

З. БРАНДТ СТАТИСТИЧЕСКИЕ

МЕТОДЫ

АНАЛИЗА

НАБЛЮДЕНИЙ





**STATISTICAL AND SIEGMUND
COMPUTATIONAL BRANDT
METHODS
IN DATA ANALYSIS**

INSTITUTE OF HIGH
ENERGY PHYSICS

HEIDELBERG UNIVERSITY
1970

North Holland Publishing
Company — Amsterdam · London
American Elsevier Publishing Company,
INC. — New York

З. БРАНДТ СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА НАБЛЮДЕНИЙ

Перевод с английского
Г. А. ПОГРЕВИНСКОГО

Под редакцией
В. Ф. ПИСАРЕНКО

Книга представляет собой краткое введение в статистический анализ данных наблюдений и предназначена для лиц, не являющихся математиками, но которым в своей практической работе приходится иметь дело с использованием статистических методов обработки данных.

Книга состоит из 12 глав и 6 небольших приложений. Первые главы посвящены кратко изложению основ теории вероятностей и математической статистике. В последующих главах изложены вопросы статистики, чаще всего встречающиеся в приложениях: метод максимального правдоподобия, статистическая проверка гипотез, метод наименьших квадратов, дисперсионный анализ, линейная регрессия. В приложениях собраны элементы программирования на ФОРТРАНе, основные формулы матричного исчисления, комбинаторный анализ и наиболее употребительные формулы и таблицы.

В книге имеется много иллюстраций и численных примеров, ее легко читать даже лицам, впервые встречающимся с математической статистикой. Она, несомненно, принесет большую пользу широкому кругу читателей, сталкивающихся со статистическим анализом данных.

*Редакция космических исследований,
астрономии и геофизики*

Б $\frac{20204-110}{041(01)-75}$ 110 — 75 © Перевод на русский язык, «Мир», 1975

З. Брандт

СТАТИСТИЧЕСКИЕ МЕТОДЫ
АНАЛИЗА НАБЛЮДЕНИИ

Редактор Э. Медушевская
Художник О. Шанецкий
Художественный редактор И. Шаврова
Технический редактор Н. Борисова
Корректоры И. Алексеева и Т. Лаврова

Сдано в набор 19/VI 1974 г. Подписано к печати 22/1 1975 г. Бумага тип. № 2 $60 \times 90^{1/16} = 9,75$ бум. л. Печ. л. 19,5. Уч.-изд. л. 16,29. Изд. № 27/7388. Цена 1 р. 33 к. Зак. 569

Издательство «Мир». Москва, 1-й Рижский пер., 2

Отпечатано в ордена Трудового Красного Знамени Ленинградской типографии № 2 имени Евгении Соколовой Союзполиграфпрома при Государственном комитете Совета Министров СССР по делам издательств, полиграфии и книжной торговли, 198052, Ленинград, Л-52, Измайловский проспект, 29, с матриц книжной фабрики им. М. В. Фрунзе Республиканского промислава, объединения «Полиграфкнига» Госкомиздата УССР, Харьков, Донец-Захаржевская, 6/8

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

В связи с резким увеличением за последние годы применений математики в самых различных областях исследований заметно возросла потребность в руководствах по прикладной статистике. Этим вопросам и посвящена предлагаемая в русском переводе советскому читателю книга Брандта «Статистические методы анализа наблюдений». Книга рассчитана на геофизиков, физиков, экономистов, биологов, медиков и других специалистов, имеющих дело с практическим использованием статистических методов при обработке наблюдений.

Автор не приводит строгих математических доказательств формулируемых им утверждений. При рассмотрении различных статистических задач упор сделан, скорее, на рецептурную сторону рассматриваемых вопросов. Вместе с тем, нельзя назвать эту книгу просто сборником формул и статистических правил, так как автор уделяет большое внимание выяснению наглядного смысла статистических понятий и методов и их иллюстрации на многочисленных примерах. Из имеющихся на русском языке руководств по прикладной статистике книга Брандта несколько похожа на книгу Д. Худсона «Статистика для физиков», М., «Мир», 1970 г. и на книгу Л. З. Румшицкого «Математическая обработка результатов наблюдений», М., «Наука», 1971 г. Однако в отличие от перечисленных книг она охватывает более широкий круг статистических задач, и, главное, в ней гораздо больше внимания уделено вычислительной стороне рассматриваемых методов. В частности, для многих статистических формул приведены вычислительные программы на ФОРТРАНе, а в приложении изложены основные сведения по программированию на ФОРТРАНе.

Изложение материала в книге Брандта является доступным даже для тех, кто впервые сталкивается с математической статистикой. Книга снабжена большим количеством иллюстраций и численных примеров. Поэтому она позволяет читателю ознакомиться с основами математической статистики, так сказать, «с минимальной затратой энергии». Для более глубокого ознакомления с методами математической статистики в форме, доступной для прикладников, можно рекомендовать книгу Ван дер Вардена «Математическая статистика», М., ИЛ., 1960 г.

Следует отметить, что в книге Брандта рассмотрены лишь наиболее важные статистические методы, чаще всего используемые на практике, такие, как метод максимального правдоподобия, метод наименьших квадратов, статистическая теория проверки гипотез, дисперсионный анализ, линейная регрессия и т. д.

В конце книги приводится сводка основных статистических формул и наиболее употребительные статистические таблицы. Благодаря этому книга может служить не только введением в прикладную статистику, но и удобным кратким статистическим справочником. Думается, что книга принесет пользу широкому кругу специалистов не только как практическое руководство по применению статистики, но и, возможно, как ступень для более глубокого овладения статистическими методами, освещенными в специальных монографиях.

При переводе на русский язык были устранены некоторые опечатки и неточности, имеющиеся в оригинале.

В. Ф. Писаренко

ПРЕДИСЛОВИЕ

Настоящая книга основана на лекциях, прочитанных в 1967/1968 г. студентам-физикам и специалистам по физике элементарных частиц Гейдельбергского университета. В ней обсуждаются разделы математической статистики, наиболее полезные для анализа наблюдений. Книга предназначена для студентов и научных работников в области естественных наук, медицины, техники и экономики, сталкивающихся с проблемой обработки экспериментальных данных.

Книга рассчитана на читателя, которому нужно применять методы математической статистики. Математическая строгость не преследовалась, однако это не просто список рекомендаций для различных практических приложений, а попытка объяснить идеи и принципы обсуждаемых статистических методов. Часть материала, представленного в книге, основана на лекциях и обзорных статьях, написанных физиками и для физиков [2, 15, 22].

Основные сведения из анализа читателю предполагаются известными. Другой необходимый математический аппарат, в частности теория вероятностей, кратко изложен в книге. Существенная особенность изложения состоит в использовании матричных обозначений, позволяющих компактно изложить многие вопросы, например метод наименьших квадратов. Основы матричного исчисления даны в приложении.

Наиболее сложные задачи анализа наблюдений решаются сейчас с помощью ЭВМ; для нескольких таких задач в книге приведены программы, написанные на ФОРТРАНе. В приложении даны основные сведения о ФОРТРАНе. Там же приведена краткая библиотека подпрограмм для оперирования с матрицами, составленная на основе программ, написанных Бёком в ЦЕРНе (Женева).

Предполагается, что эта книга послужит не только введением в статистический анализ, но и будет использоваться в повседневной работе. Поэтому в книге содержится несколько статистических таблиц и сводка наиболее важных формул, с тем чтобы их можно было быстро найти.

Мне бы хотелось выразить благодарность проф. Филтусу за проявленный интерес к этой теме, что послужило толчком к

чению курса лекций. Я благодарен группе моих коллег по Гейдельбергскому университету: д-рам Бендеру, Брокейту, Гизеке, Роте и Шнейдеру за обсуждение книги и в особенности д-ру Шаху за прочтение рукописи и внесение большого числа полезных замечаний и улучшений. Хочется поблагодарить также д-ра Теннера (Амстердам) за плодотворные советы по расположению материала в книге. Примеры 8.2, 11.1 и 11.2 принадлежат д-ру Иммичу (Гейдельберг). Гравюру для титульного листа сделал д-р Френк (Ветцлар).

Примеры и статистические таблицы напечатаны на ЭВМ IBM-360. Рис. 5.8 выполнен с помощью графопостроителя по программе, написанной д-ром Шнейдером.

З. Брандт

Гейдельберг, январь 1970 г.

ВВЕДЕНИЕ

В любой отрасли экспериментальной науки вслед за более ранней стадией качественного описания интересующего нас явления следует стадия его количественного изучения, т. е. измерения. Наряду с планированием и проведением эксперимента важной задачей является точная оценка получаемых данных и полное их использование. Перечислим несколько типичных проблем.

1. Изучается воздействие различных средств на увеличение веса в группе подопытных животных. После применения вещества А к 25 животным наблюдалось среднее увеличение веса на 5%. Вещество В, применявшееся к 10 животным, дало увеличение веса на 3%. Является ли средство А более эффективным? Средние цифры 5 и 3% практически не дают ответа на этот вопрос, так как меньшая цифра могла быть вызвана влиянием какого-то одного животного с самым маленьким весом. Следовательно, необходимо изучить *распределение* весов животных и их отклонения от среднего значения. Кроме того, надо решить, позволит ли количество взятых для испытания животных с определенной точностью различить эффекты от применяемых веществ.

2. В экспериментах по выращиванию кристаллов основной проблемой является поддержание точного соотношения между различными компонентами. Из группы, содержащей 500 кристаллов, отбираются и анализируются 20 кристаллов. Какие выводы можно сделать относительно остальных 480 кристаллов? Такая задача *выборки* возникает, например, при контроле на производстве, при испытании на надежность автоматических измерительных устройств, при опросе общественного мнения.

3. Получен некоторый экспериментальный результат. Требуется выяснить, не противоречит ли он уже известному теоретическому значению или ранее проведенным экспериментам? Тем самым эксперимент используется для *проверки гипотезы*.

4. Известен общий вид закона, выражающего зависимость между измеряемыми величинами; необходимо получить из эксперимента значения параметров этого закона. Например, при радиоактивном распаде число N атомов, распадающихся за секунду, убывает во времени по экспоненциальному закону $N(t) =$

$= \text{const} \times \exp(-\lambda t)$. Постоянная распада λ и ошибка ее измерения должны быть определены с помощью нескольких наблюдений $N_1(t_1), N_2(t_2), \dots$. Эта проблема *оценивания параметра*, возможно, является наиболее интересной для многих экспериментаторов.

Все эти примеры иллюстрируют характерные особенности анализа данных. В частности, мы видим, что исход эксперимента не однозначно определяется условиями его проведения, а содержит элемент случайности, т. е. является *случайной величиной*. Эта стохастическая тенденция лежит либо в природе эксперимента (подопытные животные неизбежно различны, радиоактивность является случайным явлением), либо в неизбежных погрешностях, вносимых применяемым оборудованием, т. е. в ошибках измерения. По этой причине следующая глава посвящена обзору наиболее важных понятий теории вероятности.

В 3-й и 4-й главах вводится понятие случайной величины. Рассматривается распределение случайной величины и параметры, которыми оно характеризуется, такие, например, как среднее значение и дисперсия. Особое внимание уделяется рассмотрению зависимости нескольких случайных величин. В гл. 5 изучаются распределения, представляющие специальный интерес для приложений; в частности, подробно рассматриваются свойства нормального, или гауссовского, распределения.

На практике распределение должно быть определено с помощью конечного числа наблюдений, т. е. *выборки*. Различные случаи выборочных процедур рассматриваются в гл. 6. Функции от выборки, т. е. функции от отдельных наблюдений, могут быть использованы для оценки параметров, характеризующих распределение. Выводятся требования, которым должна удовлетворять хорошая выборочная оценка. На этом этапе вводится величина χ^2 , являющаяся мерой среднего отклонения отдельных наблюдений из выборки от среднего значения и, следовательно, служащая показателем качества измерений.

Метод наибольшего правдоподобия, рассмотренный в гл. 7, составляет ядро современного статистического анализа. Он позволяет строить оценки с оптимальными свойствами. Метод рассмотрен для случаев одного и нескольких параметров и иллюстрирован рядом примеров.

Глава 8 посвящена способам проверки гипотез. Она содержит наиболее часто используемые критерии F , t и χ^2 и основные сведения из теории.

В гл. 9 предметом обсуждения является *метод наименьших квадратов* — наиболее широко применяемый статистический метод. Сначала подробно рассмотрены частные случаи прямых и косвенных измерений, а также измерений со связями, которые часто встречаются на практике. Затем обсуждается общий случай. Для общей задачи о наименьших квадратах приведена программа на

ФОРТРАНе, и ее применение проиллюстрировано серией примеров. Любую задачу о наименьших квадратах можно свести к задаче о нахождении минимума функции нескольких переменных. Это же верно и для оценивания параметров методом наибольшего правдоподобия. В гл. 10 кратко описаны несколько вычислительных методов нахождения таких минимумов.

Дисперсионный анализ (гл. 11) можно рассматривать как обобщение F -критерия. Он широко используется в биологии и медицине для изучения зависимости или, точнее, для проверки независимости измеряемых величин от условий эксперимента, характеризующихся другими переменными. Для нескольких переменных возникают довольно сложные ситуации. С помощью программы на ФОРТРАНе посчитано несколько простых численных примеров.

Линейная регрессия, рассматриваемая в последней главе, представляет собой специальный случай метода наименьших квадратов и поэтому встречается уже в гл. 9. До появления вычислительных машин обычно поддавались решению лишь линейные задачи наименьших квадратов. В линейной регрессии все еще используется специальная терминология, что оправдывает выделение этого вопроса в отдельную главу. В то же время гл. 10 является развитием гл. 9. Так, например, определяются доверительные интервалы для решения и изучаются соотношения между регрессией и дисперсионным анализом.

ВЕРОЯТНОСТИ

2.1. ЭКСПЕРИМЕНТЫ, СОБЫТИЯ,
ВЫБОРОЧНОЕ ПРОСТРАНСТВО

Так как эта книга посвящена анализу данных, получаемых из эксперимента, нам необходимо сначала выяснить, что понимается под экспериментом и его результатом. Аналогично лабораторному опыту мы определяем эксперимент как строгую последовательность заранее определенных действий, ведущих к получению одной или множества величин, представляющих результат эксперимента. Эти величины могут изменяться непрерывно (температура, длина, электрический ток) или дискретно (число частиц, день рождения человека, один из трех возможных цветов). Независимо от точности выполнения условий, при которых проводится эксперимент, результаты повторных экспериментов будут, вообще говоря, различны. Это вызывается либо внутренней природой исследуемого явления, либо ограниченной точностью измерения. Следовательно, для каждой величины возможные результаты будут лежать в некоторой ограниченной области. Множество этих областей для всех величин, составляющих результат эксперимента, образует *выборочное пространство* этого эксперимента. Так как точное определение допустимых областей для всех величин, измеряемых в данном эксперименте, затруднительно и часто вообще невозможно, фактически пространство выборок берется несколько большим, содержащим истинное выборочное пространство в качестве подпространства. В дальнейшем мы будем пользоваться этим более широким определением выборочного пространства.

Пример 2.1. При изготовлении сопротивлений важно выдержать заданные значения величин R (электрическое сопротивление, измеряемое в омах) и N (максимальное рассеяние мощности, измеряемое в ваттах). Выборочным пространством для R и N может служить координатная плоскость с осями R и N . Так как обе величины всегда положительны, то первый квадрант также является выборочным пространством.

Пример 2.2. На практике точные значения R и N несущественны, пока они находятся в заданных интервалах (например, $99 \text{ кОм} < R < 101 \text{ кОм}$; $0,49 \text{ Вт} < N < 0,60 \text{ Вт}$). В таких случаях будем говорить, что сопротивление обладает свойствами

R_n, N_n . Если значение выходит за нижнюю или верхнюю границу интервала, то индекс n будем заменять соответственно на $-$ или $+$. Следовательно, возможные значения электрического сопротивления и рассеянной мощности будут соответственно R_-, R_n, R_+ и N_-, N_n, N_+ . Выборочное пространство состоит в этом случае из 9 элементов:

$$\begin{array}{ccc} R_-N_-, & R_-N_n, & R_-N_+, \\ R_nN_-, & R_nN_n, & R_nN_+, \\ R_+N_-, & R_+N_n, & R_+N_+. \end{array}$$

Часто представляют интерес некоторые подпространства выборочного пространства. Например, элемент пространства выборок R_nN_n из примера 2.2 соответствует случаю, когда сопротивление полностью удовлетворяет техническим условиям. Обозначим такие подпространства A, B, \dots и будем говорить, что если результат эксперимента попал внутрь какого-либо подпространства, то произошло соответствующее событие A (или B, C, \dots). Если событие A не произошло, мы будем говорить об осуществлении *дополнительного* события \bar{A} (не A). Все выборочное пространство соответствует событию, которое осуществляется в каждом эксперименте. Обозначим его через E . В конце этой главы мы определим, что понимается под вероятностью осуществления события.

2.2. ПОНЯТИЕ ВЕРОЯТНОСТИ

Рассмотрим простейший эксперимент — бросание монеты. Как и подобные ему эксперименты с игральными костями и картами, он не представляет практического интереса, но весьма полезен для целей обучения. Какова вероятность того, что при первом бросании идеальной монеты выпадет «орел»? Интуиция подсказывает нам, что эта вероятность равна $\frac{1}{2}$. Мы исходим из того,

что все точки выборочного пространства (здесь мы имеем только две точки: «орел» и «решка») равновероятны и событие E (в данном случае заключающееся в выпадении «орла» или «решки») наступает с вероятностью 1. Этот способ определения вероятности может быть применен лишь в случае абсолютно симметричных экспериментов и, следовательно, имеет очень малое практическое применение. (Однако он крайне важен в статистической физике и квантовой механике, где постулат о равных вероятностях всех допустимых состояний является основой весьма успешных теорий.)

Если же абсолютная симметрия отсутствует — например, в случае обычной «физической» монеты, — то целесообразно поступить следующим образом. Пусть в большом числе экспериментов

N событие A имело место n раз. Определим вероятность появления события A как предел

$$P(A) = \lim_{N \rightarrow \infty} \frac{n}{N}. \quad (2.1)$$

Это несколько неточное *частотное определение* вероятности удобно для практических целей, хотя оно и не является математически строгим. Одна из трудностей, возникающих при таком определении, заключается в необходимости проведения бесконечного числа экспериментов, что, естественно, невозможно сделать и теоретически трудно объяснить. Хотя фактически в этой книге мы будем пользоваться частотным определением вероятности, хотелось бы привести основные понятия аксиоматической теории вероятности, принадлежащей Колмогорову [12]. Обычно используется следующий минимальный набор аксиом:

а) Каждому событию A соответствует некоторое неотрицательное число — его вероятность

$$P(A) \geq 0. \quad (2.2)$$

б) Событие E имеет вероятность, равную единице:

$$P(E) = 1. \quad (2.3)$$

в) Если события A и B несовместны, то вероятность наступления события A или B , т. е. события $(A + B)$, равна

$$P(A + B) = P(A) + P(B). \quad (2.4)$$

Из этих аксиом сразу получаются следующие полезные соотношения. Из (б) и (в)

$$P(A + \bar{A}) = P(A) + P(\bar{A}) = 1; \quad (2.5)$$

кроме того, из (а)

$$0 \leq P(A) \leq 1. \quad (2.6)$$

Из (в) можно легко получить более общую теорему для взаимно несовместных событий A, B, C, \dots

$$P(A + B + C + \dots) = P(A) + P(B) + P(C) + \dots \quad (2.7)$$

Необходимо заметить, что единственным способом комбинирования событий является уже упоминавшееся *исключающее «или»*. Любая другая комбинация должна быть получена с помощью этой логической операции. В эксперименте с бросанием игральной кости событие A может означать выпадание четного числа очков, B — нечетного числа очков, C — выпадание менее 4-х очков и D — 4-х или более очков. Представляет интерес вероятность события $(A$ или $C)$ (A и C , очевидно, не являются несов-

местными). Образовав событие (A и C), т. е. (AC), и затем соответственно AD , BC и BD , являющиеся взаимно несовместными, можно найти для события (A «или» C) (иногда пишут $\dot{A} + C$) выражение $AC + AD + BC$.

Интересно заметить, что аксиомы не дают метода получения конкретной вероятности $P(A)$.

В заключение отметим, что слово «вероятность» часто используется в повседневном языке в смысле, совершенно отличном или даже противоположном рассмотренному нами. Мы имеем в виду *субъективную вероятность*, т. е. вероятность события, определяемую мерой нашей уверенности в том, что оно осуществится. Например: «Вероятность того, что туннель под Ла-Маншем будет построен до 1984 г., равна $1/3$ ». В качестве другого примера рассмотрим такой случай: в ядерной эмульсии может остаться след от протона или пиона. Часто говорят: «След был оставлен пионом с вероятностью $1/2$ ». Но так как событие уже произошло и след мог быть оставлен лишь одной из двух частиц, рассматриваемая вероятность равна либо 0, либо 1, хотя истинное ее значение неизвестно.

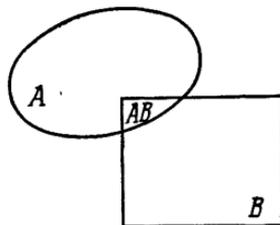
2.3. ПРАВИЛА ИСЧИСЛЕНИЯ ВЕРОЯТНОСТЕЙ; УСЛОВНАЯ ВЕРОЯТНОСТЬ

Пусть результат эксперимента обладает некоторым свойством A . Вопрос: какова вероятность того, что эксперимент будет обладать также и свойством B , т. е. какова вероятность B при условии осуществления A ? Определим эту *условную вероятность* как

$$P(B|A) = \frac{P(AB)}{P(A)}. \quad (3.1)$$

Отсюда следует, что

$$P(AB) = P(A)P(B|A). \quad (3.2)$$



Из примера, показанного на рис. 2.1, видно, что такое определение разумно. Событие (A или B) при этом осуществляется, если точка лежит либо в области A , либо в B . Если точка лежит в области пересечения, то осуществляется событие AB (A и B). Предположим, что площади различных областей пропорциональны вероятностям соответствующих событий. Тогда вероятность B при условии A равна отношению площадей AB к A , т. е. равна 1, если A содержится в B , и 0, если области не пересекаются.

Используя понятие условной вероятности, мы можем теперь сформулировать *правило полной вероятности*. Пусть некоторый

Рис. 2.1. Иллюстрация условной вероятности.

эксперимент приводит к одному из n взаимно несовместных событий

$$E = A_1 + A_2 + \dots + A_n. \quad (3.3)$$

Вероятность осуществления любого события B равна

$$P(B) = \sum_{i=1}^n P(A_i) P(B|A_i), \quad (3.4)$$

что может быть легко получено из выражений (3.2) и (2.7).

Теперь мы можем дать определение *независимости событий*. Будем говорить, что события A и B независимы, если осуществление A не влияет на вероятность появления B и наоборот, т. е. если

$$P(B|A) = P(B) \quad (3.5)$$

или, используя выражение (3.2),

$$P(AB) = P(A)P(B), \quad (3.6)$$

Вообще несколько разбиений типа (3.3)

$$\begin{aligned} E &= A_1 + A_2 + \dots + A_n, \\ E &= B_1 + B_2 + \dots + B_m, \\ &\vdots \\ E &= Z_1 + Z_2 + \dots + Z_l \end{aligned} \quad (3.7)$$

независимы, если для всех возможных комбинаций $\alpha, \beta, \dots, \omega$ выполняется условие

$$P(A_\alpha B_\beta \dots Z_\omega) = P(A_\alpha) P(B_\beta) \dots P(Z_\omega). \quad (3.8)$$

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ; РАСПРЕДЕЛЕНИЕ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

3.1. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Мы не будем сейчас интересоваться вероятностями наблюдения каких-либо конкретных событий, а рассмотрим сами события и попробуем найти наиболее простой способ их классификации; поставим, например, в соответствие событию «орел» число 0, а событию «решка» — число 1. В общем мы можем классифицировать события A_i из разбиения (3.3) гл. 2, поставив в соответствие каждому A_i его номер i . При таком способе классификации каждое событие можно характеризовать одним из возможных значений *случайной величины*. Случайные величины могут быть дискретными и непрерывными. Будем обозначать их символами x, y, \dots .

Пример 3.1. Интересно определить число монет, все еще находящихся в обращении, как функцию их возраста. Очевидно, что удобнее всего в качестве (дискретной) случайной величины использовать год чеканки каждой монеты, т. е. $x = \dots, 1873, 1874, 1875, \dots$.

Пример 3.2. Все процессы измерения или производства подвержены в большей или меньшей мере изменениям, или флуктуациям, что приводит к отклонениям в результатах. Поэтому названные результаты описываются одной или несколькими случайными величинами. Так, величины электрического сопротивления и максимума рассеянной мощности, характеризующие резистор, есть (непрерывные) случайные величины.

3.2. РАСПРЕДЕЛЕНИЕ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Вернемся от классификации событий к рассмотрению вероятностей. Рассмотрим случайную величину x и действительное число x , принадлежащее интервалу $(-\infty, +\infty)$, и будем искать вероятность события ($x < x$). Эта вероятность является функцией от x и называется *функцией распределения* вероятностей случайной величины x :

$$F(x) = P(x < x), \quad (2.1)$$

Если случайная величина x принимает лишь конечное число дискретных значений (например, число очков на гранях игральной кости), то функция распределения вероятностей этой случайной величины представляет собой ступенчатую функцию. Для рассмотренного выше примера функция распределения показана на рис. 3.1. Очевидно, любая функция распределения вероятностей является *монотонной* и *неубывающей*.

Как видно из выражения (2.3) гл. 2, в пределе при $x \rightarrow \infty$ получаем

$$\lim_{x \rightarrow \infty} F(x) = \lim_{x \rightarrow \infty} P(x < x) = P(E) = 1. \quad (2.2)$$

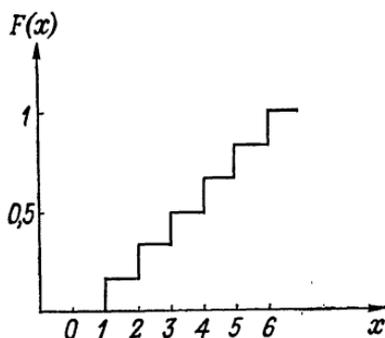


Рис. 3.1. Функция распределения вероятностей для бросания идеальной симметричной игральной кости.

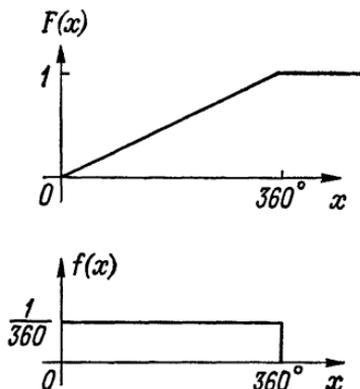


Рис. 3.2. Функция распределения вероятностей и плотность вероятности для угла положения часовой стрелки.

Применяя (2.5) гл. 2 к (2.1), получим

$$P(x \geq x) = 1 - F(x) = 1 - P(x < x) \quad (2.3)$$

и, следовательно,

$$\begin{aligned} \lim_{x \rightarrow -\infty} F(x) &= \lim_{x \rightarrow -\infty} P(x < x) = 1 - \lim_{x \rightarrow -\infty} P(x \geq x), \\ \lim_{x \rightarrow -\infty} F(x) &= 0. \end{aligned} \quad (2.4)$$

Особый интерес представляют непрерывные и дифференцируемые функции распределения вероятностей. Первая производная

$$f(x) = \frac{dF(x)}{dx} = F'(x) \quad (2.5)$$

называется *плотностью вероятности* случайной величины x .

Плотность вероятности пропорциональна вероятности события ($x \leq \mathbf{x} < x + dx$). Из выражения (2.1) и (2.5) следует, что

$$P(\mathbf{x} < a) = F(a) = \int_{-\infty}^a f(x) dx, \quad (2.6)$$

$$P(a \leq \mathbf{x} < b) = \int_a^b f(x) dx = F(b) - F(a) \quad (2.7)$$

и, в частности,

$$\int_{-\infty}^{\infty} f(x) dx = 1. \quad (2.8)$$

Простейший пример непрерывного распределения дает положение часовой стрелки на циферблате в случайные моменты времени. Здесь мы получаем постоянную плотность вероятности (рис. 3.2).

3.3. ФУНКЦИИ ОДНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ, МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ, ДИСПЕРСИЯ, МОМЕНТЫ

Кроме распределения случайной величины \mathbf{x} нас часто интересует распределение функции от этой случайной величины. Функция от случайной величины также является случайной величиной

$$\mathbf{y} = H(\mathbf{x}). \quad (3.1)$$

Следовательно, \mathbf{y} так же, как и \mathbf{x} , обладает функцией распределения вероятностей и плотностью вероятности.

В двух простых примерах предыдущего раздела вследствие симметрии природы явления мы могли сразу же определить функцию распределения вероятностей. Обычно это не удается. Вместо этого мы вынуждены получать функцию распределения из эксперимента. Часто ограничиваются получением нескольких характерных параметров вместо полного распределения.

Средним значением или *математическим ожиданием* дискретной случайной величины называется сумма всех возможных значений x_i случайной величины \mathbf{x} , умноженных на соответствующие вероятности

$$E(\mathbf{x}) = \hat{x} = \sum_{i=1}^n x_i P(\mathbf{x} = x_i). \quad (3.2)$$

Заметим, что \hat{x} является не случайной, а вполне определенной детерминированной величиной. Соответственно математическое ожидание функции (3.1) определяется как

$$E\{H(\mathbf{x})\} = \sum_{i=1}^n H(x_i) P(\mathbf{x} = x_i). \quad (3.3)$$

В случае непрерывной случайной величины (с дифференцируемой функцией распределения) по аналогии определим

$$E(\mathbf{x}) = \hat{x} = \int_{-\infty}^{\infty} x f(x) dx \quad (3.4)$$

и

$$E\{H(\mathbf{x})\} = \int_{-\infty}^{\infty} H(x) f(x) dx, \quad (3.5)$$

Взяв специальную функцию

$$H(\mathbf{x}) = (\mathbf{x} - c)^l, \quad (3.6)$$

мы получим математические ожидания

$$\alpha_l = E\{(\mathbf{x} - c)^l\}, \quad (3.7)$$

которые называются *l-ми моментами* случайной величины относительно точки *c*. Особый интерес представляют *моменты относительно среднего значения*

$$\mu_l = E\{(\mathbf{x} - \hat{x})^l\}. \quad (3.8)$$

Первые два момента, очевидно, равны

$$\mu_0 = 1, \quad \mu_1 = 0. \quad (3.9)$$

Величина

$$\mu_2 = \sigma^2(\mathbf{x}) = \text{var}(\mathbf{x}) = E\{(\mathbf{x} - \hat{x})^2\} \quad (3.10)$$

является низшим моментом, содержащим информацию о среднем отклонении случайной величины \mathbf{x} от своего среднего значения. Этот момент называется *дисперсией* случайной величины \mathbf{x} .

Попробуем теперь физически представить себе математическое ожидание и дисперсию случайной величины \mathbf{x} . Рассмотрим изменение некоторой величины (например, длины x_0 мелкого кристалла с помощью микроскопа). Вследствие влияния различных факторов, таких, как несовершенство деталей микроскопа и ошибки наблюдений, повторные измерения будут давать несколько различные результаты \mathbf{x} . Однако отдельные измерения имеют тенденцию к расположению в окрестности истинного значения измеряемой величины, т. е. более вероятно получить значение случайной величины \mathbf{x} вблизи точки x_0 , нежели вдали от нее, если только нет систематического смещения. Плотность вероятности случайной величины \mathbf{x} будет, следовательно, иметь колоколообразную форму, как показано на рис. 3.3, хотя она не обязательно должна быть симметричной. Кажется разумным — особенно в случае симметричной плотности вероятности — интерпретировать математи-

ческое ожидание (3.4) как истинное значение измеряемой величины. Интересно заметить, что (3.4) есть математическое выражение координаты центра тяжести, т. е. мы можем представить себе \hat{x} как абсциссу центра тяжести массы, расположенной под кривой, являющейся плотностью вероятности $f(x)$.

Дисперсия (3.10)

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

является математическим выражением момента инерции и служит мерой ширины или рассеяния плотности вероятности вокруг среднего значения \hat{x} . Если она мала, то отдельные измерения лежат вблизи \hat{x} (рис. 3.3, а); если велика — они будут, вообще говоря, более удалены от среднего значения (рис. 3.3, б). Положительный квадратный корень из дисперсии

$$\sigma = +\sqrt{\sigma^2(\mathbf{x})} \quad (3.11)$$

называется *стандартным отклонением* случайной величины \mathbf{x} . Как и дисперсия, стандартное отклонение \mathbf{x} представляет собой меру среднего отклонения измерений случайной величины \mathbf{x} от математического ожидания.

Так как стандартное отклонение имеет ту же размерность, что и \mathbf{x} (в нашем примере они имеют размерность длины), его можно взять в качестве характеристики *ошибки* измерения

$$\sigma(\mathbf{x}) = \Delta x.$$

Такое определение ошибки измерения более подробно обсуждается в гл. 5, разд. 5.12. Следует заметить, что определения (3.4) и (3.10) не дают прямого метода вычисления ошибки измерения, так как плотность вероятности, описывающая измерение, обычно неизвестна.

Третий момент относительно среднего значения μ_3 иногда называют *асимметрией*. Мы же будем называть *асимметрией* случайной величины \mathbf{x} безразмерный параметр

$$\gamma = \mu_3/\sigma^3. \quad (3.12)$$

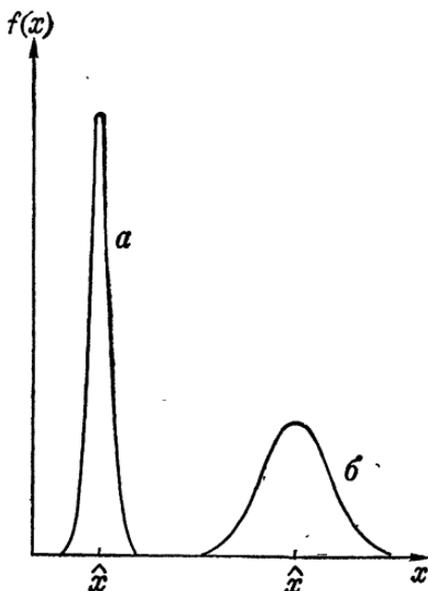


Рис. 3.3. Распределение с малой дисперсией (а) и с большой дисперсией (б).

Он несет информацию о возможном различии между положительным и отрицательным отклонением от среднего значения. Для симметричных распределений γ обращается в нуль.

Выведем теперь несколько важных свойств, относящихся к математическому ожиданию и дисперсии. В случае

$$H(\mathbf{x}) = c\mathbf{x}, \quad c = \text{const}, \quad (3.13)$$

сразу же получаем

$$\begin{aligned} E(c\mathbf{x}) &= cE(\mathbf{x}), \\ \sigma^2(c\mathbf{x}) &= c^2\sigma^2(\mathbf{x}) \end{aligned} \quad (3.14)$$

и, следовательно,

$$\sigma^2(\mathbf{x}) = E\{(\mathbf{x} - \hat{x})^2\} = E\{\mathbf{x}^2 - 2\mathbf{x}\hat{x} + \hat{x}^2\} = E(\mathbf{x}^2) - \hat{x}^2. \quad (3.15)$$

Рассмотрим функцию специального вида

$$\mathbf{u} = \frac{\mathbf{x} - \hat{x}}{\sigma(\mathbf{x})}. \quad (3.16)$$

Она имеет математическое ожидание

$$E(\mathbf{u}) = \frac{1}{\sigma(\mathbf{x})} E(\mathbf{x} - \hat{x}) = \frac{1}{\sigma(\mathbf{x})} (\hat{x} - \hat{x}) = 0 \quad (3.17)$$

и дисперсию

$$\sigma^2(\mathbf{u}) = \frac{1}{\sigma^2(\mathbf{x})} E\{(\mathbf{x} - \hat{x})^2\} = \frac{\sigma^2(\mathbf{x})}{\sigma^2(\mathbf{x})} = 1. \quad (3.18)$$

Функция \mathbf{u} , которая также является случайной величиной, обладает очень простыми свойствами, и поэтому ее предпочтительней использовать для вычислений. Такую случайную величину (имеющую математическое ожидание нуль и дисперсию, равную единице) мы будем называть *приведенной случайной величиной*. Ее также называют *стандартной, нормированной* или *безразмерной* случайной величиной.

Хотя распределение математически наиболее просто описывается своим математическим ожиданием, дисперсией и высшими моментами (практически любое распределение полностью определяется этими величинами, см. разд. 5.6), часто бывает удобно определить следующие величины, позволяющие лучше представить себе форму распределения. *Мода* x_m дискретного распределения определяется как значение случайной величины, соответствующее максимальной вероятности

$$P(\mathbf{x} = x_m) = \text{max}. \quad (3.19)$$

Если распределение имеет дифференцируемую плотность вероятности, то мода, соответствующая ее максимуму, легко определяется из условий

$$\frac{d}{dx} f(x) = 0, \quad \frac{d^2}{dx^2} f(x) < 0. \quad (3.20)$$

Во многих случаях существует только один максимум; тогда говорят, что распределение *унимодально*. В противном случае распределение называется *многовершинным*.

Медиана $x_{0,5}$ определяется как значение аргумента, при котором функция распределения вероятностей равна 0,5:

$$F(x_{0,5}) = P(\mathbf{x} < x_{0,5}) = 0,5. \quad (3.21)$$

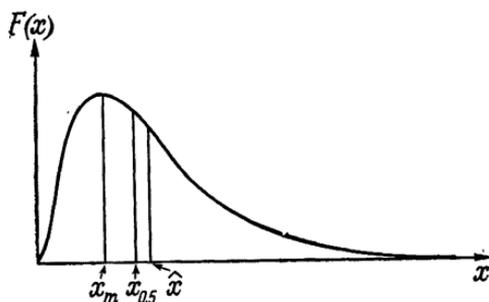


Рис. 3.4. Мода (x_m), среднее значение (\bar{x}) и медиана ($x_{0,5}$) в случае асимметричного распределения.

В случае непрерывной плотности вероятности выражение (3.21) принимает вид

$$\int_{-\infty}^{x_{0,5}} f(x) dx = 0,5, \quad (3.22)$$

т. е. медиана делит всю область изменения случайной величины на две области, в которые случайная величина попадает с равными вероятностями.

Из этих определений ясно, что в случае унимодального распределения с непрерывной плотностью вероятности, симметричной относительно своего максимума, значения среднего, моды и медианы совпадают. Однако для асимметричных распределений это не так (рис. 3.4).

Определение (3.21) легко может быть обобщено. Величины $x_{0,25}$ и $x_{0,75}$, определяемые выражениями

$$F(x_{0,25}) = 0,25, \quad F(x_{0,75}) = 0,75, \quad (3.23)$$

называются соответственно нижней и верхней *квартилью*. Аналогично можно определить *децили* $x_{0,1}$, $x_{0,2}$, ..., $x_{0,9}$ или, в более общем случае, *квантили* x_q с помощью формулы

$$F(x_q) = \int_{-\infty}^{x_q} f(x) dx = q, \quad (3.24)$$

где $0 < q < 1$. Определение квантилей проиллюстрировано на рис. 3.5. На графике функции распределения вероятностей квантиль x_q можно интерпретировать как абсциссу, соответствующую значению ординаты q . Зависимость квантилей x_q от вероятности q представляет собой функцию, обратную функции распределения вероятностей.

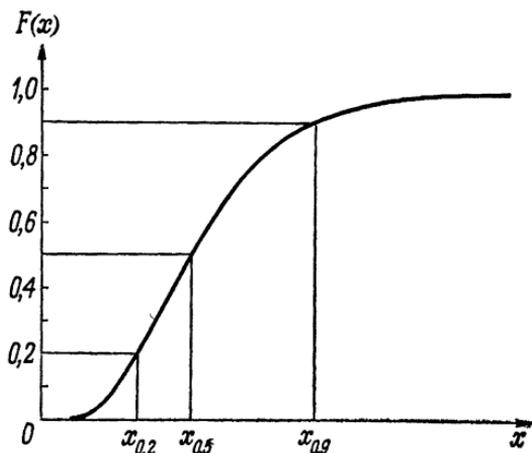


Рис. 3.5. Медиана и квантили для непрерывного распределения.

3.4. НЕРАВЕНСТВО ЧЕБЫШЕВА

Очевидно, случайная величина \mathbf{x} , как правило, принимает значения в некоторой окрестности среднего \hat{x} . С увеличением отклонения случайной величины от среднего вероятность значения, как правило, уменьшается. Этот факт находит свое количественное выражение в неравенстве Чебышева.

Если k — произвольное положительное действительное число, то вероятность абсолютного отклонения от среднего более чем на k стандартных отклонений будет меньше, чем k^{-2} :

$$P(|\mathbf{x} - \hat{x}| > k\sigma) < k^{-2}. \quad (4.1)$$

Докажем это неравенство для случая непрерывной случайной величины. Пусть $g(t)$ — плотность вероятности случайной величины $t = (\mathbf{x} - \hat{x})^2$. Тогда

$$P = P(|\mathbf{x} - \hat{x}| > k\sigma) = P((\mathbf{x} - \hat{x})^2 > k^2\sigma^2),$$

$$P = \int_{k^2\sigma^2}^{\infty} g(t) dt$$

и

$$\sigma^2 = E \{(\mathbf{x} - \hat{x})^2\} = \int_{-\infty}^{\infty} t g(t) dt, \quad (4.2)$$

$$\sigma^2 = \int_0^{k^2\sigma^2} t g(t) dt + \int_{k^2\sigma^2}^{\infty} t g(t) dt.$$

Так как интегрирование проводится только для положительных значений t и $g(t)$ как плотность вероятности неотрицательна, то величина каждого из интегралов в последнем равенстве не увеличится, если под знаком интеграла t заменить на значение, равное нижнему пределу интегрирования. Поэтому из выражения (4.2) сразу же получаем

$$\sigma^2 > k^2\sigma^2 \int_{k^2\sigma^2}^{\infty} g(t) dt = k^2\sigma^2 P,$$

что эквивалентно неравенству Чебышева (4.1).

В общем случае это неравенство довольно слабое. Фактически в большинстве интересных для практики случаев выполняется неравенство

$$P(|\mathbf{x} - \hat{x}| > k\sigma) \ll k^{-2}$$

(см., например, разд. 5.9).

РАСПРЕДЕЛЕНИЕ НЕСКОЛЬКИХ СЛУЧАЙНЫХ ВЕЛИЧИН

4.1. ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ И ПЛОТНОСТЬ ВЕРоятНОСТИ ДЛЯ ДВУХ СЛУЧАЙНЫХ ВЕЛИЧИН. УСЛОВНАЯ ВЕРоятНОСТЬ

Рассмотрим теперь две случайные величины \mathbf{x} и \mathbf{y} . Нас будет интересовать вероятность того, что одновременно $\mathbf{x} < x$ и $\mathbf{y} < y$. Так же, как и в случае одной случайной величины, мы предположим существование функции распределения (рис. 4.1)

$$F(x, y) = P(\mathbf{x} < x, \mathbf{y} < y). \quad (1.1)$$

Мы не будем вдаваться здесь в подробности аксиоматики и в условия существования F , так как эти условия всегда выполнены в случаях, представляющих практический интерес. Если F — дифференцируемая функция двух переменных, то назовем *совместной плотностью вероятности* функцию

$$f(x, y) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} F(x, y). \quad (1.2)$$

Тогда

$$P(a \leq x < b, c \leq y < d) = \int_a^b \int_c^d f(x, y) dx dy. \quad (1.3)$$

Часто мы сталкиваемся со следующей экспериментальной проблемой. В результате измерения получено множество значений пар случайных величин \mathbf{x}, \mathbf{y} . С помощью этих измерений мы приближенно вычисляем совместную функцию распределения $F(x, y)$, но нас интересует лишь поведение величины \mathbf{y} (без учета \mathbf{x}). (Например, плотность вероятности появления некоторого инфекционного заболевания может быть дана как функция времени года и географического положения. Для некоторых исследований зависимость от времени года может не представлять интереса.)

Проинтегрировав равенство (1.3) по всей области изменения y , получим

$$P(a \leq x < b, -\infty < y < \infty) = \int_a^b \left[\int_{-\infty}^{\infty} f(x, y) dy \right] dx = \int_a^b g(x) dx.$$

Здесь

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (1.4)$$

— плотность вероятности случайной величины \mathbf{x} . $g(x)$ называется *безусловным (маргинальным) распределением \mathbf{x}* . Соответствующее распределение для \mathbf{y} запишется так:

$$h(y) = \int_{-\infty}^{\infty} f(x, y) dx. \quad (1.5)$$

Аналогично понятию независимости событий [равенство (3.6) гл. 2] мы можем определить теперь *независимость случайных величин*. Случайные величины \mathbf{x} и \mathbf{y} называются независимыми, если

$$f(x, y) = g(x) h(y). \quad (1.6)$$

С помощью маргинального распределения можно определить также условную вероятность для \mathbf{y} , т. е. вероятность события $y \leq \mathbf{y} < y + dy$ при условии, что \mathbf{x} известно:

$$P(y \leq \mathbf{y} < y + dy | x \leq \mathbf{x} < x + dx). \quad (1.7)$$

Условная плотность вероятности по определению равна

$$f(y|x) = \frac{f(x, y)}{g(x)}. \quad (1.8)$$

Вероятность (1.7) записывается теперь в виде

$$f(y|x) dy.$$

Точно так же можно определить правило полной вероятности для распределений

$$h(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_{-\infty}^{\infty} f(y|x) g(x) dx. \quad (1.9)$$

Для независимых случайных величин, т. е. величин, для которых выполняется равенство (1.6), очевидно, справедливо

$$f(y|x) = \frac{f(x, y)}{g(x)} = \frac{g(x) h(y)}{g(x)} = h(y). \quad (1.10)$$

Этого следовало ожидать, так как в случае независимых случайных величин никакие сведения, относящиеся к одной из них, не увеличивают информации о распределении вероятности другой.

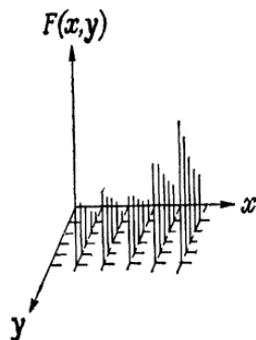


Рис. 4.1. Функция распределения двух случайных величин.

4.2. МАТЕМАТИЧЕСКИЕ ОЖИДАНИЯ, ДИСПЕРСИИ, КОВАРИАЦИИ И КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Аналогично равенству (3.5) гл. 3 определим математическое ожидание функции от случайных величин $H(\mathbf{x}, \mathbf{y})$:

$$E\{H(\mathbf{x}, \mathbf{y})\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x, y) f(x, y) dx dy \quad (2.1)$$

и дисперсию

$$\sigma^2\{H(\mathbf{x}, \mathbf{y})\} = E\{[H(\mathbf{x}, \mathbf{y}) - E(H(\mathbf{x}, \mathbf{y}))]^2\}. \quad (2.2)$$

В простом случае $H(\mathbf{x}, \mathbf{y}) = a\mathbf{x} + b\mathbf{y}$ формула (2.1) приводит к равенству

$$E(a\mathbf{x} + b\mathbf{y}) = aE(\mathbf{x}) + bE(\mathbf{y}). \quad (2.3)$$

Рассмотрим теперь функцию

$$H(\mathbf{x}, \mathbf{y}) = \mathbf{x}^l \mathbf{y}^m \quad (l, m — неотрицательные целые числа). \quad (2.4)$$

Математическое ожидание такой функции назовем lm -м моментом относительно нуля:

$$\lambda_{lm} = E(\mathbf{x}^l \mathbf{y}^m). \quad (2.5)$$

В более общем случае

$$H(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - a)^l (\mathbf{y} - b)^m \quad (2.6)$$

математическое ожидание

$$\alpha_{lm} = E\{(\mathbf{x} - a)^l (\mathbf{y} - b)^m\} \quad (2.7)$$

называется lm -м моментом относительно точки a, b . Особый интерес представляют моменты относительно точки $\lambda_{10}, \lambda_{01}$. Так же, как и в случае одной случайной величины, особенно важны младшие моменты*:

$$\begin{aligned} \mu_{00} &= \lambda_{00} = 1, \\ \mu_{10} &= \mu_{01} = 0; \\ \lambda_{10} &= E(\mathbf{x}) = \hat{x}, \\ \lambda_{01} &= E(\mathbf{y}) = \hat{y}; \\ \mu_{11} &= E\{(\mathbf{x} - \hat{x})(\mathbf{y} - \hat{y})\} = \text{cov}(\mathbf{x}, \mathbf{y}), \\ \mu_{20} &= E\{(\mathbf{x} - \hat{x})^2\} = \sigma^2(\mathbf{x}), \\ \mu_{02} &= E\{(\mathbf{y} - \hat{y})^2\} = \sigma^2(\mathbf{y}). \end{aligned} \quad (2.9)$$

* Момент μ_{11} называется ковариацией случайных величин \mathbf{x} и \mathbf{y} . — Прим. ред.

Теперь мы можем выразить дисперсию функции $(ax + by)$ через эти величины:

$$\begin{aligned}\sigma^2(ax + by) &= E \{[(ax + by) - E(ax + by)]^2\} = \\ &= E \{[a(x - \hat{x}) + b(y - \hat{y})]^2\} = \\ &= E \{a^2(x - \hat{x})^2 + b^2(y - \hat{y})^2 + 2ab(x - \hat{x})(y - \hat{y})\}, \\ \sigma^2(ax + by) &= a^2\sigma^2(x) + b^2\sigma^2(y) + 2ab \operatorname{cov}(x, y). \quad (2.10)\end{aligned}$$

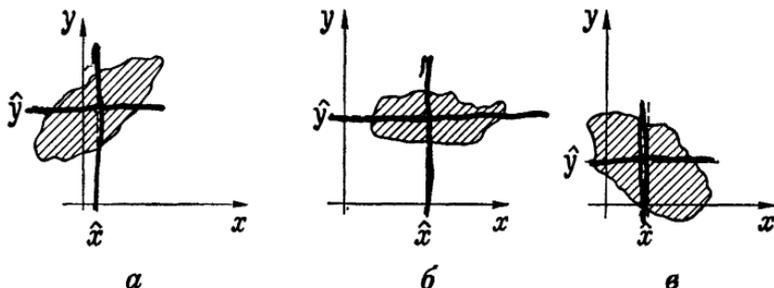


Рис. 4.2. Ковариация между случайными величинами.
а — $\operatorname{cov}(x, y) > 0$; б — $\operatorname{cov}(x, y) \approx 0$; в — $\operatorname{cov}(x, y) < 0$.

При выводе (2.10) мы использовали соотношение (3.14) гл. 3. В качестве другого примера рассмотрим

$$H(x, y) = xy. \quad (2.11)$$

Для получения математического ожидания допустим, что x и y независимы. Тогда

$$E(xy) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy g(x) h(y) dx dy = \left(\int_{-\infty}^{\infty} xg(x) dx \right) \left(\int_{-\infty}^{\infty} yh(y) dy \right). \quad (2.12)$$

Следовательно,

$$E(xy) = E(x)E(y). \quad (2.13)$$

Величины $E(x)$, $E(y)$, $\sigma^2(x)$, $\sigma^2(y)$ очень схожи с аналогичными величинами, полученными для одной переменной. Что же касается величины $\mu_{11} = \operatorname{cov}(x, y)$, называемой *ковариацией*, то она требует особого отношения ввиду значительной важности ее для понимания многих вещей, рассматриваемых ниже. Из определения видно, что ковариация положительна, если значения $x > \hat{x}$ ($x < \hat{x}$) появляются чаще со значениями $y > \hat{y}$ ($y < \hat{y}$), и отрицательна в противном случае. Если, наконец, знание значения x не дает информации о возможном значении y , то ковариация обращается в нуль. Рис. 4.2 иллюстрирует эти случаи в графической форме.

Во многих случаях вместо ковариации удобнее пользоваться коэффициентом корреляции:

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{\text{cov}(\mathbf{x}, \mathbf{y})}{\sigma(\mathbf{x})\sigma(\mathbf{y})}. \quad (2.14)$$

Ковариация и коэффициент корреляции дают грубую оценку взаимной зависимости \mathbf{x} и \mathbf{y} . Исследованием этого займемся ниже. Теперь же рассмотрим две случайные величины, преобразованные по формуле (3.16) гл. 3, и найдем дисперсию их суммы, пользуясь равенством (2.10):

$$\sigma^2(\mathbf{u} + \mathbf{v}) = \sigma^2(\mathbf{u}) + \sigma^2(\mathbf{v}) + 2\rho(\mathbf{u}, \mathbf{v})\sigma(\mathbf{u})\sigma(\mathbf{v}).$$

Согласно (3.18) гл. 3, имеем

$$\sigma^2(\mathbf{u}) = \sigma^2(\mathbf{v}) = 1.$$

Следовательно,

$$\sigma^2(\mathbf{u} + \mathbf{v}) = 2(1 + \rho(\mathbf{u}, \mathbf{v})) \quad (2.15)$$

и соответственно

$$\sigma^2(\mathbf{u} - \mathbf{v}) = 2(1 - \rho(\mathbf{u}, \mathbf{v})). \quad (2.16)$$

Но для любой случайной величины $\sigma^2 \geq 0$, поэтому

$$-1 \leq \rho(\mathbf{u}, \mathbf{v}) \leq 1.$$

Нетрудно показать, что

$$\rho(\mathbf{u}, \mathbf{v}) = \rho(\mathbf{x}, \mathbf{y}). \quad (2.17)$$

Итак, окончательно получаем

$$-1 \leq \rho(\mathbf{x}, \mathbf{y}) \leq 1. \quad (2.18)$$

Разберем теперь граничные случаи $\rho = \pm 1$. При $\rho(\mathbf{u}, \mathbf{v}) = 1$ дисперсия $\sigma^2(\mathbf{u} - \mathbf{v})$ равна нулю согласно (2.16), т. е. случайная величина $(\mathbf{u} - \mathbf{v})$ есть константа:

$$\mathbf{u} - \mathbf{v} = \frac{\mathbf{x} - \hat{x}}{\sigma(\mathbf{x})} - \frac{\mathbf{y} - \hat{y}}{\sigma(\mathbf{y})} = \text{const}. \quad (2.19)$$

Это равенство выполняется, если

$$\mathbf{y} = a + b\mathbf{x}, \quad (2.20)$$

где a — произвольная константа и b — любое положительное число. Следовательно, в случае линейной зависимости (b положительно) между \mathbf{x} и \mathbf{y} коэффициент корреляции равен $+1$. Соответственно можно показать, что при отрицательной линейной зависимости (b отрицательно) $\rho(\mathbf{x}, \mathbf{y}) = -1$.

