqrt{\frac{-\ln\beta}{2n}} \quad (\beta - \text{заданный уровень значимости}),$$

для которой истинный коэффициент доверия несколько больше заданной величины $1 - \beta$. Асимптотические границы, указанные в таблице, можно даже уменьшить на $1/(6n)$ и при этом коэффициент доверия по-прежнему не будет превосходить $1 - \beta$. Таким образом, применение асимптотических границ лишь увеличивает надежность критерия.

Таблица 4 заимствована из статьи: Birnbaum Z. W. and Tingey F. H., One-sided confidence contours for probability distribution functions, Ann. Math. Statist., 22 (1959) 595.

Таблица 5

Критерий А. Н. Колмогорова. Точные и асимптотические границы для верхней грани модуля разности истинной и эмпирической функций распределения

n	Уровень значимости 5%			Уровень значимости 1%		
	точная граница	асимптотическая граница	отношение	точная граница	асимптотическая граница	отношение
5	0,5633	0,6074	1,078	0,6685	0,7279	1,089
10	0,4087	0,4295	1,051	0,4864	0,5147	1,058
15	0,3375	0,3507	1,039	0,4042	0,4202	1,040
20	0,2939	0,3037	1,033	0,3524	0,3639	1,033
25	0,2639	0,2716	1,029	0,3165	0,3255	1,028
30	0,2417	0,2480	1,026	0,2898	0,2972	1,025
40	0,2101	0,2147	1,022	0,2521	0,2574	1,021
50	0,1884	0,1921	1,019	0,2260	0,2302	1,018
60	0,1723	0,1753	1,018	0,2067	0,2101	1,016
70	0,1597	0,1623	1,016	0,1917	0,1945	1,015
80	0,1496	0,1518	1,015	0,1795	0,1820	1,014
90	0,1412	0,1432	1,014			
100	0,1340	0,1358	1,013			

При $n > 100$ следует применять асимптотические границы

$$\varepsilon_{0,05} = \frac{1,36}{\sqrt{n}} \quad \text{и} \quad \varepsilon_{0,01} = \frac{1,63}{\sqrt{n}},$$

для которых истинные коэффициенты доверия несколько больше заданных величин 0,95 и 0,99 соответственно. Асимптотические границы, указанные в таблице, можно даже уменьшить на $1/(6n)$ и при этом коэффициенты доверия по-прежнему не будут превосходить $1 - \beta$. Таким образом, применение асимптотических границ лишь увеличивает надежность критерия.

Таблица 5 заимствована из статьи: Birnbaum Z. W., Numerical tabulation of the distribution of Kolmogorov's statistic, J. Amer. Statist. Assoc., 47 (1952) 431.

Доверительные границы для χ^2 с f степенями свободы

f	5%	1%	0,1%	f	5%	1%	0,1%
1	3,84	6,63	10,8	26	38,9	45,6	54,1
2	5,99	9,21	13,8	27	40,1	47,0	55,5
3	7,81	11,3	16,3	28	41,3	48,3	56,9
4	9,49	13,3	18,5	29	42,6	49,6	58,3
5	11,1	15,1	20,5	30	43,8	50,9	59,7
6	12,6	16,8	22,5	31	45,0	52,2	61,1
7	14,1	18,5	24,3	32	46,2	53,5	62,5
8	15,5	20,1	26,1	33	47,4	54,8	63,9
9	16,9	21,7	27,9	34	48,6	56,1	65,2
10	18,3	23,2	29,6	35	49,8	57,3	66,6
11	19,7	24,7	31,3	36	51,0	58,6	68,0
12	21,0	26,2	32,9	37	52,2	59,9	69,3
13	22,4	27,7	34,5	38	53,4	61,2	70,7
14	23,7	29,1	36,1	39	54,6	62,4	72,1
15	25,0	30,6	37,7	40	55,8	63,7	73,4
16	26,3	32,0	39,3	41	56,9	65,0	74,7
17	27,6	33,4	40,8	42	58,1	66,2	76,1
18	28,9	34,8	42,3	43	59,3	67,5	77,4
19	30,1	36,2	43,8	44	60,5	68,7	78,7
20	31,4	37,6	45,3	45	61,7	70,0	80,1
21	32,7	38,9	46,8	46	62,8	71,2	81,4
22	33,9	40,3	48,3	47	64,0	72,4	82,7
23	35,2	41,6	49,7	48	65,2	73,7	84,0
24	36,4	43,0	51,2	49	66,3	74,9	85,4
25	37,7	44,3	52,6	50	67,5	76,2	86,7

Таблицы 6 и 7 заимствованы из сборника: Hald A., *Statistical Tables and Formulas*, John Wiley and Sons, New York, 1952. Три значения в последнем столбце таблицы 7 исправлены согласно данным, указанным в таблице 12 сборника: Pearson E. S. and Hartley H. O., *Biometrika, Tables for Statisticians*, vol. 1.

Таблица 6 (продолжение)

Доверительные границы для χ^2

f	5 %	1 %	0,1%	f	5 %	1 %	0,1%
51	68,7	77,4	88,0	76	97,4	107,6	119,9
52	69,8	78,6	89,3	77	98,5	108,8	121,1
53	71,0	79,8	90,6	78	99,6	110,0	122,3
54	72,2	81,1	91,9	79	100,7	111,1	123,6
55	73,3	82,3	93,2	80	101,9	112,3	124,8
56	74,5	83,5	94,5	81	103,0	113,5	126,1
57	75,6	84,7	95,8	82	104,1	114,7	127,3
58	76,8	86,0	97,0	83	105,3	115,9	128,6
59	77,9	87,2	98,3	84	106,4	117,1	129,8
60	79,1	88,4	99,6	85	107,5	118,2	131,0
61	80,2	89,6	100,9	86	108,6	119,4	132,3
62	81,4	90,8	102,2	87	109,8	120,6	133,5
63	82,5	92,0	103,4	88	110,9	121,8	134,7
64	83,7	93,2	104,7	89	112,0	122,9	136,0
65	84,8	94,4	106,0	90	113,1	124,1	137,2
66	86,0	95,6	107,3	91	114,3	125,3	138,4
67	87,1	96,8	108,5	92	115,4	126,5	139,7
68	88,3	98,0	109,8	93	116,5	127,6	140,9
69	89,4	99,2	111,1	94	117,6	128,8	142,1
70	90,5	100,4	112,3	95	118,8	130,0	143,3
71	91,7	101,6	113,6	96	119,9	131,1	144,6
72	92,8	102,8	114,8	97	121,0	132,3	145,8
73	93,9	104,0	116,1	98	122,1	133,5	147,0
74	95,1	105,2	117,3	99	123,2	134,6	148,2
75	96,2	106,4	118,6	100	124,3	135,8	149,4

При $f > 100$ доверительную границу для χ^2 с уровнем значимости β следует вычислять по формуле

$$\chi_{\beta}^2 = \frac{1}{2} [\sqrt{2f-1} - \Psi(1-\beta)]^2.$$

Таблица 7

Критерий Стьюдента. Доверительные границы для t с f степенями свободы

f ↓	Двусторонние границы				f ↓	Двусторонние границы			
	5%	2%	1%	0,1%		5%	2%	1%	0,1%
1	12,71	31,82	63,66	636,6	20	2,086	2,528	2,845	3,850
2	4,303	6,965	9,925	31,60	21	2,080	2,518	2,831	3,819
3	3,182	4,541	5,841	12,92	22	2,074	2,508	2,819	3,792
4	2,776	3,747	4,604	8,610	23	2,069	2,500	2,807	3,767
5	2,571	3,365	4,032	6,869	24	2,064	2,492	2,797	3,745
6	2,447	3,143	3,707	5,959	25	2,060	2,485	2,787	3,725
7	2,365	2,998	3,499	5,408	26	2,056	2,479	2,779	3,707
8	2,306	2,896	3,355	5,041	27	2,052	2,473	2,771	3,690
9	2,262	2,821	3,250	4,781	28	2,048	2,467	2,763	3,674
10	2,228	2,764	3,169	4,587	29	2,045	2,462	2,756	3,659
11	2,201	2,718	3,106	4,437	30	2,042	2,457	2,750	3,646
12	2,179	2,681	3,055	4,318	40	2,021	2,423	2,704	3,551
13	2,160	2,650	3,012	4,221	50	2,009	2,403	2,678	3,495
14	2,145	2,624	2,977	4,140	60	2,000	2,390	2,660	3,460
15	2,131	2,602	2,947	4,073	80	1,990	2,374	2,639	3,415
16	2,120	2,583	2,921	4,015	100	1,984	2,365	2,626	3,389
17	2,110	2,567	2,898	3,965	200	1,972	2,345	2,601	3,339
18	2,101	2,552	2,878	3,922	500	1,965	2,334	2,586	3,310
19	2,093	2,539	2,861	3,883	∞	1,960	2,326	2,576	3,291
f ↑	2,5%	1%	0,5%	0,05%	f ↑	2,5%	1%	0,5%	0,05%
	Односторонние границы					Односторонние границы			

Линейная интерполяция в таблице 7 позволяет получать лишь два верных десятичных знака¹.

¹ Если вместо f в качестве нового аргумента выбрать $\varphi = 1/f$, то при линейной интерполяции по φ все три десятичных знака будут получаться верными. Например, в последнем столбце таблицы 7 находим, что при $f = 50$ $t = 3,495$ и при $f = 100$ $t = 3,389$, поэтому $\varphi_0 = 0,01$, $\varphi_1 = 0,02$ и $t(\varphi_0) = 3,389$, $t(\varphi_1) = 3,495$. Согласно формуле линейной интерполяции

$$t(\varphi) = t(\varphi_0) + u [t(\varphi_1) - t(\varphi_0)], \text{ где } u = \frac{\varphi - \varphi_0}{\varphi_1 - \varphi_0},$$

т. е. в нашем случае $t(\varphi) = 3,389 + 0,106 \cdot u$. Пусть $f = 60$ и 80 , тогда $u = 2/3$ и $1/4$, поэтому $t = 3,460$ и $3,415$. Вычисленные значения t совпадают с табличными. — Прим. перев.

