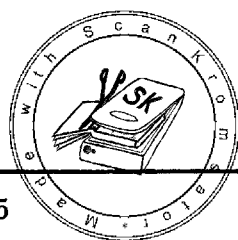


Дж. Тейлор

ВВЕДЕНИЕ В ТЕОРИЮ ОШИБОК

Перевод с английского
канд. физ.-мат. наук Л. Г. Деденко



Москва «Мир» 1985

A Series of Books in Physics
Eugene D. Commins, Editor

AN INTRODUCTION TO ERROR ANALYSIS

The Study of Uncertainties in Physical Measurements

John R. Taylor
Professor of Physics University of Colorado



University Science Books Mill Valley, California 1982

ББК 22.172

Т 30

УДК 53.088

Тейлор Дж.

Т 30 Введение в теорию ошибок. Пер. с англ.— М.: Мир, 1985. — 272 с., ил.

Книга профессора Колорадского университета (США) Дж. Тейлора является пособием по математической обработке результатов измерений в учебных физических лабораториях. Подробно разъясняются неизбежность ошибок измерений, способы фиксирования результатов измерений и на основе нормального распределения рассматриваются элементы статистической обработки случайных ошибок, обсуждаются проблема «промахов», «взвешивание» результатов различных измерений, метод наименьших квадратов, корреляции, распределение Пуассона и биномиальное распределение, критерий χ^2 . В конце каждой главы приведены задачи, для большинства которых в конце книги имеются ответы и решения.

Для студентов и преподавателей вузов, сотрудников измерительных лабораторий, а также учащихся средних специальных учебных заведений и старшеклассников

Т 2103000000—213
041(01)—85 151—85, ч. 1

ББК 22.172
517.8

Редакция литературы по физике

Джон Тейлор

Введение в теорию ошибок

Научн. редактор В. И. Самсонова
Художник В. А. Скерсис
Технический редактор Е. Н. Подчепаева

Мл. научн. редактор Г. Г. Сорокина
Художественный редактор В. А. Захаров
Корректор В. И. Постнова

ИБ № 4097

Сдано в набор 30.07.84. Подписано к печати 15.01.85. Формат 60×90¹/₁₆. Бумага ки.-журн. Гарнитура литературная. Печать высокая. Объем 8,50 бум. л. Усл. печ. л. 17,00. Усл. кр.-отт. 17,25. Уч.-изд. л. 15,97. Изд. № 2/3349. Тираж 30 000 экз. Заказ 294. Цена 1 р. 20 к.

ИЗДАТЕЛЬСТВО «МИР»

Москва, 129820, ГСП, И-110, 1-й Рижский пер., 2.

Ленинградская типография № 2 головное предприятие объединения «Ленинградская печать» им. Евгения Соколовой Союзполиграфпрома при Государственном Комитете СССР по делам издательств, полиграфии и книжной торговли. 198052, г. Ленинград, Л-52. Издательский проспект, 29.

© 1982 by University Science Books
© Перевод на русский язык, «Мир», 1985

ПРЕДИСЛОВИЕ ПЕРЕВОДЧИКА

Студенты физических и инженерных специализаций университетов и технических вузов обычно уже с первого семестра слушают курс общей физики и работают в лабораториях физического (или измерительного) практикума. Успешная работа в лаборатории наряду с овладением навыками измерений предполагает также знакомство с методами математической обработки результатов измерений. Современные учебные планы и программы высших учебных заведений уделяют этим вопросам недостаточное внимание. Практически дело сводится к двум-трем лекциям и рекомендациям использовать имеющиеся на русском языке пособия (см. список литературы, добавленной при переводе), которые давно стали библиографической редкостью и мало доступны новым поколениям студентов. Поэтому студенты младших курсов нуждаются в элементарном руководстве, в котором на доступных примерах объяснялись бы основные идеи и понятия и выводились рабочие формулы на основе, как правило, здравого смысла и элементарных понятий теории вероятностей и математической статистики.

Книга профессора Колорадского университета (США) Дж. Тейлора и является таким элементарным пособием. Отмечая, что в современных программах обучения экспериментальной физике в США практически нет места и времени для основательного обучения методам обработки результатов измерений, проф. Дж. Тейлор предлагает возместить этот недостаток официальной программы самостоятельной работой студентов.

В соответствии с этой задачей автор с особой тщательностью подходит к выбору и способу изложения материала книги. Отдельные главы строятся по возможности таким образом, чтобы их можно было изучать независимо друг от друга. В целях большей ясности изложения автор иллюстрирует основные идеи и понятия на большом числе простых примеров (отсюда значительный объем книги). Чтобы привлечь внимание на основные результаты, Дж. Тейлор выделяет их в тексте книги. Для контроля в конце каждой главы приведены задачи. С автором можно только согласиться, что решение задач — один из лучших методов усвоения материала. Для самоконтроля в конце книги приведены решения и ответы для большинства задач. Наконец, автор предполагает, что

студенты будут широко использовать микрокалькуляторы и другие современные вычислительные средства. Для удобства в конце книги приведены таблицы различных распределений. Хотя автор справедливо считает, что студенты должны работать с книгой самостоятельно, однако некоторое число лекций и домашних заданий придадут работе студентов целенаправленный характер.

Книга состоит из двух частей. Первокурсники могут ограничиться изучением первой части. Две первые главы, не содержащие математических трудностей, можно рекомендовать старшеклассникам и учащимся средних специальных учебных заведений. Вторую часть можно рекомендовать студентам второго курса, а некоторые разделы — и старшекурсникам. В книге не рассматриваются вопросы, связанные с оформлением лабораторных работ. В литературе, добавленной при переводе книги, можно найти ответы на эти вопросы.

Книга предназначена для студентов и преподавателей вузов; она может быть полезной сотрудникам измерительных лабораторий, а также учащимся средних специальных учебных заведений и старшеклассникам.

Л. Г. Деденко

ПРЕДИСЛОВИЕ

Результаты всех измерений, как бы тщательно и на каком бы научном уровне они ни выполнялись, подвержены некоторым погрешностям. Теория ошибок — наука, занимающаяся изучением и оценкой погрешностей; эти две ее функции позволяют ученому определить, насколько велики погрешности в его измерениях, и помогают уменьшить их, когда это необходимо. Анализ погрешностей, или «ошибок», является существенной частью любого научного эксперимента, и поэтому теория ошибок занимает важное место в любом университетском курсе обучения экспериментальным наукам. Она может быть даже одной из наиболее интересных частей этого курса. Оценка погрешностей и поиск возможности их уменьшения до уровня, позволяющего сделать надлежащий вывод, могут превратить всю систему скучных и рутинных измерений в подлинно интересное упражнение.

Эта книга служит введением в теорию ошибок, предназначенным для использования при изучении вводного университетского курса по экспериментальной физике, который обычно читается студентам первого или второго курса научных или инженерных специальностей. Я не берусь утверждать, что теория ошибок является самой (или тем более единственно) важной частью такого курса, но, как показали мои исследования, она обычно представляет собой ту часть, которой наиболее часто пренебрегают или которую неправильно используют. Во многих таких курсах теорию ошибок «преподают» вручением студенту пары страничек, содержащих несколько формул, после чего предполагается, что студент справится с работой самостоятельно. Результатом подобного подхода к теории ошибок становится бессмысленный ритуал, при котором студент добавляет несколько строчек вычислений в конце каждого отчета по лабораторной работе не потому, что он понимает смысл проделанного, а просто по той причине, что так требует преподаватель.

Я написал эту книгу с убеждением, что любому студенту, даже такому, который никогда не слышал об этом предмете,

необходимо знать, что представляет собой теория ошибок, почему она интересна и важна и каким образом следует применять ее основные понятия в отчетах по лабораторным работам. Первая часть книги (гл. 1—5) содержит материал, отвечающий на все эти вопросы и иллюстрированный многочисленными примерами экспериментов, встречающихся в учебных лабораториях. Студент, овладевший этим материалом, будет знать и понимать почти все аспекты теории ошибок, с которыми он может встретиться в лаборатории первого года обучения: расчет ошибок в случае косвенных измерений, использование элементарной статистики и обоснование этих расчетов на базе нормального распределения.

Вторая часть книги охватывает более сложные темы: аппроксимацию методом наименьших квадратов, коэффициент корреляции, критерий χ^2 и др. Эти темы почти наверное не будут официально включены в лабораторные работы первокурсников, хотя некоторые студенты могут заинтересоваться отдельными вопросами. Однако студентам второго года обучения необходимо знать некоторые из этих тем, и главным образом по этой причине они включены в данную книгу.

Я понимаю, что при выполнении лабораторных работ практически не остается времени на изучение такого предмета, как теория ошибок. В Колорадском университете мы читаем одночасовую лекцию в неделю в течение первых шести недель лабораторных занятий первокурсников. Эти лекции вместе с небольшим числом домашних заданий, в которых используются задачи, приведенные в конце каждой главы, позволяют нам подробно охватить материал, изложенный в гл. 1—4 и кратко — в гл. 5. Все это дает студентам активные знания расчета ошибок в случае косвенных измерений и элементов статистики плюс поверхностное знакомство с теорией нормального распределения, лежащей в основе этих расчетов.

Из высказываний нескольких студентов нашего университета мы поняли ясно, что эти лекции — ненужное излишество по крайней мере для некоторых студентов, которые познакомились с предметом по рекомендуемой литературе и по наборам задач. Я твердо верю, что материал этой книги может быть освоен без помощи каких-либо лекций.

Для изучения ч. II может потребоваться несколько лекций в начале второго года лабораторных работ (также сопровождаемых домашними заданиями). Но даже в большей мере, чем ч. I, она предназначена для самостоятельного изучения студентом в любое время, когда его собственные нужды и интересы могли бы потребовать этого. Все семь глав этой части почти полностью независимы друг от друга, что облегчает самостоятельное изучение каждой темы.

В конце каждой главы приводится набор задач; читателю и в самом деле следует поработать с некоторыми из них, чтобы получить соответствующие навыки. Большинство расчетов ошибок выполняется совсем просто. Студент, который вдруг обнаружит, что он увяз в многочисленных сложных расчетах (будь то задачи из этой книги или лабораторная работа), почти наверное делает что-то вовсе не обязательным и излишне трудным способом. Чтобы показать преподавателям и читателям, как надо вычислять ошибки, я включил в книгу значительно больше задач, чем нужно для среднего читателя. Читатель, который решит хотя бы треть задач, во всем хорошо разберется.

На обложках приведены сводки всех основных формул. Я надеюсь, что они будут полезны читателю как во время изучения, так и после. Эти сводки сделаны по главам и, как я полагаю, должны служить краткими обзорами, к которым читатель может обратиться после изучения каждой главы.

В тексте книги некоторые формулы и правила даны в рамках. Таким образом выделены только наиболее важные утверждения, сформулированные в окончательном виде (т. е. в таком, который впоследствии уже не изменяется). Читателю, несомненно, рекомендуется запомнить эти утверждения; они и выделены для того, чтобы привлечь к себе внимание.

От читателя книги требуется постепенно возрастающий уровень математической подготовки. Для чтения двух первых глав достаточно знания только алгебры; для понимания гл. 3 нужно уметь дифференцировать (включая умение вычислять частные производные в разд. 3.9, который не обязателен); гл. 5 требует умения интегрировать и знакомства с экспоненциальной функцией. В ч. II предполагается, что читатель свободно владеет всеми этими понятиями.

В книге приведены многочисленные примеры физических экспериментов, однако понимание лежащей в основе этих экспериментов теории не является обязательным. Примеры в основном взяты из элементарной механики и оптики, в этом случае более высока вероятность того, что студент уже изучал соответствующую теорию. Читатель, которому это будет необходимо, может найти обзор нужной теории в любом вводном курсе физики.

Теория ошибок — предмет, при обсуждении которого часто возникают споры, и ни одно изложение не может быть таким, чтобы с ним все согласились. Мое собственное убеждение состоит в том, что когда надо сделать выбор между доступностью изложения и его абсолютной строгостью, то в физических книгах следует отдать предпочтение первому. Например, по спорному вопросу, следует ли использовать квадратичное сложение ошибок или складывать обычным образом

их абсолютные значения, я предпочёл изложить сначала обычное сложение их абсолютных величин, так как студент легко может понять доводы, обосновывающие такое сложение.

За последние несколько лет в учебных лабораториях произошли существенные изменения в связи с появлением карманных калькуляторов. Это имело несколько неблагоприятных последствий, и наиболее значительное из них — ужасная привычка выписывать совсем *незначащие* цифры только потому, что калькулятор их выдает. Однако почти с любой точки зрения это нововведение имеет огромное значение, особенно в теории ошибок. Карманный калькулятор позволяет за несколько секунд рассчитать средние и стандартные отклонения, на что раньше требовались часы. Он делает ненужными многие таблицы, так как с его помощью можно рассчитать функции, подобные гауссовой, быстрее, чем найти их значения в этих таблицах. Я пытался использовать это чудесное устройство везде, где только можно.

Мне доставляет удовольствие поблагодарить некоторых людей за их полезные замечания и предложения. Предварительное издание этой книги использовалось в нескольких университетах, и я признателен многим студентам и коллегам за их критику. Особенно полезными были замечания Джона Моррисона и Дэвида Несбитта из Колорадского университета, профессоров Пратта и Шрёдера из Мичиганского университета, проф. Шугарта из Калифорнийского университета в г. Беркли и проф. Симона из Бэйтского университета. Диана Каспарян, Линда Фруэ и Конни Геруле отлично и быстро отпечатали черновые наброски. Без помощи моей тещи Фрэнсис Кретшманн корректура никогда не была бы прочитана вовремя. Я признателен всем этим людям за их помощь; но самую большую благодарность приношу моей жене, которая тщательным и безжалостным редактированием способствовала значительному улучшению книги.

Дж. Р. Тейлор

1 ноября 1981 г.
Боулдер, Колорадо

ЧАСТЬ I

1. Предварительное знакомство с теорией ошибок
2. Как приводить и использовать погрешности
3. Погрешности в косвенных измерениях
4. Статистический анализ случайных погрешностей
5. Нормальное распределение

В ч. I вводятся основные понятия теории ошибок в том объеме, который требуется студенту первого года обучения для работы в типичной университетской лаборатории по физике. Две первые главы дают представление о том, что такое теория ошибок, почему она важна и как можно ее использовать в типичном отчете по лабораторной работе. В гл. 3 описан расчет ошибок в случае косвенных измерений, когда погрешности в прямых измерениях «распространяются» в процессе вычислений на погрешности в конечных рассчитанных значениях. В гл. 4 и 5 вводятся статистические методы, с помощью которых могут быть оценены так называемые случайные погрешности.

Глава 1

Предварительное знакомство с теорией ошибок

Теория ошибок — изучение и оценка погрешности в измерениях. Опыт показывает, что ни одно измерение, как бы тщательно оно ни проводилось, не может быть совершенно свободно от ошибок. Поскольку в основе любой науки и ее применений лежат измерения, исключительно важно уметь рассчитывать эти ошибки и сводить их к минимуму.

В этой главе описываются некоторые простые измерения, на которых иллюстрируется неизбежность появления экспериментальных ошибок и показывается, насколько важно знать, как велики эти ошибки. Затем мы объясняем, по крайней мере для некоторых простых случаев, как можно правильно рассчитать величины экспериментальных ошибок, главным образом не более чем на основе здравого смысла.

1.1. Ошибки как погрешности

В науке слово «ошибка» не имеет обычного значения чего-то неправильного. «Ошибка» в научном измерении означает неизбежную погрешность, которая сопутствует всем измерениям. Ошибки как таковые нельзя отнести к промахам экспериментатора; вы не можете избежать их, стараясь быть очень внимательными. Лучшее, на что вы можете рассчитывать, — это свести ошибки к возможному минимуму и надежно рассчитать их величины. В большинстве учебников вводятся дополнительные определения слова «ошибка». Некоторые из них мы обсудим позднее. Пока, однако, мы будем использовать слово «ошибка» исключительно в значении «погрешность», считая эти два слова равнозначными.

1.2. Неизбежность погрешностей

Чтобы показать неизбежность появления ошибок, мы должны лишь тщательно проанализировать любое повседневное измерение. Рассмотрим, например, действия плотника, который,

чтобы установить дверь, должен измерить высоту дверного проема. Делая прикидку, он мог бы просто взглянуть на дверной проем и оценить его высоту в 210 см. Это грубое «измерение» определенно содержит погрешность. При необходимости плотник мог бы учесть эту погрешность, допуская, что высота может быть и меньше (205 см), и больше (215 см).

Если бы он захотел произвести более строгое измерение, он мог бы использовать рулетку и определить, что высота равна 211,3 см. Это измерение определенно является более точным, чем его первоначальная прикидка, но и оно, очевидно, содержит некоторую погрешность, поскольку *невероятно*, чтобы он мог знать, что высота равна точно 211,3000 см, а не, например, 211,3001 см.

Имеется много причин, влияющих на эту остающуюся погрешность. Часть из них мы будем рассматривать в этой книге. Некоторые из источников ошибок можно было бы устранить, если бы плотник проявил больше внимания к измерению. Например, одним из источников ошибки могло служить плохое освещение, затрудняющее считывание с рулетки. Эту причину ошибки можно было бы устранить, улучшив освещение.

С другой стороны, некоторые из источников ошибки присущи самому процессу измерения и никогда не могут быть полностью устранены. Например, предположим, что рулетка плотника проградуирована полусантиметровыми делениями. Верх дверного проема, по всей вероятности, не совпадает точно ни с одним из полусантиметровых делений. В этом случае плотник должен *оценить* положение верха проема между двумя делениями. Если же верх проема совпал с одним из делений, то, учитывая, что само деление имеет ширину порядка миллиметра, он должен оценить положение верха в пределах деления. В любом случае плотник должен в конечном счете оценить, где лежит верх дверного проема относительно делений на его рулетке, и это приводит к некоторой ошибке в его отсчете.

Купив другую рулетку с чаще расположенными и более тонкими делениями, плотник может уменьшить ошибку, но не может ее полностью устранить. Если бы он преисполнился решимости определить высоту проема с наилучшей точностью, допускаемой современным техническим уровнем, он мог бы купить дорогой лазерный интерферометр. Но даже точность интерферометра ограничена величиной порядка длины волны света (около $0,5 \cdot 10^{-6}$ м). Хотя теперь плотник был бы в состоянии проводить измерения с фантастической точностью, ему все же не удалось бы *точно* определить высоту дверного проема

Более того, стремясь достигнуть все более высокой точности, наш плотник столкнется с важной и принципиальной проблемой. Он определенно обнаружит, что высота в разных местах различна. Даже в одном и том же месте он найдет, что высота изменяется, если меняются температура и влажность или даже если он случайно сотрет тонкий слой пыли. Другими словами, он обнаружит, что нет такой величины, как *высота дверного проема*. Такого рода проблема называется *проблемой определения* (высота дверного проема не является точно определяемой количественной характеристикой). Она играет важную роль во многих научных измерениях.

Опыты нашего плотника иллюстрируют известную истину. Ни одну физическую величину (длину, время, температуру и т. д.) нельзя измерить с полной определенностью. Ценой особых усилий мы можем свести ошибки до очень малых значений, но исключить их полностью невозможно.

В повседневных измерениях мы обычно не затрудняем себя обсуждением ошибок. Иногда ошибки просто не имеют значения. Если мы говорим, что расстояние между домом и школой равно 3 км, то (в большинстве случаев) не важно, значит ли это, что оно лежит «между 2,5 и 3,5 км» или «между 2,99 и 3,01 км». Часто ошибки важны, но их нельзя оценить интуитивно без точного анализа. Когда наш плотник начнет подгонять дверь, он должен знать ее высоту с ошибкой порядка 1 мм. В конце концов, пока ошибка столь мала, дверь (во всех практических случаях) будет отлично подогнана и его интерес к теории ошибок пропадет.

1.3. Как важно знать погрешности

Наш пример с плотником, измеряющим дверной проем, иллюстрирует возникновение ошибок в измерениях. Теперь мы рассмотрим пример, который более отчетливо показывает, насколько важно знать величину этих ошибок.

Предположим, что мы столкнулись с проблемой, которую, как говорят, решил Архимед, а именно: нас попросили определить, изготовлена ли корона из 18-каратного золота¹⁾, как об этом заявили, или же из более дешевого сплава. Следуя Архимеду, мы решили определить плотность материала короны, зная, что плотности 18-каратного золота и подозреваемого сплава равны соответственно

$$\rho_{\text{золото}} = 19,3 \text{ г/см}^3$$

и

$$\rho_{\text{сплав}} = 13,8 \text{ г/см}^3.$$

¹⁾ 18-каратное золото — сплав, на 24 части которого приходится 18 частей драгоценного металла и 6 частей цветных металлов. — *Прим. перев.*

Таблица 1.1. Плотность короны (в г/см³)

Измеренная величина	Эксперт А	Эксперт Б
Наилучшая оценка для $\rho_{\text{корона}}$	15	13,9
Вероятный интервал $\rho_{\text{корона}}$	13,5—16,5	13,7—14,1

Если бы мы могли измерить плотность короны $\rho_{\text{корона}}$, то (в соответствии с гипотезой Архимеда) можно было бы решить, действительно ли это золотая корона, сравнивая $\rho_{\text{корона}}$ с известными плотностями $\rho_{\text{золото}}$ и $\rho_{\text{сплав}}$.

Предположим, что мы обратились к двум экспертам по определению плотности. Эксперт А мог быстро измерить $\rho_{\text{корона}}$ и сообщить, что его наилучшая оценка для $\rho_{\text{корона}}$ равна 15 и что $\rho_{\text{корона}}$ практически достоверно лежит в интервале между 13,5 и 16,5 г/см³. Эксперт Б мог поработать немного больше и затем объявить наилучшей оценкой 13,9 и вероятный интервал от 13,7 до 14,1 г/см³. Результаты наших двух экспертов можно свести в табл. 1.1.

Относительно этих результатов следует сделать два замечания. Во-первых, хотя измерения эксперта Б значительно точнее, данные эксперта А, вероятно, также правильны. Каждый эксперт приводит интервал, в котором, как он уверен, лежит $\rho_{\text{корона}}$, и эти интервалы перекрываются. Таким образом, вполне вероятно (и фактически так оно и есть), что оба измерения правильны.

Во-вторых, ошибка в измерениях эксперта А столь велика, что его результаты просто бесполезны. Значения плотности 18-каратного золота и сплава лежат в полученном им интервале от 13,5 до 16,5 г/см³, так что измерения этого эксперта не позволяют сделать никакого заключения. С другой стороны, измерения эксперта Б ясно показывают, что корона не подлинная. Плотность предполагаемого сплава 13,8 как раз находится внутри определенного экспертом Б интервала от 13,7 до 14,1, в то время как плотность 18-каратного золота, 15,5, явно не попадает в этот интервал. Очевидно, если из измерений необходимо делать определенные выводы, то экспериментальные ошибки не должны быть слишком велики. Однако нет необходимости в том, чтобы ошибки были очень малы. В этом отношении наш пример типичен для многих научных измерений, в которых ошибки должны быть разумно малы (возможно, несколько процентов от измеряемой величины), но чрезмерная точность часто является излишней.

Так как наше решение основывается на результатах эксперта Б, состоящих в том, что $\rho_{\text{корона}}$ лежит между 13,7 и

14,1 г/см³, то важно получить от эксперта Б убедительные доказательства, чтобы мы поверили в его результаты. Другими словами, экспериментатор должен доказать правильность величины определенного им интервала. Этот момент часто игнорируется начинающим студентом, который просто утверждает, что его ошибка равна 1 мм или 2 с и т. п., опуская какие-либо доказательства. Простое утверждение ошибки без краткого объяснения способа ее оценки чаще всего бесполезно.