Можно ожидать, что ковариация двух независимых случайных величин была равна нулю. Из (1.6) действительно получаем

$$\begin{aligned} \text{cov}(\mathbf{x}, \mathbf{y}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \hat{x})(y - \hat{y}) g(x) h(y) dx dy = \\ &= \left(\int_{-\infty}^{\infty} (x - \hat{x}) g(x) dx \right) \left(\int_{-\infty}^{\infty} (y - \hat{y}) h(y) dy \right) = \\ &= 0. \end{aligned}$$

4.3. СЛУЧАЙ БОЛЕЕ ЧЕМ ДВУХ СЛУЧАЙНЫХ ВЕЛИЧИН; ВЕКТОРНАЯ И МАТРИЧНАЯ ЗАПИСЬ

По аналогии с (1.1) мы определим функцию распределения n случайных величин $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ как

$$F(x_1, x_2, \dots, x_n) = P(\mathbf{X}_1 < x_1, \mathbf{X}_2 < x_2, \dots, \mathbf{X}_n < x_n). \quad (3.1)$$

Если функция F удовлетворяет определенным условиям и имеет частные производные по x_i , то совместная плотность вероятности равна

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F(x_1, x_2, \dots, x_n). \quad (3.2)$$

Маргинальное распределение

$$g_r(x_r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_{r-1} dx_{r+1} \dots dx_n \quad (3.3)$$

можно интерпретировать как плотность вероятности одной случайной величины \mathbf{X}_r . Если $H(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)$ — функция от случайных величин, то ее математическое ожидание равно

$$\begin{aligned} E\{H(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n)\} &= \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} H(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n. \end{aligned} \quad (3.4)$$

Для $H = \mathbf{X}_r$ получаем

$$\begin{aligned} E(\mathbf{X}_r) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_r f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n, \\ E(\mathbf{X}_r) &= \int_{-\infty}^{\infty} x_r g_r(x_r) dx_r. \end{aligned} \quad (3.5)$$

Случайные величины *независимы*, если

$$f(x_1, x_2, \dots, x_n) = g_1(x_1) g_2(x_2) \dots g_n(x_n). \quad (3.6)$$

Теперь можно построить *совместное маргинальное распределение* для любых l из n случайных величин*, производя интегрирование в (3.3) только по $n - l$ переменным

$$g(x_1, x_2, \dots, x_l) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_{l+1} \dots dx_n. \quad (3.7)$$

Эти l переменных *независимы*, если

$$g(x_1, x_2, \dots, x_l) = g_1(x_1) g_2(x_2) \dots g_l(x_l). \quad (3.8)$$

Моменты порядков l_1, \dots, l_n относительно нуля суть математические ожидания функций

$$H = x_1^{l_1} x_2^{l_2} \dots x_n^{l_n}, \quad (3.9)$$

т. е.

$$\lambda_{l_1 l_2 \dots l_n} = E(\mathbf{x}_1^{l_1} \mathbf{x}_2^{l_2} \dots \mathbf{x}_n^{l_n}), \quad (3.10)$$

в частности

$$\begin{aligned} \lambda_{100 \dots 0} &= E(\mathbf{x}_1) = \hat{x}_1, \\ \lambda_{010 \dots 0} &= E(\mathbf{x}_2) = \hat{x}_2, \\ \lambda_{000 \dots 1} &= E(\mathbf{x}_n) = \hat{x}_n. \end{aligned} \quad (3.11)$$

Моменты относительно $(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$ равны

$$\mu_{l_1 l_2 \dots l_n} = E\{(\mathbf{x}_1 - \hat{x}_1)^{l_1} (\mathbf{x}_2 - \hat{x}_2)^{l_2} \dots (\mathbf{x}_n - \hat{x}_n)^{l_n}\}. \quad (3.12)$$

Тогда дисперсии \mathbf{x}_i равны соответственно

$$\begin{aligned} \mu_{200 \dots 0} &= E\{(\mathbf{x}_1 - \hat{x}_1)^2\} = \sigma^2(\mathbf{x}_1), \\ \mu_{020 \dots 0} &= E\{(\mathbf{x}_2 - \hat{x}_2)^2\} = \sigma^2(\mathbf{x}_2), \\ &\vdots \\ \mu_{000 \dots 2} &= E\{(\mathbf{x}_n - \hat{x}_n)^2\} = \sigma^2(\mathbf{x}_n). \end{aligned} \quad (3.13)$$

При $l_i = l_j = 1$ и $l_k = 0$ ($i \neq k \neq j$) момент относительно $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$ является *ковариацией* между \mathbf{x}_i и \mathbf{x}_j :

$$c_{ij} = \text{cov}(\mathbf{x}_i, \mathbf{x}_j) = E\{(\mathbf{x}_i - \hat{x}_i)(\mathbf{x}_j - \hat{x}_j)\}. \quad (3.14)$$

Все эти свойства указывают на возможность представления случайных величин x_1, x_2, \dots, x_n как компонент вектора \mathbf{x} в n -мер-

* Мы можем считать без потери общности, что этими l величинами являются $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l$.

ном пространстве. Функция распределения (3.1) запишется теперь в виде

$$F = F(\mathbf{x}), \quad (3.15)$$

и соответственно плотность вероятности (3.2) как

$$f(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} F(\mathbf{x}). \quad (3.16)$$

Запись математического ожидания функции $H(\mathbf{x})$ упростится и будет иметь вид

$$E\{H(\mathbf{x})\} = \int H(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}. \quad (3.17)$$

Дисперсии и ковариации можно теперь собрать в виде матрицы, которая носит название *ковариационная матрица*:

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix}. \quad (3.18)$$

Здесь c_{ij} определяются с помощью (3.14). Диагональные элементы суть дисперсии $c_{ii} = \sigma^2(\mathbf{x}_i)$. Вследствие равенства

$$c_{ij} = c_{ji} \quad (3.19)$$

матрица симметрична. Если теперь написать математические ожидания \mathbf{x}_i в виде вектора

$$E(\mathbf{x}) = \hat{\mathbf{x}}, \quad (3.20)$$

то легко заметить, что каждый элемент ковариационной матрицы

$$c_{ij} = E\{(\mathbf{x}_i - \hat{x}_i)(\mathbf{x}_j - \hat{x}_j)\}$$

можно интерпретировать как математическое ожидание ij -го элемента произведения вектора-строки $(\mathbf{x} - \hat{\mathbf{x}})$ на вектор-столбец $(\mathbf{x} - \hat{\mathbf{x}})^T$. Записываются они следующим образом*:

$$\mathbf{x}^T = (x_1, x_2, \dots, x_n), \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Следовательно, ковариационная матрица равна

$$C = E\{(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T\}. \quad (3.21)$$

* Подробнее о матричной записи можно прочитать в приложении Б.

4.4. ПРЕОБРАЗОВАНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН

Как уже отмечалось в гл. 3, разд. 3, функция от случайной величины сама является случайной величиной:

$$y = y(x).$$

Возникает вопрос: какова плотность вероятности $g(y)$ случайной величины y , если известна плотность $f(x)$? Задача иллюстрирована на рис. 4.3. Очевидно, что вероятность $g(y) dy$ для y из

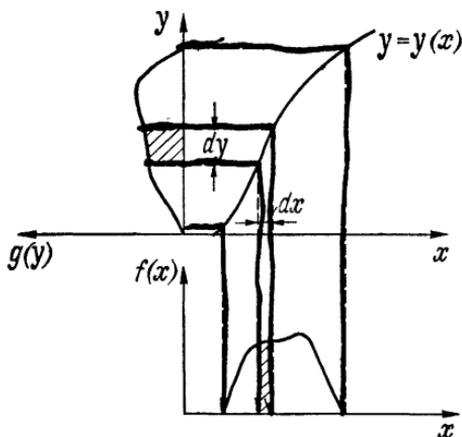


Рис. 4.3. Преобразование случайной величины x в y .

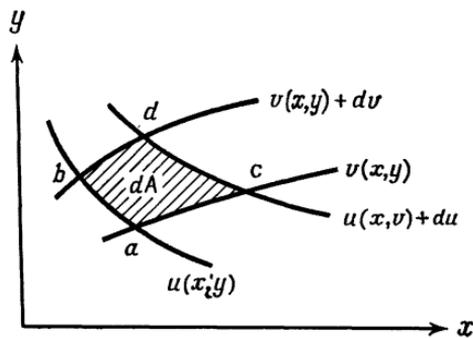


Рис. 4.4. Преобразование случайных величин x, y в u, v .

малого интервала dy равна $f(x) dx$, если x принадлежит «соответствующему» интервалу dx . Это соответствие показано на рис. 4.3. Имеем

$$dy = \left| \frac{dy}{dx} \right| dx \text{ или } dx = \left| \frac{dx}{dy} \right| dy.$$

Абсолютная величина здесь берется потому, что интервалы dx, dy не имеют направлений. Только при таком условии вероятности $f(x) dx, g(y) dy$ будут всегда положительны. Таким образом, связь между двумя плотностями вероятности следующая:

$$g(y) = \left| \frac{dx}{dy} \right| f(x). \quad (4.1)$$

Легко видеть, что $g(y)$ определена формулой (4.1) только для однозначных функций $y = y(x)$.

Многозначные функции нужно рассматривать специальным образом. Например, если $y = \sqrt{x}$, то следует рассматривать только

ветвь $\mathbf{y} = +\sqrt{\mathbf{x}}$. Из равенства (4.1) также следует, что для преобразованной плотности вероятности сохраняется равенство

$$\int_{-\infty}^{\infty} g(y) dy = \int_{-\infty}^{\infty} f(x) dx = 1.$$

Рассмотрим далее преобразование двух независимых переменных в новые:

$$\mathbf{u} = \mathbf{u}(\mathbf{x}, \mathbf{y}), \quad \mathbf{v} = \mathbf{v}(\mathbf{x}, \mathbf{y}). \quad (4.2)$$

Попробуем найти такую функцию J , с помощью которой можно перейти от плотности вероятностей $f(x, y)$ к новой $\mathfrak{F}(u, v)$:

$$g(u, v) = f(x, y) \left| J \left(\frac{x, y}{u, v} \right) \right|. \quad (4.3)$$

На плоскости x, y (рис. 4.4) изображены две пары кривых, соответствующих $u = \text{const}$, и кривые $v = \text{const}$. Они ограничивают элемент площади, который соответствует элементу площади $dx dy$ при отображении (4.2). Так как мы рассматриваем бесконечно малый элемент $dx dy$, то можно принять площадь dA (рис. 4.4) равной площади параллелограмма с вершинами a, b, c, d . Координаты первых трех вершин равны соответственно

$$\begin{aligned} x_a &= x(u, v), & y_a &= y(u, v), \\ x_b &= x(u, v + dv), & y_b &= y(u, v + dv), \\ x_c &= x(u + du, v), & y_c &= y(u + du, v). \end{aligned}$$

Разлагая эти функции в ряд Тейлора, получим (отбросив члены высших порядков)

$$\begin{aligned} x_b &= x(u, v) + \frac{\partial x}{\partial v} dv, & y_b &= y(u, v) + \frac{\partial y}{\partial v} dv, \\ x_c &= x(u, v) + \frac{\partial x}{\partial u} du, & y_c &= y(u, v) + \frac{\partial y}{\partial u} du. \end{aligned}$$

Площадь dA параллелограмма с точностью до знака, который здесь несуществен, так как в (4.3) входит модуль, равна определителю

$$dA = \begin{vmatrix} 1 & x_a & y_a \\ 1 & x_b & y_b \\ 1 & x_c & y_c \end{vmatrix},$$

т. е.

$$dA = \frac{\partial x}{\partial u} du \frac{\partial y}{\partial v} dv - \frac{\partial y}{\partial u} du \frac{\partial x}{\partial v} dv,$$

или в виде определителя второго порядка

$$dA = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} du dv = J \left(\begin{matrix} x, y \\ u, v \end{matrix} \right) du dv. \quad (4.4)$$

Определитель

$$J \left(\begin{matrix} x, y \\ u, v \end{matrix} \right) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} \quad (4.5)$$

называется *якобианом* преобразования (4.2).

В общем случае n переменных $\mathbf{x} = (x_1, x_2, \dots, x_n)$ и преобразования

$$\begin{aligned} y_1 &= y_1(\mathbf{x}), \\ y_2 &= y_2(\mathbf{x}), \\ &\dots \\ y_n &= y_n(\mathbf{x}) \end{aligned} \quad (4.6)$$

плотность вероятности преобразованных величин равна

$$g(\mathbf{y}) = J \left(\begin{matrix} \mathbf{x} \\ \mathbf{y} \end{matrix} \right) f(\mathbf{x}). \quad (4.7)$$

Якобиан в этом случае равен

$$J \left(\begin{matrix} \mathbf{x} \\ \mathbf{y} \end{matrix} \right) = J \left(\begin{matrix} x_1, x_2, \dots, x_n \\ y_1, y_2, \dots, y_n \end{matrix} \right) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} & \dots & \frac{\partial x_n}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} & \dots & \frac{\partial x_n}{\partial y_2} \\ \dots & \dots & \dots & \dots \\ \frac{\partial x_1}{\partial y_n} & \frac{\partial x_2}{\partial y_n} & \dots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}. \quad (4.8)$$

Якобиан существует, конечно, при условии существования и единственности всех частных производных, входящих в него.

4.5. ЛИНЕЙНЫЕ И ОРТОГОНАЛЬНЫЕ ПРЕОБРАЗОВАНИЯ; РАСПРОСТРАНЕНИЕ ОШИБОК

На практике часто приходится иметь дело с линейными преобразованиями случайных величин. Причиной этого является простота обращения с ними. Поэтому по возможности мы стараемся перейти от нелинейных преобразований к линейным, используя лишь линейный член разложения в ряд Тейлора.

где

$$T = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \dots & \frac{\partial y_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial y_r}{\partial x_1} & \frac{\partial y_r}{\partial x_2} & \dots & \frac{\partial y_r}{\partial x_n} \end{pmatrix}. \quad (5.6)$$

Если пренебречь членами выше первого порядка и ввести матрицу T , содержащую частные производные, в (5.4), то получим закон распространения ошибок. Заметим, в частности, что ошибки для y (т. е. диагональные элементы C_y) зависят не только от ошибок (или дисперсий) x , но и от ковариаций между разными x_i . Нельзя полагаться на точность результатов, не учитывая ковариаций, так как они вносят существенный вклад в ошибку.

Пренебречь ковариациями можно только в случае взаимной независимости x_i , т. е. тогда, когда они обращаются в нуль. При этом матрица C_x принимает диагональный вид. Диагональные элементы матрицы C_y в этом случае имеют простой вид:

$$\sigma^2(y_i) = \sum_{j=1}^n \left(\frac{\partial y_i}{\partial x_j} \right)_{x=\hat{x}}^2 \sigma^2(x_j). \quad (5.7)$$

Если теперь стандартное отклонение, т. е. положительный квадратный корень из дисперсии, обозначить через Δ , то получим формулу

$$\Delta y_i = \sqrt{\sum_{j=1}^n \left(\frac{\partial y_i}{\partial x_j} \right)^2 (\Delta x_j)^2}, \quad (5.8)$$

известную как закон распространения ошибок. Подчеркнем, что выражение (5.8) корректно только в случае нулевых ковариаций. Проиллюстрируем это на следующем примере:

Пример 4.1. Координаты точки x, y измеряются в декартовой системе с помощью координатографа, который дает y с ошибкой в три раза большей, чем ошибка x . Измерения x и y независимы. Следовательно, можно написать ковариационную матрицу (с точностью до множителя

$$C_{xy} = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix}.$$

Найдем теперь матрицу ошибок в полярных координатах:

$$r = \sqrt{x^2 + y^2}, \quad \varphi = \arctg \frac{y}{x}.$$

Матрица преобразования (5.6) для этого случая равна

$$T = \begin{pmatrix} \frac{x}{r} & \frac{y}{r} \\ -\frac{y}{r^2} & \frac{x}{r^2} \end{pmatrix}.$$

Для упрощения вычислений возьмем конкретную точку $(x, y) = (1, 1)$. Тогда

$$T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

и, следовательно,

$$C_{r\varphi} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{2} \\ \frac{1}{\sqrt{2}} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 5 & \frac{4}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} & 5 \end{pmatrix}.$$

Теперь мы можем вернуться к первоначальной декартовой системе координат

$$x = r \cos \varphi, \quad y = r \sin \varphi$$

с помощью матрицы преобразования

$$T' = \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -1 \\ \frac{1}{\sqrt{2}} & 1 \end{pmatrix}.$$

Получаем

$$C_{xy} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -1 \\ \frac{1}{\sqrt{2}} & 1 \end{pmatrix} \begin{pmatrix} 5 & \frac{4}{\sqrt{2}} \\ \frac{4}{\sqrt{2}} & 5 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 9 \end{pmatrix},$$

что и следовало ожидать. Если бы мы для вычисления C'_{xy} использовали формулу (5.8), т. е. пренебрегли ковариациями при преобразовании r, φ в x, y , то результат был бы

$$C'_{xy} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -1 \\ \frac{1}{\sqrt{2}} & 1 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix},$$

т. е. сильно отличался бы от первоначальной матрицы. Этот пример подчеркивает важность учета ковариаций при распространении ошибок.

В заключение мы обсудим особый вид линейного преобразования. Пусть n функций y зависят от n переменных x . Потребуем, чтобы $a = 0$ в (5.2). Тогда имеем

$$y = Rx, \quad (5.9)$$

где R — квадратная матрица. Потребуем теперь, чтобы модуль вектора был инвариантен относительно преобразования (5.9), т. е.

$$y^2 = \sum_{i=1}^n y_i^2 = x^2 = \sum_{i=1}^n x_i^2. \quad (5.10)$$

С помощью (В 2.20) можно записать

$$y^T y = (Rx)^T (Rx) = x^T R^T R x = x^T x.$$

Для этого необходимо, чтобы

$$R^T R = I,$$

или, в компонентах матрицы,

$$\sum_{i=1}^n r_{ik} r_{il} = \delta_{kl} = \begin{cases} 0 & \text{при } i \neq k, \\ 1 & \text{при } i = k. \end{cases} \quad (5.11)$$

Преобразование типа (5.9), удовлетворяющее условиям (5.11), называется *ортогональным*.

Рассмотрим определитель матрицы преобразования

$$D = \begin{vmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & & & \\ r_{n1} & r_{n2} & \dots & r_{nn} \end{vmatrix}$$

и его квадрат. Пользуясь правилами алгебраических действий с определителями из (5.11), получаем

$$D^2 = \begin{vmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \\ \vdots & & & \\ 0 & 0 & \dots & 1 \end{vmatrix},$$

т. е. $D = \pm 1$. Определитель D является также якобианом преобразования (5.9):

$$J\left(\frac{y}{x}\right) = \pm 1, \quad (5.12)$$

Если равенство (5.9) умножим слева на R^T , то получим

$$R^T \mathbf{y} = R^T R \mathbf{x},$$

или вследствие (5.11)

$$\mathbf{x} = R^T \mathbf{y}, \quad (5.13)$$

т. е. матрица преобразования, обратного к ортогональному, просто равна транспонированной матрице.

Важное свойство линейных преобразований вида

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

заключается в следующем. Если удовлетворяется условие

$$\sum_{i=1}^n r_{1i}^2 = 1,$$

можно добавить к y_1 такую систему функций y_2, y_3, \dots, y_n аналогичного вида, что в результате получится ортогональное преобразование.

НЕКОТОРЫЕ ВАЖНЫЕ РАСПРЕДЕЛЕНИЯ И ТЕОРЕМЫ

Обсудим теперь некоторые специальные распределения. Эту главу можно рассматривать, следовательно, как собрание примеров. Впрочем, эти распределения имеют большую практическую важность. Кроме того, изучение их приведет нас к ряду важных теорем.

5.1. БИНОМИАЛЬНОЕ И ПОЛИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЯ

Пусть некоторый эксперимент может приводить к одному из двух несовместных исходов A, \bar{A} . В этом случае он характеризуется разложением

$$E = A + \bar{A}. \quad (1.1)$$

Вероятности исходов обозначим соответственно

$$P(A) = p, \quad P(\bar{A}) = 1 - p = q. \quad (1.2)$$

Результат такого эксперимента можно выразить случайной величиной x_i , принимающей значение 1 (0), если произошло событие A (\bar{A}). Индекс i — номер эксперимента в серии. Повторим эксперимент n раз и рассмотрим распределение вероятностей случайной величины

$$x = \sum_{i=1}^n x_i.$$

Вероятность того, что первые k экспериментов имеют исход A , а остальные — исход \bar{A} , равна, согласно (3.8) гл. 2,

$$p^k q^{n-k}.$$

Событие, состоящее в появлении исхода A k раз в n экспериментах, может осуществляться

$$\binom{n}{k} = n! / [k! (n - k)!]$$

различными способами* в зависимости от порядка следования A и \bar{A} в серии. Таким образом, вероятность этого события равна

$$W_k^n = \binom{n}{k} p^k q^{n-k}. \quad (1.3)$$

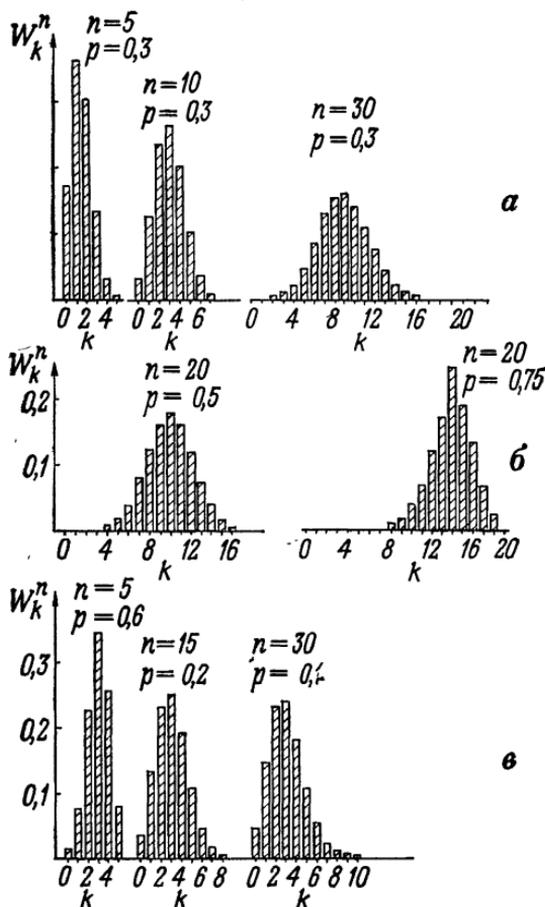


Рис. 5.1. Биномиальное распределение. а — для фиксированного p и различных значений n ; б — для фиксированного n и различных значений p ; в — для различных значений n и p , таких, что $n \cdot p = \text{const}$.

Рассмотрим математическое ожидание и дисперсию x . Определим эти величины сначала для случайной величины x_i , соответствующей одному эксперименту. Согласно (3.2) гл. 3, мы имеем

$$E(x_i) = 1 \cdot p + 0 \cdot q \quad (1.4)$$

* См. приложение В.

и

$$\begin{aligned}\sigma^2(\mathbf{x}_i) &= E\{(x_i - p)^2\} = (1 - p)^2 p + (0 - p)^2 q, \\ \sigma^2(\mathbf{x}_i) &= pq.\end{aligned}\quad (1.5)$$

Обобщая формулу (2.3) гл. 4, получаем для $\mathbf{x} = \sum \mathbf{x}_i$:

$$E(\mathbf{x}) = \sum_{i=1}^n p = np. \quad (1.6)$$

Из (2.10) гл. 4 получаем

$$\sigma^2(\mathbf{x}) = npq, \quad (1.7)$$

так как из-за независимости \mathbf{x}_i все ковариации обращаются в нуль.

Распределение W_k^n показано на рис. 5.1, а для фиксированного p и различных n и на рис. 5.1, б для фиксированного n и различных p . Наконец, на рис. 5.1, в n и p меняются оба так, что произведение их остается постоянным. Эти рисунки помогут нам обнаружить сходство между биномиальным распределением (названным так потому, что каждый эксперимент имеет два возможных исхода) и другими распределениями.

Дальнейшее обобщение биномиального распределения относится к экспериментам, имеющим более чем два возможных исхода. Вместо (1.1) имеем

$$E = A_1 + A_2 + \dots + A_l. \quad (1.8)$$

Пусть вероятности каждого из взаимно несовместных событий A_j заданы:

$$P(A_j) = p_j, \quad \sum_{j=1}^l p_j = 1. \quad (1.9)$$

Произведем n испытаний. Вероятность того, что событие A_j имело место k_j раз ($j = 1, 2, \dots, l$), равна

$$W_{(k_1, k_2, \dots, k_l)}^n = \frac{n!}{\prod_{j=1}^l k_j!} \prod_{j=1}^l p_j^{k_j}. \quad (1.10)$$

Оставляем доказательство (1.10) читателю.

Распределение (1.10) называется полиномиальным распределением. Если положить \mathbf{x}_{ij} равным 1, когда в i -м эксперименте имел место исход A_j , и 0 в противном случае, то для

$$\mathbf{x}_i = \sum_{j=1}^l \mathbf{x}_{ij}$$

получим математическое ожидание

$$E(\mathbf{x}_j) = \hat{x}_j = np_j. \quad (1.11)$$

Элементы ковариационной матрицы для \mathbf{x}_j равны*

$$c_{ij} = np_i(\delta_{ij} - p_j). \quad (1.12)$$

Мы получили, что внедиагональные элементы этой матрицы не равны нулю. Этого и следовало ожидать, так как случайные величины \mathbf{x}_j не являются независимыми вследствие (1.9).

5.2. ЧАСТОТА; ЗАКОН БОЛЬШИХ ЧИСЕЛ

Обычно вероятности различных типов событий (например, p_j для полиномиального распределения) не известны и должны быть найдены из эксперимента. Определим частоту событий A_j в n испытаниях:

$$\mathbf{h} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_{ij} = \frac{1}{n} \mathbf{x}_j. \quad (2.1)$$

В отличие от вероятности частота является случайной величиной, так как она зависит от результата данных n экспериментов. Используя (1.11), (1.12) и (3.4) гл. 3, получаем

$$E(\mathbf{h}) = \hat{\mathbf{h}} = E\left(\frac{\mathbf{x}_j}{n}\right) = p_j \quad (2.2)$$

и

$$\sigma^2(\mathbf{h}) = \sigma^2\left(\frac{\mathbf{x}_j}{n}\right) = \frac{1}{n^2} \sigma^2(\mathbf{x}_j) = \frac{1}{n} p_j(1 - p_j). \quad (2.3)$$

Произведение $p_j(1 - p_j)$ в равенстве (2.3) не превосходит $\frac{1}{4}$. Мы получили, таким образом, что математическое ожидание частоты события равно вероятности его появления, а стандартное отклонение стремится к нулю с ростом n . При фиксированном n оно не превосходит $1/2\sqrt{n}$. Это свойство частоты обычно называется *законом больших чисел*** . Закон больших чисел служит обоснованием частотного определения вероятности (2.1) гл. 2.

Часто целью экспериментальных исследований является лишь получение вероятности появления некоторого события. Согласно (2.2), можно пользоваться частотой как приближенным значением

* В (1.12) $\delta_{ij} = 1$ при $i = j$, $\delta_{ij} \neq 0$ при $i \neq j$. — Прим. ред.

** Точнее, закон больших чисел выражают в виде предела $P(|h - p_j| > \epsilon) \rightarrow 0$ при $n \rightarrow \infty$ для любого $\epsilon > 0$. Это утверждение следует из (2.2), (2.3) и неравенства Чебышева (4.1) гл. 3. — Прим. ред.

для вероятности. Квадрат ошибки такого приближения обратно пропорционален количеству экспериментов. Такая ошибка, являющаяся лишь следствием конечности числа экспериментов, называется *статистической ошибкой*. Это понятие особенно важно для приложений, имеющих дело с подсчетом каких-либо событий, например подсчет ядерных частиц, проходящих через счетчик, животных с определенной чертой наследственности, дефектных деталей в контрольной выборке и т. д.

Пример 5.1. Пусть из предварительных экспериментов известно, что если подвергнуть облучению определенной дозой рентгеновских лучей какое-то количество дрозофил, то у R -й доли, составляющей приблизительно $1/200$ часть выборки, будет развиваться некоторое свойство A . Требуется с помощью эксперимента найти R с точностью 1% . Как велика должна быть первоначальная выборка, чтобы эта точность была достигнута? Положим $p_i = 0,005$, $(1 - p_i) \approx 1$ и воспользуемся равенством (2.3) для определения такого n , чтобы $\sigma(h)/h = 200\sigma(h) = 0,01$, или $\sigma(h) = 0,00005$ и $\sigma^2(h) = 0,25 \cdot 10^{-8}$. Равенство (2.3) дает

$$0,25 \cdot 10^{-8} = \frac{1}{n} \times 0,005,$$

откуда $n = 2 \cdot 10^6$, т. е. мы должны были бы взять 2 млн. дрозофил, что практически невозможно. Определение R с точностью 10% потребовало бы 20 тыс. дрозофил.

5.3. ГИПЕРГЕОМЕТРИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ

Ниже мы дадим точное определение случайной выборки, а пока разберем типичную задачу с этим понятием. Рассмотрим ящик (в математике принято называть его *урной*) с K белыми и $L = N - K$ черными шарами. Требуется найти вероятность того, что из n вынутых (без возвращения) шаров k окажутся белыми и $l = n - k$ — черными. Задача усложняется тем фактом, что выбор шара какого-либо цвета изменяет соотношение между числом оставшихся белых и черных шаров и поэтому влияет на исход следующего выбора. Следовательно, в данном случае нельзя рассуждать так же, как в случае биномиального распределения. Известно, однако, что число различных способов выборки n шаров из N равно $\binom{N}{n}$, и поэтому вероятность каждой из этих выборок $1/\binom{N}{n}$. Соответственно $\binom{K}{k}$ и $\binom{L}{l}$ — количества различных способов выбора k из K белых шаров и l из L черных.

Отсюда получаем искомую вероятность

$$W_k = \frac{\binom{K}{k} \binom{L}{l}}{\binom{N}{n}}. \quad (3.1)$$

Так же, как в разд. 5.1, определим случайную величину

$$\mathbf{x} = \sum_{i=1}^n \mathbf{x}_i,$$

где $x_i = 1$ (0), если i -й шар оказался белым (черным). Другими словами, мы определяем k как случайную величину \mathbf{x} .

Чтобы вычислить математическое ожидание \mathbf{x} , возвратимся к определению (3.2) гл. 3. Получаем

$$\begin{aligned} E(\mathbf{x}) &= \frac{1}{\binom{N}{n}} \sum_{i=1}^n i \binom{K}{i} \binom{N-K}{n-i} = \\ &= \frac{(N-n)! n!}{N!} \sum_{i=1}^n \frac{i K! (N-K)!}{i! (K-i)! (n-i)! (N-K-n+i)!} = \frac{n(n-1)! (N-n)!}{N(N-1)!} \times \\ &\times \sum_{i=1}^n \frac{K! (N-K)!}{(i-1)! [K-1-(i-1)]! (n-1-(i-1))! [N-K-n+1+(i-1)]!}. \end{aligned}$$

Обозначим $i-1 = j$; тогда

$$\begin{aligned} E(\mathbf{x}) &= n \frac{K}{N} \frac{(n-1)! (N-n)!}{(N-1)!} \times \\ &\times \sum_{j=0}^{n-1} \frac{(K-1)! (N-K)!}{j! (K-1-j)! (n-1-j)! (N-K-n+1+j)!} = \\ &= n \frac{K}{N} \frac{1}{\binom{N-1}{n-1}} \sum_{j=0}^{n-1} \binom{K-j}{j} \binom{N-K}{n-1-j}. \end{aligned}$$

Воспользовавшись равенством (B.5), окончательно получаем

$$E(\mathbf{x}) = n \frac{K}{N}. \quad (3.2)$$

Вычисление дисперсии проводится тем же способом, но требует длинных выкладок. Приведем результат:

$$\sigma^2(\mathbf{x}) = \frac{nK(N-K)(N-n)}{N^2(N-1)}. \quad (3.3)$$

Несколько примеров гипергеометрического распределения (3.1) показано на рис. 5.2. В случае $n \ll N$ результат одной выборки незначительно влияет на исход следующей. Поэтому можно ожидать, что в этом случае W_k будет вести себя подобно биномиальному распределению с $p = K/N$ и $q = (N - K)/N$. Математическое ожидание при этом равно

$$E(\mathbf{x}) = n \frac{K}{N} = np,$$

так же как и для биномиального распределения. Дисперсия (3.3) может быть записана

$$\sigma^2(\mathbf{x}) = \frac{npq(N-n)}{N-1}$$

и для $n \ll N$ становится равной

$$\sigma^2(\mathbf{x}) = npq.$$

Эту связь биномиального и гипергеометрического распределений можно увидеть, сравнивая рис. 5.2 с левой частью рис. 5.1, а.

Гипергеометрическое распределение имеет разнообразные применения. Голосования, контроль качества и т. п. основаны на выборке объектов без возвращения из генеральной совокупности. Гипергеометрическое распределение можно обобщать далее в двух направлениях. Первое из них состоит в рассмотрении более чем двух объектов (черные шары, белые шары) разного качества. Это напоминает переход от биномиального распределения к полиномиальному. Генеральная совокупность (популяция) в этом случае состоит из N элементов, обладающих одним из l свойств:

$$N = N_1 + N_2 + \dots + N_l.$$

Вероятность обнаружить в выборке величины n_j ($j = 1, 2, \dots, l$) элементов n -го типа равна

$$W_{n_1, n_2, \dots, n_l} = \frac{\binom{N_1}{n_1} \binom{N_2}{n_2} \dots \binom{N_l}{n_l}}{\binom{N}{n}} \quad (3.4)$$

аналогично (3.1).

Другое обобщение гипергеометрического распределения получается следующим образом. Мы видели выше, что последователь-

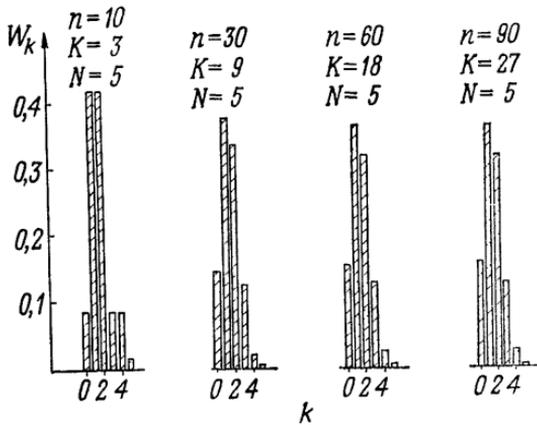


Рис. 5.2. Гипергеометрическое распределение.

ные выборки не являлись независимыми вследствие того, что шары не возвращались обратно. Если теперь каждый раз, когда вынимается шар одного из типов, добавлять в урну фиксированное число шаров того же типа, то зависимость одной выборки от предыдущей усиливается. В этом случае получаем *распределение Пойа*, имеющее важное значение для изучения возникновения эпидемий; здесь появление одного случая заболевания увеличивает вероятность аналогичных заболеваний.

Пример 5.2. Интересно применение гипергеометрического распределения для изучения зоологических популяций. Из водоема выловлены и отмечены каким-то образом K рыб. Затем они отпущены обратно в водоем. Через некоторое время пойманы n рыб, из которых k оказались отмеченными. Этот процесс описывается гипергеометрическим распределением, если предположить, что водоем содержит N рыб и из них K отмечены. Мы вернемся к этой задаче в примере 7.3 с целью оценки величины популяции N в водоеме.

5.4. РАСПРЕДЕЛЕНИЕ ПУАССОНА

Рассмотрение рис. 5.1, в наводит на мысль, что если $n \rightarrow \infty$ так, что произведение $np = \lambda$ остается постоянным, то биномиальное распределение стремится к некоторому фиксированному распределению. Перепишем равенство (1.3):

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

В пределе при больших n каждый сомножитель дроби, стоящей в числителе выражения, стремится к единице. Вследствие того что

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda},$$

окончательно получаем

$$\lim_{n \rightarrow \infty} W_k^n = f(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad (4.1)$$

Величина $f(k)$ дает вероятности распределения Пуассона. На рис. 5.3 она построена для разных значений λ . Этот рисунок можно сравнить с рис. 5.1, в. Так же, как и ранее рассмотренные распределения, распределение Пуассона определено только для целых

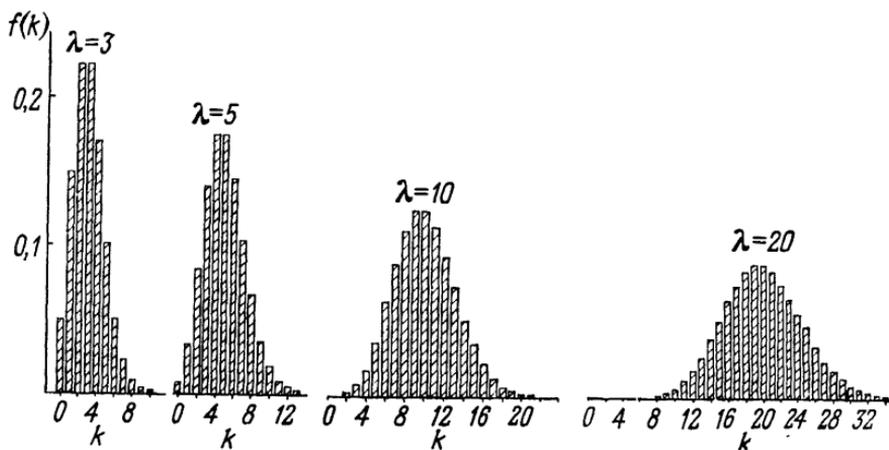


Рис. 5.3. Распределение Пуассона.

значений k . Полная вероятность, т. е. вероятность появления любого значения k , должна быть равна единице:

$$\begin{aligned} \sum_{k=0}^{\infty} f(k) &= \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} = \\ &= e^{-\lambda} \left(\lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right) = \\ &= e^{-\lambda} e^{\lambda} = 1, \end{aligned}$$

т. е.

$$\sum_{k=0}^{\infty} f(k) = 1. \quad (4.2)$$

Найдем теперь математическое ожидание, дисперсию и асимметрию распределения Пуассона. Из определения (3.2) гл. 3 имеем (используя обозначение \mathbf{x} для случайной величины k)

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

$$= \sum_{k=1}^{\infty} \frac{\lambda \lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda},$$

откуда, приняв во внимание (4.2), получаем

$$E(\mathbf{x}) = \lambda. \quad (4.3)$$

Рассмотрим далее $E(\mathbf{x}^2)$. Аналогичные выкладки дают:

$$\begin{aligned} E(\mathbf{x}^2) &= \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \\ &= \lambda \sum_{j=0}^{\infty} (j+1) \frac{\lambda^j}{j!} e^{-\lambda} = \lambda \left(\sum_{j=0}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda} + 1 \right), \end{aligned}$$

или

$$E(\mathbf{x}^2) = \lambda(\lambda + 1). \quad (4.4)$$

Математическое ожидание \mathbf{x}^3 равно

$$E(\mathbf{x}^3) = \lambda + 3\lambda^2 + \lambda^3. \quad (4.5)$$

Из (3.15) гл. 3, используя (4.3) и (4.4), получаем дисперсию \mathbf{x} :

$$\sigma^2(\mathbf{x}) = E(\mathbf{x}^2) - \{E(\mathbf{x})\}^2 = \lambda(\lambda + 1) - \lambda^2,$$

т. е.

$$\sigma^2(\mathbf{x}) = \lambda. \quad (4.6)$$

Асимметрия (3.12) гл. 3 равна

$$\gamma = \frac{\mu_3}{\sigma^3} = \frac{\lambda}{\lambda^{3/2}} = \lambda^{-1/2}, \quad (4.7)$$

т. е. распределение Пуассона становится более симметричным с возрастанием λ . Это иллюстрирует рис. 5.3.

Распределение Пуассона получается из биномиального распределения при большом n и малом p , поскольку $\lambda = np$ — фиксированная величина. Следовательно, его можно применять к процессам с большим количеством исходов, из которых только малая часть обладает некоторыми интересующими нас свойствами (т. е. много испытаний, но мало «успехов»).

Пример 5.3. Рассмотрим очень большое число n атомных ядер в радиоактивном источнике и найдем вероятность $\lambda(n, \Delta t)$ того, что одно из ядер распадется за интервал времени Δt . Вероятность не изменится, если умножить n на константу c и рассматривать теперь распады за интервал времени $\Delta t/c$. Причиной этому служит то, что распады ядер происходят независимо и поэтому $\lambda(n, \Delta t)$ равна $np(\Delta t)$, где $p(\Delta t)$ — вероятность распада отдельного ядра за время Δt . [Мы рассматриваем очень малые элементы времени

по сравнению с периодом полураспада ядра. Поэтому величина $p(\Delta t)$ прямо пропорциональна Δt .] Следовательно, можно ожидать, что частота f_k ($k = 0, 1, 2, \dots$) распада ядер за некоторый промежуток времени подчиняется закону Пуассона. И наоборот, мы можем утверждать, что если частота подчиняется закону Пуассона, то индивидуальные распады происходят независимо друг от друга. Таким путем была установлена статистическая природа распада известными экспериментами Резерфорда и Гейгера. Фактические цифровые данные этих экспериментов приведены во многих учебниках по статистике (см., например, Фримен [8], Фиш [6]).

Аналогично число звезд на небольшом участке неба или число изюминок в кусочке из целого кекса распределены по закону Пуассона, но, например, частота появления k животных в данном районе, если животные перемещаются стадами, не распределена по Пуассону, так как в этом случае не удовлетворяется условие независимости.

В качестве количественного примера распределения Пуассона во многих учебниках рассматривается число прусских кавалеристов, убитых ударом копыта лошади за период 20 лет. Этот пример впервые был приведен Борткевичем [4]. Мы предпочитаем обратиться к менее мрачным примерам, заимствованным из лекций Прайса [21].

Пример 5.4. Автор сначала строит модель яблони с 1000 яблоками и 1000 сборщиками с завязанными глазами, которые пытаются одновременно сорвать по яблоку. Поскольку мы имеем дело с моделью, то можно предположить, что они не отталкивают друг друга, но может случиться, что два или несколько из них схватят одно и то же яблоко. Число яблок, каждое из которых схвачено одновременно k сборщиками, подчиняется тогда распределению Пуассона. Прайс также сообщает, что количество научных открытий, сделанных независимо дважды, трижды и т. д., подчинено закону Пуассона (табл. 5.1). Он делает вывод, что

Таблица 5.1

Распределение Пуассона и одновременные открытия

Число одновременных открытий	Случай одновременного открытия	Предсказание распределения Пуассона
0	Неопределенное	368
1	Нет данных	368
2	179	184
3	51	61
4	17	15
5	6	3
≥6	8	1

ученых, подобно слепым сборщикам яблок, мало заботит деятельность коллег, так как «ученые испытывают сильное побуждение писать статьи, но относительно слабое стремление читать их».

Вернемся теперь к связи биномиального и пуассоновского распределений, которую мы начали обсуждать в этом разделе. Можно полагать, что даже если число экспериментов не бесконечно, эти два распределения очень похожи. Один из примеров продемонстрирован в табл. 5.2. Следовательно, при отсутствии таблиц биномиального распределения (что было обычным до появления электронных машин) для достаточно больших n и малых p можно использовать распределение Пуассона с $\lambda = np$ в случаях, когда, строго говоря, справедливо только биномиальное распределение.

Таблица 5.2

Сравнение биномиальных распределений
при $np = 3$ и распределения Пуассона с $\lambda = 3$

k	Биномиальное распределение W_k^n						Распределение Пуассона $\lambda = 3$
	$p = 0,5$ $n = 6$	$p = 0,2$ $n = 15$	$p = 0,1$ $n = 30$	$p = 0,05$ $n = 60$	$p = 0,02$ $n = 150$	$p = 0,01$ $n = 300$	
0	0,0156	0,0352	0,0424	0,0461	0,0483	0,0490	0,0498
1	0,0937	0,1319	0,1413	0,1455	0,1478	0,1486	0,1494
2	0,2344	0,2309	0,2276	0,2259	0,2248	0,2244	0,2240
3	0,3125	0,2501	0,2361	0,2298	0,2263	0,2252	0,2240
4	0,2344	0,1876	0,1771	0,1724	0,1697	0,1689	0,1680
5	0,0937	0,1032	0,1023	0,1016	0,1011	0,1010	0,1008
6	0,0156	0,0430	0,0474	0,0490	0,0499	0,0501	0,0504
7	0,0000	0,0138	0,0180	0,0199	0,0209	0,0213	0,0216
8	0,0000	0,0035	0,0058	0,0069	0,0076	0,0079	0,0081
9	0,0000	0,0007	0,0016	0,0021	0,0025	0,0026	0,0027
10	0,0000	0,0001	0,0004	0,0006	0,0007	0,0008	0,0008
11	0,0000	0,0000	0,0001	0,0001	0,0002	0,0002	0,0002
12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001

5.5. РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ

До сих пор мы обсуждали только дискретные распределения одной или (случай полиномиального распределения) нескольких случайных величин. Рассмотрим теперь простейший случай *непрерывной* функции распределения. Он был уже приведен в качестве примера в разд. 3.2.

Назовем распределение равномерным, если его плотность вероятности постоянна на интервале $a \leq x < b$ и равна нулю вне этого интервала:

$$f(x) = \begin{cases} c & a \leq x < b, \\ 0 & x < a, x \geq b. \end{cases} \quad (5.1)$$

Из равенства полной вероятности единице получаем

$$\int_{-\infty}^{\infty} f(x) dx = c \int_a^b dx = c(b-a) = 1$$

и, следовательно,

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x < b, \\ 0, & x < a, \quad x \geq b. \end{cases} \quad (5.2)$$

Функция распределения в этом случае равна

$$\begin{aligned} F(x) &= \int_a^x \frac{dx}{b-a} = \frac{x-a}{b-a}, & a \leq x < b, \\ F(x) &= 0, & x < a, \\ F(x) &= 1, & x \geq b. \end{aligned} \quad (5.3)$$

Из соображений симметрии можно заключить, что математическое ожидание равно среднему арифметическому от значений границ интервала (a, b) . Действительно, из (3.4) гл. 3 видно:

$$E(x) = \hat{x} = \frac{1}{b-a} \int_a^b x dx = \frac{1}{2} \frac{1}{b-a} (b^2 - a^2) = \frac{b+a}{2}. \quad (5.4)$$

Аналогично из (3.10) гл. 3 получаем

$$\sigma^2(x) = \frac{1}{12} (b-a)^2. \quad (5.5)$$

Равномерное распределение само по себе мало существенно для практических приложений. Однако, являясь простейшим непрерывным распределением, оно очень удобно для вычислений. Поэтому целесообразно преобразовать данное распределение в равномерное или, наоборот, получить его путем преобразования переменных из равномерного распределения. Такая техника широко используется в методе Монте-Карло.

5.6. ХАРАКТЕРИСТИЧЕСКАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

До сих пор мы рассматривали случайные величины, принимающие только действительные значения. В разд. 3.1 мы ввели понятие случайной величины как величины, принимающей какое-либо действительное значение, которое соответствует событию.

Не меняя ничего в этом определении, можно построить комплексную случайную величину из двух действительных:

$$\mathbf{z} = \mathbf{x} + i\mathbf{y}. \quad (6.1)$$

При этом ее математическое ожидание определяется как

$$E(\mathbf{z}) = E(\mathbf{x}) + iE\mathbf{y}. \quad (6.2)$$

Аналогично действительным случайным величинам комплексные величины независимы, если независимы их действительные и мнимые части.

Пусть \mathbf{x} — действительная случайная величина, имеющая функцию распределения $F(x)$ и плотность вероятности $f(x)$; назовем *характеристической функцией* математическое ожидание величины $\exp(it\mathbf{x})$:

$$\varphi(t) = E\{\exp(it\mathbf{x})\}. \quad (6.3)$$

Из равенства (6.3) видно, что в случае непрерывной случайной величины характеристическая функция представляет собой интеграл Фурье от плотности

$$\varphi(t) = \int_{-\infty}^{\infty} \exp(itx) f(x) dx. \quad (6.4)$$

Свойства преобразования Фурье известны. Для дискретной случайной величины получаем так же, как в (3.2) гл. 3,

$$\varphi(t) = \sum_i \exp(itx_i) P(\mathbf{x} = x_i). \quad (6.5)$$

Рассмотрим теперь моменты \mathbf{x} относительно нуля:

$$\lambda_n = E(\mathbf{x}^n) = \int_{-\infty}^{\infty} x^n f(x) dx. \quad (6.6)$$

Легко показать, что их можно получить просто дифференцированием характеристической функции n раз в точке $t = 0$:

$$\varphi^{(n)}(t) = \frac{d^n \varphi(t)}{dt^n} = i^n \int_{-\infty}^{\infty} x^n \exp(itx) f(x) dx,$$

т. е.

$$\varphi^{(n)}(0) = i^n \lambda_n. \quad (6.7)$$

Если теперь ввести сдвиг координат

$$\mathbf{y} = \mathbf{x} - \hat{x} \quad (6.8)$$

и построить характеристическую функцию для y :

$$\varphi_y(t) = \int_{-\infty}^{\infty} \exp\{it(x - \hat{x})\} f(x - \hat{x}) dx, \quad (6.9)$$

то ее n -я производная (с точностью до степени i) будет равна n -му моменту x относительно математического ожидания \hat{x} [ср. (3.8) гл. 3]:

$$\varphi_y^{(n)}(0) = i^n \mu_n = i^n E\{(x - \hat{x})^n\} \quad (6.10)$$

и, в частности,

$$\sigma^2(x) = -\varphi_y''(0). \quad (6.11)$$

С помощью обратного преобразования Фурье можно выразить плотность вероятности через характеристическую функцию

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx) \varphi(t) dt. \quad (6.12)$$

Можно показать, что распределение *однозначно* определяется своей характеристической функцией. Это верно также и для дискретных случайных величин; для них равенство (6.12) заменяется равенством

$$F(b) - F(a) = \frac{i}{2\pi} \int_{-\infty}^{\infty} \frac{\exp(itb) - \exp(-ita)}{t} \varphi(t) dt, \quad (6.13)$$

так как плотность вероятности в этом случае не определена. Часто более удобно пользоваться характеристической функцией вместо исходной функции распределения. Благодаря их однозначной связи можно пользоваться в разных местах либо той, либо другой функцией.

Рассмотрим теперь сумму двух независимых случайных величин:

$$\mathbf{w} = \mathbf{x} + \mathbf{y};$$

ее характеристическая функция равна

$$\varphi_w(t) = E[\exp\{it(\mathbf{x} + \mathbf{y})\}] = E\{\exp(it\mathbf{x}) \exp(it\mathbf{y})\}.$$

Обобщая формулу (2.13) гл. 4 для комплексных случайных величин, получим

$$\varphi_w(t) = E\{\exp(itx)\} E\{\exp(ity)\} = \varphi_x(t) \varphi_y(t), \quad (6.14)$$

т. е. характеристическая функция суммы двух независимых случайных величин равна произведению соответствующих характеристических функций.

Пример 5.5. Согласно равенствам (6.5) и (4.1), характеристическая функция распределения Пуассона равна

$$\begin{aligned}\varphi(t) &= \sum_{k=0}^{\infty} \exp(itk) \frac{\lambda^k}{k!} \exp(-\lambda) = \exp(-\lambda) \sum_{k=0}^{\infty} \frac{(\lambda \exp(it))^k}{k!} = \\ &= \exp(-\lambda) \exp(\lambda e^{it}) = \exp\{\lambda(e^{it} - 1)\}.\end{aligned}\quad (6.15)$$

Найдем характеристическую функцию суммы двух независимых распределений Пуассона со средними значениями соответственно λ_1 и λ_2 :

$$\begin{aligned}\varphi_{\text{sum}}(t) &= \exp\{\lambda_1(e^{it} - 1)\} \exp\{\lambda_2(e^{it} - 1)\} = \\ &= \exp\{(\lambda_1 + \lambda_2)(e^{it} - 1)\}.\end{aligned}\quad (6.16)$$

Полученное выражение аналогично (6.15). Следовательно, распределение суммы двух независимых пуассоновских распределений также является пуассоновским. Его среднее значение равно сумме средних значений каждого из распределений.

5.7. ЛАПЛАСОВСКАЯ МОДЕЛЬ ОШИБОК

В 1783 г. Лаплас сделал следующие замечания, касающиеся источника ошибок наблюдений. Пусть истинное значение величины, которую надо измерить, равно m_0 . Предположим, что измерение искажается большим числом n независимых помех, каждая из которых имеет величину ϵ . Каждое возмущение с равной вероятностью приводит к отклонению от истинной величины на $+\epsilon$ или $-\epsilon$. Ошибка измерения складывается из суммы отдельных возмущений. Очевидно, что в данной модели распределение вероятностной ошибки измерения будет описываться биномиальным распределением. Тем не менее интересно проследить модель несколько дальше, так как это приводит к известному треугольнику Паскаля.

На рис. 5.4 распределение вероятностей развивается следующим образом. В начальной точке помехи отсутствуют, и, следовательно, вероятность измерения величины m_0 равна 1. С появлением одной помехи эта вероятность равно распределяется между соседними возможностями $m_0 + \epsilon$ и $m_0 - \epsilon$. Аналогично происходит с каждым следующим возмущением. Очевидно, что вероятности одних и тех же измеренных значений должны быть просуммированы. Каждая строка результирующего треугольника содержит распределение W_k^n ($k = 0, 1, \dots, n$) [равенство (1.3)] для случая $p = q = \frac{1}{2}$. Умножение на $(p^k q^{n-k})^{-1} = 2^n$ приводит к строке биномиальных коэффициентов треугольника Паскаля (см. приложение В).

Перейдем к пределу при числе возмущений $n \rightarrow \infty$. В отличие от подобной процедуры в случае распределения Пуассона положим p и q равными постоянной величине. Хотя в модели Лапласа $p = q = \frac{1}{2}$, здесь мы рассмотрим общий случай произвольного p . Из равенств (1.3) и (6.5) получаем характеристическую функцию биномиального распределения

$$\varphi(t) = \sum_{j=0}^n \exp(itj) \binom{n}{j} p^j q^{n-j}.$$

Число возмущений n	Наблюдаемая величина ($m_0 + \dots$)							
	-3ϵ	-2ϵ	$-\epsilon$	0	$+\epsilon$	$+2\epsilon$	$+3\epsilon$	
0	1							
1	$\frac{1}{2}$		$\frac{1}{2}$		$\frac{1}{2}$		$\frac{1}{2}$	
2	$\frac{1}{4}$		$\frac{1}{4}, \frac{1}{4}$		$\frac{1}{4}$		$\frac{1}{4}$	
3	$\frac{1}{8}$		$\frac{1}{8}, \frac{1}{4}$		$\frac{1}{4}, \frac{1}{8}$		$\frac{1}{8}$	

Рис. 5.4. Связь модели Лапласа с биномиальным распределением.