Таблица 8А

Доверительные границы для $F = s_1^2/s_2^2$ с уровнем значимости 5%; f_1 — число степеней свободы числителя, f_2 — число степеней свободы знаменателя

f_2 ↓	f_1 (число степеней свободы числителя)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161	200	216	225	230	234	237	239	241	242	243	244	245	245	246
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,22	2,20	2,18
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20	2,17	2,15
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,23	2,20	2,18	2,15	2,13
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,21	2,18	2,15	2,13	2,11
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14	2,11	2,09
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12	2,09	2,07
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13	2,10	2,08	2,06
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,09	2,06	2,04
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,08	2,05	2,03
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06	2,04	2,01
32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,10	2,07	2,04	2,01	1,99
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,08	2,05	2,02	1,99	1,97
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,07	2,03	2,00	1,98	1,95
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02	1,99	1,96	1,94
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,92
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	2,03	1,99	1,96	1,93	1,91
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	2,01	1,98	1,95	1,92	1,90
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,15	2,09	2,04	2,00	1,97	1,94	1,91	1,89
48	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	2,03	1,99	1,96	1,93	1,90	1,88
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92	1,89	1,87
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89	1,86	1,84
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86	1,84	1,81
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84	1,82	1,79
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,90	1,86	1,83	1,80	1,78
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82	1,79	1,77
125	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,96	1,91	1,87	1,83	1,80	1,77	1,75
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,82	1,79	1,76	1,73
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80	1,77	1,74	1,72
300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86	1,82	1,78	1,75	1,72	1,70
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,77	1,74	1,71	1,69
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,80	1,76	1,73	1,70	1,68

Если $f_2 > 1000$, то следует воспользоваться границей, соответствующей $f_2 = 1000$.

Таблица 8А (продолжение)

Доверительные границы для $F = s_1^2/s_2^2$ с уровнем значимости 5%;
 f_1 — число степеней свободы числителя, f_2 — число степеней свободы
 знаменателя

f_2 ↓	f_1 (число степеней свободы числителя)														
	16	17	18	19	20	22	24	26	28	30	40	50	60	80	100
1	246	247	247	248	248	249	249	249	250	250	251	252	252	252	253
2	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5
3	8,69	8,68	8,67	8,67	8,66	8,65	8,64	8,63	8,62	8,62	8,59	8,58	8,57	8,56	8,55
4	5,84	5,83	5,82	5,81	5,80	5,79	5,77	5,76	5,75	5,75	5,72	5,70	5,69	5,67	5,66
5	4,60	4,59	4,58	4,57	4,56	4,54	4,53	4,52	4,50	4,50	4,46	4,44	4,43	4,41	4,41
6	3,92	3,91	3,90	3,88	3,87	3,86	3,84	3,83	3,82	3,81	3,77	3,75	3,74	3,72	3,71
7	3,49	3,48	3,47	3,46	3,44	3,43	3,41	3,40	3,39	3,38	3,34	3,32	3,30	3,29	3,27
8	3,20	3,19	3,17	3,16	3,15	3,13	3,12	3,10	3,09	3,08	3,04	3,02	3,01	2,99	2,97
9	2,99	2,97	2,96	2,95	2,94	2,92	2,90	2,89	2,87	2,86	2,83	2,80	2,79	2,77	2,76
10	2,83	2,81	2,80	2,78	2,77	2,75	2,74	2,72	2,71	2,70	2,66	2,64	2,62	2,60	2,59
11	2,70	2,69	2,67	2,66	2,65	2,63	2,61	2,59	2,58	2,57	2,53	2,51	2,49	2,47	2,46
12	2,60	2,58	2,57	2,56	2,54	2,52	2,51	2,49	2,48	2,47	2,43	2,40	2,38	2,36	2,35
13	2,51	2,50	2,48	2,47	2,46	2,44	2,42	2,41	2,39	2,38	2,34	2,31	2,30	2,27	2,26
14	2,44	2,43	2,41	2,40	2,39	2,37	2,35	2,33	2,32	2,31	2,27	2,24	2,22	2,20	2,19
15	2,38	2,37	2,35	2,34	2,33	2,31	2,29	2,27	2,26	2,25	2,20	2,18	2,16	2,14	2,12
16	2,33	2,32	2,30	2,29	2,28	2,25	2,24	2,22	2,21	2,19	2,15	2,12	2,11	2,08	2,07
17	2,29	2,27	2,26	2,24	2,23	2,21	2,19	2,17	2,16	2,15	2,10	2,08	2,06	2,03	2,02
18	2,25	2,23	2,22	2,20	2,19	2,17	2,15	2,13	2,12	2,11	2,06	2,04	2,02	1,99	1,98
19	2,21	2,20	2,18	2,17	2,16	2,13	2,11	2,10	2,08	2,07	2,03	2,00	1,98	1,96	1,94
20	2,18	2,17	2,15	2,14	2,12	2,10	2,08	2,07	2,05	2,04	1,99	1,97	1,95	1,92	1,91
21	2,16	2,14	2,12	2,11	2,10	2,07	2,05	2,04	2,02	2,01	1,96	1,94	1,92	1,86	1,88
22	2,13	2,11	2,10	2,08	2,07	2,05	2,03	2,01	2,00	1,98	1,94	1,91	1,89	1,86	1,85
23	2,11	2,09	2,07	2,06	2,05	2,02	2,00	1,99	1,97	1,96	1,91	1,88	1,86	1,84	1,82
24	2,09	2,07	2,05	2,04	2,03	2,00	1,98	1,97	1,95	1,94	1,89	1,86	1,84	1,82	1,80
25	2,07	2,05	2,04	2,02	2,01	1,98	1,96	1,95	1,93	1,92	1,87	1,84	1,82	1,80	1,78
26	2,05	2,03	2,02	2,00	1,99	1,97	1,95	1,93	1,91	1,90	1,85	1,82	1,80	1,78	1,76
27	2,04	2,02	2,00	1,99	1,97	1,95	1,93	1,91	1,90	1,88	1,84	1,81	1,79	1,76	1,74
28	2,02	2,00	1,99	1,97	1,96	1,93	1,91	1,90	1,88	1,87	1,82	1,79	1,77	1,74	1,73
29	2,01	1,99	1,97	1,96	1,94	1,92	1,90	1,88	1,87	1,85	1,81	1,77	1,74	1,71	1,70
30	1,99	1,98	1,96	1,95	1,93	1,91	1,89	1,87	1,85	1,84	1,79	1,76	1,74	1,71	1,70
32	1,97	1,95	1,94	1,92	1,91	1,88	1,86	1,85	1,83	1,82	1,77	1,74	1,71	1,69	1,67
34	1,95	1,93	1,92	1,90	1,89	1,86	1,84	1,82	1,80	1,80	1,75	1,71	1,69	1,66	1,65
36	1,93	1,92	1,90	1,88	1,87	1,85	1,82	1,81	1,79	1,78	1,73	1,69	1,67	1,64	1,62
38	1,92	1,90	1,88	1,87	1,85	1,83	1,81	1,79	1,77	1,76	1,71	1,68	1,65	1,62	1,61
40	1,90	1,89	1,87	1,85	1,84	1,81	1,79	1,77	1,76	1,74	1,69	1,66	1,64	1,61	1,59
42	1,89	1,87	1,86	1,84	1,83	1,80	1,78	1,76	1,74	1,73	1,68	1,65	1,62	1,59	1,57
44	1,88	1,86	1,84	1,83	1,81	1,79	1,77	1,75	1,73	1,72	1,67	1,63	1,61	1,58	1,56
46	1,87	1,85	1,83	1,82	1,80	1,78	1,76	1,74	1,72	1,71	1,65	1,62	1,60	1,57	1,55
48	1,86	1,84	1,82	1,81	1,79	1,77	1,75	1,73	1,71	1,70	1,64	1,61	1,59	1,56	1,54
50	1,85	1,83	1,81	1,80	1,78	1,76	1,74	1,72	1,70	1,69	1,63	1,60	1,58	1,54	1,52
60	1,82	1,80	1,78	1,76	1,75	1,72	1,70	1,68	1,66	1,65	1,59	1,56	1,53	1,50	1,48
70	1,79	1,77	1,75	1,74	1,72	1,70	1,67	1,65	1,64	1,62	1,57	1,53	1,50	1,47	1,45
80	1,77	1,75	1,73	1,72	1,70	1,68	1,65	1,63	1,62	1,60	1,54	1,51	1,48	1,45	1,43
90	1,76	1,74	1,72	1,70	1,69	1,66	1,64	1,62	1,60	1,59	1,53	1,49	1,46	1,43	1,41
100	1,75	1,73	1,71	1,69	1,68	1,65	1,63	1,61	1,59	1,57	1,52	1,48	1,45	1,41	1,39
125	1,72	1,70	1,69	1,67	1,65	1,63	1,60	1,58	1,57	1,55	1,49	1,45	1,42	1,39	1,36
150	1,71	1,69	1,67	1,66	1,64	1,61	1,59	1,57	1,55	1,53	1,48	1,44	1,41	1,37	1,34
200	1,69	1,67	1,66	1,64	1,62	1,60	1,57	1,55	1,53	1,52	1,46	1,41	1,39	1,35	1,32
300	1,68	1,66	1,64	1,62	1,61	1,58	1,55	1,53	1,51	1,50	1,43	1,39	1,36	1,32	1,30
500	1,66	1,64	1,62	1,61	1,59	1,56	1,54	1,52	1,50	1,48	1,42	1,38	1,34	1,30	1,28
1000	1,65	1,63	1,61	1,60	1,58	1,55	1,53	1,51	1,49	1,47	1,41	1,36	1,33	1,29	1,26

Если $f_2 > 1000$, то следует воспользоваться границей, соответствующей $f_2 = 1000$.