Наиболее важный вывод относительно измерений наших двух экспертов состоит в следующем: подобно большинству научных измерений, оба измерения были бы бесполезны, если бы они не содержали надежных сведений об их ошибках. Действительно, если бы мы располагали только информацией, содержащейся в верхней строке табл. 1.1, то мы не только не могли бы сделать какое-либо правильное заключение, но фактически были бы введены в заблуждение, так как результат эксперта А (15 г/см³) наталкивал бы на предположение, что корона подлинная.

1.4. Другие примеры

Примеры, обсуждаемые в предыдущих двух разделах, представляют хорошее введение к некоторым основным положениям теории ошибок. Они были выбраны не вследствие их собственной большой важности, и читателя можно извинить, если он сочтет их немного искусственными. Однако легко привести примеры, которые исключительно важны почти в любой области прикладной или фундаментальной науки.

В прикладных науках инженер, конструируя ядерную силовую установку, должен знать характеристики материалов и ядерного горючего, которые он собирается использовать. Производителю карманного калькулятора необходимо знать характеристики его различных электронных компонент. В любом случае кто-то должен измерить требуемые параметры, а измерив их, установить достоверность своих результатов. Здесь то и требуется применение теории ошибок. Инженеры, занятые обеспечением безопасности самолетов, поездов или автомобилей, должны разбираться в ошибках времени реакции водителей, в тормозных путях и еще во множестве других вещей. И ошибка в расчетах погрешностей может привести к несчастным случаям, подобным изображенному на обложке этой книги. Даже в области, далекой от науки, такой, как пошив одежды, теория ошибок в виде *контроля качества* играет решающую роль.

В фундаментальных науках теория ошибок имеет еще более важное значение. Когда предлагается любая новая тео-

рия, она должна быть проверена наряду с более ранними в одном или нескольких экспериментах, для которых новая и старые теории предсказывают различные результаты. В принципе просто ставится эксперимент, результаты которого позволяют сделать выбор между соперничающими теориями. На практике положение осложняется вследствие неизбежных экспериментальных ошибок. Все эти ошибки необходимо тщательно учитывать и уменьшать до тех пор, пока эксперимент не позволит выбрать одну приемлемую теорию. Другими словами, экспериментальные результаты вместе с их ошибками должны находиться в *согласии* с предсказаниями одной теории и *расходиться* с данными всех известных альтернативных вариантов. Очевидно, успех такой процедуры решающим образом зависит от понимания ученым теории ошибок и его способности убедить других в правильности своего понимания.

Известный пример такого рода проверки научной теории — измерение отклонения луча света, проходящего вблизи Солнца. Когда Эйнштейн в 1916 г. опубликовал свою общую теорию относительности, он отметил, что, согласно предсказаниям этой теории, свет от звезды, проходя вблизи края Солнца, будет отклоняться на угол $\alpha = 1,8''$. Простейшая классическая теория не предсказывала никакого отклонения ($\alpha = 0$), а более тщательное рассмотрение с классических позиций давало (как указал сам Эйнштейн в 1911 г.) отклонение на угол $\alpha = 0,9''$. В принципе необходимо было лишь наблюдать звезду, когда она сравняется с краем диска Солнца, и измерить угол отклонения α . Если бы результат был $\alpha = 1,8''$, то общая теория относительности была бы подтверждена (по крайней мере для этого явления). Если бы угол α был равен 0 или $0,9''$, то общая теория относительности оказалась бы неверна, а правильной была бы одна из более ранних теорий.

На практике измерение отклонения света Солнцем исключительно затруднено и возможно только во время солнечных затмений. Тем не менее оно было успешно измерено в 1919 г. Дайсоном, Эддингтоном и Дэвидсоном. Их наилучшая оценка составила $\alpha = 2''$, и они нашли, что с вероятностью 95 % отклонение лежит в пределах от $1,7''$ до $2,3''$ ¹⁾. Очевидно, этот результат оказался в соответствии с общей теорией относительности и противоречил любым более ранним предсказаниям. Следовательно, он убедительно подтвердил эйнштейновскую общую теорию относительности.

¹⁾ Это упрощенное рассмотрение основано на оригинальной статье Дайсона, Эддингтона и Дэвидсона (Philosophical Transactions of the Royal Society, 220A, 291 (1920)). Я перевел *вероятную ошибку*, приведенную в оригинале, в 95 %-ный доверительный интервал. Точный смысл такого доверительного интервала будет установлен в гл. 5.

В течение какого-то времени этот результат считался спорным. Многие предполагали, что ошибки сильно недооценены и, следовательно, эксперимент неубедителен. Последующие эксперименты подтвердили предсказание Эйнштейна и оправдали заключение Дайсона, Эддингтона и Дэвидсона. Важный момент состоит в том, что решение всего вопроса зависело от способности экспериментатора достоверно оценить все ошибки и убедить других, что это сделано правильно.

Студент в лаборатории вводного курса физики обычно не будет в состоянии проводить решающие испытания новых теорий. С другой стороны, многие эксперименты в лаборатории вводного курса физики задуманы как тесты существующих физических теорий. Например, теория тяготения Ньютона предсказывает, что тела падают с постоянным ускорением g (при соответствующих условиях), и студент может проводить эксперименты по проверке правильности этого предсказания. С первого взгляда такой тип эксперимента может показаться искусственным и бессодержательным, поскольку эти теории, очевидно, были проверены много раз со значительно большей точностью, чем это возможно в учебной лаборатории. Тем не менее если студент понимает решающую роль теории ошибок и принимает вызов по выполнению наиболее точных проверок, возможных с имеющимся оборудованием, то такие эксперименты могут быть интересными и поучительными упражнениями.

1.5. Оценка погрешностей при считывании со шкалы

Итак, мы рассмотрели несколько примеров, которые иллюстрируют, почему каждое измерение содержит погрешности и почему важно знать их величину. Но мы еще не обсуждали, как фактически можно оценить величину ошибки. На практике такая оценка может быть довольно сложна; она и составляет главный предмет данной книги. К счастью, для ряда простых измерений легко с приемлемой точностью оценить погрешность часто на основе лишь здравого смысла. Здесь и в разд. 1.6 мы рассмотрим два примера таких простых измерений. Понимание этих примеров позволит студенту приступить к использованию теории ошибок в своих экспериментах и послужит нам основой для дальнейшего изложения.

Наш первый пример — измерение с использованием маркированной шкалы, такой, как у линейки на рис. 1.1 или у вольтметра на рис. 1.2. Чтобы измерить длину карандаша на рис. 1.1, мы должны сначала совместить торец карандаша с нулем линейки и затем определить, где окажется его острие



Рис. 1.1. Измерение длины линейкой.

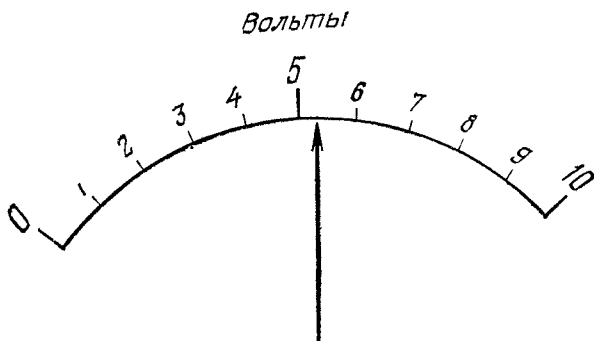


Рис. 1.2. Считывание со шкалы вольтметра.

на шкале линейки. Чтобы измерить напряжение согласно рис. 1.2, мы должны определить то место на шкале вольтметра, куда указывает стрелка. Если допустить, что правильность показаний линейки и вольтметра гарантируется¹⁾, то главная задача в каждом из этих двух случаев — определить, где располагается определенная точка по отношению к меткам шкалы. (Конечно, если существует вероятность того, что правильность показаний линейки или вольтметра не гарантируется, то мы должны будем это учесть.)

Метки на линейке на рис. 1.1 довольно близки друг к другу (с интервалом в 1 мм). Экспериментатор вполне разумно мог бы решить, что искомая длина, без сомнения, ближе к 36 мм, чем к 35 или 37 мм, и что более точный отсчет невозможен. Следовательно, он мог бы сформулировать свой вывод как

$$\begin{aligned} \text{наилучшая оценка длины} &= 36 \text{ мм}, \\ \text{вероятный интервал} & \quad 35,5\text{—}36,5 \text{ мм} \end{aligned} \quad (1.1)$$

¹⁾ В нашей стране такая гарантия реализуется системой поверочных измерений на основе соответствующих ГОСТов. Эта система состоит в проверке данного прибора по контрольному, который в свою очередь поверяется по еще более точным приборам и т. д., пока не будет обеспечено сравнение с эталоном измерения данной величины. Однако учебные и демонстрационные приборы, как правило, не поверяются, т. е. правильность показаний таких приборов не гарантируется, — *Прим. перев.*

и сказал бы, что он измерил длину до ближайшего миллиметрового деления.

Такой тип заключения — что величина лежит ближе к данной метке, чем к любой из соседних, — является довольно общим. По этой причине многие ученые следуют соглашению, в соответствии с которым утверждение $l = 36$ мм без дополнительных пояснений означает, что l ближе к 36, чем к 35 или 37 мм, т. е.

$$l = 36 \text{ мм}$$

означает

$$35,5 \text{ мм} \leq l \leq 36,5 \text{ мм}.$$

Подобным же образом запись типа $x = 1,27$ без указания какой-либо погрешности в соответствии с соглашением означает, что x лежит между 1,265 и 1,275. В данной книге мы не будем следовать этому соглашению и всегда будем указывать наши погрешности явным образом. Тем не менее для студента важно понимать это соглашение и знать, что оно используется по отношению к любому числу, приведенному без погрешности. Особенно важно знать об этом соглашении в наш век карманных калькуляторов, которые часто показывают много цифр¹⁾. Если студент слепо перепишет со своего калькулятора, скажем, число 123,456 без какого-либо объяснения, то человек, читающий это число, обязан принять, что число определено верно до шести значащих цифр, а это представляется весьма невероятным.

Метки на шкале вольтметра, показанного на рис. 1.2, расположены гораздо реже, чем на линейке. В этом случае большинство наблюдателей согласились бы, что можно сделать больше, чем просто идентифицировать метку, к которой стрелка ближе всего. Поскольку промежутки между метками больше, можно уверенно оценить, в каком месте между метками находится стрелка. Таким образом, разумное заключение об измеренном напряжении может иметь вид

$$\begin{aligned} \text{наилучшая оценка напряжения} &= 5,3 \text{ В}, \\ \text{вероятный интервал} &5,2-5,4 \text{ В}. \end{aligned} \quad (1.2)$$

Процесс определения положений между метками шкалы называется *интерполяцией*. Этот важный навык совершенствуется с практикой.

Другие наблюдатели могли бы не согласиться с оценками точности, даваемыми соотношениями (1.1) и (1.2). В частности, кто-то мог бы решить, что можно прибегнуть к интерполяции при измерении длины на рис. 1.1 и измерить ее с мень-

¹⁾ В карманных калькуляторах используется форма представления чисел с фиксированным и большим числом разрядов. — *Прим. перев.*

шей погрешностью, чем приведено в соотношении (1.1). Тем не менее лишь меньшинство стало бы отрицать, что соотношения (1.1) и (1.2) *разумно* оценивают соответствующие величины и их вероятные погрешности. Таким образом, мы видим, что приближенная оценка погрешностей представляет довольно легкую задачу в случае, когда единственной проблемой является определение положения точки на маркированной шкале.

1.6. Оценка погрешностей в случае многократных измерений

Многие измерения содержат погрешности, которые значительно труднее оценить, чем ошибки, связанные с определением положения точки на шкале. Например, когда мы измеряем временной интервал с помощью секундомера, главным источником погрешностей является не считывание с циферблата, а наше собственное неизвестное время реакции при запуске и остановке секундомера. Такого рода погрешности иногда можно надежно оценить, если повторить измерение несколько раз. Предположим, например, что мы измеряем период колебаний математического маятника один раз и получаем в результате 2,3 с. Из одного измерения нельзя много сказать об экспериментальной погрешности. Но если мы повторим измерение и получим 2,4 с, то можно немедленно сказать, что погрешность, вероятно, порядка 0,1 с. Если последовательность четырех измерений дает результаты (в секундах)

$$2,3; 2,4; 2,5; 2,4, \quad (1.3)$$

то мы можем сделать довольно правдоподобные оценки.

Во-первых, естественно предположить, что наилучшей оценкой периода будет *среднее значение* 2,4 с¹⁾.

Во-вторых, представляется довольно разумным предположение, что правильное значение для периода лежит где-то между наименьшей величиной 2,3 и наибольшей 2,5. Таким образом, мы могли бы вполне резонно заключить, что

$$\begin{aligned} \text{наилучшая оценка} &= \text{среднее} = 2,4 \text{ с}, \\ \text{вероятный интервал} & \quad 2,3-2,5 \text{ с}. \end{aligned} \quad (1.4)$$

В случаях, когда мы можем повторить одно и то же измерение несколько раз, разброс в измеренных значениях дает

¹⁾ Мы покажем в гл. 5, что наилучшая оценка, основанная на нескольких измерениях какой-то величины, есть почти всегда среднее результатов измерений.

ценное указание о погрешности в наших измерениях. В гл. 4 и 5 мы обсудим статистические методы обработки результатов таких многократных измерений. При соответствующих условиях эти статистические методы дают более правильную оценку погрешности, чем соотношение (1.4), полученное только на основе здравого смысла. Правильная статистическая обработка обладает также тем преимуществом, что дает объективную величину для погрешности, не зависящую от мнения индивидуального наблюдателя¹⁾. Тем не менее оценка (1.4) дает простое и реалистическое заключение, полученное на основании четырех измерений (1.3).

На результаты многократных измерений, такие, как (1.3), не всегда можно опираться для обнаружения погрешности. Во-первых, мы должны быть уверены, что измеряемая величина действительно есть *та же самая* величина в каждом случае. Предположим, например, что мы измеряем разрывное усилие для двух предположительно идентичных проволок, подвергая их разрыву (процедуре, которую мы не можем выполнить более чем один раз для каждой проволоки). В случае получения двух различных ответов эта разница *может* указывать на то, что наши измерения выполнены с погрешностью *или* что две проволоки в действительности не идентичны. Сама по себе разница между двумя ответами ничего не говорит о надежности наших измерений.

Даже в случае, когда мы можем быть уверены, что измеряем каждый раз одну и ту же величину, многократные измерения не всегда укажут на погрешность. Например, предположим, что секундомер, используемый при получении результатов (1.3), имел ход на 5% быстрее правильного. В этом случае все времена, получаемые с его помощью, будут на 5% больше, и никакое количество повторений (с тем же секундомером) не обнаружит этого дефекта. Погрешности такого рода, которые оказывают одно и то же влияние на все измерения, называются *систематическими* ошибками. Эти ошибки трудно обнаружить, как мы увидим в гл. 4. В нашем примере выходом из положения могла быть проверка данного секундомера относительно более надежного. В общем случае должно быть ясно, что если у кого-то имеются основания сомневаться в правильности показаний какого-либо измерительного прибора (секундомера, рулетки, вольтметра), он должен попы-

¹⁾ Правильная статистическая обработка обычно дает также *меньшую* погрешность, чем полный интервал между наименьшим и наибольшим наблюдаемыми значениями. В самом деле, по четырем результатам в (1.3) мы заключили, что период, «вероятно», лежит где-то между 2,3 и 2,5 с. Статистические методы, изложенные в гл. 4 и 5, позволят нам утверждать, что с вероятностью 70% он лежит в меньшем интервале: от 2,36 до 2,44 с.

таться поверить его относительно прибора, о котором известно, что его показания более надежны¹⁾.

Примеры, обсуждаемые в этом и предыдущем разделах, показывают, что в некоторых случаях экспериментальные погрешности могут быть легко оценены. С другой стороны, имеется много измерений, для которых оценить ошибки *не так* легко. В конце концов, мы также хотим получить более точные значения для погрешностей, чем те, которые могут нам дать обсуждаемые выше простые оценки. Эта тема будет занимать нас в последующих главах, начиная с гл. 3. В гл. 2 временно предполагается, что нам известны методы оценки погрешности для всех величин, представляющих интерес, так что мы можем обсуждать, как лучше записывать погрешности и как их использовать при получении экспериментальных выводов.

¹⁾ См. примечание переводчика на с. 19. — *Прим. перев.*

Глава 2

Как приводить и использовать погрешности

Итак, у нас имеются некоторые представления о том, насколько важны экспериментальные погрешности, каковы причины их появления. Мы также видели, как можно их оценить для ряда простых ситуаций. В этой главе будут представлены некоторые основные соображения и правила теории ошибок и приведены примеры их использования в нескольких типичных экспериментах физической лаборатории. Наша главная цель — познакомить вас с основным словарем теории ошибок и с его применением в учебной лаборатории. После этого, начиная с гл. 3, мы будем готовы перейти к изучению реальной оценки погрешностей.

В разд. 2.1—2.3 определяются несколько основных положений теории ошибок и обсуждаются некоторые общие правила представления погрешностей. В разд. 2.4—2.6 мы обсудим, как эти определения могли бы быть использованы в некоторых типичных экспериментах в учебной физической лаборатории. Наконец, в разд. 2.7—2.9 вводится еще одно основное определение — относительная погрешность — и обсуждается ее значение.

2.1. Наилучшая оценка $\pm$ погрешность

Мы видели, что корректный способ представления результата любого измерения состоит в том, что экспериментатор указывает свою наилучшую оценку измеряемой величины и интервал, в котором, как он уверен, она лежит. Например, результат измерения периодов, обсуждаемый в разд. 1.6, был представлен как

$$\begin{array}{ll} \text{наилучшая оценка времени} = 2,4 \text{ с,} & \\ \text{вероятный интервал} & 2,3—2,5 \text{ с.} \end{array} \quad (2.1)$$

В этом случае наилучшая оценка 2,4 с лежит в середине оцененного интервала вероятных значений от 2,3 до 2,5 с, так же,

как это было во всех наших примерах. Такое положение, очевидно, очень естественно и относится почти ко всем измерениям. Оно позволяет выразить результаты измерений в наиболее компактном виде. Например, измерение времени, зафиксированное в (2.1), обычно выражают следующим образом:

$$\text{измеренное значение времени} = 2,4 \pm 0,1 \text{ с.} \quad (2.2)$$

Это выражение точно эквивалентно двум в (2.1).

В общем случае результат любого измерения величины x приводится как

$(\text{измеренная величина } x) = x_{\text{наил}} \pm \delta x.$

(2.3)

Это утверждение означает, что, во-первых, наилучшая оценка экспериментатора для измеряемой величины есть число $x_{\text{наил}}$ и, во-вторых, он до определенной степени уверен, что эта величина лежит где-то между $x_{\text{наил}} - \delta x$ и $x_{\text{наил}} + \delta x$. Число δx называется *погрешностью* или *ошибкой* в измерении x . Погрешность δx принято считать положительной величиной, так что $x_{\text{наил}} + \delta x$ есть всегда *наибольшее* вероятное значение измеряемой величины и $x_{\text{наил}} - \delta x$ — *наименьшее*.

Мы умышленно оставили смысл понятия интервала от $x_{\text{наил}} - \delta x$ до $x_{\text{наил}} + \delta x$ в какой-то степени неопределенным. В некоторых измерениях его можно определить более точно. В случае простого измерения, подобного определению высоты дверного проема, мы можем легко указать интервал от $x_{\text{наил}} - \delta x$ до $x_{\text{наил}} + \delta x$, внутри которого, как мы *абсолютно* уверены, лежит измеряемая величина. К сожалению, для большинства научных измерений очень затруднительно сделать такое утверждение. В частности, если мы хотим быть *вполне* уверены в том, что измеряемая величина лежит между $x_{\text{наил}} - \delta x$ и $x_{\text{наил}} + \delta x$, обычно необходимо выбрать для δx такое значение, которое слишком велико, чтобы представлять практический интерес. Чтобы избежать этого, мы можем иногда выбирать такое значение δx , для которого вероятность того, что действительное значение лежит между $x_{\text{наил}} - \delta x$ и $x_{\text{наил}} + \delta x$, будет равна, например, 70%. Однако этого, конечно, нельзя сделать без детального знания статистических законов, которым подчиняются процессы измерения. Мы вернемся к этому вопросу в гл. 4, а пока довольствуемся определением погрешности δx , согласно которому мы «до некоторой степени» уверены в том, что измеряемая величина лежит где-то между $x_{\text{наил}} - \delta x$ и $x_{\text{наил}} + \delta x$.

2.2. Значащие цифры

Следует отметить несколько основных правил записи погрешностей. Во-первых, поскольку величина δx служит оценкой погрешности, ее, очевидно, нельзя приводить с очень большой точностью. Если мы измеряем ускорение силы тяжести g , было бы абсурдом представлять результат, подобно следующему:

$$(\text{измеренное значение } g) = 9,82 \pm 0,02385 \text{ м/с}^2. \quad (2.4)$$

Невероятно, чтобы погрешность в измерении могла быть известна до четырех значащих цифр. В случае высокоточных измерений иногда приводятся погрешности с двумя значащими цифрами, но для учебной лаборатории мы можем сформулировать следующее правило ¹⁾:

Правило приведения погрешностей

В начальной учебной лаборатории экспериментальные погрешности обычно должны округляться до одной значащей цифры. (2.5)

Таким образом, если некоторый расчет дает для погрешности $\delta g = 0,02385 \text{ м/с}^2$, то это значение должно быть округлено до $\delta g = 0,02 \text{ м/с}^2$, и вывод (2.4) следует переписать как

$$(\text{измеренное значение } g) = 9,82 \pm 0,02 \text{ м/с}^2. \quad (2.6)$$

Важное практическое следствие этого правила состоит в том, что многие расчеты ошибок можно выполнить в уме, без помощи калькулятора или даже карандаша и бумаги.

Есть только одно важное исключение из правила (2.5). Если первая цифра в погрешности δx есть 1, то, возможно, лучше сохранить две значащие цифры в δx . Например, предположим, что некоторый расчет дал для погрешности $\delta x = 0,14$. Округлить это значение до $\delta x = 0,1$ — значит на 40 % уменьшить ошибку; так что более правильным было бы сохранить две цифры и привести $\delta x = 0,14$. Тот же аргумент, вероятно, можно было бы использовать, если первая цифра есть 2, но уже определенно нельзя, если она больше.

Когда погрешность в измерении рассчитана, необходимо проанализировать, какие цифры в измеренной величине являются значащими. Утверждение типа

$$\text{измеренная скорость} = 6051,78 \pm 30 \text{ м/с} \quad (2.7)$$

¹⁾ Для удобства ссылок на такие правила они включены в нумерованную последовательность соотношений независимо от того, содержат ли они уравнения.

очевидно нелепо. Погрешность 30 означает, что цифра 5 на третьем месте от начала числа 6051,78 могла быть в действительности равна 2 или 8. Ясно, что последующие цифры 1,7 и 8 вовсе не имеют значения и должны быть округлены. Таким образом, корректная запись (2.7) есть

$$\text{измеренная скорость} = 6050 \pm 30 \text{ м/с.} \quad (2.8)$$

Ясно, что общее правило выражается следующим образом:

Правило приведения результатов

Последняя значащая цифра в любом приводимом результате обычно должна быть того же порядка величины (находиться в той же десятичной позиции), что и погрешность. (2.9)

Например, результат 92,81 с погрешностью 0,3 должен быть округлен до

$$92,8 \pm 0,3.$$

Если же ошибка равна 3, то тот же результат следует представить как

$$93 \pm 3,$$

а если ошибка равна 30, то как

$$90 \pm 30.$$

Однако используемые в расчетах числа должны, как правило, содержать *на одну значащую цифру больше*, чем это оправдано. Это уменьшит неточности, возникающие при округлении чисел. В конце расчета окончательный ответ следует округлить и избавиться от этой добавочной (и незначащей) цифры¹⁾.