С помощью биномиальной теоремы (В.6) это выражение можно преобразовать:

$$\varphi(t) = \{\exp(it) p + q\}^n. \quad (7.1)$$

Вместо \mathbf{x} воспользуемся нормированной переменной

$$\mathbf{u} = \frac{\mathbf{x} - \hat{x}}{\sigma(\mathbf{x})} = \frac{\mathbf{x} - np}{\sigma},$$

которую можно рассматривать как сумму двух независимых случайных величин \mathbf{x}/σ и $-\hat{x}/\sigma$, хотя последняя фактически является постоянной. Характеристическая функция для \mathbf{u} , согласно (6.14),

$$\varphi_u(t) = \exp\left(-\frac{itnp}{\sigma}\right) \left\{ \exp\left(\frac{it}{\sigma}\right) p + q \right\}^n.$$

Прологарифмировав (и используя равенство $p = 1 - q$), получаем

$$\ln \varphi_u(t) = -\frac{itnp}{\sigma} + n \ln \left[1 + p \left\{ \exp\left(\frac{it}{\sigma}\right) - 1 \right\} \right].$$

Поскольку $t/\sigma = t/\sqrt{npq}$ стремится к нулю при $n \rightarrow \infty$, разложим выражение в фигурных скобках и получим

$$\ln \varphi_u(t) = -\frac{itnp}{\sigma} + n \ln \left[1 + p \left\{ \frac{it}{\sigma} - \frac{1}{2} \left(\frac{t}{\sigma} \right)^2 + \dots \right\} \right].$$

Воспользуемся разложением

$$\ln(1 + \varepsilon) = \varepsilon - \frac{\varepsilon^2}{2} + \frac{\varepsilon^3}{3} - \dots$$

и найдем

$$\begin{aligned} \ln \varphi_u(t) &= -\frac{itnp}{\sigma} + n \left\{ \frac{itp}{\sigma} - \frac{1}{2} \left(\frac{t}{\sigma} \right)^2 (p - p^2) + \mathcal{O}(\sigma^{-3}) \right\} = \\ &= -\frac{1}{2} t^2 \frac{np(1-p)}{\sigma^2} + \mathcal{O}(n\sigma^{-3}) = \\ &= -\frac{1}{2} t^2 + \mathcal{O}(n\sigma^{-3}). \end{aligned}$$

В пределе при $n \rightarrow \infty$ логарифм стремится к $-\frac{1}{2} t^2$ и, следовательно, характеристическая функция принимает вид

$$\varphi_u(t) = \exp\left(-\frac{1}{2} t^2\right). \quad (7.2)$$

Для получения соответствующей плотности вероятности воспользуемся равенством (6.12):

$$\begin{aligned} f(u) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itu) \exp\left(-\frac{1}{2} t^2\right) dt = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} (t + iu)^2 - \frac{1}{2} u^2\right\} dt. \end{aligned} \quad (7.3)$$

Возьмем $r = (t + iu)$ в качестве новой переменной интегрирования и получим

$$f(u) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} u^2\right) \int_{-\infty + iu}^{\infty + iu} \exp\left(-\frac{1}{2} r^2\right) dr.$$

Интегрирование проводится вдоль линии, параллельной действительной оси, но вследствие «хорошего поведения» подынтегральной функции в бесконечности можно интегрировать вдоль самой действительной оси. Воспользовавшись (Г.4), окончательно получаем

$$f(u) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} u^2\right) \sqrt{2} \Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} u^2\right). \quad (7.4)$$

Распределение (7.4) получено нами из биномиального распределения с использованием нормированной переменной, имеющей

нулевое среднее и единичную дисперсию. Таким образом мы получили

$$\begin{aligned} E(\mathbf{u}) = \hat{u} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \exp\left(-\frac{1}{2}u^2\right) du = \\ &= \left[-\exp\left(-\frac{1}{2}u^2\right)\right]_{-\infty}^{\infty} = 0 \end{aligned} \quad (7.5)$$

и, воспользовавшись соотношением (6.11),

$$\sigma^2(u) = -\varphi''(0) = 1. \quad (7.6)$$

5.8. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Более общая форма плотности вероятности (7.4) имеет вид

$$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}b} \exp\left(-\frac{(x-a)^2}{2b^2}\right), \quad (8.1)$$

Математическое ожидание и дисперсия равны соответственно

$$\hat{x} = a, \quad \sigma^2(x) = b^2. \quad (8.2)$$

Это распределение является предельным для биномиального распределения со средним значением и дисперсией (8.2) при $n \rightarrow \infty$. Оно занимает центральное место в математической статистике и особенно в теории ошибок и называется *нормальным*, или *гауссовским, распределением*. Ввиду его важности мы должны подробно изучить его свойства и количественное поведение.

Для обозначения плотности вероятности нормального распределения мы используем специальный символ $\phi(x)$. *Стандартное нормальное распределение*, т. е. нормальное распределение с нулевым математическим ожиданием и единичной дисперсией, с которым мы уже сталкивались в равенстве (7.4), обозначим через

$$\phi_0(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right). \quad (8.3)$$

Из равенства (6.4) следует, что характеристическая функция ϕ равна

$$\varphi(t) = \frac{1}{\sqrt{2\pi}b} \int_{-\infty}^{\infty} \exp(itx) \exp\left(-\frac{(x-a)^2}{2b^2}\right) dx.$$

Дополняя выражение в скобках до полного квадрата, находим

$$\varphi(t) = \exp(ita) \exp\left(-\frac{1}{2}b^2t^2\right). \quad (8.4)$$

При $a = 0$ получаем интересный факт:

Характеристическая функция нормального распределения с нулевым математическим ожиданием пропорциональна некоторому нормальному распределению с нулевым математическим ожиданием. Произведение дисперсий этих двух распределений равно единице.

Рассматривая сумму независимых нормальных распределений, получаем из (6.14), что ее характеристическая функция снова имеет вид (8.4). Таким образом, сумма независимых нормально распределенных случайных величин также распределена нормально. Подобные свойства были обнаружены в примере 5.5 для пуассоновского распределения.

Для нахождения высших моментов гауссовского распределения воспользуемся равенством (6.10). Характеристическая функция центрированной случайной величины $y = x - \hat{x}$ равна

$$\varphi_y(t) = \exp\left(-\frac{1}{2} b^2 t^2\right).$$

Тогда

$$\mu_n = \frac{1}{i^n} \frac{d^n}{dt^n} \left[\exp\left(-\frac{1}{2} b^2 t^2\right) \right]_{t=0}; \quad (8.5)$$

в частности,

$$\begin{aligned} \mu_1 &= 0, & \mu_2 &= b^2 = \sigma^2, \\ \mu_3 &= 0, & \mu_4 &= 3\sigma^4 \end{aligned}$$

и в общем случае (для целого k)

$$\mu_{2k+1} = 0, \quad (8.6)$$

как и следовало ожидать ввиду симметричности распределения, и

$$\mu_{2k} = \frac{(2k)!}{2^k k!} \sigma^{2k}. \quad (8.7)$$

Таким образом, все высшие моменты выражаются через σ^2 .

5.9. КОЛИЧЕСТВЕННЫЕ СВОЙСТВА НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

На рис. 5.5 приведены графики плотности вероятности стандартного нормального распределения $\phi_0(x)$ [равенство (8.3)] и соответствующей функции распределения

$$\psi_0(x) = \int_{-\infty}^x \phi_0(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{1}{2} x^2\right) dx. \quad (9.1)$$

График плотности вероятности имеет форму знаменитого гауссовского «колокола». Легко проверить, что точками перегиба служат точки $x = \pm 1$ (в общем случае нормального распределения точки $x = a \pm b$). Функция распределения $\psi_0(x)$ определяет вероятность

того, что случайная величина \mathbf{x} примет значение меньшее, чем x :

$$\psi_0(x) = P(\mathbf{x} < x). \quad (9.2)$$

Из соображений симметрии следует, что

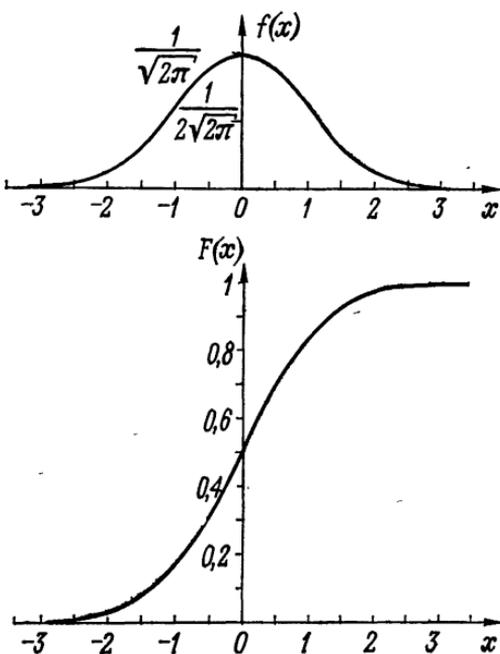
$$P(|\mathbf{x}| > x) = 2\psi_0(-|x|) = 2\{1 - \psi_0(|x|)\}, \quad (9.3)$$

или, наоборот, вероятность наблюдения случайной величины x в интервале ширины $2x$ с центром в нуле (математическом ожидании) равна

$$P(|\mathbf{x}| \leq x) = 2\psi_0(|x|) - 1. \quad (9.4)$$

Значения величин (9.2)–(9.4) можно найти во многих таблицах (ссылки на ряд статистических таблиц даны в библиографии). В табл. Е.2а и Е.2б приведены табулированные значения (9.2) и (9.4).

Рис. 5.5. Плотность вероятности $f(x)$ и функция распределения вероятностей $F(x)$ стандартного нормального распределения.



Полученные выше соотношения можно легко обобщить для случая общего гауссовского распределения. При этом функция распределения равна

$$\begin{aligned} \psi(x) &= \frac{1}{\sqrt{2\pi}b} \int_{-\infty}^x \exp\left(-\frac{(x-a)^2}{b^2}\right) dx = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-a)/b} \exp\left(-\frac{1}{2}u^2\right) du = \psi_0\left(\frac{x-a}{b}\right). \end{aligned} \quad (9.5)$$

Попытаемся найти вероятность отклонения случайной величины от математического ожидания a на величину, кратную стандартному отклонению $\sigma = b$:

$$P(|\mathbf{x} - a| \leq n\sigma) = 2\psi_0\left(\frac{n\sigma}{b}\right) - 1 = 2\psi_0(n) - 1. \quad (9.6)$$

Из табл. Е.26 находим

$$\begin{aligned} P(|\mathbf{x} - a| \leq \sigma) &= 68,2\%, & P(|\mathbf{x} - a| > \sigma) &= 31,8\%, \\ P(|\mathbf{x} - a| \leq 2\sigma) &= 95,4\%, & P(|\mathbf{x} - a| > 2\sigma) &= 4,6\%, \\ P(|\mathbf{x} - a| \leq 3\sigma) &= 99,8\%, & P(|\mathbf{x} - a| > 3\sigma) &= 0,2\%. \end{aligned} \quad (9.7)$$

Интересно сравнить эти оценки с оценками, получающимися с помощью неравенства Чебышева, не требующего каких-либо предположений о виде функции распределения:

$$\begin{aligned} P(|\mathbf{x} - a| > \sigma) &< 1, \\ P(|\mathbf{x} - a| > 2\sigma) &< 25\%, \\ P(|\mathbf{x} - a| > 3\sigma) &< 11\%. \end{aligned}$$

Как уже отмечалось в разд. 3.4, оценка с помощью неравенства Чебышева довольно слабая и может быть полезна лишь в ситуациях, когда о распределении ничего не известно.

Позже мы увидим, что во многих случаях с большой точностью можно предположить, что ошибки измерения распределены нормально вокруг истинного значения a , т. е. вероятность того, что измеряемая величина будет лежать в интервале $[x, x + dx]$, определяется с помощью формулы

$$P(x \leq \mathbf{x} < x + dx) = \phi(x) dx.$$

Стандартное отклонение σ распределения $\phi(x)$ может быть приближенно найдено из ряда измерений и затем использовано для характеристики ошибки измерения. Принимаемое обычно соглашение об отождествлении стандартного отклонения с ошибкой измерения является очень важным. Из соотношения (9.7) мы знаем, что вероятность отклонения результата измерения от истинного значения не более чем на одно стандартное отклонение равна только 68,2%. Эта величина кажется довольно низкой. Поэтому многие экспериментаторы имеют обыкновение умножать стандартное отклонение на более или менее произвольный множитель, например на 3, что увеличивает вероятность до 99,8%. Хотя этот метод и кажется разумным, он вводит в заблуждение и часто вреден. Если множитель явно не установлен, то сравнение различных измерений одной и той же величины и особенно вычисление взвешенного среднего (см. пример 7.2) становится невозможным или ошибочным. Функция, обратная по отношению к функции нормального распределения, также представляет большой интерес. Обозначим ее через

$$x = \Omega(P), \quad (9.8)$$

Функция x для заданного P равна такому значению, для которого стандартная нормальная случайная величина \mathbf{x} меньше x с вероятностью P :

$$P(\mathbf{x} < \Omega(P)) = P. \quad (9.9)$$

Определим также функцию Ω' , такую, что

$$P(|\mathbf{x}| < \Omega'(P)) = P. \quad (9.10)$$

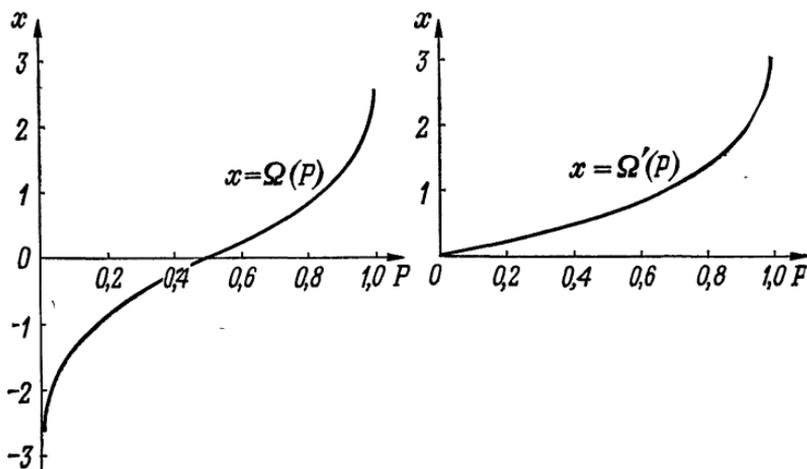


Рис. 5.6. Функции, обратные нормальной функции распределения.

Очевидно,

$$P = \int_{-\infty}^{\Omega(P)} \phi_0(x) dx \quad (9.11)$$

или

$$P = 2 \int_0^{\Omega'(P)} \phi_0(x) dx \quad (9.12)$$

соответственно. Функции $\Omega(P)$ и $\Omega'(P)$ табулированы в табл. Е.3а и Е.3б. На рис. 5.6, где изображены обе эти функции, легко проследить очевидное соотношение между ними:

$$\Omega'(P) = \Omega\left(\frac{1}{2}(1 + P)\right). \quad (9.13)$$

5.10. МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Совместное нормальное распределение n случайных величин $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ определяется следующим образом:

$$\phi(\mathbf{x}) = k \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{a})^T B(\mathbf{x} - \mathbf{a})\right\}. \quad (10.1)$$

Здесь k выбирается из условия равенства полной вероятности единице; \mathbf{a} так же, как \mathbf{x} , является n -мерным вектором и B — матрица размера $n \times n$. Так как $\phi(\mathbf{x})$, очевидно, симметрична относительно $(\mathbf{x} - \mathbf{a})$, имеем

$$\int_{-\infty}^{\infty} (\mathbf{x} - \mathbf{a}) \phi(\mathbf{x}) d\mathbf{x} = 0, \quad (10.2)$$

т. е.

$$E(\mathbf{x} - \mathbf{a}) = 0$$

или

$$E(\mathbf{x}) = \mathbf{a}. \quad (10.3)$$

Таким образом, вектор математического ожидания равен вектору \mathbf{a} . Продифференцируем равенство (10.2) по \mathbf{a} :

$$\int_{-\infty}^{\infty} [I - (\mathbf{x} - \mathbf{a})(\mathbf{x} - \mathbf{a})^T B] \phi(\mathbf{x}) d\mathbf{x} = 0,$$

откуда следует, что математическое ожидание величины, заключенной в квадратных скобках, равно 0, т. е.

$$BE\{(\mathbf{x} - \mathbf{a})(\mathbf{x} - \mathbf{a})^T\} = I,$$

или

$$C = E\{(\mathbf{x} - \mathbf{a})(\mathbf{x} - \mathbf{a})^T\} = B^{-1}. \quad (10.4)$$

Сравнение этого результата с равенством (3.21) гл. 4 показывает, что C является ковариационной матрицей случайных величин $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$.

Рассмотрим подробно нормальное распределение двух случайных величин. Имеем

$$C = B^{-1} = \begin{pmatrix} \sigma_1^2 & \text{cov}(\mathbf{x}_1, \mathbf{x}_2) \\ \text{cov}(\mathbf{x}_1, \mathbf{x}_2) & \sigma_2^2 \end{pmatrix}. \quad (10.5)$$

Обращая эту матрицу, получаем

$$B = \frac{1}{\sigma_1^2 \sigma_2^2 - \text{cov}(\mathbf{x}_1, \mathbf{x}_2)^2} \begin{pmatrix} \sigma_2^2 & -\text{cov}(\mathbf{x}_1, \mathbf{x}_2) \\ -\text{cov}(\mathbf{x}_1, \mathbf{x}_2) & \sigma_1^2 \end{pmatrix}. \quad (10.6)$$

В случае нулевой ковариации, т. е. если $\mathbf{x}_1, \mathbf{x}_2$ независимы, B принимает вид

$$B_0 = \begin{pmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{pmatrix}. \quad (10.7)$$

Подставляя B_0 в (10.1), получаем, как и следовало ожидать, произведение двух одномерных нормальных распределений:

$$\phi = k \exp\left(-\frac{1}{2} \frac{(x_1 - a_1)^2}{\sigma_1^2}\right) \exp\left(-\frac{1}{2} \frac{(x_2 - a_2)^2}{\sigma_2^2}\right). \quad (10.8)$$

k в этом случае равно

$$k_0 = \frac{1}{2\pi\sigma_1\sigma_2},$$

что получается путем интегрирования (10.8) или, проще, путем сравнения с равенством (8.1). В общем случае n случайных величин и ненулевых ковариаций

$$k = \frac{1}{(2\pi)^{\frac{n}{2}} (\det B)^{\frac{1}{2}}}, \quad (10.9)$$

где $\det B$ — определитель матрицы B .

В случае зависимых случайных величин, т. е. ненулевых ковариаций, нормальное распределение двух случайных величин более сложно. Для рассмотрения этого случая воспользуемся следующими нормированными величинами:

$$u_i = \frac{x_i - a_i}{\sigma_i}, \quad i = 1, 2,$$

и коэффициентом корреляции

$$\rho = \frac{\text{cov}(x_1, x_2)}{\sigma_1\sigma_2} = \text{cov}(u_1, u_2).$$

Теперь равенство (10.1) принимает простой вид

$$\phi(u_1, u_2) = k \exp\left(-\frac{1}{2} u^T B u\right), \quad (10.10)$$

где

$$B = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}. \quad (10.11)$$

Линии, на которых плотность вероятности постоянна, определяются из условия равенства постоянной величине показателя экспоненты в равенстве (10.10):

$$\frac{1}{(1-\rho^2)} (u_1^2 + u_2^2 - 2u_1u_2\rho) = c. \quad (10.12)$$

Положим сначала $c = 1$: Тогда для исходных случайных величин равенство (10.12) примет вид

$$\frac{(x_1 - a_1)^2}{\sigma_1^2} - 2\rho \frac{x_1 - a_1}{\sigma_1} \frac{x_2 - a_2}{\sigma_2} + \frac{(x_2 - a_2)^2}{\sigma_2^2} = 1 - \rho^2. \quad (10.13)$$

Это уравнение эллипса с центром в точке (a_1, a_2) . Главные оси этого эллипса образуют угол α с осями x_1 и x_2 . Этот угол и полуоси p_1 и p_2 могут быть найдены из равенства (10.13), если

воспользоваться хорошо известными свойствами конических сечений:

$$\operatorname{tg} 2\alpha = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}, \quad (10.14)$$

$$\rho_1^2 = \frac{\sigma_1^2\sigma_2^2(1 - \rho^2)}{\sigma_2^2 \cos^2 \alpha - 2\rho\sigma_1\sigma_2 \sin \alpha \cos \alpha + \sigma_1^2 \sin^2 \alpha}, \quad (10.15)$$

$$\rho_2^2 = \frac{\sigma_1^2\sigma_2^2(1 - \rho^2)}{\sigma_2^2 \sin^2 \alpha + 2\rho\sigma_1\sigma_2 \sin \alpha \cos \alpha + \sigma_1^2 \cos^2 \alpha}. \quad (10.16)$$

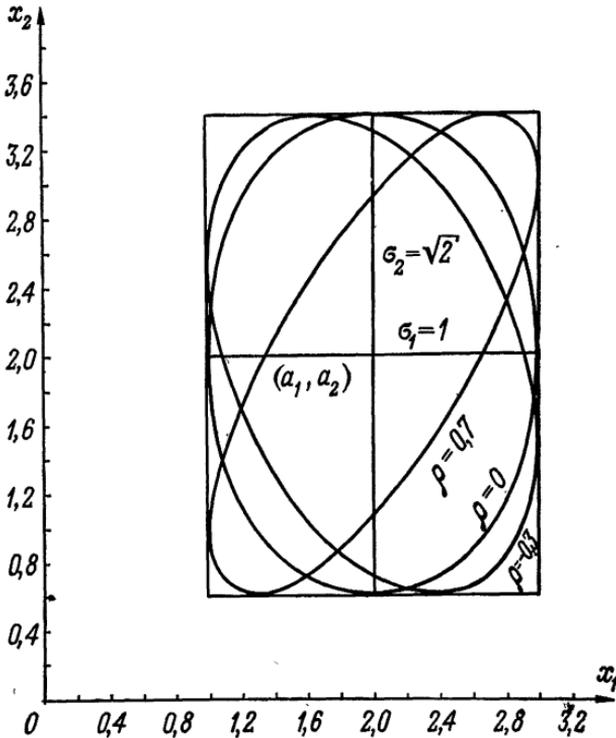


Рис. 5.7. Ковариационные эллипсы.

Эллипс, обладающий описанными выше свойствами, называется *ковариационным эллипсом* двумерного нормального распределения. Для частного случая $\sigma_1^2 = 1$, $\sigma_2^2 = 2$ три таких эллипса показаны на рис. 5.7 для $\rho = 0, 0, 7$ и $-0,3$ соответственно. Ковариационный эллипс всегда расположен внутри прямоугольника, с центром в точке (a_1, a_2) и со сторонами, равными $2\sigma_1, 2\sigma_2$. Он касается этого прямоугольника в четырех точках. В предельных случаях, когда $\rho = \pm 1$, эллипс переходит в прямую, совпадающую с одной из диагоналей прямоугольника.

Как следует из (10.13), другие линии постоянной вероятности также представляют собой эллипсы, подобные ковариационному эллипсу и расположенные внутри (снаружи) этого эллипса для большей (меньшей) вероятности. Следовательно, двумерное нормальное распределение соответствует поверхности в трехмерном пространстве (рис. 5.8), горизонтальными сечениями которой являются концентрические эллипсы. Для наибольшей вероятности этот

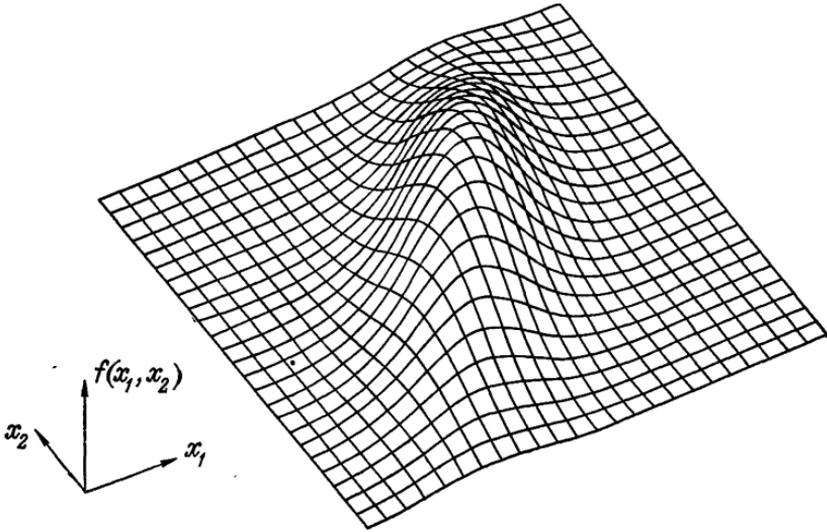


Рис. 5.8. Плотность вероятности двумерного гауссовского распределения.

эллипс вырождается в точку (a_1, a_2) . Вертикальные сечения, проходящие через центр, имеют форму гауссовской функции, ширина которой прямо пропорциональна оси ковариационного эллипса, вдоль которого производится сечение.

Вероятность наблюдения пары x_1, x_2 случайных величин внутри ковариационного эллипса равна интегралу

$$\int_A k \phi(\mathbf{x}) d\mathbf{x} = 1 - e^{-1/2}, \quad (10.17)$$

где область интегрирования A является ковариационный эллипс (10.13). Соотношение (10.17) можно получить, применяя преобразование случайных величин $\mathbf{y} = T\mathbf{x}$, где $T^{-1} = |\overline{B}^{-1}|$, к распределению $\phi(\mathbf{x})$. Получающееся распределение обладает следующими свойствами:

$$\sigma(y_1) = \sigma(y_2) = 1, \quad \text{cov}(y_1, y_2) = 0,$$

т. е. оно имеет вид (10.8). Область интегрирования A при этом преобразуется в единичный круг с центром (a_1, a_2) .

Рассматривая одномерное нормальное распределение ошибок измерения, мы обнаружили, что интервал $a - \sigma \leq x \leq a + \sigma$ является областью, содержащей значения $f(x)$ больше некоторой величины. Интеграл по такой области не зависел от σ . В случае двух случайных величин в качестве такой области берется ковариационный эллипс, который определен величинами σ_1 , σ_2 и ρ , а не внешним прямоугольником (рис. 5.7), как это часто принимается. Важность ковариационного эллипса становится ясной при рассмотрении рис. 5.9. Точки 1 и 2, будучи расположены на ковариационном эллипсе, соответствуют равным вероятностям $P(1) = P(2) = P_e$, хотя расстояния от центра до точки 1 меньше по обеим координатам. Кроме того, вероятность в точке 4 меньше P_e , а в точке 3 больше ($P(4) < P_e$, $P(3) > P_e$), хотя точка 4 ближе к точке (a_1, a_2) , чем точка 3.

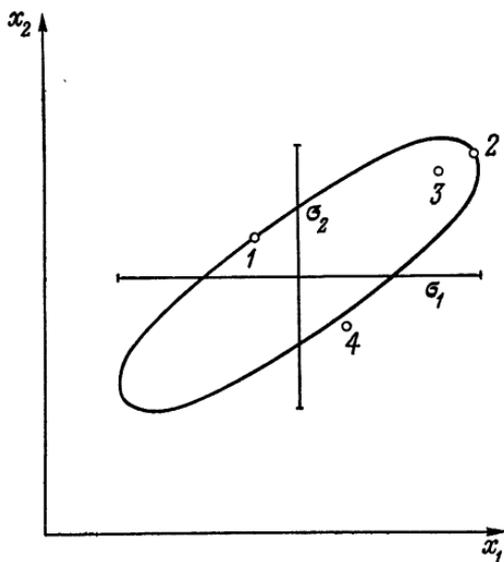


Рис. 5.9. Сравнение вероятностей в различных точках для двумерного нормального распределения: $P_1 = P_2 = P_e$, $P_3 > P_e$, $P_4 < P_e$.

5.11. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

При выводе нормального распределения мы руководствовались следующими соображениями. Мы начали с простого эксперимента с двумя возможными исходами, который описывался случайной величиной x_i , принимающей значения 0 или 1. Конечное число n таких экспериментов привело к биномиальному распределению случайной величины $\sum x_i$. Наконец, переходя к пределу при $n \rightarrow \infty$, мы пришли к нормальному распределению для

$$\mathbf{x} = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbf{x}_i. \quad (11.1)$$

Займемся теперь более общими суммами такого типа. Допустим, что все x_i независимы и порождены распределением со средним значением a и дисперсией σ . Предположим, что нам ничего не известно о распределении каждой x_i . Обозначим через $\varphi(t)$ характеристическую функцию такого распределения. Тогда, согласно равенству (6.14), характеристическая функция суммы n случайных

величин равна $\{\varphi(t)\}^n$. Чтобы упростить вычисления, положим $a = 0$ (общий случай можно свести к этому простым переносом координат $\mathbf{x}'_i = \mathbf{x}_i - a$).

Первая и вторая производные функции $\varphi(A)$ в точке $t = 0$ нам известны из (6.10);

$$\varphi'(0) = 0, \quad \varphi''(0) = -\sigma^2,$$

и мы можем выписать первые члены в разложении Тейлора

$$\varphi_{x'}(t) = 1 - \frac{1}{2} \sigma^2 t^2 + \dots$$

Введем теперь случайную величину \mathbf{u}_i , определенную следующим образом:

$$\mathbf{u}_i = \frac{\mathbf{x}'_i}{\sigma \sqrt{n}} = \frac{\mathbf{x}_i - a}{\sigma \sqrt{n}}. \quad (11.2)$$

Ее характеристическая функция имеет вид

$$\varphi_{u_i}(t) = E \{ \exp(itu_i) \} = E \left\{ \exp \left(it \frac{\mathbf{x}_i - a}{\sigma \sqrt{n}} \right) \right\} = \varphi_{x'_i} \left(\frac{t}{\sigma \sqrt{n}} \right),$$

поэтому

$$\varphi_{u_i}(t) = 1 - \frac{t^2}{2n} + \dots,$$

где опущенные члены имеют порядок $n^{-3/2}$. Перейдем к пределу при $n \rightarrow \infty$ и обозначим

$$\mathbf{u} = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbf{u}_i = \lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{\mathbf{x}_i - a}{\sigma \sqrt{n}}.$$

Тогда

$$\varphi_{\mathbf{u}}(t) = \lim_{n \rightarrow \infty} \{\varphi_{u_i}(t)\}^n = \lim_{n \rightarrow \infty} \left(1 - \frac{t^2}{2n} + \dots \right)^n,$$

т. е.

$$\varphi_{\mathbf{u}}(t) = \exp \left(-\frac{1}{2} t^2 \right). \quad (11.3)$$

Полученная функция является характеристической функцией стандартного нормального распределения $\Phi_0(\mathbf{u})$. Возвращаясь к исходной случайной величине \mathbf{x} и используя равенства (11.2) и (9.5), получаем центральную предельную теорему:

Если \mathbf{x}_i — множество независимых случайных величин, распределенных со средним значением a и дисперсией σ^2 , то их сумма

$\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ при $n \rightarrow \infty$ подчиняется гауссовскому распределению со средним a и дисперсией σ^2/n .

Можно показать, что суммы случайных величин подчиняются гауссовскому распределению даже при более слабых предположениях, чем те, которые использовались нами для доказательства, в частности, в тех случаях, когда не все x_i имеют одинаковое распределение.

5.12. ОШИБКИ ЭКСПЕРИМЕНТА И НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ; МОДЕЛЬ ГЕРШЕЛЯ

С помощью центральной предельной теоремы лапласовскую модель ошибок можно теперь истолковать как модель для суммы малых индивидуальных эффектов (разд. 5.7). Первоначальное условие, по которому все возмущения имели одинаковые амплитуды, можно теперь отбросить, и тем не менее в результате получится нормальное распределение ошибок. Интересно рассмотреть модель Джона Гершеля, который вывел нормальное распределение при довольно слабых предположениях. Пусть из точки O на плоскость падают маленькие зерна; вычислим вероятность попадания одного из зерен на элемент площади dA . Этот элемент может рассматриваться в декартовых или полярных координатах (рис. 5.10). Модель построена так, чтобы были невозможны какие-либо систематические отклонения зерен от центра, т. е. для данного r все углы одинаково вероятны, и поэтому плотность вероятности для попадания зерен в элемент dA на расстоянии r от центра O является функцией, зависящей только от r . Назовем ее $h(r)$.

Сделаем теперь первое предположение, состоящее в том, что попадания зерна в полосы между x и $x + dx$ и между y и $y + dy$ являются независимыми событиями. Вследствие симметрии по θ плотности вероятностей по x и y должны иметь одинаковый вид, т. е. $f(x, y) = f(x)f(y)$. Таким образом, вероятность попадания зерна в элемент площади dA равна

$$h(r) dA = f(x) f(y) dA. \quad (12.1)$$

Взяв логарифм от обеих частей (12.1), получаем

$$\ln h(r) = \ln f(x) + \ln f(y)$$

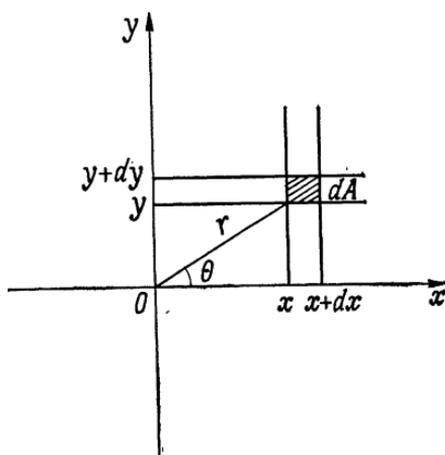


Рис. 5.10. Координатная система для рассмотрения модели Гершеля.

или

$$\ln h(r) = \ln f(r \cos \theta) + \ln f(r \sin \theta). \quad (12.2)$$

Дифференцирование по θ дает

$$0 = -r \sin \theta \frac{f'(r \cos \theta)}{f(r \cos \theta)} + r \cos \theta \frac{f'(r \sin \theta)}{f(r \sin \theta)},$$

или, перейдя снова к x , y ,

$$\frac{f'(x)}{xf(x)} = \frac{f'(y)}{yf(y)}. \quad (12.3)$$

Так как x и y не связаны между собой, это равенство может удовлетворяться, только если обе части равны постоянной величине, т. е.

$$f'(x)/f(x) = cx. \quad (12.4)$$

Интегрированием находим

$$\ln f(x) = \frac{1}{2} cx^2 + \text{const}$$

или

$$f(x) = k \exp\left(\frac{1}{2} cx^2\right). \quad (12.5)$$

Далее, поскольку плотность вероятности $f(x)$ должна быть конечной при x , стремящемся к бесконечности, c должно быть отрицательным. Обозначим $c = -\sigma^{-2}$, тогда

$$f(x) = k \exp\left(-\frac{1}{2} x^2/\sigma^2\right), \quad (12.6)$$

т. е. $f(x)$ имеет форму гауссовского распределения. Вследствие (12.3) также имеем

$$f(y) = k \exp\left(-\frac{1}{2} y^2/\sigma^2\right)$$

и, используя (12.1),

$$g(r) = k^2 \exp\left(-\frac{1}{2} \frac{x^2 + y^2}{\sigma^2}\right) = k^2 \exp\left(-\frac{1}{2} \frac{r^2}{\sigma^2}\right). \quad (12.7)$$

Из очень простых предположений симметрии и независимости мы получили нормальное распределение для отклонений при поражении цели.

Эта модель Гершеля еще раз демонстрирует правдоподобность использования гауссовской формы для распределения ошибок измерения. отождествление распределения ошибок с гауссовским распределением служит решающим фактором во многих задачах, особенно в теории наименьших квадратов (гл. 9). Тем не менее нормальное распределение ошибок не является законом природы.

Причины экспериментальных ошибок слишком многообразны для того, чтобы кто-либо мог принять во внимание все возможные их проявления. Симметрия и независимость не всегда имеют место. Поэтому нормальность распределения необходимо проверять. Это можно сделать, например, применяя критерий χ^2 к исследуемому распределению (гл. 8, разд. 6), до проведения обширных вычислений, имеющих смысл для гауссовского распределения.

5.13. СВЕРТКА РАСПРЕДЕЛЕНИЙ

Мы уже рассматривали суммы случайных величин в различных случаях и, например, при доказательстве центральной предельной теоремы нашли, что применение в этих ситуациях характеристических функций оказывается полезным. Однако их использование приводит к некоторой потере непосредственности в рассуждениях. Поэтому в этом небольшом разделе мы обсудим простой случай распределения суммы двух случайных величин, не обращаясь к характеристическим функциям.

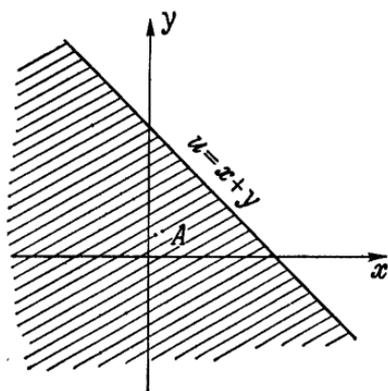


Рис. 5.11. Область интегрирования в равенстве (13.4).

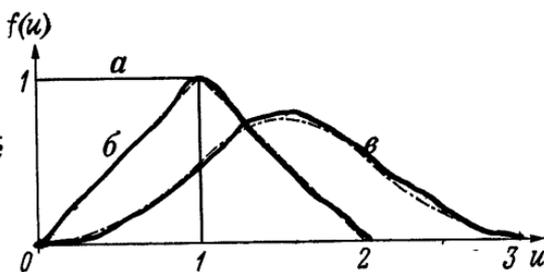


Рис. 5.12. Свертка равномерных распределений.

На опыте мы часто наблюдаем сумму двух распределений. Распределение углов разлета при вторичном распаде элементарных частиц важно для определения некоторых свойств этих частиц. Наблюдаемый угол складывается в этом случае из двух случайных величин: угла распада при случайном процессе радиоактивного распада и ошибки измерения. Для того чтобы сформулировать новое распределение, мы введем понятие о *свертке**.

* Вместо термина «свертка» (convolution) некоторые авторы предпочитают немецкое слово Faltung или французское composition.

Пусть \mathbf{x} , \mathbf{y} — исходные величины и \mathbf{u} — их сумма

$$\mathbf{u} = \mathbf{x} + \mathbf{y}. \quad (13.1)$$

Для дальнейших рассмотрений необходимо предположить независимость \mathbf{x} и \mathbf{y} , так как только в этом случае совместную плотность вероятности можно представить произведением соответствующих плотностей

$$f(x, y) = f_x(x) f_y(y). \quad (13.2)$$

Функция распределения для \mathbf{u} , т. е.

$$F(u) = P(\mathbf{u} < u) = P(\mathbf{x} + \mathbf{y} < u), \quad (13.3)$$

определяется с помощью интегрирования (13.2) по области, заштрихованной на рис. 5.11:

$$\begin{aligned} F(u) &= \iint_A f_x(x) f_y(y) dx dy = \\ &= \int_{-\infty}^{\infty} f_x(x) dx \int_{-\infty}^{u-x} f_y dy = \int_{-\infty}^{\infty} f_y dy \int_{-\infty}^{u-y} f_x(x) dx. \end{aligned} \quad (13.4)$$

Плотность вероятности получается дифференцированием:

$$\begin{aligned} f(u) &= \frac{dF(u)}{du} = \int_{-\infty}^{\infty} f_x(x) f_y(u-x) dx = \\ &= \int_{-\infty}^{\infty} f_y(y) f_x(u-y) dy. \end{aligned} \quad (13.5)$$

Это так называемые *интегралы-свертки*.

Пример 5.6. Хорошо известен пример последовательной свертки равномерных распределений. Результат показан на рис. 5.12. Свертка трех равномерных распределений выглядит похожей на гауссовское распределение. (В действительности она состоит из трех сопряженных парабол, в чем читатель может легко убедиться.)

ВЫБОРОЧНЫЙ МЕТОД

В предыдущей главе мы обсудили ряд распределений, но не уточняли, как они реализуются в каждом отдельном случае. Находилась только вероятность того, что случайная величина x лежит внутри интервала $x < \mathbf{x} \leq x + dx$. Теперь же эта вероятность будет зависеть от некоторых параметров, описывающих ее распределение (например, λ в случае распределения Пуассона), которые, как правило, неизвестны. Следовательно, мы не будем знать точного распределения вероятностей и должны аппроксимировать его *частотным распределением*, полученным экспериментально. Совокупность экспериментов, производимых для этой цели, называется *выборкой*. Число их обязательно должно быть конечным. Прежде чем приступить к изложению некоторых элементов выборочного метода, введем несколько новых определений.

6.1. СЛУЧАЙНЫЙ ВЫБОР, РАСПРЕДЕЛЕНИЕ ВЫБОРКИ; ВЫБОРОЧНЫЕ ОЦЕНКИ

Выборка производится из множества элементов. Размер этого множества обычно бесконечен. Оно может состоять, например, из всевозможных экспериментов или событий данного типа. Такое множество называется *популяцией*. Если произведена выборка n элементов, то говорят, что она имеет *объем n* . Пусть распределение случайной величины x в популяции описывается плотностью вероятности $f(x)$. Нас теперь будут интересовать значения x , которые принимают элементы выборки. Предположим, что последовательно произведено l выборок размера n и при этом получены следующие значения x :

$$\begin{aligned}
 &1\text{-я выборка: } \mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \dots, \mathbf{x}_n^{(1)}, \\
 &\quad \vdots \\
 &j\text{-я выборка: } \mathbf{x}_1^{(j)}, \mathbf{x}_2^{(j)}, \dots, \mathbf{x}_n^{(j)}, \\
 &\quad \vdots \\
 &l\text{-я выборка: } \mathbf{x}_1^{(l)}, \mathbf{x}_2^{(l)}, \dots, \mathbf{x}_n^{(l)}.
 \end{aligned}$$

Сгруппируем результат одной выборки в n -мерную случайную величину

$$\mathbf{x}^{(j)} = (\mathbf{x}_1^{(j)}, \mathbf{x}_2^{(j)}, \dots, \mathbf{x}_n^{(j)}), \quad (1.1)$$

которую можно рассматривать как вектор n -мерного выборочного пространства (гл. 2, разд. 1). Его распределение соответствует плотности вероятности

$$g(\mathbf{x}) = g(x_1, x_2, \dots, x_n). \quad (1.2)$$

Эта плотность вероятности описывает процесс *случайного выбора* только в том случае, если она удовлетворяет двум следующим условиям:

а) Случайные величины \mathbf{x}_i должны быть независимыми, т. е.

$$g(\mathbf{x}) = g_1(x_1) g_2(x_2) \dots g_n(x_n). \quad (1.3)$$

б) Каждая плотность вероятности должна быть равна плотности вероятности распределения x в популяции:

$$g_1(x) = g_2(x) = \dots = g_n(x) = f(x). \quad (1.4)$$

При сопоставлении этих условий с (1.2) становится очевидным тот факт, что простая связь между популяцией и выборкой имеет место только в случае их выполнения. Если явно не утверждается обратное, то под словом «выборка» мы будем понимать результат случайного выбора.

Следует подчеркнуть, что в реальном процессе выбора часто бывает очень трудно осуществить его случайность. Из-за того что применения выборочного метода бывают крайне многообразны, нельзя дать общего правила для обеспечения этого условия. Чтобы получить из выборки достоверные результаты, необходимо строго соблюдать выполнение условий (1.3) и (1.4). В то время как независимость (1.3) можно определенным образом проконтролировать, сопоставляя экспериментальные распределения для 1-го, 2-го, ... элементов, взятых из большого числа выборок, мы, вообще говоря, не можем быть уверены в том, что данная выборка, взятая из популяции, описывается плотностью вероятности $f(x)$. Если элементы популяции можно некоторым образом пронумеровать, то для получения выборки иногда лучше отбирать элементы, пользуясь таблицей случайных чисел (табл. Е.8).

Предположим теперь, что n элементов выборки упорядочены в соответствии со значениями случайной величины \mathbf{x} , лежащими на оси x , и определим n_x как число элементов, для которых выполняется соотношение $\mathbf{x} < x$. Функция

$$W_n(x) = n_x/n \quad (1.5)$$

играет роль эмпирической функции распределения вероятностей. Это ступенчатая функция, возрастающая на $1/n$ всякий раз, когда

x становится равным значению какого-то элемента выборки. Назовем эту функцию *выборочным распределением*. Очевидно, $W_n(x)$ является приближением функции распределения вероятностей популяции $F(x)$, к которой она стремится при $n \rightarrow \infty$. Любая выборочная функция от (1.1), т. е. функция случайной величины \mathbf{x} , сама являющаяся случайной величиной, называется *статистикой*. Хорошо известным примером является *выборочное среднее*

$$\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n), \quad (1.6)$$

определяемое как среднее арифметическое величин x_i . Типичной задачей анализа наблюдений является следующая. Известен общий вид плотности вероятности популяции. Из выборки должно быть получено численное значение одного (или нескольких) параметров. Например, при радиоактивном распаде число ядер, распавшихся за интервал времени между $t = 0$ и $t = \tau$, выражается формулой

$$N_\tau = N_0 [1 - \exp(-\lambda\tau)];$$

N_0 — начальное количество ядер в момент $t = 0$. С помощью конечной выборки, а именно, регистрируя времена конечного числа распадов, можно оценить параметр λ . Из-за конечного объема выборки результат не может быть точен. Следовательно, возникает задача *оценивания параметров*.

Так как оцениваемое значение получается с помощью выборки, оно является статистикой, которую мы будем называть *оценкой*

$$\mathbf{S} = \mathbf{S}(x_1, x_2, \dots, x_n). \quad (1.7)$$

Впрочем, иногда мы будем использовать термин «статистика» («статистика» \mathbf{S} служит для оценки параметра λ). Оценка называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру для выборки любого размера:

$$E\{\mathbf{S}(x_1, x_2, \dots, x_n)\} = \lambda \text{ для любого } n. \quad (1.8)$$

Оценка называется *состоятельной*, если при неограниченном увеличении размера выборки точность результата неограниченно возрастает*, т. е. если

$$\lim_{n \rightarrow \infty} \sigma(\mathbf{S}) = 0. \quad (1.9)$$

Не все оценки равным образом эффективны. Для сравнения относительной эффективности двух оценок может быть использовано отношение

$$\eta = \frac{E\{(\mathbf{S}_1 - \lambda)^2\}}{E\{(\mathbf{S}_2 - \lambda)^2\}} = \frac{\sigma^2(\mathbf{S}_1)}{\sigma^2(\mathbf{S}_2)}.$$

* Обычно оценку называют состоятельной, если она сходится по вероятности к оцениваемому параметру.— *Прим. ред.*

Иногда можно установить нижний предел дисперсии любой оценки параметра (гл. 7, разд. 3). Если существует оценка, на которой этот предел достигается, то, естественно, она предпочтительней других.

6.2. ВЫБОР ИЗ НЕПРЕРЫВНЫХ ПОПУЛЯЦИЙ

На практике наиболее интересен случай выбора из непрерывной популяции, описываемой плотностью вероятности $f(x)$. Рассмотрим сначала *выборочное среднее*, уже определенное посредством (1.6) как среднее арифметическое всех значений из выборки:

$$\bar{\mathbf{x}} = \frac{1}{n} (\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_n). \quad (2.1)$$

Будучи функцией от случайных величин, выборочное среднее само является случайной величиной. Ее математическое ожидание равно

$$E(\bar{\mathbf{x}}) = \frac{1}{n} \{E(\mathbf{x}_1) + E(\mathbf{x}_2) + \dots + E(\mathbf{x}_n)\} = \hat{x}, \quad (2.2)$$

т. е. равно математическому ожиданию \mathbf{x} . Так как это справедливо для любого n , выборочное среднее является несмещенной оценкой математического ожидания \mathbf{x} в популяции, т. е. *среднего значения популяции*. Характеристическая функция выборочного среднего (которая нам потребуется в дальнейшем), согласно (2.1), равна

$$\varphi_{\bar{\mathbf{x}}}(t) = \frac{1}{n} \{\varphi_{\mathbf{x}}(t)\}^n = \left\{ \varphi_{\mathbf{x}} \left(\frac{t}{n} \right) \right\}^n. \quad (2.3)$$

Далее нас будет интересовать дисперсия $\bar{\mathbf{x}}$:

$$\begin{aligned} \sigma^2(\bar{\mathbf{x}}) &= E \{(\bar{\mathbf{x}} - E(\bar{\mathbf{x}}))^2\} = E \left\{ \left(\frac{\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_n}{n} - \hat{x} \right)^2 \right\} = \\ &= \frac{1}{n^2} E \{((\mathbf{x}_1 - \hat{x}) + (\mathbf{x}_2 - \hat{x}) + \dots + (\mathbf{x}_n - \hat{x}))^2\}. \end{aligned}$$

Так как \mathbf{x}_i взаимно независимы, все смешанные члены типа

$$E \{(\mathbf{x}_i - \hat{x})(\mathbf{x}_j - \hat{x})\},$$

т. е. ковариации, обращаются в нуль, и мы получаем

$$\sigma^2(\bar{\mathbf{x}}) = \frac{1}{n} \sigma^2(\mathbf{x}). \quad (2.4)$$

Следовательно, выборочное среднее является состоятельной оценкой \hat{x} . Найдем теперь оценку дисперсии популяции. В первом приближении определим *выборочную дисперсию* как среднее арифметическое среднеквадратических отклонений от выборочного среднего

$$s'^2 = \frac{1}{n} \{(\mathbf{x}_1 - \bar{\mathbf{x}})^2 + (\mathbf{x}_2 - \bar{\mathbf{x}})^2 + \dots + (\mathbf{x}_n - \bar{\mathbf{x}})^2\}. \quad (2.5)$$

Математическое ожидание этой величины равно

$$\begin{aligned}
 E(\mathbf{s}'^2) &= \frac{1}{n} E \left\{ \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 \right\} = \\
 &= \frac{1}{n} E \left\{ \sum_{i=1}^n (\mathbf{x}_i - \hat{x} + \hat{x} - \bar{\mathbf{x}})^2 \right\} = \\
 &= \frac{1}{n} \sum_{i=1}^n \{ E((\mathbf{x}_i - \hat{x})^2) - E((\bar{\mathbf{x}} - \hat{x})^2) \} = \\
 &= \frac{1}{n} \left\{ n\sigma^2(\mathbf{x}) - n \left(\frac{1}{n} \sigma^2(\mathbf{x}) \right) \right\}, \\
 E(\mathbf{s}'^2) &= \frac{n-1}{n} \sigma^2(\mathbf{x}). \tag{2.6}
 \end{aligned}$$

Видно, что \mathbf{s}'^2 не является несмещенной оценкой $\sigma^2(\mathbf{x})$. Однако, заменив определение (2.5) на

$$\mathbf{s}^2 = \frac{1}{n-1} \{ (\mathbf{x}_1 - \bar{\mathbf{x}})^2 + (\mathbf{x}_2 - \bar{\mathbf{x}})^2 + \dots + (\mathbf{x}_n - \bar{\mathbf{x}})^2 \}, \tag{2.7}$$

можно получить несмещенную оценку.

Выражение $(n-1)$ в знаменателе хотя математически и оправдано, на первый взгляд выглядит несколько странно. Заметим, однако, что при $n=1$ выборочное среднее равно значению \mathbf{x} единственного элемента выборки ($\bar{\mathbf{x}} = \mathbf{x}$) и, следовательно, выборочная дисперсия, определенная с помощью выражения (2.5), обратится в нуль. Причина в том, что в (2.5) — как и в (2.7) — нам пришлось использовать выборочное среднее $\bar{\mathbf{x}}$ вместо среднего значения популяции \hat{x} , так как последнее неизвестно. Информация, содержащаяся в выборке, прежде всего должна быть использована для определения выборочного среднего. При вычислении выборочной дисперсии эта информация теряется. Таким образом, эффективное число элементов выборки уменьшается. Этот факт можно учесть уменьшением знаменателя среднего арифметического (2.5). Доказательство повторяется детально в разд. 6.4.

6.3. ВЫБОР ИЗ СОСТАВНЫХ РАСПРЕДЕЛЕНИЙ

Часто более предпочтительно разбить популяцию G (например, число всех студентов в Европе) на различные подпопуляции G_1, G_2, \dots, G_t (число студентов в университетах $1, 2, \dots, t$). Величину \mathbf{x} можно описать плотностями вероятности $f_1(x), f_2(x), \dots, f_t(x)$ каждой из популяций. Соответствующие функции

распределения вероятностей имеют вид

$$F_i(x) = \int_{-\infty}^x f_i(x) dx = P(\mathbf{x} < x | \mathbf{x} \in G_i). \quad (3.1)$$

Они имеют характер условных вероятностей, так как каждая из них определяется при условии, что элемент принадлежит соответствующей подпопуляции. Применим правило полной вероятности (3.4) гл. 2 для получения функции распределения вероятностей всей популяции

$$F(x) = P(\mathbf{x} < x | \mathbf{x} \in G) = \sum_{i=1}^t P(\mathbf{x} < x | \mathbf{x} \in G_i) P(\mathbf{x} \in G_i),$$

т. е.

$$F(x) = \sum_{i=1}^t P(\mathbf{x} \in G_i) F_i(x). \quad (3.2)$$

Соответственно плотность вероятности равна

$$f(x) = \sum_{i=1}^t P(\mathbf{x} \in G_i) f_i(x). \quad (3.3)$$

Заменяя $P(\mathbf{x} \in G_i)$ на p_i , можно получить выражение для среднего значения популяции:

$$\begin{aligned} \hat{x} = E(\mathbf{x}) &= \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} \sum_{i=1}^t x p_i f_i(x) dx = \\ &= \sum_{i=1}^t p_i \int_{-\infty}^{\infty} x f_i(x) dx, \end{aligned}$$

т. е.

$$\hat{x} = \sum_{i=1}^t p_i \hat{x}_i. \quad (3.4)$$

Математическое ожидание популяции равно среднему математическим ожиданиям подпопуляций, взятых со своими вероятностями. Для дисперсии популяции получим

$$\begin{aligned} \sigma^2(\mathbf{x}) &= \int_{-\infty}^{\infty} (x - \hat{x})^2 f(x) dx = \int_{-\infty}^{\infty} (x - \hat{x})^2 \sum_{i=1}^t p_i f_i(x) dx = \\ &= \sum_{i=1}^t p_i \int_{-\infty}^{\infty} \{(x - \hat{x}_i) + (\hat{x}_i - \hat{x})\}^2 f_i(x) dx. \end{aligned}$$

Так как отдельные x_i независимы, то члены с ковариациями равны нулю, следовательно,

$$\sigma^2(\mathbf{x}) = \sum_{i=1}^t p_i \left\{ \int_{-\infty}^{\infty} (x - \hat{x}_i)^2 f_i dx + (\hat{x}_i - \hat{x})^2 \int_{-\infty}^{\infty} f_i(x) dx \right\}, \quad (3.5)$$

$$\sigma^2(\mathbf{x}) = \sum_{i=1}^t p_i \{ \sigma_i^2 + (\hat{x}_i - \hat{x})^2 \}.$$

Выражение (3.5) — средневзвешенное значение сумм двух членов, первый из которых является дисперсией подпопуляции, второй — отклонением среднего значения подпопуляции от среднего значения популяции.

Возьмем теперь из каждой подпопуляции выборку объема n_i , причем

$$\sum_{i=1}^t n_i = n.$$

Среднее арифметическое полной составной выборки равно

$$\bar{\mathbf{x}}_p = \frac{1}{n} \sum_{i=1}^t \sum_{j=1}^{n_i} \mathbf{x}_{ij} = \frac{1}{n} \sum_{i=1}^t n_i \bar{\mathbf{x}}_i. \quad (3.6)$$

Оно имеет математическое ожидание

$$E(\bar{\mathbf{x}}_p) = \frac{1}{n} \sum_{i=1}^t n_i \hat{x}_i \quad (3.7)$$

и дисперсию

$$\begin{aligned} \sigma^2(\bar{\mathbf{x}}_p) &= E\{(\bar{\mathbf{x}}_p - E(\bar{\mathbf{x}}_p))^2\} = \\ &= E\left\{\left(\sum_{i=1}^t \frac{n_i}{n} (\bar{\mathbf{x}}_i - \hat{x}_i)\right)^2\right\} = \\ &= \frac{1}{n^2} \sum_{i=1}^t n_i^2 E\{(\bar{\mathbf{x}}_i - \hat{x}_i)^2\}, \\ \sigma^2(\bar{\mathbf{x}}_p) &= \frac{1}{n^2} \sum_{i=1}^t n_i^2 \sigma^2(\bar{\mathbf{x}}_i). \end{aligned} \quad (3.8)$$

Используя (2.4), получим

$$\sigma^2(\bar{\mathbf{x}}_p) = \frac{1}{n} \sum_{i=1}^t \frac{n_i}{n} \sigma_i^2. \quad (3.9)$$

Аналогичные результаты можно получить с помощью закона пространства ошибок, применяя (5.7) гл. 4 к (3.6).