Таблица 8Б

Доверительные границы для $F = s_1^2/s_2^2$ с уровнем значимости 1%;
 f_1 — число степеней свободы числителя, f_2 — число степеней свободы
знаменателя

f_2 ↓	f_1 (число степеней свободы числителя)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,4
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	27,0	27,0	26,9	26,9
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,4	14,4	14,3	14,2	14,2
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,96	9,89	9,82	9,77	9,72
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66	7,60	7,56
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41	6,36	6,31
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67	5,61	5,56	5,52
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05	5,00	4,96
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71	4,65	4,60	4,56
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,46	4,40	4,34	4,29	4,25
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,10	4,05	4,01
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,91	3,86	3,82
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75	3,70	3,66
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61	3,56	3,52
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55	3,50	3,45	3,41
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46	3,40	3,35	3,31
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37	3,32	3,27	3,23
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,24	3,19	3,15
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23	3,18	3,13	3,09
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17	3,12	3,07	3,03
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12	3,07	3,02	2,98
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07	3,02	2,97	2,93
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03	2,98	2,93	2,89
25	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,22	3,13	3,06	2,99	2,94	2,89	2,85
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96	2,90	2,86	2,82
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,99	2,93	2,87	2,82	2,78
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,90	2,84	2,79	2,75
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,93	2,87	2,81	2,77	2,73
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84	2,79	2,74	2,70
32	7,50	5,34	4,46	3,97	3,65	3,43	3,26	3,13	3,02	2,93	2,86	2,80	2,74	2,70	2,66
34	7,44	5,29	4,42	3,93	3,61	3,39	3,22	3,09	2,98	2,89	2,82	2,76	2,70	2,66	2,62
36	7,40	5,25	4,38	3,89	3,57	3,35	3,18	3,05	2,95	2,86	2,79	2,72	2,67	2,62	2,58
38	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,92	2,83	2,75	2,69	2,64	2,59	2,55
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66	2,61	2,56	2,52
42	7,28	5,15	4,29	3,80	3,49	3,27	3,10	2,97	2,86	2,78	2,70	2,64	2,59	2,54	2,50
44	7,25	5,12	4,26	3,78	3,47	3,24	3,08	2,95	2,84	2,75	2,68	2,62	2,56	2,52	2,47
46	7,22	5,10	4,24	3,76	3,44	3,22	3,06	2,93	2,82	2,73	2,66	2,60	2,54	2,50	2,45
48	7,19	5,08	4,22	3,74	3,43	3,20	3,04	2,91	2,80	2,72	2,64	2,58	2,53	2,48	2,44
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,79	2,70	2,63	2,56	2,51	2,46	2,42
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	2,66	2,59	2,53	2,47	2,42	2,38
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50	2,44	2,39	2,35
70	7,01	4,92	4,08	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45	2,40	2,35	2,31
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64	2,55	2,48	2,42	2,36	2,31	2,27
90	6,93	4,85	4,01	3,54	3,23	3,01	2,84	2,72	2,61	2,52	2,45	2,39	2,33	2,29	2,24
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37	2,31	2,26	2,22
125	6,84	4,78	3,94	3,47	3,17	2,95	2,79	2,66	2,55	2,47	2,39	2,33	2,28	2,23	2,19
150	6,81	4,75	3,92	3,45	3,14	2,92	2,76	2,63	2,53	2,44	2,37	2,31	2,25	2,20	2,16
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27	2,22	2,17	2,13
300	6,72	4,68	3,85	3,38	3,08	2,86	2,70	2,57	2,47	2,38	2,31	2,24	2,19	2,14	2,10
500	6,69	4,65	3,82	3,36	3,05	2,84	2,68	2,55	2,44	2,36	2,28	2,22	2,17	2,12	2,07
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,27	2,20	2,15	2,10	2,06

Если $f_2 = 1$, то доверительная граница для F равна квадрату двусторонней 1%-границы для t с f_1 степенями свободы (таблица 7).

Доверительные границы для $F = s_1^2/s_2^2$ с уровнем значимости 1%;
 f_1 — число степеней свободы числителя, f_2 — число степеней свободы
 знаменателя

f_2	f_1 (число степеней свободы числителя)														
	16	17	18	19	20	22	24	26	28	30	40	50	60	80	100
2	99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5	99,5	99,5	99,5	99,5	99,5
3	26,8	26,8	26,8	26,7	26,7	26,6	26,6	26,6	26,5	26,5	26,4	26,4	26,3	26,3	26,2
4	14,2	14,1	14,1	14,0	14,0	14,0	13,9	13,9	13,9	13,8	13,7	13,7	13,6	13,6	13,6
5	9,68	9,64	9,61	9,58	9,55	9,51	9,47	9,43	9,40	9,38	9,29	9,24	9,20	9,16	9,13
6	7,52	7,48	7,45	7,42	7,40	7,35	7,31	7,28	7,25	7,23	7,14	7,09	7,06	7,01	6,99
7	6,27	6,24	6,21	6,18	6,16	6,11	6,07	6,04	6,02	5,99	5,91	5,86	5,82	5,78	5,75
8	5,48	5,44	5,41	5,38	5,36	5,32	5,28	5,25	5,22	5,20	5,12	5,07	5,03	4,99	4,96
9	4,92	4,89	4,86	4,83	4,81	4,77	4,73	4,70	4,67	4,65	4,57	4,52	4,48	4,44	4,42
10	4,52	4,49	4,46	4,43	4,41	4,36	4,33	4,30	4,27	4,25	4,17	4,12	4,08	4,04	4,01
11	4,21	4,18	4,15	4,12	4,10	4,06	4,02	3,99	3,96	3,94	3,86	3,81	3,78	3,73	3,71
12	3,97	3,94	3,91	3,88	3,86	3,82	3,78	3,75	3,72	3,70	3,62	3,57	3,54	3,49	3,47
13	3,78	3,75	3,72	3,69	3,66	3,62	3,59	3,56	3,53	3,51	3,43	3,38	3,34	3,30	3,27
14	3,62	3,59	3,56	3,53	3,51	3,46	3,43	3,40	3,37	3,35	3,27	3,22	3,18	3,14	3,11
15	3,49	3,45	3,42	3,40	3,37	3,33	3,29	3,26	3,24	3,21	3,13	3,08	3,05	3,00	2,98
16	3,37	3,34	3,31	3,28	3,26	3,22	3,18	3,15	3,12	3,10	3,02	2,97	2,93	2,89	2,86
17	3,27	3,24	3,21	3,18	3,16	3,12	3,08	3,05	3,03	3,00	2,92	2,87	2,83	2,79	2,76
18	3,19	3,16	3,13	3,10	3,08	3,03	3,00	2,97	2,94	2,92	2,84	2,78	2,75	2,70	2,68
19	3,12	3,08	3,05	3,03	3,00	2,96	2,92	2,89	2,87	2,84	2,76	2,71	2,67	2,63	2,60
20	3,05	3,02	2,99	2,96	2,94	2,90	2,86	2,83	2,80	2,78	2,69	2,64	2,61	2,56	2,54
21	2,99	2,96	2,93	2,90	2,88	2,84	2,80	2,77	2,74	2,72	2,64	2,58	2,55	2,50	2,48
22	2,94	2,91	2,88	2,85	2,83	2,78	2,75	2,72	2,69	2,67	2,58	2,53	2,50	2,45	2,42
23	2,89	2,86	2,83	2,80	2,78	2,74	2,70	2,67	2,64	2,62	2,54	2,48	2,45	2,40	2,37
24	2,85	2,82	2,79	2,76	2,74	2,70	2,66	2,63	2,60	2,58	2,49	2,44	2,40	2,36	2,33
25	2,81	2,78	2,75	2,72	2,70	2,66	2,62	2,59	2,56	2,54	2,45	2,40	2,36	2,32	2,29
26	2,78	2,74	2,72	2,69	2,66	2,62	2,58	2,55	2,53	2,50	2,42	2,36	2,33	2,28	2,25
27	2,75	2,71	2,68	2,66	2,63	2,59	2,55	2,52	2,49	2,47	2,38	2,33	2,29	2,25	2,22
28	2,72	2,68	2,65	2,63	2,60	2,56	2,52	2,49	2,46	2,44	2,35	2,30	2,26	2,22	2,19
29	2,69	2,66	2,63	2,60	2,57	2,53	2,49	2,46	2,44	2,41	2,33	2,27	2,23	2,19	2,16
30	2,66	2,63	2,60	2,57	2,55	2,51	2,47	2,44	2,41	2,39	2,30	2,25	2,21	2,16	2,13
32	2,62	2,58	2,55	2,53	2,50	2,46	2,42	2,39	2,36	2,34	2,25	2,20	2,16	2,11	2,08
34	2,58	2,55	2,51	2,49	2,46	2,42	2,38	2,35	2,32	2,30	2,21	2,16	2,12	2,07	2,04
36	2,54	2,51	2,48	2,45	2,43	2,38	2,35	2,32	2,29	2,26	2,17	2,12	2,08	2,03	2,00
38	2,51	2,48	2,45	2,42	2,40	2,35	2,32	2,28	2,26	2,23	2,14	2,09	2,05	2,00	1,97
40	2,48	2,45	2,42	2,39	2,37	2,33	2,29	2,26	2,23	2,20	2,11	2,06	2,02	1,97	1,94
42	2,46	2,43	2,40	2,37	2,34	2,30	2,26	2,23	2,20	2,18	2,09	2,03	1,99	1,94	1,91
44	2,44	2,40	2,37	2,35	2,32	2,28	2,24	2,21	2,18	2,15	2,06	2,01	1,97	1,92	1,89
46	2,42	2,38	2,35	2,33	2,30	2,26	2,22	2,19	2,16	2,13	2,04	1,99	1,95	1,90	1,86
48	2,40	2,37	2,33	2,31	2,28	2,24	2,20	2,17	2,14	2,12	2,02	1,97	1,93	1,88	1,84
50	2,38	2,35	2,32	2,29	2,27	2,22	2,18	2,15	2,12	2,10	2,01	1,95	1,91	1,86	1,82
55	2,34	2,31	2,28	2,25	2,23	2,18	2,15	2,11	2,08	2,06	1,97	1,91	1,87	1,81	1,78
60	2,31	2,28	2,25	2,22	2,20	2,15	2,12	2,08	2,05	2,03	1,94	1,88	1,84	1,78	1,75
70	2,27	2,23	2,20	2,18	2,15	2,11	2,07	2,03	2,01	1,98	1,89	1,83	1,78	1,73	1,70
80	2,23	2,20	2,17	2,14	2,12	2,07	2,03	2,00	1,97	1,94	1,85	1,79	1,75	1,69	1,66
90	2,21	2,17	2,14	2,11	2,09	2,04	2,00	1,97	1,94	1,92	1,82	1,76	1,72	1,66	1,62
100	2,19	2,15	2,12	2,09	2,07	2,02	1,98	1,94	1,92	1,89	1,80	1,73	1,69	1,63	1,60
125	2,15	2,11	2,08	2,05	2,03	1,98	1,94	1,91	1,88	1,85	1,76	1,69	1,65	1,59	1,55
150	2,12	2,09	2,06	2,03	2,00	1,96	1,92	1,88	1,85	1,83	1,73	1,66	1,62	1,56	1,52
200	2,09	2,06	2,02	2,00	1,97	1,93	1,89	1,85	1,82	1,79	1,69	1,63	1,58	1,52	1,48
300	2,06	2,03	1,99	1,97	1,94	1,89	1,85	1,82	1,79	1,76	1,66	1,59	1,55	1,48	1,44
500	2,04	2,00	1,97	1,94	1,92	1,87	1,83	1,79	1,76	1,74	1,63	1,56	1,52	1,45	1,41
1000	2,02	1,98	1,95	1,92	1,90	1,85	1,81	1,77	1,74	1,72	1,61	1,54	1,50	1,43	1,38

Если $f_2 > 1000$, то следует воспользоваться границей, соответствующей $f_2 = 1000$.