Заметим, что погрешность в любой измеренной величине имеет ту же размерность, что и сама измеренная величина. Следовательно, будет понятнее и более экономно писать единицы измерения (м/с^2 , см^2 и т. д.) после результата и погрешности, как в выражениях (2.6) и (2.8). Подобным же образом, если измеренное число настолько велико или мало, что оно

¹⁾ Имеется еще одно небольшое исключение из правила (2.9): если первая цифра в погрешности мала (1 или 2), то может быть более правильным сохранить одну дополнительную цифру в конечном результате. Например, такая запись, как

$$\text{измеренная длина} = 27,6 \pm 1 \text{ см,}$$

вполне приемлема, поскольку, как можно показать, ее округление до 28 ± 1 см означало бы потерю информации.

требует «научной записи» (т. е. использования формы $3 \cdot 10^3$ вместо 3000), то проще и нагляднее приводить результат и погрешность в одинаковом виде. Например, результат

$$\text{измеренный заряд} = (1,61 \pm 0,05) \cdot 10^{-19} \text{ Кл}$$

гораздо проще прочитать и понять в такой форме записи, чем в виде

$$\text{измеренный заряд} = 1,61 \cdot 10^{-19} \pm 5 \cdot 10^{-21} \text{ Кл.}$$

2.3. Различие

Прежде чем обратиться к вопросу о том, как использовать ошибки в экспериментальных отчетах, необходимо ввести и определить несколько важных терминов. Во-первых, если два измерения одной и той же величины различаются, то мы будем говорить, что между ними имеется *различие*. Численно определим различие между двумя измерениями как их разность:

$$\text{различие} = \text{разность между двумя измеренными значениями одной и той же величины.} \quad (2.10)$$

Важно иметь в виду, что различие может быть значимым или незначимым. Если два студента измеряют одно и то же сопротивление и получают результаты

$$40 \pm 5 \text{ Ом}$$

и

$$42 \pm 8 \text{ Ом,}$$

то различие в 2 Ом меньше, чем погрешности их результатов, так что два эти измерения, очевидно, согласуются. В этом случае мы бы сказали, что различие является *незначимым*. С другой стороны, если бы два результата были

$$35 \pm 2 \text{ Ом}$$

и

$$45 \pm 1 \text{ Ом,}$$

то оказалось бы, что два измерения явно расходятся, и различие в 10 Ом было бы *значимым*. В этом случае требуется ряд тщательных проверок, чтобы обнаружить, какой из результатов является неверным.

В учебной лаборатории часто измеряют величины (такие, как скорость света c или заряд электрона e), которые прежде много раз тщательно измерялись и для которых очень точное

принятое значение известно и опубликовано в учебниках. Это принятое значение, конечно, не является абсолютно точным; оно представляет собой результат измерений и, подобно всем экспериментальным результатам, обладает некоторой погрешностью. Тем не менее в большинстве случаев принятое значение намного точнее того, которое студент может получить сам. Например, принятое значение величины скорости света c есть

$$(\text{принятое значение } c) = 299\,792\,458 \pm 1 \text{ м/с.} \quad (2.11)$$

Как ожидалось, этот результат *имеет* погрешность, но она исключительно мала по стандартам большинства учебных лабораторий¹⁾.

Хотя имеется много экспериментов, в которых измеряют величины, принятые значения которых известны, тем не менее лишь в небольшом числе случаев известен *«истинный ответ»*²⁾. Фактически истинное значение измеряемой величины *никогда* не может быть точно известно, и его в действительности трудно определить. Однако иногда полезно обсуждать разницу между измеренным значением и соответствующим истинным значением, и некоторые авторы называют эту разницу *истинной ошибкой*.

2.4. Сравнение измеренного и принятого значений

Мало смысла в выполнении эксперимента, если из него не делается какого-либо вывода. Лишь очень небольшое число экспериментов имеет целью главным образом качественный результат — наблюдение интерференционной картины на поверхности кюветы с водой или наблюдение цвета светового луча, прошедшего через некоторую оптическую систему, — в то время как огромное большинство экспериментов приводит к *количественным* выводам, т. е. к утверждению численных результатов. Поэтому важно осознать, что *не представляет никакого интереса результат, представленный в виде единственного измеренного значения*. Утверждения, подобные тому, что для плотности некоторого металла измерения дали

¹⁾ Однако так бывает не всегда. Например, если кто-то определяет показатель преломления стекла, то может получать значения в интервале от 1,5 до 1,9 в зависимости от состава стекла. Следовательно, в эксперименте по определению показателя преломления стекла, состав которого неизвестен, «принятое» значение есть не более чем грубый ориентир к ожидаемому результату.

²⁾ Так как читатель может быть в затруднении придумать *любой* такой эксперимент, приведем пример. Если кто-то измеряет отношение длины окружности к ее диаметру, то истинный ответ есть точно π . Очевидно, такие эксперименты очень надуманны.

$9,3 \pm 0,2$ г/см³ или что для импульса тележки измерения дали $0,051 \pm 0,004$ кг·м/с, сами по себе не интересны. В выводе, представляющем интерес, *должны сравниваться два или более значений*: измеренное и принятое значения; результат измерения и теоретически предсказанное значение или результаты нескольких измерений, чтобы было видно, в каком отношении друг к другу они находятся в соответствии с некоторым физическим законом. Именно при таком сравнении чисел особенно важна теория ошибок. В этом и двух последующих разделах мы обсудим три типичных эксперимента, чтобы проиллюстрировать, как используются погрешности при получении выводов из эксперимента.

По-видимому, простейший тип эксперимента — измерение величины, принятое значение которой известно. Как мы уже обсуждали, это несколько искусственный, характерный для учебной лаборатории, эксперимент. В нем измеряют величину, оценивают экспериментальные погрешности и, наконец, сравнивают их с принятым значением. Например, эксперимент по определению скорости звука в воздухе (при нормальных температуре и давлении) мог бы привести к выводу, что

$$\text{измеренная скорость} = 329 \pm 5 \text{ м/с} \quad (2.12)$$

сравнивается с выражением

$$\text{принятая скорость} = 331 \text{ м/с.} \quad (2.13)$$

Этот численный вывод студент мог бы, вероятно, прокомментировать тем, что поскольку принятое значение скорости лежит внутри полученного им интервала скоростей, то измерение было удовлетворительным, на чем его отчет мог бы и закончиться.

Смысл погрешности δx заключается в том, что правильное значение x , «вероятно», лежит между $x_{\text{наил}} - \delta x$ и $x_{\text{наил}} + \delta x$, и, конечно, *возможно*, что правильное значение слегка выходит за рамки этого интервала. Следовательно, измерение можно рассматривать как удовлетворительное, даже если принятое значение слегка выходит за рамки измеренного интервала. Например, измеренное значение 325 ± 5 м/с можно считать совместимым с принятым значением 331 м/с. С другой стороны, если принятое значение выходит далеко за рамки измеренного интервала (скажем, различие намного больше, чем удвоенная погрешность), то имеются основания полагать, что где-то допущена ошибка. Таким образом, незадачливый студент, который получит

$$\text{измеренная скорость} = 345 \pm 2 \text{ м/с} \quad (2.14)$$

по сравнению с

$$\text{принятая скорость} = 331 \text{ м/с,} \quad (2.15)$$

должен проверить свои измерения и расчеты, чтобы обнаружить допущенные огрехи.

К сожалению, очень сложно проследить появление его ошибки, так как существует множество разных возможностей. Он мог допустить ошибку в измерениях или расчетах, которые привели к результату 345 м/с. Он мог неправильно оценить погрешность своего эксперимента. (Результат 345 ± 10 м/с был бы вполне приемлемым.) Он мог сравнивать свои измерения с ошибочным принятым значением. Например, принятое значение 331 м/с есть скорость звука при нормальных температуре и давлении. Так как нормальная температура равна 0°C , то вполне вероятно, что измеренная скорость в (2.14) получена *не* при нормальной температуре. На самом деле, если измерение было выполнено при 20°C (т. е. при обычной комнатной температуре), то правильное принятое значение для скорости звука составит 343 м/с и результат измерения окажется вполне приемлемым.

Наконец (и, возможно, наиболее вероятно), различие, подобное полученному между (2.14) и (2.15), может указывать на некоторый необнаруженный источник систематической погрешности (как в случае секундомера, ход которого быстрее нормального, что обсуждалось в гл. 1). Обнаружение таких систематических погрешностей (которые изменяют результат в одном направлении) потребует тщательной проверки калибровки всех приборов и детального рассмотрения всех процессов.

2.5. Сравнение двух измеренных значений

Во многих экспериментах измеряют два значения, которые, согласно теории, должны быть равны. Например, закон сохранения импульса утверждает, что полный импульс изолированной системы есть величина постоянная. Чтобы проверить это, мы могли бы выполнить серию экспериментов с двумя тележками, которые могут сталкиваться при движении без трения по скамье. Мы могли бы измерить полный импульс двух тележек перед столкновением (p) и после столкновения (p') и затем проверить, выполняется ли равенство $p = p'$ в пределах экспериментальных погрешностей. Для одной пары измерений наши результаты могли бы иметь вид

начальный импульс $p = 1,49 \pm 0,04$ кг · м/с

и

конечный импульс $p' = 1,56 \pm 0,06$ кг · м/с.

В этом случае интервал, в котором, возможно, лежит p (от 1,45 до 1,53), *перекрывается* с интервалом, в котором, воз-

Таблица 2.1. Измеренные импульсы (в кг·м/с)

Начальный импульс p ($\pm 0,04$)	Конечный импульс p' ($\pm 0,06$)	Начальный импульс p ($\pm 0,04$)	Конечный импульс p' ($\pm 0,06$)
1,49 2,10	1,56 2,12	1,16 и т. д.	1,05 и т. д.

можно, лежит p' (от 1,50 до 1,62). Следовательно, это измерение находится в согласии с законом сохранения импульса. Если бы, с другой стороны, оказалось, что два вероятных интервала даже не близки к тому, чтобы перекрываться, то измерение не находилось бы в согласии с законом сохранения импульса и мы должны были бы искать ошибки в наших измерениях или расчетах, определять возможные систематические погрешности и проверять возможность того, что какие-то внешние силы (такие, как сила тяжести или трения) изменяют импульс системы.

Предположим, что мы повторяем подобные пары измерений несколько раз. Каков лучший способ представить наши результаты? Во-первых, почти всегда удобнее всего представить последовательность подобных измерений в виде таблицы, а не как отдельные результаты. Во-вторых, наша погрешность часто очень мало изменяется от одного измерения к другому. Например, мы могли бы принять, что погрешность во всех измерениях начального импульса p есть $\delta p \approx 0,04$ кг·м/с и что ошибка конечного импульса p' есть $\delta p' \approx 0,06$ кг·м/с. В этом случае хороший способ представить наши измерения был бы такой, как показано в табл. 2.1. Для каждой пары измерений вероятный интервал значений p перекрывается (или почти перекрывается) с интервалом значений p' . Если бы это оставалось верным для всех измерений, то мы могли бы считать, что наши результаты согласуются с законом сохранения импульса.

Немного поразмыслив, мы могли бы представить наши результаты в виде, который сделает наш вывод даже яснее. Например, закон сохранения импульса требует, чтобы *разность* $p - p'$ была равна нулю. Если бы мы добавили в нашу таблицу столбец со значениями $p - p'$, то в любом месте этого столбца стояли бы величины, не отличающиеся от нуля. Единственная трудность здесь заключается в том, что мы должны знать, как рассчитать погрешность для разности $p - p'$. Но это можно легко сделать. Предположим, что мы произвели измерения и получили

$$(\text{измеренное значение } p) = p_{\text{нап}} \pm \delta p$$

и

$$(\text{измеренное значение } p') = p'_{\text{наил}} \pm \delta p'.$$

Числа $p_{\text{наил}}$ и $p'_{\text{наил}}$ — наши наилучшие оценки для p и p' . Следовательно, наилучшая оценка для разности $(p - p')$ есть $(p_{\text{наил}} - p'_{\text{наил}})$. Чтобы найти погрешность в $(p - p')$, мы должны определить наибольшее и наименьшее вероятные значения величины $(p - p')$. Наибольшее значение величины $(p - p')$ получилось бы, если бы величина p имела *наибольшее* вероятное значение $p_{\text{наил}} + \delta p$ и в то же время p' имела *наименьшее* вероятное значение $p'_{\text{наил}} - \delta p'$. Таким образом, наибольшее вероятное значение $p - p'$ есть

$$\text{наибольшее вероятное значение} = (p_{\text{наил}} - p'_{\text{наил}}) + (\delta p + \delta p'). \quad (2.16)$$

Аналогично наименьшее вероятное значение получается, когда величина p минимальна $(p_{\text{наил}} - \delta p)$, а p' — максимальна $(p'_{\text{наил}} + \delta p')$. Это дает

$$\text{наименьшее вероятное значение} = (p_{\text{наил}} - p'_{\text{наил}}) - (\delta p + \delta p'). \quad (2.17)$$

Сопоставляя (2.16) и (2.17), мы видим, что погрешность в разности $(p - p')$ есть *сумма $\delta p + \delta p'$ начальных погрешностей*. Например, если

$$p = 1,49 \pm 0,04 \text{ кг} \cdot \text{м/с}$$

и

$$p' = 1,56 \pm 0,06 \text{ кг} \cdot \text{м/с},$$

то

$$p - p' = -0,07 \pm 0,1 \text{ кг} \cdot \text{м/с}.$$

Затем мы можем добавить лишний столбец для $p - p'$ в табл. 2.1 и получить табл. 2.2.

Теперь можно с первого взгляда определить, находятся ли наши результаты в согласии с законом сохранения импульса, проверяя, согласуются ли с нулем числа в последнем столбце (т. е. меньше ли они или сравнимы с погрешностью 0,1).

Таблица 2.2. Измеренные импульсы (в кг · м/с)

Начальный импульс p ($\pm 0,04$)	Конечный импульс p' ($\pm 0,06$)	Разность $p - p'$ ($\pm 0,1$)	Начальный импульс p ($\pm 0,04$)	Конечный импульс p' ($\pm 0,06$)	Разность $p - p'$ ($\pm 0,1$)
1,49	1,56	-0,07	1,16	1,05	0,11
2,10	2,12	-0,02	и т. д.	и т. д.	и т. д.

Другим способом получения того же эффекта было бы табулирование *отношений* p'/p , которые должны быть в согласии с величиной $p'/p = 1$. (В этом случае мы были бы вынуждены рассчитать погрешность в p'/p ; эту проблему мы рассмотрим в гл. 3.)

Наш вывод погрешности в $p - p'$, очевидно, применим к разности любых двух измеренных чисел. Таким образом, мы установили следующее общее правило:

Погрешность разности

Если величины x и y измерены с погрешностями δx и δy и если измеренные значения x и y используются для расчета разности $q = x - y$, то *погрешность в q есть сумма погрешностей в x и y :*

(2.18)

$$\delta q \approx \delta x + \delta y.$$

Мы использовали знак приближенного равенства ($\approx$), чтобы подчеркнуть два момента. Во-первых, у нас до сих пор нет точного определения погрешностей, с которыми мы имеем дело, так что было бы абсурдом утверждать, что δq *точно* равняется $\delta x + \delta y$. Во-вторых, в разд. 3.4 мы увидим, что погрешность δq часто несколько меньше, чем дает (2.18); лучшая оценка — это так называемая «квадратичная сумма» δx и δy , определенная в (3.13). Таким образом, знак « $\approx$ » в (2.18) использован как напоминание о том, что мы позднее заменим (2.18) лучшей оценкой.

Результат (2.18) — первый в серии правил *вычисления погрешностей в случае косвенных измерений*¹⁾. Когда мы рассчитаем величину q в единицах измеренных величин x и y , нам нужно будет узнать, как погрешности в x и y «распространяются» и приводят к погрешности в q . В гл. 3 детально обсуждается проблема расчета погрешностей в случае косвенных измерений.

2.6. Проверка пропорциональности с помощью графика

Многие физические закономерности предполагают, что одна величина должна быть пропорциональна другой. Закон

¹⁾ Автор использует термин «propagation of errors», что иногда переводится как «распространение ошибок». Однако в литературе на русском языке для этого выражения принята терминология «вычисление ошибок в случае косвенных измерений». — *Прим. перев.*

Гука утверждает, что растяжение пружины пропорционально силе, растягивающей ее; согласно закону Ньютона, ускорение тела пропорционально полной приложенной силе, и это только два из бесчисленного множества примеров. Многие эксперименты в учебной лаборатории организованы так, чтобы проверять этот вид пропорциональности.

Когда одна величина y пропорциональна некоторой другой x , то график зависимости y от x есть прямая линия, проходящая через начало координат. Таким образом, можно проверить пропорциональность y и x , если нанести измеренные значения y для данных x и посмотреть, лежат ли в действительности полученные точки на прямой линии, проходящей через начало координат. Поскольку прямая линия очень легко распознается, то такой путь является простейшим и эффективным способом проверки пропорциональности.

Чтобы проиллюстрировать такое использование графиков, представим себе эксперимент по проверке закона Гука. Этот закон, обычно записываемый в виде $F = kx$, утверждает, что растяжение пружины x пропорционально силе F , которая ее растягивает, т. е. $x = F/k$, где k есть коэффициент упругости пружины. Простой способ проверки этого закона — повесить пружину вертикально и подвешивать к ней различные массы m . В этом случае сила F есть вес груза mg , так что растяжение равно

$$x = \frac{mg}{k} = \left(\frac{g}{k}\right) m. \quad (2.19)$$

Растяжение x должно быть пропорционально нагрузке m , и график x от m должен представлять прямую линию, проходящую через начало координат.

Если мы будем измерять x для набора различных грузов m и откладывать на графике эту зависимость x от m , то в высшей степени невероятно, чтобы измеренные значения в *точности* легли на прямую линию. Предположим, например, что мы измеряем растяжение x для восьми различных грузов m и получаем результаты, представленные в табл. 2.3. Эти значения приведены на рис. 2.1, *a*, где мы также начертили возможную прямую линию, которая проходит через начало координат и примерно одинаково близка ко всем восьми точкам. Как и ожидалось, восемь точек не лежат точно на одной

Таблица 2.3. Нагрузка и растяжение

Нагрузка m , г (δm пренебрежимо мало)	200	300	400	500	600	700	800	900
Растяжение x , см ($\pm 0,3$)	1,1	1,5	1,9	2,8	3,4	3,5	4,6	5,4

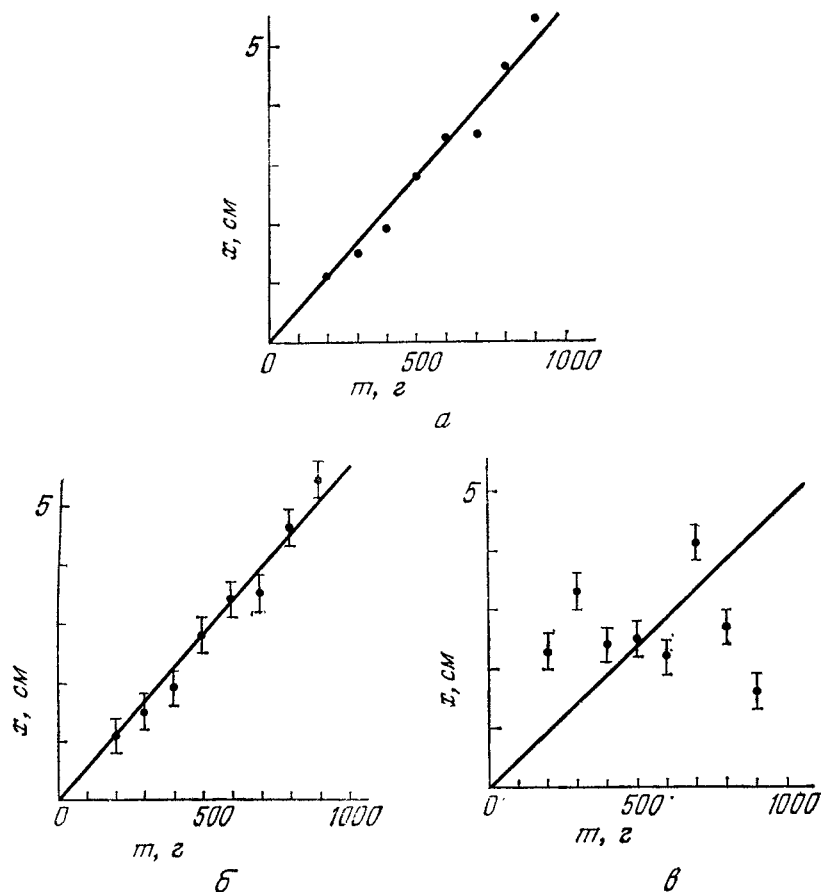


Рис. 2.1. Три графика, на которых изображена зависимость растяжения x пружины от нагрузки m .

a — данные табл. 23 без черточек ошибок; *б* — те же данные с черточками ошибок, которые показывают погрешности в x . (Погрешности в m предполагаются пренебрежимо малыми.) Эти данные находятся в согласии с ожидаемой пропорциональностью x и m ; *в* — другой набор данных, который не подтверждает пропорциональность x и m .

прямой. Возникает вопрос, обусловлено ли это экспериментальными погрешностями (как хотелось бы надеяться), или же мы наделали ошибок, а может быть, даже растяжение x не пропорционально m . Чтобы выяснить это, мы должны рассмотреть наши погрешности.

Естественно, что измеренные значения растяжения x и массы m подвержены некоторым погрешностям. Для простоты предположим сначала, что массы известны с очень высокой точностью, так что погрешность в m пренебрежимо мала.

С другой стороны, предположим, что все измерения x имеют погрешность порядка 0,3 см (как показано в табл. 2.3). Например, для груза в 200 г растяжение, вероятно, будет находиться где-то в интервале $1,1 \pm 0,3$ см. Наша первая экспериментальная точка ляжет на вертикальную линию $m = 200$ г в середине интервала между $x = 0,8$ и $x = 1,4$ см. Это отображено на рис. 2.1, б, где мы представили вертикальными черточками ошибок интервалы, в пределах которых вероятно лежит каждое значение. В данном случае мы сможем найти прямую, проходящую через начало координат и *через черточки ошибок или вблизи них*. На рис. 2.1, б приведена подобная прямая; таким образом, мы могли бы сделать вывод, что данные, на основании которых построен рис. 2.1, б, находятся в согласии с пропорциональностью между x и m .

Мы видели из уравнения (2.19), что наклон графика зависимости x от m есть g/k . Измеряя наклон прямой на рис. 2.1, б, можно найти коэффициент упругости пружины k . Проводя возможно более крутую и более пологую линии, которые все еще хорошо отображали бы экспериментальные данные, мы могли бы также найти погрешность для этого значения k (см. задачу 2.8).

Если бы наилучшая прямая проходила в стороне от большей доли черточек ошибок или на слишком большом расстоянии от некоторых из них (по сравнению с длиной интервала ошибок), то наши результаты *не согласовались* бы с пропорциональностью между x и m . Этот случай отображен на рис. 2.1, в. Результаты, представленные на этом графике, таковы, что мы должны были бы перепроверить наши измерения и вычисления (включая расчет погрешностей) и подумать, нет ли оснований к тому, что величина x может быть *не* пропорциональной m .

До сих пор мы предполагали, что погрешность в массе (значения которой откладываются по горизонтальной оси)

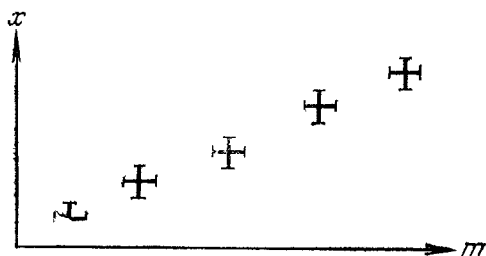


Рис. 2.2. Изображение результатов измерений с учетом погрешностей в x и m в виде крестиков, составленных из одной черточки ошибок для x и одной — для m .