Интуитивно ясно, что среднее арифметическое $\bar{\mathbf{x}}_p$ не может в общем случае служить оценкой среднего популяции $\hat{\mathbf{x}}$, так как $\bar{\mathbf{x}}_p$ зависит от произвольно выбранных объемов каждой выборки n_i . Сопоставление (3.7) с (3.4) показывает, что это возможно лишь в случае $p_i = n_i/n$.

Среднее значение популяции $\hat{\mathbf{x}}$ можно оценить, определив сначала выборочные средние $\bar{\mathbf{x}}_i$ подпопуляций и определив далее выражение

$$\bar{\mathbf{x}} = \sum_{i=1}^t p_i \bar{\mathbf{x}}_i \quad (3.10)$$

согласно (3.4). С помощью закона распространения ошибок получим для дисперсии

$$\sigma^2(\bar{\mathbf{x}}) = \sum_{i=1}^t p_i^2 \sigma^2(\bar{\mathbf{x}}_i) = \sum_{i=1}^t \frac{p_i^2}{n_i} \sigma_i^2. \quad (3.11)$$

Дисперсии подпопуляций в общем случае не бывают известны. Их можно оценить, используя (2.7). Тогда получим оценку

$$s_x = \sum_{i=1}^t \frac{p_i^2}{n_i} \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^2. \quad (3.12)$$

На практике мы стараемся получить оценку среднего популяции с возможно меньшей дисперсией для данного n . Для этого мы должны взять выборки таких объемов n_i , чтобы (3.11) приняло минимальное значение. Хотя в общем случае они не известны, их можно оценить путем предварительной выборки, используя (2.7). Во многих практических случаях важно изучить влияние отдельных подпопуляций на среднее значение популяции.

Обсуждение разбиения популяций будет продолжено в гл. 11.

6.4 ВЫБОР БЕЗ ВОЗВРАЩЕНИЯ ИЗ КОНЕЧНОЙ ПОПУЛЯЦИИ; СРЕДНЕКВАДРАТИЧНОЕ ОТКЛОНЕНИЕ; СТЕПЕНИ СВОБОДЫ

Впервые мы столкнулись с понятием «выбор» при рассмотрении гипергеометрического распределения в гл. 5, разд. 3. Указывалось, что независимость последовательных исходов не имеет места, так как отдельные элементы не возвращаются в конечную (дискретную) популяцию. В связи с этим теперь мы не будем рассматривать абсолютно случайный выбор.

Для дальнейшего рассмотрения возьмем популяцию, состоящую из N элементов y_1, y_2, \dots, y_N . Возьмем из нее выборку объема n с элементами $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. (В случае гипергеометрических

распределений y_j и вместе с ними x_i могли принимать значения только 0 и 1.) Так как все элементы популяции имеют одинаковую вероятность быть выбранными, среднее значение популяции равно

$$E(y) = \hat{y} = \bar{y} = \frac{1}{N} \sum_{j=1}^N y_j. \quad (4.1)$$

Не будучи случайной величиной, \hat{y} является средним арифметическим конечного числа элементов популяции. Таким образом, при определении дисперсии популяции мы сталкиваемся с трудностями, обсужденными в конце разд.

6.2. Согласно (2.7), определим

$$\begin{aligned} \sigma^2(y) &= \frac{1}{N-1} \sum_{j=1}^N (y_j - \bar{y})^2 = \\ &= \frac{1}{N-1} \left\{ \sum_{j=1}^N y_j^2 - \frac{1}{N} \left(\sum_{j=1}^N y_j \right)^2 \right\}. \end{aligned} \quad (4.2)$$

Рассмотрим теперь сумму квадратов

$$\sum_{j=1}^N (y_j - \bar{y})^2. \quad (4.3)$$

Так как мы не накладывали на популяцию никаких ограничений, y_j могут принимать любые значения. Следовательно, первый член суммы (4.3) также может принять любое значение. То же справедливо для 2-го, 3-го, ..., $(N-1)$ -го членов. Однако значение N -го члена определено, так как

$$\sum_{j=1}^N (y_j - \bar{y}) = 0. \quad (4.4)$$

В таком случае говорят, что число степеней свободы суммы (4.3) равно $(N-1)$. Это можно проиллюстрировать геометрически. Рассмотрим простой случай $\bar{y} = 0$ и образуем N -мерное векторное пространство с координатами y_i . При этом сумма квадратов будет соответствовать квадрату модуля вектора в этом пространстве. Вследствие равенства (4.4) конец этого вектора может перемещаться только в пространстве размерности $(N-1)$. Размерность такого подпространства в механике называется числом степеней свободы. На рис. 6.1 показан случай $N=2$. Здесь положение вектора ограничивается прямой $y_2 = -y_1$. Сумма квадратов [например,

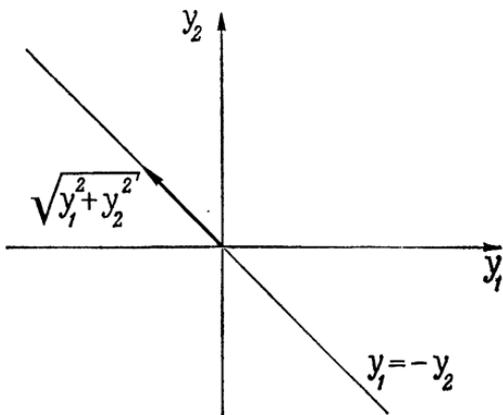


Рис. 6.1. Выборка объема 2 дает сумму квадратов с одной степенью свободы.

(4.3)], деленная на число степеней свободы, называется *средним квадратичным* или, более точно, *среднеквадратичным отклонением*. Это определение относится не только к популяциям, но и к выборкам. В последнем случае оно используется для характеристики отклонений отдельных измерений. Квадратный корень из среднеквадратичного отклонения (имеющий размерность измеряемой величины, т. е. среднего значения) называется *стандартным отклонением*.

Вернемся к выборке x_1, x_2, \dots, x_n . Чтобы облегчить рассуждения, введем символ Кронекера, описывающий процесс выбора. Определим

$$\delta_i^j = \begin{cases} 1, & \text{если } x_i \text{ равно элементу популяции } y_j; \\ 0, & \text{во всех других случаях.} \end{cases} \quad (4.5)$$

Это означает, что

$$x_i = \sum_{j=1}^N \delta_i^j y_j. \quad (4.6)$$

Так как каждый y_j с равной вероятностью может быть i -м элементом выборки, имеем

$$P(\delta_i^j = 1) = 1/N. \quad (4.7)$$

Величина δ_i^j , описывающая случайный процесс выбора, естественно является случайной. Ее математическое ожидание легко получить из выражения (3.2) гл. 3 (где $n = 2$, $x_1 = 0$, $x_2 = 1$):

$$E(\delta_i^j) = P(\delta_i^j = 1) = 1/N. \quad (4.8)$$

Если известен один элемент выборки, скажем x_i , то для другого элемента число возможных способов выбора из популяции равно $(N - 1)$. Следовательно,

$$P(\delta_i^j \delta_k^l = 1) = \frac{1}{N} \frac{1}{N-1} = E(\delta_i^j \delta_k^l). \quad (4.9)$$

Так как выборка производится без возвращения, один и тот же элемент не может быть выбран дважды, т. е. $j \neq l$ или

$$\delta_i^j \delta_k^j = 0. \quad (4.10)$$

Соответственно имеем

$$\delta_i^j \delta_i^l = 0, \quad (4.11)$$

так как один элемент выборки не может соответствовать двум различным элементам популяции. Математическое ожидание перво-

го элемента выборки \mathbf{x}_1 равно

$$\begin{aligned} E(\mathbf{x}_1) &= E\left\{\sum_{j=1}^N \delta_1^j y_j\right\} = \sum_{j=1}^N y_j E(\delta_1^j) = \\ &= \frac{1}{N} \sum_{j=1}^N y_j = \bar{y}. \end{aligned} \quad (4.12)$$

Теперь, поскольку элемент \mathbf{x}_1 ничем не отличается от других, математические ожидания любых других элементов выборки, а также их среднего арифметического принимают то же самое значение:

$$E(\bar{\mathbf{x}}) = \frac{1}{n} \sum_{i=1}^n E(\mathbf{x}_i) = \bar{y}, \quad (4.13)$$

которое, следовательно, дает несмещенную оценку среднего популяции.

Для того чтобы выразить математическое ожидание выборочной дисперсии

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2, \quad (4.14)$$

используем тождество

$$\sum_{i=1}^n \mathbf{x}_i^2 = n\bar{\mathbf{x}}^2 + \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 \quad (4.15)$$

и определим математические ожидания $\sum \mathbf{x}_i^2$ и $n\bar{\mathbf{x}}^2$. Из (4.6) имеем

$$\mathbf{x}_i^2 = \sum_{j=1}^N \delta_i^j y_j^2$$

и, следовательно,

$$E(\mathbf{x}_1^2) = \frac{1}{N} \sum_{j=1}^N y_j^2,$$

что с помощью (4.2) преобразуется так:

$$E(\mathbf{x}_1^2) = \frac{1}{N} \{(N-1)\sigma^2(y) + N\bar{y}^2\} = \bar{y}^2 + \left(1 - \frac{1}{N}\right)\sigma^2(y).$$

Используя снова то, что \mathbf{x}_1 ничем не отличается от других \mathbf{x}_i , получаем

$$E\left\{\sum_{i=1}^n \mathbf{x}_i^2\right\} = nE(\mathbf{x}_1^2) = n\left\{\bar{y}^2 + \left(1 - \frac{1}{N}\right)\sigma^2(y)\right\}. \quad (4.16)$$

Для определения математического ожидания $n\bar{x}^2$ рассмотрим сначала

$$\begin{aligned} E(\mathbf{x}_1 \mathbf{x}_2) &= E\left(\sum_{j \neq l} \delta_1^j \delta_2^l y_j y_l\right) = \\ &= \sum_{j \neq l} E\{\delta_1^j \delta_2^l y_j y_l\}. \end{aligned}$$

В силу (4.9)

$$E(\mathbf{x}_1 \mathbf{x}_2) = \frac{1}{N(N-1)} \sum_{j \neq l} y_j y_l.$$

Сумма берется по всем неравным индексам и содержит любую комбинацию дважды ($y_j y_l$ и $y_l y_j$). Можно написать

$$\begin{aligned} E(\mathbf{x}_1 \mathbf{x}_2) &= \frac{1}{N(N-1)} \left\{ \left(\sum_{k=1}^N y_k \right)^2 - \sum_{k=1}^N y_k^2 \right\} = \\ &= \frac{1}{N(N-1)} \{ (N\bar{y})^2 - [(N-1)\sigma^2(y) + N\bar{y}^2] \} = \\ &= \bar{y}^2 - \frac{\sigma^2(y)}{N}. \end{aligned}$$

Вследствие симметрии этот результат справедлив для $E(\mathbf{x}_i \mathbf{x}_k)$ при любых $i, k, i \neq k$. Тогда

$$E\left(\sum_{i \neq k} \mathbf{x}_i \mathbf{x}_k\right) = n(n-1) \left\{ \bar{y}^2 - \frac{\sigma^2(y)}{N} \right\} \quad (4.17)$$

и, наконец,

$$\begin{aligned} E(n\bar{x}^2) &= nE(\bar{x}^2) = \frac{1}{n} E\left\{ \left(\sum_{i=1}^n \mathbf{x}_i \right)^2 \right\} = \\ &= \frac{1}{n} E\left\{ \sum_{i=1}^n \mathbf{x}_i^2 + \sum_{i \neq k} \mathbf{x}_i \mathbf{x}_k \right\} = \\ &= \frac{1}{n} E\left(\sum_{i=1}^n \mathbf{x}_i^2 \right) + \frac{1}{n} E\left(\sum_{i \neq k} \mathbf{x}_i \mathbf{x}_k \right). \end{aligned}$$

Сопоставляя (4.16) и (4.17), найдем

$$\begin{aligned} E(n\bar{x}^2) &= \bar{y}^2 + \left(1 - \frac{1}{N}\right) \sigma^2(y) + (n-1) \bar{y}^2 - \frac{n-1}{N} \sigma^2(y) = \\ &= \left(1 - \frac{n}{N}\right) \sigma^2(y) + n\bar{y}^2. \end{aligned} \quad (4.18)$$

Используя (4.18) и (4.16), можно теперь определить математическое ожидание выборочной дисперсии (4.14):

$$\begin{aligned}
 E(\mathbf{s}_x^2) &= \frac{1}{n-1} E\left\{\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2\right\} = \frac{1}{n-1} \left\{E\left(\sum_{i=1}^n \mathbf{x}_i^2\right) - E(n\bar{\mathbf{x}}^2)\right\} = \\
 &= \frac{1}{n-1} \left\{n\left[\bar{y}^2 + \left(1 - \frac{1}{N}\right)\sigma^2(y)\right] - \left(1 - \frac{n}{N}\right)\sigma^2(y) - n\bar{y}^2\right\} = \\
 &= \frac{1}{n-1} \left\{n - \frac{n}{N} - 1 + \frac{n}{N}\right\} \sigma^2(y), \\
 E(\mathbf{s}_x^2) &= \sigma^2(y). \tag{4.19}
 \end{aligned}$$

Она является несмещенной оценкой дисперсии популяции.

Рассмотрим, наконец, дисперсию выборочного среднего

$$\sigma^2(\bar{\mathbf{x}}) = E\{(\bar{\mathbf{x}} - E(\bar{\mathbf{x}}))^2\}.$$

Поскольку $E(\bar{\mathbf{x}}) = \bar{y}$ является не случайной, а детерминированной величиной, то

$$\begin{aligned}
 \sigma^2(\bar{\mathbf{x}}) &= E(\bar{\mathbf{x}}^2) - \bar{y}^2 = \frac{1}{n} \left\{ \left(1 - \frac{n}{N}\right)\sigma^2(y) + n\bar{y}^2 \right\} - \bar{y}^2, \\
 \sigma^2(\bar{\mathbf{x}}) &= \frac{\sigma^2(y)}{n} \left(1 - \frac{n}{N}\right). \tag{4.20}
 \end{aligned}$$

Сравнение со случаем бесконечной популяции (2.4) обнаруживает появление добавочного члена $(1 - n/N)$. При $n = N$, когда выборочное среднее и среднее популяции становятся идентичными, из-за этого члена дисперсия (4.20) обращается в нуль.

6.5. ВЫБОРКА ИЗ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ; РАСПРЕДЕЛЕНИЕ χ^2

В этом разделе мы вернемся к непрерывным популяциям и в особенности остановимся на популяции, описываемой нормальным распределением со средним a и дисперсией σ^2 . Согласно равенству (8.4) гл. 5, характеристическая функция такого распределения равна

$$\varphi_x(t) = \exp(ita) \exp\left(-\frac{1}{2}\sigma^2 t^2\right). \tag{5.1}$$

Произведем из популяции выборку объема n . Характеристическая функция выборочного среднего записывается через характеристическую функцию популяции согласно выражению (2.3). Следовательно, имеем

$$\varphi_{\bar{x}}(t) = \left\{ \exp\left(i \frac{t}{n} a - \frac{\sigma^2}{2} \left(\frac{t}{n}\right)^2\right) \right\}^n. \tag{5.2}$$

Если вместо \bar{x} рассмотреть $(\bar{x} - a) = (\bar{x} - \hat{x})$, получим

$$\varphi_{\bar{x}-a}(t) = \exp\left(-\frac{\sigma^2 t^2}{2n}\right). \quad (5.3)$$

Эта функция также является характеристической функцией нормального распределения с дисперсией

$$\sigma^2(\bar{x}) = \sigma^2(\mathbf{x})/n. \quad (5.4)$$

В простейшем случае стандартного нормального распределения ($a = 0, \sigma^2 = 1$) имеем

$$\varphi_{\bar{x}}(t) = \exp(-t^2/2n). \quad (5.5)$$

Произведем выборку

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$$

из этого распределения и образуем сумму квадратов элементов выборки

$$\chi^2 = \mathbf{x}_1^2 + \mathbf{x}_2^2 + \dots + \mathbf{x}_n^2. \quad (5.6)$$

Покажем, что величина* χ^2 имеет функцию распределения:

$$F(\chi^2) = \frac{1}{\Gamma(\lambda) 2^\lambda} \int_0^{\chi^2} u^{\lambda-1} e^{-\frac{1}{2}u} du, \quad (5.7)$$

где

$$\lambda = \frac{1}{2}n. \quad (5.8)$$

Здесь n называется *числом степеней свободы*. Обозначив

$$\frac{1}{\Gamma(\lambda) 2^\lambda} = k, \quad (5.9)$$

найдем плотность вероятности

$$f(\chi^2) = k(\chi^2)^{\lambda-1} e^{-\frac{1}{2}\chi^2}. \quad (5.10)$$

Докажем сначала справедливость (5.7) для простого случая одной степени свободы, т. е. $\lambda = \frac{1}{2}$. Функция распределения имеет вид

$$\begin{aligned} F(\chi^2) &= P(\mathbf{x}^2 < \chi^2) = P(-\sqrt{\chi^2} < \mathbf{x} < +\sqrt{\chi^2}) = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{\chi^2}}^{+\sqrt{\chi^2}} e^{-\frac{1}{2}x^2} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{\chi^2}} e^{-\frac{1}{2}x^2} dx. \end{aligned}$$

* Символ χ^2 (хи-квадрат) был введен К. Пирсоном. Хотя он содержит показатель 2, напоминая о том, что он связан с суммой квадратов, мы будем рассматривать в наших уравнениях χ^2 как исходную величину, а не как квадрат какого-то числа.

Произведя замену $x^2 = u$, т. е. $du = 2x dx$, будем иметь

$$F(x^2) = \frac{1}{\sqrt{2\pi}} \int_0^{x^2} u^{-\frac{1}{2}} e^{-\frac{1}{2}u} du, \quad (5.11)$$

равную (5.7) при $\lambda = 1/2$. Прежде чем провести основное доказательство, напомним характеристическую функцию

$$\varphi_{\chi^2}(t) = \int_0^{\infty} k(\chi^2)^{\lambda-1} \exp\left\{-\frac{\chi^2}{2} + it\chi^2\right\} d\chi^2, \quad (5.12)$$

которая после преобразования интеграла с помощью подстановки $v = \left(\frac{1}{2} - it\right)\chi^2$ принимает вид

$$\varphi_{\chi^2}(t) = 2^\lambda (1 - 2it)^{-\lambda} k \int_0^{\infty} v^{\lambda-1} e^{-v} dv.$$

Согласно приложению (Г. 1), последний интеграл равен $\Gamma(\lambda)$. Тогда

$$\varphi_{\chi^2}(t) = (1 - 2it)^{-\lambda}. \quad (5.13)$$

Очевидно, характеристическая функция другого распределения χ^2 с λ' степенями свободы равна

$$\varphi'_{\chi^2}(t) = (1 - 2it)^{-\lambda'}.$$

Так как характеристическая функция суммы равна произведению характеристических функций отдельных слагаемых, приходим к важной теореме:

Распределение суммы двух различных χ^2 -распределений соответственно с n_1 и n_2 степенями свободы также является χ^2 -распределением с $n = n_1 + n_2$ степенями свободы.

Применяя эту теорему к (5.11), немедленно получаем доказательство (5.7), так как отдельные слагаемые в сумме квадратов (5.6) независимы и ее можно представить как сумму различных распределений χ^2 с одной степенью свободы.

Для вычисления математического ожидания и дисперсии распределения χ^2 используются производные характеристической функции согласно (6.7) гл. 5:

$$E(\chi^2) = i\varphi'(0) = 2\lambda$$

или

$$E(\chi^2) = n. \quad (5.14)$$

Далее,

$$E\{(\chi^2)^2\} = -\varphi''(0) = 4\lambda^2 + 4\lambda,$$

что дает

$$\begin{aligned}\sigma^2(\chi^2) &= E\{(\chi^2)^2\} - \{E(\chi^2)\}^2 = 4\lambda, \\ \sigma^2(\chi^2) &= 2n.\end{aligned}\quad (5.15)$$

Среднее χ^2 -распределения равно числу степеней свободы; дисперсия — удвоенному числу степеней свободы. На рис. 6.2 приведены распределения χ^2 для различных значений n . Как видно из (5.10),

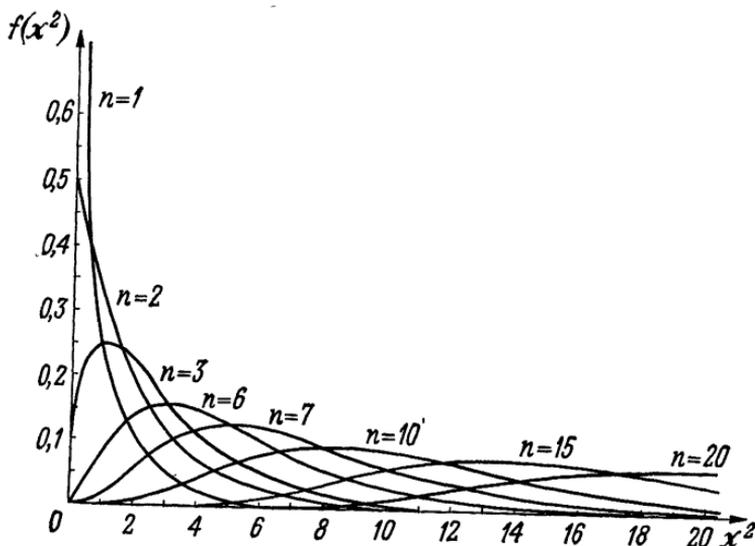


Рис. 6.2. Плотность вероятности χ^2 .

в случае $n = 1$ распределение обращается в ∞ в точке $\chi^2 = 0$. В случае $n = 2$ $f(\chi^2)$ равна $1/2$ при $\chi^2 = 0$; для $n \geq 3$ $f(\chi^2)$ в этой точке обращается в нуль.

Для трех и более степеней свободы график распределения имеет типичную скошенную колоколообразную форму и с увеличением n становится все более симметричным. Некоторые χ^2 -распределения приведены в табл. Е.4.

Распределение χ^2 особенно важно в тех применениях, где величина χ^2 используется как мера достоверности некоторого частного результата (разд. 8.6). Функция распределения

$$F(\chi^2) = P(\chi^2 < \chi^2) \quad (5.16)$$

дает вероятность того, что случайная величина χ^2 не превышает заданного значения χ^2 . На практике в качестве меры достоверности используется величина

$$W(\chi^2) = 1 - F(\chi^2). \quad (5.17)$$

$W(\chi^2)$ иногда называют уровнем значимости.

Уровень значимости велик при малых χ^2 и уменьшается с увеличением χ^2 . На рис. 6.3 представлена функция распределения (5.16) для нескольких значений n . На том же рисунке на оси справа отмечены значения $W(\chi^2)$. Обратная функция, описывающая квантили χ^2 -распределения,

$$\chi_F^2 = \chi^2(F) = \chi^2(1 - W) \quad (5.18)$$

обычно используется при «проверке гипотез» для установления интервала, внутри которого гипотеза принимается (разд. 8.6). Табулированные значения функции χ_F^2 представлены в табл. Е. 5.

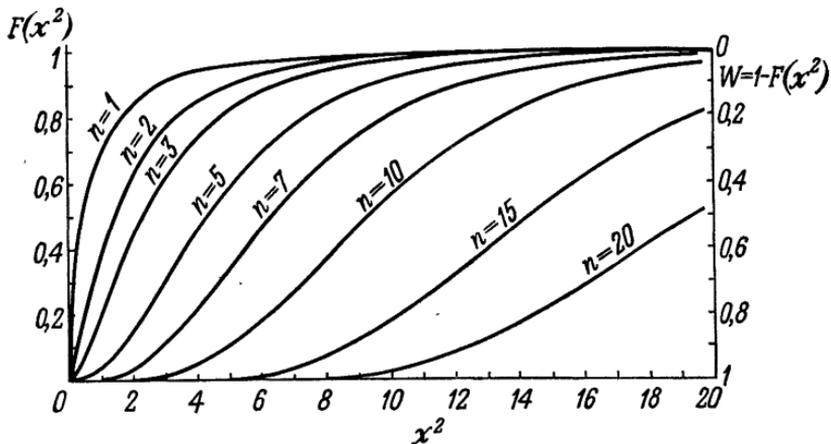


Рис. 6.3. Функция распределения χ^2 .

Остается отметить, что в общем случае нормально распределенной популяции со средним a и дисперсией σ^2 распределению χ^2 подчиняется не сумма квадратов (5.6), а величина

$$\chi^2 = \frac{(x_1 - a)^2 + (x_2 - a)^2 + \dots + (x_n - a)^2}{\sigma^2}. \quad (5.19)$$

Это сразу видно из (9.5) гл. 5.

6.6. χ^2 И ЭМПИРИЧЕСКАЯ ДИСПЕРСИЯ

При рассмотрении несмещенной состоятельной оценки дисперсии популяции мы определили величину (2.7):

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (6.1)$$

Рассмотрим теперь специальный случай, когда выборка независимых x_i производится из нормального распределения со средним

$\hat{\lambda} = a$ и дисперсией σ^2 . Мы собираемся показать, что случайная величина

$$\frac{n-1}{\sigma^2} \mathbf{s}^2 \quad (6.2)$$

подчиняется χ^2 -распределению с $\lambda = n - 1$ степенями свободы.

Сделаем сначала ортогональное преобразование (разд. 4.5) n переменных \mathbf{x}_i :

$$\begin{aligned} \mathbf{y}_1 &= \frac{1}{\sqrt{1-2}} (\mathbf{x}_1 - \mathbf{x}_2), \\ \mathbf{y}_2 &= \frac{1}{\sqrt{2-3}} (\mathbf{x}_1 + \mathbf{x}_2 - 2\mathbf{x}_3), \\ \mathbf{y}_3 &= \frac{1}{\sqrt{3-4}} (\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 - 3\mathbf{x}_4) \\ &\vdots \\ \mathbf{y}_{n-1} &= \frac{1}{\sqrt{(n-1)n}} (\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_{n-1} - (n-1)\mathbf{x}_n), \\ \mathbf{y}_n &= \frac{1}{\sqrt{n}} (\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_n) = \sqrt{n} \bar{\mathbf{x}}. \end{aligned} \quad (6.3)$$

Легко проверить, что это преобразование в самом деле ортогонально, т. е.

$$\sum_{i=1}^n \mathbf{x}_i^2 = \sum_{i=1}^n \mathbf{y}_i^2. \quad (6.4)$$

Так как сумма (а также и разность) независимых нормально распределенных величин также нормально распределена (разд. 5.8), все \mathbf{y}_i распределены нормально. Числовые множители в (6.3) обеспечивают их нулевое среднее и единичную дисперсию. Можно теперь переписать (6.1):

$$\begin{aligned} (n-1) \mathbf{s}^2 &= \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 = \sum_{i=1}^n \mathbf{x}_i^2 - 2\bar{\mathbf{x}} \sum_{i=1}^n \mathbf{x}_i + n\bar{\mathbf{x}}^2 = \\ &= \sum_{i=1}^n \mathbf{x}_i^2 - n\bar{\mathbf{x}}^2 = \sum_{i=1}^n \mathbf{y}_i^2 - \mathbf{y}_n^2 = \sum_{i=1}^{n-1} \mathbf{y}_i^2. \end{aligned}$$

Это выражение является лишь суммой $(n-1)$ независимых квадратов. Сравнение с (5.19) показывает, что величина (6.2) на самом деле подчиняется χ^2 -распределению с $(n-1)$ степенями свободы.

Квадраты $(\mathbf{x}_i - \bar{\mathbf{x}})^2$ не являются независимыми, так как $(\mathbf{x}_i - \bar{\mathbf{x}})^2$ связаны уравнением

$$\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}) = 0.$$

Можно показать, что любое дополнительное уравнение, связывающее эти квадраты, уменьшает число степеней свободы на единицу. Мы будем часто пользоваться этим общим результатом, приведенным здесь без доказательства.

МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

7.1. ОТНОШЕНИЕ ПРАВДОПОДОБИЯ; ФУНКЦИЯ ПРАВДОПОДОБИЯ

В предыдущей главе был рассмотрен вопрос об оценивании параметра. Мы также описали требуемые свойства оценок, хотя и не останавливались на методах их построения в частных случаях. Оценки были получены только для таких важных величин, как математическое ожидание и дисперсия. Рассмотрим теперь решение более общей задачи.

Пусть p параметров

$$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$$

полностью характеризуют распределение случайных величин

$$\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n).$$

Запишем плотность вероятности этих величин в виде

$$f = f(\mathbf{x}; \lambda). \quad (1.1)$$

Пусть одиночный эксперимент, т. е. единственное измерение вектора \mathbf{x} , или выборка единичного объема*, приводит к результату $\mathbf{x}^{(i)}$. Этому эксперименту можно поставить в соответствие число

$$dP^{(i)} = f(\mathbf{x}^{(i)}; \lambda) d\mathbf{x}. \quad (1.2)$$

Число $dP^{(i)}$ имеет характер *апостериорной вероятности*** . Оно показывает — после того как результат эксперимента известен, — с какой вероятностью можно было ожидать этот результат. Для выборки из N (независимых) элементов или экспериментов эта вероятность определяется в соответствии с (3.6) гл. 2

$$dP = \prod_{i=1}^N f(\mathbf{x}^{(i)}; \lambda) d\mathbf{x}, \quad (1.3)$$

причем dP является функцией от параметров λ . Существуют случаи, когда известно, что популяции соответствует один из двух

* Выборка состоит из одного вектора, имеющего n компонент. — Прим. ред.

**Если λ является случайной величиной с равномерным распределением в некоторой области, то dP пропорционально в этой области апостериорной вероятности λ при фиксированном \mathbf{x} . — Прим. ред.

возможных, отличных друг от друга наборов параметров λ_1 или λ_2 . Такой случай можно привести, например, из ядерной физики, когда при проверке состояния оно может оказаться как «четным», так и «нечетным». образуем отношение

$$Q = \frac{\prod_{j=1}^N f(x^{(j)}; \lambda_1)}{\prod_{j=1}^N f(x^{(j)}; \lambda_2)}. \quad (1.4)$$

Результат выбора можно выразить фразой: «Набор параметров λ_1 в Q раз вероятнее, чем набор λ_2 ». Выражение (1.4) называется *отношением правдоподобия*, а функция

$$L = \prod_{j=1}^N f(x^{(j)}; \lambda) \quad (1.5)$$

— *функцией правдоподобия*. Очень важно подчеркнуть различие между функцией типа (1.3) и обычной априорной вероятностью*.

Пример 7.1. Асимметричная монета бросается несколько раз. Постараемся решить, исходя из результата эксперимента, принадлежит ли эта монета к классу А или В, каждый из которых характеризуется следующими вероятностями:

	Класс А	Класс В
«Орел»	$\frac{1}{3}$	$\frac{2}{3}$
«Решка»	$\frac{2}{3}$	$\frac{1}{3}$

Из пяти бросаний один раз выпал «орел» и четыре раза — «решка». Тогда

$$L_A = \frac{1}{3} \left(\frac{2}{3}\right)^4, \quad L_B = \frac{2}{3} \left(\frac{1}{3}\right)^4$$

и

$$Q = L_A/L_B = 8.$$

Это говорит о том, что монета с большей вероятностью может быть отнесена к классу А.

* Автор, по-видимому, хочет подчеркнуть, что в функции правдоподобия (1.5) векторы $x^{(j)}$ фиксированы, а аргументом является λ . — *Прим. ред.*

7.2. ПРИНЦИП МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Теперь становится ясно, как обобщить рассмотренное выше понятие отношения правдоподобия. Мы получаем, что наибольшего доверия заслуживают значения параметров, при которых функция правдоподобия (1.5) принимает максимальное значение.

На рис. 7.1 приведены несколько возможных функций правдоподобия, зависящих от одного параметра. Часто встречается наиболее желательный случай симметричной колоколообразной функции правдоподобия. В этом случае значение $\tilde{\lambda}$ параметра λ , при котором функция правдоподобия $L(\lambda)$ максимальна, несомненно, можно принять в качестве лучшей оценки λ . Квадратный корень из дисперсии задаваемого функцией правдоподобия распределения относительно $\tilde{\lambda}$ можно принять в качестве ошибки оценки*. В случае асимметричного распределения $\tilde{\lambda}$ также является наилучшей оценкой, но дисперсия, хотя и не теряет своего математического смысла, служит неудовлетворительным параметром для описания поведения функции правдоподобия в окрестности своего максимума. В этом случае желательно вместе с $\tilde{\lambda}$ приводить и само распределение $L(\lambda)$. Может случиться, что $L(\lambda)$ имеет несколько локальных максимумов. Будем в этом случае отдавать предпочтение наибольшему максимуму функции правдоподобия. Однако вычисления могут быть весьма сложными, особенно в случае нескольких параметров λ . При нескольких максимумах приблизительно равной величины и малом числе наблюдений необходима крайняя осторожность.

Для определения положения максимума можно просто использовать стандартные методы дифференциального исчисления, т. е. найти первую производную от функции правдоподобия по λ и приравнять ее нулю. Однако дифференцирование произведения, имеющего большое число сомножителей, крайне затруднительно. Поэтому сначала возьмем логарифм от L

$$l = \ln L = \sum_{i=1}^N \ln f(x^{(i)}; \lambda). \quad (2.1)$$

Эту функцию часто также называют функцией правдоподобия.

* Плотность этого распределения задается функцией $kL(\lambda)$, где k — нормирующий множитель, такой, что интеграл от $kL(\lambda)$ по всей области возможных значений λ равен 1. — *Прим. ред.*

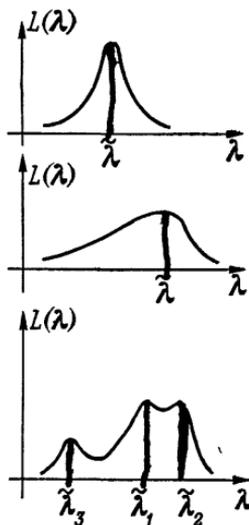


Рис. 7.1. Функции правдоподобия.

Для того чтобы различать определения (1.5) и (2.1), будем называть последнее *логарифмической функцией правдоподобия*. Очевидно, что положения максимумов (1.5) и (2.1) одинаковы. Следовательно, для оценки наиболее правдоподобного значения λ достаточно продифференцировать (2.1).

Сначала рассмотрим случай одного параметра, т. е. $\lambda = \lambda$. Задача тогда упрощается и сводится к решению так называемого *уравнения правдоподобия*

$$l' = dl/d\lambda = 0. \quad (2.2)$$

Используя (2.1), имеем

$$l' = \sum_{j=1}^N \frac{d}{d\lambda} \ln f(\mathbf{x}^{(j)}; \lambda) = \sum_{j=1}^N \frac{f'}{f} = \sum_{j=1}^N \varphi(\mathbf{x}^{(j)}; \lambda). \quad (2.3)$$

Здесь

$$\varphi(\mathbf{x}^{(j)}; \lambda) = \frac{\frac{d}{d\lambda} f(\mathbf{x}^{(j)}; \lambda)}{f(\mathbf{x}^{(j)}; \lambda)} \quad (2.4)$$

есть принятое сокращенное обозначение *логарифмической производной* функции f по λ . В общем случае p параметров уравнение правдоподобия (2.2) превращается в систему p уравнений

$$\frac{\partial l}{\partial \lambda_i} = 0, \quad i = 1, 2, \dots, p. \quad (2.5)$$

Пример 7.2. Рассмотрим важный случай оценивания среднего значения при *повторных измерениях*. Измерения $\mathbf{x}^{(j)}$ в общем случае имеют различную точность: например, они могут быть получены с помощью разной аппаратуры или различных методов. Предположим, что ошибки измерений распределены нормально. Следовательно, отдельное измерение соответствует выборке единичного объема из нормального распределения со средним λ и дисперсией σ_j^2 . Тогда вероятность значения $\mathbf{x}^{(j)}$ равна

$$f(\mathbf{x}^{(j)}; \lambda) dx = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(\mathbf{x}^{(j)} - \lambda)^2}{2\sigma_j^2}\right) dx.$$

Если произведены N измерений, то соответствующая функция правдоподобия равна

$$L = \prod_{j=1}^N \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(\mathbf{x}^{(j)} - \lambda)^2}{2\sigma_j^2}\right) \quad (2.6)$$

и ее логарифм

$$l = -\frac{1}{2} \sum_{j=1}^N \frac{(\mathbf{x}^{(j)} - \lambda)^2}{\sigma_j^2} + \text{const.} \quad (2.7)$$

Уравнение правдоподобия принимает вид

$$\frac{dl}{d\lambda} = \sum_{j=1}^N \frac{x^{(j)} - \lambda}{\sigma_j^2} = 0.$$

Его решение:

$$\tilde{\lambda} = \frac{\sum_{j=1}^N \frac{x^{(j)}}{\sigma_j^2}}{\sum_{j=1}^N \frac{1}{\sigma_j^2}}. \quad (2.8)$$

Результат, соответствующий максимальному правдоподобию, есть среднее от отдельных измерений, взятых с весами, обратными дисперсиям.

Пример 7.3. Подобно тому как это было в задаче с монетой, рассмотренной в начале этой главы, оцениваемый параметр может иногда принимать только дискретные значения. В примере 5.2 обсуждался вопрос оценки зоологической популяции с помощью меток с возвращением. Согласно выражению (3.1) гл. 5, вероятность среди n рыб, пойманных из пруда, найти k меченых, определяется как

$$L(k; n, K, N) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}},$$

где N — неизвестное число рыб в пруду, K — известное число меченых рыб в пруду. Требуется найти значение N , при котором функция L достигает максимума. Рассмотрим отношение

$$\frac{L(k; n, K, N)}{L(k; n, K, N-1)} = \frac{(N-n)(N-K)}{(N-n-K+k)N} \begin{cases} > 1, \text{ если } Nk < nK, \\ < 1, \text{ если } Nk > nK. \end{cases}$$

Следовательно, наибольшее значение L принимает тогда, когда N — ближайшее целое число к nK/k .

7.3. НЕРАВЕНСТВО ИНФОРМАЦИИ; ОЦЕНКИ С МИНИМАЛЬНОЙ ДИСПЕРСИЕЙ И ДОСТАТОЧНЫЕ ОЦЕНКИ

Вернемся к вопросу построения оценок с желаемыми свойствами. В разд. 6.1 мы определили оценку \mathbf{S} как несмещенную, если для любой выборки смещение равно нулю

$$B(\lambda) = E(\mathbf{S}) - \lambda = 0. \quad (3.1)$$

Однако для «хорошей» оценки несмещенности недостаточно. Потребуем также, чтобы дисперсия

$$\sigma^2(\mathbf{S})$$

была возможно меньшей. Часто необходимо искать компромисс между минимальностями смещения и дисперсии, так как эти величины связаны *неравенством информации**. То, что между ними существует связь, видно из того, что всегда можно построить оценку с $\sigma^2(\mathbf{S}) = 0$, просто выбрав в качестве \mathbf{S} любую константу. Рассмотрим оценку $\mathbf{S}(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)})$, как и прежде, как функцию выборки $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$. Совместная плотность вероятности (случайной) выборки, согласно (1.3) и (1.4) гл. 6, равна

$$f(x^{(1)}, x^{(2)}, \dots, x^{(N)}; \lambda) = f(x^{(1)}; \lambda) f(x^{(2)}; \lambda) \dots f(x^{(N)}; \lambda).$$

Математическое ожидание \mathbf{S} равно

$$E(\mathbf{S}) = \int \mathbf{S}(x^{(1)}, x^{(2)}, \dots, x^{(N)}) f(x^{(1)}; \lambda) f(x^{(2)}; \lambda) \dots f(x^{(N)}; \lambda) \times \\ \times dx^{(1)} dx^{(2)} dx^{(N)}. \quad (3.2)$$

Согласно (3.1), имеем также

$$E(\mathbf{S}) = B(\lambda) + \lambda.$$

Предположим теперь, что подынтегральное выражение дифференцируемо по λ . Дифференцируя, получим

$$1 + B'(\lambda) = \int \mathbf{S} \sum_{i=1}^N \frac{f'(x^{(i)}; \lambda)}{f(x^{(i)}; \lambda)} f(x^{(1)}; \lambda) f(x^{(2)}; \lambda) \dots f(x^{(N)}; \lambda) \times \\ \times dx^{(1)} dx^{(2)} \dots dx^{(N)},$$

что эквивалентно

$$1 + B'(\lambda) = E \left\{ \mathbf{S} \sum_{i=1}^N \frac{f'(x^{(i)}; \lambda)}{f(x^{(i)}; \lambda)} \right\} = E \left\{ \mathbf{S} \sum_{i=1}^N \varphi(x^{(i)}; \lambda) \right\}.$$

Из равенства (2.3) имеем

$$I' = \sum_{i=1}^N \varphi(x^{(i)}; \lambda).$$

Следовательно,

$$1 + B'(\lambda) = E\{\mathbf{S}I'\}. \quad (3.3)$$

Очевидно, равенство

$$\int f(x^{(1)}; \lambda) f(x^{(2)}; \lambda) \dots f(x^{(N)}; \lambda) dx^{(1)} dx^{(2)} \dots dx^{(N)} = 1.$$

* Это неравенство было независимо найдено Г. Крамером, М. Фреше, С. Р. Рао и другими авторами. Его также называют неравенством Крамера—Рао или неравенством Фреше.

Дифференцируя его, получаем

$$\int \sum_{j=1}^N \frac{f'(x^{(j)}; \lambda)}{f(x^{(j)}; \lambda)} f(x^{(1)}; \lambda) \dots f(x^{(N)}; \lambda) dx^{(1)} \dots dx^{(N)} = E(l') = 0.$$

Умножим это равенство на $E(\mathbf{S})$ и результат вычтем из (3.3); будем иметь

$$1 + B'(\lambda) = E\{\mathbf{S}l'\} - E(\mathbf{S})E(l') = E\{[\mathbf{S} - E(\mathbf{S})]l'\}. \quad (3.4)$$

Для оценки этого выражения необходимо рассмотреть неравенство типа Коши — Шварца в следующем виде:

если \mathbf{x} и \mathbf{y} — случайные величины, такие, что \mathbf{x}^2 и \mathbf{y}^2 имеют конечные математические ожидания, то

$$\{E(\mathbf{xy})\}^2 \leq E(\mathbf{x}^2)E(\mathbf{y}^2). \quad (3.5)$$

Для доказательства этого неравенства рассмотрим выражение

$$E((a\mathbf{x} + \mathbf{y})^2) = a^2E(\mathbf{x}^2) + 2aE(\mathbf{xy}) + E(\mathbf{y}^2) \geq 0, \quad (3.6)$$

которое неотрицательно для любых значений a . Равенство этого выражения нулю приводит к квадратному уравнению для a с решением

$$a_{1,2} = \frac{E(\mathbf{xy})}{E(\mathbf{x}^2)} \pm \sqrt{\left(\frac{E(\mathbf{xy})}{E(\mathbf{x}^2)}\right)^2 - \frac{E(\mathbf{y}^2)}{E(\mathbf{x}^2)}}. \quad (3.7)$$

Неравенство (3.6) выполняется, если дискриминант в (3.7) меньше или равен нулю

$$\left(\frac{E(\mathbf{xy})}{E(\mathbf{x}^2)}\right)^2 - \frac{E(\mathbf{y}^2)}{E(\mathbf{x}^2)} \leq 0.$$

Это неравенство эквивалентно (3.5). Применяя теперь (3.5) к (3.4), получим

$$\{1 + B'(\lambda)\}^2 \leq E\{[\mathbf{S} - E(\mathbf{S})]^2\} E(l'^2). \quad (3.8)$$

Воспользуемся (2.3) и перепишем выражение для

$$\begin{aligned} E(l'^2) &= E\left\{\left(\sum_{j=1}^N \varphi(\mathbf{x}^{(j)}; \lambda)\right)^2\right\} = \\ &= E\left\{\sum_{j=1}^N (\varphi(\mathbf{x}^{(j)}; \lambda))^2\right\} + E\left\{\sum_{i \neq j} \varphi(\mathbf{x}^{(i)}; \lambda) \varphi(\mathbf{x}^{(j)}; \lambda)\right\}. \end{aligned}$$

Все смешанные произведения в правой части выражения равны нулю, так как

$$E \{ \varphi(\mathbf{x}^{(i)}; \lambda) \varphi(\mathbf{x}^{(j)}; \lambda) \} = E \{ \varphi(\mathbf{x}^{(i)}; \lambda) \} E \{ \varphi(\mathbf{x}^{(j)}; \lambda) \},$$

$$E \{ \varphi(\mathbf{x}; \lambda) \} = \int_{-\infty}^{\infty} \frac{f'(x; \lambda)}{f(x; \lambda)} f(x; \lambda) dx = \int_{-\infty}^{\infty} f'(x; \lambda) dx$$

и

$$\int_{-\infty}^{\infty} f(x; \lambda) dx = 1.$$

Дифференцируя последнее равенство по λ , получаем

$$\int_{-\infty}^{\infty} f'(x; \lambda) dx = 0.$$

Следовательно,

$$E(l'^2) = E \left\{ \sum_{j=1}^N (\varphi(\mathbf{x}^{(j)}; \lambda))^2 \right\} = E \left\{ \sum_{j=1}^N \left(\frac{f'(\mathbf{x}^{(j)}; \lambda)}{f(\mathbf{x}^{(j)}; \lambda)} \right)^2 \right\}.$$

Так как слагаемые суммы независимы, математическое ожидание суммы равно сумме математических ожиданий, которые все одинаковы. Следовательно,

$$I(\lambda) = E(l'^2) = NE \left\{ \left(\frac{f'(\mathbf{x}; \lambda)}{f(\mathbf{x}; \lambda)} \right)^2 \right\}.$$

Это выражение называется *информацией выборки* относительно λ . Она всегда неотрицательна и равна нулю, лишь если функция правдоподобия не зависит от параметра λ .

Иногда полезно иметь выражение для информации в иной форме. Для этого продифференцируем по λ выражение

$$E \left(\frac{f'(\mathbf{x}; \lambda)}{f(\mathbf{x}; \lambda)} \right) = \int_{-\infty}^{\infty} \frac{f'(x; \lambda)}{f(x; \lambda)} f(x; \lambda) d\lambda = 0$$

и получим

$$0 = \int_{-\infty}^{\infty} \left\{ \frac{f'^2}{f} + f \left(\frac{f'}{f} \right)' \right\} dx = \int_{-\infty}^{\infty} \left\{ \left(\frac{f'}{f} \right)^2 + \left(\frac{f'}{f} \right)' \right\} f dx =$$

$$= E \left\{ \left(\frac{f'}{f} \right)^2 \right\} + E \left\{ \left(\frac{f'}{f} \right)' \right\}.$$

Следовательно, информацию можно записать в следующем виде:

$$I(\lambda) = NE \left\{ \left(\frac{f'(\mathbf{x}; \lambda)}{f(\mathbf{x}; \lambda)} \right)^2 \right\} = -NE \left\{ \left(\frac{f'(\mathbf{x}; \lambda)}{f(\mathbf{x}; \lambda)} \right)' \right\}$$

или

$$I(\lambda) = E(l'^2) = -E(l''), \quad (3.9)$$

Теперь можно переписать (3.8)

$$\{1 + B'(\lambda)\}^2 \leq \sigma^2(\mathbf{S}) I(\lambda)$$

или

$$\sigma^2(\mathbf{S}) \geq \frac{\{1 + B'(\lambda)\}^2}{I(\lambda)}. \quad (3.10)$$

Это выражение представляет собой *неравенство информации*. Оно устанавливает отношение между смещением, дисперсией оценки и информацией, содержащейся в выборке. Заметим, что при выводе неравенства оценка была взята произвольной. Следовательно, правая часть неравенства (3.10) является нижней границей дисперсии данной оценки. Она называется *минимальной дисперсией*, или границей Крамера — Рао. В случаях когда смещение не зависит от λ , в частности в случае нулевого смещения, неравенство (3.10) сводится к следующему:

$$\sigma^2(\mathbf{S}) \geq 1/I(\lambda). \quad (3.11)$$

Это соотношение показывает, что большая информация содержится в выборке с меньшей дисперсией оценки.

Поставим теперь такой вопрос: при каких условиях достигается минимум дисперсии? Более точно — когда в выражении (3.10) имеет место равенство? В выражении (3.6) равенство имеет место, когда $(a\mathbf{x} + \mathbf{y})$ равно нулю, так как только при этом условии $E\{(a\mathbf{x} + \mathbf{y})^2\} = 0$ для любых a , \mathbf{x} и \mathbf{y} . Применяя это условие к (3.8), получаем

$$l' + a(\mathbf{S} - E(\mathbf{S})) = 0,$$

или

$$l' = A(\lambda)(\mathbf{S} - E(\mathbf{S})). \quad (3.12)$$

Здесь A — любая величина, не зависящая от выборки $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}$, хотя, возможно, являющаяся функцией от λ . Интегрированием находим

$$l = \int l' d\lambda = B(\lambda)\mathbf{S} + C(\lambda) + D, \quad (3.13)$$

и окончательно

$$L = d \exp\{B(\lambda)\mathbf{S} + G\}. \quad (3.14)$$

Величины d и D не зависят от λ .

Таким образом, мы видим, что оценки, к которым приводят функции правдоподобия специального вида (3.14), имеют минимальную дисперсию, определяемую выражением (3.10). Они называются *наилучшими оценками*.

В случае несмещенной наилучшей оценки из (3.11) имеем

$$\sigma^2(\mathbf{S}) = \frac{1}{I(\lambda)} = \frac{1}{E(I'^2)}. \quad (3.15)$$

Подставляя (3.12), имеем

$$\sigma^2(\mathbf{S}) = \frac{1}{(A(\lambda))^2 E\{(\mathbf{S} - \lambda)^2\}} = \frac{1}{(A(\lambda))^2 \sigma^2(\mathbf{S})}$$

или

$$\sigma^2(\mathbf{S}) = \frac{1}{A(\lambda)}. \quad (3.16)$$

Если вместо (3.14) выполняется лишь более слабое условие

$$L = g(\mathbf{S}, \lambda) c(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}), \quad (3.17)$$

оценка \mathbf{S} называется *достаточной* для λ . Можно показать (см. [11]), что если условие (3.17) удовлетворено, то никакая другая оценка не увеличивает информации о λ , уже содержащейся в \mathcal{S} . Отсюда название — достаточная оценка (или статистика).

Пример 7.4. Рассмотрим *распределение Пуассона* (4.1) гл. 5

$$f(k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Функция правдоподобия для выборки $k^{(1)}, k^{(2)}, \dots, k^{(N)}$ равна

$$l = \sum_{j=1}^N \{k^{(j)} \ln \lambda - \ln(k^{(j)}!) - \lambda\},$$

и ее производная по λ

$$\begin{aligned} \frac{dl}{d\lambda} = l' &= \sum_{j=1}^N \left\{ \frac{k^{(j)}}{\lambda} - 1 \right\} = \frac{1}{\lambda} \sum_{j=1}^N \{k^{(j)} - \lambda\}, \\ l' &= \frac{N}{\lambda} (\bar{k} - \lambda). \end{aligned} \quad (3.18)$$

Сравнение полученного результата с (3.12) и (3.16) показывает, что среднее арифметическое значение \bar{k} является несмещенной наилучшей оценкой с дисперсией λ/N .

Пример 7.5. Функция правдоподобия для *биномиального распределения* с параметрами $p = \lambda$, $q = 1 - \lambda$ дается непосредственно выражением (1.3) гл. 5

$$L(k, \lambda) = \binom{n}{k} \lambda^k (1 - \lambda)^{n-k}.$$

Тогда

$$l = \ln L = k \ln \lambda + (n - k) \ln (1 - \lambda) + \ln \binom{n}{k},$$

$$l' = \frac{k}{\lambda} - \frac{n - k}{1 - \lambda} = \frac{n}{\lambda(1 - \lambda)} \left(\frac{k}{n} - \lambda \right).$$

Сравнивая этот результат с (3.12) и (3.16), видим, что среднее арифметическое k/n является наилучшей оценкой с дисперсией $\lambda(1 - \lambda)/n$.

Пример 7.6. Вернемся к задаче примера 7.2, т. е. к повторным измерениям одной и той же величины с различной точностью, или, говоря иначе, к выборке из нормальных распределений с равными средними λ и различными, но известными дисперсиями σ_j . Из (2.7) имеем

$$\frac{dl}{d\lambda} = l' = \sum_{j=1}^N \frac{x^{(j)} - \lambda}{\sigma_j^2}.$$

Перепишем это в виде

$$l' = \sum \frac{x^{(j)}}{\sigma_j^2} - \sum \frac{\lambda}{\sigma_j^2} =$$

$$= \sum_{j=1}^N \frac{1}{\sigma_j^2} \left\{ \frac{\sum x^{(j)}}{\sum \frac{1}{\sigma_j^2}} - \lambda \right\}.$$

В примере 7.2 мы установили, что

$$\mathbf{s} = \tilde{\lambda} = \frac{\sum \frac{x^{(j)}}{\sigma_j^2}}{\sum \frac{1}{\sigma_j^2}} \quad (3.19)$$

является несмещенной оценкой λ . Сравнение с (3.12) показывает, что она также наилучшая. Из (3.16) можно выразить соответствующую дисперсию

$$\sigma^2(\tilde{\lambda}) = \left(\sum_{j=1}^N \frac{1}{\sigma_j^2} \right)^{-1}. \quad (3.20)$$

Выражение (3.20) называют иногда *законом комбинации ошибок*. Этот закон можно также получить, применяя закон распространения ошибок (5.7) гл. 4 к (3.19). Идентифицируя $\sigma(\tilde{\lambda})$ с ошиб-

кой оценки $\tilde{\lambda}$ и σ_j — с ошибкой j -го измерения, можно записать его в обычной форме

$$\Delta \tilde{\lambda} = \left(\frac{1}{(\Delta x_1)^2} + \frac{1}{(\Delta x_2)^2} + \dots + \frac{1}{(\Delta x_n)^2} \right)^{-\frac{1}{2}} \quad (3.21)$$

Если все измерения проводятся с одинаковой точностью $\sigma = \sigma_j$, (3.19) и (3.20) превращаются соответственно в $\tilde{\lambda} = \bar{x}$, $\sigma^2(\tilde{\lambda}) = \sigma^2/n$, что уже было найдено в разд. 6.2.

7.4. АСИМПТОТИЧЕСКИЕ СВОЙСТВА ФУНКЦИИ ПРАВДОПОДОБИЯ И ОЦЕНОК МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Рассмотрим теперь несколько важных свойств функции правдоподобия и оценок максимального правдоподобия для очень больших выборок, т. е. для $N \rightarrow \infty$. Оценка $\mathbf{S} = \tilde{\lambda}$ была определена как решение уравнения правдоподобия

$$l'(\lambda) = \sum_{j=1}^N \left(\frac{f'(\mathbf{x}^{(j)}; \lambda)}{f(\mathbf{x}^{(j)}; \lambda)} \right)_{\tilde{\lambda}} = 0. \quad (4.1)$$

Предположив, что производную $l'(\lambda)$ можно дифференцировать далее, разложим ее в ряд Тейлора в окрестности точки $\lambda = \tilde{\lambda}$

$$l'(\lambda) = l'(\tilde{\lambda}) + (\lambda - \tilde{\lambda}) l''(\tilde{\lambda}) + \dots \quad (4.2)$$

Первое слагаемое в правой части (4.2) равно нулю вследствие (4.1). Во втором слагаемом можно явно выразить $l''(\tilde{\lambda})$:

$$l''(\tilde{\lambda}) = \sum_{j=1}^N \left(\frac{f'(\mathbf{x}^{(j)}; \lambda)}{f(\mathbf{x}^{(j)}; \lambda)} \right)'_{\tilde{\lambda}}.$$

Это выражение имеет вид выборочного среднего. При достаточно больших N его можно заменить соответствующим математическим ожиданием (разд. 6.2)

$$l''(\tilde{\lambda}) = NE \left\{ \left(\frac{f'(\mathbf{x}; \lambda)}{f(\mathbf{x}; \lambda)} \right)'_{\tilde{\lambda}} \right\}. \quad (4.3)$$

Пользуясь (3.9), запишем

$$l''(\tilde{\lambda}) = E(l''(\tilde{\lambda})) = -E(l'^2(\tilde{\lambda})) = -I(\tilde{\lambda}) = -1/b^2. \quad (4.4)$$

Таким образом мы выразили производную $l''(\tilde{\lambda})$, являющуюся функцией от частной выборки $x^{(1)}, x^{(2)}, \dots, x^{(N)}$, через величину

ну — $1/b^2$, которая зависит лишь от плотности f и оценки $\tilde{\lambda}$. Пренебрегая членами высших порядков, (4.2) можно записать в виде

$$l'(\lambda) = -\frac{1}{b^2}(\lambda - \tilde{\lambda}). \quad (4.5)$$

Интегрируя, получим

$$l(\lambda) = -\frac{1}{2b^2}(\lambda - \tilde{\lambda})^2 + c$$

или

$$L(\lambda) = k \exp\{-(\lambda - \tilde{\lambda})^2/2b^2\}. \quad (4.6)$$

Здесь c и k — постоянные. Функция правдоподобия $L(\lambda)$ имеет вид нормального распределения* со средним $\tilde{\lambda}$ и дисперсией b^2 .