Таблица 8В

Доверительные границы для $F = s_1^2/s_2^2$ с уровнем значимости 0,1%;
 f_1 — число степеней свободы числителя, f_2 — число степеней свободы
 знаменателя

f_2 ↓	f_1 (число степеней свободы числителя)														
	1	2	3	4	5	6	7	8	9	10	15	20	30	50	100
2	998	999	999	999	999	999	999	999	999	999	999	999	999	999	999
3	168	148	141	137	135	133	132	131	130	129	127	126	125	125	124
4	74,1	61,2	56,2	53,4	51,7	50,5	49,7	49,0	48,5	48,0	46,8	46,1	45,4	44,9	44,5
5	47,0	36,6	33,2	31,1	29,8	28,8	28,2	27,6	27,2	26,9	25,9	25,4	24,9	24,4	24,1
6	35,5	27,0	23,7	21,9	20,8	20,0	19,5	19,0	18,7	18,4	17,6	17,1	16,7	16,3	16,0
7	29,2	21,7	18,8	17,2	16,2	15,5	15,0	14,6	14,3	14,1	13,3	12,9	12,5	12,2	11,9
8	25,4	18,5	15,8	14,4	13,5	12,9	12,4	12,0	11,8	11,5	10,8	10,5	10,1	9,80	9,57
9	22,9	16,4	13,9	12,6	11,7	11,1	10,7	10,4	10,1	9,89	9,24	8,90	8,55	8,26	8,04
10	21,0	14,9	12,6	11,3	10,5	9,92	9,52	9,20	8,96	8,75	8,13	7,80	7,47	7,19	6,98
11	19,7	13,8	11,6	10,4	9,58	9,05	8,66	8,35	8,12	7,92	7,32	7,01	6,68	6,41	6,21
12	18,6	13,0	10,8	9,63	8,89	8,38	8,00	7,71	7,48	7,29	6,71	6,40	6,09	5,83	5,63
13	17,8	12,3	10,2	9,07	8,35	7,86	7,49	7,21	6,98	6,80	6,23	5,93	5,62	5,37	5,17
14	17,1	11,8	9,73	8,62	7,92	7,43	7,08	6,80	6,58	6,40	5,85	5,56	5,25	5,00	4,80
15	16,6	11,3	9,34	8,25	7,57	7,09	6,74	6,47	6,26	6,08	5,53	5,25	4,95	4,70	4,51
16	16,1	11,0	9,00	7,94	7,27	6,81	6,46	6,19	5,98	5,81	5,27	4,99	4,70	4,45	4,26
17	15,7	10,7	8,73	7,68	7,02	6,56	6,22	5,96	5,75	5,58	5,05	4,78	4,48	4,24	4,05
18	15,4	10,4	8,49	7,46	6,81	6,35	6,02	5,76	5,56	5,39	4,87	4,59	4,30	4,06	3,87
19	15,1	10,2	8,28	7,26	6,61	6,18	5,84	5,59	5,39	5,22	4,70	4,43	4,14	3,90	3,71
20	14,8	9,95	8,10	7,10	6,46	6,02	5,69	5,44	5,24	5,08	4,56	4,29	4,01	3,77	3,58
22	14,4	9,61	7,80	6,81	6,19	5,76	5,44	5,19	4,99	4,83	4,32	4,06	3,79	3,53	3,34
24	14,0	9,34	7,55	6,59	5,98	5,55	5,23	4,99	4,80	4,64	4,14	3,87	3,59	3,35	3,16
26	13,7	9,12	7,36	6,41	5,80	5,38	5,07	4,83	4,64	4,48	3,99	3,72	3,45	3,20	3,01
28	13,5	8,93	7,19	6,25	5,66	5,24	4,93	4,69	4,50	4,35	3,86	3,60	3,32	3,08	2,89
30	13,3	8,77	7,05	6,12	5,53	5,12	4,82	4,58	4,39	4,24	3,75	3,49	3,22	2,98	2,79
40	12,6	8,25	6,60	5,70	5,13	4,73	4,43	4,21	4,02	3,87	3,40	3,15	2,87	2,64	2,44
50	12,2	7,95	6,34	5,46	4,90	4,51	4,22	4,00	3,82	3,67	3,20	2,95	2,68	2,44	2,24
60	12,0	7,76	6,17	5,31	4,76	4,37	4,09	3,87	3,69	3,54	3,08	2,83	2,56	2,31	2,11
80	11,7	7,54	5,97	5,13	4,58	4,21	3,92	3,70	3,53	3,39	2,93	2,68	2,40	2,16	1,95
100	11,5	7,41	5,85	5,01	4,48	4,11	3,83	3,61	3,44	3,30	2,84	2,59	2,32	2,07	1,87
200	11,2	7,15	5,64	4,81	4,29	3,92	3,65	3,43	3,26	3,12	2,67	2,42	2,15	1,90	1,68
500	11,0	7,01	5,51	4,69	4,18	3,82	3,54	3,33	3,16	3,02	2,58	2,33	2,05	1,80	1,57
∞	10,8	6,91	5,42	4,62	4,10	3,74	3,47	3,27	3,10	2,96	2,51	2,27	1,99	1,73	1,49

Приближенные доверительные границы, вычисленные с помощью линейной интерполяции табличных значений, получают несколько грубыми¹. Граница при $f_2 = 1000$ приблизительно равна арифметическому среднему границ при $f_2 = 500$ и ∞ .

Таблица 8А, Б и В заимствованы из сборника: Hald A., Statistical Tables and Formulas, John Wiley and Sons, New York, 1952.

¹ Если вместо f_1 и f_2 в качестве новых аргументов выбрать $\varphi_1 = 1/f_1$ и $\varphi_2 = 1/f_2$, то результаты линейной интерполяции по φ_1 и φ_2 будут практически иметь такую же погрешность, какую имеют табличные значения. — Прим. перев.

Границы критической области для критерия знаков

n	Односторонние границы						n	Односторонние границы					
	2,5 %		1 %		0,5 %			2,5 %		1 %		0,5 %	
5	0	5	0	5	0	5	45	16	29	15	30	14	31
6	1	5	0	6	0	6	46	16	30	15	31	14	32
7	1	6	1	6	0	7	47	17	30	16	31	15	32
8	1	7	1	7	1	7	48	17	31	16	32	15	33
9	2	7	1	8	1	8	49	18	31	16	33	16	33
10	2	8	1	9	1	9	50	18	32	17	33	16	34
11	2	9	2	9	1	10	51	19	32	17	34	16	35
12	3	9	2	10	2	10	52	19	33	18	34	17	35
13	3	10	2	11	2	11	53	19	34	18	35	17	36
14	3	11	3	11	2	12	54	20	34	19	35	18	36
15	4	11	3	12	3	12	55	20	35	19	36	18	37
16	4	12	3	13	3	13	56	21	35	19	37	18	38
17	5	12	4	13	3	14	57	21	36	20	37	19	38
18	5	13	4	14	4	14	58	22	36	20	38	19	39
19	5	14	5	14	4	15	59	22	37	21	38	20	39
20	6	14	5	15	4	16	60	22	38	21	39	20	40
21	6	15	5	16	5	16	61	23	38	21	40	21	40
22	6	16	6	16	5	17	62	23	39	22	40	21	41
23	7	16	6	17	5	18	63	24	39	22	41	21	42
24	7	17	6	18	6	18	64	24	40	23	41	22	42
25	8	17	7	18	6	19	65	25	40	23	42	22	43
26	8	18	7	19	7	19	66	25	41	24	42	23	43
27	8	19	8	19	7	20	67	26	41	24	43	23	44
28	9	19	8	20	7	21	68	26	42	24	44	23	45
29	9	20	8	21	8	21	69	26	43	25	44	24	45
30	10	20	9	21	8	22	70	27	43	25	45	24	46
31	10	21	9	22	8	23	71	27	44	26	45	25	46
32	10	22	9	23	9	23	72	28	44	26	46	25	47
33	11	22	10	23	9	24	73	28	45	27	46	26	47
34	11	23	10	24	10	24	74	29	45	27	47	26	48
35	12	23	11	24	10	25	75	29	46	27	48	26	49
36	12	24	11	25	10	26	76	29	47	28	48	27	49
37	13	24	11	26	11	26	77	30	47	28	49	27	50
38	13	25	12	26	11	27	78	30	48	29	49	28	50
39	13	26	12	27	12	27	79	31	48	29	50	28	51
40	14	26	13	27	12	28	80	31	49	30	50	29	51
41	14	27	13	28	12	29	81	32	49	30	51	29	52
42	15	27	14	28	13	29	82	32	50	31	51	29	53
43	15	28	14	29	13	30	83	33	50	31	52	30	53
44	16	28	14	30	14	30	84	33	51	31	53	30	54
n	5 %		2 %		1 %		n	5 %		2 %		1 %	
	Двусторонние границы							Двусторонние границы					

Таблица 9 (продолжение)

n	Односторонние границы						n	Односторонние границы					
	2,5%		1%		0,5%			2,5%		1%		0,5%	
85	33	52	32	53	31	54	93	37	56	35	58	34	59
86	34	52	32	54	31	55	94	38	56	36	58	35	59
87	34	53	33	54	32	55	95	38	57	36	59	35	60
88	35	53	33	55	32	56	96	38	58	37	59	35	61
89	35	54	34	55	32	57	97	39	58	37	60	36	61
90	36	54	34	56	33	57	98	39	59	38	60	36	62
91	36	55	34	57	33	58	99	40	59	38	61	37	62
92	37	55	35	57	34	58	100	40	60	38	62	37	63
n	Двусторонние границы			n	Двусторонние границы								
	5%	2%	1%		5%	2%	1%						

Если наблюдаемое количество знаков (+) находится вне указанных границ, то гипотеза $P(+)=P(-)$ отвергается.