ничтожна, а все погрешности могут содержаться только в x , что отображено вертикальными черточками ошибок. Для случая, когда как x , так и m подвержены заметным погрешностям, имеются различные способы отобразить их. Простейший состоит в том, чтобы начертить как вертикальные, так и горизонтальные черточки ошибок в каждой точке, причем длина половины каждой черточки должна равняться соответствующей погрешности, как показано на рис. 2.2. Каждый крест на этом графике соответствует одному измерению x и m , причем x , вероятно, лежит в интервале, определенном вертикальной чертой креста, а m , вероятно, в интервале, определенном горизонтальной чертой.

Несколько более сложная ситуация возникает тогда, когда ожидается, что одна физическая величина пропорциональна некоторой степени другой. Рассмотрим путь x , пройденный телом за время t при свободном падении. Этот путь равен $x = 1/2gt^2$ и пропорционален квадрату t . Если графически представить зависимость x от t , то экспериментальные точки должны лечь на параболу. Однако визуально трудно проверить, лежат ли точки на параболе (или на любой другой кривой, кроме прямой линии). Намного проще проследить зависимость $x \sim t^2$, когда график зависимости x от t^2 должен представлять собой прямую линию, проходящую через начало координат, и такой график позволит проверить, согласуются ли данные с прямой линией так же, как в предыдущем примере. Аналогично если величина x пропорциональна экспоненциальной функции e^{At} (где A — некоторая постоянная), то график зависимости $\ln x$ от t должен представлять собой прямую линию, и такой график легко позволит проверить пропорциональность $x \sim e^{At}$. (Этот вопрос мы обсудим в гл. 8.)

Имеются другие, неграфические, методы проверки пропорциональности двух величин. Например, если $x \sim m$, то отношение x/m должно быть постоянным. Можно было бы просто добавить к таблице значений x и m один дополнительный ряд или столбец, показывающий отношения x/m , и таким образом легко проверить, остаются ли эти отношения постоянными в пределах их погрешностей. Другой способ — с помощью программируемого калькулятора просчитать специально написанную программу, что позволит автоматически проверить, насколько хорошо совокупность измерений согласуется с прямой линией. Однако даже в случае применения каких-либо других способов проверки пропорциональности $x \sim m$ весьма желательно было бы воспользоваться также и графическим методом. Графики, подобные изображенным на рис. 2.1, б и в, ясно показывают, насколько хорошо предсказания подтверждаются измерениями; вычерчивание таких графиков помогает понять эксперимент и связанные с ним физические законы.

2.7. Относительные погрешности

Погрешность δx в измерении

$$(\text{измеренное значение } x) = x_{\text{наил}} \pm \delta x$$

показывает надежность или точность измерения. Однако погрешность сама по себе не раскрывает всей картины. Погрешность в 1 см для расстояния 1 км означала бы необычайно точное измерение, в то время как погрешность в 1 см для расстояния в 3 см означала бы лишь грубую оценку. Очевидно, что качество измерения характеризуется не только самой погрешностью δx , но также и *отношением* δx к $x_{\text{наил}}$, и это обстоятельство заставляет нас рассматривать *относительную погрешность*

$$\text{относительная погрешность} = \frac{\delta x}{|x_{\text{наил}}|}. \quad (2.20)$$

(Относительная погрешность также называется *точностью*.) В этом определении символ $|x_{\text{наил}}|$ обозначает абсолютную величину¹⁾ $x_{\text{наил}}$.

Чтобы избежать недоразумений с относительной погрешностью, саму погрешность δx иногда называют *абсолютной погрешностью*.

В большинстве серьезных измерений погрешность δx намного меньше измеряемой величины $x_{\text{наил}}$. Поскольку при этом относительная погрешность $\delta x/|x_{\text{наил}}|$ представляет собой обычно малое число, часто удобно умножать ее на 100 и приводить как *погрешность в процентах*. Например, результат измерения

$$\text{длина } l = 50 \pm 1 \text{ см} \quad (2.21)$$

имеет относительную погрешность

$$\frac{\delta l}{|l_{\text{наил}}|} = \frac{1}{50} = 0,02$$

¹⁾ Абсолютная величина $|x|$ числа x равна самому числу x , если x — положительная величина, и получается отбрасыванием знака минус, если x — отрицательная величина. Мы использовали абсолютную величину в (2.20), чтобы гарантировать, что относительная погрешность, подобно самой погрешности δx , всегда положительна независимо от того, положительна или отрицательна величина $x_{\text{наил}}$. На практике обычно делается так, чтобы измеряемые числа были положительными, и в этом случае знак абсолютно-го значения в (2.20) может быть опущен.

и погрешность, выраженную в процентах, 2%. Таким образом, результат (2.21) мог быть представлен как

$$\text{длина } l = 50 \text{ см} \pm 2 \text{ \%}.$$

Следует обратить внимание на то, что, в то время как абсолютная погрешность δl измеряется в тех же единицах, что и l , относительная погрешность $\delta l / |l_{\text{наил}}|$ является *безразмерной* величиной. Учет этого различия поможет вам избежать обычных ошибок, когда путают абсолютную погрешность с относительной.

Относительная погрешность приближенно характеризует качество измерений независимо от значения измеряемой величины. Относительная погрешность в 10% или около того — это обычно характеристика довольно грубых измерений. (Грубое измерение 10 см могло бы иметь погрешность в 1 см; грубое измерение 10 км могло бы иметь погрешность в 1 км.) Относительная погрешность в 1 или 2% характеризует уже довольно точные измерения, и это, пожалуй, лучшее, на что можно надеяться во многих экспериментах, выполняемых в учебной физической лаборатории. Относительных погрешностей, значительно меньших 1%, обычно трудно добиться, и они редки в учебной лаборатории.

Эти оценки, конечно, весьма условны; при некоторых чрезвычайно простых измерениях без труда можно получить относительную погрешность в 0,1%. С хорошей рулеткой расстояние в 3 м легко измерить с погрешностью в 0,3 см, или приблизительно 0,1%, с хорошими часами отрезок времени в один час легко измерить с погрешностью меньшей, чем секунда, или 0,03%. С другой стороны, для многих величин, которые очень трудно измерить, погрешность в 10% рассматривалась бы как экспериментальный триумф. Следовательно, большая относительная погрешность *не* означает, что измерение в научном смысле бесполезно. Многие важные измерения в истории физики имели экспериментальные погрешности 10% или более. В учебной физической лаборатории с оборудованием, которое позволяет получать результаты с минимальной погрешностью порядка нескольких процентов, можно изучать многие интересные физические закономерности.

2.8. Значащие цифры и относительные погрешности

Концепция относительной погрешности тесно связана с обычным понятием значащих цифр. Действительно, число значащих цифр в численном значении какой-то величины прибли-

Таблица 2.4. Приблизительное соответствие между значащими цифрами и относительной погрешностью

Число значащих цифр	Соответствующая относительная погрешность	
	лежит между	или очень приближенно равна
1	5 и 50%	10%
2	0,5 и 5%	1%
3	0,05 и 0,5%	0,1%

женно указывает на относительную погрешность этого значения. Например, рассмотрим два числа

510 и 0,51,

точность которых до двух значащих цифр удостоверена. Поскольку число 510 (с двумя значащими цифрами) означает

510 ± 5 , или $510 \pm 1\%$,

и 0,51 означает

$0,51 \pm 0,005$, или $0,51 \pm 1\%$,

то мы видим, что оба числа определены с точностью до 1%. Другими словами, утверждение, что числа 510 и 0,51 имеют две значащие цифры, эквивалентно высказыванию, что они определены с точностью до 1%. Аналогично число 510 с тремя значащими цифрами характеризовалось бы относительной погрешностью 0,1% и т. д.

К сожалению, это полезное соответствие является лишь приблизительным. Число 110, данное с двумя значащими цифрами, означает

110 ± 5 , или $110 \pm 5\%$,

в то время как число 910 (тоже с двумя значащими цифрами) означает

910 ± 5 , или $910 \pm 0,5\%$.

Мы видим, что относительная погрешность, связанная с двумя значащими цифрами, изменяется от 0,5 до 5% в зависимости от первой цифры рассматриваемого числа. Итог наших рассуждений отражен в табл. 2.4.

2.9. Умножение двух измеренных значений

Пожалуй, наиболее важная особенность в понятии относительной погрешности проявляется при умножении измеренных значений друг на друга. Например, чтобы найти импульс тела, мы могли бы измерить его массу m и скорость v , затем

перемножить их и получить импульс $p = mv$. Обе величины m и v обладают погрешностями, которые мы должны будем оценить. Затем возникнет задача найти погрешность в p , которая является следствием известных погрешностей в m и v .

Во-первых, для удобства запишем число в стандартном виде

$$(\text{измеренное значение } x) = x_{\text{наил}} \pm \delta x$$

и, используя понятие относительной погрешности,

$$(\text{измеренное значение } x) = x_{\text{наил}} \left(1 \pm \frac{\delta x}{|x_{\text{наил}}|} \right). \quad (2.22)$$

Например, если относительная погрешность составляет 3%, то, следуя (2.22), имеем

$$(\text{измеренное значение } x) = x_{\text{наил}} \left(1 \pm \frac{3}{100} \right),$$

т. е. погрешность в 3% означает, что x , вероятно, лежит где-то между значениями $x_{\text{наил}}$, умноженным на 0,97, и $x_{\text{наил}}$, умноженным на 1,03, т. е.

$$(0,97) \cdot x_{\text{наил}} \leq x \leq (1,03) \cdot x_{\text{наил}}.$$

Мы увидим, что это полезная форма представления числа, которое мы собираемся умножать.

Вернемся теперь к нашей задаче вычисления $p = mv$, когда m и v были измерены как

$$(\text{измеренное значение } m) = m_{\text{наил}} \left(1 \pm \frac{\delta m}{|m_{\text{наил}}|} \right) \quad (2.23)$$

и

$$(\text{измеренное значение } v) = v_{\text{наил}} \left(1 \pm \frac{\delta v}{|v_{\text{наил}}|} \right). \quad (2.24)$$

Поскольку $m_{\text{наил}}$ и $v_{\text{наил}}$ — наши наилучшие оценки для m и v , то наилучшая оценка для $p = mv$ есть

$$p_{\text{наил}} = (\text{наилучшая оценка для } p) = m_{\text{наил}} v_{\text{наил}}.$$

Наибольшие вероятные значения m и v даются выражениями (2.23) и (2.24) со знаком плюс. Таким образом, наибольшее вероятное значение для $p = mv$ есть

$$(\text{наибольшее значение } p) = m_{\text{наил}} v_{\text{наил}} \left(1 + \frac{\delta m}{|m_{\text{наил}}|} \right) \left(1 + \frac{\delta v}{|v_{\text{наил}}|} \right). \quad (2.25)$$

Наименьшее вероятное значение для p дается аналогичным выражением с двумя знаками минус. Теперь результат произведения скобок в (2.25) может быть представлен как

$$\begin{aligned} & \left(1 + \frac{\delta m}{|m_{\text{наил}}|} \right) \left(1 + \frac{\delta v}{|v_{\text{наил}}|} \right) = \\ & = 1 + \frac{\delta m}{|m_{\text{наил}}|} + \frac{\delta v}{|v_{\text{наил}}|} + \frac{\delta m}{|m_{\text{наил}}|} \frac{\delta v}{|v_{\text{наил}}|}. \end{aligned} \quad (2.26)$$

Поскольку две относительные погрешности $\delta t/|t_{\text{наил}}|$ и $\delta v/|v_{\text{наил}}|$ — малые числа (возможно, порядка нескольких процентов), то их произведение очень мало. Следовательно, последним членом в (2.26) можно пренебречь. Возвращаясь к (2.25), мы получаем

$$(\text{наибольшее значение } p) = t_{\text{наил}} v_{\text{наил}} \left(1 + \frac{\delta t}{|t_{\text{наил}}|} + \frac{\delta v}{|v_{\text{наил}}|} \right).$$

Наименьшее вероятное значение дается аналогичным выражением с двумя знаками минус. Наши измерения t и v приводят, следовательно, к значению $p = tv$, определяемому выражением

$$(\text{значение } p) = t_{\text{наил}} v_{\text{наил}} \left(1 \pm \left[\frac{\delta t}{|t_{\text{наил}}|} + \frac{\delta v}{|v_{\text{наил}}|} \right] \right).$$

Сравнивая это выражение с общей формой записи

$$(\text{значение } p_{\text{наил}}) = p_{\text{наил}} \left(1 \pm \frac{\delta p}{|p_{\text{наил}}|} \right),$$

мы видим, что наилучшая оценка для p есть $p_{\text{наил}} = t_{\text{наил}} v_{\text{наил}}$ (как нам уже известно) и что *относительная погрешность p равна сумме относительных погрешностей t и v* :

$$\frac{\delta p}{|p_{\text{наил}}|} = \frac{\delta t}{|t_{\text{наил}}|} + \frac{\delta v}{|v_{\text{наил}}|}.$$

Если, например, у нас были следующие измеренные значения для t и v :

$$t = 0,53 \pm 0,01 \text{ кг}$$

и

$$v = 9,1 \pm 0,3 \text{ м/с},$$

то наилучшая оценка для $p = tv$ равна

$$p_{\text{наил}} = t_{\text{наил}} v_{\text{наил}} = (0,53) \cdot (9,1) = 4,82 \text{ кг} \cdot \text{м/с}.$$

Чтобы рассчитать погрешность в p , мы сначала вычислим относительные погрешности

$$\frac{\delta t}{|t_{\text{наил}}|} = \frac{0,01}{0,53} = 0,02 = 2 \%$$

и

$$\frac{\delta v}{|v_{\text{наил}}|} = \frac{0,3}{9,1} = 0,03 = 3 \%.$$

Относительная погрешность в p есть сумма

$$\frac{\delta p}{|p_{\text{наил}}|} = 2 \% + 3 \% = 5 \%.$$

Если мы захотим узнать абсолютную погрешность в p , необходимо умножить полученное число на $p_{\text{наил}}$:

$$\delta p = \frac{\delta p}{p_{\text{наил}}} \cdot p_{\text{наил}} = 0,05 \cdot 4,82 = 0,241.$$

Затем, округляя δp и $p_{\text{наил}}$, получаем окончательный ответ

$$(\text{значение } p) = 4,8 \pm 0,2 \text{ кг} \cdot \text{м/с}.$$

Предыдущее рассмотрение применимо к любому произведению двух измеренных величин. Таким образом, мы получили второе общее правило для косвенных измерений. Если мы измеряем две величины и ищем их произведение, то погрешности в исходных двух величинах «распространяются» и образуют погрешность в их произведении. Эта погрешность дается следующим правилом:

Погрешность в произведении

Если величины x и y измерены с малыми относительными погрешностями $\delta x/|x_{\text{наил}}|$ и $\delta y/|y_{\text{наил}}|$ и если измеренные величины x и y используются для вычисления произведения $q = xy$, то *относительная погрешность q равна сумме относительных погрешностей x и y :*

(2.27)

$$\frac{\delta q}{|q_{\text{наил}}|} \approx \frac{\delta x}{|x_{\text{наил}}|} + \frac{\delta y}{|y_{\text{наил}}|}.$$

Мы использовали знак приближенного равенства в (2.27), поскольку, как и в случае правила вычисления погрешности для разности, мы позже заменим (2.27) более точным правилом. Следует подчеркнуть также две другие особенности этого правила. Во-первых, при выводе выражения (2.27) предполагалось, что относительные погрешности x и y должны быть достаточно малы, чтобы их произведением можно было пренебречь. На практике это условие почти всегда выполняется, так что мы всегда будем его предполагать. Тем не менее следует помнить, что если относительные погрешности ненамного меньше единицы, то правило (2.27) неприменимо. Во-вторых, даже если x и y имеют различную размерность, то в уравнение (2.27) входят безразмерные величины, поскольку все относительные погрешности безразмерны.

В физике мы постоянно перемножаем числа, поэтому очевидно, что правило (2.27) для нахождения погрешности в произведении играет важную роль в теории ошибок. В данный момент наша главная цель — подчеркнуть, что погрешность

любого произведения $q = xy$ наиболее просто выражается через относительные погрешности выражением типа (2.27).

Задачи

Напоминание: звездочка у номера задачи означает, что задача решается или ее ответ приводится в разделе «Ответы» в конце книги.

2.1 (разд. 2.1). В гл. 1 плотник представил результаты своего измерения высоты дверного проема в виде утверждения, что его наилучшая оценка высоты равна 210 см и что, по его убеждению, высота может составлять величину, лежащую где-то между 205 и 215 см. Перепишите этот результат в стандартной форме $x_{\text{наил}} \pm \delta x$. Прodelайте то же для измерений, отраженных соотношениями (1.1) и (1.2) и (1.4).

*2.2 (разд. 2.2). Перепишите следующие ответы в наиболее наглядном виде с нужным числом значащих цифр:

- а) измеренная высота = $5,03 \pm 0,04329$ м;
- б) измеренное время = $19,5432 \pm 1$ с;
- в) измеренный заряд = $-3,21 \cdot 10^{-19} \pm 2,67 \cdot 10^{-20}$ Кл;
- г) измеренная длина волны = $0,000000563 \pm 0,00000007$ м;
- д) измеренный импульс = $3,267 \cdot 10^3 \pm 42$ г·см/с.

*2.3 (разд. 2.3).

а. Студент измеряет плотность жидкости пять раз и получает результаты (в г/см³): 1,8; 2,0; 2,0; 1,9; 1,8. Что вы могли бы предположить о наилучшей оценке и погрешности, основываясь на его измерениях?

б. Ему сказали, что принятое значение равно 1,85 г/см³. Каково различие (между его наилучшей оценкой и принятым значением)? Считаете ли вы его значимым?

2.4 (разд. 2.5). Время десяти оборотов диска проигрывателя измеряется путем фиксирования моментов времени начала и конца вращения при помощи второсортных ручных часов с последующим вычитанием одной величины из другой. Если время начала и время конца вращения имеют погрешность по ± 1 с, то какова погрешность времени десяти оборотов?

*2.5 (разд. 2.5). В эксперименте по проверке закона сохранения момента импульса студент получил для начального и конечного моментов импульса (L и L') вращающейся системы результаты, представленные в табл. 2.5. Добавьте к табл. 2.5 дополнительный столбец для разности $L - L'$ и ее погрешности. Согласуются ли результаты студента с законом сохранения момента импульса?

2.6 (разд. 2.5). Экспериментатор измеряет массы M и m автомобиля и прицепа. Он приводит свои результаты в стандартной форме $M_{\text{наил}} \pm \delta M$ и $m_{\text{наил}} \pm \delta m$. Какова будет его наилучшая оценка для полной массы $M + m$? Рассматривая наибольшее и наименьшее вероятные значения полной массы, покажите, что его оценка погрешности полной массы равна.

Таблица 2.5. Момент импульса (в кг·м²/с)

Начальный L	Конечный L'	Начальный L	Конечный L'
$3,0 \pm 0,3$	$2,7 \pm 0,6$	25 ± 2	24 ± 2
$7,4 \pm 0,5$	$8,0 \pm 1$	32 ± 2	31 ± 2
$14,3 \pm 1$	$16,5 \pm 1$	37 ± 2	41 ± 2

Таблица 2.6. Значения высоты и скорости

h , м ($\pm 0,05$)	v^2 , м ² /с ²	h , м ($\pm 0,05$)	v^2 , м ² /с ²
0,4	7 ± 3	2,6	45 ± 5
0,8	17 ± 3	3,4	62 ± 5
1,4	25 ± 3	3,8	72 ± 6
2,0	38 ± 4		

сумме δM и δm . Приводите свои результаты последовательно и аргументированно, а не просто записывайте ответ.

2.7 (разд. 2.6). Используя данные задачи 2.5, постройте график зависимости конечного момента импульса L' от начального L для описанного там эксперимента. Нарисуйте вертикальные и горизонтальные черточки ошибок. (Как и при составлении всех графиков, отчетливо разметьте ваши оси, указывая названия величин и единицы измерения. Используйте соответствующую миллиметровую бумагу. Выберите масштаб таким образом, чтобы график заполнял разумную долю площади бумаги и включал начало координат.)

На какую кривую, по-вашему, будут ложиться все точки? Лежат ли точки на ожидаемой кривой (с учетом экспериментальных погрешностей)?

*2.8 (разд. 2.7). Если камень бросить вверх со скоростью v , он должен подняться до высоты h , определяемой уравнением $v^2 = 2gh$. В частности, величина v^2 должна быть пропорциональна h . Чтобы проверить это, студент измеряет v^2 и h для семи различных бросков и получает результаты, приведенные в табл. 2.6.

а. Постройте график зависимости v^2 от h , включая вертикальные и горизонтальные черточки ошибок. (Как обычно, разметьте оси координат, используйте миллиметровую бумагу и разумно выберите масштаб.) Согласуется ли ваш график с предсказанием, что $v^2 \sim h$?

б. Наклон вашего графика должен быть равен $2g$. Чтобы найти наклон, проведите наилучшую, по вашему мнению, прямую через начало координат и все остальные точки и затем определите ее наклон. Чтобы найти погрешность в определении наклона, проведите наиболее крутую и наименее крутую прямые, которые еще разумно совпадают с точками. Наклоны этих прямых дадут наибольшее и наименьшее вероятные значения наклона. Согласуются ли ваши результаты с принятым значением $2g = 19,6$ м/с²?

*2.9 (разд. 2.6).

а. В эксперименте с математическим маятником студент решает проверить, действительно ли период T не зависит от амплитуды A (определенной как наибольший угол, на который отклоняется маятник от вертикали во время его колебаний). Он получает результаты, представленные в табл. 2.7. Постройте график зависимости T от A .

Таблица 2.7. Амплитуда и период колебаний маятника

Амплитуда A , град	Период T , с	Амплитуда A , град	Период T , с
5 ± 2	$1,932 \pm 0,005$	40 ± 4	$2,01 \pm 0,01$
17 ± 2	$1,94 \pm 0,01$	53 ± 4	$2,04 \pm 0,01$
25 ± 2	$1,96 \pm 0,01$	67 ± 6	$2,12 \pm 0,02$

Обратите внимание на выбор масштаба. Если почувствуете затруднения, постройте два графика: один, включающий начало координат $A = 0$, $T = 0$, и второй, на котором показаны только значения T между 1,9 и 2,2 с. Должен ли студент сделать вывод, что период не зависит от амплитуды?

б. Рассмотрите, как изменились бы результаты «а», если бы все измеренные значения T имели погрешность $\pm 0,3$ с.

2.10 (разд. 2.7). Рассчитайте погрешности в процентах для пяти измерений, приведенных в задаче 2.2 (не забудьте округлить до разумного числа значащих цифр).

2.11 (разд. 2.7). С помощью деревянной линейки можно произвести отсчет с точностью до миллиметра, а с помощью измерительного микроскопа — до 0,1 мм. Предположим, вы хотите измерить длину 2 см с точностью 1 %. Можно ли это сделать с помощью деревянной линейки? А с помощью микроскопа?

*2.12 (разд. 2.7). Чтобы рассчитать ускорение тележки, студент измеряет ее начальную и конечную скорости v_i и v_f и вычисляет разность ($v_f - v_i$). Его данные для двух независимых испытаний (в см/с) приведены в табл. 2.8. Все четыре результата измерения характеризуются погрешностью 1 %.

а. Вычислите абсолютные погрешности всех четырех измерений, найдите разность ($v_f - v_i$) и ее погрешность для каждого испытания.

б. Вычислите погрешность в процентах для каждого из двух значений ($v_f - v_i$). (Ваши ответы в этом задании, особенно в случае второго испытания, иллюстрируют отрицательные последствия методики измерения малых чисел с помощью разности двух гораздо больших чисел.)