Сравним теперь (4.5) с (3.12) и (3.16). Так как при оценке параметра мы получили $\mathbf{S} = \tilde{\lambda}$ и $E(\mathbf{S}) = \lambda$, следовательно, $\tilde{\lambda}$ является несмещенной наилучшей оценкой с дисперсией

$$\sigma^2(\tilde{\lambda}) = b^2 = \frac{1}{I(\tilde{\lambda})} = \frac{1}{E(l'^2(\tilde{\lambda}))} = -\frac{1}{E(l''(\tilde{\lambda}))}. \quad (4.7)$$

Так как оценка $\tilde{\lambda}$ обладает этим свойством только в пределе при $N \rightarrow \infty$, мы будем называть ее *асимптотически несмещенной* и состоятельной (разд. 6.1). Аналогично функция правдоподобия называется *асимптотически нормальной*.

В разд. 7.2 мы установили, что функция правдоподобия $L(\lambda)$ является мерой вероятности того, что истинное значение параметра равно λ . Результат оценивания часто записывается в упрощенной форме:

$$\lambda = \tilde{\lambda} \pm \sigma(\tilde{\lambda}) = \tilde{\lambda} \pm \Delta\lambda.$$

Так как функция правдоподобия является асимптотически нормальной, это можно интерпретировать следующим образом (разд. 5.9): вероятность того, что истинное значение λ_0 лежит в интервале

$$\tilde{\lambda} - \Delta\tilde{\lambda} < \lambda_0 < \tilde{\lambda} + \Delta\tilde{\lambda},$$

равна 68,2%.

На практике приведенные выше соотношения используются в случае больших, но конечных выборок. К сожалению, не существует общих правил для определения объема выборки, при котором эти соотношения уже можно применять. Очевидно, что для конечных N (4.3) является приближением, точность которого зависит не только от N , но и от конкретной плотности вероятности $f(x; \lambda)$.

* Вывод автора не строг. Следовало еще доказать, что $(\lambda - \tilde{\lambda})$ стремится к нулю при $N \rightarrow \infty$, так что в (4.2) действительно можно пренебречь отброшенными членами. — Прим. ред.

7.5. РЕШЕНИЕ УРАВНЕНИЯ ПРАВДОПОДОБИЯ С ПОМОЩЬЮ ИТЕРАЦИЙ

Часто уравнение правдоподобия не может быть решено аналитически. В простых случаях для получения максимума функций правдоподобия L или l можно пользоваться графическими методами. Можно также применять различные численные методы получения решения с помощью последовательных приближений. Для описания одного из таких методов мы будем использовать некоторые соотношения предыдущего раздела. Начнем с уравнения правдоподобия

$$l' = \sum_{j=1}^N \frac{f'(x^{(j)}; \lambda)}{f(x^{(j)}; \lambda)} = 0 \quad (5.1)$$

и предположим, что имеется некоторое первое приближение λ_0 оцениваемого параметра (из предварительного эксперимента или предположения). Обозначив второе приближение через λ_1 , можно записать

$$\lambda_1 = \lambda_0 + \delta\lambda_0. \quad (5.2)$$

Поправка $\delta\lambda_0$ может быть найдена из разложения

$$l'(\lambda_1) = l'(\lambda_0) + \delta\lambda_0 l''(\lambda_0) + \dots \quad (5.3)$$

Пренебрегая членами высших порядков, имеем

$$\delta\lambda_0 = -\frac{l'(\lambda_0)}{l''(\lambda_0)}. \quad (5.4)$$

Из предположения (5.2) левая часть в (5.3) равна нулю, т. е. второе приближение является решением уравнения правдоподобия. Теперь, так же как в разд. 7.4, можно заменить

$$-l''(\lambda_0) \approx -E(l''(\lambda_0)) = E(l'^2(\lambda_0)) = I(\lambda_0). \quad (5.5)$$

Хотя для конечных выборок равенство (5.5) неточно, его можно принять, так как λ_1 всего лишь приближение. Процесс можно повторить итерациями

$$\begin{aligned} \lambda_1 &= \lambda_0 + \delta\lambda_0 = \lambda_0 + \frac{l'(\lambda_0)}{I(\lambda_0)}, \\ \lambda_2 &= \lambda_1 + \delta\lambda_1 = \lambda_1 + \frac{l'(\lambda_1)}{I(\lambda_1)}, \\ &\vdots \end{aligned}$$

Приближение λ_i можно считать достаточно точным, когда $|\delta\lambda_{i-1}|$ будет меньше некоторого заданного числа.

7.6. СОВМЕСТНАЯ ОЦЕНКА НЕСКОЛЬКИХ ПАРАМЕТРОВ

Мы уже сталкивались [равенство (2.5)] с системой уравнений, позволяющей найти совместную оценку p параметров $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$. Мы покажем, что сложности вызывает не нахождение самих параметров, а, скорее, оценивание их ошибок. При этом надо обязательно учитывать ковариации между оценками различных параметров.

Обобщим рассуждения разд. 7.4 для рассмотрения асимптотических свойств функции правдоподобия в случае нескольких параметров. Логарифмическую функцию правдоподобия

$$l(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}; \lambda) = \sum_{i=1}^N \ln f(\mathbf{x}^{(i)}; \lambda) \quad (6.1)$$

можно разложить в ряд Тейлора в окрестности точки

$$\tilde{\lambda} = (\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_p), \quad (6.2)$$

$$l(\lambda) = l(\tilde{\lambda}) + \sum_{k=1}^p \left(\frac{\partial l}{\partial \lambda_k} \right)_{\tilde{\lambda}} (\lambda_k - \tilde{\lambda}_k) + \\ + \frac{1}{2} \sum_{l=1}^p \sum_{m=1}^p \left(\frac{\partial^2 l}{\partial \lambda_l \partial \lambda_m} \right)_{\tilde{\lambda}} (\lambda_l - \tilde{\lambda}_l) (\lambda_m - \tilde{\lambda}_m) + \dots, \quad (6.3)$$

так как из определения $\tilde{\lambda}$

$$\left(\frac{\partial l}{\partial \lambda_k} \right)_{\tilde{\lambda}} = 0 \quad k = 1, 2, \dots, p,$$

то для всех k разложение (6.3) сводится к

$$l(\lambda) = l(\tilde{\lambda}) + \frac{1}{2} (\lambda - \tilde{\lambda})^T A (\lambda - \tilde{\lambda}) + \dots, \quad (6.4)$$

где

$$A = \begin{bmatrix} \frac{\partial^2 l}{\partial \lambda_1^2} & \frac{\partial^2 l}{\partial \lambda_1 \partial \lambda_2} & \dots & \frac{\partial^2 l}{\partial \lambda_1 \partial \lambda_p} \\ \frac{\partial^2 l}{\partial \lambda_1 \partial \lambda_2} & \frac{\partial^2 l}{\partial \lambda_2^2} & \dots & \frac{\partial^2 l}{\partial \lambda_2 \partial \lambda_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \lambda_1 \partial \lambda_p} & \frac{\partial^2 l}{\partial \lambda_2 \partial \lambda_p} & \dots & \frac{\partial^2 l}{\partial \lambda_p^2} \end{bmatrix}_{\lambda = \tilde{\lambda}}. \quad (6.5)$$

В пределе при $N \rightarrow \infty$ элементы матрицы A , которые зависят от конкретной выборки, можно заменить соответствующими матема-

тическими ожиданиями

$$B = E(A) = \left[\begin{array}{cccc} E\left(\frac{\partial^2 l}{\partial \lambda_1^2}\right) & E\left(\frac{\partial^2 l}{\partial \lambda_1 \partial \lambda_2}\right) & \cdots & E\left(\frac{\partial^2 l}{\partial \lambda_1 \partial \lambda_p}\right) \\ E\left(\frac{\partial^2 l}{\partial \lambda_1 \partial \lambda_2}\right) & E\left(\frac{\partial^2 l}{\partial \lambda_2^2}\right) & \cdots & E\left(\frac{\partial^2 l}{\partial \lambda_2 \partial \lambda_p}\right) \\ \vdots & \vdots & \ddots & \vdots \\ E\left(\frac{\partial^2 l}{\partial \lambda_1 \partial \lambda_p}\right) & E\left(\frac{\partial^2 l}{\partial \lambda_2 \partial \lambda_p}\right) & \cdots & E\left(\frac{\partial^2 l}{\partial \lambda_p^2}\right) \end{array} \right]_{\lambda=\tilde{\lambda}}. \quad (6.6)$$

Пренебрегая членами высших порядков, можно записать выражение для нелогарифмической функции правдоподобия

$$L = k \exp\left\{\frac{1}{2}(\lambda - \tilde{\lambda})^T B(\lambda - \tilde{\lambda})\right\}. \quad (6.7)$$

Сравнение с (10.1) гл. 5 показывает, что (6.7) является p -мерным нормальным распределением со средним $\tilde{\lambda}$ и ковариационной матрицей

$$C = -B^{-1}. \quad (6.8)$$

Диагональные элементы матрицы (6.8) являются дисперсиями оценок максимального правдоподобия $\tilde{\lambda}_1, \tilde{\lambda}_2, \dots, \tilde{\lambda}_p$. Недиагональные элементы являются ковариациями между всевозможными парами оценок

$$\sigma^2(\tilde{\lambda}_i) = c_{ii}, \quad (6.9)$$

$$\text{cov}(\tilde{\lambda}_j, \tilde{\lambda}_k) = c_{jk}. \quad (6.10)$$

Можно определить коэффициент корреляции между оценками

$$\rho(\tilde{\lambda}_j, \tilde{\lambda}_k) = \frac{\text{cov}(\tilde{\lambda}_j, \tilde{\lambda}_k)}{\sigma(\tilde{\lambda}_j) \sigma(\tilde{\lambda}_k)}. \quad (6.11)$$

Как и в случае одного параметра, квадратные корни из дисперсий называются ошибками или стандартными отклонениями оценок

$$\Delta \tilde{\lambda}_i = \sigma(\tilde{\lambda}_i) = \sqrt{c_{ii}}. \quad (6.12)$$

В разд. 7.4 было показано, что с помощью оценки максимального правдоподобия и ее ошибки может быть найдена область, содержащая истинное значение с вероятностью 68,2%. Так как в многопараметрическом случае функция правдоподобия является многомерным гауссовским распределением, эта область зависит не только от ошибок, но и от всей ковариационной матрицы. В частном случае двух параметров такой областью является ковариационный эллипс, рассмотренный в разд. 5.10*.

* Согласно формуле (10.7) гл. 5, ковариационный эллипс содержит истинные значения параметров с вероятностью $1 - e^{-1/2} \approx 0,39$. — Прим. ред.

Пример 7.7. Из выборки объема N требуется найти среднее значение λ_1 и дисперсию λ_2 нормального распределения. Эта задача возникает, например, при изучении свободного пробега α -частиц в веществе. Вследствие статистической природы энергии, выделяющейся при большом числе отдельных независимых соударений, величина пробега нормально распределена вокруг среднего значения. Измеряя длину пробега $\mathbf{x}^{(j)}$ для N отдельных частиц, можно оценить среднее λ_1 и «постоянную разброса» λ_2 . Получаем функции правдоподобия

$$L = \prod_{j=1}^N \frac{1}{\sqrt{2\pi} \lambda_2} \exp\left(-\frac{(\mathbf{x}^{(j)} - \lambda_1)^2}{2\lambda_2^2}\right)$$

и

$$l = -\frac{1}{2} \sum_{j=1}^N \frac{(\mathbf{x}^{(j)} - \lambda_1)^2}{\lambda_2^2} - N \ln \lambda_2 - \text{const.}$$

Система уравнений правдоподобия превращается в

$$\begin{aligned} \frac{\partial l}{\partial \lambda_1} &= \sum_{j=1}^N \frac{\mathbf{x}^{(j)} - \lambda_1}{\lambda_2^2} = 0, \\ \frac{\partial l}{\partial \lambda_2} &= \frac{1}{\lambda_2^3} \sum_{j=1}^N (\mathbf{x}^{(j)} - \lambda_1)^2 - \frac{N}{\lambda_2} = 0. \end{aligned}$$

Решение системы имеет вид

$$\begin{aligned} \tilde{\lambda}_1 &= \frac{1}{N} \sum_{j=1}^N \mathbf{x}^{(j)}, \\ \tilde{\lambda}_2 &= \sqrt{\frac{\sum_{j=1}^N (\mathbf{x}^{(j)} - \tilde{\lambda}_1)^2}{N}}. \end{aligned}$$

Метод максимального правдоподобия дает в качестве оценки математического ожидания среднее арифметическое отдельных измерений; для дисперсии он дает величину \mathbf{s}'^2 [равенство (2.5) гл. 6], которая несколько смещена в отличие от несмещенной оценки \mathbf{s} [выражение (2.7) гл. 6].

Найдем теперь матрицу B . Вторые производные равны

$$\frac{\partial^2 l}{\partial \lambda_1^2} = -\frac{N}{\lambda_2^2},$$

$$\frac{\partial^2 l}{\partial \lambda_1 \partial \lambda_2} = -\frac{2 \sum (\mathbf{x}^{(j)} - \lambda_1)}{\lambda_2^3},$$

$$\frac{\partial^2 l}{\partial \lambda_2^2} = -\frac{3 \sum (\mathbf{x}^{(j)} - \lambda_1)^2}{\lambda_2^4} + \frac{N}{\lambda_2^2}.$$

По аналогии с (6.6) заменим λ_1, λ_2 на $\tilde{\lambda}_1, \tilde{\lambda}_2$ и найдем

$$-B = \begin{pmatrix} N/\tilde{\lambda}_2^2 & 0 \\ 0 & 2N/\tilde{\lambda}_2^2 \end{pmatrix},$$

откуда ковариационная матрица равна

$$C = -B^{-1} = \begin{pmatrix} \tilde{\lambda}_2^2/N & 0 \\ 0 & \tilde{\lambda}_2^2/2N \end{pmatrix}.$$

Диагональные элементы можно принять в качестве ошибок соответствующих параметров, т. е.

$$\Delta \tilde{\lambda}_1 = \tilde{\lambda}_2/\sqrt{N},$$

что в данном случае совпадает с выражением (2.4) гл. 6 и

$$\Delta \tilde{\lambda}_2 = \tilde{\lambda}_2/\sqrt{2N}.$$

Корреляции между λ_1 и λ_2 нет.

Пример 7.8. Рассмотрим популяцию, подчиняющуюся двумерному нормальному распределению с плотностью вероятности (разд. 5.10)

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times$$

$$\times \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(x_1-a_1)^2}{\sigma_1^2} + \frac{(x_2-a_2)^2}{\sigma_2^2} - 2\rho\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2}\right\}\right].$$

Решив систему уравнений правдоподобия относительно пяти параметров $a_1, a_2, \sigma_1^2, \sigma_2^2, \rho$, находим оценки максимального правдоподобия для этих параметров:

$$\bar{\mathbf{x}}_1 = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_1^{(j)}, \quad \bar{\mathbf{x}}_2 = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_2^{(j)},$$

$$\mathbf{s}'_1{}^2 = \frac{1}{N} \sum_{j=1}^N (\mathbf{x}_1^{(j)} - \bar{\mathbf{x}}_1)^2, \quad \mathbf{s}'_2{}^2 = \frac{1}{N} \sum_{j=1}^N (\mathbf{x}_2^{(j)} - \bar{\mathbf{x}}_2)^2,$$

$$\mathbf{r} = \frac{\sum_{j=1}^N (\mathbf{x}_1^{(j)} - \bar{\mathbf{x}}_1)(\mathbf{x}_2^{(j)} - \bar{\mathbf{x}}_2)}{N\mathbf{s}'_1\mathbf{s}'_2}. \quad (6.13)$$

Как и в примере 7.7, оценки s_1^2 и s_2^2 дисперсий являются смещенными. Это справедливо также для выражения (6.13), *выборочного коэффициента корреляции*. Подобно всем оценкам максимального правдоподобия, оценка r состоятельна, т. е. хорошо оценивает ρ для очень больших выборок. При $N \rightarrow \infty$ распределение r стремится к гауссовскому распределению со средним ρ и дисперсией

$$\sigma^2(r) = (1 - \rho^2)^2/N. \quad (6.14)$$

Для конечных выборок распределение оценки r асимметрично. Следовательно, при применении (6.13) важно убедиться в том, что выборка достаточно велика. Как правило, рекомендуется брать N равным по меньшей мере 500.

7.7. ОДНОЗНАЧНОСТЬ МЕТОДА; ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ

Обычно любой процесс или эксперимент можно описать одной или несколькими величинами (параметрами), такими, как время полураспада, среднее время жизни или постоянная распада; скорость, импульс или энергия; объем или масса и т. д., которые связаны друг с другом с помощью известных величин или констант. Очевидно, результат оценки не должен зависеть от выбора того или иного вида параметра (например, $\alpha(\lambda)$ или λ).

Для λ имеем следующее уравнение правдоподобия:

$$\frac{\partial l}{\partial \lambda} = 0. \quad (7.1)$$

Взяв вместо λ другой однозначно связанный с ним параметр $\alpha = \alpha(\lambda)$, мы получим

$$\frac{\partial l}{\partial \alpha} = 0. \quad (7.2)$$

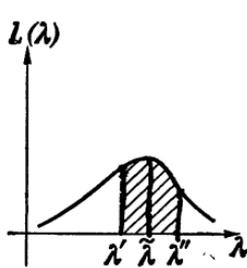
Но так как

$$\frac{\partial l}{\partial \lambda} = \frac{\partial l}{\partial \alpha} \frac{\partial \alpha}{\partial \lambda},$$

то вследствие (7.1) равенство (7.2) заведомо выполняется. Следовательно, независимо от параметризации метод максимального правдоподобия определяет *то же самое решение*.

Менее удовлетворительная ситуация возникает при рассмотрении ошибок. Введем новую меру неопределенности оценки, называемую *доверительным интервалом*. Мы хотим знать вероятность

того, что истинное значение параметра λ лежит между двумя фиксированными значениями λ' и λ'' (рис. 7.2)*:



$$P(\lambda' < \lambda < \lambda'') = \frac{\int_{\lambda'}^{\lambda''} L(\lambda) d\lambda}{\int_{-\infty}^{\infty} L(\lambda) d\lambda}, \quad (7.3)$$

Рис. 7.2. Доверительный интервал.

Преимущество этой формулы в том, что в ней используется некоторая информация о форме функции правдоподобия. Это особенно полезно, когда функция правдоподобия не совпадает с гауссовской кривой. Часто задаются каким-то значением вероятности (например, $P = 95\%$) и выбирают пределы так, чтобы (7.3) выполнялось.

К сожалению, вероятность (7.3) зависит от выбора λ . Взяв вместо λ параметр $\alpha(\lambda)$, получим (разд. 4.4)**

$$P(\lambda' < \lambda < \lambda'') = \frac{\int_{\lambda'}^{\lambda''} L(\lambda) d\lambda}{\int_{-\infty}^{\infty} L(\lambda) d\lambda} = \frac{\int_{\alpha'}^{\alpha''} L \left| \frac{d\lambda}{d\alpha} \right| d\alpha}{\int_{-\infty}^{\infty} L d\lambda} \neq \frac{\int_{\alpha'}^{\alpha''} L d\alpha}{\int_{-\infty}^{\infty} L d\alpha}.$$

7.8. ФУНКЦИЯ БАРТЛЕТТА

Чтобы избежать некоторых из рассмотренных выше трудностей, Бартлетт [1] ввел следующую функцию:

$$S(\lambda) = \frac{1}{C} \frac{\partial l}{\partial \lambda}, \quad (8.1)$$

где

$$C^2 = - \int_{\lambda_{\min}}^{\lambda_{\max}} \frac{\partial^2 l}{\partial \lambda^2} L(\lambda) d\lambda. \quad (8.2)$$

Интегрирование производится по всей области изменения λ или, другими словами, по области, для которой $L(\lambda) \neq 0$ ***. Функция S

* Формула (7.3), строго говоря, справедлива лишь тогда, когда параметр λ можно считать случайной величиной с равномерным априорным распределением по некоторой области. Для очень больших выборок форма априорного распределения становится несущественной. — Прим. ред.

** Это объясняется тем, что если параметр λ имел равномерное априорное распределение, то параметр $\alpha(\lambda)$ будет распределен неравномерно. — Прим. ред.

*** Предполагается, что $L(\lambda)$ всюду имеет вторую производную и обращается в нуль вне интервала $\lambda_{\min} < \lambda < \lambda_{\max}$. — Прим. ред.

имеет нулевое среднее и единичную дисперсию независимо от способа параметризации:

$$\begin{aligned} E(S) &= \frac{1}{C} \int_{\lambda_{\min}}^{\lambda_{\max}} \frac{\partial L}{\partial \lambda} L d\lambda = \frac{1}{C} \int_{\lambda_{\min}}^{\lambda_{\max}} \frac{\partial L}{\partial \lambda} d\lambda = \frac{1}{C} \int_{\lambda_{\min}}^{\lambda_{\max}} dL = \\ &= \frac{1}{C} \{L(\lambda_{\max}) - L(\lambda_{\min})\}. \end{aligned}$$

Так как в граничных точках интервала интегрирования $L(\lambda) = 0$, получаем

$$E(S) = 0. \quad (8.3)$$

Выражение для дисперсии принимает вид

$$\begin{aligned} \sigma^2(S) &= E\{S^2 - (E(S))^2\} = E\{S^2\} = \frac{1}{C^2} \int_{\lambda_{\min}}^{\lambda_{\max}} \left(\frac{\partial L}{\partial \lambda}\right)^2 L d\lambda = \\ &= \frac{1}{C^2} \int \frac{1}{L^2} \left(\frac{\partial L}{\partial \lambda}\right)^2 L d\lambda = \frac{\int \left(\frac{\partial L}{\partial \lambda}\right)^2 \frac{1}{L} d\lambda}{-\int \frac{\partial}{\partial \lambda} \left(\frac{1}{L} \frac{\partial L}{\partial \lambda}\right) L d\lambda} = \\ &= \frac{\int \left(\frac{\partial L}{\partial \lambda}\right)^2 \frac{1}{L} d\lambda}{-\int \frac{\partial^2 L}{\partial \lambda^2} d\lambda + \int \left(\frac{\partial L}{\partial \lambda}\right)^2 \frac{1}{L} d\lambda}. \end{aligned}$$

Вне области возможных значений λ , т. е. для $\lambda < \lambda_{\min}$ и $\lambda > \lambda_{\max}$, функция правдоподобия $L(\lambda)$ тождественно равна нулю. Следовательно, ее производная $\partial L/\partial \lambda$ также равна нулю. Отсюда

$$-\int_{\lambda_{\min}}^{\lambda_{\max}} \frac{\partial^2 L}{\partial \lambda^2} d\lambda = \left(\frac{\partial L}{\partial \lambda}\right)_{\lambda_{\min}} - \left(\frac{\partial L}{\partial \lambda}\right)_{\lambda_{\max}} = 0,$$

т. е. просто получаем

$$\sigma^2(S) = 1. \quad (8.4)$$

Мы получили среднее значение и дисперсию функции $S(\lambda)$, но еще не выяснили, как она выглядит в конкретных случаях. Предположим, что функция правдоподобия имеет форму гауссовской кривой по λ со средним $\bar{\lambda}$ и дисперсией ϵ^2 . Используя (4.5) и (4.6), легко получить

$$C^2 = \frac{1}{\epsilon^2}, \quad S(\lambda) = \frac{\lambda - \bar{\lambda}}{\epsilon}.$$

В этом случае S -функция Бартлетта представляет собой прямую линию, проходящую через нуль при $\lambda = \tilde{\lambda}$ и принимающую значения $\pm 1, \pm 2, \dots$, когда λ отклоняется от $\tilde{\lambda}$ на $1, 2, \dots$ стандартных отклонения. Найдя вид S -функции, можно, следовательно, разрешить уравнения

$$S(\lambda_-) = +1, \quad S(\lambda_+) = -1$$

и определить

$$\Delta_- = \tilde{\lambda} - \lambda_-, \quad \Delta_+ = \lambda_+ - \tilde{\lambda}.$$

Если Δ_- и Δ_+ примерно равны, будем считать ошибку симметричной. В противном случае результат анализа запишем в виде

$$\lambda = \tilde{\lambda} + \frac{\Delta_+}{-\Delta_-}.$$

Пример 7.9. Для иллюстрации использования S -функции Бартлетта рассмотрим измерение величины τ — среднего времени жизни Λ -гиперона. Это — тяжелая элементарная частица, распадающаяся согласно закону радиоактивного распада. Вероятность того, что частица распадается в некоторый промежуток времени t после своего образования, равна

$$P(t) = \int_0^t \frac{1}{\tau} \exp(-x/\tau) dx.$$

Отсюда плотность вероятности, пропорциональная вероятности распада в интервале времени от t до $t + dt$, равна

$$f(t) = \frac{1}{\tau} \exp(-t/\tau).$$

Если частица движется в некотором направлении со скоростью v , а скорость света обозначить через c , то из-за релятивистского замедления времени мы получим

$$f(t) = \frac{V\sqrt{1-(v/c)^2}}{\tau} \exp\left[-\frac{1}{\tau} t V\sqrt{1-(v/c)^2}\right].$$

Λ -частицы могут быть получены при бомбардировке водородной пузырьковой камеры пучком K^- -мезонов. Момент образования Λ -частицы можно заметить, поскольку путь K^- -мезона в точке образования внезапно обрывается. Λ -частица, будучи электрически

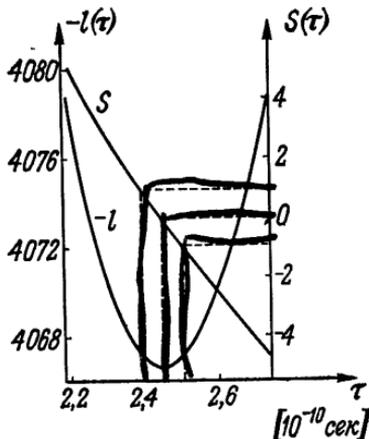


Рис. 7.3. Логарифмическая функция правдоподобия и S -функция Бартлетта для примера 7.9.

нейтральной, не производит ионизации, но частицы, образующиеся в результате ее распада, дадут характерный рисунок. Их импульсы можно измерить, и в сумме они дадут импульс Λ -частицы и скорость ($v^{(i)}$). Затем из расстояния между точками образования и распада можно определить время распада $t^{(i)}$. Каждый индивидуальный распад имеет плотность вероятности

$$f_i = \frac{\sqrt{1 - (v^{(i)}/c)^2}}{\tau} \exp \left[-\frac{1}{\tau} t^{(i)} \sqrt{1 - (v^{(i)}/c)^2} \right],$$

которая должна быть потом изменена, чтобы учесть некоторые экспериментальные поправки (на незамеченные распады и т. д.).

Наконец, перемножая вероятности всех наблюдаемых исходов, мы получаем функции правдоподобия L или $l = \ln L$ и S -функцию Бартлетта. На рис. 7.3 показаны последние две функции, полученные в результате 2213 распадов в эксперименте Энгелманна, Филтута и др. [5]. Считывая с графика максимум функции правдоподобия (здесь построена кривая $-l$) и значения τ для $S = \pm 1$, получаем в результате

$$\tau_{\Lambda} = \left(2,452 \begin{array}{c} +0,056 \\ -0,054 \end{array} \right) \times 10^{-10} \text{ с.}$$

Задача статистического анализа часто состоит в определении параметров, которые нельзя считать совершенно неизвестными. Вполне вероятно заранее ожидать определенных значений параметров на основании предварительных результатов, каких-то моделей или теории. Целью выборки (или измерений) является проверка этой *гипотезы*. Например, при контроле качества продукции мы предполагаем, что некоторые критические величины нормально распределены вокруг их номинального значения. Для проверки этой гипотезы производится выборка продукции. Мы хотим проиллюстрировать применение статистических критериев на этом примере. Рассмотрим для простоты гипотезу о том, что переменная x имеет стандартное гауссовское распределение. Для проверки мы делаем выборку из 10 элементов.

Пусть анализ выборки дает арифметическое среднее $\bar{x} = 0,154$. Если мы предполагаем, что гипотеза верна, случайная величина \bar{x} имеет нормальное распределение с нулевым средним и дисперсией $1/n$. Теперь зададим вопрос: какова вероятность наблюдения величины $|\bar{x}| \geq 0,154$ с таким распределением? Согласно (9.5) гл. 5, из табл. E. 2 находим

$$P(|\bar{x}| \geq 0,154) = 2 \{1 - \phi_0(0,154 \sqrt{10})\} = 0,62.$$

Поэтому, даже если наша гипотеза верна, вероятность того, что выборка, состоящая из 10 элементов, дает среднее, отличающееся на 0,154 или более от среднего популяции, равна 62%.

Теперь мы оказываемся в трудном положении, пытаюсь ответить на простой вопрос: эта гипотеза верна или ложна? Нельзя прямо дать четкий ответ. Этого следовало ожидать, потому что все результаты в математической статистике являются вероятностными результатами. Мы можем несколько искусственно выйти из этого положения, вводя понятие *уровня значимости*. Этот выход заключается в следующем. Перед анализом выборки мы фиксируем некоторую (обычно малую) вероятность α . Затем мы спрашиваем: если предполагать гипотезу верной, то вероятность получения выборки с отмеченными свойствами больше или меньше α ? В нашем случае вопрос был бы таким: верно ли, что $P(|\bar{x}| \geq$

$\geq 0,154) < \alpha$? Если бы эта вероятность была действительно меньше, чем α , то мы должны были бы заключить, что очень неправдоподобно получить такую выборку из популяции с предложенным распределением, и поэтому *отвергнуть* гипотезу. Обратное рассуждение невозможно. Если вероятность, о которой идет речь, превосходит α , нельзя сказать, что «гипотеза верна», а лишь только, что «она совместима с результатом выборки».

Выбор уровня значимости зависит от исследуемой задачи. Например, в задаче контроля качества карандашей можно удовлетвориться 1%-ным уровнем значимости. С другой стороны, если мы хотим определить суммы страховой компании, которые она выплачивает, так чтобы вероятность ее разорения была меньше, чем α , то 0,01% все еще слишком высокий уровень значимости. При анализе научных данных обычно берутся значения 5, 1 и 0,1%. Из табл. Е. 5 мы можем вычислить границы для величины $|\bar{x}|$, соответствующие этим уровням:

$$0,05 = 2 \{1 - \psi_0(1,96)\} = 2 \{1 - \psi_0(0,62 \sqrt{10})\},$$

$$0,01 = 2 \{1 - \psi_0(2,58)\} = 2 \{1 - \psi_0(0,82 \sqrt{10})\},$$

$$0,001 = 2 \{1 - \psi_0(3,29)\} = 2 \{1 - \psi_0(1,04 \sqrt{10})\}.$$

Для того чтобы отвергнуть нашу гипотезу, нужно было бы, чтобы при этих значениях α $|\bar{x}|$ превысило соответственно 0,62, 0,82 или 1,04.

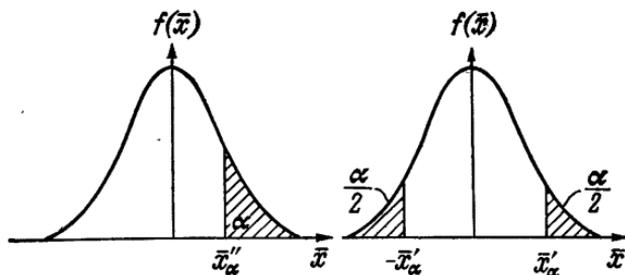


Рис. 8.1. Односторонний и двусторонний критерии.

В некоторых случаях знак рассматриваемой величины (в нашем примере \bar{x}) является существенным. Булочник, который выпекает булку с излишком веса, теряет прибыль; булки с недовесом могут быть причиной потери им лицензии. Поэтому мы могли бы проверить отклонения только в одну сторону и спросить, будет ли выполняться неравенство

$$P(\bar{x} \geq x'_\alpha) < \alpha.$$

Такой критерий называют односторонним критерием в противоположность уже упомянутому двустороннему критерию (рис. 8.1).

В общем случае представляют интерес и некоторые величины, отличные от среднего. Поступим следующим образом. Сначала мы определяем функцию от выборки, подходящую для проверки нашей гипотезы. Она называется *статистикой T , лежащей в основе критерия*. Затем мы фиксируем уровень значимости α и определяем подобласть U внутри области изменения T , такую, что

$$P_H(T \in U) = \alpha.$$

Индекс H означает, что эта вероятность вычислена при условии, что гипотеза H верна. Теперь мы делаем выборку. Пусть, например, на этой выборке статистика принимает некоторое значение T' . Если T' попадает внутрь критической области U , то гипотеза H отвергается.

В следующих разделах мы сначала обсудим несколько специальных критериев, а затем возвратимся к более строгой трактовке теории статистических критериев.

8.1. КРИТЕРИЙ F ДЛЯ РАВЕНСТВА ДИСПЕРСИЙ

По мере развития техники измерений и техники производства часто стало необходимо сравнивать дисперсии совокупностей, у которых равны средние. Так, например, пусть величина измерена двумя различными приборами, не имеющими систематических ошибок; имеют ли наблюдения одну и ту же дисперсию, т. е. являются ли приборы равнокачественными.

Для того чтобы проверить эту гипотезу, мы предполагаем, что совокупности имеют нормальные распределения, а объемы выборок равны N_1 и N_2 соответственно.

Для каждой выборки найдем эмпирическую дисперсию и рассмотрим *отношение дисперсий*

$$Z = \frac{s_1^2}{s_2^2}. \quad (1.1)$$

Если гипотеза о равенстве дисперсий верна, то отношение должно быть близко к единице. Мы знаем из разд. 6.6, что для каждой выборки мы можем построить статистику, которая будет иметь χ^2 -распределение. Имеем

$$\chi_1^2 = \frac{(N_1 - 1) s_1^2}{\sigma_1^2} = \frac{f_1 s_1^2}{\sigma_1^2},$$

$$\chi_2^2 = \frac{(N_2 - 1) s_2^2}{\sigma_2^2} = \frac{f_2 s_2^2}{\sigma_2^2}.$$

Распределения этих двух величин имеют $f_1 = (N_1 - 1)$ и $f_2 = (N_2 - 1)$ степеней свободы соответственно. Если предположить, что наша

гипотеза верна, т. е. $\sigma_1^2 = \sigma_2^2$, то отношение Z будет равно

$$Z = \frac{f_2}{f_1} \frac{X_1^2}{X_2^2}.$$

Плотность вероятности χ^2 -распределения с f степенями свободы равна [равенство (5.10) гл. 6]

$$f(\chi^2) = \frac{1}{\Gamma\left(\frac{1}{2}f\right) 2^{\frac{1}{2}f}} (\chi^2)^{\frac{1}{2}(f-2)} e^{-\frac{1}{2}\chi^2}.$$

Теперь вычислим вероятность

$$V(Q) = P\left(\frac{X_1^2}{X_2^2} < Q\right),$$

т. е. вероятность того, что отношение X_1^2/X_2^2 будет меньше, чем Q :

$$V(Q) = \frac{1}{\Gamma\left(\frac{1}{2}f_1\right) \Gamma\left(\frac{1}{2}f_2\right) 2^{\frac{1}{2}(f_1+f_2)}} \int_{\substack{x>0 \\ y>0 \\ x/y < Q}} x^{f_1-1} e^{-\frac{1}{2}x} y^{f_2-1} e^{-\frac{1}{2}y} dx dy.$$

Этот интеграл сводится к

$$V(Q) = \frac{\Gamma\left(\frac{1}{2}f\right)}{\Gamma\left(\frac{1}{2}f_1\right) \Gamma\left(\frac{1}{2}f_2\right)} \int_0^Q t^{\frac{1}{2}f_1-1} (t+1)^{-\frac{1}{2}f} dt, \quad (1.2)$$

где $f = f_1 + f_2$. Возвращаясь теперь к первоначальному отношению (1.1), получаем функцию распределения случайной величины Z :

$$W(F) = P\left(\frac{\sigma_1^2}{\sigma_2^2} < F\right) = P\left(\frac{X_1^2}{X_2^2} < \frac{f_1 F}{f_2}\right) = V\left(\frac{f_1 F}{f_2}\right).$$

Это распределение называется F -распределением Фишера*. Оно зависит от параметров f_1 и f_2 .

Плотность вероятности F -распределения равна

$$f(F) = \left(\frac{f_1}{f_2}\right)^{\frac{1}{2}f_1} \frac{\Gamma\left(\frac{1}{2}(f_1+f_2)\right)}{\Gamma\left(\frac{1}{2}f_1\right) \Gamma\left(\frac{1}{2}f_2\right)} F^{\frac{1}{2}f_1-1} \left(1 + \frac{f_1}{f_2} F\right)^{-\frac{1}{2}(f_1+f_2)}. \quad (1.3)$$

* Иногда пользуются другими названиями: ν^2 -распределение, ω^2 -распределение, распределение Снедекора.

На рис. 8.2. изображена плотность для конкретных значений f_1 и f_2 . Это распределение похоже на χ^2 -распределение; его плотность отлична от нуля только для положительных значений F и стремится к нулю при $F \rightarrow \infty$. Следовательно, она несимметрична. Можно легко показать, что для $f_2 > 2$ математическое ожидание Z равно

$$E(Z) = f_1 / (f_2 - 2).$$

Теперь мы можем определить границу F'_α , потребовав (рис. 8.2) чтобы

$$P\left(\frac{s_1^2}{s_2^2} > F'_\alpha\right) = \alpha. \quad (1.4)$$

Из этого равенства видно, что граница F'_α равна *квантили* $F_{1-\alpha}$ (разд. 3.3) F -распределения

$$\begin{aligned} P\left(\frac{s_1^2}{s_2^2} < F'_\alpha\right) &= P\left(\frac{s_1^2}{s_2^2} < F_{1-\alpha}\right) = \\ &= 1 - \alpha. \end{aligned} \quad (1.5)$$

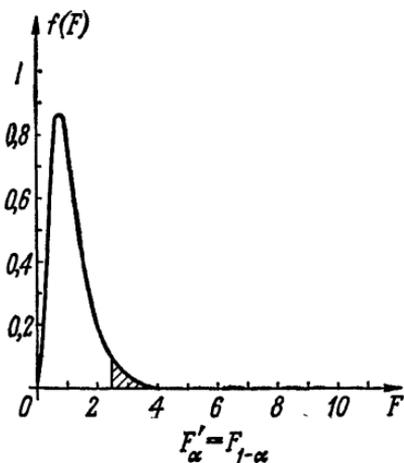


Рис. 8.2. Плотность вероятности F -распределения для $f_1 = 10$ и $f_2 = 20$. Граница $F_{1-\alpha}$ выбрана для $\alpha = 0,05$.

Если эта граница превышена, мы скажем, что $\sigma_1^2 > \sigma_2^2$ с уровнем значимости α . В табл. Е.6 значения $F_{1-\alpha}$ протабулированы для различных пар (f_1, f_2) . Как правило, мы хотим получить двусторонний критерий, т. е. критерий, когда отношение Z попадает между двумя

границами F''_α и F'''_α , определенными условиями

$$P\left(\frac{s_1^2}{s_2^2} > F''_\alpha\right) = \frac{1}{2} \alpha, \quad P\left(\frac{s_1^2}{s_2^2} < F'''_\alpha\right) = \frac{1}{2} \alpha. \quad (1.6)$$

Так как Z есть отношение двух чисел, то условие

$$s_1^2 / s_2^2 < F''_\alpha(f_1, f_2)$$

эквивалентно

$$s_2^2 / s_1^2 > F'''_\alpha(f_2, f_1).$$

Здесь всюду первый аргумент обозначает число степеней свободы числителя, а второй — знаменателя. Поэтому мы можем заменить

условия (1,6) на следующие:

$$P\left(\frac{s_1^2}{s_2^2} > F''_{\alpha}(f_1, f_2)\right) = \frac{1}{2} \alpha,$$

$$P\left(\frac{s_2^2}{s_1^2} > F''_{\alpha}(f_2, f_1)\right) = \frac{1}{2} \alpha. \quad (1.7)$$

Сравнение (1.7) и (1.4) показывает, что

$$F''_{\alpha}(f_1, f_2) = F_{1-\frac{1}{2}\alpha}(f_1, f_2).$$

Поэтому мы можем использовать табл. Е.6 как для одностороннего, так и для двустороннего F -критерия.

Из табл. Е.6 мы находим, что $F_{1-\alpha/2} > 1$ для всех разумных значений α , скажем $\alpha \leq 0,1$. Поэтому нам нужно проверить лишь выполнение неравенства

$$s_L^2/s_S^2 > F_{1-\frac{1}{2}\alpha}(f_L, f_S). \quad (1.8)$$

Здесь индексы L и S обозначают соответственно большее и меньшее значения дисперсий выборок, т. е. $s_L^2 > s_S^2$. Если неравенство (1.8) выполняется, то гипотеза равенства дисперсий отвергается.

Таблица 8.1

**F -критерий для проверки равенства дисперсий;
данные примера 8.1**

Номер измерения	Измерения в микронах, выполненные на	
	1-м приборе	2-м приборе
1	100	97
2	101	102
3	103	103
4	98	96
5	97	100
6	98	101
7	102	100
8	101	
9	99	
10	101	
<hr/>		
Среднее	100	99,8
Число степеней свободы	9	6
s^2	$34/9 = 3,7$	$39/6 = 6,5$
<hr/>		
$F = s_L^2/s_S^2 = 6,5/3,7 = 1,8$		

Пример 8.1. Стандартная длина (объектный микрометр 100 мкм) многократно измеряется на двух различных микроскопах. Эти измерения и вычисления приведены в табл. 8.1. Из табл. Е.6 находим для двустороннего F -критерия границу 10%-го уровня значимости

$$F_{0,1}^{\alpha}(6, 9) = F_{0,95}(6, 9) = 3,37.$$

Следовательно, гипотеза равенства дисперсий не может быть отвергнута.

8.2. КРИТЕРИЙ СТЬЮДЕНТА; СРАВНЕНИЕ СРЕДНИХ

Рассмотрим случайную величину \mathbf{x} , имеющую нормальное распределение. Произведем выборку объема N , пусть ее среднее равно $\bar{\mathbf{x}}$. Из равенства (2.4) гл. 6 определяется дисперсия этого среднего

$$\sigma^2(\bar{\mathbf{x}}) = \sigma^2(\mathbf{x})/N. \quad (2.1)$$

По центральной предельной теореме для достаточно большой выборки выборочное среднее $\bar{\mathbf{x}}$ распределено нормально с математическим ожиданием $\hat{\mathbf{x}}$ и дисперсией $\sigma^2(\bar{\mathbf{x}})$, т. е.

$$\mathbf{y} = (\bar{\mathbf{x}} - \hat{\mathbf{x}})/\sigma(\bar{\mathbf{x}}) \quad (2.2)$$

имеет стандартное гауссовское распределение. Однако обычно $\sigma(\bar{\mathbf{x}})$ не известно. Вместо этого мы имеем лишь выборочную оценку для $\sigma^2(\mathbf{x})$:

$$\mathbf{s}_x^2 = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})^2. \quad (2.3)$$

Используя равенство (2.1), мы можем получить оценку для $\sigma^2(\bar{\mathbf{x}})$:

$$\mathbf{s}_{\bar{x}}^2 = \frac{1}{N(N-1)} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})^2. \quad (2.4)$$

Далее мы спрашиваем: насколько (2.2) отличается от стандартного гауссовского распределения, если $\sigma(\bar{\mathbf{x}})$ заменить на $\mathbf{s}_{\bar{x}}$? При помощи сдвига начала отсчета мы всегда можем получить $\hat{\mathbf{x}} = 0$. Поэтому рассмотрим распределение величины

$$\mathbf{t} = \bar{\mathbf{x}}/\mathbf{s}_{\bar{x}} = \bar{\mathbf{x}}\sqrt{N}/\mathbf{s}_x. \quad (2.5)$$

Так как $(N-1)\mathbf{s}_x^2 = f\mathbf{s}_x^2$ имеет χ^2 -распределение с $f = N-1$ степенями свободы, мы можем записать

$$\mathbf{t} = \bar{\mathbf{x}}\sqrt{N}/\sqrt{f}\chi,$$

где $\chi^2 = f\mathbf{s}_x^2$.

Функция распределения величины \mathbf{t} задается вероятностью

$$F(t) = P(\mathbf{t} < t) = P\left(\frac{\bar{x}\sqrt{N}\sqrt{f}}{\chi} < t\right). \quad (2.6)$$

После не очень длинных вычислений находим, что

$$F(t) = \frac{\Gamma\left(\frac{1}{2}(f+1)\right)}{\Gamma\left(\frac{1}{2}f\right)\sqrt{\pi}\sqrt{f}} \int_{-\infty}^t \left(1 + \frac{t^2}{f}\right)^{-\frac{1}{2}(f+1)} dt. \quad (2.7)$$

Соответствующая плотность вероятности равна

$$f(t) = \frac{\Gamma\left(\frac{1}{2}(f+1)\right)}{\Gamma\left(\frac{1}{2}f\right)\sqrt{\pi}\sqrt{f}} \left(1 + \frac{t^2}{f}\right)^{-\frac{1}{2}(f+1)}. \quad (2.8)$$

Сравнение функции $f(t)$ с плотностью стандартного гауссовского распределения дано на рис. 8.3. Когда $N \rightarrow \infty$, выражение (2.8), как и следовало ожидать, стремится к стандартному гауссовскому

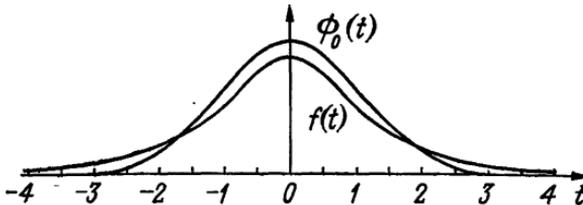


Рис. 8.3. Сравнение плотностей распределения Стьюдента $f(t)$ с двумя степенями свободы и стандартного нормального распределения $\phi_0(t)$.

распределению $\phi_0(t)$. Они похожи и тем, что имеют колоколообразную форму и симметричны относительно нуля. Согласно равенству (9.4) гл. 5

$$P(|\mathbf{t}| \leq t) = 2F(|t|) - 1. \quad (2.9)$$

Далее мы снова можем выбрать границы $\pm t'_\alpha$, соответствующие заданному уровню значимости α , из условия

$$\int_0^{t'_\alpha} f(t) dt = \frac{1}{2}(1 - \alpha), \quad (2.10)$$

т. е. $t'_\alpha = t_{1-\alpha/2}$. Квантили $t_{1-\alpha/2}$ протабулированы в табл. Е.7 для различных значений α и f .

Проиллюстрируем применения критерия Стьюдента*.

Согласно гипотезе, математическое ожидание равно некоторой величине λ_0 . Делается выборка объема N , которая имеет среднее \bar{x} и эмпирическую дисперсию s_x^2 . Если выполняется неравенство

$$|t| = \frac{|\bar{x} - \lambda_0| \sqrt{N}}{s_x} > t'_\alpha = t_{1 - \frac{1}{2}\alpha}, \quad (2.11)$$

то гипотеза должна быть отвергнута.

Это двусторонний критерий. Если важны отклонения только в одном направлении, то критерием с тем же уровнем значимости является

$$t = \frac{(\bar{x} \pm \lambda_0) \sqrt{N}}{s_x} > t'_{2\alpha} = t_{1-\alpha}. \quad (2.12)$$

Теперь мы хотим обобщить этот критерий и применить его для сравнения средних двух выборок. Пусть из двух совокупностей X и Y сделаны выборки объема N_1 и N_2 соответственно. Мы ищем статистику для проверки гипотезы равенства их средних, т. е.

$$\hat{x} = \hat{y}.$$

Согласно центральной предельной теореме, выборочные средние имеют приблизительно нормальное распределение. Их дисперсии равны

$$\sigma^2(\bar{x}) = \frac{1}{N_1} \sigma^2(x), \quad \sigma^2(\bar{y}) = \frac{1}{N_2} \sigma^2(y). \quad (2.13)$$

Они оцениваются величинами

$$s_x^2 = \frac{1}{N_1(N_1 - 1)} \sum_{j=1}^{N_1} (x_j - \bar{x})^2, \\ s_y^2 = \frac{1}{N_2(N_2 - 1)} \sum_{j=1}^{N_2} (y_j - \bar{y})^2. \quad (2.14)$$

Согласно разд. 5.8, разность

$$\Delta = \bar{x} - \bar{y} \quad (2.15)$$

также распределена приблизительно нормально с дисперсией

$$s_\Delta = s_x^2 + s_y^2. \quad (2.16)$$

* Распределение t было введено и применялось в задачах статистических гипотез Госсетом, который публиковал свои работы под псевдонимом Стьюдент (студент).

Если наша гипотеза о равенстве средних верна, то, конечно, среднее значение $\hat{\Delta}$ случайной величины Δ равно нулю. Отношение

$$\Delta/\sigma(\Delta) \quad (2.17)$$

тогда имело бы стандартное гауссовское распределение. Поэтому, используя равенство (9.4) гл. 5, мы могли бы немедленно записать вероятность того, что гипотеза верна, если бы $\sigma(\Delta)$ было известно. Взамен этого из выборок можно получить лишь s_{Δ} . Соответствующее отношение

$$\Delta/s_{\Delta}, \quad (2.18)$$

как правило, бывает немного больше.

Обычно кроме гипотезы $\hat{x} = \hat{y}$ предполагают, что \bar{x} и \bar{y} получены из одной и той же совокупности. Поэтому $\sigma^2(\mathbf{x})$ и $\sigma^2(\mathbf{y})$ совпадают и в качестве самой лучшей оценки для $\sigma^2(\mathbf{x}) = \sigma^2(\mathbf{y})$ можно использовать взвешенное среднее s^2 выборочных оценок s_x^2 и s_y^2 . Соответствующие веса равны $N_1 - 1$ и $N_2 - 1$ и

$$s^2 = \frac{(N_1 - 1) s_x^2 + (N_2 - 1) s_y^2}{(N_1 - 1) + (N_2 - 1)}. \quad (2.19)$$

Образуем

$$S_x^2 = \frac{s^2}{N_1}, \quad S_y^2 = \frac{s^2}{N_2}$$

и

$$s_{\Delta}^2 = S_x^2 + S_y^2 = \frac{N_1 + N_2}{N_1 N_2} s^2. \quad (2.20)$$

Можно доказать (см. Кендалл и Стьюарт [11]), что отношение (2.18) имеет распределение Стьюдента с $f = N_1 + N_2 - 2$ степенями свободы.

Теперь равенство средних может быть проверено с помощью критерия различия Стьюдента. По результатам двух выборок вычисляется величина (2.18). Ее абсолютная величина сравнивается с квантилью распределения Стьюдента с $f = N_1 + N_2 - 2$ степенями свободы, соответствующей уровню значимости α . Если

$$|t| = \frac{|\Delta|}{s_{\Delta}} = \frac{|\bar{x} - \bar{y}|}{s_{\Delta}} \geq t'_{\alpha} = t_{1 - \frac{1}{2}\alpha}, \quad (2.21)$$

то гипотеза равенства средних должна быть отвергнута. В этом случае мы можем принять, что $\hat{x} > \hat{y}$ или $\hat{x} < \hat{y}$ в зависимости от знака величины $\Delta = \bar{x} - \bar{y}$.

Пример 8.2. В табл. 8.2 приведены измерения (в произвольных единицах) концентраций нейраминовой кислоты в эритроцитах (красных кровяных тельцах) больных, страдающих от некоторой

Таблица 8.2

Критерий различия Стьюдента для проверки гипотезы равенства средних; данные примера 8.2

x	y
21	16
24	20
18	22
19	19
25	18
17	19
18	19
22	19
21	
23	
18	
13	
16	
23	
22	
24	
<hr/>	
$N_1=16$	$N_2=7$
$\bar{x}=20,3$	$\bar{y}=19,0$
$s_x^2=171,8/15$	$s_y^2=20/6$

болезни крови (колонка x), и для проверки группы здоровых людей (y). Находим величины выборочных средних и дисперсий

$$|\Delta| = |\bar{x} - \bar{y}| = 1,3,$$

$$s^2 = \frac{15s_x^2 + 6s_y^2}{21} = 9,15,$$

$$s_{\Delta}^2 = \frac{23}{112} s^2 = 1,88.$$

Для $\alpha = 10\%$ и $f = 21$ мы находим, что $t_{1-\frac{1}{2}\alpha} = 2,08$. Поэтому мы заключаем, что экспериментального материала недостаточно, чтобы обнаружить какую-либо связь между заболеванием и концентрацией.

8.3. НЕКОТОРЫЕ АСПЕКТЫ ОБЩЕЙ ТЕОРИИ КРИТЕРИЕВ

До сих пор мы использовали формулы для некоторых критериев, не давая точного обоснования. Например, мы не изложили причин нашего специального выбора критических областей. В этом

разделе дан набросок основных идей более строгой теории критериев. Конечно, полное изложение вопроса выходит за рамки этой книги. Читатель, интересующийся этим вопросом, может обратиться к работам, указанным в библиографии.