Таблица 10

Распределение статистики Вилкоксона

Пусть имеются две выборки x_1, \dots, x_g и y_1, \dots, y_h объемов g и h , и пусть из всех x_i и y_k составлен общий вариационный ряд. Статистикой Вилкоксона U называют число инверсий в этом вариационном ряду, равное сумме инверсий для каждого x (число инверсий для данного x_i определяется как число тех y_k , которые удовлетворяют условию $y_k < x_i$). В таблице указаны вероятности $p(u)$ событий $U \leq u$, выраженные в процентах и не превосходящие 5%. Эти вероятности вычислены в предположении, что «нулевая гипотеза» верна (согласно нулевой гипотезе, все x_i и y_k независимы и подчиняются одинаковому непрерывному распределению). Если окажется, что $p(U) \leq \beta$, то нулевую гипотезу следует отвергнуть. Истинный уровень значимости одностороннего критерия не превосходит β , а истинный уровень значимости двустороннего критерия не превосходит 2β . При $U > g h/2$ обозначения x и y следует поменять местами.

u	Объемы выборок g и h													
	2; 5	2; 6	2; 7	2; 8	2; 9	2; 10	3; 3	3; 4	3; 5	3; 6	3; 7	3; 8	3; 9	3; 10
0	4,76	3,57	2,78	2,22	1,82	1,52	5,00	2,86	1,79	1,19	0,83	0,61	,045	0,35
1				2,22	1,82	1,52			3,57	2,38	1,67	1,21	0,91	0,70
2				4,44	3,64	3,03				4,76	3,33	2,42	1,82	1,40
3												4,24	3,18	2,45
4													5,00	3,85

Таблица 10 (продолжение)

u ↓	Объемы выборок g и h													
	4; 4	4; 5	4; 6	4; 7	4; 8	4; 9	4; 10	5; 5	5; 6	5; 7	5; 8	5; 9	5; 10	6; 6
0	1,43	0,79	0,48	0,30	0,20	0,14	0,10	0,40	0,22	0,13	0,08	0,05	0,03	0,11
1	2,86	1,59	0,95	0,61	0,40	0,28	0,20	0,79	0,43	0,25	0,16	0,10	0,07	0,22
2		3,17	1,90	1,21	0,81	0,56	0,40	1,59	0,87	0,51	0,31	0,20	0,13	0,43
3			3,33	2,12	1,41	0,98	0,70	2,78	1,52	0,88	0,54	0,35	0,23	0,76
4				3,64	2,42	1,68	1,20	4,76	2,60	1,52	0,93	0,60	0,40	1,30
5					3,64	2,52	1,80		4,11	2,40	1,48	0,95	0,63	2,06
6						3,78	2,70			3,66	2,25	1,45	0,97	3,25
7							3,80				3,26	2,10	1,40	4,65
8											4,66	3,00	2,00	
9												4,15	2,76	
10													3,76	
11													4,96	

u ↓	Объемы выборок g и h													
	6; 7	6; 8	6; 9	6; 10	7; 7	7; 8	7; 9	7; 10	8; 8	8; 9	8; 10	9; 9	9; 10	10; 10
0	0,06	0,03	0,02	0,01	0,03	0,02	0,01	0,01	0,01	0,00				
1	0,12	0,07	0,04	0,02	0,06	0,03	0,02	0,01	0,02	0,01	0,00	0,00		
2	0,23	0,13	0,08	0,05	0,12	0,06	0,03	0,02	0,03	0,02	0,01	0,01	0,00	
3	0,41	0,23	0,14	0,09	0,20	0,11	0,06	0,04	0,05	0,03	0,02	0,01	0,01	0,00
4	0,70	0,40	0,24	0,15	0,35	0,19	0,10	0,06	0,09	0,05	0,03	0,02	0,01	0,01
5	1,11	0,63	0,38	0,24	0,55	0,30	0,17	0,10	0,15	0,08	0,04	0,04	0,02	0,01
6	1,75	1,00	0,60	0,37	0,87	0,47	0,26	0,15	0,23	0,12	0,07	0,06	0,03	0,02
7	2,56	1,47	0,88	0,55	1,31	0,70	0,39	0,23	0,35	0,19	0,10	0,09	0,05	0,02
8	3,67	2,13	1,28	0,80	1,89	1,03	0,58	0,34	0,52	0,28	0,15	0,14	0,07	0,04
9		2,96	1,80	1,12	2,65	1,45	0,82	0,48	0,74	0,39	0,22	0,20	0,10	0,05
10		4,06	2,48	1,56	3,64	2,00	1,15	0,68	1,03	0,56	0,31	0,28	0,15	0,08
11			3,32	2,10	4,87	2,70	1,56	0,93	1,41	0,76	0,43	0,39	0,21	0,10
12			4,40	2,80		3,61	2,09	1,25	1,90	1,03	0,58	0,53	0,28	0,14
13				3,63		4,69	2,74	1,65	2,49	1,37	0,78	0,71	0,38	0,19
14				4,67			3,56	2,15	3,25	1,80	1,03	0,94	0,51	0,26
15							4,54	2,77	4,15	2,32	1,33	1,22	0,66	0,34
16								3,51		2,96	1,71	1,57	0,86	0,45
17								4,39		3,72	2,17	2,00	1,10	0,57
18										4,64	2,73	2,52	1,40	0,73
19											3,38	3,13	1,75	0,93
20											4,16	3,85	2,17	1,16
21												4,70	2,67	1,44
22													3,26	1,77
23													3,94	2,16
24													4,74	2,62
25														3,15
26														3,76
27														4,46

Если $\min(g, h) \geq 4$ и $g + h \geq 20$, то вероятности $p(u)$ достаточно точно приближаются формулой

$$p(u) = \Phi \left(\frac{u + \frac{1}{2} - \frac{1}{2}gh}{\sqrt{\frac{1}{12}gh(g+h+1)}} \right)$$

Таблица 10 представляет собой сокращенный вариант таблицы из отчета: van der Vaart H. R., Gebruiksaanwijzing voor de toets van Wilcoxon, Mathematisch Centrum Amsterdam, Rapport S. 32 (1952).

Таблица 11

Границы критической области для критерия X

Односторонние 2,5 %-границы				Односторонние 1 %-границы				Односторонние 0,5 %-границы			
n	$g-h=$ 0 или 1	$g-h=$ 2 или 3	$g-h=$ 4 или 5	n	$g-h=$ 0 или 1	$g-h=$ 2 или 3	$g-h=$ 4 или 5	n	$g-h=$ 0 или 1	$g-h=$ 2 или 3	$g-h=$ 4 или 5
6	∞	∞	∞	6	∞	∞	∞	6	∞	∞	∞
7	∞	∞	∞	7	∞	∞	∞	7	∞	∞	∞
8	2,40	2,30	∞	8	∞	∞	∞	8	∞	∞	∞
9	2,38	2,20	∞	9	2,80	∞	∞	9	∞	∞	∞
10	2,60	2,49	2,30	10	3,00	2,90	2,80	10	3,20	3,10	∞
11	2,72	2,58	2,40	11	3,20	3,00	2,90	11	3,40	3,40	∞
12	2,86	2,79	2,68	12	3,29	3,30	3,20	12	3,60	3,58	3,40
13	2,96	2,91	2,78	13	3,50	3,36	3,18	13	3,71	3,68	3,50
14	3,11	3,06	3,00	14	3,62	3,55	3,46	14	3,94	3,88	3,76
15	3,24	3,19	3,06	15	2,74	3,68	3,57	15	4,07	4,05	3,88
16	3,39	3,36	3,28	16	3,92	3,90	3,80	16	4,26	4,25	4,12
17	3,49	3,44	3,36	17	4,06	4,01	3,90	17	4,44	4,37	4,23
18	3,63	3,60	3,53	18	4,23	4,21	4,14	18	4,60	4,58	4,50
19	3,73	3,69	3,61	19	4,37	4,32	4,23	19	4,77	4,71	4,62
20	3,86	3,84	3,78	20	5,52	4,50	4,44	20	4,94	4,92	4,85
21	3,96	3,92	3,85	21	4,66	4,62	4,53	21	5,10	5,05	4,96
22	4,08	4,06	4,01	22	4,80	4,78	4,72	22	5,26	5,24	5,17
23	4,18	4,15	4,08	23	4,92	4,89	4,81	23	5,40	5,36	5,27
24	4,29	4,27	4,23	24	5,06	5,04	4,99	24	5,55	5,53	5,48
25	4,39	4,36	4,30	25	5,18	5,14	5,08	25	5,68	5,65	5,58
26	4,50	4,48	4,44	26	5,30	5,29	5,24	26	5,83	5,81	5,76
27	4,59	4,56	4,51	27	5,42	5,39	5,33	27	5,95	5,92	5,85
28	4,69	4,68	4,64	28	5,54	5,52	5,48	28	6,09	6,07	6,03
29	4,78	4,76	4,72	29	5,65	5,62	5,57	29	6,22	6,19	6,13
30	4,88	4,87	4,84	30	5,77	5,75	5,72	30	6,35	6,34	6,30
31	4,97	4,95	4,91	31	5,87	5,85	5,80	31	6,47	6,44	6,39
32	5,07	5,06	5,03	32	5,99	5,97	5,94	32	6,60	6,58	6,55
33	5,15	5,13	5,10	33	6,09	6,07	6,02	33	6,71	6,69	6,64
34	5,25	5,24	5,21	34	6,20	6,19	6,16	34	6,84	6,82	6,79
35	5,33	5,31	5,28	35	6,30	6,28	6,24	35	6,95	6,92	6,88
36	5,42	5,41	5,38	36	6,40	6,39	6,37	36	7,06	7,05	7,02
37	5,50	5,48	5,45	37	6,50	6,48	6,45	37	7,17	7,15	7,11
38	5,59	5,58	5,55	38	6,60	6,59	6,57	38	7,28	7,27	7,25
39	5,67	5,65	5,62	39	6,70	6,68	6,65	39	7,39	7,37	7,33
40	5,75	5,74	5,72	40	6,80	6,79	6,77	40	7,50	7,49	7,47
41	5,83	5,81	5,79	41	6,89	6,88	6,85	41	7,62	7,60	7,56
42	5,91	5,90	5,88	42	6,99	6,98	6,96	42	7,72	7,71	7,69
43	5,99	5,97	5,95	43	7,08	7,07	7,04	43	7,82	7,81	7,77
44	6,06	6,06	6,04	44	7,17	7,17	7,14	44	7,93	7,92	7,90
45	6,14	6,12	6,10	45	7,26	7,25	7,22	45	8,02	8,01	7,98
46	6,21	6,21	6,19	46	7,35	7,35	7,32	46	8,13	8,12	8,10
47	6,29	6,27	6,25	47	7,44	7,43	7,40	47	8,22	8,21	8,18
48	6,36	6,35	6,34	48	7,53	7,52	7,50	48	8,32	8,31	8,29
49	6,43	6,42	6,39	49	7,61	7,60	7,57	49	8,41	8,40	8,37
50	6,50	6,50	6,48	50	7,70	7,69	7,68	50	8,51	8,50	8,48

Двусторонние 5 %-границы

Двусторонние 2 %-границы

Двусторонние 1 %-границы

Если случайная величина X находится вне указанных границ, то «нулевая гипотеза» отвергается.