2.13 (разд. 2.8).

а. Калькулятор студента показывает результат 123,123. Если студент решил, что это число в действительности имеет только три значащие цифры, оцените, каковы его абсолютная и относительная погрешности.

б. Сделайте то же для числа 0,123123.

в. Сделайте то же для числа 321,321.

г. Лежит ли относительная погрешность в интервале, ожидаемом для случая трех значащих цифр?

*2.14 (разд. 2.9).

а. Студент измеряет две величины a и b и получает $a = 11,5 \pm 0,2$ см и $b = 25,4 \pm 0,2$ см. Затем он вычисляет произведение $q = ab$. Получите его ответ и приведите абсолютное значение его погрешности, а также погрешность в процентах.

б. Повторите действия задания «а» для измерений $a = 10 \pm 1$ см и $b = 27,2 \pm 0,1$ с

в. Повторите задание «а» для $a = 0,8 \text{ м} \pm 8 \%$ и $b = 1,5 \text{ кг} \pm 2 \%$.

*2.15 (разд. 2.9).

а. Студент измеряет два числа x и y и находит $x = 10 \pm 1$, $y = 20 \pm 1$. Какова его наилучшая оценка для произведения $q = xy$? Используя наибольшие вероятные значения для x и y (11 и 21), вычислите

Таблица 2.8. Начальные и конечные скорости

	v_i	v_f
Первое испытание	14,0	18,0
Второе испытание	19,0	19,6

наибольшее вероятное значение для q . Аналогично найдите наименьшее вероятное значение q и, следовательно, интервал, в котором, вероятно, лежит q . Сравните ваш результат с тем, что дает правило (2.27).

- б. Сделайте то же для измерений $x = 10 \pm 8$, $y = 20 \pm 15$.

Напоминаем: правило (2.27) было получено в предположении, что относительные погрешности намного меньше единицы.

2.16 (разд. 2.9). Согласно известному правилу, при перемножении двух чисел результат будет более надежным, если его округлить до количества значащих цифр в наименее точном из двух исходных чисел.

- а. Используя правило (2.27) и тот факт, что значащие цифры определяют относительную погрешность, докажите, что это «известное правило» *приближенно* верно. (Для определенности рассмотрите случай, когда наименее точное число имеет две значащие цифры.)
- б. Покажите на примере, что ответ в действительности несколько менее точен, чем дает «известное правило». (Это особенно справедливо при перемножении одинаковых чисел.)

Глава 3

Погрешности в косвенных измерениях

Большинство физических величин обычно невозможно измерить непосредственно, и их определение включает два различных этапа. Сначала измеряют одну или более величин $x, y, \dots$, которые могут быть непосредственно измерены и с помощью которых можно вычислить интересующую нас величину. Затем, используя измеренные значения $x, y, \dots$, вычисляют саму искомую величину. Например, чтобы найти площадь прямоугольника, обычно измеряют его длину l и высоту h и затем рассчитывают его площадь A по формуле $A = lh$. Аналогично наиболее очевидный способ определения скорости v некоторого тела состоит в том, чтобы измерить путь d , пройденный телом, и затраченное на это время t , а затем вычислить v по формуле $v = d/t$. Для читателя, проработавшего какое-то время в учебной лаборатории, не составит труда привести и другие примеры. Действительно, если хоть немного задуматься, то станет ясно, что почти все важные измерения включают эти два различных этапа, состоящих из простых измерений и последующих расчетов.

Если измерение включает эти два этапа, то и оценка погрешностей также включает их. Сначала надо оценить погрешности в величинах, которые измеряются непосредственно, а затем определить, как эти погрешности «распространяются» в расчетах и приводят к погрешности в конечном результате¹⁾. Это «распространение ошибок», или расчет погрешностей в случае косвенных измерений, составляет главную тему данной главы.

Фактически мы уже рассмотрели некоторые примеры расчета погрешностей в случае косвенных измерений в гл. 2.

¹⁾ В гл. 4 мы рассмотрим другой способ, с помощью которого иногда вычисляют результирующую погрешность. Если все измерения можно повторить несколько раз и если есть уверенность, что все погрешности по природе случайны, то погрешность в интересующей величине можно оценить, исследуя разброс в ответах. Но даже когда этот метод применим, его обычно лучше использовать как проверку двухэтапной процедуры, описанной в этой главе.

В разд. 2.5 обсуждался случай, когда измеряются два числа x и y , которые используются для расчета разности $q = x - y$. Мы нашли, что погрешность в q равна сумме $\delta q \approx \delta x + \delta y$ погрешностей в x и y . В разд. 2.9 мы рассмотрели произведение $q = xy$ и в задаче 2.6 предложили рассмотреть сумму $q = x + y$. Эти случаи обсуждаются также в разд. 3.2. В конце главы мы рассмотрим более общие методы расчета погрешностей в косвенных измерениях и приведем несколько примеров.

В разд. 3.1, прежде чем коснуться расчета погрешностей в косвенных измерениях, мы кратко обсудим оценку погрешностей в величинах, которые измеряются непосредственно. Будет сделан обзор методов, рассмотренных в гл. 1, и затем рассмотрено еще несколько примеров оценки погрешности в прямых измерениях.

Начиная с разд. 3.2, рассматриваются погрешности в косвенных измерениях. Мы обнаружим, что почти все проблемы расчета погрешностей в косвенных измерениях могут быть решены с помощью трех простых правил. Мы сформулируем также единственное, более сложное, правило, которое пригодно во всех случаях и из которого могут быть получены три более простых правила.

Это довольно длинная глава. Однако читатель может опустить два последних раздела без какой-либо потери непрерывности изложения.

3.1. Погрешности в прямых измерениях

Почти все прямые измерения включают считывание со шкалы (например, линейки, часов или вольтметра) или с цифрового табло (например, цифровых часов или цифрового вольтметра). Некоторые проблемы считывания со шкалы уже обсуждались в разд. 1.5. Иногда главными источниками погрешностей служат считывание со шкалы и необходимость интерполяции между метками шкалы. В таких ситуациях разумная оценка погрешностей не составляет труда. Например, если кто-то должен измерить некоторую точно определенную длину l при помощи линейки, проградуированной миллиметровыми делениями, то он мог бы вполне разумно решить, что длина может быть измерена до ближайшего миллиметра, и не лучше. В этом случае погрешность δl составляла бы $\delta l = 0,5$ мм. Если метки шкалы отстоят друг от друга дальше (например, на 5 мм), то экспериментатор вполне разумно мог бы решить, что он может считывать, например, с точностью до одной пятой деления. В любом случае погрешности считывания со шкалы, очевидно, могут быть довольно легко и разумно оценены.

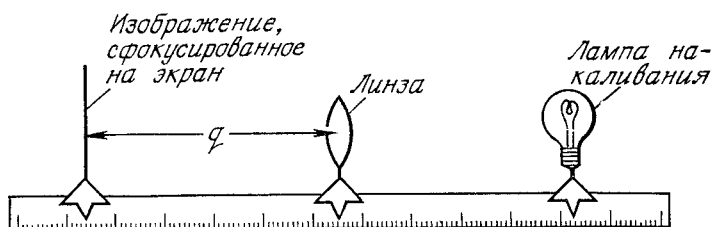


Рис. 3.1. Фокусирование изображения лампы накаливания, помещенной справа, на экран, установленный слева.

К сожалению, часто имеются другие источники ошибок, которые гораздо более существенны, чем любые трудности, обусловленные считыванием со шкалы. При измерении расстояния между двумя точками главной проблемой может стать определение истинного положения этих точек. Например, в оптическом эксперименте желательно измерить расстояние q от центра линзы до сфокусированного изображения, как показано на рис. 3.1. На практике линза обычно имеет толщину несколько миллиметров, так что определение ее центра вызовет затруднения, а если линза смонтирована в массивной оправе, как это часто бывает, то определение центра становится еще более сложной задачей. Более того, может оказаться, что изображение будет хорошо сфокусированным на длине порядка многих миллиметров. Даже если вся система смонтирована на оптической скамье, которая проградуирована отчетливыми миллиметровыми делениями, погрешность в расстоянии от линзы до изображения могла бы, следовательно, легко составлять величину порядка сантиметра. Поскольку погрешность обусловлена тем, что положение двух представляющих интерес точек точно не определено, то проблема такого рода называется *проблемой определения*.

Этот пример указывает на серьезную опасность при оценке погрешностей. Если смотреть только на шкалу и забыть о других источниках погрешностей, то можно очень сильно недооценить полную погрешность. Фактически наиболее частая ошибка начинающего студента — это игнорирование некоторых источников погрешности и, следовательно, *недооценка* погрешностей часто на порядок или более. Однако важно и не *переоценивать* ошибки. Экспериментатор, который решил бы перестраховаться, приводя избыточные погрешности, может избежать неприятной несогласованности во всех измерениях, но его измерения не будут представлять большого интереса для других. Очевидно, в идеальном случае следовало бы учесть все возможные источники погрешности и аккуратно

оценить их влияние, что обычно не так трудно, как может показаться.

Считывание с цифрового табло легче считывания с обычной шкалы. Если такой прибор исправен, он показывает только значащие цифры. Цифровые часы, которые показывают секунды с двумя знаками после запятой, могли бы показать время t , равное 8,03 с, что означает (в худшем случае), что

$$t = 8,03 \pm 0,01 \text{ с.}$$

В зависимости от того, как устроены часы, погрешность могла бы быть и вдвое меньше, т. е. 8,03 может означать $8,03 \pm 0,005$.

Цифровое табло даже в большей степени, чем обычная шкала, может создать ложное впечатление большой точности. Например, кто-то мог бы использовать цифровые часы для определения времени движения грузов на машине Атвуда или подобном приборе. Если часы показывают три цифры после запятой, то они могут показать время $t = 8,036$ с, и в этом случае время движения есть

$$t = 8,036 \pm 0,001 \text{ с.} \quad (3.1)$$

Однако осторожный студент, который повторяет эксперимент при как можно более идентичных условиях, мог бы найти, что результат второго измерения равен

$$t = 8,13 \text{ с.}$$

Одно вероятное объяснение этого различия состоит в том, что погрешности в процедуре запуска изменяют начальные условия и, следовательно, время движения. В любом случае ясно, что точность, представленная в (3.1), до абсурда хороша. Судя по проделанным до сих пор двум измерениям, более реалистичный результат был бы

$$t = 8,07 \pm 0,05 \text{ с.}$$

Этот пример возвращает нас к другому моменту, упомянутому в гл. 1. Если измерения могут быть повторены, их обычно следует выполнить несколько раз. Получающийся разброс величин часто служит хорошей мерой погрешностей, и среднее измеренных значений всегда более заслуживает доверия, чем результат любого единичного измерения. В гл. 4 и 5 мы коснемся вопросов статистической обработки многократных измерений. Сейчас лишь подчеркнем, что если измерение можно повторить, его следует повторить как для получения более точного результата (при усреднении), так и, что даже более важно, в связи с возможностью оценить погрешности. К сожалению, как уже было отмечено в гл. 1, повторение измерений не всегда приводит к обнаружению погрешностей. Если измеряемая величина подвержена некоторым систематиче-

ским ошибкам, которые смещают все результаты в одну и ту же сторону (подобно часам, которые отстают), то разброс в результатах не будет отражать этой систематической ошибки. Исключение таких систематических ошибок потребует тщательной проверки калибровки приборов и других процедур.

Наконец, имеется еще один, совсем особый тип измерения, для которого погрешности легко могут быть оценены. Некоторые эксперименты включают счет событий, которые происходят случайно, но с определенной средней скоростью. Например, в образце радиоактивного вещества каждое индивидуальное ядро распадается в случайный момент времени, но существует определенная средняя скорость распада, согласно которой мы бы ожидали увидеть какое-то число распадов в единицу времени, происходящих во всем образце. Мы можем попытаться измерить этот средний темп, наблюдая число распадов, происходящих за определенный временной интервал, например за одну минуту. (Это может быть сделано, например, с помощью счетчика Гейгера, который регистрирует заряженные частицы, испускаемые при распаде каждым ядром.) Предположим, мы нашли, что произошло ν распадов, пока мы считали их в течение одной минуты. Поскольку распады происходят случайным образом, мы не можем быть уверены, что ν есть действительно верное среднее число распадов, которое ожидается за одну минуту. Вопрос, конечно, заключается в том, насколько недостоверна величина ν как мера ожидаемого среднего числа событий.

Теория вопросов, связанных с таким подсчетом, обсуждается в гл. 11, но ответ исключительно прост и может быть приведен сейчас. Если мы сосчитаем число событий за время T и получим в результате ν , то в качестве меры ожидаемого среднего числа событий за время T наш результат ν имеет погрешность $\sqrt{\nu}$. Таким образом, наш вывод (основанный на этом одном измерении) следует представить в виде

$$(\text{среднее число событий за время } T) = \nu \pm \sqrt{\nu}. \quad (3.2)$$

Например, если бы мы насчитали 15 распадов от образца радиоактивного урана за одну минуту, то могли бы сделать вывод, что в среднем в нашем образце происходит $15 \pm \sqrt{15}$, или 15 ± 4 распадов в минуту.

3.2. Суммы и разности; произведения и частные

До конца этой главы мы будем предполагать, что мы провели измерения одной или более величин x , y , ... с соответствующими погрешностями δx , δy , ... и что теперь мы хотим по измеренным значениям x , y , ... вычислить величину q , ко-

торая нас в действительности интересует. Расчет величины q обычно выполняется непосредственно; проблема, которую мы должны обсудить, заключается в том, каким образом погрешности $\delta x, \delta y, \dots$, «распространяясь» через вычисления, приводят к погрешности δq косвенных измерений окончательной величины q .

Суммы и разности

В гл. 2 мы обсуждали, что происходит, когда измеряются две величины x и y и вычисляется их сумма $x + y$ или их разность $x - y$. Чтобы оценить погрешность как в сумме, так и в разности, мы должны были только определить их наибольшие и наименьшие вероятные значения. Наибольшее и наименьшее вероятные значения x равны $x_{\text{наил}} \pm \delta x$, а аналогичные величины для y равны $y_{\text{наил}} \pm \delta y$. Следовательно, наибольшее вероятное значение $x + y$ есть

$$x_{\text{наил}} + y_{\text{наил}} + (\delta x + \delta y)$$

и наименьшее вероятное значение

$$x_{\text{наил}} + y_{\text{наил}} - (\delta x + \delta y).$$

Таким образом, наилучшая оценка для $q = x + y$ есть

$$q_{\text{наил}} = x_{\text{наил}} + y_{\text{наил}}$$

и ее погрешность

$$\delta q \approx \delta x + \delta y. \quad (3.3)$$

Аналогичные аргументы (вы должны быть уверены, что сможете воспроизвести их) приводят к тому, что погрешность в разности $x - y$ дается той же самой формулой (3.3). Таким образом, погрешность как в сумме $x + y$, так и в разности $x - y$ представляет собой сумму $\delta x + \delta y$ погрешностей в x и y .

В случае нескольких чисел $x, \dots, w$, которые надо складывать или вычитать, повторные применения (3.3) приводят к следующему правилу:

Погрешности в суммах и разностях

Если несколько величин $x, \dots, w$ измерены с погрешностями $\delta x, \dots, \delta w$ и используются для вычисления

$$q = x + \dots + z - (u + \dots + w),$$

то погрешность в рассчитанной величине q есть сумма

$$\delta q \approx \delta x + \dots + \delta z + \delta u + \dots + \delta w \quad (3.4)$$

всех исходных погрешностей.

Другими словами, когда складывают или вычитают любое число измеренных величин, то погрешности этих величин всегда *складываются*. Как и прежде, мы используем знак $\approx$, чтобы подчеркнуть, что вскоре мы улучшим это правило.

Пример

В качестве простого примера применения правила (3.4) предположим, что экспериментатор смешивает жидкости из двух фляг, предварительно измерив по отдельности массы этих наполненных и затем пустых фляг и получив в результате

M_1 = масса первой фляги и ее содержимого = 540 ± 10 г;

m_1 = масса первой пустой фляги = 72 ± 1 г;

M_2 = масса второй фляги и ее содержимого = 940 ± 20 г;

m_2 = масса второй пустой фляги = 97 ± 1 г.

Затем он рассчитывает полную массу жидкости как

$$M = M_1 - m_1 + M_2 - m_2 = (540 - 72 + 940 - 97) \text{ г} = 1311 \text{ г}.$$

В соответствии с правилом (3.4) погрешность в его результате есть сумма всех четырех погрешностей:

$$\delta M \approx \delta M_1 + \delta m_1 + \delta M_2 + \delta m_2 = (10 + 1 + 20 + 1) \text{ г} = 32 \text{ г}.$$

Таким образом, его конечный результат (надлежащим образом округленный) имеет вид

$$\text{полная масса жидкости} = 1310 \pm 30 \text{ г}.$$

Заметьте, что существенно меньшие погрешности в массах пустых фляг вносят ничтожную добавку в конечную погрешность. Это очень важный эффект, который мы обсудим позднее. С опытом студент сможет научиться заранее выявлять те погрешности, которые пренебрежимо малы и поэтому могут быть исключены из рассмотрения. Часто это может очень существенно упростить расчет погрешностей.

Произведения и частные

В разд. 2.9 мы рассмотрели погрешности в произведении $q = xy$ двух измеренных величин. Мы нашли, что при условии малых относительных погрешностей исходных величин *относительная* погрешность в $q = xy$ есть сумма *относительных* погрешностей в x и y . Вместо того чтобы повторно обсудить вывод этого результата, рассмотрим сейчас подобный же случай частного $q = x/y$. Как мы увидим, погрешность в частном дается тем же самым правилом, что и для произведения, т. е. относительная погрешность в $q = x/y$ равна сумме относительных погрешностей в x и y .

Поскольку погрешности в произведениях и частных наилучшим образом выражаются в терминах относительных погрешностей, то удобно для последующего ввести более краткое обозначение. Напомним, что если мы измеряем некоторую величину x как

$$(\text{измеренное значение } x) = x_{\text{наил}} \pm \delta x,$$

т. е. обычным образом, то относительная погрешность в x определяется как

$$(\text{относительная погрешность в } x) = \frac{\delta x}{|x_{\text{наил}}|}.$$

(Абсолютное значение в знаменателе всегда обеспечивает положительность относительной погрешности, даже когда величина $x_{\text{наил}}$ отрицательна.) Поскольку выражение $\delta x/|x_{\text{наил}}|$ неудобно писать и читать, с этого момента мы будем использовать его сокращенное написание, в котором опустим индекс «наил»:

$$(\text{относительная погрешность в } x) = \frac{\delta x}{|x|}.$$

Результат измерения любой величины x может быть выражен через его относительную погрешность $\delta x/|x|$ как

$$(\text{значение } x) = x_{\text{наил}} (1 \pm \delta x/|x|).$$

Следовательно, значение $q = x/y$ может быть переписано как

$$(\text{значение } q) = \frac{x_{\text{наил}}}{y_{\text{наил}}} \frac{1 \pm \delta x/|x|}{1 \pm \delta y/|y|}.$$

Наша задача теперь — найти экстремальные вероятные значения второго множителя справа. Этот множитель максимален, например, когда числитель равен его *наибольшему* значению, $1 + \delta x/|x|$, а знаменатель равен его *наименьшему* значению, $1 - \delta y/|y|$. Таким образом, наибольшее вероятное значение для $q = x/y$ равно

$$(\text{наибольшее значение } q) = \frac{x_{\text{наил}}}{y_{\text{наил}}} \frac{1 + \delta x/|x|}{1 - \delta y/|y|}. \quad (3.5)$$

Последний множитель в выражении (3.5) имеет форму $(1+a)/(1-b)$, где числа a и b обычно малы (т. е. много меньше единицы). Эту форму можно упростить с помощью двух приближений. Во-первых, поскольку число b мало, биномиальная теорема¹⁾ дает

$$\frac{1}{(1-b)} \approx 1 + b. \quad (3.6)$$

¹⁾ Биномиальная теорема позволяет выразить $1/(1-b)$ через бесконечный ряд $1 + b + b^2 + b^3 + \dots$. Если b много меньше 1, то $1/(1-b) \approx 1 + b$, как в (3.6). Читатель, незнакомый с биномиальной теоремой, может найти дополнительные сведения в задаче 3.7.

Следовательно,

$$\frac{1+a}{1-b} \approx (1+a)(1+b) = 1+a+b+ab \approx 1+a+b,$$

где в последнем выражении мы пренебрегли произведением двух малых величин ab . Возвращаясь к (3.5) и используя эти приближения, мы получаем для наибольшего вероятного значения $q = x/y$

$$(\text{наибольшее значение } q) = \frac{x_{\text{наил}}}{y_{\text{наил}}} \left(1 + \frac{\delta x}{|x|} + \frac{\delta y}{|y|} \right).$$

Аналогичное рассмотрение показывает, что наименьшее вероятное значение дается подобным же выражением с двумя знаками минус. Объединяя эти результаты, находим

$$(\text{значение } q) = \frac{x_{\text{наил}}}{y_{\text{наил}}} \left(1 \pm \left[\frac{\delta x}{|x|} + \frac{\delta y}{|y|} \right] \right).$$

Сравнивая это выражение со стандартной записью

$$(\text{значение } q) = q_{\text{наил}} \left(1 \pm \frac{\delta q}{|q|} \right),$$

мы видим, что наилучшее значение q равно $q_{\text{наил}} = x_{\text{наил}}/y_{\text{наил}}$, как мы могли бы ожидать, и что относительная погрешность есть

$$\frac{\delta q}{|q|} \approx \frac{\delta x}{|x|} + \frac{\delta y}{|y|}. \quad (3.7)$$

Мы приходим к выводу, что при делении или умножении двух измеренных значений x и y относительная погрешность результата равна сумме относительных погрешностей x и y , как в (3.7). Если теперь умножать или делить целый ряд чисел, то повторные применения этого результата приведут нас к следующему общему правилу:

Погрешность в произведениях и частных

Если несколько величин $x, \dots, w$ измерены с малыми погрешностями $\delta x, \dots, \delta w$ и измеренные значения используются для расчета

$$q = \frac{x \times \dots \times z}{u \times \dots \times w},$$

то относительная погрешность рассчитанной величины q равна сумме

$$\begin{aligned} \frac{\delta q}{|q|} \approx & \frac{\delta x}{|x|} + \dots + \frac{\delta z}{|z|} + \\ & + \frac{\delta u}{|u|} + \dots + \frac{\delta w}{|w|} \end{aligned} \quad (3.8)$$

относительных погрешностей в $x, \dots, w$.

Итак, при умножении или делении величин *относительные погрешности складываются*.