Каждая выборка объема N дает N точек в выборочном пространстве, определенном в разд. 2.1. Здесь мы для простоты рассмотрим случай одной непрерывной случайной величины x , при этом результаты выборки фиксируются в виде N точек $(x^{(1)}, x^{(2)}, \dots, x^{(N)})$ на оси x . В случае r случайных величин мы будем иметь N точек в r -мерном пространстве. Результат выборки может быть также представлен в виде одноточечной выборки в Nr -мерном пространстве. Таким образом, например, выборка объема 2 одной переменной может быть мысленно представлена в виде точки на плоскости с координатами $(x^{(1)}, x^{(2)})$. Такое пространство мы будем сокращенно обозначать E . Любая гипотеза* H состоит из предположения о плотности вероятности рассматриваемой случайной величины:

$$f(x; \lambda_1, \lambda_2, \dots, \lambda_p) = f(x; \lambda). \quad (3.1)$$

Гипотеза называется *простой*, если она полностью характеризует плотность вероятности f , т. е. если гипотеза определяет все параметры λ_i ; она называется *сложной*, если общий математический вид f известен, но точное значение по крайней мере одного параметра не определено. Например, простая гипотеза может состоять в том, что f является стандартной гауссовской плотностью, тогда как сложная гипотеза могла бы заключаться в том, что f — это гауссовская плотность с нулевым средним, но с неизвестной дисперсией. Проверяемая гипотеза H_0 называется *нулевой гипотезой*. Иногда мы будем записывать ее в явном виде

$$H_0 (\lambda = \lambda_0) = H_0 (\lambda_1 = \lambda_{10}, \lambda_2 = \lambda_{20}, \dots, \lambda_p = \lambda_{p0}). \quad (3.2)$$

Любая другая возможная гипотеза называется *альтернативной гипотезой*, например

$$H_1 (\lambda = \lambda_1) = H_1 (\lambda_1 = \lambda_{11}, \lambda_2 = \lambda_{21}, \dots, \lambda_p = \lambda_{p1}). \quad (3.3)$$

Часто мы должны проверять простую нулевую гипотезу против сложной альтернативной гипотезы

$$H_1 (\lambda \neq \lambda_0) = H_1 (\lambda_1 \neq \lambda_{10}, \lambda_2 \neq \lambda_{20}, \dots, \lambda_p \neq \lambda_{p0}). \quad (3.4)$$

Так как нулевая гипотеза задает распределение вероятностей в выборочном пространстве, она также определяет вероятность наблюдения точки $X = (x^{(1)}, x^{(2)}, \dots, x^{(N)})$ в любой области про-

*Автор рассматривает здесь только параметрические гипотезы. — Прим. ред.

странства E . Мы можем определить критическую область S_c с уровнем значимости α условием

$$P(X \in S_c | H_0) = \alpha, \quad (3.5)$$

т. е. вероятность попадания точки X внутрь S_c при условии, что гипотеза H_0 верна, будет равна α . Если обнаруживается, что в действительности X попало внутрь S_c , то гипотеза H_0 отвергается. Отметим, что условие (3.5) не определяет критическую область S_c единственным образом.

Хотя использование пространства E выглядит изящным, на самом деле применять его критерии неудобно. Поэтому мы образуем статистику, на основе которой строится критерий,

$$T = T(X) = T(x^{(1)}, x^{(2)}, \dots, x^{(N)}), \quad (3.6)$$

и определим область U значений T , которая соответствует критической области S_c , с помощью отображения

$$X \rightarrow T(X), \quad S_c(X) \rightarrow U(X). \quad (3.7)$$

Если $T \in U$, то нулевая гипотеза отвергается.

Конечно, нулевая гипотеза может быть верна, хотя она и была отвергнута из-за того, что $X \in S_c$. Вероятность такой ошибки, называемой *ошибкой первого рода*, в точности равна α . Имеется другая возможность принять ошибочное решение. Мы можем отвергнуть гипотезу H_1 из-за того, что X не попало в S_c , хотя в действительности нулевая гипотеза ошибочна, а верна альтернативная гипотеза H_1 . Такая ошибка называется *ошибкой второго рода*. Ее вероятность

$$P(X \notin S_c | H_1) = \beta, \quad (3.8)$$

конечно, зависит от конкретной альтернативной гипотезы H_1 . Теперь мы знаем, как надо определять S_c . Ясно, что критерий будет особенно полезным, если для заданной вероятности α критическая область определена таким образом, что вероятность ошибки второго рода β принимает наименьшее значение. Критическая область и сам критерий, естественно, зависят от возможных альтернативных гипотез.

После того как критическая область S_c выбрана, мы можем рассматривать вероятность отвержения нулевой гипотезы как функцию от «истинной» гипотезы или описывающих ее параметров. Эта вероятность по аналогии с равенством (3.5) равна

$$M(S_c, \lambda) = P(X \in S_c | H) = P(X \in S_c | \lambda). \quad (3.9)$$

Эта вероятность, являющаяся функцией от S_c и параметров λ , называется *функцией мощности* критерия. Дополнительная ве-

роятность, а именно вероятность принятия* нулевой гипотезы как функции от параметров истинной гипотезы,

$$L(S_c, \lambda) = 1 - M(S_c, \lambda) \quad (3.10)$$

называется *вероятностью принятия* или *оперативной характеристикой* критерия.

Очевидно, мы имеем

$$\begin{aligned} M(S_c, \lambda_0) &= \alpha, & M(S_c, \lambda_1) &= 1 - \beta, \\ L(S_c, \lambda_0) &= 1 - \alpha, & L(S_c, \lambda_1) &= \beta. \end{aligned} \quad (3.11)$$

Теперь естественно определить *наиболее мощный критерий* простой гипотезы H_0 относительно простой альтернативной гипотезы H_1 как критерий, для которого

$$M(S_c, \lambda_1) = 1 - \beta = \max, \quad (3.12)$$

Может существовать *равномерно наиболее мощный критерий*. Это такой критерий, который является наиболее мощным относительно любой возможной альтернативной гипотезы.

Критерий называется *несмещенным*, если его мощность больше или равна α для любой альтернативной гипотезы

$$M(S_c, \lambda_1) \geq \alpha. \quad (3.13)$$

Это определение является разумным, так как вероятность отвергнуть нулевую гипотезу должна быть наименьшей в том случае, когда эта гипотеза истинна. *Несмещенный наиболее мощный критерий* является наиболее мощным среди всех несмещенных критериев. *Несмещенный равномерно наиболее мощный критерий* определяется соответственно. Методика построения критериев с указанными выше желательными свойствами обсуждается в следующем разделе. Перед тем как возвратиться к этой проблеме, мы проиллюстрируем только что введенные определения на примере.

Пример 8.3. Выборка объема N получена из нормально распределенной совокупности с известной дисперсией σ^2 и неизвестным средним λ . Будет проверяться гипотеза H_0 ($\lambda = \lambda_0$). В качестве статистики мы используем арифметическое среднее $\bar{x} = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$ (в примере 8.4 мы убедимся, что эта статистика является лучшей для нашей цели). Из разд. 6.2 мы

*Мы употребляем выражение «принять» для краткости, хотя должны были бы использовать выражение «отсутствие причин отвергнуть».

знаем, что \bar{x} имеет нормальное распределение со средним λ и дисперсией σ^2/n , т. е. плотность вероятности \bar{x} в случае $\lambda = \lambda_0$ равна

$$f(\bar{x}; \lambda_0) = \frac{\sqrt{n}}{\sqrt{2\pi\sigma}} \exp\left(-\frac{n}{2\sigma^2}(\bar{x} - \lambda_0)^2\right). \quad (3.14)$$

На рис. 8.4 показаны четыре различные критические области U , соответствующие одному и тому же уровню значимости α . Они равны

$$U_1: \bar{x} < \lambda^I \text{ и } \bar{x} > \lambda^{II}, \quad \text{где} \quad \int_{-\infty}^{\lambda^I} f(\bar{x}) d\bar{x} = \int_{\lambda^{II}}^{\infty} f(\bar{x}) d\bar{x} = \frac{1}{2} \alpha;$$

$$U_2: \bar{x} > \lambda^{III}, \quad \text{где} \quad \int_{\lambda^{III}}^{\infty} f(\bar{x}) d\bar{x} = \alpha;$$

$$U_3: \bar{x} < \lambda^{IV}, \quad \text{где} \quad \int_{-\infty}^{\lambda^{IV}} f(\bar{x}) d\bar{x} = \alpha;$$

$$U_4: \lambda^V < \bar{x} < \lambda^{IV}, \quad \text{где} \quad \int_{\lambda^V}^{\lambda^{IV}} f(\bar{x}) d\bar{x} = \int_{\lambda_0}^{\lambda^{IV}} f(\bar{x}) d\bar{x} = \frac{1}{2} \alpha.$$

Чтобы получить функции мощности, соответствующие каждой из этих областей, мы должны теперь изменять λ . Плотность вероятности \bar{x} при любом значении λ по аналогии с равенством (3.14) равна

$$f(\bar{x}; \lambda) = \frac{\sqrt{n}}{\sqrt{2\pi\sigma}} \exp\left[-\frac{n}{2\sigma^2}(\bar{x} - \lambda)^2\right]. \quad (3.15)$$

Пунктирная кривая на рис. 8.4, б изображает эту плотность для $\lambda = \lambda_1 = \lambda_0 + 1$. Далее, функция мощности (3.9) просто равна

$$P(\bar{x} \in U | \lambda) = \int_U f(\bar{x}; \lambda) d\bar{x}. \quad (3.16)$$

Функция мощности, полученная таким образом для критических областей U_1, U_2, U_3, U_4 , изображена на рис. 8.4, в для $n = 2$ и $n = 10$ (соответственно сплошные и пунктирные кривые).

Теперь мы можем сравнить эффективность четырех построенных критериев. Непосредственно из рис. 8.4, в мы видим, что U_1 соответствует несмещенному критерию, так как условие (3.13), очевидно, выполняется.

С другой стороны, область U_2 отвечает наиболее мощному критерию, если альтернативной гипотезой будет гипотеза $H_1(\lambda_1 > \lambda_0)$, и является очень плохой для гипотезы $H_1(\lambda_1 < \lambda_0)$.

Для области U_3 справедливы утверждения с противоположными неравенствами. И наконец, область U_4 дает критерий, у которого вероятность отвергнуть нулевую гипотезу является наибольшей,

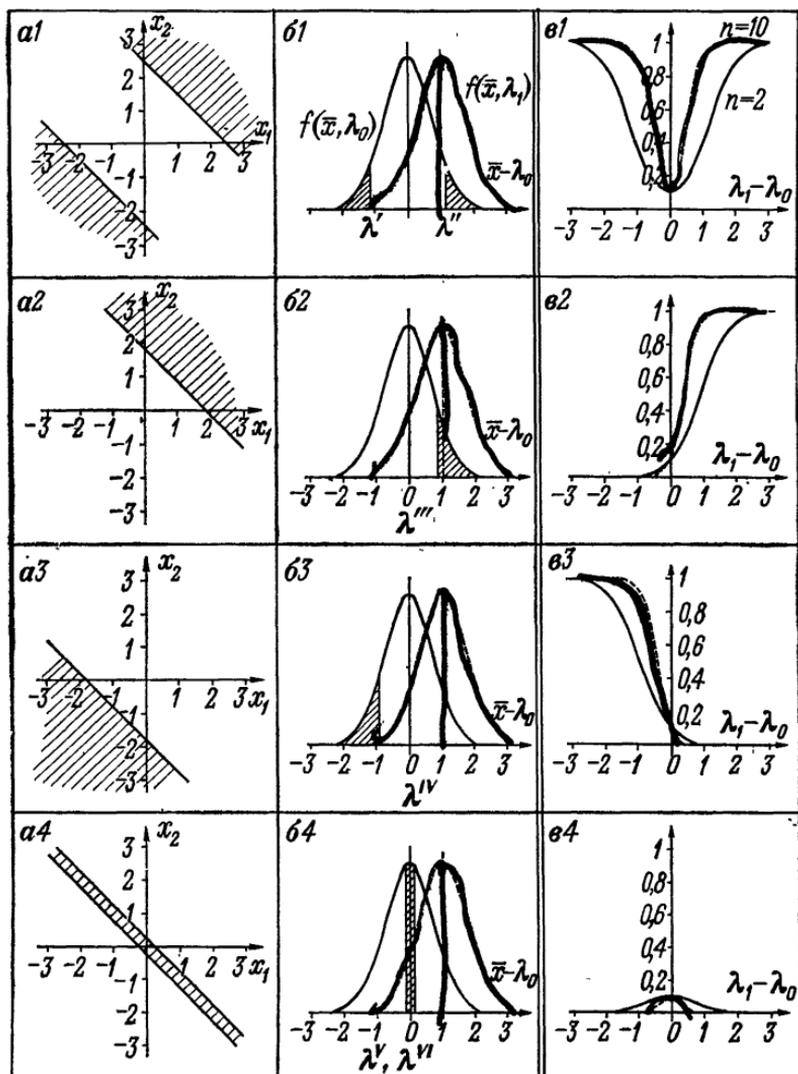


Рис. 8.4. a — критические области в пространстве E ; b — критические области для статистики; v — функции мощности для критерия примера 8.3.

если она истинна. Конечно, это нежелательно. Этот критерий построен исключительно для иллюстрации. Сравнивая первые три критерия, мы видим, что ни один из них не является более 5*

мощным, чем два оставшихся, для всех значений λ_1 . Следовательно, нам не удалось найти равномерно наиболее мощный критерий. В примере 8.4, в котором мы продолжим обсуждение настоящего примера, мы обнаружим, что в нашей задаче такого критерия не существует.

8.4. ТЕОРЕМА НЕЙМАНА — ПИРСОНА И ЕЕ ПРИМЕНЕНИЯ

В предыдущем разделе мы ввели пространство E , в котором выборка описывается одной точкой X . Вероятность попадания точки X внутрь критической области S_c при условии, что справедлива нулевая гипотеза, была уже определена равенством (3.5)

$$P(X \in S_c | H_0) = \alpha. \quad (4.1)$$

Теперь мы определим условную плотность вероятности в пространстве E

$$f(X | H_0).$$

Очевидно, имеем

$$\int_{S_c} f(X | H_0) dX = P(X \in S_c | H_0) = \alpha. \quad (4.2)$$

Рис. 8.5. Области S и S_c .

Имеет место следующая теорема Неймана — Пирсона:

Критерий простой гипотезы H_0 относительно простой гипотезы H_1 является наиболее мощным, если критическая область S_c в пространстве E выбрана так, что

$$\left. \begin{array}{l} \frac{f(X | H_0)}{f(X | H_1)} \leq c \text{ для любого } X \in S_c, \\ \frac{f(X | H_0)}{f(X | H_1)} \geq c \text{ для любого } X \notin S_c. \end{array} \right\} \quad (4.3)$$

Здесь c — постоянная, зависящая от уровня значимости.

Для доказательства теоремы мы кроме S_c рассмотрим еще одну область S . Она может частично перекрываться с S_c (рис. 8.5). Мы так выбираем размеры области S , чтобы она могла служить критической областью с тем же самым уровнем значимости, т. е.

$$\int_S f(X | H_0) dX = \int_{S_c} f(X | H_0) dX = \alpha.$$

Пользуясь обозначениями на рис. 8.5, мы можем записать

$$\int_A f(X | H_0) dX = \int_{S_c} f(X | H_0) dX - \int_C f(X | H_0) dX =$$

$$\begin{aligned}
 &= \int_S f(X | H_0) dX - \int_C f(X | H_0) dX = \\
 &= \int_B f(X | H_0) dX.
 \end{aligned}$$

Так как A содержится в S_c , то можно воспользоваться (4.3)

$$\int_A f(X | H_0) dX < c \int_A f(X | H_1) dX.$$

Соответственно, так как B лежит вне S_c ,

$$\int_B f(X | H_0) dX \geq c \int_B f(X | H_1) dX.$$

Теперь мы можем выразить функцию мощности (3.7) через эти интегралы:

$$\begin{aligned}
 M(S_c, \lambda_1) &= \int_{S_c} f(X | H_1) dX = \int_A f(X | H_1) dX + \int_C f(X | H_1) dX \geq \\
 &\geq \frac{1}{c} \int_A f(X | H_0) dX + \int_C f(X | H_1) dX \geq \\
 &\geq \int_B f(X | H_1) dX + \int_C f(X | H_1) dX \geq \\
 &\geq \int_S f(X | H_1) dX = M(S, \lambda_1),
 \end{aligned}$$

или непосредственно

$$M(S_c, \lambda_1) \geq M(S, \lambda_1). \quad (4.4)$$

Это не что иное, как свойство (3.12) наиболее мощного критерия. Так как мы не делали никакого специального предположения относительно альтернативной гипотезы H_1 ($\lambda = \lambda_1$), или области S , то на самом деле мы доказали, что условие (4.3) обеспечивает *равномерно наиболее мощный критерий*, если это условие справедливо для всех альтернативных гипотез.

Пример 8.4. Мы продолжаем рассмотрение примера 8.3, т. е. мы рассматриваем критерий для выборки объема N из нормальной совокупности с дисперсией σ^2 и неизвестным средним λ . Условная плотность вероятности в точке $X = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)})$ из пространства E есть совместная плотность вероятности величин $\mathbf{x}^{(l)}$ для данного значения λ , т. е.

$$f(X | H_0) = \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^N \exp \left[-\frac{N}{2\sigma^2} \sum_{j=1}^N (x^{(j)} - \lambda_0)^2 \right] \quad (4.5)$$

и

$$f(X|H_1) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left[-\frac{N}{2\sigma^2} \sum_{j=1}^N (x^{(j)} - \lambda_1)^2\right] \quad (4.6)$$

соответственно для нулевой и альтернативной гипотез. Отношение (4.3) принимает вид

$$\begin{aligned} Q &= \frac{f(X|H_0)}{f(X|H_1)} = \exp\left[-\frac{1}{2\sigma^2} \left\{ \sum_{j=1}^N (x^{(j)} - \lambda_0)^2 - \sum_{j=1}^N (x^{(j)} - \lambda_1)^2 \right\}\right] = \\ &= \exp\left[-\frac{1}{2\sigma^2} \left\{ N(\lambda_0^2 - \lambda_1^2) - 2(\lambda_0 - \lambda_1) \sum_{j=1}^N x^{(j)} \right\}\right]. \end{aligned}$$

Выражение

$$\exp\left[-\frac{N}{2\sigma^2} (\lambda_0^2 - \lambda_1^2)\right] = k \geq 0$$

есть неотрицательная постоянная. Поэтому условие (4.3) принимает вид

$$k \exp\left[\frac{\lambda_0 - \lambda_1}{\sigma^2} \sum_{j=1}^N x^{(j)}\right] \begin{cases} \leq c & \text{для } X \in S_c, \\ \geq c & \text{для } X \notin S_c. \end{cases}$$

Оно эквивалентно условию

$$(\lambda_0 - \lambda_1) \bar{x} \begin{cases} \leq c' & \text{для } X \in S_c, \\ \geq c' & \text{для } X \notin S_c. \end{cases} \quad (4.7)$$

с постоянной c' , отличной от c . Соотношение (4.7) не только налагает условия на S_c , но также показывает, что в качестве статистики для критерия надо использовать \bar{x} . Для любого заданного λ_1 , т. е. для любой простой альтернативной гипотезы $H_1 (\lambda = \lambda_1)$, соотношение (4.7) дает ясное определение S_c или, в терминах U , критической области статистики \bar{x} .

Для $\lambda_1 < \lambda_0$ соотношение (4.7) принимает вид

$$\bar{x} \begin{cases} \leq c'' & \text{для } X \in S_c, \\ \geq c'' & \text{для } X \notin S_c. \end{cases}$$

Это соответствует ситуации, рассмотренной на рис. 8.4, б3, с $c'' = \lambda_1^N$. Соответственно для любой альтернативной гипотезы с $\lambda_1 > \lambda_0$ критическая область наиболее мощного критерия задается неравенством

$$\bar{x} \geq c'''$$

(ср. рис. 8.4, б2 с $c''' = \lambda_1^N$). Однако здесь не существует равномерно наиболее мощного критерия, так как множитель $(\lambda_0 - \lambda_1)$ в соотношении (4.7) при $\lambda_1 = \lambda_0$ меняет знак,

8.5. МЕТОД ОТНОШЕНИЯ ПРАВДОПОДОБИЯ

Теорема Неймана — Пирсона дает условия существования наиболее мощного критерия. Однако такой критерий не существует, если альтернативная гипотеза включает значения параметра как большие, так и меньшие, чем значения параметра нулевой гипотезы. Мы обнаружили это в примере 8.4. Можно показать, что это справедливо и в общем случае. Теперь возникает вопрос: каков наилучший критерий, которым можно воспользоваться, если не существует равномерно наиболее мощного критерия? Конечно, понятие «наилучший» определить нелегко, и мы не пытаемся это сделать. Вместо этого мы опишем процедуру, которая дает критерии, обладающие желательными свойствами, и которые имеют то преимущество, что ими легко пользоваться.

Рассмотрим общий случай p параметров $\lambda = (\lambda_1, \dots, \lambda_p)$. Результат выбора, т. е. точка $X = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)})$ в пространстве E , используется для проверки некоторой гипотезы. Сложная нулевая гипотеза определяется некоторой областью значений каждого параметра. Мы можем образовать p -мерное пространство, используя $\lambda_1, \lambda_2, \dots, \lambda_p$ в качестве координат, и рассмотреть совокупность всех областей, соответствующих нулевой гипотезе, как область в пространстве параметров. Обозначим эту область ω . Пусть область Ω в пространстве параметров описывает все возможные значения параметров. Тогда наиболее общая альтернативная гипотеза описывается той частью Ω , которая не содержит ω . Для краткости будем обозначать ее $\Omega - \omega$. Напомним, что в гл. 7 мы ввели оценку максимального правдоподобия $\tilde{\lambda}$ параметра λ . Она была определена как значение λ , для которого функция правдоподобия принимала максимальное значение. Молчаливо предполагалось, что область значений λ , в которой ищется этот максимум, содержала все возможные значения λ . Если имеется в виду максимум лишь в ограниченной области, скажем ω , то он будет обозначаться $\tilde{\lambda}^{(\omega)}$. Определим теперь критерий отношения правдоподобия с помощью статистики

$$T = \frac{f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}; \tilde{\lambda}_1^{(\Omega)}, \tilde{\lambda}_2^{(\Omega)}, \dots, \tilde{\lambda}_p^{(\Omega)})}{f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}; \tilde{\lambda}_1^{(\omega)}, \tilde{\lambda}_2^{(\omega)}, \dots, \tilde{\lambda}_p^{(\omega)})}. \quad (5.1)$$

Здесь $f(x^{(1)}, x^{(2)}, \dots, x^{(N)}; \lambda_1, \lambda_2, \dots, \lambda_p)$ есть совместная плотность вероятности величин $x^{(j)}$ ($j = 1, 2, \dots, N$), т. е. функция правдоподобия (1.4) гл. 7. Критерий отношения правдоподобия состоит в том, что мы отвергаем нулевую гипотезу, если

$$T > T_{1-\alpha}, \quad (5.2)$$

где $T_{1-\alpha}$ определяется из равенства

$$P(T > T_{1-\alpha} | H_0) = \int_{T_{1-\alpha}}^{\infty} g(T | H_0) dT; \quad (5.3)$$

здесь $g(T | H_0)$ — условная плотность вероятности статистики T . Теорема Уилкса [23] дает асимптотическую функцию распределения T , или, вернее, $2 \ln T$ для больших выборок.

Если совокупность описывается плотностью вероятности $f(x; \lambda_1, \lambda_2, \dots, \lambda_r)$, которая удовлетворяет естественным требованиям непрерывности, и если нулевая гипотеза задает r из p параметров

$$H_0 (\lambda_1 = \lambda_{10}, \lambda_2 = \lambda_{20}, \dots, \lambda_r = \lambda_{r0}), \quad r \leq p,$$

то статистика $2 \ln T$, определенная выборкой объема N , при $N \rightarrow \infty$ подчиняется χ^2 -распределению с r степенями свободы*. В случае простой нулевой гипотезы, т. е. когда $r = p$, число степеней свободы равно p .

Теперь применим этот метод к задаче исследования нормальной совокупности с известной дисперсией σ и неизвестным средним λ , которая уже рассматривалась в примерах 8.3 и 8.4.

Пример 8.5. Так как для простой гипотезы $H_0 (\lambda = \lambda_0)$ область ω вырождается в одну точку $\lambda = \lambda_0$, имеем

$$\tilde{\lambda}^{(\omega)} = \lambda_0. \quad (5.4)$$

Рассматривая наиболее общую альтернативную гипотезу $H_1 (\lambda \equiv \equiv \lambda_1 \neq \lambda_0)$, получаем в качестве оценки максимального правдоподобия λ выборочное среднее \bar{x} . Поэтому отношение правдоподобия (5.1) получается таким:

$$T = \frac{f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}; \bar{x})}{f(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}; \lambda_0)}. \quad (5.5)$$

Совместная плотность вероятности задается равенством (2.6) гл. 7:

$$= \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^N \exp \left[-\frac{N}{2\sigma^2} \sum_{i=1}^N (x^{(i)} - \lambda)^2 \right]. \quad (5.6)$$

* Имеется в виду распределение статистики $2 \ln T$ при нулевой гипотезе. — *Прим. ред.*

Следовательно,

$$\begin{aligned} T &= \exp \left[\frac{1}{2\sigma^2} \left\{ - \sum_{j=1}^N (\mathbf{x}^{(j)} - \bar{\mathbf{x}})^2 + \sum_{j=1}^N (\mathbf{x}^{(j)} - \lambda_0)^2 \right\} \right] = \\ &= \exp \left[\frac{1}{2\sigma^2} \sum_{j=1}^N (\bar{\mathbf{x}} - \lambda_0)^2 \right] = \\ &= \exp \left[\frac{N}{2\sigma^2} (\bar{\mathbf{x}} - \lambda_0)^2 \right]. \end{aligned}$$

Теперь мы должны вычислить $T_{1-\alpha}$ и отвергнуть гипотезу H_0 , если выполняется неравенство (5.2). Поскольку логарифм — монотонная функция T , в качестве статистики критерия можно использовать

$$T' = 2 \ln T = \frac{N}{\sigma^2} (\bar{\mathbf{x}} - \lambda_0)^2,$$

и тогда гипотеза H_0 отвергается, если

$$T' > T'_{1-\alpha},$$

где

$$\int_{T'_{1-\alpha}}^{\infty} h(T' | H_0) dT' = \alpha.$$

Чтобы вычислить плотность вероятности $h(T' | H_0)$ статистики T' мы сначала вычислим плотность $f(\bar{x})$ выборочного среднего при условии, что $\lambda = \lambda_0$; она равна

$$f(\bar{x} | H_0) = \left[\frac{N}{2\pi\sigma^2} \right]^{1/2} \exp \left(- \frac{N}{2\sigma^2} (\bar{x} - \lambda_0)^2 \right).$$

Чтобы произвести преобразование переменных, заданное равенством (4.1) гл. 4, нам необходима ещё производная

$$\left| \frac{d\bar{x}}{dT'} \right| = \frac{1}{2} \left(\frac{\sigma^2}{N} \right)^{1/2} T'^{-1/2},$$

которая легко получается из определения T' . Тогда

$$h(T' | H_0) = \left| \frac{d\bar{x}}{dT'} \right| f(\bar{x} | H_0) = \frac{1}{\sqrt{2\pi}} T'^{-1/2} e^{-\frac{1}{2} T'}. \quad (5.7)$$

Это в действительности χ^2 -распределение с одной степенью свободы. В нашем примере теорема Уилкса верна даже для конечного N . Кроме того, мы видим, что критерий отношения правдоподобия дает несмещенный критерий, изображенный на рис. 8.4, б1. Критерий

$$T' = - \frac{N}{\sigma^2} (\bar{x} - \lambda_0)^2 > T'_{1-\alpha}$$

эквивалентен критерию

$$\left(\frac{N}{\sigma^2}\right)^{\frac{1}{2}} (\bar{x} - \lambda_0) < \lambda', \quad \left(\frac{N}{\sigma^2}\right)^{\frac{1}{2}} (\bar{x} - \lambda_0) > \lambda'', \quad (5.8)$$

где

$$\lambda' = -\lambda'' = (-T'_{1-\alpha})^{\frac{1}{2}} = (\chi^2_{1-\alpha})^{\frac{1}{2}} = \chi_{1-\alpha}.$$

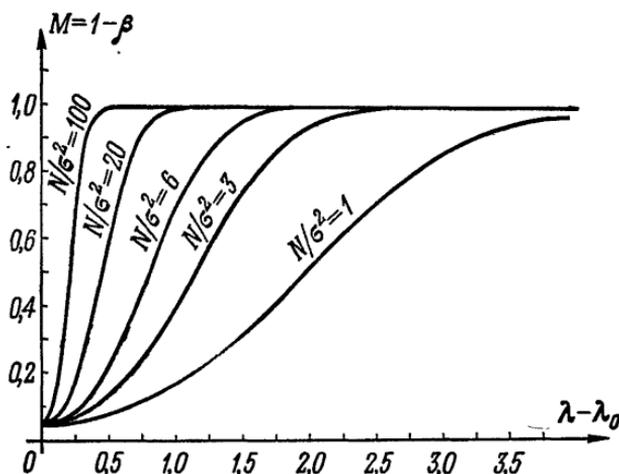


Рис. 8.6. Функции мощности для критерия в примере 8.5.

Используя этот результат, мы можем точно вычислить функцию мощности нашего критерия. При любом значении среднего совокупности плотность выборочного среднего равна

$$f(\bar{x}; \lambda) = \left(\frac{N}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp\left[-\frac{N(\bar{x} - \lambda)^2}{2\sigma^2}\right] = \phi_0\left(\frac{\bar{x} - \lambda}{\sigma/\sqrt{N}}\right).$$

Подставляя (5.8) в (3.9), имеем

$$M(S_c; \lambda) = \int_{-\infty}^A f(\bar{x}; \lambda) d\bar{x} + \int_B^{\infty} f(\bar{x}; \lambda) d\bar{x} = \\ = \psi_0\left(\chi_{1-\alpha} - \frac{\lambda - \lambda_0}{\sigma/\sqrt{N}}\right) + \psi_0\left(\chi_{1-\alpha} + \frac{\lambda - \lambda_0}{\sigma/\sqrt{N}}\right), \quad (5.9)$$

где

$$A = -\chi_{1-\alpha}\sigma/\sqrt{N} - \lambda_0, \quad B = \chi_{1-\alpha}\sigma/\sqrt{N} - \lambda_0.$$

Здесь ϕ_0 и ψ_0 обозначают плотность вероятности и функцию распределения стандартного нормального распределения. На рис. 8.6 представлена функция мощности для $\alpha = 0,05$ и различных значений N/σ^2 .

Пример 8.6. Теперь мы рассмотрим сложную нулевую гипотезу. Вновь произведем выборку из нормальной совокупности. Однако на этот раз и среднее λ и дисперсия σ^2 неизвестны. Нулевая гипотеза H_0 ($\lambda = \lambda_0$) не накладывает ограничения на область изменений σ^2 . Из примера 7.7 мы знаем оценки максимального правдоподобия в полном пространстве параметров:

$$\begin{aligned}\tilde{\lambda}^{(\omega)} &= \bar{\mathbf{x}}, \\ \sigma^{2(\omega)} &= \frac{1}{N} \sum_{j=1}^N (\mathbf{x}^{(j)} - \bar{\mathbf{x}})^2 = \mathbf{s}'^2.\end{aligned}$$

Если описывать нулевую гипотезу в пространстве параметров, имеем

$$\begin{aligned}\tilde{\lambda}^{(\omega)} &= \lambda_0, \\ \sigma^{2(\omega)} &= \sigma^2(\omega) = \frac{1}{N} \sum_{j=1}^N (\mathbf{x}^{(j)} - \lambda_0)^2.\end{aligned}$$

Отношение правдоподобия (5.1) будет таким:

$$\begin{aligned}T &= \left(\frac{\sum (\mathbf{x}^{(j)} - \lambda_0)^2}{\sum (\mathbf{x}^{(j)} - \bar{\mathbf{x}})^2} \right)^{\frac{1}{2}N} \exp \left(- \frac{\sum (\mathbf{x}^{(j)} - \bar{\mathbf{x}})^2}{\sum (\mathbf{x}^{(j)} - \bar{\mathbf{x}})^2} + \frac{\sum (\mathbf{x}^{(j)} - \lambda_0)^2}{\sum (\mathbf{x}^{(j)} - \lambda_0)^2} \right) = \\ &= \left(\frac{\sum (\mathbf{x}^{(j)} - \lambda_0)^2}{\sum (\mathbf{x}^{(j)} - \bar{\mathbf{x}})^2} \right)^{\frac{1}{2}N}.\end{aligned}$$

Вновь заменим эту статистику на статистику T' , которая является монотонной функцией от T . T' определяется таким образом:

$$\begin{aligned}T' = T^{2/N} &= \frac{\sum (\mathbf{x}^{(j)} - \lambda_0)^2}{\sum (\mathbf{x}^{(j)} - \bar{\mathbf{x}})^2} = \frac{\sum (\mathbf{x}^{(j)} - \bar{\mathbf{x}})^2 + N(\bar{\mathbf{x}} - \lambda_0)^2}{\sum (\mathbf{x}^{(j)} - \bar{\mathbf{x}})^2}, \\ T' &= 1 + \frac{t^2}{N-1}.\end{aligned}\tag{5.10}$$

Здесь

$$t = N^{\frac{1}{2}} \frac{\bar{\mathbf{x}} - \lambda_0}{\left(\frac{\sum (\mathbf{x}^{(j)} - \bar{\mathbf{x}})^2}{N-1} \right)^{\frac{1}{2}}} = N^{\frac{1}{2}} \frac{\bar{\mathbf{x}} - \lambda_0}{\mathbf{s}_x}$$

является статистикой Стьюдента, рассмотренной в разд. 8.2.

Из равенства (5.10) мы можем, следовательно, вычислить значение t для конкретной выборки и отвергнуть гипотезу, если

$$|t| > t_{1 - \frac{1}{2}\alpha}$$

Самый общий метод отношения правдоподобия привел нас к критерию Стьюдента, который был построен специально для проверки того, что выборка происходит из совокупности с нормальным распределением с известным средним и неизвестной дисперсией.

8.6. КРИТЕРИЙ СОГЛАСИЯ χ^2

До сих пор мы устанавливали критерии для гипотез, которые определяли значение одного или нескольких параметров совокупности. Их можно назвать *параметрическими критериями*. Процедура состояла в том, чтобы сначала найти статистику, на основе которой строится критерий, а затем, вычислив эту статистику, т. е. определив по выборке единственное число, принять или отвергнуть гипотезу. Другой класс критериев сравнивает функцию распределения выборки прямо с распределением совокупности — последнее должно основываться на некоторой гипотезе. Такие критерии представляют собой *критерии согласия*. Мы рассмотрим только важнейший из них, так называемый *критерий χ^2* .

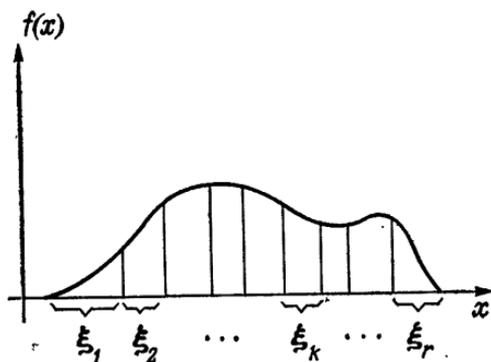


Рис. 8.7. Разбиение значений случайной величины x на интервалы ξ_k .

Обозначим функцию распределения и плотность вероятности совокупности соответственно через $F(x)$ и $f(x)$. Полную область изменения случайной переменной x можно разбить на r интервалов

$$\xi_1, \xi_2, \dots, \xi_k, \dots, \xi_r.$$

Это показано на рис. 8.7. Интегрируя $f(x)$ на отдельных интервалах, получаем вероятности попадания наблюдения x в ξ_k :

$$p_k = P(x \in \xi_k) = \int_{\xi_k} f(x) dx, \quad \sum_{k=1}^r p_k = 1. \quad (6.1)$$

Теперь произведем выборку объема n и обозначим через n_k число

элементов выборки, которые попали внутрь интервала ξ_k . Разумеется,

$$\sum_{k=1}^r n_k = n. \quad (6.2)$$

Исходя из предполагаемой плотности вероятности, мы должны ожидать, что n_k примет значение np_k . В качестве меры отклонения распределения выборки от гипотетического распределения кажется разумным использовать статистику

$$\chi^2 = \sum_{k=1}^r \frac{(n_k - np_k)^2}{np_k} = \frac{1}{n} \sum_{k=1}^r \frac{n_k^2}{p_k} - n. \quad (6.3)$$

Статистика χ^2 асимптотически (т. е. при $n \rightarrow \infty$) распределена как χ^2 с $r - 1$ степенью свободы.

Имеется несколько совершенно различных доказательств этого важного свойства χ^2 . Кратко изложим здесь довольно изящное доказательство, принадлежащее Фишеру. Заметим, что равенства (6.1) и (6.2) и рис. 8.7 соответствуют полиномиальному распределению (разд. 5.1). Согласно (1.10) гл. 5, совместная вероятность наблюдения n_1 элементов выборки в интервале ξ_1 , n_2 элементов в ξ_2 и т. д. такова:

$$P = \frac{n!}{\prod_{k=1}^r n_k!} \prod_{k=1}^r p_k^{n_k}. \quad (6.4)$$

Теперь рассмотрим r независимых пуассоновских распределений с параметрами np_1, np_2, \dots, np_r , т. е. предположим, что значение n_k из интервала ξ_k описывается распределением Пуассона с параметром np_k . Тогда вероятность наблюдения n_1 элементов в ξ_1 , n_2 в ξ_2 и т. д. равна

$$P(n_1, n_2, \dots, n_r) = \prod_{k=1}^r \frac{(np_k)^{n_k}}{n_k!} e^{-np_k} = e^{-n} n^n \prod_{k=1}^r \frac{p_k^{n_k}}{n_k!}. \quad (6.5)$$

Условие (6.2), разумеется, должно быть выполнено. Согласно примеру 5.5, сумма r независимых величин, распределенных по закону Пуассона, снова имеет пуассоновское распределение, параметр которого задается как сумма параметров слагаемых

$$P(n) = \frac{n^n}{n!} e^{-n}. \quad (6.6)$$

Вероятность получить n_1 наблюдений в ξ_1 , n_2 в ξ_2 и т. д. при условии, что всего наблюдалось n элементов, теперь получается

просто в виде отношения (6.5) к (6.6) [см. равенство (3.1) гл. 2]

$$P(n_1, n_2, \dots, n_r | n) = \frac{n!}{\prod_{k=1}^r n_k!} \prod_{k=1}^r (p_k)^{n_k}. \quad (6.7)$$

Это совпадает с (6.4). Наш случай, следовательно, описывается полиномиальным распределением, или, что то же самое, r независимыми величинами, подчиняющимися распределению Пуассона со средними значениями и дисперсиями, равными np_k . Вместо n_k можно ввести преобразованные переменные

$$u_k = \frac{n_k - np_k}{\sqrt{np_k}}, \quad (6.8)$$

для которых среднее равно нулю, дисперсия — единице.

Согласно центральной предельной теореме, асимптотический предел величин u_k подчиняется нормальному гауссовскому распределению. Выражение (6.3)

$$X^2 = \sum_{k=1}^r \frac{(n_k - np_k)^2}{np_k} = \sum_{k=1}^r u_k^2 \quad (6.9)$$

в пределе представляет собой сумму квадратов нормальных гауссовских величин с одним дополнительным условием [равенства (6.2) и (6.8)]

$$\sum_{k=1}^r \sqrt{np_k} u_k = \sum_{k=1}^r n_k - \sum_{k=1}^r np_k = 0. \quad (6.10)$$

В разд. 6.5 и 6.6 мы получили, что такая сумма квадратов подчиняется распределению χ^2 с $r - 1$ степенью свободы.

Критерий χ^2 применяется следующим образом. Делается n наблюдений x_1, x_2, \dots, x_n случайной величины x . Разбиение области значений x на интервалы $\xi_1, \xi_2, \dots, \xi_r$ выбирается таким, чтобы число наблюдений n_k в каждом интервале было достаточно велико и X^2 можно было с уверенностью рассматривать как величину, распределенную по χ^2 . С другой стороны, должно быть достаточное количество интервалов, потому что в противном случае ступенчатая функция от результатов наблюдений уже не будет хорошо описывать плотность вероятности $f(x)$. Обычно придерживаются эмпирического правила, что n_k должно быть равно по крайней мере 4. Далее выбирается уровень значимости α , подсчитывается величина X^2 и сравнивается с квантилью $\chi_{1-\alpha}^2$ с $r - 1$ степенью свободы. Если $X^2 > \chi_{1-\alpha}^2$, то гипотеза, состоящая в том, что плотность равна $f(x)$, отвергается.

До сих пор мы предполагали, что эта гипотеза проста, т. е. $f(x)$ полностью известна. Иногда бывает нужно проверить *сложную* гипотезу. В таком случае может быть известен общий вид $f(x)$, однако значения одного или нескольких параметров $f(x)$ еще должны быть получены из наблюдений. Опишем кратко общий случай p неизвестных параметров $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$, где $p < r$. Поскольку вероятности p_k есть функции $f(x; \lambda)$ [равенство (6.1)], имеем

$$p_k = p_k(\lambda).$$

Функция правдоподобия задается непосредственно совместной вероятностью (6.4)

$$L(\lambda) = \frac{n!}{r \prod_{k=1}^r n_k!} \prod_{k=1}^r p_k(\lambda)^{n_k}.$$

Получаем систему уравнений правдоподобия

$$\frac{\partial L}{\partial \lambda_i} = \sum_{k=1}^r \frac{n_k}{p_k} \frac{\partial p_k}{\partial \lambda_i} = 0, \quad i = 1, 2, \dots, p. \quad (6.11)$$

Эти соотношения накладывают на переменные (6.8) дальнейшие ограничения такого же типа, как ограничение (6.10). Таким образом, число степеней свободы суммы квадратов (6.9) уменьшается на p (разд. 6.6). Итак, мы получили следующее правило использования критерия χ^2 для сложной гипотезы.

Если параметры $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_p)$ оцениваются из наблюдений методом максимального правдоподобия, то величина

$$\chi^2 = \sum_{k=1}^r \frac{(n_k - np_k(\lambda))^2}{np_k(\lambda)} \quad (6.12)$$

(асимптотически) подчиняется χ^2 -распределению с $f = r - p - 1$ степенями свободы.

Пример 8.7. При экспериментальном изучении взаимодействия фотонов и протонов водородная пузырьковая камера подвергается действию пучка фотонов (γ -квантов) высокой энергии. Фотоны также дают увеличение электронно-позитронных пар, которые в этом эксперименте не представляют непосредственного интереса, однако их средние числа на фотографии пузырьковой камеры можно использовать как меру интенсивности фотонного пучка. Число фотографий с 0, 1, 2, ... парами должно подчиняться распределению Пуассона (см. примеры в разд. 5.4). Если отклонения от распределения Пуассона значительны, надо сделать вывод, что имеют место потери наблюдений и, следовательно, систематические ошибки в эксперименте.

Таблица 8.3

Данные для применения критерия χ^2 к примеру 8.7

Число электронных пар на фотографиях, k	Число фотографий с k электронными парами, n_k	Значения, предсказанные по распределению Пуассона, np_k
0	47	34,4
1	69	80,2
2	84	93,7
3	76	72,8
4	49	42,6
5	16	19,9
6	11	7,8
7	3	2,5
8	—	(0,7)

Во второй колонке табл. 8.3 и на рис. 8.8 даны результаты наблюдений $n = 355$ фотографий. Из примера 7.4 мы знаем, что оценка максимального правдоподобия для параметров распределения Пуассона задается таким образом:

$$\tilde{\lambda} = \frac{\sum_k kn_k}{\sum_k n_k}$$

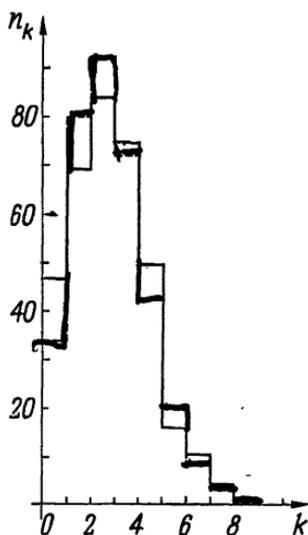


Рис. 8.8. Сравнение распределения Пуассона (пунктир) с выборочным распределением (сплошная линия).

Находим, что $\tilde{\lambda} = 2,33$. Из табл. Е.1 можно получить значения p_k пуассоновского распределения с таким параметром; эти значения, умноженные на n , даны в третьей колонке табл. 8.3. Значение $\chi^2 = 11,47$ получается, если просуммировать квадраты разностей второй и третьей колонок, разделенных на значения третьей колонки. Степеней свободы 6, так как $r = 8$, $p = 1$. Мы выбираем $\alpha = 1\%$ и из табл. Е.5 находим $\chi_{0,99}^2 = 16,81$. Следовательно, нет оснований отвергать гипотезу о том, что имеет место распределение Пуассона.

Пример 8.8. Пусть проведено n экспериментов, результаты которых характеризуются значениями двух случайных величин x и y . Предположим, что величины дискретные и что они могут принимать значения $x_1, x_2, \dots, x_k; y_1, y_2, \dots, y_l$. (Непрерывную величину можно приблизить дискретными, разделив область ее значений на интервалы, как на рис. 8.7.) Число экспериментов, в которых получились значения $x = x_i$,

а $\mathbf{y} = y_i$, обозначим n_{ij} . Из этих чисел можно образовать матрицу, которая называется *таблицей сопряженности признаков* (табл. 8.4).

Таблица 8.4

Таблица сопряженности признаков

	y_1	y_2	...	y_l
x_1	n_{11}	n_{12}	...	n_{1l}
x_2	n_{21}	n_{22}	...	n_{2l}
...
x_k	n_{k1}	n_{k2}	...	n_{kl}

Обозначим вероятность получить $\mathbf{x} = x_i$ через p_i и аналогично вероятность $\mathbf{y} = y_j$ — через q_j , т. е. $p_i = P(\mathbf{x} = x_i)$ и $q_j = P(\mathbf{y} = y_j)$. Если величины \mathbf{x} и \mathbf{y} независимы, то вероятность того, что $\mathbf{x} = x_i$ и одновременно $\mathbf{y} = y_j$, равна произведению $p_i q_j$. Оценки максимального правдоподобия для p и q таковы:

$$\tilde{p}_i = \frac{1}{n} \sum_{j=1}^l n_{ij}, \quad \tilde{q}_j = \frac{1}{n} \sum_{i=1}^k n_{ij}. \quad (6.13)$$

Так как

$$\sum_{j=1}^l q_j = \sum_{i=1}^k p_i = 1, \quad (6.14)$$

то фактически имеется лишь $k + l - 2$ неизвестных параметров, которые надо оценить методом максимального правдоподобия.

Теперь элементы таблицы сопряженности признаков можно выписать в один ряд:

$$n_{11}, n_{12}, \dots, n_{1l}, n_{21}, n_{22}, \dots, n_{2l}, \dots, n_{kl}$$

и применить χ^2 -критерий. Мы должны вычислить величину

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - n \tilde{p}_i \tilde{q}_j)^2}{n \tilde{p}_i \tilde{q}_j} \quad (6.15)$$

и сравнить ее с квантилью $\chi_{1-\alpha}^2$ распределения χ^2 , соответствующего заданному уровню значимости α . Число степеней свободы получается уменьшением числа элементов таблицы сопряженности признаков на единицу и вычитанием числа оцененных параметров

$$f = kl - 1 - (k + l - 2) = (k - 1)(l - 1). \quad (6.16)$$

Если случайные величины \mathbf{x} и \mathbf{y} зависимы, то n_{ij} будет, вообще говоря, отличным от $n \tilde{p}_i \tilde{q}_j$. Поэтому окажется, что

$$\chi^2 > \chi_{1-\alpha}^2,$$

и гипотеза независимости должна быть отвергнута.

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Метод наименьших квадратов восходит к работам Лежандра и Гаусса соответственно 1805 и 1809 гг. Он основан на правиле, которое в простейшем случае принимает следующий вид.

Результаты повторных наблюдений y_j могут быть рассмотрены как сумма (неизвестной) величины x и погрешностей наблюдения ϵ_j :

$$y_j = x + \epsilon_j.$$

Теперь нужно определить величину x таким образом, чтобы сумма квадратов погрешностей ϵ_j достигала своего минимума:

$$\sum_j \epsilon_j^2 = \sum_j (x - y_j)^2 = \min.$$

Мы обнаружим, что во многих случаях это правило можно получить из метода максимального правдоподобия, который был развит значительно позднее. Однако даже в тех случаях, когда метод наименьших квадратов не совпадает с методом максимального правдоподобия, он часто приводит к удовлетворительным результатам по сравнению с другими возможными способами оценивания. Метод наименьших квадратов, который из всех статистических методов наиболее употребителен в практических задачах, можно использовать и в том случае, когда наблюдения величины не прямо связаны с неизвестной величиной x , а косвенно, т. е. когда их надо рассматривать как функции нескольких неизвестных. Мы намереваемся сначала обсудить различные частные случаи, а не начинать рассмотрение с наиболее общего случая. Поскольку эти частные случаи представляют большой практический интерес, мы их рассмотрим довольно подробно и разберем некоторые примеры.

9.1. ПРЯМЫЕ РАВНОТОЧНЫЕ И НЕРАВНОТОЧНЫЕ НАБЛЮДЕНИЯ

Простейший случай уже был описан в общих чертах. Пусть произведено всего n измерений неизвестной величины x . Измеренные значения y_j содержат в себе погрешности измерения ϵ_j , о ко-

торых мы предполагаем, что они *нормально* распределены вокруг нуля:

$$y_j = x + \varepsilon_j, \quad E(\varepsilon_j) = 0, \quad E(\varepsilon_j^2) = \sigma^2. \quad (1.1)$$

По-видимому, во многих случаях это предположение оправдывается центральной предельной теоремой.

Вероятность получить в результате измерения значение в интервале от y_j до $y_j + dy$ равна

$$f_j dy = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(y_j - x)^2}{2\sigma^2}\right] dy.$$

Поэтому логарифмическая функция правдоподобия для всех n измерений будет равна (см. пример 7.2)

$$l = -\frac{1}{2\sigma^2} \sum_{j=1}^n (y_j - x)^2 + \text{const.} \quad (1.2)$$

Поэтому условие максимума правдоподобия $l = \max$ эквивалентно следующему:

$$M = \sum_{j=1}^n (y_j - x)^2 = \sum_{j=1}^n \varepsilon_j^2 = \min; \quad (1.3)$$

это как раз и есть условие наименьших квадратов. Как мы уже показали в примерах 7.2 и 7.4, из (1.3) следует, что наилучшей оценкой для x является среднее арифметическое значений y_j :

$$\tilde{x} = \bar{y} = \frac{1}{n} \sum_{j=1}^n y_j. \quad (1.4)$$

Дисперсия этой величины равна

$$\sigma^2(\bar{y}) = \sigma^2/n, \quad (1.5)$$

или, определяя ошибку как стандартное отклонение, получаем

$$\Delta\tilde{x} = \Delta y/\sqrt{n}. \quad (1.6)$$

Более общий случай *прямых неравноточных наблюдений* также был уже рассмотрен в примере 7.4. Мы снова предполагаем, что погрешности наблюдений нормально распределены вокруг нуля, т. е.

$$y_j = x + \varepsilon_j, \quad E(\varepsilon_j) = 0, \quad E(\varepsilon_j^2) = \sigma_j^2 = 1/g_j. \quad (1.7)$$

Сравнение с равенством (2.7) гл. 7 показывает, что метод максимального правдоподобия приводит к условию

$$\begin{aligned} M &= \sum_{j=1}^n \frac{(y_j - x)^2}{\sigma_j^2} = \sum_{j=1}^n g_j (y_j - x)^2 = \\ &= \sum_{j=1}^n g_j \varepsilon_j^2 = \min. \end{aligned} \quad (1.8)$$

Члены суммы квадратов теперь *взвешены* обратно пропорционально дисперсиям. Наилучшей оценкой для x в этом случае является [см. равенство (2.8) гл. 7]

$$\tilde{x} = \frac{\sum_{j=1}^n g_j y_j}{\sum_{j=1}^n g_j}, \quad (1.9)$$

т. е. взвешенное среднее индивидуальных наблюдений.

Можно убедиться, что чем больше дисперсия наблюдения, тем меньше его вклад в конечный результат. Кроме того, из равенства (3.20) гл. 7 нам известна дисперсия этой оценки; она равна

$$\sigma^2(\tilde{x}) = \left(\sum_{j=1}^n \frac{1}{\sigma_j^2} \right)^{-1} = \left(\sum_{j=1}^n g_j \right)^{-1}. \quad (1.10)$$

Мы можем использовать формулу (1.9) для вычисления наилучших выборочных оценок $\tilde{\varepsilon}_j$ исходных погрешностей ε_j , определенных равенством (1.1). Получаем

$$\tilde{\varepsilon}_j = y_j - \tilde{x}.$$

Эти величины распределены нормально с нулевым средним и с дисперсией σ_j^2 , т. е. величины ε_j/σ_j имеют нормальное гауссовское распределение. Тогда, согласно разд. 6.6, сумма

$$M = \sum_{j=1}^n \left(\frac{\tilde{\varepsilon}_j}{\sigma_j} \right)^2 = \sum_{j=1}^n \frac{(y_j - \tilde{x})^2}{\sigma_j^2} = \sum_{j=1}^n g_j (y_j - \tilde{x})^2 \quad (1.11)$$

будет иметь распределение χ^2 с $n - 1$ степенью свободы.

Это свойство величины M может быть теперь использовано для проверки критерием χ^2 справедливости нашего предположения (1.7). Если для выбранного уровня значимости α величина M превышает $\chi_{1-\alpha}^2$, мы должны будем пересмотреть предположения (1.7). Как правило, не подвергается сомнению, что величины y_j действительно являются наблюдениями неизвестной величины x . Погрешности ε_j

могут, однако, не иметь нормального распределения. В частности, наблюдения могут быть смещенными, т. е. математическое ожидание погрешностей ϵ_j не обращается в нуль.

Обычно такие систематические погрешности, изменяющиеся от одного наблюдения к другому, приводят к тому, что критерий χ^2 не выполняется.

Пример 9.1. Наилучшие значения наиболее важных констант обычно получаются при помощи взвешенных средних для измерений, выполненных различными экспериментальными группами. Для свойств элементарных частиц такие средние были найдены Розенфельдом и другими (см., например, Розенфельд и др. [18]). Рассмотрим полученные при помощи таких вычислений средние значения масс нейтральных К-мезонов (K^0). При усреднении использованы результаты четырех экспериментов, которые были выполнены с помощью различной техники. Вычисления могут быть произведены по схеме, указанной в табл. 9.1. Величина M

Таблица 9.1

Взвешенные средние 4-х измерений массы нейтрального К-мезона

№ измерений	y_j (масса K^0 в МэВ)	σ_j	$1/\sigma_j^2 = g_j$	$y_j g_j$	$y_j - \bar{x}$	$(y_j - \bar{x})^2 g_j$
1	498,1	0,4	6,3	3038,0	0,2	0,3
2	497,44	0,33	10	4974,4	-0,46	2,1
3	498,9	0,5	4	1995,6	1,0	4,0
4	497,44	0,5	4	1989,8	-0,46	0,8

$$\sum g_j = 24,3 \quad \sum y_j g_j = 11997,8$$

$$M = \sum (y_j - \bar{x})^2 g_j = 7,2$$

$$\bar{x} = \frac{\sum y_j g_j}{\sum g_j} = 497,9$$

$$\Delta \bar{x} = (\sum g_j)^{-1/2} = 0,20$$

принимает значение 7,2. Мы выбираем $\alpha = 5\%$. Из табл. Е.5 находим, что для трех степеней свободы $\chi_{0,95}^2 = 7,82$. Поэтому мы согласны принять значение $m_{K^0} = (497,9 \pm 0,2)$ МэВ в качестве наилучшей оценки массы K^0 до тех пор, пока не будут проведены новые эксперименты.

Рассмотрим теперь случай, когда критерий χ^2 не выполняется. Как мы указывали выше, в этом случае можно подозревать, что по крайней мере одно из наблюдений является смещенным. При рассмотрении отдельных наблюдений иногда можно заметить, что одно или два наблюдения очень сильно отличаются от всех остальных.