Вспомогательная таблица для критерия X

$$Q = \frac{1}{n} \sum_{r=1}^n \left[\Psi \left(\frac{r}{n+1} \right) \right]^2$$

n	Q	n	Q	n	Q
1	0,000	51	0,872	101	0,923
2	0,186	52	0,874	102	0,924
3	0,303	53	0,876	103	0,924
4	0,386	54	0,877	104	0,925
5	0,449	55	0,879	105	0,926
6	0,497	56	0,880	106	0,926
7	0,537	57	0,882	107	0,927
8	0,570	58	0,884	108	0,927
9	0,598	59	0,885	109	0,928
10	0,622	60	0,887	110	0,928
11	0,642	61	0,888	111	0,929
12	0,661	62	0,889	112	0,929
13	0,677	63	0,891	113	0,930
14	0,692	64	0,892	114	0,930
15	0,705	65	0,893	115	0,931
16	0,716	66	0,894	116	0,931
17	0,727	67	0,895	117	0,932
18	0,737	68	0,897	118	0,932
19	0,746	69	0,898	119	0,932
20	0,755	70	0,899	120	0,933
21	0,763	71	0,900	121	0,933
22	0,770	72	0,901	122	0,934
23	0,777	73	0,902	123	0,934
24	0,783	74	0,903	124	0,935
25	0,789	75	0,904	125	0,935
26	0,794	76	0,905	126	0,935
27	0,799	77	0,906	127	0,936
28	0,804	78	0,907	128	0,936
29	0,809	79	0,908	129	0,937
30	0,813	80	0,908	130	0,937
31	0,817	81	0,909	131	0,937
32	0,821	82	0,910	132	0,938
33	0,825	83	0,911	133	0,938
34	0,829	84	0,912	134	0,938
35	0,833	85	0,913	135	0,939
36	0,836	86	0,913	136	0,939
37	0,839	87	0,914	137	0,939
38	0,842	88	0,915	138	0,940
39	0,845	89	0,916	139	0,940
40	0,848	90	0,916	140	0,940
41	0,850	91	0,917	141	0,941
42	0,853	92	0,918	142	0,941
43	0,855	93	0,918	143	0,941
44	0,858	94	0,919	144	0,942
45	0,860	95	0,920	145	0,942
46	0,862	96	0,920	146	0,942
47	0,864	97	0,921	147	0,943
48	0,866	98	0,922	148	0,943
49	0,868	99	0,922	149	0,943
50	0,870	100	0,923	150	0,944

Таблица 13

Доверительные границы для выборочного коэффициента корреляции r

В случае общей корреляции $f = n - 2$, в случае частной корреляции $f = n - k - 2$, где k — количество исключенных величин.

f ↓	Двусторонние границы				f ↓	Двусторонние границ			
	5%	2%	1%	0,1%		5%	2%	1%	0,1%
1	0,997	1,000	1,000	1,000	16	0,468	0,543	0,590	0,708
2	0,950	0,980	0,990	0,999	17	0,456	0,529	0,575	0,693
3	0,878	0,934	0,959	0,991	18	0,444	0,516	0,561	0,679
4	0,811	0,882	0,917	0,974	19	0,433	0,503	0,549	0,665
5	0,754	0,833	0,875	0,951	20	0,423	0,492	0,537	0,652
6	0,707	0,789	0,834	0,925	25	0,381	0,445	0,487	0,597
7	0,666	0,750	0,798	0,898	30	0,349	0,409	0,449	0,554
8	0,632	0,715	0,765	0,872	35	0,325	0,381	0,418	0,519
9	0,602	0,685	0,735	0,847	40	0,304	0,358	0,393	0,490
10	0,576	0,658	0,708	0,823	45	0,288	0,338	0,372	0,465
11	0,553	0,634	0,684	0,801	50	0,273	0,322	0,354	0,443
12	0,532	0,612	0,661	0,780	60	0,250	0,295	0,325	0,408
13	0,514	0,592	0,641	0,760	70	0,232	0,274	0,302	0,380
14	0,497	0,574	0,623	0,742	80	0,217	0,257	0,283	0,357
15	0,482	0,558	0,606	0,725	90	0,205	0,242	0,267	0,338
					100	0,195	0,230	0,254	0,321
↑ f	2,5%	1%	0,5%	0,05%	↑ f	2,5%	1%	0,5%	0,05%
	Односторонние границы					Односторонние границы			

При $f > 100$ следует вычислить

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{f}$$

и применить таблицу 7.

Таблица 13 заимствована из сборника: Pearson E. S. and Hartley H. O., *Biometrika, Tables for Statisticians v. 1*, Cambridge Univ. Press (1954), Table 13, p. 138.

Примеры, упорядоченные по областям их применений

Число, стоящее после названия, указывает страницу

Теория вероятностей

1. Вероятности при бросании игральной кости 13
3. Выборка без возвращения 16
21. Оценка вероятности 185
28. Оценка вероятности 202
43. Проверка предполагаемых значений вероятностей 314

Теория ошибок, нормальное и родственное ему распределения

4. Среднее значение и квадратичное отклонение гауссовой функции ошибок 26
5. Суммы нормально распределенных случайных величин 33
13. Округление результатов наблюдений 105
22. Оценка среднего значения и дисперсии 186
25. Оценка среднего значения 200
26. Оценка дисперсии 201
27. Метод наименьших квадратов 201
29. Распределение χ^2 с множителем α 212
30. Прямоугольное распределение 214
42. Наиболее мощный критерий для проверки гипотезы $\hat{x} = 0$ 313

Физика и химия

10. Космическое излучение 65
16. Период колебаний физического маятника 138
23. Многократные измерения концентраций 187
24. Оценка положения источника излучения 191
39. Анализ газов 292

Астрономия, геодезия и метеорология

12. Осадки в Ротемстеде 100
14. Ошибки округления в сатурновых таблицах 131
15. Широта Капштадта 135
18. Обработка повторных наблюдений при триангуляции 161
19. Византийские солнечные таблицы 170

Биология и психология

11. Селекционные опыты по Юхансену 86
32. Сцепленные наследственные признаки 226
34. Группы крови, оценка генных частот 250
35. Группы крови, проверка гипотезы Бернштейна 255
36. Реакция кроликов на гельземицин 262

37. Отклонения от закона Менделя 286
38. Отклонение распределения размеров бобов от нормального распределения 289
40. Бактерии в пробах почвы 302
41. Пересаживание имагинальных дисков 305
44. Облучение яиц мухи-дрозофилы 323
46. Корреляция между размером цветочной пыльцы и количеством пор для выхода пыльцевых трубок 361

Медицина и гигиена

7. Смертность от раздутия бронхов 48
9. Новая терапия при тромбозах 59
17. Прибавление школьников в весе 154

Статистика народонаселения и экономическая статистика

6. Рождение девочек и мальчиков 43

8. Выборочный метод 53
20. Выплавка чугуна 179
47. Связь между количеством осадков и урожаем пшеницы 368

Технические приложения

31. Статистический контроль качества продукции 215
45. Сокращение времени простоев 357

Статистическая теория стрельбы

2. Стрельба в круг 14
33. Стрельба по точечной цели 228

Теория корреляции и экспериментальная психология

48. Доверительные границы для истинного коэффициента корреляции 381
49. Сравнение двух коэффициентов корреляции 382
50. Корреляция между экзаменационными оценками 393

Краткий англо-русский словарь
статистических терминов, использованных в этой книге

- Accept**, не отвергать
Analysis of variance, дисперсионный анализ
Asymptotically efficient, асимптотически эффективный
— **normal**, асимптотически нормальный
- Best**, наилучший
— **asymptotically normal (BAN)**, наилучший асимптотически нормальный (НАН)
Bias, смещение, систематическая ошибка
Binomial distribution, биномиальное распределение
Bio-assay, биологические испытания
- Central limit theorem**, центральная предельная теорема
Characteristic function, характеристическая функция
Chi-square, хи-квадрат, χ^2
Complete set, полная система
Composite, сложный (составной)
Conditional expectation, условное математическое ожидание
— **probability**, условная вероятность
Confidence limits, доверительные пределы
Consistent, состоятельный
Contingency table, таблица сопряженности признаков
- Continuous**, непрерывный
Convergence in probability, сходимость по вероятности
Correlation, корреляция
— **coefficient**, коэффициент корреляции
Covariance, ковариация
Cramér—Rao inequality, неравенство Крамера—Рао (неравенство Фреше)
Critical region, критическая область
— **set**, критическое множество
— **value**, критическое значение
- Degrees of freedom**, степени свободы
Dependent, зависимый
Distribution, распределение
— **free**, не зависящий от распределения, непараметрический
— **function**, функция распределения
Dosis—mortality curve, кривая эффекта
- Effect curve**, кривая эффекта
Efficient, эффективный
Error first kind, ошибка первого рода
— **probability**, вероятность ошибки
Error second kind, ошибка второго рода
Estimable, допускающий оценку
Estimate, оценка
Event, событие
Excess, эксцесс