Пример

При съемке местности иногда приходится определять недоступную непосредственному измерению длину l (такую, как высота большого дерева) при помощи измерений трех других длин l_1, l_2, l_3 , которые дают

$$l = \frac{l_1 l_2}{l_3}.$$

Предположим, что мы выполняем такой эксперимент и получаем результаты (в метрах)

$$l_1 = 50 \pm 0,5; \quad l_2 = 1,5 \pm 0,03; \quad l_3 = 5,0 \pm 0,2.$$

Наша наилучшая оценка для l равна

$$l_{\text{наил}} = \frac{50 \cdot 1,5}{5} = 15 \text{ м.}$$

В соответствии с (3.8) относительная погрешность этого результата равна сумме относительных погрешностей в l_1, l_2, l_3 , которые равны соответственно 1, 2 и 4%. Таким образом,

$$\frac{\delta l}{l} \approx \frac{\delta l_1}{l_1} + \frac{\delta l_2}{l_2} + \frac{\delta l_3}{l_3} = (1 + 2 + 4) \% = 7\%.$$

и наш окончательный результат имеет вид

$$l = 15 \pm 1 \text{ м.}$$

Измеренная величина умножается на точное число

Два важных частных случая применения правила (3.8) заслуживают отдельного упоминания. Во-первых, предположим, что мы измеряем величину x и используем ее для вычисления произведения $q = Bx$, где число B не содержит погрешности. Например, мы могли бы измерять диаметр окружности и затем вычислять ее периметр как $c = \pi \times d$, или мы могли бы измерять толщину T 100 идентичных листов бумаги и затем определять толщину одного листа как $t = (1/100) \times T$. В соответствии с правилом (3.8) относительная погрешность в $q = Bx$ равна сумме относительных погрешностей для величин B и x . Поскольку $\delta B = 0$, то

$$\frac{\delta q}{|q|} = \frac{\delta x}{|x|}.$$

Умножая на $|q| = |Bx|$, мы находим, что $\delta q = |B|\delta x$, т. е. получаем следующее полезное правило:

**Измеренная величина умножается
на точное число**

Если величина x измеряется с погрешностью δx и используется для вычисления произведения

$$q = Bx,$$

в котором B не имеет погрешности, то погрешность в q равна $|B|$, умноженному на погрешность в x :

$$\delta q = |B|\delta x. \quad (3.9)$$

Это правило особенно полезно в случае, когда надо измерять что-то необычно малое, но имеющееся в большом количестве, такое, как толщина листа бумаги или время оборота быстро вращающегося колеса. Например, если мы измеряем толщину T 100 листов бумаги и получаем результат

$$\text{толщина 100 листов} = T = 30 \pm 3 \text{ мм},$$

то отсюда сразу же вытекает, что толщина t одного листа равна

$$\text{толщина одного листа} = t = \frac{1}{100} \times T = 0,3 \pm 0,03 \text{ мм}.$$

Заметьте, как такой прием (измерение толщины нескольких идентичных листов и последующее деление на число листов) делает легко выполнимым измерение, которое в противном случае потребовало бы довольно сложного оборудования, и приводит также к исключительно малой погрешности. Конечно, необходима уверенность, что все листы имеют одинаковую толщину.

Возведение в степень

Второй частный случай правила (3.8) касается оценки степени некоторой измеренной величины. Например, мы могли бы измерять скорость v некоторого тела и затем для определения кинетической энергии $(1/2)mv^2$ вычислять v^2 . Поскольку $v^2 = v \times v$, то из (3.8) следует, что относительная погрешность в v^2 равна удвоенной относительной погрешности в v .

Если обобщить, то из (3.8) ясно, что общее правило для любой целой степени будет следующим:

Погрешность при возведении в степень

Если величина x измеряется с погрешностью δx и измеренное значение используется для вычисления степени этого числа

$$q = x^n,$$

то относительная погрешность в q в n раз больше относительной погрешности в x ,

$$\frac{\delta q}{|q|} = n \frac{\delta x}{|x|}. \quad (3.10)$$

Наш вывод этого правила требовал, чтобы n было целым и положительным числом. Фактически, однако, правило можно обобщить на случай *любых* показателей степени n [см. (3.26)].

Пример

Предположим, что студент определяет ускорение свободного падения g , измеряя время t падения камня с высоты h . После нескольких измерений времени он находит

$$t = 1,6 \pm 0,1 \text{ с}$$

и измеряет высоту h как

$$h = 14,1 \pm 0,1 \text{ м.}$$

Поскольку h определяется известной формулой $h = (1/2)gt^2$, то он вычисляет g как

$$g = \frac{2h}{t^2} = \frac{2 \cdot 14,1 \text{ м}}{(1,6 \text{ с})^2} = 11 \text{ м/с}^2.$$

Погрешность этого результата может быть найдена при помощи сформулированного выше правила. Здесь нам необходимо знать относительные погрешности в каждом из множителей выражения $g = 2h/t^2$, используемого для расчета g . Множитель 2 не содержит погрешности. Относительные погрешности в h и t равны

$$\frac{\delta h}{h} = \frac{0,1}{14,1} = 0,7 \%$$

и

$$\frac{\delta t}{t} = \frac{0,1}{1,6} = 6,3 \%.$$

В соответствии с правилом (3.10) относительная погрешность в t^2 в 2 раза больше, чем в t . Следовательно, применяя пра-

вило (3.8) для произведений и частных к формуле $g = 2h/t^2$, мы получим для относительной погрешности

$$\frac{\delta g}{g} = \frac{\delta h}{h} + 2 \frac{\delta t}{t} = 0,7 \% + 2 \cdot (6,3 \%) = 13,3 \%, \quad (3.11)$$

и, следовательно, погрешность

$$\delta g = (11 \text{ м/с}^2) \cdot \frac{13,3}{100} = 1,46 \text{ м/с}^2.$$

Таким образом, окончательный результат нашего студента (необходимым образом округленный) составляет

$$g = 11 \pm 1 \text{ м/с}^2.$$

Этот пример показывает, насколько простой часто может быть оценка погрешностей. Отсюда видно также, каким образом теория ошибок не только позволяет оценить величину погрешностей, но подсказывает пути их уменьшения. В данном случае из (3.11) ясно, что наибольший вклад в погрешность обусловлен измерением времени. Если мы хотим иметь более точное значение g , то необходимо улучшить именно измерение t ; любая попытка улучшить измерение h была бы в значительной мере напрасным усилием.

3.3. Независимые погрешности в сумме

Правила, которые мы пока нашли, могут быть сформулированы кратко: когда измеряемые величины складываются или вычитаются, *погрешности складываются*; когда измеряемые величины умножаются или делятся, *складываются относительные погрешности*. В этом и следующем разделах мы рассмотрим, как при определенных условиях погрешности, рассчитанные на основании этих правил, могут оказаться неоправданно большими. Точнее, мы увидим, что если исходные погрешности *независимы и случайны*, то более реалистичная (и меньшая) оценка окончательной погрешности дается аналогичными правилами, в которых погрешности (или относительные погрешности) *складываются квадратично* (т. е. в соответствии с процедурой, которую мы вскоре определим).

Рассмотрим сначала вычисление суммы $q = x + y$ двух чисел x и y , которые были измерены обычным образом:

$$(\text{измеренное значение } x) = x_{\text{наил}} \pm \delta x$$

и аналогично для y . Способ, который использовался в последнем разделе, выглядел следующим образом. Во-первых, наилучшая оценка $q = x + y$ есть, очевидно, $q_{\text{наил}} = x_{\text{наил}} + y_{\text{наил}}$. Во-вторых, поскольку наибольшие вероятные значения для x и y равны соответственно $x_{\text{наил}} + \delta x$ и $y_{\text{наил}} + \delta y$, то

наибольшее вероятное значение величины q есть

$$x_{\text{наил}} + y_{\text{наил}} + \delta x + \delta y. \quad (3.12)$$

Аналогично наименьшее вероятное значение q есть

$$x_{\text{наил}} + y_{\text{наил}} - \delta x - \delta y.$$

Отсюда мы делаем вывод, что величина q , вероятно, лежит между этими двумя значениями, и погрешность в q равна

$$\delta q \approx \delta x + \delta y.$$

Чтобы увидеть, почему эта формула скорее всего переоценивает δq , рассмотрим, в каком случае фактическая величина q сравнивается с наибольшим значением (3.12). Очевидно, это может случиться, если мы недооценили x на полную величину δx и недооценили y на полную величину δy . Однако это весьма маловероятно. Если x и y измеряются независимо и наши ошибки случайны по природе, в 50% случаев *недооценка* x будет сопровождаться *переоценкой* y или *наоборот*. Тогда ясно, что вероятность недооценки как x , так и y на полные величины δx и δy довольно мала. Следовательно, значение $\delta q \approx \delta x + \delta y$ переоценивает нашу возможную ошибку.

А что же тогда будет лучшей оценкой для δq ? Это зависит от того, что мы понимаем под ошибкой (т. е. что мы подразумеваем, утверждая, что q , «вероятно», лежит где-то между $q_{\text{наил}} - \delta q$ и $q_{\text{наил}} + \delta q$). Это также зависит от того, каковы статистические законы, которым подчиняются наши ошибки в измерениях. В гл. 5 мы обсудим нормальное распределение, или распределение Гаусса, которое описывает измерения, подверженные случайным погрешностям. Мы увидим, что если измерения x и y выполняются независимо и если они оба подчиняются нормальному распределению, то погрешность в $q = x + y$ дается выражением

$$\delta q = \sqrt{(\delta x)^2 + (\delta y)^2}. \quad (3.13)$$

Когда комбинируют два числа, возводя их в квадрат, складывая квадраты и затем извлекая квадратный корень, как в (3.13), то говорят, что числа *складывают квадратично*. Таким образом, правило, которое содержится в (3.13), может быть сформулировано следующим образом: если измерения x и y независимы и подвержены только случайным погрешностям, то погрешность δq в рассчитанном значении $q = x + y$ равна *сумме квадратично сложенных* погрешностей δx и δy , или их *квадратичной сумме*.

Важно сравнить новое выражение (3.13) для погрешности в $q = x + y$ с нашим старым выражением

$$\delta q \approx \delta x + \delta y. \quad (3.14)$$

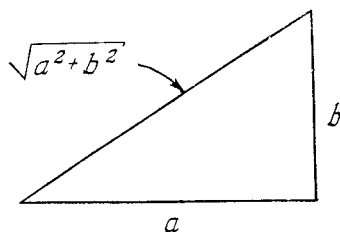


Рис 3.2. Так как любая сторона треугольника меньше суммы двух других сторон, то всегда верно неравенство $\sqrt{a^2 + b^2} < a + b$.

Во-первых, новое выражение (3.13) всегда меньше, чем старое (3.14), как можно видеть из простых геометрических соображений. Для любых двух положительных чисел a и b числа a , b и $\sqrt{a^2 + b^2}$ соответствуют трем сторонам прямоугольного треугольника (рис. 3.2). Поскольку длина любой стороны треугольника всегда меньше суммы двух других сторон, то ясно, что $\sqrt{a^2 + b^2} < a + b$ и, следовательно, (3.13) всегда меньше, чем (3.14).

Так как выражение (3.13) для погрешности в $q = x + y$ всегда меньше, чем (3.14), то всегда, когда это применимо, следует использовать выражение (3.13). Однако оно *не всегда* применимо. Выражение (3.13) отражает возможность того, что переоценка x может быть как-то скомпенсирована недооценкой в y или наоборот. Легко можно привести пример измерения, где такая компенсация невозможна.

Предположим, например, что $q = x + y$ есть сумма двух длин x и y , измеренных одной и той же стальной рулеткой. Предположим также, что главный источник погрешности заключается в том, что, как мы опасаемся, рулетка предназначена для использования при температуре, отличающейся от данной. Если мы не знаем этой температуры (и не имеем надежной рулетки для сравнения), то мы должны будем признать, что наша рулетка может быть длиннее или короче, чем ее калиброванная длина, и что, следовательно, она может давать недооцененные или переоцененные значения длины. Эту погрешность легко учесть¹⁾. Однако смысл заключается в том, что если рулетка несколько длиннее, то мы *недооцениваем оба значения x и y* , а если рулетка несколько короче, то мы *переоцениваем оба значения x и y* . Таким образом, нет

¹⁾ Предположим, например, что рулетка характеризуется коэффициентом расширения $\alpha = 10^{-5}$ на градус, и пусть мы решили, что разность между температурой, при которой производилась калибровка, и фактической температурой, по всей вероятности, не больше 10 град. Тогда маловероятно, чтобы длина рулетки более чем на 10^{-4} , или на 0,01 %, отличалась от правильной, и наша погрешность, следовательно, составляет 0,01 %.

шансов для взаимной компенсации ошибок, которая оправдывает использование квадратичной суммы для вычисления погрешности в $q = x + y$.

Позднее (в гл. 9) мы докажем, что независимо от того, являются ли наши ошибки независимыми и случайными, погрешность в $q = x + y$ *определенно не больше*, чем простая сумма $\delta x + \delta y$:

$$\delta q \leq \delta x + \delta y, \quad (3.15)$$

т. е. наше старое выражение (3.14) для δq представляет собой в действительности *верхний предел*, который справедлив во всех случаях. Если у нас имеются какие-либо основания подозревать, что ошибки в x и y *не* независимы и случайны (как в примере с измерением стальной рулеткой), то использование нами квадратичной суммы (3.13) для δq не будет оправданным. С другой стороны, предел (3.15) гарантирует, что δq *определенно не больше*, чем $\delta x + \delta y$, и более надежным будет использование старого правила

$$\delta q \approx \delta x + \delta y.$$

Фактически часто не существенно, каким образом складывать погрешности: квадратично или непосредственно. Например, предположим, что x и y — длины, измеренные обе с погрешностями $\delta x = \delta y = 2$ мм. Если бы мы были уверены, что эти погрешности независимы и случайны, то мы оценили бы ошибку в $x + y$ квадратичной суммой

$$\sqrt{(\delta x)^2 + (\delta y)^2} = \sqrt{4 + 4} \text{ мм} = 2,8 \text{ мм} \approx 3 \text{ мм},$$

а если бы мы подозревали, что погрешности могут не быть независимыми, то были бы вынуждены использовать обычную сумму

$$\delta x + \delta y \approx (2 + 2) \text{ мм} = 4 \text{ мм}.$$

Во многих экспериментах оценка погрешностей настолько груба, что различие между этими двумя результатами (3 и 4 мм) не важно. С другой стороны, иногда квадратичная сумма значительно меньше, чем обычная сумма. Кроме того, как это ни удивительно, квадратичную сумму иногда легче вычислить, чем обычную сумму. Мы встретимся с примерами этих эффектов в следующем разделе.

3.4. Еще о независимых погрешностях

В последнем разделе мы видели, как независимые случайные погрешности в двух величинах x и y вызывают погрешность в сумме $x + y$. Было показано, что для такого типа погрешностей две ошибки должны складываться квадратично. Естественно, можно рассмотреть соответствующую задачу для разностей, произведений и частных. Как мы докажем позже,

можно показать, что во всех случаях наши предыдущие правила (3.4) и (3.8) модифицируются только в том отношении, что суммы ошибок (или относительных ошибок) заменяются квадратичными суммами. В дальнейшем мы докажем, что старые выражения (3.4) и (3.8) являются фактически верхними пределами, которые всегда справедливы независимо от того, независимы и случайны ли погрешности. Таким образом, конечный вариант наших двух главных правил выглядит следующим образом:

Погрешность в суммах и разностях

Предположим, что $x, \dots, w$ измерены с погрешностями $\delta x, \dots, \delta w$ и что измеренные значения используются для вычисления

$$q = x + \dots + z - (u + \dots + w).$$

Если известно, что погрешности в $x, \dots, w$ *независимы и случайны*, то погрешность в q равна квадратичной сумме

$$\delta q = \sqrt{(\delta x)^2 + \dots + (\delta z)^2 + (\delta u)^2 + \dots + (\delta w)^2} \quad (3.16)$$

исходных погрешностей. В любом случае δq никогда не больше, чем их обычная сумма

$$\delta q \leq \delta x + \dots + \delta z + \delta u + \dots + \delta w. \quad (3.17)$$

Погрешности в произведениях и частных

Предположим, что $x, \dots, w$ измерены с погрешностями $\delta x, \dots, \delta w$ и что измеренные значения используются для вычисления

$$q = \frac{x \times \dots \times z}{u \times \dots \times w}.$$

Если погрешности в $x, \dots, w$ *независимы и случайны*, то относительная погрешность в q равна квадратичной сумме исходных относительных погрешностей

$$\frac{\delta q}{|q|} = \sqrt{\left(\frac{\delta x}{x}\right)^2 + \dots + \left(\frac{\delta z}{z}\right)^2 + \dots + \left(\frac{\delta u}{u}\right)^2 + \dots + \left(\frac{\delta w}{w}\right)^2}. \quad (3.18)$$

В любом случае она никогда не больше, чем их обычная сумма

$$\frac{\delta q}{|q|} \leq \frac{\delta x}{|x|} + \dots + \frac{\delta z}{|z|} + \frac{\delta u}{|u|} + \dots + \frac{\delta w}{|w|}. \quad (3.19)$$

Заметим, что мы не обосновали возможность использования квадратичного сложения для независимых случайных погрешностей. Мы только высказали соображения, что когда различные погрешности независимы и случайны, то имеется вероятность взаимной компенсации ошибок, и что результирующая погрешность (или относительная погрешность) должна быть меньше, чем простая сумма исходных погрешностей (или относительных погрешностей). Квадратичная сумма действительно обладает таким свойством. Мы приведем надлежащее доказательство ее применимости в гл. 5. Предельные соотношения (3.17) и (3.19) будут доказаны в гл. 9.

Пример

Как мы уже отмечали, иногда нет существенного различия между погрешностями, полученными как квадратичные суммы, и погрешностями, вычисленными простым сложением. С другой стороны, иногда имеется существенная разница и, что довольно удивительно, квадратичную сумму часто намного проще вычислить. Чтобы увидеть, как это происходит, рассмотрим следующий пример.

Предположим, что мы желаем определить коэффициент полезного действия электрического мотора постоянного тока, используя этот мотор для того, чтобы поднять массу m на высоту h . Совершенная работа равна mgh , а электрическая энергия, подведенная к мотору, равна VIt , где V — приложенное напряжение, I — ток и t — время, в течение которого работал мотор. В этом случае коэффициент полезного действия равен

$$\text{коэффициент полезного действия } e = \frac{\text{работа, совершенная мотором}}{\text{энергия, подведенная к мотору}} = \frac{mgh}{VIt}.$$

Предположим, что m , h , V и I могут быть измерены все с точностью 1%:

$$(\text{относительная погрешность } m, h, V \text{ и } I) = 1\%$$

и что время t имеет погрешность 5%

$$(\text{относительная погрешность } t) = 5\%.$$

(Конечно, величина g известна с ничтожной погрешностью.) Если мы теперь вычислим коэффициент полезного действия e , то в соответствии с нашим старым правилом («относительные ошибки складываются») погрешность будет равна

$$\frac{\delta e}{e} \approx \frac{\delta m}{m} + \frac{\delta h}{h} + \frac{\delta V}{V} + \frac{\delta I}{I} + \frac{\delta t}{t} = (1 + 1 + 1 + 1 + 5)\% = 9\%.$$

С другой стороны, если мы уверены, что различные погрешности независимы и случайны, то мы можем вычислить $\delta e/e$ как квадратичную сумму, что дает

$$\frac{\delta e}{e} = \sqrt{\left(\frac{\delta m}{m}\right)^2 + \left(\frac{\delta h}{h}\right)^2 + \left(\frac{\delta V}{V}\right)^2 + \left(\frac{\delta I}{I}\right)^2 + \left(\frac{\delta t}{t}\right)^2} = \\ = \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 5^2} \% = \sqrt{29} \% \approx 5 \%.$$

Ясно, что квадратичная сумма приводит к значительно меньшей оценке для δe . Более того, как можно увидеть, с точностью до одной значащей цифры погрешности в m , h , V и I *совсем не вносят вклад* в погрешность e , рассчитанную таким способом, т. е. с точностью до одной значащей цифры мы нашли (в этом примере)

$$\frac{\delta e}{e} = \frac{\delta t}{t}.$$

Это поразительное упрощение легко понять. Когда числа складываются квадратично, они сначала возводятся в квадрат, а затем суммируются. Процесс возведения в квадрат сильно преувеличивает влияние больших чисел. Например, если одно число в 5 раз больше любого другого (как в нашем примере), то его квадрат уже в 25 раз больше аналогичных величин для других чисел и мы можем обычно полностью пренебречь другими числами.

Этот пример показывает, что лучше, а часто и легче, складывать ошибки квадратично. Он иллюстрирует также, какого рода проблемы возникают в случае, когда ошибки *действительно* независимы и когда оправданно квадратичное сложение. (Пока мы принимаем на веру, что ошибки случайны. Этот более трудный вопрос обсуждается в гл. 4.) Пять измеренных величин (m , h , V , I и t) являются физически различными величинами с разными единицами измерения, и измеряются они с помощью совершенно различных методов. Поэтому практически невероятно, чтобы источники ошибок для одной какой-либо величины были коррелированы с источниками для любой другой. Следовательно, ошибки можно рассматривать как независимые и складывать квадратично.

3.5. Произвольная функция одной переменной

Теперь мы знаем, как оценивать независимые и зависимые погрешности для сумм, разностей, произведений и частных. Однако многие расчеты включают и более сложные операции, такие, как вычисление синуса, косинуса или квадратного корня, и нам следует знать, как оценивать ошибки и в этих случаях косвенных измерений.

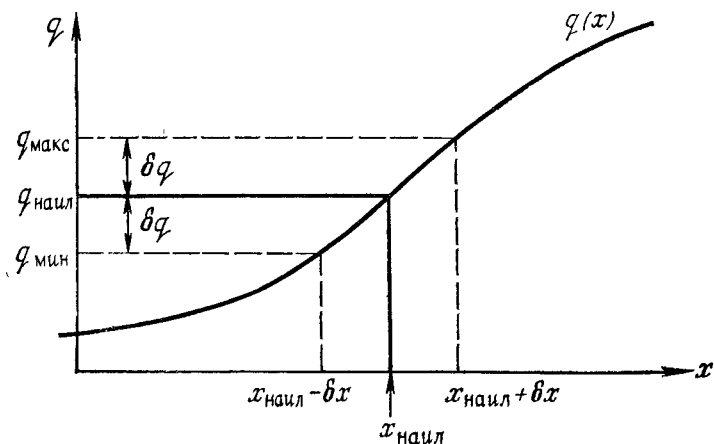


Рис. 3.3. График зависимости $q(x)$ от x .

Если величина x измерена как $x_{\text{наил}} \pm \delta x$, то наилучшая оценка $q(x)$ есть $q_{\text{наил}} = q(x_{\text{наил}})$. Наибольшее и наименьшее вероятные значения $q(x)$ соответствуют значениям $x_{\text{наил}} \pm \delta x$ величины x .

В качестве примера можно привести задачу, в которой определяется показатель преломления стекла n методом измерения предельного угла θ полного внутреннего отражения. Из элементарной оптики известно, что $n = 1/\sin \theta$. Если мы можем измерить угол θ , то легко вычислить показатель преломления n . Но затем мы должны найти погрешность δn в определении $n = 1/\sin \theta$, зная погрешность $\delta \theta$ в нашем измерении θ .

В более общем случае предположим, что мы измерили величину x стандартным образом в виде $x_{\text{наил}} \pm \delta x$ и хотим вычислить некоторую известную функцию $q(x)$, например такую, как $q(x) = 1/\sin x$ или $q(x) = \sqrt{x}$. Одним из простых способов решения этой задачи является построение графика $q(x)$, как показано на рис. 3.3. Наилучшая оценка для $q(x)$ есть, конечно, $q_{\text{наил}} = q(x_{\text{наил}})$, и значения $x_{\text{наил}}$ и $q_{\text{наил}}$ на рис. 3.3 соединены жирными линиями.

Чтобы определить погрешность δq , воспользуемся обычным способом. Наибольшее вероятное значение x есть $x_{\text{наил}} + \delta x$; используя график, можно немедленно найти наибольшее вероятное значение для q , которое обозначено $q_{\text{макс}}$. Аналогично можно найти и наименьшее вероятное значение, обозначенное $q_{\text{мин}}$. Если погрешность δx мала (как мы всегда предполагаем), то отрезок графика, используемый в наших построениях, приближенно представляет собой прямую линию, и тогда легко видеть, что $q_{\text{макс}}$ и $q_{\text{мин}}$ находятся на равных расстояниях от $q_{\text{наил}}$. Погрешность δq в этом случае может быть

определена из графика, как любое из расстояний, указанных стрелками, и, таким образом, мы нашли значение q в стандартном виде $q_{\text{наил}} \pm \delta q$.