ных. Такой случай представлен на рис. 9.1, *а*, на котором изображены результаты различных наблюдений вместе с их ошибками. Результаты наблюдений откладываются по оси ординат, номера наблюдений — по оси абсцисс. Критерий χ^2 не выполняется, если использовать все наблюдения рис. 9.1, *а*; но он будет выполнен, если наблюдения № 4 и 6 исключить.

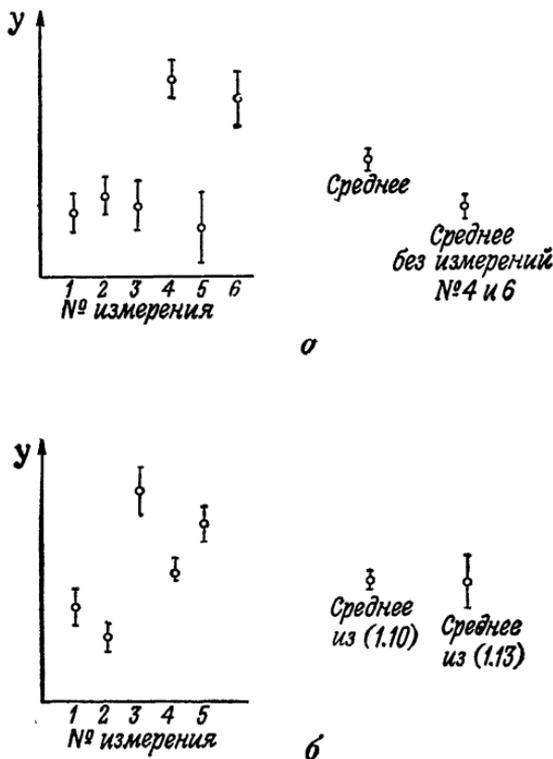


Рис. 9.1. Случай взвешенных средних с невыполняющимся критерием χ^2 . *а* — аномальное отклонение нескольких наблюдений, *б* — стандартные отклонения отдельных наблюдений, видимо, слишком малы.

К сожалению, часто ситуация совсем не так проста. На рис. 9.1, *б* дан пример, в котором критерий χ^2 тоже не выполняется. Однако здесь нет единственного наблюдения, из-за которого это происходило бы. В этом случае оправданным с математической точки зрения было бы решение не делать попыток усреднения и нахождения наилучшей оценки неизвестного параметра до тех пор, пока не будут сделаны дальнейшие измерения.

На практике этим, конечно, нельзя удовлетвориться. Розенфельд и др. [18] предложили метод увеличения индивидуальных погреш-

ностей, использующий нормирующие множители $\sqrt{M/(n-1)}$, т. е. замену σ_j на

$$\sigma_j = \sigma_j \sqrt{\frac{M}{n-1}}. \quad (1.12)$$

Если рассматривать погрешности с такими дисперсиями, то взвешенное среднее \bar{x} не отличается от результата выражения (1.9). Его дисперсия, однако, отлична от (1.10):

$$\sigma'^2(\bar{x}) = \frac{M}{n-1} \left(\sum_{j=1}^n \frac{1}{\sigma_j} \right)^{-1}. \quad (1.13)$$

Вычислим теперь выражение, соответствующее (1.11):

$$M' = \frac{n-1}{M} \sum_{j=1}^n \frac{(y_j - \bar{x})^2}{\sigma_j^2} = \frac{n-1}{M} M = n-1. \quad (1.14)$$

Это равно математическому ожиданию χ^2 с $(n-1)$ степенью свободы. Мы ввели нормировку (1.12) специально для того, чтобы получить равенство (1.14). Напоминаем читателю, что эта нормировка не базируется на строгой математической основе. Этот метод необходимо использовать с осторожностью, так как он подавляет эффект систематического различия между наблюдениями. С другой стороны, в случаях, подобных рис. 9.1, б, получаются разумные ошибки усредненных величин, в то время как непосредственное употребление равенства (1.10) дает ошибки, которые слишком малы, чтобы отразить разброс индивидуальных наблюдений. Оба решения приведены на рис. 9.1, б.

9.2. НЕПРЯМЫЕ (КОСВЕННЫЕ) ИЗМЕРЕНИЯ

9.2.1. ЛИНЕЙНЫЙ СЛУЧАЙ

Теперь обсудим более общий случай нескольких неизвестных величин x_i ($i = 1, 2, \dots, r$). Часто интересующие нас величины непосредственно не измеряются, а вместо этого может быть измерен ряд линейных функций от величин x_i :

$$\eta_j = p_{j0} + p_{j1}x_1 + p_{j2}x_2 + \dots + p_{jr}x_r. \quad (2.1)$$

Мы предпочитаем записывать эти равенства в несколько измененном виде

$$\eta_j + a_{j0} + a_{j1}x_1 + a_{j2}x_2 + \dots + a_{jr}x_r = 0. \quad (2.2)$$

Можно определить вектор-столбец

$$a_j = \begin{pmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jr} \end{pmatrix} \quad (2.3)$$

и записать равенство (2.2) в компактном виде

$$\eta_j + a_{j0} + a_j^T x = 0, \quad j = 1, 2, \dots, n. \quad (2.4)$$

Определим далее

$$\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix}, \quad a_0 = \begin{pmatrix} a_{10} \\ a_{20} \\ \vdots \\ a_{n0} \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ a_{21} & a_{22} & \dots & a_{2r} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nr} \end{pmatrix}. \quad (2.5)$$

Тогда система (2.4) может быть записана таким образом:

$$\eta + a_0 + Ax = 0. \quad (2.6)$$

Мы вновь предполагаем, что каждое измерение содержит погрешность ϵ_j , которая нормально распределена. Фактические измерения мы вновь обозначаем* через y_j :

$$\left. \begin{aligned} y_j &= \eta_j + \epsilon_j, \\ E(\epsilon_j) &= 0, \\ E(\epsilon_j^2) &= \sigma_j^2 = 1/g_j. \end{aligned} \right\} \quad (2.7)$$

Поскольку y_j — *независимые* измерения, можно образовать *диагональную* ковариационную матрицу величин y_j или ϵ_j с σ_j^2 на диагонали:

$$C_y = C_\epsilon = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \dots & \\ 0 & & & \sigma_n^2 \end{pmatrix}. \quad (2.8)$$

По аналогии с равенством (1.7) назовем обратную матрицу весовой матрицей:

$$G_y = G_\epsilon = C_y^{-1} = C_\epsilon^{-1} = \begin{pmatrix} g_1 & & & 0 \\ & g_2 & & \\ & & \dots & \\ 0 & & & g_n \end{pmatrix}. \quad (2.9)$$

* Для простоты мы больше не записываем случайные величины буквами специального типа. Из контекста всегда будет ясно, какие из рассматриваемых величин являются случайными.

Образуя также n -мерные векторы измерений и ошибок, из равенства (2.7) имеем

$$\mathbf{y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad (2.10)$$

затем, используя равенство (2.6), получаем

$$\mathbf{y} - \boldsymbol{\varepsilon} + \mathbf{a}_0 + \mathbf{A}\mathbf{x} = \mathbf{0}. \quad (2.11)$$

Попробуем решить эту систему уравнений относительно \mathbf{x} , используя метод максимального правдоподобия. В соответствии с нашим предположением (2.7) измерения y_j распределены по нормальному закону с плотностью вероятности

$$\begin{aligned} f(y_j) &= \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{(y_j - \eta_j)^2}{2\sigma_j^2}\right] = \\ &= \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{\varepsilon_j^2}{2\sigma_j^2}\right). \end{aligned} \quad (2.12)$$

Для всех n измерений, следовательно, получаем функции правдоподобия

$$L = \prod_{i=1}^n f(y_i) = (2\pi)^{-\frac{1}{2}n} \left(\prod_{i=1}^n \sigma_i^{-1}\right) \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{\varepsilon_i^2}{\sigma_i^2}\right), \quad (2.13)$$

$$l = \ln L = -\frac{1}{2}n \ln 2\pi + \ln\left(\prod_{i=1}^n \sigma_i^{-1}\right) - \frac{1}{2} \sum_{i=1}^n \frac{\varepsilon_i^2}{\sigma_i^2}. \quad (2.14)$$

Ясно, что последнее выражение принимает максимальное значение, когда

$$M = \sum_{i=1}^n \frac{\varepsilon_i^2}{\sigma_i^2} = \sum_{i=1}^n \frac{(y_i + \mathbf{a}_i^T \mathbf{x} + a_{i0})^2}{\sigma_i^2} = \min. \quad (2.15)$$

Используя равенства (2.9) и (2.11), можно переписать это выражение в виде

$$M = \boldsymbol{\varepsilon}^T G_y \boldsymbol{\varepsilon} = \min, \quad (2.16)$$

или

$$M = (\mathbf{y} + \mathbf{a}_0 + \mathbf{A}\mathbf{x})^T G_y (\mathbf{y} + \mathbf{a}_0 + \mathbf{A}\mathbf{x}) = \min. \quad (2.17)$$

Функция M достигает минимума, если все частные производные по x_i обращаются в нуль одновременно, т. е.

$$\frac{\partial M}{\partial x_i} = 0, \quad i = 1, 2, \dots, r. \quad (2.18)$$

Это требование приводит к

$$2\mathbf{A}^T G_y (\mathbf{y} + \mathbf{a}_0 + \mathbf{A}\mathbf{x}) = \mathbf{0}. \quad (2.19)$$

Если $r \leq n$, эту систему можно разрешить (см. приложение в разд. Б3). Это решение имеет вид

$$\tilde{x} = -(A^T G_y A)^{-1} A^T G_y (y + a_0). \quad (2.20)$$

Мы выразили наилучшие оценки \tilde{x} неизвестных величин x через известные коэффициенты A и a_0 , измеренные величины y и их ковариационную матрицу G_y^{-1} . Чтобы согласовать обозначения со следующими разделами, мы вводим сокращение:

$$c = y + a_0. \quad (2.21)$$

Тогда равенство (2.11) можно записать следующим образом:

$$Ax + c - \varepsilon = 0, \quad (2.22)$$

а решение (2.20) принимает вид

$$\tilde{x} = -(A^T G_y A)^{-1} A^T G_y c. \quad (2.23)$$

Такое решение, разумеется, включает специальный случай, рассмотренный в разд. 9.1. В случае прямых неравноточных измерений вектор x имеет только одну компоненту, $a_0 = 0$, а A сводится к n -мерному вектору-столбцу, все элементы которого равны -1 . Можно легко проверить, что равенство

$$\tilde{x} = \left[(1, 1, \dots, 1) \begin{pmatrix} g_1 & & 0 \\ & g_2 & \\ & & \dots \\ 0 & & & g_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \right]^{-1} \otimes \\ \otimes \begin{pmatrix} g_1 & & 0 \\ & g_2 & \\ & & \dots \\ 0 & & & g_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

в действительности эквивалентно равенству (1.9). Если, более того, все диагональные элементы матрицы G_y одинаковы, результат превращается в (1.4).

Теперь займемся изучением влияния ошибок измерений на неизвестные параметры x . Поскольку равенство (2.20) задает линейную зависимость между \tilde{x} и измерениями y , мы можем использовать закон распространения ошибок, найденный в разд. 4.5. Из равенств (5.2) и (5.4) гл. 4 немедленно получаем, что

$$G_{\tilde{x}}^{-1} = [(A^T G_y A)^{-1} A^T G_y] G_y^{-1} [(A^T G_y A)^{-1} A^T G_y]^T.$$

Матрицы G_y , G_y^{-1} и $(A^T G_y A)$ симметричны, т. е. каждая из них совпадает со своей транспонированной. Поэтому, используя пра-

вило (2.14) из приложения Б, мы можем упростить предыдущее выражение:

$$\begin{aligned} G_x^{-1} &= (A^T G_y A)^{-1} A^T G_y G_y^{-1} G_y A (A^T G_y A)^{-1} = \\ &= (A^T G_y A)^{-1} (A^T G_y A) (A^T G_y A)^{-1}, \\ G_x^{-1} &= (A^T G_y A)^{-1}. \end{aligned} \quad (2.24)$$

Мы получили простую форму ковариационной матрицы наилучших выборочных оценок \tilde{x} для неизвестных x . Квадратные корни из диагональных элементов можно рассматривать как «ошибки измерения» \tilde{x} , хотя эти величины не измеряются непосредственно.

Кроме того, используя формулу (2.23), теперь можно уточнить первоначальные измерения y . Подставляя (2.23) в (2.22), определяем вектор наилучших оценок погрешностей измерений $\tilde{\varepsilon}$:

$$\tilde{\varepsilon} = A\tilde{x} + c = -A(A^T G_y A)^{-1} A^T G_y c + c. \quad (2.25)$$

Эти оценки могут быть использованы для исправления исходных измерений y и для вычисления вектора уточненных измерений

$$\begin{aligned} \tilde{\eta} &= y - \tilde{\varepsilon} = y + A(A^T G_y A)^{-1} A^T G_y c - c, \\ \tilde{\eta} &= A(A^T G_y A)^{-1} A^T G_y c - a_0. \end{aligned} \quad (2.26)$$

Это выражение снова линейно относительно y . Поэтому мы можем использовать закон распространения ошибок для того, чтобы определить матрицу ковариации уточненных измерений

$$\begin{aligned} G_{\tilde{\eta}}^{-1} &= [A(A^T G_y A)^{-1} A^T G_y] G_y^{-1} [A(A^T G_y A)^{-1} A^T G_y]^T, \\ G_{\tilde{\eta}}^{-1} &= A(A^T G_y A)^{-1} A^T = A G_x^{-1} A^T. \end{aligned} \quad (2.27)$$

Уточненные измерения $\tilde{\eta}$ полностью удовлетворяют равенствам (2.1), если неизвестные заменить на решение \tilde{x} , полученное методом наименьших квадратов.

В качестве первого примера рассмотрим определение «наилучшей» прямой линии, проходящей через несколько замеренных точек. Такого рода задача построения *эмпирической кривой* по точкам типична для анализа наблюдений.

Пример 9.2. Рассмотрим тот случай, когда измеренные величины являются функциями переменной t . Значения t_j независимой переменной t известны точно, т. е. с нулевыми ошибками. Пусть имеется линейная зависимость

$$\eta_j = y_j - \varepsilon_j = x_1 + x_2 t_j, \text{ или } \eta - x_1 - x_2 t = 0. \quad (2.28)$$

Мы хотим определить неизвестные

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

Таблица 9.2

Данные примера 9.2

j	1	2	3	4
t_j	0,0	1,0	2,0	3,0
y_j	1,4	1,5	3,7	4,1
σ_j	0,5	0,5	0,5	0,5

исходя из 4 измерений табл. 9.2. Ошибка измерения y равна 0,5, т. е. имеем

$$C_y = \begin{pmatrix} 0,5^2 & & 0 \\ & 0,5^2 & \\ 0 & & 0,5^2 \end{pmatrix} = 0,25I, \quad G_y = C_y^{-1} = 4I.$$

Мы должны построить для нашей задачи матрицу A и вектор коэффициентов \mathbf{a}_0 . Из равенства (2.28) имеем

$$\eta_j - x_1 - x_2 t_j = 0.$$

Сравнивая с (2.2), получаем

$$\eta + A\mathbf{x} = 0,$$

т. е. $\mathbf{a}_0 = 0$ и

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}.$$

Поскольку

$$\mathbf{y} = \begin{pmatrix} 1,4 \\ 1,5 \\ 3,7 \\ 4,1 \end{pmatrix},$$

уравнение (2.16) приводит к решению

$$\mathbf{x} = \left[\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1,4 \\ 1,5 \\ 3,7 \\ 4,1 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 6 & 14 \end{pmatrix}^{-1} \begin{pmatrix} 10,7 \\ 21,2 \end{pmatrix}.$$

Используя равенство (3.12) из приложения Б для того, чтобы обратить квадратную матрицу 2-го порядка, получаем

$$\tilde{x} = \frac{1}{20} \begin{pmatrix} 14 & -6 \\ -6 & 4 \end{pmatrix} \begin{pmatrix} 10,7 \\ 21,2 \end{pmatrix} = \begin{pmatrix} 1,13 \\ 1,03 \end{pmatrix},$$

или подробнее

$$\tilde{x}_1 = 1,13, \quad \tilde{x}_2 = 1,03.$$

Теперь применим (2.24), чтобы получить матрицу ковариации выборочных оценок \tilde{x} :

$$\begin{aligned} G_{\tilde{x}}^{-1} = C_{\tilde{x}} &= 0,25 \left[\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \right]^{-1} = \\ &= \begin{pmatrix} 0,175 & -0,075 \\ -0,075 & 0,050 \end{pmatrix}. \end{aligned}$$

Используя равенство (2.28), можно просто вычислить уточненные измерения

$$\tilde{\eta} = -A\tilde{x} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 1,13 \\ 1,03 \end{pmatrix} = \begin{pmatrix} 1,13 \\ 2,16 \\ 3,19 \\ 4,22 \end{pmatrix}.$$

Они лежат на прямой, записываемой уравнением $\tilde{\eta} = -A\tilde{x}$, которое, разумеется, в общем случае будет отличаться от «истинного» решения. *Остаточные ошибки* величин $\tilde{\eta}$, полученные из равенства (2.19), равны

$$\begin{aligned} G_{\tilde{\eta}}^{-1} = C_{\tilde{\eta}} &= \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 0,175 & -0,075 \\ -0,075 & 0,05 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} = \\ &= \begin{pmatrix} 0,175 & -0,075 \\ 0,1 & -0,025 \\ 0,025 & -0,025 \\ -0,5 & -0,175 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} = \\ &= \begin{pmatrix} 0,175 & 0,1 & 0,025 & -0,05 \\ 0,1 & 0,075 & 0,05 & 0,025 \\ 0,025 & 0,05 & 0,075 & 0,1 \\ -0,05 & 0,025 & 0,01 & 0,175 \end{pmatrix}. \end{aligned}$$

Квадратные корни из диагональных элементов равны

$$\begin{aligned} \Delta\tilde{\eta}_1 &= 0,42, & \Delta\tilde{\eta}_2 &= 0,27, & \Delta\tilde{\eta}_3 &= 0,27, \\ \Delta\tilde{\eta}_4 &= 0,42. \end{aligned}$$

Процедура подгонки, которая использует большее число измерений (4), чем это необходимо для определения неизвестных (2), дает существенное уменьшение индивидуальных ошибок измерений,

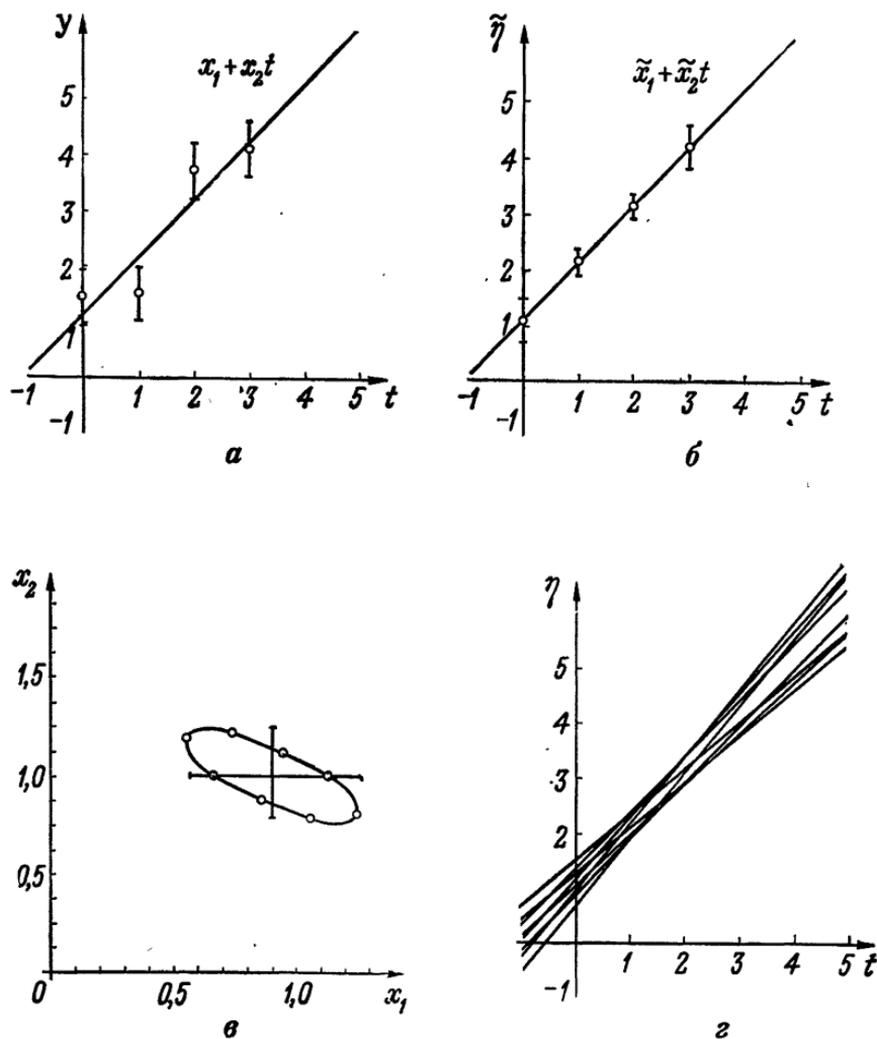


Рис. 9.2. Подгонка прямой линии к данным примера 9.2. а — исходные измерения и стандартные отклонения; б — уточненные измерения и остаточные ошибки; в — ковариационный эллипс подогнанных величин x_1 и x_2 ; г — прямые, соответствующие нескольким точкам на ковариационном эллипсе.

которые были равны 0,5 для всех точек. Результаты этой процедуры представлены на рис. 9.2. Измерения y_j изображены как функционально зависящие от переменной t . Вертикальными чер-

точками обозначены стандартные отклонения. Ими отмечена область $y_j \pm \sigma(y_j)$. Прямая линия соответствует значениям \tilde{x}_1, \tilde{x}_2 . Уточненные измерения лежат на этой линии. Они изображены на рис. 9.2, б вместе с остаточными ошибками $\Delta\eta_j$. Чтобы проиллюстрировать ошибки величин \tilde{x}_1, \tilde{x}_2 , рассмотрим ковариационную матрицу $C_{\tilde{x}}$. Она определяет эллипс ковариации (разд. 5.10) в плоскости \tilde{x}_1, \tilde{x}_2 . Эллипс вычерчен на рис. 9.2, в. Точки на эллипсе соответствуют равной вероятности. Каждая из них определяет некоторую линию на t, y -плоскости. Несколько точек отмечено кружочками на рис. 9.2, в, соответствующие им линии начерчены на рис. 9.2, г. Таким образом, точки из ковариационного эллипса определяют пучок прямых линий. Линии, соответствующие истинным значениям неизвестных, лежат внутри этого пучка с вероятностью $1 - e^{-1/2}$ [см. равенство (10.17) гл. 5].

9.2.2. НЕЛИНЕЙНЫЙ СЛУЧАЙ

До сих пор мы предполагали, что соотношения между измеряемыми величинами и неизвестными параметрами имеют вид (2.2), т. е. линейны. Однако часто эти соотношения нелинейны. В общем случае имеем

$$f_j(\mathbf{x}, \eta) = \eta_j - g_j(x_1, x_2, \dots, x_r) = 0, \quad j = 1, 2, \dots, n, \quad (2.29)$$

или

$$f(\mathbf{x}, \eta) = 0. \quad (2.30)$$

Теперь мы можем свести общий случай к линейному, производя разложение f_j в ряд Тейлора и рассматривая только линейные члены этого разложения. Произведем разложение в «точке» $\mathbf{x}_0 = (x_{10}, x_{20}, \dots, x_{r0})$, которая описывает первое приближение неизвестных, полученное каким-либо образом:

$$f_j(\mathbf{x}, \eta) = f_j(\mathbf{x}_0, \eta) + \left(\frac{\partial f_j}{\partial x_1}\right)_{\mathbf{x}_0} (x_1 - x_{10}) + \dots + \left(\frac{\partial f_j}{\partial x_r}\right)_{\mathbf{x}_0} (x_r - x_{r0}) + \dots \quad (2.31)$$

Вводя определения

$$\xi = \mathbf{x} - \mathbf{x}_0 = \begin{pmatrix} x_1 - x_{10} \\ x_2 - x_{20} \\ \vdots \\ x_r - x_{r0} \end{pmatrix}, \quad (2.32)$$

$$a_{jl} = \left(\frac{\partial f_j}{\partial x_l}\right)_{\mathbf{x}_0} = - \left(\frac{\partial g_j}{\partial x_l}\right)_{\mathbf{x}_0},$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nr} \end{pmatrix}, \quad (2.33)$$

$$c_j = f_j(\mathbf{x}_0, \mathbf{y}) = y_j - g_j(\mathbf{x}_0), \quad \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \quad (2.34)$$

и используя соотношения (2.10), получаем

$$f_j(\mathbf{x}_0, \boldsymbol{\eta}) = f_j(\mathbf{x}_0, \mathbf{y} - \boldsymbol{\varepsilon}) = f_j(\mathbf{x}_0, \mathbf{y}) - \boldsymbol{\varepsilon}. \quad (2.35)$$

Тогда эти соотношения дают нам возможность представить систему (2.31) в виде

$$\mathbf{f} = A\bar{\boldsymbol{\xi}} + \mathbf{c} - \boldsymbol{\varepsilon}. \quad (2.36)$$

Это матричное уравнение полностью аналогично уравнению (2.22). Решение, найденное методом наименьших квадратов для поправок $\bar{\boldsymbol{\xi}}$, является, следовательно, таким:

$$\bar{\boldsymbol{\xi}} = -(A^T G_y A)^{-1} A^T G_y \mathbf{c}. \quad (2.37)$$

Соответствующая ковариационная матрица имеет вид [см. равенство (2.27)]

$$G_{\bar{\boldsymbol{\xi}}}^{-1} = C_{\bar{\boldsymbol{\xi}}} = (A^T G_y A)^{-1}. \quad (2.38)$$

Как и в линейном случае, мы можем получить вектор ошибок измерения $\boldsymbol{\varepsilon}$, вектор уточненных измерений $\boldsymbol{\eta}$ и их ковариационную матрицу $G_{\boldsymbol{\eta}}^{-1}$. Обозначения выбраны таким образом, чтобы результат сразу задавался равенствами (2.25)—(2.27).

Поскольку \mathbf{x}_0 фиксировано, ковариационная матрица $\mathbf{x}_1 = \mathbf{x}_0 + \bar{\boldsymbol{\xi}}$ совпадает с $G_{\bar{\boldsymbol{\xi}}}^{-1}$:

$$G_{\mathbf{x}_1}^{-1} = G_{\bar{\boldsymbol{\xi}}}^{-1} = (A^T G_y A)^{-1}. \quad (2.39)$$

Теперь мы можем заменить \mathbf{x}_0 на \mathbf{x}_1 и повторить процедуру, чтобы получить лучшее приближение неизвестных. Этот итерационный процесс может быть повторен до тех пор, пока поправки $\bar{\boldsymbol{\xi}}$ не станут так малы, что уточнения, получающиеся в результате очередного шага, будут пренебрежимы. Предположим, мы произ-

вели с итерационных шагов и затем рассмотрели окончательные результаты, которые имеют вид

$$\begin{aligned}\tilde{x} &= x_s, \\ G_{\tilde{x}}^{-1} &= G_{\xi}^{-1} = (A^T G_y A)^{-1}, \\ \tilde{\eta} &= y + A (A^T G_y A)^{-1} A^T G_y c,\end{aligned}$$

где элементы A и c вычислены с использованием промежуточного результата x_{s-1} .

Мы не показали, что эта процедура приводит к единственному решению x . Это действительно не может быть доказано в общем случае. В каждом случае должна быть проведена проверка сходимости к нулю последовательности разностей между решениями, получающимися на соседних шагах. Очевидно, функции f_j не должны сильно отклоняться от линейных в некоторой области, включающей точки x_0 и \tilde{x} . Другими словами, если f_j отличаются от линейных лишь немного, то x_0 (начальное значение) может достаточно сильно отличаться от конечного решения; если f_j сильно отличаются от линейных функций, процедура будет сходиться, только если x_0 уже представляет собой очень хорошее приближение. Фактически искусство использования метода наименьших квадратов в нелинейных задачах состоит в умении найти достаточно хорошее приближение. Нет общего правила, по которому его можно было бы отыскать, так как ясно, что это целиком зависит от природы задачи.

Пример 9.3. Величина η измерена как функция переменной t , которая предполагается свободной от погрешностей, как и в примере 9.2. Пусть предполагаемая зависимость η от t имеет вид

$$\eta = a \sin(2\pi \nu t),$$

т. е. мы предполагаем, что это колебание с амплитудой a и частотой ν . Запишем данное уравнение в виде (2.29):

$$f_j = \eta_j - x_1 \sin(2\pi x_2 t_j), \quad j = 1, 2, \dots, n.$$

В соответствии с (2.33) и (2.34) получаем

$$A = \begin{pmatrix} -\sin(2\pi x_{20} t_1) & -2\pi t_1 x_{10} \cos(2\pi x_{20} t_1) \\ -\sin(2\pi x_{20} t_2) & -2\pi t_2 x_{10} \cos(2\pi x_{20} t_2) \\ \vdots & \vdots \\ -\sin(2\pi x_{20} t_n) & -2\pi t_n x_{10} \cos(2\pi x_{20} t_n) \end{pmatrix}$$

и

$$c = \begin{pmatrix} y_1 - x_{10} \sin(2\pi x_{20} t_1) \\ y_2 - x_{10} \sin(2\pi x_{20} t_2) \\ \vdots \\ y_n - x_{10} \sin(2\pi x_{20} t_n) \end{pmatrix}.$$

Пусть заданы пять измерений табл. 9.3. Из предположения, что относительная ошибка измерения составляет 10%, получены σ_j , т. е. $\sigma_j = y_j/10$. Измерения изображены на рис. 9.3, а.

Таблица 9.3
Данные примера 9.3

j	1	2	3	4	5
t_j	0,1	0,2	0,3	0,4	0,5
y_j	1,2	1,9	2,0	2,0	1,4
σ_j	0,12	0,19	0,20	0,20	0,14

Как обычно, вертикальными черточками обозначаются ошибки измерений, т. е. они отмечают область $y_j \pm \sigma_j$. Вычертив грубо

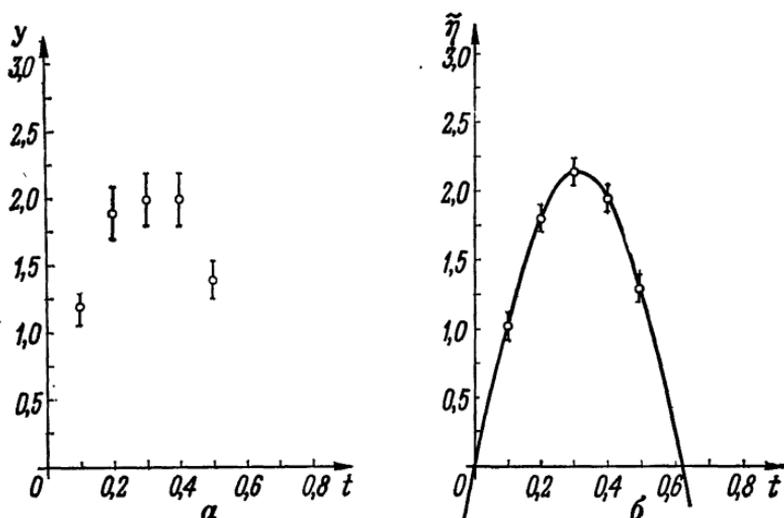


Рис. 9.3. Подгонка синусоиды к данным примера 9.3. а — исходные измерения и стандартные отклонения; б — подогнанная кривая вместе с уточненными измерениями и остаточными ошибками.

от руки синусоиду, проходящую через измеренные точки, получим в качестве первого приближения

$$a_0 = x_{10} = 2, \quad \gamma_0 = x_{20} = 1.$$

Поскольку достаточно обременительно вновь вычислять A и c и осуществлять операции над матрицами (2.37), мы для этого

используем следующую вычислительную программу:

```

C
C   SINE FIT USING      LEAST SQUARES
C
  DIMENSION T(5),Y(5),GY(25),CY(25),X(2),XI(2),CX(4)
  DIMENSION EPSLN(5),ETA(5),CETA(25),A(10),C(5)
  DIMENSION S1(25),S2(25)

C
C   DEFINE NUMBER OF MEASUREMENTS (N) AND UNKNOWNNS (NR)
C
  N=5
  NR=2

C
C   INTRODUCE DATA
C
  DATA T,Y /0.1,0.2,0.3,0.4,0.5,1.2,1.9,2.0,2.0,1.4/
  NN=N**2
  DO 10 I=1,NN
10  CY(I)=0.
  DO 20 I=1,N
  II=(I-1)*N+I
20  CY(II)=(Y(I)/10.)***2
  CALL MTXINV (CY,GY,N)

C
C   FIRST APPROXIMATIONS OF UNKNOWNNS
C
  X(1)=2.
  X(2)=1.

C
C   PERFORM 3 STEPS OF ITERATION
C
  DO 40 NSTEP=1,3
C
C   CALCULATE MATRIX OF DERIVATIVES A AND VECTOR C
C
  DO 30 I=1,N
  IK=(I-1)*NR+1
  A(IK)=-SIN(6.28318*X(2)*T(I))
  A(IK+1)=-6.28318*T(I)*X(1)*COS(6.28318*X(2)*T(I))
30  C(I)=Y(I)-X(1)*SIN(6.28318*X(2)*T(I))
  WRITE(6,1000) NSTEP
  CALL MTXWRT(A,N,NR)
  WRITE(6,1100)
  CALL MTXWRT(C,1,N)

C
C   CALCULATE VECTOR XI OF CORRECTIONS (EQS. (2.37) AND (2.24))
C
  CALL MTXMLT (GY,A,S1,N,N,NR)
  CALL MTXMAT (A,S1,S2,NR,N,NR)
  CALL MTXINV (S2,CX,NR)
  CALL MTXMBT (CX,A,S2,NR,NR,N)
  CALL MTXMLT (S2,GY,S1,NR,N,N)
  CALL MTXMLT (S1,C,XI,NR,N,1)
  CALL MTXMSC (XI,XI,-1.,NR,1)

C
C   CALCULATE NEW VALUE FOR X
C
  CALL MTXADD (X,XI,X,NR,1)

C
C   CALCULATE MINIMUM FUNCTION XM
C
  CALL MTXMLT (A,XI,S1,N,NR,1)
  CALL MTXADD (S1,C,EPSLN,N,N,1)
  CALL MTXMAT (EPSLN,GY,S1,1,N,N)
  CALL MTXMLT (S1,EPSLN,XM,1,N,1)

```

```

C
C      WRITE RESULTS OF THIS STEP
C
C      40 WRITE (6,1200) XM,X
C
C      CALCULATE ETA AND COVARIANCE MATRIX CETA (EQS. (2.26) AND (2.27))
C
C      CALL MTXSUB (Y,EPSLN,ETA,N,1)
C      CALL MTXMLT (A,CX,S1,N,NR,NR)
C      CALL MTXMBT (S1,A,CETA,N,NR,N)
C
C      WRITE FINAL RESULTS
C
C      WRITE (6,1300) ETA
C      WRITE (6,1400)
C      CALL MTXWRT (CX,NR,NR)
C      WRITE (6,1500)
C      CALL MTXWRT (CETA,N,N)
C      STOP
C
C      FORMAT STATEMENTS
C
1000 FORMAT(/15H ITERATION STEP,12,10H, MATRIX A/)
1100 FORMAT(/9H VECTOR C/)
1200 FORMAT(/19H MINIMUM FUNCTION =F8.5,14H, UNKNOWNNS X =,2F10.5)
1300 FORMAT(/29H IMPROVED MEASUREMENTS ETA = 5F10.5)
1400 FORMAT(/30H COVARIANCE MATRIX OF UNKNOWNNS/)
1500 FORMAT(/43H COVARIANCE MATRIX OF IMPROVED MEASUREMENTS/)
      END

```

Программа состоит из трех частей. Широко используются оперирующие с матрицей подпрограммы приложения разд. Б.4. На первом этапе задаются число измерений и число неизвестных и вводятся величины t_j и y_j . Далее вычисляется ковариационная матрица $C_y = G_y^{-1}$ (в программе она обозначается через CY). В конце этой части получается первое приближение к неизвестным параметрам.

Во второй части выполняется трехшаговый итерационный процесс. На каждом шаге при вычислении матрицы производных A и вектора c используется текущее приближение x . Для каждого шага эти величины печатаются, при этом используется подпрограмма MTXWRT. Далее, вычисленный вектор поправок (коррекции) прибавляется к текущему приближению x и образуется новое приближение. Величина минимума функции M (обозначенная в программе через XM) вычисляется и печатается вместе со значениями неизвестных на каждом шаге. Используя вспомогательные массивы, обозначенные через S1 и S2, накапливают промежуточные матрицы.

В третьей части программы уточненные измерения $\tilde{\eta}$ и их ковариационная матрица $C_{\tilde{\eta}} = G_{\tilde{\eta}}^{-1}$ (обозначенные через ETA и CETA соответственно) вычисляются и печатаются вместе с ковариационной матрицей $C_{\tilde{x}} = G_{\tilde{x}}^{-1}$ неизвестных параметров. Выходная печать программы воспроизведена в табл. 9.4.

Отметим, что удовлетворительный результат достигается после двух шагов, так как на третьем шаге минимум функции уменьшается только на очень малую величину. Это также можно

Таблица 9.4

Печать машины для примера 9.3

ITERATION STEP 1, MATRIX A

-0.58778	-1.01664
-0.95106	-0.77665
-0.95106	1.16495
-0.58779	4.06655
-0.00000	6.28318

VECTOR C

0.02443	-0.00211	0.09789	0.82443	1.35999
---------	----------	---------	---------	---------

MINIMUM FUNCTION = 4.27797, UNKNOWNNS X = 2.12551 0.79193

ITERATION STEP 2, MATRIX A

-0.47730	-1.17355
-0.83885	-1.45392
-0.99696	-0.31236
-0.91328	2.17602
-0.60810	5.30098

VECTOR C

0.18548	0.11701	-0.11904	0.05882	0.10747
---------	---------	----------	---------	---------

MINIMUM FUNCTION = 2.67794, UNKNOWNNS X = 2.23210 0.79245

ITERATION STEP 3, MATRIX A

-0.47759	-1.23219
-0.83920	-1.52537
-0.99703	-0.32394
-0.91275	2.29179
-0.60681	5.57373

VECTOR C

0.13397	0.02681	-0.22548	-0.03734	0.04553
---------	---------	----------	----------	---------

MINIMUM FUNCTION = 2.67793, UNKNOWNNS X = 2.23211 0.79243

IMPROVED MEASUREMENTS ETA = 1.06601 1.87317 2.22548 2.03740 1.35458

COVARIANCE MATRIX OF UNKNOWNNS

0.01119	0.00083
0.00083	0.00059

COVARIANCE MATRIX OF IMPROVED MEASUREMENTS

0.00443	0.00706	0.00672	0.00323	-0.00241
0.00706	0.01139	0.01115	0.00606	-0.00245
0.00672	0.01115	0.01172	0.00809	0.00124
0.00323	0.00606	0.00809	0.00895	0.00836
-0.00241	-0.00245	0.00124	0.00836	0.01686

Усмотреть из поведения элементов c , которые не сильно меняются на третьем шаге. По определению [равенство (2.34)] эти элементы (невязки) показывают отклонение от точного выполнения уравнений (2.29). Если на некотором шаге эти элементы не могут быть

уменьшены, то не имеет смысла продолжать итерационный процесс. В более общей программе мы, следовательно, могли бы повторять итерации до тех пор, пока величины M на двух последовательных шагах не будут отличаться только на величину меньшую, чем некоторое заранее выбранное малое число. Такая программа и, кроме того, критерии сходимости рассматриваются в разд. 9.5.

Если извлечь квадратные корни из диагональных элементов матрицы C_x , то получим $\tilde{a} = \tilde{x}_1 = 2,23 \pm 0,10$ и $\tilde{v} = \tilde{x}_2 = 0,80 \pm 0,02$. Кривая с этими параметрами показана на рис. 9.3, б. Уточненные измерения $\tilde{\eta}$ лежат на этой кривой. Их ошибки — квадратные корни из диагональных элементов последней матрицы в табл. 9.4 — также изображены на рис. 9.3.

9.2.3. СВОЙСТВА РЕШЕНИЯ, ПОЛУЧЕННОГО МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ; КРИТЕРИЙ χ^2

До сих пор метод наименьших квадратов рассматривался только как применение принципа наибольшего правдоподобия для решения линейных или линеаризуемых задач. Требование минимальности суммы квадратов (2.15) было получено непосредственно в результате максимизации функции правдоподобия (2.13). Для построения функции правдоподобия плотность вероятности наблюдений предполагалась известной. Мы считали, что наблюдения распределены по нормальному закону. Если не предполагать, что наблюдения нормально распределены вокруг истинного значения, то все же можно попытаться использовать соотношение (2.15), а вместе с ним и другие формулы этого раздела, хотя может показаться, что для этого нет достаточных теоретических обоснований. Однако теорема Гаусса—Маркова утверждает, что даже в этом случае метод наименьших квадратов приводит к результатам с хорошими свойствами. Прежде чем приступить к рассмотрению этих результатов, коротко ознакомимся со свойствами решения уравнения максимального правдоподобия.

а) Решение \tilde{x} является несмещенной оценкой, т. е.

$$E(\tilde{x}_i) = x_i, \quad i = 1, 2, \dots, r.$$

б) Оно имеет минимальную дисперсию

$$\sigma^2(\tilde{x}_i) = E\{(\tilde{x}_i - x_i)^2\} = \min.$$

в) Величина (2.16)

$$M = \varepsilon^T G_y \varepsilon$$

подчиняется распределению χ^2 с $(n - r)$ степенями свободы.

Свойства (а) и (б) легко получаются из результатов гл. 7. Покажем справедливость (в) в простом случае прямых наблюдений ($r = 1$), для которого G_y — диагональная матрица:

$$G_y = \begin{pmatrix} 1/\sigma_1^2 & & & 0 \\ & 1/\sigma_2^2 & & \\ & & \ddots & \\ 0 & & & 1/\sigma_n^2 \end{pmatrix}.$$

Величина M тогда превращается просто в сумму квадратов:

$$M = \sum_{j=1}^n \varepsilon_j^2 / \sigma_j^2. \quad (2.40)$$

Поскольку каждое ε_j имеет нормальное распределение с нулевым средним и дисперсией σ_j^2 , величины ε_j / σ_j подчиняются стандартному гауссовскому распределению. В этом случае сумма квадратов (2.40) подчиняется распределению χ^2 с $(n - 1)$ степенью свободы (разд. 6.6).

Если распределение ошибок ε_j неизвестно, то решение, полученное методом наименьших квадратов, обладает следующими свойствами:

- а) Решение является несмещенной оценкой.
- б) Среди всех x^* , которые являются несмещенными оценками x и линейными комбинациями наблюдений y , решение, полученное методом наименьших квадратов, имеет наименьшую дисперсию (теорема Гаусса—Маркова).
- в) Математическое ожидание величины

$$M = \varepsilon^T G_y \varepsilon$$

равно

$$E(M) = n - r$$

(эта величина равна математическому ожиданию χ^2 с $n - r$ степенями свободы).

Величину M часто называют просто χ^2 , хотя она не обязательно подчиняется χ^2 -распределению. Вместе с матрицами C_x и C_y она задает удобную меру качества подгонки наименьшими квадратами. Если значение, полученное для M , много больше $n - r$, то нам придется внимательно пересмотреть предположения, на которых основывался подсчет.

Величина $f = n - r$ называется *числом степеней свободы подгонки*, а r — *числом связей подгонки*. Эти выражения можно понять, обратясь к частному случаю, рассмотренному в разд. 9.3. Является очевидным (см. приложение разд. Б.3), что задача наименьших квадратов может быть решена лишь для $f \geq 0$. Но

только для $f > 0$ величина M имеет смысл и ее можно использовать в качестве критерия качества подгонки.

В тех случаях, когда ошибки распределены нормально, подгонку методом наименьших квадратов можно объединить с критерием χ^2 . Мы не принимаем результата подгонки, если

$$M = \varepsilon^T G_y \varepsilon > \chi^2_{1-\alpha}(n-r), \quad (2.41)$$

т. е. если M превышает критическое значение χ^2 , соответствующее уровню значимости α и $n-r$ степеням свободы. Отказ мог бы быть обусловлен одной из следующих причин (помимо возможности ошибки первого рода);

а) Предполагаемая функциональная зависимость $f(x, \eta) = 0$ между измеряемыми величинами η и неизвестными параметрами x не осуществляется. Либо функция $f(x, \eta)$ выбрана вообще неверно, либо некоторые параметры, которые считались известными, заданы неправильно.

б) Функциональная зависимость $f(x, \eta)$ имеет место, но ряд Тейлора всего с одним членом не представляет ее адекватно в требуемой области.

в) Первое приближение x_0 слишком далеко от истинного решения x . Лучшее значение x_0 может привести к приемлемому значению M . Очевидно, этот пункт тесно связан с п. (б).

г) Ковариационная матрица G_y^{-1} измеренных переменных, которая часто основывается только на грубых оценках или даже на догадках, неверна.

Эти четыре пункта должны быть внимательно рассмотрены, чтобы метод наименьших квадратов мог привести к успеху. Часто вычисления методом наименьших квадратов повторяются много раз с различными множествами исходных измерений. Тогда можно изучить распределение M и сравнить его с χ^2 -распределением с соответствующим числом степеней свободы. Это сравнение может быть полезным для оценки G_y^{-1} . В общем случае, если начат новый эксперимент, то измерительный прибор, который дает значения y_j , проверяется при помощи измерений известных величин. Таким образом, x окажется известным для некоторых множеств исходных данных. Ковариационная матрица тогда может быть подобрана так, чтобы распределение M для измерений с известными результатами совпадало с распределением χ^2 .

9.3 ИЗМЕРЕНИЯ С ОГРАНИЧЕНИЯМИ

Теперь вернемся к случаю, рассмотренному в разд. 9.1, когда интересующие нас величины измерялись непосредственно. Однако теперь мы не будем предполагать, что все n наблюдений не связаны друг с другом. Они могут, например, подчиняться q урав-

нениям связи. В качестве примера рассмотрим три угла треугольника, которые могут быть непосредственно измерены. Уравнение связи состоит в том, что их сумма должна составлять 180° . Мы снова хотим определить наилучшие значения величин $\tilde{\eta}_j$.

Измерения дают величины

$$y_j = \eta_j + \varepsilon_j, \quad j = 1, 2, \dots, n. \quad (3.1)$$

Как и раньше, предполагаем, что погрешности ε_j имеют нормальное распределение с нулевым средним:

$$E(\varepsilon_j) = 0, \quad E(\varepsilon_j^2) = \sigma_j^2.$$

Имеется q уравнений связи вида

$$f_k(\eta) = 0, \quad k = 1, 2, \dots, q. \quad (3.2)$$

Сначала рассмотрим простой случай, когда уравнения связи линейны. Уравнения (3.2) имеют вид

$$\begin{aligned} b_{10} + b_{11}\eta_1 + b_{12}\eta_2 + \dots + b_{1n}\eta_n &= 0, \\ b_{20} + b_{21}\eta_1 + b_{22}\eta_2 + \dots + b_{2n}\eta_n &= 0, \\ \vdots & \\ b_{q0} + b_{q1}\eta_1 + b_{q2}\eta_2 + \dots + b_{qn}\eta_n &= 0, \end{aligned} \quad (3.3)$$

или в матричной записи

$$B\eta + b_0 = 0. \quad (3.4)$$

9.3.1. МЕТОД ЭЛЕМЕНТОВ

Эти q уравнений мы можем использовать, чтобы исключить q из n величин η . Оставшиеся $n - q$ величин α_i ($i = 1, 2, \dots, n - q$) назовем *элементами*. Они могут быть выбраны произвольно среди исходных величин η или могут быть их линейными комбинациями. Тогда мы можем снова выразить полный вектор η как множество линейных комбинаций элементов:

$$\eta_j = f_{j0} + f_{j1}\alpha_1 + f_{j2}\alpha_2 + \dots + f_{j, n-q}\alpha_{n-q}, \quad j = 1, 2, \dots, n. \quad (3.5)$$

или

$$\eta = F\alpha + f_0. \quad (3.6)$$

Уравнение (3.6) имеет тот же вид, что и уравнение (2.2); следовательно, решение должно иметь вид (2.20), т. е.

$$\tilde{\alpha} = (F^T G_y F)^{-1} F^T G_y (y - f_0). \quad (3.7)$$

Уравнения (3.7) задают выборочные оценки элементов α , полученные методом наименьших квадратов. Соответствующая ковариационная матрица равна

$$G_{\alpha}^{-1} = (F^T G_y F)^{-1} \quad (3.8)$$

[см. уравнение (2.24)]. Уточненные измерения получаются в результате подстановки (3.7) в (3.6):

$$\tilde{\eta} = F\tilde{\alpha} + f_0 = F (F^T G_y F)^{-1} F^T G_y (y - f_0) + f_0. \quad (3.9)$$

Используя закон распространения ошибок, получаем ковариационную матрицу

$$G_{\tilde{\eta}}^{-1} = F (F^T G_y F)^{-1} F^T = F G_{\alpha}^{-1} F^T. \quad (3.10)$$

Пример 9.4. Пусть заданы следующие результаты измерения углов треугольника: $y_1 = 89^\circ$, $y_2 = 31^\circ$, $y_3 = 62^\circ$, т. е.

$$y = \begin{pmatrix} 89 \\ 31 \\ 62 \end{pmatrix}.$$

Линейное уравнение связи имеет вид

$$\eta_1 + \eta_2 + \eta_3 = 180.$$

Оно может быть записано так:

$$B\eta + b_0 = 0,$$

где

$$B = (1, 1, 1), \quad b_0 = b_0 = -180.$$

В качестве элементов выбираем η_1 и η_2 . Система (3.5) получается такой:

$$\eta_1 = \alpha_1, \quad \eta_2 = \alpha_2, \quad \eta_3 = 180 - \alpha_1 - \alpha_2,$$

или

$$\eta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \alpha + \begin{pmatrix} 0 \\ 0 \\ 180 \end{pmatrix},$$

т. е.

$$F = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}, \quad f_0 = \begin{pmatrix} 0 \\ 0 \\ 180 \end{pmatrix}.$$

Допустим, что при измерении углов ошибка равна 1° , т. е.

$$C_y = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I, \quad G_y = C_y^{-1} = I.$$

Применяя (3.7), получаем

$$\begin{aligned}\tilde{\alpha} &= \left[\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} I \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} I \begin{pmatrix} 89 \\ 31 \\ -118 \end{pmatrix} = \\ &= \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 207 \\ 149 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 207 \\ 149 \end{pmatrix} = \\ &= \begin{pmatrix} 88\frac{1}{3} \\ 30\frac{1}{3} \end{pmatrix}.\end{aligned}$$

Используя уравнение (3.9), имеем

$$\tilde{\eta} = F\tilde{\alpha} + f_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} 88\frac{1}{3} \\ 30\frac{1}{3} \end{pmatrix} = \begin{pmatrix} 88\frac{1}{3} \\ 30\frac{1}{3} \\ 61\frac{1}{3} \end{pmatrix}.$$

Разумеется, этого результата надо было ожидать. Излишек в 2° , полученный при измерении углов, был устранен, при этом из всех измерений вычли поровну. Однако если допустить разную дисперсию для углов, то получится другой результат. Советуем читателю провести вычисления в этом случае.

Применяя уравнение (3.10), мы можем определить остаточные ошибки уточненных измерений

$$\begin{aligned}G_{\tilde{\eta}}^{-1} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \left[\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} I \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \right]^{-1} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} = \\ &= \frac{1}{3} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} = \\ &= \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.\end{aligned}$$

Остаточная ошибка для каждого угла равна $\sqrt{2/3} \approx 0,82$.

Здесь мы хотим сделать общее утверждение относительно измерений, которые подчиняются нескольким уравнениям связи. Хотя до сих пор статистический метод не помог нам исправить или даже обнаружить *систематические ошибки*, уравнения связи дают нам возможность с их помощью сделать это. Если, например,

в большом числе измерений сумма углов превышает 180° намного чаще, чем допустимо, то это подтвердит наше подозрение, что имеет место постоянная неисправность в измерительном приборе.

9.3.2. МЕТОД МНОЖИТЕЛЕЙ ЛАГРАНЖА

Вместо того чтобы выполнять вычисления с помощью элементов, мы можем использовать метод множителей Лагранжа. Хотя, разумеется, оба метода приводят к одинаковым результатам, последний имеет то преимущество, что обрабатывает все неизвестные единообразно, тем самым избавляя нас от выбора элементов. Метод множителей Лагранжа часто используют в дифференциальном исчислении.