- Expectation value, математическое ожидание, среднее значение
- Fourfold table, таблица сопряженности признаков 2×2
- Frequency, частота
- Fréchet inequality, неравенство Фреше (неравенство Крамера — Рао)
- Gene, ген
- Hypergeometric distribution, гипергеометрическое распределение
- Hypothesis, simple, гипотеза простая — composite, гипотеза сложная
- Independent, независимый
- Information, информация — inequality, неравенство информации (неравенство Фреше — Крамера — Рао)
- Intraclass correlation, корреляция между классами
- Lag, запаздывание
- Level, уровень
- Likely, правдоподобный
- Likelihood, правдоподобие — ratio, отношение правдоподобия — — test, критерий отношения правдоподобия
- Logistic curve, логистическая кривая
- Logit, логит
- Loss, проигрыш, убыток
- Maximum Likelihood, наибольшее правдоподобие
- Mean of sample, выборочное среднее значение — — population, среднее значение, математическое ожидание
- Measurable, измеримый
- Median, медиана
- Minimum variance estimate, оценка с наименьшей дисперсией, наилучшая оценка
- Moment, момент
- Most powerful, наиболее мощный — — unbiased test, наиболее мощный несмещенный критерий
- Nonparametric, непараметрический — test, непараметрический критерий
- Normal distribution, нормальное распределение
- Null hypothesis, нулевая гипотеза, основная гипотеза
- One-sided, односторонний
- Order statistics, порядковые статистики — test, порядковый критерий
- Partial correlation, частная корреляция
- Poisson distribution, распределение Пуассона
- Population, совокупность — mean, среднее значение, математическое ожидание
- Power, мощность
- Powerful, мощный
- Probability, вероятность — density, плотность вероятности, плотность распределения
- Probable error, вероятная ошибка
- Probit, пробит
- Random sample, случайная выборка — variable, случайная величина
- Range, размах
- Rank correlation, ранговая корреляция
- Rectangular, прямоугольный
- Region, область
- Regression, регрессия
- Regular, регулярный — best asymptotically normal (RBAN) estimate, регулярная наилучшая асимптотически нормальная (РНАН) оценка
- Reject, отвергать

- Sample**, *выборка*
 — **mean**, *выборочное среднее значение*
 — **space**, *выборочное пространство*
- Sampling**, *выбор, выборочный метод*
- Score**, *взнос, доля*
- Second limit theorem**, *вторая предельная теорема, теорема Маркова*
- Sign test**, *критерий знаков*
- Significance**, *значимость*
 — **level**, *уровень значимости*
 — **test**, *критерий значимости*
- Similar to sample space**, *подобный пространству выборок (обладающий уровнем значимости, в точности равным β)*
- Simple hypothesis**, *простая гипотеза*
- Skewness**, *асимметрия*
- Standard deviation**, *квадратичное отклонение*
 — **error**, *средняя ошибка*
- Step function**, *ступенчатая функция*
- Statistic**, *статистика (функция результатов наблюдений)*
- Stochastic**, *стохастический, вероятностный*
 — **approximation**, *стохастическая аппроксимация*
- Student's test**, *критерий Стьюдента*
- Sufficient estimate**, *достаточная оценка*
 — **statistic**, *достаточная статистика*
- Superefficient**, *сверхэффективный*
- Tail**, *хвост, шлейф (часть графика функции распределения, соответствующая большим абсолютным величинам аргумента)*
- Test**, *критерий, тест*
 — **of significance**, *критерий значимости*
 — **statistic**, *статистика, на основе которой построен критерий*
- Testing**, *проверка*
- Tie**, *связь (случайное осуществление равенства нескольких результатов наблюдений)*
- Trend**, *тренд (уравнение регрессии, в котором аргументом является время)*
- Two-sample problem**, *задача сравнения двух выборок*
- Two-sided**, *двусторонний*
- Unbiased**, *несмещенный*
- Uniformly most powerful**, *равномерно наиболее мощный*
- Up and down**, *вверх и вниз*
- Variance**, *дисперсия*
 — **ratio test**, *критерий отношения дисперсий*
- Within classes**, *внутри классов*

Указатель

- Аксиома непрерывности 15
Аксиомы Колмогорова 15
Аномалия истинная 171
— средняя 171
Арифметическое среднее 98
Асимметрия 281
Асимптотическая дисперсия 221
Асимптотически несмещенный 222
— нормальный 122
— эквивалентный 249
— эффективный 223, 246
Асимптотическое разложение 22, 64
— распределение 221, 331
— среднее значение 221
- Бернштейн С. Н. 41, 52
Бернштейн Ф. 250, 255
Берэнс 259, 266
Бета-функция 75
— неполная 49, 292
Биологические испытания 257
Булевская алгебра 14
- Вальд 221, 284, 352
Варден, ван дер 48, 170, 175, 263,
344, 351, 356
Вариационный ряд 93
Варт, ван дер 337, 418
Вектор 79
Вероятная ошибка 101
Вероятное отклонение 108
Вероятностная бумага 264
Вероятность 13
— полная 17
— условная 16
Верхний вектор 79
Вес 136
Вилкоксона критерий 328
Вольфовитц 195, 221, 352
Выборка 11, 84, 98
Выборочный метод 51, 53
- Гальтон 86
Гамма-функция 71
- Гамма-функция неполная 71,
143
Гармонический анализ 174
Гаусс 12, 134, 155, 166
Гауссова теория ошибок 134
— функция ошибок 20
Гипергеометрическое распределение
52
Гиперплоскость 241
Гипотеза простая 315
— сложная 315
Гиппарх 174
Гнеденко Б. В. 97, 123
Границы двусторонние 43
— для F 292, 411
— — R 388
— — r 367, 421
— — s^2 144
— — t 148, 410
— — U 328, 417
— — X 349, 419
— — χ^2 144, 406, 409
— односторонние 43
Графическая оценка кривой эффекта
263
Группировка 84, 283
- Дантциг, ван 338
Диксон 267
Дисперсионное отношение 290
Дисперсионный анализ 296
Дисперсия 26
— внутри классов 296
— выборочная 98
— между классами 296
Доверительные границы 44, 65
— — для медианы 325
— — точные 48
Доверительный уровень 44
Дуб 221
Дюге 122
- Задача двух выборок 325
Закон больших чисел 37, 122
— — — усиленный 123

- Знаков критерий 321, 416
 — — двусторонний 322
 — — односторонний 322
- Измеримая функция 18
 Измеримое множество 18, 29
 Инварианты 80
 Инверсия 328
 Интеграл Лебега 24
 — несобственный 70
 — Римана 30
 — Стильбеса 24
 — Фурье 111
 Интегральное среднее 103
 — уравнение 211
 Интерполяция квадратичная 132
 — линейная 131
 Информация 190, 222
 Истинная точка 165
- Каратеодори 14, 18, 25
 Качественные признаки 383
 Квадратичное отклонение 26, 98
 133
 Квартиль 94, 107
 Кендалл М. Дж. 154, 380, 384, 394
 Кетле 84
 Клоппер 49, 315
 Ковариация 358
 — выборочная 360
 Колмогоров А. Н. 15, 25, 86, 93,
 123, 125, 191, 205, 209, 217
 Колмогорова критерий 92
 Контрагреднентный 79
 Корреляция 358
 — внутри классов 304
 Кочрен 284, 286
 Коши 106
 Коэффициент корреляции 358, 362
 — — выборочный 360
 — — истинный 358
 — — частный 369
 — — регрессии 177, 358
 — — выборочный 177, 361
 Крайние значения 93
 Крамер 74, 103, 114, 120, 126, 195,
 234, 242, 250, 281, 289, 316, 320
 Кривая Кетле 84
 — — эффекта 257
 — — логарифмическая 257
 — — логистическая 267
 — — нормальная 258
 Критерий Вилкоксона 328
 — знаков 321, 416
 — — двусторонний 322
- Критерий Вилкоксона односторон-
 ный 322
 — Колмогорова 92
 — отношения правдоподобия 311
 — симметрии 325
 — Смирнова 327
 — Стьюдента 145, 154
 — F' 299
 — X 346
 — — двусторонний 349
 — — односторонний 349
 — Δ 92
 — χ^2 55, 67, 252, 272
 Критическая область 308
- Лаплас 39, 124
 Леви 114, 120, 124
 Лежандр 155
 Лё Кам 223
 Леманн 195, 203, 215, 218, 332
 Линдсберг 125
 Линия регрессии 138, 175
 — — эмпирическая 177, 360
 Логистическая кривая 267
 Логит 267
 Ляпунов А. М. 124
- Марков А. А. 39, 124, 127
 Математическое ожидание 24, 98,
 110
 — — условное 203
 Матрица обратная 80
 Медиана выборочная 93, 105
 — истинная 93, 105, 325
 Менделя закон 226, 286
 Мера 18, 29
 Метод двух точек 266
 — моментов 281, 332
 — наибольшего правдоподобия
 184, 224
 — наименьших квадратов 155, 175
 — Ньютона 189
 — одной точки 266
 — площадей 259
 — повторных наблюдений 161
 — последовательных приближений
 189
 — пробитов 259
 — χ^2 282
 Метрика 234
 Минимум χ^2 233
 — χ_0^2 234
 — χ_x^2 236
 Момент 112

- Момент центральный 281
 Мощность 308
 Муавр 39
- Наблюденная точка 164, 234
 Наиболее мощный 307
 — — несмещенный критерий 316
 Наибольшее правдоподобие 184, 224
 Наилучшая оценка 194, 210
 Наименьшие квадраты 155
 Независимость 17, 25, 110, 143
 Нейман 229, 250, 272, 278, 307, 316, 317
 Непараметрический критерий 321
 Непрерывность характеристической функции 111
 Неравенство Буняковского 195
 — информации 195
 — Крамера—Рао 195
 — Фреше 195
 — Чебышева 27
 — Шварца 195
 Несмещенный критерий 163
 — наиболее мощный критерий 316
 Несобственный интеграл 70
 Нижний вектор 79
 Николдм 205
 Нормальное распределение 23, 116
 — — двумерное 375
 Нормальные уравнения 157
 Нулевая гипотеза 326, 337
 Ньютона метод 189
- Область 68
 — критическая 308
 Обработка результатов биологических испытаний 257
 Обратная матрица 80
 — функция 23, 403
 Объем выборки 224
 Ограниченно-полная система 317
 Округление 100, 105
 Опыт 17
 Ортогонализация 176
 Отношение правдоподобия 311
 Оценка 41, 162, 183, 197, 225
 — вероятности 185, 202
 — дисперсии 168, 186, 201, 218
 — достаточная 197
 — квадратичного отклонения 41
 — несмещенная 43, 163, 194
 — параметров по наблюдаемым частотам 224
 — регулярная 223, 248
 — сверхэффективная 223
 — состоятельная 122, 221
- Оценка среднего значения 200
 — функции распределения 86
 Оценки поправок 160
 Ошибка второго рода 308
 — округления 129
 — первого рода 308
 — систематическая 43, 134
 — случайная 21, 134
 — средняя 134
 — элементарная 138
- Пирсон Е. 12, 49, 278, 307, 316, 408, 421
 Пирсон К. 86, 118, 242, 277, 393
 Плотность распределения вероятностей 20
 — — — двумерная 30
 Площадь поверхности многомерной сферы 73
 Полная система функций 212
 Полярные координаты 69
 Поправки Шеппарда 101
 Порядковые статистики 93
 Порядковый критерий 321
 Порядок величины 38, 340
 Последовательное интегрирование 68
 — приближение 189
 Правдоподобия метод 184, 224
 — отношение 311
 — уравнение 188, 225
 — функция 184, 225
 Правдоподобное значение 184
 Предельная теорема 120
 — — вторая 127
 — — Крамера—Леви 120
 — — центральная 124
 — — элементарная 127
 Преобразование ортогональное 77
 — Рао 42
 — Фишера 380
 Приемочный контроль 215
 Пробит 264
 — метод 259
 Проверка гипотез 272
 — независимости 276
 — нормальности 281, 284
 Пуассона распределение 117
 — формула 62
- Равновозможные последовательности 329
 — сочетания 349
 Равномерно наиболее мощный 314, 315