Иногда погрешности действительно вычисляют из графика, как описано выше (см., например, задачу 3.10). Однако обычно функция $q(x)$ известна в явном виде [например, $q(x) = \sin x$ или $q(x) = \sqrt{x}$], и погрешность δq может быть выражена аналитически. Из рис. 3.3 ясно, что

$$\delta q = q(x_{\text{наил}} + \delta x) - q(x_{\text{наил}}). \quad (3.20)$$

Теперь, согласно основному приближенному выражению математического анализа, для любой функции $q(x)$ и любого достаточно малого приращения u можно написать

$$q(x + u) - q(x) = \frac{dq}{dx} u.$$

Таким образом, при условии, что погрешность δx мала (как мы всегда предполагаем), можно переписать разность (3.20) и получить

$$\delta q = \frac{dq}{dx} \delta x. \quad (3.21)$$

Итак, чтобы найти погрешность δq , мы должны вычислить производную dq/dx и умножить ее на погрешность δx .

Правило (3.21) имеет еще не окончательный вид. Оно выведено для функции, показанной на рис. 3.3, с положительным наклоном. На рис. 3.4 показана функция с отрицательным

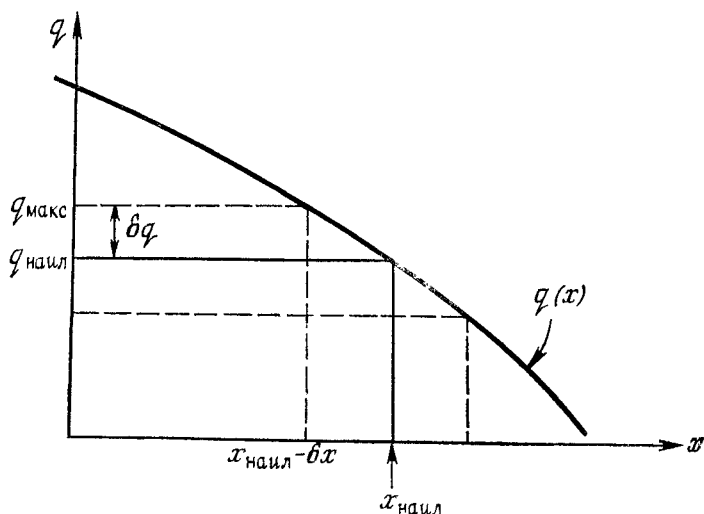


Рис. 3.4. Если наклон графика $q(x)$ отрицателен, то максимальное вероятное значение q соответствует минимальному значению x , и наоборот.

наклоном. В этом случае максимальное вероятное значение $q_{\text{макс}}$, очевидно, соответствует минимальному значению $x_{\text{наил}}$ — δx величины x , так что

$$\delta q = -\frac{dq}{dx} \delta x. \quad (3.22)$$

Поскольку производная dq/dx отрицательна, мы можем записать $-dq/dx$ как $|dq/dx|$ и получим следующее общее правило:

**Погрешность в произвольной функции
одной переменной**

Если величина x измерена с погрешностью δx и используется для вычисления функции $q(x)$, то погрешность δq равна

$$\delta q = \left| \frac{dq}{dx} \right| \delta x. \quad (3.23)$$

В качестве простого примера применения этого правила предположим, что мы измерили угол θ

$$\theta = 20 \pm 3 \text{ град}$$

и хотим найти $\cos \theta$. Наша наилучшая оценка для $\cos \theta$ составляет $\cos 20^\circ = 0,94$, а в соответствии с (3.23) погрешность равна

$$\delta(\cos \theta) = \left| \frac{d \cos \theta}{d\theta} \right| \delta\theta = |\sin \theta| \delta\theta \text{ (в рад)}. \quad (3.24)$$

Мы указали, что погрешность $\delta\theta$ должна быть выражена в радианах, поскольку производная от $\cos \theta$ равна $-\sin \theta$, только если угол θ выражен в радианах. Следовательно, перепишем $\delta\theta = 3^\circ$ в виде $\delta\theta = 0,05$ рад; тогда (3.24) дает

$$\delta(\cos \theta) = (\sin 20^\circ) \times 0,05 = 0,34 \times 0,05 = 0,02.$$

Таким образом, наш конечный результат имеет вид

$$\cos \theta = 0,94 \pm 0,02.$$

В качестве второго примера применения правила (3.23) мы можем вновь вывести (и обобщить) результат, полученный в разд. 3.2. Предположим, что мы измерили величину x и затем вычисляем степенную функцию $q(x) = x^n$ (где n — любое известное фиксированное число — положительное или отрицательное). В соответствии с (3.23) погрешность в q есть

$$\delta q = \left| \frac{dq}{dx} \right| \delta x = |n x^{n-1}| \delta x.$$

Если разделить обе части этого равенства на $|q| = |x^n|$, то получим

$$\frac{\delta q}{|q|} = |n| \frac{\delta x}{|x|}, \quad (3.25)$$

т. е. относительная погрешность в $q = x^n$ в $|n|$ раз больше, чем в x . Это и есть правило (3.10), полученное ранее. Однако в данном случае наш результат является более общим, поскольку n теперь может быть любым числом. Например, если $n = 1/2$, то $q = \sqrt{x}$ и

$$\frac{\delta q}{|q|} = \frac{1}{2} \frac{\delta x}{|x|},$$

т. е. относительная погрешность в $\sqrt{x}$ равна только *половине* погрешности в самом x . Аналогично относительная погрешность в $1/x = x^{-1}$ та же самая, что и в самом x .

Результат (3.25)—это лишь частный случай правила (3.23). Однако он достаточно важен и заслуживает отдельного рассмотрения, как следующее общее правило:

Погрешность в степенной функции

Если величина x измерена с погрешностью δx и используется для вычисления степенной функции $q = x^n$ (где n — фиксированное известное число), то относительная погрешность в q в $|n|$ раз больше, чем в x :

$$\frac{\delta q}{|q|} = |n| \frac{\delta x}{|x|}. \quad (3.26)$$

3.6. Метод «шаг за шагом»

Теперь мы имеем достаточно правил, чтобы справиться почти с любой задачей вычисления ошибок в случае косвенных измерений. Любой расчет может быть представлен как последовательность определенных шагов, каждый из которых включает только один из следующих видов операций: 1) нахождение сумм и разностей; 2) расчет произведений и частных; 3) вычисление функции одного переменного, например x^n , $\sin x$, e^x или $\ln x$. Так, мы могли бы рассчитать

$$q = x(y - z \sin u) \quad (3.27)$$

по измеренным значениям x , y , z и u в следующей последовательности шагов: вычисление функции $\sin u$, затем расчет произведения z и $\sin u$, потом определение разности y и $z \sin u$ и, наконец, произведения x и $(y - z \sin u)$.

Мы знаем, как вычисляются погрешности для каждой из этих отдельных операций. Таким образом, при условии, что все величины, с которыми мы имеем дело, независимы, мы можем вычислить погрешность конечного результата в серии последовательных шагов, исходя из погрешностей в исходных измерениях¹⁾. Например, если величины x , y , z и u в (3.27) были измерены с соответствующими погрешностями δx , ..., ..., δu , то мы могли бы вычислить погрешность q следующим образом. Сначала найдем погрешность в функции $\sin u$; зная ее, определим погрешность в произведении $z \sin u$ и затем — в разности $y - z \sin u$; наконец, найдем погрешность в произведении (3.27).

Прежде чем привести примеры вычисления ошибок с помощью этого метода «шаг за шагом», подчеркнем два основных момента. Во-первых, поскольку погрешности в суммах и разностях вычисляются в терминах абсолютных погрешностей (подобно δx), в то время как в произведениях и частных — в терминах относительных погрешностей (подобно $\delta x/|x|$), то наши расчеты потребуют, как мы увидим, некоторых средств, позволяющих переходить от абсолютных погрешностей к относительным и наоборот.

Во-вторых, важным упрощающим фактором во всех этих расчетах является то, что (как мы уже несколько раз подчеркивали) погрешности редко требуется вычислять с более чем одной значащей цифрой. Следовательно, многие расчеты можно выполнить очень быстро в уме и многими меньшими погрешностями можно полностью пренебречь. В типичном эксперименте, состоящем из нескольких попыток измерить что-либо, возможно, будет необходимо тщательно рассчитать на бумаге ошибки косвенного измерения только для первой попытки. Если это было сделано, то часто легко видеть, что все попытки достаточно подобны друг другу, и тогда не потребуется никаких дополнительных расчетов или в худшем случае для последующих попыток можно в уме модифицировать расчеты, выполненные для первой попытки.

3.7. Примеры

В этом и следующем разделах мы детально рассмотрим три примера типичных расчетов, встречающихся в учебной лаборатории. Ни один из этих примеров не является особенно

¹⁾ В разд. 3.9 мы обсудим, почему метод «шаг за шагом» иногда неудовлетворителен, например когда различные величины не независимы, как в случае функции, подобной $q = x(y - x \sin y)$, в которой x и y встречаются дважды. В этом случае вычисления «шаг за шагом» погрешности δq могут иногда дать завышенную оценку δq .

сложным, и фактически лишь небольшое число встречающихся на практике задач значительно сложнее, чем эти рассмотренные примеры.

Измерение g с помощью математического маятника

В качестве первого примера рассмотрим измерение g , ускорения свободного падения, с помощью математического маятника. Как хорошо известно, период колебаний такого маятника равен $T = 2\pi \sqrt{l/g}$, где l — длина маятника. Таким образом, если l и T измерены, мы можем найти g :

$$g = 4\pi^2 l / T^2. \quad (3.28)$$

Это выражение позволяет представить g в виде произведения двух множителей $4\pi^2$ и l и частного $4\pi^2 l / T^2$. Если различные погрешности независимы и случайны, то относительная погрешность в нашем результате равна квадратичной сумме относительных погрешностей в этих множителях. Множитель $4\pi^2$ не имеет погрешности, а относительная погрешность в T^2 в два раза больше, чем в T :

$$\frac{\delta(T^2)}{T^2} = 2 \frac{\delta T}{T}.$$

Таким образом, относительная погрешность нашего результата для g будет

$$\frac{\delta g}{g} = \sqrt{\left(\frac{\delta l}{l}\right)^2 + \left(2 \frac{\delta T}{T}\right)^2}. \quad (3.29)$$

Предположим, что мы измеряем период T для одного значения длины l и получаем результаты¹⁾

$$l = 92,95 \pm 0,1 \text{ см},$$

$$T = 1,936 \pm 0,004 \text{ с}.$$

Наша наилучшая оценка для g легко находится из (3.28) как

$$g_{\text{наил}} = \frac{4\pi^2 (92,95 \text{ см})}{(1,936 \text{ с})^2} = 979 \text{ см/с}^2.$$

Чтобы найти погрешность в g согласно (3.29), нам необходимо знать относительные погрешности в l и T . Они легко рассчитываются (в уме) как

$$\frac{\delta l}{l} = 0,1 \% \text{ и } \frac{\delta T}{T} = 0,2 \ \%.$$

¹⁾ Хотя погрешность $\delta T = 0,004$ с на первый взгляд могла бы показаться неправдоподобно малой, она легко достижима, если измерять время нескольких колебаний. Если измерения выполнять с точностью 0,1 с, что вполне возможно в случае использования секундомера, то, измеряя время 25 колебаний, можно найти период T с точностью 0,004 с.

Подставляя в (3.29), находим

$$\frac{\delta g}{g} = \sqrt{(0,1)^2 + (2 \cdot 0,2)^2} \% = 0,4 \%$$

и, следовательно,

$$\delta g = 0,004 \cdot 979 \text{ см/с}^2 = 4 \text{ см/с}^2.$$

Таким образом, наш конечный результат, основанный на этих измерениях, равен

$$g = 979 \pm 4 \text{ см/с}^2.$$

Если теперь эксперимент повторить (как это необходимо для большинства таких экспериментов) с другими значениями параметров, то не обязательно повторять расчет погрешностей во всех деталях. Немного подумав, можно легко свести разные значения l , T и g и соответствующих погрешностей в одну общую таблицу (см. задачу 3.13).

Определение показателя преломления из закона Снелла

Если луч света проходит из воздуха в стекло, то можно определить углы падения (i) и преломления (r) (рис. 3.5). Эти углы связаны законом Снелла $\sin i = n \sin r$, где n — показатель преломления стекла. Таким образом, если измерить углы i и r , то можно рассчитать показатель преломления n как

$$n = \sin i / \sin r. \quad (3.30)$$

Погрешность этого результата легко вычисляется. Так как n — частное от деления $\sin i$ на $\sin r$, то относительная погрешность дается квадратичной суммой

$$\frac{\delta n}{n} = \sqrt{\left(\frac{\delta \sin i}{\sin i}\right)^2 + \left(\frac{\delta \sin r}{\sin r}\right)^2} \quad (3.31)$$

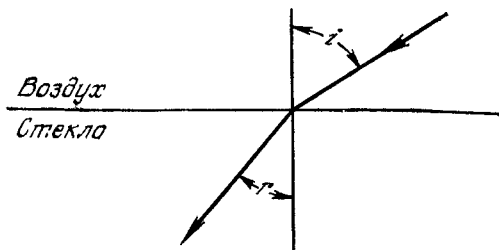


Рис. 3.5. Углы падения i и преломления r при прохождении луча света из воздуха в стекло.

Таблица 3.1. Определение показателя преломления

i , град (± 1)	r , град (± 1)	$\sin i$	$\sin r$	n	$\frac{\delta \sin i}{ \sin i }$, %	$\frac{\delta \sin r}{ \sin r }$, %	$\frac{\delta n}{n}$, %
20	13	0,342	0,225	1,52	5	8	9
40	23,5	0,643	0,399	1,61	2	4	5

относительных погрешностей в $\sin i$ и $\sin r$. Чтобы найти относительную погрешность в значении синуса любого угла θ , заметим, что

$$\delta \sin \theta = \left| \frac{d \sin \theta}{d\theta} \right| \delta \theta = |\cos \theta| \delta \theta \text{ (в рад).}$$

Таким образом, относительная погрешность равна

$$\frac{\delta \sin \theta}{|\sin \theta|} = |\operatorname{ctg} \theta| \delta \theta \text{ (в рад).} \quad (3.32)$$

Предположим, что теперь мы измеряем угол r для двух значений i и получаем результаты, приведенные в двух первых столбцах табл. 3.1 (с найденными во всех измерениях погрешностями ± 1 град, или 0,02 рад). Расчет $n = \sin i / \sin r$ легко выполняется, как видно из трех следующих столбцов табл. 3.1. Погрешность в n находится, как показано в трех последних столбцах; относительные погрешности для $\sin i$ и $\sin r$ вычисляются по формуле (3.32) и, наконец, для n — по (3.31).

Прежде чем выполнять серию измерений, подобных двум представленным в табл. 3.1, тщательно продумайте, как наилучшим образом привести данные и расчеты. Аккуратное представление данных, подобно показанному в табл. 3.1, позволяет легко записывать результаты и уменьшает опасность появления ошибок в расчетах. Читающему также будет легче проследить за записью и проверить расчеты.

3.8. Более сложный пример

Два рассмотренных выше примера типичны для экспериментов в учебной физической лаборатории. Однако небольшое число экспериментов требует более сложных расчетов. В качестве примера такого эксперимента рассмотрим измерение ускорения тележки, скатывающейся по наклонной плоскости¹⁾.

¹⁾ Читатель, если желает, может пропустить этот раздел без потери непрерывности изложения и вернуться к его изучению в связи с задачей 3.15.

Ускорение тележки, скатывающейся по наклонной плоскости

Рассмотрим тележку, скатывающуюся по наклонной плоскости с углом наклона θ , как показано на рис. 3.6. Ожидаемое ускорение равно $g \sin \theta$, и если измерить θ , то легко можно вычислить ожидаемое ускорение и его погрешность (задача 3.15). Мы можем измерить фактическое ускорение a , определяя времена, за которые тележка проходит каждый из двух фотоэлементов, соединенных с часами. Если тележка имеет длину l и за время t_1 проходит первый фотоэлемент, то ее скорость равна $v_1 = l/t_1$. Аналогично $v_2 = l/t_2$. (Строго говоря, эти скорости представляют собой *средние* скорости тележки за время прохождения фотоэлементов. Однако, пока длина l мала, различие между средней и мгновенной скоростями незначительно.) Если расстояние между фотоэлементами равно s , то в соответствии с известной формулой $v_2^2 = v_1^2 + 2as$ находим

$$a = \frac{v_2^2 - v_1^2}{2s} = \left(\frac{l^2}{2s} \right) \left(\frac{1}{t_2^2} - \frac{1}{t_1^2} \right). \quad (3.33)$$

С помощью этой формулы и измеренных значений l , s , t_1 и t_2 легко найти наблюдаемое ускорение и его погрешность.

Пусть данные для этого эксперимента (числа в скобках характеризуют соответствующие относительные погрешности в процентах, как легко можно проверить) выглядят следующим образом:

$$\begin{aligned} l &= 5,0 \pm 0,05 \text{ см} \quad (1 \%), \\ s &= 100,0 \pm 0,2 \text{ см} \quad (0,2 \%), \\ t_1 &= 0,054 \pm 0,001 \text{ с} \quad (2 \%), \\ t_2 &= 0,031 \pm 0,001 \text{ с} \quad (3 \%). \end{aligned} \quad (8.34)$$

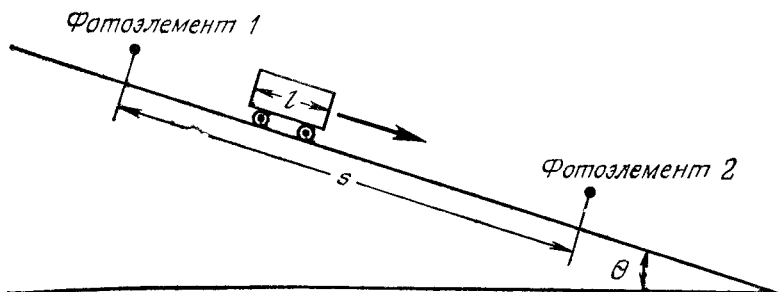


Рис. 3.6. Тележка, скатывающаяся вниз по наклонной плоскости с углом наклона θ .

Каждый фотоэлемент соединен с часами, отмечающими интервал времени, в течение которого тележка проходит мимо него.

Из этих данных можно сразу вычислить первый множитель в (3.33) $l^2/2s = 0,125$ см. Поскольку относительные погрешности в l и s равны соответственно 1 и 0,2%, аналогичная величина для $l^2/2s$ есть

$$\sqrt{(2 \times 1)^2 + (0,2)^2} \% = 2 \%.$$

(Обратите внимание на то, что погрешность в s не дает заметного вклада и ею можно пренебречь.) Следовательно,

$$l^2/2s = 0,125 \text{ см} \pm 2 \%. \quad (3.35)$$

Чтобы вычислить второй множитель в (3.33) и его погрешность, будем продолжать расчет по этапам. Так как относительная погрешность в t_1 составляет 2%, то аналогичная величина для $1/t_1^2$ составляет 4%. Таким образом, поскольку $t_1 = 0,054$ с,

$$1/t_1^2 = 343 \pm 14 \text{ с}^{-2}.$$

Аналогично относительная погрешность в $1/t_2^2$ составляет 6%, и

$$1/t_2^2 = 1041 \pm 62 \text{ с}^{-2}.$$

Вычитая эти значения (и складывая ошибки квадратично), находим

$$\frac{1}{t_2^2} - \frac{1}{t_1^2} = 698 \pm 64 \text{ с}^{-2} \quad (9 \%). \quad (3.36)$$

Наконец, в соответствии с (3.33) искомое ускорение равно произведению (3.35) и (3.36). Перемножая эти значения (и складывая квадратично относительные погрешности), получаем

$$a = (0,125 \text{ см} \pm 2 \%) \cdot (698 \text{ с}^{-2} \pm 9 \%) = 87,3 \text{ см/с}^2 \pm 9 \%,$$

или

$$a = 87 \pm 8 \text{ см/с}^2. \quad (3.37)$$

Этот результат можно было бы сравнить с ожидаемым ускорением $g \sin \theta$, если бы оно было рассчитано.

При внимательном изучении расчета, приведшего к (3.37), можно отметить несколько интересных особенностей. Во-первых, 2%-ная погрешность множителя $l^2/2s$ полностью перекрывается 9%-ной погрешностью в $(1/t_2^2) - (1/t_1^2)$. В случае расчетов для последующих испытаний погрешности в l и s можно игнорировать (если прикидка показывает, что они все еще неважны).

Другой важной особенностью нашего расчета является возрастание 2- и 3%-ных погрешностей в t_1 и t_2 при вычислении $1/t_1^2$ и $1/t_2^2$ и разности $(1/t_2^2) - (1/t_1^2)$, так что конечная по-

грешность становится равной 9%. Этот рост частично обусловлен возведением в квадрат и частично вычислением разности больших чисел. Мы могли бы представить себе некоторое расширение эксперимента для проверки постоянства a , когда тележке дается начальный толчок, так что скорости v_1 и v_2 возрастают. В этом случае времена t_1 и t_2 стали бы меньше и ошибки возросли бы (см. задачу 3.15).

3.9. Общая формула для вычисления ошибок в косвенных измерениях ¹⁾

Итак, мы установили три основных правила для расчета ошибок в случае косвенных измерений: правило для сумм и разностей, правило для произведений и частных и правило для произвольной функции одного переменного. В последних трех разделах мы видели, как вычисление сложной функции часто может быть разбито на отдельные элементы и как погрешность в рассчитываемой функции можно оценить методом «шаг за шагом», используя три наших простых правила.

В этом заключительном разделе мы приведем одну общую формулу, из которой могут быть получены все три упомянутых правила и с помощью которой может быть решена любая задача вычисления ошибок в косвенных измерениях. Хотя эта формула на практике довольно громоздка, теоретически она весьма полезна. Более того, имеется ряд задач, для которых лучше проделать вычисления в один прием с помощью общей формулы, чем рассчитывать погрешность методом «шаг за шагом», как в последних трех разделах.

Чтобы проиллюстрировать тип задач, для которых расчет в один прием предпочтительнее, предположим, что мы измеряем три величины x , y , z и должны вычислить функцию типа

$$q = \frac{x+y}{x+z}, \quad (3.38)$$

в которой какая-то переменная появляется более чем один раз (в данном случае x). Если бы мы намеревались вычислить погрешность δq методом «шаг за шагом», то сначала вычислили бы погрешности в двух суммах $x+y$ и $x+z$ и затем в их частном. Поступая таким образом, мы бы совсем не учли возможность того, что ошибка в числителе, связанная с x , может до некоторой степени компенсироваться ошибкой в знаменателе, связанной с теми же ошибками в x . Чтобы понять, как это может случиться, предположим, что x , y , z — положительные числа, и посмотрим, что произойдет, если

¹⁾ Читатель может без ущерба отложить изучение этого раздела. Материал, который в нем затрагивается, не используется вплоть до разд. 5.6.

наше измерение x подвержено ошибке. Если мы *первоцениваем* x , то мы *переоцениваем* обе суммы $x + y$ и $x + z$ и (в большой степени) эти переоценки компенсируют друг друга, когда мы вычисляем $(x + y)/(x + z)$. Аналогично *недооценка* x приводит к *недооценке* обеих сумм $x + y$ и $x + z$, что опять ведет к компенсации при вычислении частного. В любом случае ошибка в x существенно компенсируется при вычислении частного $(x + y)/(x + z)$, а наш метод «шаг за шагом» совсем не учитывает этой компенсации.

Всегда, когда функция включает одну и ту же величину более чем один раз, как в (3.38), некоторые из ошибок могут взаимно компенсироваться (эффект, который иногда называют эффектом *компенсирующихся ошибок*). Если это происходит, расчеты погрешности методом «шаг за шагом» могут привести к переоценке конечной погрешности. Единственный способ избежать этого заключается в расчете погрешности за один прием с помощью метода, который мы сейчас обсудим¹⁾.