Мы снова начнем с системы линейных уравнений связи (3.4)

$$B\eta + b_0 = 0$$

и заметим, что измеренные величины являются суммой истинных значений η и погрешностей ε :

$$y = \eta + \varepsilon.$$

Поэтому

$$By - B\varepsilon + b_0 = 0. \quad (3.11)$$

Так как y известны из измерений, а B и b_0 также состоят из известных коэффициентов, мы можем образовать вектор c , состоящий из q компонент,

$$c = By + b_0, \quad (3.12)$$

который не содержит неизвестных. Тогда уравнение (3.11) можно записать в виде

$$c - B\varepsilon = 0. \quad (3.13)$$

Теперь мы вводим другой q -компонентный вектор-столбец, элементами которого являются пока неизвестные *множители Лагранжа*

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{pmatrix}, \quad (3.14)$$

и вместо первоначальной минимизируемой функции

$$M = \varepsilon^T G_y \varepsilon$$

рассмотрим функцию

$$L = \varepsilon^T G_y \varepsilon + 2\mu^T (c - B\varepsilon). \quad (3.15)$$

Функция L называется *функцией Лагранжа*. Требование

$$M = \min$$

с учетом связей

$$c - B\varepsilon = 0$$

выполняется, если полный дифференциал функции Лагранжа обращается в нуль, т. е. если

$$dL = 2\varepsilon^T G_y d\varepsilon - 2\mu^T B d\varepsilon = 0.$$

Это эквивалентно уравнениям

$$\varepsilon^T G_y - \mu^T B = 0. \quad (3.16)$$

Система (3.16) состоит из n уравнений, которые содержат уже $n + q$ неизвестных $\varepsilon_1, \dots, \varepsilon_n$ и μ_1, \dots, μ_q . Однако в дополнение мы имеем q уравнений связи. Транспонируя (3.16), получаем

$$G_y \varepsilon = B^T \mu \quad \text{или} \quad \varepsilon = G_y^{-1} B^T \mu. \quad (3.17)$$

Подставляя это в (3.13), получаем

$$c - B G_y^{-1} B^T \mu = 0.$$

Это равенство легко разрешить относительно μ :

$$\tilde{\mu} = (B G_y^{-1} B^T)^{-1} c. \quad (3.18)$$

Используя (3.17), мы имеем оценки метода наименьших квадратов для отклонений

$$\tilde{\varepsilon} = G_y^{-1} B^T (B G_y^{-1} B^T)^{-1} c. \quad (3.19)$$

Наилучшие оценки для неизвестных η задаются (3.1):

$$\tilde{\eta} = y - \tilde{\varepsilon} = y - G_y^{-1} B^T (B G_y^{-1} B^T)^{-1} c.$$

Введение сокращения

$$G_B = (B G_y^{-1} B^T)^{-1}$$

приводит нас к равенству

$$\tilde{\eta} = y - G_y^{-1} B^T G_B c. \quad (3.20)$$

Ковариационные матрицы для μ и η легко получаются, если применить закон распространения ошибок к линейным системам (3.18) и (3.20):

$$\begin{aligned} G_{\tilde{\mu}}^{-1} &= (B G_y^{-1} B^T)^{-1} = G_B, \\ G_{\tilde{\eta}}^{-1} &= G_y^{-1} - G_y^{-1} B^T G_B B G_y^{-1}. \end{aligned} \quad (3.21)$$

Пример 9.5. Мы применим метод множителей Лагранжа к задаче примера 9.4. Имеем

$$c = By + b_0 = (1, 1, 1) \begin{pmatrix} 89 \\ 31 \\ 62 \end{pmatrix} - 180 = 182 - 180 = 2.$$

Кроме того,

$$G_B = (BG_y B^T)^{-1} = \left[(1, 1, 1) I \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right]^{-1} = 3^{-1} = \frac{1}{3}$$

и

$$G_y B^T = I \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Теперь мы можем вычислить (3.20):

$$\tilde{\eta} = \begin{pmatrix} 89 \\ 31 \\ 62 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \frac{2}{3} = \begin{pmatrix} 88\frac{1}{3} \\ 30\frac{1}{3} \\ 61\frac{1}{3} \end{pmatrix}.$$

Ковариационные матрицы равны

$$G_{\mu}^{-1} = \frac{1}{3},$$

$$\begin{aligned} G_{\eta}^{-1} &= I - I \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \frac{1}{3} (1, 1, 1) I = \\ &= I - \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}. \end{aligned}$$

Теперь мы обобщим метод множителей Лагранжа на нелинейные уравнения связи общего вида (3.2), т. е.

$$f_k(\eta) = 0, \quad k = 1, 2, \dots, q.$$

Мы можем разложить эти уравнения в ряды Тейлора в окрестности значения η_0 , которое является первым приближением к истинной величине η :

$$f_k(\eta) = f_k(\eta_0) + \left(\frac{\partial f_k}{\partial \eta_1} \right)_{\eta_0} (\eta_1 - \eta_{10}) + \dots + \left(\frac{\partial f_k}{\partial \eta_n} \right)_{\eta_0} (\eta_n - \eta_{n0}). \quad (3.22)$$

Введя

$$b_{kl} = \left(\frac{\partial f_k}{\partial \eta_l} \right)_{\eta_0}, \quad B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{q1} & b_{q2} & \dots & b_{qn} \end{pmatrix},$$

$$c_k = f_k(\eta_0), \quad c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_q \end{pmatrix},$$

$$\delta_k = (\eta_k - \eta_{k0}), \quad \delta = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{pmatrix},$$

мы можем записать (3.22) так:

$$B\delta + c = 0. \quad (3.23)$$

Нетрудно выбрать первое приближение η_0 , так как y_i являются измерениями η_j . Поэтому мы можем использовать измерения непосредственно в качестве первого приближения и полагаем

$$\eta_0 = y. \quad (3.24)$$

Вектор δ тогда становится таким:

$$\delta = \eta - \eta_0 = \eta - y = -\varepsilon. \quad (3.25)$$

Следовательно, уравнение (3.23) принимает вид

$$c - B\varepsilon = 0,$$

т. е. совпадает с (3.13). Решение $\tilde{\varepsilon}$ задается равенством (3.19). Так как в нелинейном случае за один шаг обычно нельзя получить удовлетворительного результата, приходится использовать итерации. На втором шаге простое соотношение $\delta = -\varepsilon$ уже не выполняется. Мы решаем (3.23) таким же образом, как и в линейном случае, требуя, чтобы выражение

$$M' = \delta^T G_y \delta \quad (3.26)$$

достигало минимума при выполнении уравнений связи (3.23). Это требование эквивалентно тому, чтобы взвешенная сумма квадратов разностей между двумя приближениями для η была бы минимальна. Тогда решение задается по аналогии с (3.19) в виде

$$\tilde{\varepsilon} = G_y^{-1} B^T (B G_y^{-1} B^T)^{-1} c. \quad (3.27)$$

Результат каждого шага используется для получения лучшего приближения η . Если итерации прерываются на s -м шаге, мы имеем

$$\tilde{\eta} = \eta_s = \eta_{s-1} + \tilde{\varepsilon}. \quad (3.28)$$

Соответствующая ковариационная матрица будет иметь вид (3.21), т. е.

$$G_{\tilde{\eta}}^{-1} = G_y^{-1} - G_y B^T G_B B G_y^{-1}, \quad (3.29)$$

если элементы B заменить значениями, вычисленными на последнем шаге. Выборочные оценки метода наименьших квадратов для погрешностей измерений могут быть теперь вычислены из равенства

$$\tilde{\varepsilon} = y - \tilde{\eta}. \quad (3.30)$$

Отсюда, наконец, может быть получена первоначальная минимизируемая функция

$$\tilde{M} = \tilde{\varepsilon}^T G_y \tilde{\varepsilon}.$$

Эту величину снова можно проверить критерием χ^2 с q степенями свободы.

9.4. ОБЩИЙ СЛУЧАЙ ПОДГОНКИ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ

Опираясь на предшествующие рассуждения, мы совсем просто можем обобщить метод на случай не прямых измерений с уравнениями связи.

Сначала напомним наши обозначения. Имеется r неизвестных параметров, образующих вектор x . Измеряемые величины образуют n -мерный вектор η . Результаты полученных измерений y отклоняются от η , причем погрешности образуют вектор ε . Предполагается, что каждая из величин ε_j ($j = 1, 2, \dots, n$) распределена нормально. Между x и η имеется m функциональных соотношений

$$f_k(x, \eta) = f_k(x, y + \varepsilon) = 0, \quad k = 1, 2, \dots, m. \quad (4.1)$$

Кроме того, предполагаем, что у нас уже имеется первое приближение x_0 к неизвестному вектору x . В качестве первого приближения к вектору η берем $\eta_0 = y$, как в разд. 9.3. Наконец, мы требуем, чтобы функции f_k могли быть достаточно точно приближены линейными функциями от разностей $x - x_0$ и $\eta -$

— η_0 в некоторой области, содержащей векторы (x_0, η_0) , т. е. эти функции должны определяться. Тогда можно записать

$$\begin{aligned} f_k(x, \eta) = & f_k(x_0, \eta_0) + \\ & + \left(\frac{\partial f_k}{\partial x_1} \right)_{x_0, \eta_0} (x_1 - x_{10}) + \dots + \left(\frac{\partial f_k}{\partial x_r} \right)_{x_0, \eta_0} (x_r - x_{r0}) + \\ & + \left(\frac{\partial f_k}{\partial \eta_1} \right)_{x_0, \eta_0} (\eta_1 - \eta_{10}) + \dots + \left(\frac{\partial f_k}{\partial \eta_n} \right)_{x_0, \eta_0} (\eta_n - \eta_{n0}). \end{aligned} \quad (4.2)$$

Введя сокращения

$$a_{kl} = \left(\frac{\partial f_k}{\partial x_l} \right)_{x_0, \eta_0}, \quad A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ a_{21} & a_{22} & \dots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mr} \end{pmatrix}, \quad (4.3)$$

$$b_{kl} = \left(\frac{\partial f_k}{\partial \eta_l} \right)_{x_0, \eta_0}, \quad B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix}, \quad (4.4)$$

$$c_k = f_k(x_0, \eta_0), \quad c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}. \quad (4.5)$$

$$\xi = x - x_0, \quad \delta = \eta - \eta_0, \quad (4.6)$$

систему (4.2) можно записать в виде

$$A\xi + B\delta + c = 0. \quad (4.7)$$

Функция Лагранжа равна

$$L = \delta^T G_y \delta + 2\mu^T (A\xi + B\delta + c), \quad (4.8)$$

где μ — m -мерный вектор множителей Лагранжа. Приравнявая полный дифференциал выражения (4.7) по δ нулю, получаем

$$G_y \delta + B^T \mu = 0$$

или

$$\delta = -G_y^{-1} B^T \mu. \quad (4.9)$$

Подстановка в (4.7) дает

$$A\xi - B G_y^{-1} B^T \mu + c = 0, \quad (4.10)$$

или

$$\mu = G_B (A\xi + c), \quad (4.11)$$

где, по определению,

$$G_B = (B G_y^{-1} B^T)^{-1}. \quad (4.12)$$

Благодаря (4.9) мы теперь можем написать

$$\delta = -G_y^{-1} B^T G_B (A\bar{\xi} + c). \quad (4.13)$$

Поскольку минимум функции Лагранжа ищется не только по δ , но и по $\bar{\xi}$, имеем

$$\frac{\partial L}{\partial \bar{\xi}} = 2\mu^T A = 0.$$

Транспонируя это выражение и подставляя (4.11), получаем

$$2A^T G_B (A\bar{\xi} + c) = 0,$$

или

$$\bar{\xi} = -(A^T G_B A)^{-1} A^T G_B c. \quad (4.14)$$

Подстановка (4.14) в выражения (4.13) и (4.11) дает выборочные оценки метода наименьших квадратов для отклонений δ и множителей Лагранжа μ :

$$\bar{\delta} = -G_y^{-1} B^T G_B (c - A (A^T G_B A)^{-1} A^T G_B c), \quad (4.15)$$

$$\bar{\mu} = G_B A (c - A (A^T G_B A)^{-1} A^T G_B c). \quad (4.16)$$

Выборочные оценки для параметров \tilde{x} и для уточненных измерений $\tilde{\eta}$ получаются следующие:

$$\tilde{x} = x_0 + \bar{\delta}, \quad (4.17)$$

$$\tilde{\eta} = \eta_0 + \bar{\delta}. \quad (4.18)$$

Из (4.14), (4.4) и (4.5) получаем

$$\frac{\partial \bar{\xi}}{\partial y} = -(A^T G_B A)^{-1} A^T G_B \frac{\partial c}{\partial y} = -(A^T G_B A)^{-1} A^T G_B B.$$

Используя закон распространения ошибок, получаем ковариационную матрицу

$$G_{\bar{\xi}}^{-1} = G_{\xi}^{-1} = (A^T G_B A)^{-1}. \quad (4.19)$$

Соответственно находим

$$G_{\eta}^{-1} = G_y^{-1} - G_y^{-1} B^T G_B B G_y^{-1} + G_y^{-1} B^T G_B A (A^T G_B A)^{-1} A^T G_B B G_y^{-1}. \quad (4.20)$$

Можно показать, что при условиях, которые мы всегда предполагаем выполненными, а именно при условии линейизуемости (4.1) и при условии нормального распределения ошибок, минимизируемая функция M , которая может быть также записана в виде

$$M = (B\bar{\xi})^T G_B (B\bar{\xi}), \quad (4.21)$$

подчиняется распределению χ^2 с $m - r$ степенями свободы.

Если уравнения (4.1) уже линейны, то окончательное решение задается равенствами (4.17)—(4.20). В нелинейных случаях мы обычно рассматриваем их в качестве лучших приближений; поэтому мы заменяем x_0, η_0 на $\tilde{x}, \tilde{\eta}$, вычисляем, исходя из этого, матрицы (4.3)—(4.5) и получаем еще лучшие приближения, вводя и их в (4.17)—(4.20). Этот итерационный процесс может быть повторен до тех пор, пока не будет получено удовлетворительное решение. Вообще говоря, нелегко определить, когда решение является удовлетворительным.

Простейший метод состоит в том, чтобы исследовать значение M на каждом шаге и прекратить итерационный процесс, если M опустится ниже заданного значения или уменьшится менее чем на заданную часть при переходе от одного шага к другому. Конечно, возможно также, что M совсем не уменьшается, т. е. что итерационный процесс расходится. В этом случае следует рассмотреть пункты (а)—(г), упомянутые в конце разд. 9.2. Более конкретные критерии приемлемости решения можно выработать только тогда, когда известна природа рассматриваемой задачи.

9.5. ПРОГРАММА НА ФОРТРАНе ДЛЯ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ В ОБЩЕМ СЛУЧАЕ. ПРИМЕРЫ

В этом разделе мы рассматриваем довольно простую программу, которая решает общую задачу разд. 9.4 методом наименьших квадратов с использованием итерационного метода. Эту программу можно рассматривать как обобщение программы приближения синусоидой примера 9.3 на общий случай. Кроме того, она оказывается более гибкой в следующих отношениях.

Измерения y , их ковариационная матрица $C_y = G_y^{-1}$ и функция $f(\eta, x)$ в программе не описаны. Для решения конкретной задачи следует написать две специальные подпрограммы. В первой, названной INWAL, измерения и их ковариационные матрицы располагаются в массивах Y и CY соответственно. задается также число измерений N. Подпрограмма DERIV используется на каждом шаге итерационной процедуры для подсчета матриц производных A и B и вектора C. На начальном шаге программа DERIV, кроме того, дает первое приближение неизвестных X. Существуют задачи, в которых полезно попробовать подогнать к данным различные функции $f(\eta, x)$, т. е. испробовать различные гипотезы. Целочисленное переменное NHYP, которое также является аргументом подпрограммы DERIV, используется для нумерации этих гипотез. Каждый раз, когда вызывается DERIV, выполняются операции, описанные выше для гипотезы, имеющей номер NHYP. Если все гипотезы уже проверены и DERIV вызывается

снова, NHYP получается отрицательным. В этом случае основная программа выходит на STOP.

Таким образом, все вычисления, относящиеся к конкретной задаче, хранятся вне основной программы, которая поэтому может быть использована для широкого круга задач. Программа начинается с вызова подпрограммы INVAL, описанной выше, и заносит в память измерения и их ковариационные матрицы. Остальная часть программы состоит из двух циклов: внешний цикл (начинающийся с оператора номер 7) выполняется один раз для каждой гипотезы и начинается переносом измерений в массив ETA, внутренний цикл (начинающийся с оператора номер 10) пробегает все шаги итерационного процесса, занумерованные целочисленной переменной NSTEP. Внутренний цикл начинается вызовом подпрограммы DERIV. Как было описано выше, DERIV задает первое приближение неизвестных в массиве X и вычисляет A, B и C. Она также задает число уравнений (M) и неизвестных параметров (NR). В случае связанных наблюдений, когда нет неизвестных параметров, как, например, в разд. 9.3, переменная NR полагается равной нулю. Это используется в дальнейшем с помощью операторов IF для того, чтобы пренебречь всеми членами, содержащими матрицу A, которая в этом случае не имеет смысла. Следующий раздел программы выполняет матричный подсчет, необходимый для того, чтобы получить исправления (XI, DELTA) и уточненные значения неизвестных параметров и наблюдений, которые снова засылаются в массивы X и ETA. В этой части вычисляется вектор разностей между исходными

```

C
C      PROGRAM FOR GENERAL LEAST SQUARES FITTING.
C
C      DIMENSION X(10),Y(100),CY(1000),A(1000),B(1000),C(50),
*ETA(100),XI(10),DELTA(100),EPSLN(100),CX(100),CETA(1000),GB(1000),
*S1(1000),S2(1000),S3(1000),S4(1000),S5(1000)
C
C      SUBROUTINE INVAL PREPARES INITIAL VALUES.
C      IT STORES THE MEASUREMENTS IN Y, THEIR COVARIANCE MATRIX IN CY
C      AND THE NUMBER OF MEASUREMENTS IN N.
C
C      CALL INVAL(Y,CY,N)
C
C      WRITE INITIAL VALUES
C
C      WRITE(6,1400)
C      CALL MTXWRT(Y,1,N)
C      WRITE(6,1500)
C      CALL MTXWRT(CY,N,N)
C      NHYP=0
5   NHYP=NHYP+1
C
C      SET FIRST APPROXIMATION OF ETA EQUAL TO Y.
C
C      CALL MTXTRA(Y,ETA,N,1)
C      NSTEP=0
10  NSTEP=NSTEP+1

```

```

C   SUBROUTINE DERIV SETS UP ONE HYPOTHESIS AT A TIME. THE NUMBER
C   NHYP INDICATES THE CURRENT HYPOTHESIS. IT IS SET NEGATIVE, IF
C   NO MORE HYPOTHESES ARE TO BE TRIED. DERIV DETERMINES THE NUMBER
C   NR OF UNKNOWN AND THE NUMBER M OF EQUATIONS. IT PLACES FIRST
C   APPROXIMATIONS FOR THE UNKNOWN IN X AND ALSO FIXES THE
C   MAXIMUM NUMBER MAXSTP OF ITERATION STEPS TO BE PERFORMED.
C   DERIV CALCULATES THE MATRICES A, B OF DERIVATIVES AND THE VECTOR
C   C FOR THE CURRENT VALUES OF X AND ETA.
C
C   CALL DERIV (X,ETA,A,B,C,NR,N,M,NHYP,NSTEP,MAXSTP)
C   NF=M-NR
C
C   STOP PROGRAM IF 'NHYP' NEGATIVE.
C
C   IF (NHYP) 500,11,11
500 STOP
C
C   WRITE FIRST APPROXIMATION OF UNKNOWN.
C
C   11 IF(NSTEP-1) 15,12,15
C   12 IF(NR) 13,15,13
C   13 WRITE(6,1300) NHYP,M,NR,NF
C   CALL MTXWRT (X,1,NR)
C
C   CALCULATE GB (EQ. 4.12).
C
C   15 CALL MTXMBT (CY,B,S1,N,N,M)
C   CALL MTXMLT (B,S1,S2,M,N,M)
C   CALL MTXINV (S2,GB,M)
C   IF(NR) 17,16,17
C   16 CALL MTXTRA (C,S2,M,1)
C   GO TO 18
C
C   CALCULATE MATRIX CX (EQ. 4.19).
C
C   17 CALL MTXMLT (GB,A,S1,M,M,NR)
C   CALL MTXMAT (A,S1,S2,NR,M,NR)
C   CALL MTXINV (S2,CX,NR)
C
C   CALCULATE VECTOR OF CORRECTIONS XI (EQ. 4.14).
C
C   CALL MTXMBT (CX,A,S2,NR,NR,M)
C   CALL MTXMLT (S2,GB,S1,NR,M,M)
C   CALL MTXMLT (S1,C,XI,NR,M,1)
C   CALL MTXMSC (XI,XI,-1.,NR,1)
C
C   CALCULATE NEW VALUES FOR X (EQ. 4.17).
C
C   CALL MTXADD (X,XI,X,NR,1)
C
C   CALCULATE VECTOR OF CORRECTIONS DELTA (EQ. 4.15).
C
C   CALL MTXMLT (A,XI,S1,M,NR,1)
C   CALL MTXADD (S1,C,S2,M,1)
18  CALL MTXMLT (GB,S2,S1,M,M,1)
C   CALL MTXMAT (B,S1,S2,N,M,1)
C   CALL MTXMLT (CY,S2,DELTA,N,N,1)
C
C   CALCULATE NEW VALUES FOR ETA (EQ. 4.18).
C
C   CALL MTXSUB (ETA,DELTA,ETA,N,1)
C
C   CALCULATE VECTOR EPSLN
C
C   CALL MTXSUB (ETA,Y,EPSLN,N,1)
C
C   CALCULATE MINIMUM FUNCTION XM (EQ. 4.21).
C
C   CALL MTXMLT (B,EPSLN,S1,M,N,1)
C   CALL MTXMLT (GB,S1,S2,M,M,1)
C   CALL MTXMAT (S1,S2,XM,1,M,1)

```

```

C   WRITE RESULTS OF THIS STEP
C
WRITE(6,1700) NHYP,NSTEP,XM
IF(NR) 19,20,19
19 WRITE(6,1900)
CALL MTXWRT (X,1,NR)
20 WRITE(6,1800)
CALL MTXWRT (ETA,1,N)

C   BREAK OFF ITERATION, IF MAXIMUM NUMBER OF STEPS PERFORMED.
C
IF (NSTEP-MAXSTP)25,80,80

C   IF THIS WAS FIRST STEP, STORE VALUE OF XM IN XMO AND REPEAT ITERATION.
C
25 IF (NSTEP-1) 30,30,40
30 XMO=XM
GO TO 10

C   IF MINIMUM FUNCTION DIFFERS FROM PREVIOUS VALUE BY
C   LESS THAN 0.1 PER CENT, BREAK OFF ITERATION.
C
40 TEST= (XM-XMO)/XMO
50 IF (ABS(TEST)-0.001) 90,60,60
60 XMO=XM

C   AFTER FIRST 3 STEPS COMPARE MINIMUM FUNCTION WITH RESULT OF PREVIOUS STEP.
C   GIVE UP THIS HYPOTHESIS, IF LARGER BY MORE THAN 1 PER CENT.
C   OTHERWISE PERFORM NEXT STEP.
C
IF(NSTEP-3) 10,70,70
70 IF(TEST-0.01)10,10,75
75 WRITE(6,1000) NHYP
GO TO 5
80 WRITE (6,1100)

C   CALCULATE COVARIANCE MATRIX OF ETA (EQ. 4.20).
C
90 CALL MTXMLT (B,CY,S1,M,N,N)
CALL MTXMLT (GB,S1,S2,M,M,N)
IF (NR) 92,95,92
92 CALL MTXMAT (A,S2,S3,NR,M,N)
CALL MTXMLT (CX,S3,S5,NR,NR,N)
CALL MTXMAT (S3,S5,S4,N,NR,N)
95 CALL MTXMAT (S1,S2,S5,N,M,N)
CALL MTXSUB (CY,S5,S1,N,N)
IF(NR) 97,96,97
96 CALL MTXTRA(S1,CETA,N,N)
GO TO 100
97 CALL MTXADD (S1,S4,CETA,N,N)

C   WRITE RESULTS OF CURRENT HYPOTHESIS.
C
WRITE (6,1200)
CALL MTXWRT (CX,NR,NR)
100 WRITE (6,1600)
CALL MTXWRT (CETA,N,N)

C   TRY NEXT HYPOTHESIS.
C
GO TO 5

C   FORMAT STATEMENTS

```

```

1000 FORMAT(//40H PROCESS DIVERGES FOR HYPOTHESIS NUMBER ,I2)
1100 FORMAT (//69H ITERATION WAS BROKEN OFF BECAUSE MAXIMUM NUMBER OF S
*TEPS WAS REACHED )
1200 FORMAT (//31H COVARIANCE MATRIX OF UNKNOWNNS /)
1300 FORMAT(//18H HYPOTHESIS NUMBER,I2/
*32H NUMBER OF EQUATIONS = I3/
*32H NUMBER OF UNKNOWNNS = I3/
*32H NUMBER OF DEGREES OF FREEDOM = I3//
*34H FIRST APPROXIMATION OF UNKNOWNNS /)
1400 FORMAT(15HINITIAL VALUES///15H MEASUREMENTS/)
1500 FORMAT(//36H COVARIANCE MATRIX OF MEASUREMENTS/)
1600 FORMAT(//43H COVARIANCE MATRIX OF IMPROVED MEASUREMENTS /)
1700 FORMAT(//18H HYPOTHESIS NUMBER,I2,I3H, STEP NUMBER,I2,20H, MINIMUM
* FUNCTION =,F10.5 )
1800 FORMAT(//24H IMPROVED MEASUREMENTS/)
1900 FORMAT(//11H UNKNOWNNS/)
END

```

и уточненными наблюдениями (EPSLN), который используется для определения минимизируемой функции на этом шаге (XM).

Следующая часть (начиная с оператора номер 80) определяет, необходимо ли повторить итерационный процесс. Здесь используются следующие простые критерии. Считается, что необходимая степень точности достигнута и итерации прекращаются, если минимизируемая функция перестает заметно убывать (т. е. убывает не более чем на 1%) на очередном шаге. В этом случае итерация называется *сходящейся*. Итерационный процесс также обрывается, если совершено некоторое максимальное число шагов (заданное в DERIV в виде переменной MAXSTP). В обоих случаях программа переходит к оператору 90 и далее, где вычисляется ковариационная матрица уточненных наблюдений (CETA) и записываются результаты для текущей гипотезы. Если, с другой стороны, минимизируемая функция возрастает (более чем на 1%) на данном шаге, то считается, что процесс *расходится*, а для данной гипотезы не записывается никакого результата. Возрастание минимизируемой функции допускается на первых трех шагах, поскольку итерационный процесс иногда приводит к некоторым колебаниям M относительно его окончательного значения.

Пример 9.6. Нашу общую программу можно, конечно, применять также и к линейным задачам. В этом случае решение получается за один шаг, и выбор начального приближения неизвестных не имеет значения. Мы можем положить его равным нулю в подпрограмме DERIV и задать MAXSTP равной единице. (Мы могли бы также выбрать MAXSTP больше, чтобы произвестись несколько шагов и остановиться, если результат все еще меняется после первого шага. Если имеет место такой случай, то подпрограмма DERIV содержит ошибки в программировании арифметических операций.) В примере 9.2 мы обсуждали проведение через n измеренных точек $y_j(t_j)$ прямой линии, где переменная t_j

не имела ошибок. Теперь мы изучим общую задачу построения полинома степени l по этим наблюдениям:

$$f_j(\eta, \mathbf{x}) = \eta_j - x_1 - x_2 t_j - \dots - x_{l+1} t_j^l, \quad j = 1, 2, \dots, n.$$

Эти уравнения линейны относительно η и \mathbf{x} . Матрица B является просто единичной матрицей n -го порядка. Матрица A является $n \times (l+1)$ матрицей и имеет вид

$$A = - \begin{pmatrix} 1 & t_1 & t_1^2 & \dots & t_1^l \\ 1 & t_2 & t_2^2 & & t_2^l \\ \vdots & & & & \\ 1 & t_n & t_n^2 & & t_n^l \end{pmatrix}.$$

Окончательно, так как $\eta_0 = \mathbf{y}$ и $\mathbf{x}_0 = 0$, то вектор \mathbf{c} совпадает с \mathbf{y} .

Часто интересно попробовать построить полиномы возрастающих порядков по измеренным данным, до тех пор пока число степеней свободы подгонки остается больше единицы, т. е. $n > l + 1$.

В качестве численного примера мы используем измерения, взятые из эксперимента в области физики элементарных частиц. Упругое рассеяние отрицательных K -мезонов на протонах исследуется для K -мезонов с фиксированной энергией. В системе отсчета, связанной с центром масс соударяемых частиц, распределение косинуса угла рассеяния характеризует момент количества движения возможных промежуточных состояний при столкновении. В частности, если рассматривать это распределение в виде полинома относительно $\cos \theta$, то порядок полинома может служить для определения спиновых квантовых чисел таких промежуточных состояний.

Измерения y_j ($j = 1, 2, \dots, 10$) представляют собой число наблюдаемых распадов с $\cos \theta$, находящимися в малом интервале около значения $t_j = \cos \theta_j$. Статистические ошибки, т. е. квадратные корни из наблюдаемых чисел, берутся в качестве ошибок измерений. Следовательно, дисперсии по величине совпадают с измерениями. Так как отдельные измерения предполагаются независимыми, ковариационная матрица содержит только диагональные элементы. Подпрограмма `INVAL` принимает приведенную ниже простую форму, в которую вводятся измерения при помощи оператора `DATA`.

Подпрограмма `DERIV` написана таким образом, что подгоняются полиномы от нулевого (постоянная) до 5-го порядка, т. е. всего испытывается 6 гипотез.

Первая часть выходной печати программы воспроизведена в табл. 9.5. Результаты проверки гипотез собраны в табл. 9.6

```

SUBROUTINE INVAL (Y,CY,N)
C
C POLYNOMIAL FIT.
C
DIMENSION Y(1),CY(1),A(10)
DATA A/81.,50.,35.,27.,26.,60.,106.,189.,318.,520./
N=10
CALL MTXTRA (A,Y,10,1)
CALL MTXUNT (CY,10)
DO 10 J=1,N
JJ=(J-1)*N+J
10 CY(JJ)=Y(J)
RETURN
END

SUBROUTINE DERIV(X,Y,A,B,C,NR,N,M,NHYP,NSTEP,MAXSTP)
C
C POLYNOMIAL FIT.
C
DIMENSION X(1),Y(1),A(1),B(1),C(1),T(10)
DATA T/-0.9,-0.7,-0.5,-0.3,-0.1,0.1,0.3,0.5,0.7,0.9/
IF(NHYP-6) 5,5,80
5 MAXSTP=1
M=N
CALL MTXUNT(B,M)
NR=NHYP
DO 20 J=1,NR
20 X(J)=0.
DO 60 I=1,N
IK=(I-1)*NR
A(IK+1)=-1.
C(I)=Y(I)
IF(NR-1) 60,60,40
40 DO 50 K=2,NR
50 A(IK+K)=-T(I)**(K-1)
60 CONTINUE
70 RETURN
80 NHYP=-1
GO TO 70
END

```

и воспроизведены на рис. 9.4. Мы видим, что первые две гипотезы (постоянная и прямая линия) никак не согласуются с экспериментальными данными. Это усматривается из рис. 9.4 и отражено в значениях минимизируемой функции. Третья гипотеза, полином второго порядка, дает качественное согласование. Однако большая часть измеренных точек не попадает на график параболы даже с учетом интервалов ошибок, т. е. с учетом одного стандартного отклонения. Для вычисленного значения минимизируемой функции критерий χ^2 с $\alpha = 0,001$ дает отрицательный результат. С другой стороны, для 4-, 5- и 6-й гипотез согласованность очень хорошая. Построенные для этих случаев кривые проходят через интервалы ошибок и почти совпадают между собой. Критерий χ^2 не отвергается даже при $\alpha = 0,5$. Следовательно, можно заключить, что полинома третьего порядка уже достаточно для описания этого распределения. Коэффициенты \tilde{x}_5 , \tilde{x}_6 гипотез 5 и 6 не вносят ничего существенно нового. Проверка результатов в целом показывает, что эти коэффициенты малы

INITIAL VALUES

Таблица 9.5

Часть выходной печати для примера 9.6

MEASUREMENTS

81.00000 50.00000 35.00000 27.00000 26.00000 60.00000 106.00000 189.00000 318.00000 520.00000

COVARIANCE MATRIX OF MEASUREMENTS

81.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	50.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	35.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	27.00000	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	26.00000	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	60.00000	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	106.00000	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	189.00000	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	318.00000	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	520.00000

HYPOTHESIS NUMBER 1
 NUMBER OF EQUATIONS = 10
 NUMBER OF UNKNOWNNS = 1
 NUMBER OF DEGREES OF FREEDOM = 9

FIRST APPROXIMATION OF UNKNOWNNS

0.0

HYPOTHESIS NUMBER 1, STEP NUMBER 1, MINIMUM FUNCTION = 833.54541

UNKNOWNNS

57.84526

IMPROVED MEASUREMENTS

57.84528 57.84525 57.84525 57.84523 57.84525 57.84526 57.84529 57.84532 57.84570 57.84570

Таблица 9.5 (продолжение)

ITERATION WAS BROKEN OFF BECAUSE MAXIMUM NUMBER OF STEPS WAS REACHED

COVARIANCE MATRIX OF UNKNOWNNS

8.26552	9.90999
9.91000	39.58414

COVARIANCE MATRIX OF IMPROVED MEASUREMENTS

22.49068	17.34750	12.20437	7.06124	1.91809	-3.22504	-8.36816	-13.51132	-18.65446	-23.79762
17.34750	13.78776	10.22796	6.66819	3.10841	-0.45136	-6.01114	-7.57090	-11.13069	-14.69049
12.20438	10.22797	8.25157	6.27514	4.29873	2.32232	0.34591	-1.63051	-3.60692	-5.58334
7.06124	6.66819	6.27514	5.88211	5.48905	5.09600	4.70295	4.30990	3.91685	3.52379
1.91810	3.10841	4.29873	5.48905	6.67938	7.86968	9.05999	10.25031	11.44063	12.63093
-3.22504	-0.45136	2.32232	5.09600	7.86968	10.64337	13.41703	16.19070	18.96439	21.73805
-8.36817	-4.01114	0.34591	4.70295	9.05999	13.41703	17.77411	22.13110	26.48816	30.84518
-13.51131	-7.57091	-1.63050	4.30990	10.25031	16.19070	22.13110	28.07158	34.01193	39.95233
-18.65443	-11.13069	-3.60691	3.91686	11.44063	18.96440	26.48816	34.01192	41.53596	49.05948
-23.79758	-14.69047	-5.58332	3.52381	12.63095	21.73808	30.84520	39.95233	49.05949	58.16685

по сравнению со своими ошибками, т. е. по сравнению с квадратными корнями элементов (5,5) и (6,6) ковариационной матрицы неизвестных параметров.

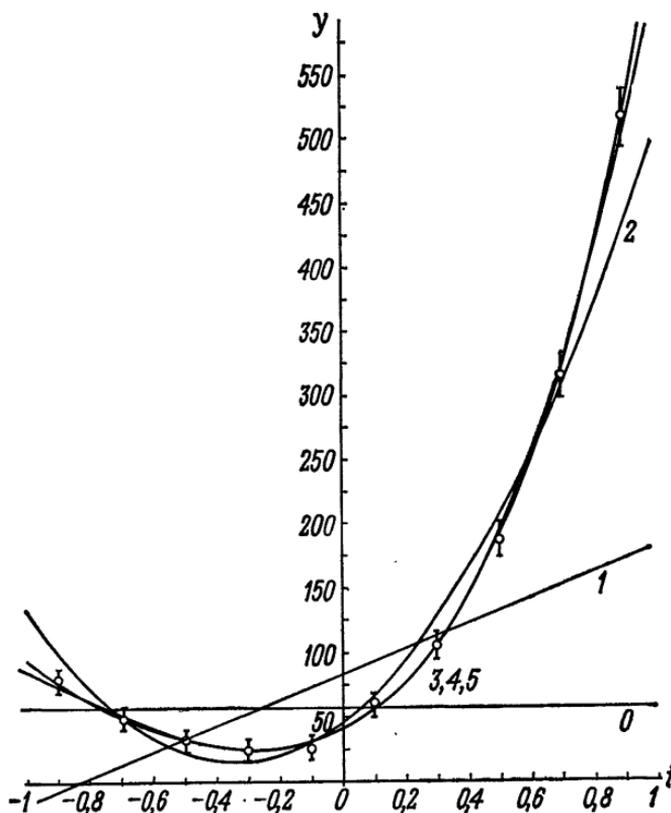


Рис. 9.4. Подгонка полиномов различного порядка (0, 1, ..., 5) к данным примера 9.6.

Таблица 9.6

Результаты примера 9.6

№ гипотезы	Неизвестные параметры						Число степеней свободы	Минимум функции
	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3	\tilde{x}_4	\tilde{x}_5	\tilde{x}_6		
1	57,85						9	833,55
2	82,66	99,10					8	585,45
3	47,27	185,96	273,61				7	36,41
4	37,94	126,55	312,02	137,59			6	2,85
5	39,62	119,10	276,49	151,91	52,60		5	1,68
6	39,88	121,39	273,19	136,58	56,90	16,72	4	1,66

Пример 9.7. В качестве примера нелинейной задачи мы рассмотрим несколько измерений точек (s_i, t_i) на плоскости с координатами s и t . Точки лежат на окружности с неизвестным центром s_0, t_0 и неизвестным радиусом r (рис. 9.5). Такая окружность описывается уравнением

$$(s - s_0)^2 + (t - t_0)^2 = r^2$$

или

$$s^2 + t^2 - 2ss_0 - 2tt_0 + s_0^2 + t_0^2 - r^2 = 0.$$

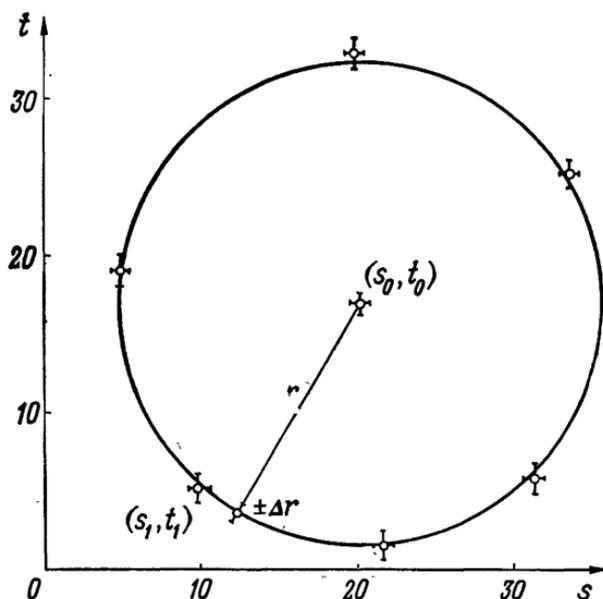


Рис. 9.5. Подгонка окружности к данным примера 9.7.

Мы можем выразить это уравнение в обозначениях, принятых нами при рассмотрении общего метода наименьших квадратов. Мы используем неизвестные параметры

$$x_1 = s_0, \quad x_2 = t_0, \quad x_3 = r.$$

Координаты s_i, t_i ($i = 1, 2, \dots, m$) измеренных точек обозначаются

$$y_{2i-1} = s_i, \quad y_{2i} = t_i,$$

т. е.

$$y_1 = s_1, \quad y_2 = t_1, \quad y_3 = s_2, \quad y_4 = t_2, \quad \dots$$

Тогда уравнения (4.1) становятся такими:

$$y_{2i-1}^2 + y_{2i}^2 - 2x_1 y_{2i-1} - 2x_2 y_{2i} + x_1^2 + x_2^2 - x_3^2 = 0, \\ i = 1, 2, \dots, m.$$

Матрицы (4.3) и (4.4) принимают вид

$$A = \begin{pmatrix} -2y_1 + 2x_{10} & -2y_2 + 2x_{20} & -2x_{30} \\ -2y_3 + 2x_{10} & -2y_4 + 2x_{20} & -2x_{30} \\ \vdots & \vdots & \vdots \\ -2y_{2m-1} + 2x_{10} & -2y_{2m} + 2x_{20} & -2x_{30} \end{pmatrix}$$

и

$$B = \begin{pmatrix} 2y_1 - 2x_{10} & 2y_2 - 2x_{20} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 2y_3 - 2x_{10} & 2y_4 - 2x_{20} & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 2y_{2m-1} - 2x_{10} & 2y_{2m} - 2x_{20} & 0 \end{pmatrix}.$$

Теперь мы нуждаемся в первом приближении для неизвестных параметров s_0 , t_0 , r . Используем простое геометрическое построение: рассмотрим первые три точки и восставим перпендикуляры из середины отрезков, соединяющих две соседние точки. Пересечение этих перпендикуляров было бы центром окружности, если бы эти три точки были измерены без ошибок. Радиус равен расстоянию между центром и любой (мы будем брать первую) точкой на окружности. Необходимые формулы хорошо известны из геометрии. Мы здесь их не выписываем, их можно прочесть в тексте программы.

Подпрограмма DERIV принимает следующий вид:

```

SUBROUTINE DERIV(X,Y,A,B,C,NR,N,M,NHYP,NSTEP,MAXSTP)
C
C FIT OF CIRCLE TO MEASURED POINT COORDINATES
C
DIMENSION X(1),Y(1),A(1),B(1),C(1)
IF(NHYP-1) 60,5,60
5 IF(NSTEP-1) 8,7,8
7 NR=3
M=N/2
MAXSTP=10
C
C DERIVE FIRST APPROXIMATION
SA=(Y(1)+Y(3))/2.
TA=(Y(2)+Y(4))/2.
SB=(Y(3)+Y(5))/2.
TB=(Y(4)+Y(6))/2.
SLOPEA=(Y(3)-Y(1))/(Y(4)-Y(2))
SLOPEB=(Y(5)-Y(3))/(Y(6)-Y(4))
SO=(TA-TB-SLOPEA*SA+SLOPEB*SB)/(SLOPEB-SLOPEA)
TO =TA+ SLOPEA*(SO-SA)
RSQU=(SO-Y(1))**2+(TO-Y(2))**2
X(1)=SO
X(2)=TO
X(3)=SQRT(RSQU)
C
C CALCULATE MATRICES A,B,C
C
8 DO 10 I=1,M
IJ=(I-1)*3+1

```

```

A(IJ)=-2.*Y(2*I-1)+2.*X(1)
A(IJ+1)=-2.*Y(2*I)+2.*X(2)
10 A(IJ+2)=-2.*X(3)
   DO 30 I=1,M
   DO 30 J=1,N
   IJ=(I-1)*N+J
   IF(2*I-1-J) 20,15,20
15 B(IJ)=2.*Y(J)-2.*X(1)
   GO TO 30
20 IF(2*I-J) 28,25,28
25 B(IJ)=2.*Y(J)-2.*X(2)
   GO TO 30
28 B(IJ)=0.
30 CONTINUE
   DO 40 I=1,M
40 C(I)=Y(2*I-1)**2+Y(2*I)**2-2.*X(1)*Y(2*I-1)-2.*X(2)*Y(2*I)
   **X(1)**2+X(2)**2 -X(3)**2
50 RETURN
60 NHYP=-1
   GO TO 50
   END

```

Для нашего примера 6 точек на окружности с неизвестным центром и радиусом были измерены с расчетной точностью 0,5 мм по оси s и 1 мм по оси t , т. е. $\sigma^2(s_i) = 0,25 \text{ мм}^2$, $\sigma^2(t_i) = 1 \text{ мм}^2$. Предполагалось, что измерения некоррелированы. Печать результатов приведена в табл. 9.7. Можно заметить, что сходимость достаточно быстрая. Действительно, заметное исправление первого приближения происходит только на первом шаге. Кроме того, заметим, что разности между уточненными и исходными измерениями (т. е. $\bar{\varepsilon}_j$), вообще говоря, больше для t -координаты. Это вытекает из нашего начального предположения, что $\sigma^2(t_i) > \sigma^2(s_i)$.

В табл. 9.8 мы приводим часть выходной печати (входные данные и окончательные результаты) вычислений, базирующихся на другом наборе из 6 точек той же самой окружности. В то время как в предыдущем примере точки были более или менее равномерно распределены на окружности, теперь они взяты на небольшой части окружности. Мы видим, что хотя ковариационная матрица измерений одна и та же в обоих случаях, элементы ковариационной матрицы неизвестных параметров теперь намного больше.

Печать примера 9.7

MEASUREMENTS

10.20000 6.40000 22.09999 1.90000 32.20000 6.10000 33.50000 26.50000 20.20000 33.29999 6.00000 19.20000

COVARIANCE MATRIX OF MEASUREMENTS

0.25000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	1.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.25000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	1.00000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.25000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	1.00000	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.25000	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.00000	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.25000	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.00000	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.25000	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.00000

HYPOTHESIS NUMBER 1
 NUMBER OF EQUATIONS = 6
 NUMBER OF UNKNOWN = 3
 NUMBER OF DEGREES OF FREEDOM = 3

FIRST APPROXIMATION OF UNKNOWN

21.35919 17.92542 16.04251

HYPOTHESIS NUMBER 1, STEP NUMBER 1, MINIMUM FUNCTION = 0.46391

UNKNOWN

21.29190 17.33159 15.49220

IMPROVED MEASUREMENTS

10.22099 6.40673 22.10045 1.06022 32.15936 6.27723 33.59576 26.77057 20.20985 32.77652 5.91352 19.22870

HYPOTHESIS NUMBER 1, STEP NUMBER 2, MINIMUM FUNCTION = 0.44326

UNKNOWN

21.29671 17.33029 15.49808

IMPROVED MEASUREMENTS

10.22159 6.48908 22.10054 1.85306 32.15958 6.27639 33.59340 26.76332 20.20961 32.79022 5.91522 19.22786

HYPOTHESIS NUMBER 1, STEP NUMBER 3, MINIMUM FUNCTION = 0.44328

UNKNOWN

21.29669 17.33029 15.49808

IMPROVED MEASUREMENTS

10.22159 6.48907 22.10054 1.85308 32.15958 6.27639 33.59340 26.76332 20.20961 32.79021 5.91522 19.22786

COVARIANCE MATRIX OF UNKNOWN

0.16513 -0.00794 0.03522
 -0.00794 0.23195 0.01188
 0.03522 0.01188 0.09978

COVARIANCE MATRIX OF IMPROVED MEASUREMENTS

0.21694 -0.12944 -0.00078 0.05992 -0.00927 0.03773 0.00904 0.02773 -0.00037 0.02105 0.03444 -0.01699
 -0.12944 0.49318 -0.00305 0.23462 -0.03629 0.14773 0.03538 0.10857 0.08244 0.08244 -0.13485 -0.06654
 -0.00078 -0.00305 0.24988 0.00891 0.00102 -0.00415 -0.00062 -0.00005 0.00003 -0.00174 -0.00014 0.00007
 0.05992 0.23462 0.00891 0.31340 -0.07858 0.31987 0.00121 0.00372 -0.00235 0.13377 0.01089 -0.00537
 -0.00927 -0.03629 0.00102 0.07858 0.07858 0.09158 0.01495 0.04589 -0.00026 -0.01493 0.01553 -0.00766
 0.03773 0.14773 0.00415 0.31987 0.29158 0.62723 -0.06087 -0.18677 -0.00107 0.06075 -0.06322 0.03120
 0.00904 0.03538 -0.00002 0.00121 0.01495 -0.06087 0.22522 -0.07605 0.00179 0.10160 0.00259 -0.00128
 0.0273 0.10857 -0.00005 0.00372 0.04589 -0.18677 0.07605 0.76666 0.00548 0.31175 0.00796 -0.00393
 -0.00037 -0.00145 0.00003 -0.00235 0.00026 -0.00107 -0.00179 -0.00568 0.24980 0.01137 0.00206 -0.00102
 0.02105 0.08244 0.00014 0.13377 -0.01493 0.06075 0.10160 0.31175 0.11137 0.35337 -0.11736 0.05791
 0.03444 0.13485 -0.00014 0.01089 0.01553 -0.06322 0.00259 0.00796 0.00206 -0.11736 0.19551 0.02689
 -0.01699 -0.06654 0.00007 -0.00537 -0.00766 0.03120 -0.00128 -0.00393 -0.00102 0.03791 0.02689 0.98673

Таблица 9.8 (продолжение)

COVARIANCE MATRIX OF IMPROVED MEASUREMENTS

0.24589	-0.01387	0.00452	0.04218	0.00039	0.00656	0.00078	-0.02884	0.00240	-0.02551	-0.00398	0.01948
-0.01387	0.95324	0.01523	0.14218	0.00132	0.02213	0.00263	-0.09721	0.00811	-0.08600	-0.01342	0.06566
0.00452	0.01523	0.24248	-0.07021	0.00193	0.03236	-0.00067	0.02464	-0.00114	0.01207	0.00288	-0.01409
0.04218	0.14218	-0.07021	0.34473	0.01800	0.30198	-0.00623	0.22999	-0.01061	0.11260	0.02687	-0.13148
0.00039	0.00132	0.00193	0.01800	0.24767	-0.03915	-0.00046	0.01688	-0.00092	0.00974	0.00139	-0.00680
0.00656	0.02213	0.03236	0.30198	-0.03915	0.34327	-0.00767	0.28321	-0.01540	0.16341	0.02330	-0.11400
0.00078	0.00263	-0.00067	-0.00623	0.00046	-0.00767	0.24949	0.01888	0.00060	-0.00639	-0.00934	0.00124
-0.02884	-0.09721	0.02464	0.22999	0.01688	0.28321	0.01888	0.30239	-0.02224	0.23593	-0.00934	0.04570
0.00240	0.00811	-0.00114	-0.01061	-0.00092	-0.01540	0.00060	-0.02224	0.2379	0.06587	0.00526	-0.02572
-0.02551	-0.08600	0.01207	0.11260	0.00974	0.16341	-0.00639	0.23593	0.06587	0.30114	-0.05578	0.27292
-0.00398	-0.01342	0.00288	0.02687	0.00139	0.02330	0.00025	-0.00934	0.00526	-0.05578	0.24420	0.02837
0.01948	0.06566	-0.01409	-0.13148	-0.00680	-0.11400	-0.00124	0.04570	-0.02572	0.27292	0.02837	0.86120

НЕКОТОРЫЕ ЗАМЕЧАНИЯ О МИНИМИЗАЦИИ

10.1. ОЦЕНКА ПАРАМЕТРОВ И МИНИМИЗАЦИЯ

В предыдущей главе мы определяли неизвестные параметры \mathbf{x} по измерениям \mathbf{y} , отыскивая тот вектор \mathbf{x} , для которого заданная скалярная функция \mathbf{x} и \mathbf{y} принимала наименьшее значение

$$M(\mathbf{x}, \mathbf{y}) = \min. \quad (1.1)$$

Для различных типов задач, использующих метод наименьших квадратов, сложность функции M получалась разной, как это приведено в табл. Д.7.

Исходный метод максимального правдоподобия, рассмотренный в гл. 7, дает экстремальное условие, аналогичное (1.1),

$$L = \max, \quad (1.2)$$

или

$$l = \ln L = \max, \quad \text{где } L = \prod_{j=1}^N f(y^{(j)}, \mathbf{x})$$

— функция правдоподобия, т. е. совместная условная плотность вероятности N наблюдений $\mathbf{y}^{(j)}$ для данных значений параметров \mathbf{x}^* . Таким образом, оценивание параметров тесно связано с задачей нахождения экстремума. Изменив знак функций (1.2), мы всегда можем превратить эту задачу в задачу о минимизации.

Такие задачи интересны с рассматриваемой точки зрения не только при обработке наблюдений, они находят и некоторые другие приложения в науке, технике и экономике. Все они имеют ту общую черту, что в них строится некоторая *целевая функция* $M(\mathbf{x})$, которая содержит ряд переменных параметров \mathbf{x} . Требуется отыскать значения \mathbf{x} , при которых $M(\mathbf{x})$ достигает минимума. В экономических проектах целевой функцией может быть прибыль (с отрицательным знаком), а в качестве параметров \mathbf{x} могут выступать такие переменные, как затраты на персонал, время доставки, финансирование и т. д. В планировании процедура минимизации имеет свои специфические традиции — она обычно относится к вопросу об *оптимизации* управляющих параметров.

* Как и в гл. 9, но в отличие от гл. 7 мы обозначаем измерения через \mathbf{y} , неизвестные параметры — через \mathbf{x} .

В принципе метод Гаусса, использованный в предыдущей главе, может быть применен для отыскания минимума любой целевой функции. Однако во многих частных случаях это бывает невозможно по целому ряду причин. Например, метод Гаусса может не сходиться для некоторых функций M , если недостаточно хорошо выбрано первое приближение неизвестных. Кроме того, M может быть такой, что ее нельзя будет аналитически продифференцировать, и поэтому ее частные производные придется приближать соответствующими конечными разностями. Наиболее важная причина заключается в том, что сходимость может быть очень медленной в случае большого количества параметров и что другие методы, специально предназначенные для конкретных задач определения точки минимума функции $M(\mathbf{x})$ пространства параметров, могут действовать много быстрее.

Литература по вычислительной математике последних десяти лет в значительной степени посвящена изложению и обсуждению разнообразных методов минимизации. Поэтому дать надлежащий обзор всех аспектов этого вопроса в данной книге совершенно не представляется возможным. Мы хотим только обсудить основные идеи нескольких методов.

Различными авторами было написано множество весьма общих минимизационных программ, которые обычно можно получить из соответствующих учреждений в виде набора фортрановских перфокарт. Выбор минимизационного метода или программы в каждом случае будет зависеть от характеристик той функции $M(\mathbf{x})$, которую собираются минимизировать. Никаких общих рекомендаций дать нельзя. Сравнение эффективности нескольких методов для ряда модельных функций проведено Флетчером [7] и Леоном [13]. Более общая библиография по оптимизации приведена Леоном в [14].

10.2. РАЗЛИЧНЫЕ МЕТОДЫ МИНИМИЗАЦИИ

Для решения основной проблемы

$$M(\mathbf{x}) = \min, \mathbf{x} = (x_1, x_2, \dots, x_r) \quad (2.1)$$

все методы используют *последовательную линейную минимизацию*, т. е. начиная с некоторого приближения \mathbf{x}_0 получают новое значение \mathbf{x}_1 (новое множество параметров), двигаясь в параметрическом пространстве по направлению \mathbf{p}_1 с шагом α_1 . Величина шага α_1 выбирается так, чтобы $M(\mathbf{x}_1) < M(\mathbf{x}_0)$. Обычно мы стараемся выбрать α_1 таким, чтобы $M(\mathbf{x}_1)$ было абсолютным минимумом $M(\mathbf{x})$ по направлению \mathbf{p}_1 . Затем совершается следующий шаг по направлению \mathbf{p}_2 . Общая формула для перехода от \mathbf{x}_j к \mathbf{x}_{j+1} имеет вид

$$\mathbf{x}_{j+1} = \mathbf{x}_j + \alpha_j \mathbf{p}_j, \quad M(\mathbf{x}_{j+1}) < M(\mathbf{x}_j). \quad (2.2)$$

Итерационный процесс продолжается до тех пор, пока не будет удовлетворен соответствующий критерий сходимости. Например, мы можем считать \mathbf{x}_{j+1} решением, если $M(\mathbf{x}_j) - M(\mathbf{x}_{j+1}) < \varepsilon$, где ε — некоторое заданное число, или если не существует направления, для которого $M(\mathbf{x}_{j+1}) < M(\mathbf{x}_j)$.

Существующие методы сильно различаются способом выбора направления и величины шага, другими словами, *стратегией* приближения к минимуму.

Возможна геометрическая интерпретация равенства (2.2). Функция $M(\mathbf{x})$ определяет гиперповерхность в $(r + 1)$ -мерном пространстве. Например, для двух переменных $M(\mathbf{x}) = M(x_1, x_2)$ образует поверхность в 3-мерном пространстве с координатными осями x_1 , x_2 и M . В этом случае равенство (2.2) описывает последовательный поиск минимума поверхности. Легко видеть, что решение задачи сильно затрудняется, если поверхность имеет несколько локальных минимумов и \mathbf{x}_0 не лежит вблизи абсолютного минимума. Для того чтобы обеспечить хорошее первое приближение, часто необходимо исследовать параметрическое пространство: функция $M(\mathbf{x})$ вычисляется в узлах решетки, покрывающей некоторую область параметрического пространства; за \mathbf{x}_0 берется точка, в которой $M(\mathbf{x})$ имеет наименьшее значение, если только поведение $M(\mathbf{x})$ в соседних точках указывает на непрерывное стремление $M(\mathbf{x})$ к единственному минимуму. В противном случае такое же исследование повторяется в окрестности точки \mathbf{x}_0 с большей детальностью.

Наиболее простая стратегия состоит в *линейной минимизации по направлениям координатных осей*. Этот метод иллюстрирован рис. 10.1. Линии уровня (линии равных значений M) показаны для случая двух переменных. Начиная с точки \mathbf{x}_0 минимум сначала ищется в направлении оси x_1 из условия

рис. 10.1. Линейная минимизация вдоль направлений координатных осей.

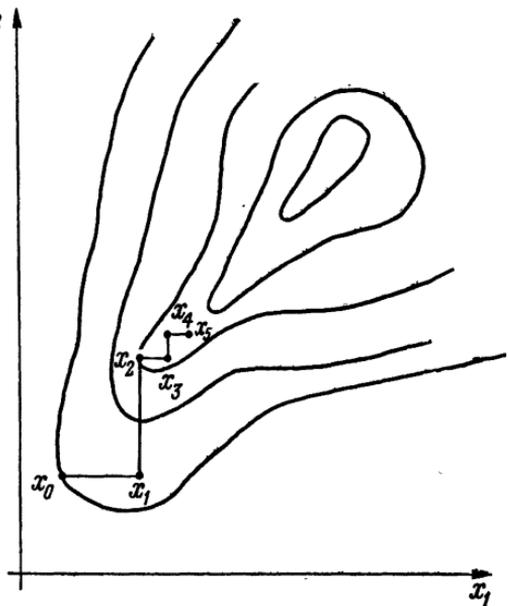


Рис. 10.1. Линейная минимизация вдоль направлений координатных осей.

$$\left(\frac{\partial M}{\partial x_1}\right)_{x_2=x_{20}} = 0.$$