- Равномерно распределенный 129
 Размах 97
 Ранговая корреляция 383, 394
 Рао 12, 42, 158, 195, 203, 212
 Распределение биномиальное 35, 115
 — Вилкоксона 336, 417
 — гипергеометрическое 52
 — Коши 160, 192
 — нормальное 23
 — прямоугольное 88, 129
 — Пуассона 117
 — равномерное 129
 — треугольное 131
 — эмпирическое 86
 — R 384
 — r 363, 375
 — r_{xy}^2 370
 — s^2 139
 — T 395
 — t 148
 — U 331
 — X 352
 — χ^2 118, 126, 139, 239
 Регрессия 175, 358
 Регулярная оценка 223, 248
 Редкие события 62, 279
 Рекомбинационное отношение 227
- Сверхэффективный 223
 Связи 323
 Секстиль 94
 Середина класса 101
 Сираждинов С. X. 217
 Систематическая ошибка 43, 134, 194
 Скалярное произведение 80
 Слуцкий Е. Е. 144
 Случайная выборка 11, 51
 — ошибка 134
 Случайные величины 18
 — — независимые 25
 Смещение 43, 194, 198
 Смирнов Н. В. 50, 54, 91, 93, 97, 154, 217, 327, 362, 406
 Смирнова критерий 327
 Событие 13
 — редкое 62
 Солнечные таблицы 170
 Состоятельный 122, 221
 Спирмен 383
 Сравнение вероятностей 54, 275
 — двух выборок 325
 — — средних 148
 — частот 65
 Среднее значение 24, 98
 — — вектора 110
- Средне значение комплексной случайной величины 110
 — отклонение 108
 Средняя ошибка 134
 Статистика 206
 — достаточная 198, 206
 — порядковая 93
 Статистики 93, 198, 206
 Статистическое понятие вероятности 13
 Степени свободы 118, 254
 Стирлинга формула 39, 74
 Стохастическая аппроксимация 269
 Ступенчатая линия 87
 — функция 20, 40
 Стьюдент 148
 Стьюдента критерий 145, 148
 — распределение 148
 Сфера многомерная 73
 Сходимость по вероятности 122
- Таблица сопряженности признаков 61, 276
 Тело множеств 14
 Тензор 79
 Теорема Пифагора 166
 — полной аддитивности 18
 — сложения 15, 18
 Теория ошибок 134
 Точки решетки 241
 Точные доверительные границы 48
 Тренд 179
- Улучшение оценок 210
 Уравнивание центра 172
 Уровень доверительный 44
 — значимости 313
 Условная вероятность 16
 Условное математическое ожидание 203
- Феллер 39, 93, 124, 125
 Финней 265
 Фишер 12, 61, 97, 100, 118, 126, 150, 184, 190, 226, 287, 302, 307, 368, 380
 Формула асимптотическая 96
 — Бернулли 35
 — обращения Крамера 114
 — — Леви 114
 — полной вероятности 17
 — Пуассона 62
 — Стирлинга 39, 74
 Фреше 24, 97, 127, 194

- Функциональное уравнение 72
Функция измеримая 18
— мощности 339
— нормального распределения 23
— ошибок 20, 21
— правдоподобия 184, 225
— распределения 19
— — эмпирическая 86
— характеристическая 111
- Хальд 54, 295, 408, 415
Характеристическая функция 111
Хартли 408, 421
Хельмерт 118, 161
Хинчин А. Я. 122
Ходжес 195, 223, 268
Хотеллинг 221
- Частная корреляция 369
Частота 13, 38, 224
Чебышева неравенство 27
Чжун 270
- Шеппарда поправка 101
- Эйлерова бета-функция 75
— гамма-функция 71
Эксцентрик 171
Эксцесс 281
Эллипс 45
Эмпирическая линия регрессии 360
Эмпирическое распределение 86
Эффективность 246
— асимптотическая 246

О Г Л А В Л Е Н И Е

Предисловие к русскому переводу	5
Предисловие	7
Введение	11
Глава I. Общие основы	13
§ 1. Основные понятия теории вероятностей	13
§ 2. Случайные величины. Функции распределения	18
§ 3. Среднее значение и квадратичное отклонение	24
§ 4. Интегральные представления средних значений и вероятностей	28
Глава II. Вероятности и частоты	35
§ 5. Биномиальное распределение	35
§ 6. Как велико может быть отклонение частоты h от вероятности p ?	39
§ 7. Доверительные границы для неизвестной вероятности ...	44
§ 8. Проблема случайного отбора. Выборочный метод	51
§ 9. Сравнение двух вероятностей	54
§ 10. Частота редких событий	62
Глава III. Математические вспомогательные средства	68
§ 11. Кратные интегралы. Переход к полярным координатам ...	68
§ 12. Бета- и гамма-функции	71
§ 13. Ортогональные преобразования	77
§ 14. Квадратичные формы и их инварианты	79
Глава IV. Оценки функций распределения, средних значений и дисперсий	84
§ 15. Кривая Кетле	84
§ 16. Оценки функций распределения	86
§ 17. Порядковые статистики	93

§ 18. Выборочное среднее значение и выборочная дисперсия	98
§ 19. Поправки Шеппарда	101
§ 20. Другие числовые характеристики распределения	105
Глава V. Интегралы Фурье и предельные теоремы	110
§ 21. Характеристические функции	110
§ 22. Примеры	115
§ 23. Распределение χ^2	118
§ 24. Предельные теоремы	120
§ 25. Прямоугольное распределение. Ошибки округления	129
Глава VI. Гауссова теория ошибок и критерий Стьюдента	134
§ 26. Гауссова теория ошибок	134
§ 27. Распределение s^2	139
§ 28. Критерий Стьюдента	145
§ 29. Сравнение двух средних значений	148
Глава VII. Метод наименьших квадратов	155
§ 30. Выравнивание ошибок наблюдений	155
§ 31. Средние значения и дисперсии оценок \tilde{y}	162
§ 32. Оценка дисперсии σ^2	168
§ 33. Линии регрессии	175
§ 34. Выяснение причин изменения экономических показателей	180
Глава VIII. Оценки неизвестных параметров	183
§ 35. Метод наибольшего правдоподобия Р. А. Фишера	184
§ 36. Вычисление максимума	188
§ 37. Неравенство Фреше	194
§ 38. Достаточные оценки и наилучшие оценки	197
§ 39. Примеры	200
§ 40. Условные математические ожидания	203
§ 41. Достаточные статистики	206
§ 42. Применение теории условных математических ожиданий к задаче отыскания наилучших несмещенных оценок	210
§ 43. Приложения	212
§ 44. Оценка дисперсии нормального распределения	218
§ 45. Асимптотические свойства	221
Глава IX. Оценка параметров по наблюдаемым частотам	224
§ 46. Метод наибольшего правдоподобия	224
§ 47. Состоятельность оценок наибольшего правдоподобия	229

§ 48. Наибольшее правдоподобие, минимум χ^2 и наименьшие квадраты	238
§ 49. Асимптотическое распределение χ^2 и \tilde{d} при $n \rightarrow \infty$	239
§ 50. Асимптотическая эффективность	246
§ 51. Критерий χ^2	252
Глава X. Обработка результатов биологических испытаний	257
§ 52. Кривая эффекта и логарифмическая кривая эффекта	257
§ 53. Метод площадей Берэнса и Кербера	259
§ 54. Методы, основанные на предположении нормальности кривой эффекта	263
§ 55. Методы «вверх и вниз»	267
Глава XI. Проверка гипотез с помощью статистических критериев 272	272
§ 56. Применения критерия χ^2	272
§ 57. Критерий, основанный на дисперсионном отношении (критерий F)	290
§ 58. Дисперсионный анализ	296
§ 59. Общие принципы. Наиболее мощные критерии	307
§ 60. Сложные гипотезы	315
Глава XII. Порядковые критерии	321
§ 61. Критерий знаков	321
§ 62. Задача двух выборок	325
§ 63. Критерий Вилкоксона	328
§ 64. Мощность критерия Вилкоксона	337
§ 65. Критерий X	346
Глава XIII. Корреляция	358
§ 66. Ковариация и коэффициент корреляции	358
§ 67. Коэффициент корреляции как признак зависимости.....	362
§ 68. Частные коэффициенты корреляции	369
§ 69. Распределение выборочного коэффициента корреляции зависимых случайных величин	375
§ 70. Коэффициент ранговой корреляции R , по Спирмену	383
§ 71. Коэффициент ранговой корреляции T , по Кендаллу	394
Т а б л и ц ы	401
Примеры, упорядоченные по областям их применений	422
Краткий англо-русский словарь статистических терминов, использованных в этой книге	424
Указатель	427

Б. Л. Ван дер Варден
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Редактор *В. И. Битюков*
Художник *И. В. Богдановский*
Художественный редактор *А. В. Вилленева*
Технический редактор *А. Г. Резоухова*
А. П. Иванова

Сдано в производство 11/II 1959 г.
Подписано к печати 11/II 1960 г.
Бумага $60 \times 92 \frac{1}{16}$ = 13,6 бум. л.
27,2 печ. л.
Уч.-изд. л. 25,3. Изд. №1/4463
Цена 19 р. 20 к. Зак. 1062

ИЗДАТЕЛЬСТВО
ИНОСТРАННОЙ ЛИТЕРАТУРЫ
Москва, Ново-Алексеевская, 52.

Едьетеми Нйомда, Будапешт.