Предположим сначала, что мы измеряем две величины x и y и затем вычисляем некоторую функцию $q = q(x, y)$. Эта функция может быть простой, как, например, $q = x + y$, или же несколько более сложной, подобно $q = (x^3 + y) \sin(xy)$. Для функции $q(x)$ одной переменной мы показали, что если наилучшая оценка для x есть число $x_{\text{наил}}$, то наилучшая оценка для $q(x)$ есть $q(x_{\text{наил}})$. Затем мы показали, что если экстремальные (т. е. наибольшее и наименьшее) вероятные значения x равны $x_{\text{наил}} \pm \delta x$, то и соответствующие экстремальные значения q равны

$$q(x_{\text{наил}} \pm \delta x). \quad (3.39)$$

Наконец, мы использовали приближение

$$q(x + u) \approx q(x) + \frac{dq}{dx} u \quad (3.40)$$

(для любого малого приращения u), чтобы переписать экстремальные вероятные значения (3.39) в виде

$$q(x_{\text{наил}}) \pm \left| \frac{dq}{dx} \right| \delta x, \quad (3.41)$$

где знак абсолютного значения используется, чтобы учесть возможность того, что величина dq/dx может быть отрицательной. Запись (3.41) означает, что $dq \approx |dq/dx| \delta x$.

¹⁾ Иногда функция, включающая некоторую переменную более чем один раз, может быть переписана в другом виде, в котором это уже исключено. Например, $q = xy - xz$ можно переписать как $q = x(y - z)$. Для второй записи погрешность δq может быть вычислена методом «шаг за шагом» без какой-либо опасности переоценки.

В случае когда q является функцией двух переменных $q(x, y)$, аргументация аналогична. Если $x_{\text{наил}}$ и $y_{\text{наил}}$ суть наилучшие оценки x и y , то мы полагаем, что наилучшей оценкой для q , как обычно, будет

$$q_{\text{наил}} = q(x_{\text{наил}}, y_{\text{наил}}).$$

Чтобы оценить погрешность этого результата, мы должны обобщить приближение (3.40) на функцию двух переменных. Требуемое обобщение имеет вид

$$q(x + u, y + v) \approx q(x, y) + \frac{\partial q}{\partial x} u + \frac{\partial q}{\partial y} v, \quad (3.42)$$

где u и v — любые достаточно малые приращения x и y , а $\partial q / \partial x$ и $\partial q / \partial y$ — так называемые *частные производные* функции q по x и y . Таким образом, $\partial q / \partial x$ есть результат дифференцирования функции q по x , при котором y считается постоянным, и наоборот для $\partial q / \partial y$. (Дополнительный материал о частных производных приведен в задачах 3.16 и 3.17.)

Экстремальные вероятные значения x и y равны $x_{\text{наил}} \pm \delta x$ и $y_{\text{наил}} \pm \delta y$. Если подставить эти значения в (3.42) и напомнить, что $\partial q / \partial x$ и $\partial q / \partial y$ могут быть как положительными, так и отрицательными, то для экстремальных значений q получим

$$q(x_{\text{наил}}, y_{\text{наил}}) \pm \left(\left| \frac{\partial q}{\partial x} \right| \delta x + \left| \frac{\partial q}{\partial y} \right| \delta y \right).$$

Эта запись означает, что погрешность в $q(x, y)$ равна

$$\delta q \approx \left| \frac{\partial q}{\partial x} \right| \delta x + \left| \frac{\partial q}{\partial y} \right| \delta y. \quad (3.43)$$

Прежде чем рассматривать различные обобщения этого нового правила, имеет смысл применить его для вывода некоторых уже известных. Допустим, например, что

$$q(x, y) = x + y, \quad (3.44)$$

т. е. величина q просто равна сумме x и y . Обе частные производные равны единице

$$\frac{\partial q}{\partial x} = \frac{\partial q}{\partial y} = 1, \quad (3.45)$$

и тогда в соответствии с (3.43)

$$\delta q \approx \delta x + \delta y, \quad (3.46)$$

а это — уже известное правило, согласно которому погрешность в $x + y$ равна сумме погрешностей в x и y .

Совершенно аналогично если q есть произведение $q = xy$, то, как легко показать, (3.43) дает старое правило, согласно

которому относительная погрешность в q равна сумме относительных погрешностей в x и y (см. задачу 3.18).

Правило (3.43) можно обобщить. Читатель не будет удивлен, когда узнает, что в случае независимых и случайных погрешностей δx и δy сумма (3.43) может быть заменена квадратичной суммой. Если же функция q зависит более чем от двух переменных, то мы просто добавляем лишний член для каждой новой переменной. Все это приводит к следующему общему правилу (надлежащее обоснование которого будет дано в гл. 5 и 9).

Погрешность функции нескольких переменных

Предположим, что $x, \dots, z$ измерены с погрешностями $\delta x, \dots, \delta z$ и что измеренные значения используются для вычисления функции $q(x, \dots, z)$. Если погрешности в $x, \dots, z$ независимы и случайны, то погрешность в q равна

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x} \delta x\right)^2 + \dots + \left(\frac{\partial q}{\partial z} \delta z\right)^2}. \quad (3.47)$$

В любом случае она никогда не больше, чем обычная сумма

$$\delta q \leq \left| \frac{\partial q}{\partial x} \right| \delta x + \dots + \left| \frac{\partial q}{\partial z} \right| \delta z. \quad (3.48)$$

Возможно, наиболее полезной особенностью этого общего правила является то, что мы можем вывести из него все наши предыдущие правила вычисления погрешностей в случае косвенных измерений (см. задачу 3.18). Непосредственное использование общего правила довольно громоздко на практике, и обычно проще, если возможно, продвигаться шаг за шагом, используя наши предыдущие, более простые правила. Однако, если функция $q(x, \dots, z)$ включает любую переменную более одного раза, могут возникнуть компенсирующиеся ошибки; в этом случае вычисления методом «шаг за шагом» могут привести к переоценке окончательной погрешности, и тогда лучше вычислять δq за один прием, используя непосредственно (3.47) или (3.48).

Задачи

Напоминание: звездочка у номера задачи означает, что задача решается или ее ответ приводится в разделе «Ответы» в конце книги.

***3.1** (разд. 3.1). Двум студентам предложили измерить скорость эмиссии α -частиц из некоторого радиоактивного образца. Студент А считал в

течение двух минут и насчитал 32 α -частицы; студент Б считал в течение часа и насчитал 786 α -частиц. (Образец распадается настолько медленно, что ожидаемую скорость эмиссии можно считать постоянной за время измерений.)

- а. Используйте формулу (3.2) для вычисления погрешности в результате студента А, равном 32 частицам, испущенным за две минуты.
- б. Какова погрешность результата студента Б, составляющего 786 частиц, испущенных за один час?
- в. Каждый студент делит свое число отсчетов на число минут, чтобы найти *скорость* распада, т. е. число распадов в минуту. Каковы их результаты и погрешности? (Хотя погрешность в общем числе отсчетов студента Б больше, чем у студента А, погрешность в скорости, полученная студентом Б, намного меньше, чем у студента А, т. е., делая отсчеты в течение более длительного времени, можно получить более точный результат для темпа отсчетов, как можно было ожидать.)

3.2 (разд. 3.2). Студент получил следующие результаты измерения:

$$\begin{aligned}a &= 5 \pm 1 \text{ см;} \\b &= 18 \pm 2 \text{ см;} \\c &= 12 \pm 1 \text{ см;} \\t &= 3,0 \pm 0,5 \text{ с;} \\m &= 18 \pm 1 \text{ г.}\end{aligned}$$

Используя правила (3.4) и (3.8), вычислите следующие величины, их погрешности и относительные погрешности в процентах: $a + b + c$; $a + b - c$; ct ; $4a$; $b/2$ (где цифры 4 и 2 не содержат погрешности) и mb/t .

*3.3 (разд. 3.2). Используя правила (3.4) и (3.8), вычислите следующие выражения:

- а) $(5 \pm 1) + (8 \pm 2) - (10 \pm 4)$;
- б) $(5 \pm 1) \times (8 \pm 2)$;
- в) $(10 \pm 1)/(20 \pm 2)$;
- г) $2\pi(10 \pm 1)$.

Числа 2 и π (см. п. г) не содержат погрешности.

*3.4 (разд. 3.2). С помощью хорошего секундомера после некоторой практики можно измерять времена примерно от одной секунды до многих минут с погрешностью порядка 0,1 с. Предположим, что мы хотим найти период τ маятника, который приблизительно равен 0,5 с. Если измерить время одного колебания, то погрешность составит около 20 %, но, измеряя время нескольких последовательных колебаний, мы можем добиться лучшего, как показывают следующие вопросы.

- а. Если мы измеряем время пяти последовательных колебаний и получаем $2,4 \pm 0,1$ с, то каков будет наш результат для τ , его абсолютной и процентной погрешности? [Помните правило (3.9).]
- б. Каков будет ответ на вопросы п. «а», если измерено время 20 колебаний и получено $9,4 \pm 0,1$ с?
- в. Может ли быть бесконечно улучшена точность измерения τ , если измерять время все большего числа периодов?

3.5 (разд. 3.2). Если для t найдено, что $t = 8,0 \pm 0,5$ с, то каковы значения и погрешности t^2 , $1/t$ и $1/t^3$?

*3.6 (разд. 3.2). Посетитель средневекового замка решает определить глубину колодца, измеряя время падения брошенного в него камня. Он определяет, что время падения равно $t = 3,0 \pm 0,5$ с. Какой вывод он делает о глубине колодца?

3.7 (разд. 3.2). Согласно биномиальной теореме, для любого числа n и любого x , удовлетворяющего условию $|x| < 1$, имеет место разложение

$$(1 + x)^n = 1 + nx + \frac{n(n-1)}{1 \cdot 2} x^2 + \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3} x^3 + \dots$$

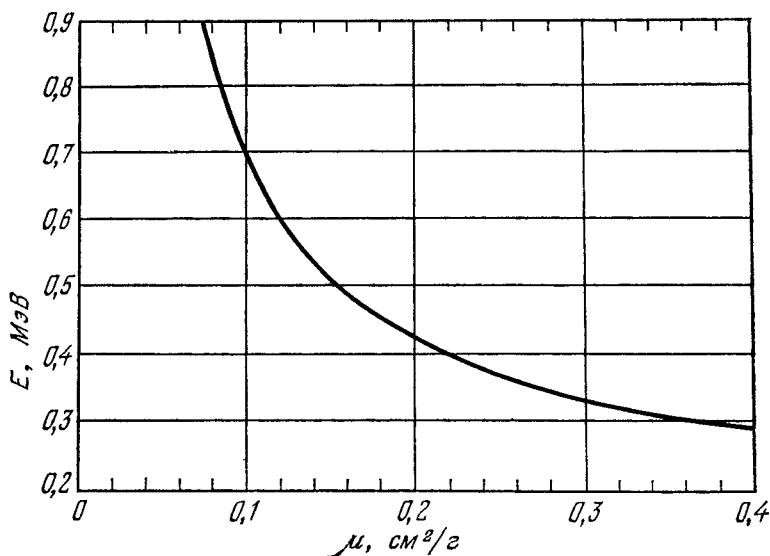


Рис. 3.7. Зависимость энергии E от коэффициента поглощения μ фотонов в свинце.

а. Покажите, что если n — целое положительное число, то этот бесконечный ряд обрывается (т.е. содержит только конечное число членов). Напишите его явное выражение для случаев $n = 2$ и $n = 3$.

б. Напишите биномиальный ряд для случая $n = -1$. Это приведет к бесконечному ряду для $1/(1+x)$; когда величина x мала, два первых члена этого бесконечного ряда дают хорошее приближение

$$1/(1+x) \approx 1-x,$$

как уже упоминалось при записи (3.6). Вычислите значения этих двух выражений для каждого из значений $x = 0,5; 0,1; 0,01$ и в каждом случае рассчитайте процентную погрешность, на которую приближение $1-x$ отличается от точного значения $1/(1+x)$.

*3.8 (разд. 3.3). Студент измеряет четыре длины: $a = 50 \pm 5$, $b = 30 \pm 3$, $c = 40 \pm 1$, $d = 7,8 \pm 0,3$ (все в сантиметрах) и вычисляет три суммы $a+b$, $a+c$, $a+d$. Найдите погрешности для этих сумм в случае, когда исходные погрешности могут не быть независимыми [«ошибки складываются», как в (3.14)], а также когда известно, что эти ошибки независимы и случайны [«ошибки складываются квадратично», как в (3.13)]. Предполагая, что погрешности надо знать только с одной значащей цифрой, в каком случае из трех погрешностей во втором слагаемом (т.е. в b , c или d) можно полностью игнорировать?

3.9 (разд. 3.4). Решите снова задачу 3.2, предполагая, что все погрешности независимы и случайны, т.е. используя квадратичное сложение, как в правилах (3.16) и (3.18) расчета ошибок для косвенных измерений¹⁾.

¹⁾ Квадратичное сложение часто можно выполнить в уме с достаточной точностью. Если вы используете калькулятор, то заметьте, что при переходе от прямоугольных к полярным координатам автоматически вычисляется величина $\sqrt{x^2 + y^2}$ для любых данных x и y .

***3.10** (разд. 3.5) В ядерной физике энергия субатомных частиц может быть измерена разными способами. Один из них состоит в измерении поглощения частиц в веществе, например таком, как свинец, и последующем сравнении с известными графиками зависимости энергии от скорости поглощения. На рис. 3.7 изображен такой график для фотонов (частиц света) в свинце. По оси ординат отложена энергия фотонов E в МэВ (миллионах электронвольт), а по оси абсцисс — соответствующий коэффициент поглощения μ в $\text{см}^2/\text{г}$. (Точное определение этого коэффициента не должно нас сейчас беспокоить; μ есть просто подходящая мера скорости поглощения фотонов в свинце.) Из этого графика легко найти энергию E фотона, если известен его коэффициент поглощения μ .

а. Студент производит опыты с пучком фотонов (одинаковой энергии) и находит, что в свинце их коэффициент поглощения равен $\mu = 0,10 \pm 0,01 \text{ см}^2/\text{г}$. Найдите по графику энергию фотонов E и ее погрешность δE (Может быть, вы сочтете полезным представить на графике линии, соединяющие разные точки, представляющие интерес, как было сделано на рис. 3.3).

б. Какой вывод сделал бы студент, если бы получил $\mu = 0,22 \pm 0,01 \text{ см}^2/\text{г}$?

***3.11** (разд. 3.5)

а. Угол θ измерен как 125 ± 2 град, и это значение используется для вычисления $\sin \theta$. Используя правило (3.23), рассчитайте $\sin \theta$ и его погрешность.

б. Если величина a измерена как $a_{\text{наил}} \pm \delta a$ и это значение используется для вычисления $f(a) = e^a$, то каковы $f_{\text{наил}}$ и δf ? Если $a = 3,0 \pm 0,1$, то чему равны e^a и ее погрешность?

в. Повторите все задание «б» для функции $f(a) = \ln a$.

3.12 (разд. 3.6). Вычислите значения следующих выражений методом «шаг за шагом», как описано в разд. 3.6 (Предположите, что все ошибки независимы и случайны)

а. $(12 \pm 1) \times [(25 \pm 3) - (10 \pm 1)]$,

б. $\sqrt{16 \pm 4} + (3,0 \pm 0,1)^3 (2,0 \pm 0,1)$,

в. $(20 \pm 2)e^{-(1,0 \pm 0,1)}$.

3.13 (разд. 3.7). Продолжите рассмотрение задачи о математическом маятнике из разд. 3.7. В условиях реального эксперимента необходимо измерять период T для разных значений длин l , и поэтому получатся разные значения g , которые можно сравнивать. Немного подумав, можно представить все данные и результаты расчетов в виде одной удобной таблицы типа табл. 3.2. Используя табл. 3.2 (или другую запись, которая вам понравится), вычислите величину g и ее погрешность δg для приведенных четырех пар данных. Объясните изменения δg с уменьшением l . (Результаты, представленные в первой строке, позволят проверить ваш метод расчета.)

***3.14** (разд. 3.7). Продолжите рассмотрение задачи об измерении показателя преломления стекла из разд. 3.7. Используя таблицу, подоб-

Таблица 3.2. Определение величины g с помощью маятника

l , см ($\pm 0,1$)	T , с ($\pm 0,001$)	g , см/с ²	$\delta l/l$, %	$\delta T/T$, %	$\delta g/g$, %	Результат $g \pm \delta g$
93,8	1,944	980	0,1	0,05	0,14	$980 \pm 1,4$
70,3	1,681					
45,7	1,358					
21,2	0,922					

Таблица 3.3. Данные для определения показателя преломления (в градусах)

$i (\pm 1)$	10	20	30	50	70
$r (\pm 1)$	6	13	19	29	38

ную 3.1, вычислите показатель преломления n и его относительную погрешность для данных из табл. 3.3. Объясните изменение погрешности. (Все углы измерены в градусах; i — угол падения, r — угол преломления.)

3.15 (разд. 3.8). Продолжите рассмотрение эксперимента из разд. 3.8, в котором тележка скатывается вниз по наклонной плоскости с углом наклона θ .

а. Если колеса у тележки гладкие и легкие, то ожидаемое ускорение равно $g \sin \theta$. Если угол θ измерен как $\theta = 5,4 \pm 0,1$ град, то каковы ожидаемое ускорение и его погрешность?

б. Если эксперимент повторяется для разных начальных толчков, сообщаемых тележке на вершине склона, то, как обычно, данные и все расчеты удобно поместить в одну таблицу, как показано в табл. 3.4. Используя формулу (3.33) для ускорения (и то же значение $l^2/2s = 0,125 \text{ см} \pm 2\%$, как прежде), вычислите a и δa для приведенных там данных. Находятся ли результаты в согласии с ожидаемым постоянством величины a и с ожидаемым значением $g \sin \theta$, вычисленным в задании «а»? Следует ли толкать тележку сильнее, чтобы проверять постоянство величины a при еще больших скоростях? Объясните.

*3.16 (разд. 3.9). Частная производная $\partial q/\partial x$ от $q(x, y)$ получается дифференцированием функции q по x , когда y считается постоянным. Найдите частные производные $\partial q/\partial x$ и $\partial q/\partial y$ для трех функций:

а) $q(x, y) = x + y$,

б) $q(x, y) = xy$,

в) $q(x, y) = x^2 y^3$.

*3.17 (разд. 3.9). Основное приближение, использованное в разд. 3.9, связывает значение функции q в точке $(x + u, y + v)$ с аналогичной величиной в соседней точке (x, y)

$$q(x + u, y + v) \approx q(x, y) + \frac{\partial q}{\partial x} u + \frac{\partial q}{\partial y} v \quad (3.49)$$

для случая, когда u и v малы. Проверьте в явном виде, что для трех функций из задачи 3.16 это — хорошее приближение, т. е. для каждой из

Таблица 3.4. Эксперимент по определению ускорения¹⁾

$t_1, \text{с}$ ($\pm 0,001$)	$t_2, \text{с}$ ($\pm 0,001$)	$\frac{1}{t_1^2}$	$\frac{1}{t_2^2}$	$\frac{1}{t_2^2} - \frac{1}{t_1^2}$	$a, \text{см/с}^2$
$0,054 \pm 2\%$	$0,031 \pm 3\%$	343 ± 14	1040 ± 62	698 ± 64	87 ± 8
0,038	0,027				
0,025	0,020				

¹⁾ Первая строка данных уже была использована в разд. 3.8. Два первых столбца содержат измеренные времена t_1 и t_2 . Погрешность во всех значениях времени равна 0,001 с; для каждого значения времени ее можно выразить как процентную погрешность.

этих трех функций запишите точные выражения для обеих частей равенства (3.49) и покажите, что они приблизительно равны, когда u и v малы. Например, если $q(x, y) = xy$, то левая часть равенства (3.49) равна

$$(x + u)(y + v) = xy + uy + xv + uv.$$

Как вы покажете, правая часть (3.49) есть

$$xy + uy + xv.$$

Если u и v малы, то величиной uv в первом выражении можно пренебречь и тогда два выражения будут приближенно равны.

3.18 (разд. 3.9).

а. Для функции $q(x, y) = xy$ запишите частные производные $\partial q/\partial x$ и $\partial q/\partial y$. Предположим, что мы измеряем x и y с погрешностями δx и δy и затем вычисляем $q(x, y)$. Используя общие правила (3.47) и (3.48), найдите погрешность δq для двух случаев, когда δx и δy независимы и случайны и когда это не так. Разделите левую и правую части этих выражений для δq на $|q| = |xy|$ и покажите, что вы получаете простые правила (3.18) и (3.19) для относительной погрешности произведения.

б. Выполните задание «а» для функции $q(x, y) = x^n y^m$, где n и m — известные фиксированные числа.

в. Какой вид примут формулы (3.47) и (3.48) в случае, когда $q(x)$ зависит только от одной переменной?

***3.19** (разд. 3.9). Если мы измеряем три независимые величины x , y , z и затем вычисляем функцию, подобную $q = (x + y)/(x + z)$, то, как мы уже указывали в начале разд. 3.9, расчеты погрешности в q методом «шаг за шагом» могут дать ее завышенное значение.

а. Рассмотрите случай, когда измеренные величины равны $x = 20 \pm 1$, $y = 2$; $z = 0$, и для простоты предположите, что δy и δz пренебрежимо малы. Вычислите погрешность δq строго, используя общее правило (3.47), и сравните полученный результат с тем, который вы получили бы, если бы рассчитывали δq методом «шаг за шагом».

б. Сделайте то же для значений $x = 20 \pm 1$; $y = -40$; $z = 0$. Объясните разницу в результатах для заданий «а» и «б».

Глава 4

Статистический анализ случайных погрешностей

Мы уже видели, что один из лучших способов оценить достоверность измерений состоит в том, чтобы повторить их несколько раз и затем сравнить между собой различные полученные значения. В этой и следующей главах мы рассмотрим статистические методы обработки результатов измерений, полученных таким образом.

Как уже отмечалось, не все виды экспериментальной погрешности можно выявить на основе статистической обработки многократных измерений. По этой причине погрешности разделяются на две группы: *случайные* погрешности, которые можно обрабатывать статистическими методами, и *систематические* погрешности, к которым эти методы *неприменимы*. Такая классификация ошибок описана в разд. 4.1. В большей части последующего материала этой главы наше внимание будет занято случайными погрешностями. В разд. 4.2 мы введем, правда без строгого обоснования, два важных определения, относящихся к ряду измеренных значений $x_1, \dots, x_N$ одной и той же величины x . Во-первых, мы определим *среднее значение*, или *среднее* $\bar{x}$, для $x_1, \dots, x_N$. При определенных условиях среднее $\bar{x}$ является наилучшей оценкой x , основанной на измеренных значениях $x_1, \dots, x_N$. Затем мы определим *стандартное отклонение* для $x_1, \dots, x_N$. Оно обозначается как σ_x и характеризует среднюю погрешность в отдельных измеренных значениях $x_1, \dots, x_N$. В разд. 4.3 приведен пример использования стандартного отклонения.

В разд. 4.4 мы введем важное понятие *стандартного отклонения среднего*. Оно обозначается $\sigma_{\bar{x}}$ и характеризует погрешность среднего $\bar{x}$ как наилучшей оценки x . В разд. 4.5 приведены некоторые примеры использования стандартного отклонения среднего. Наконец, в разд. 4.6 мы вернемся к трудной проблеме систематических ошибок.

Нигде в этой главе мы не будем пытаться дать строгое обоснование описываемых методов. Наша главная цель — ввести основные понятия и показать, как они используются. В гл. 5 мы приведем надлежащее обоснование, базирующееся на важном понятии нормального распределения.

4.1. Случайные и систематические ошибки

Экспериментальные погрешности, которые можно обнаружить с помощью многократных измерений, называются *случайными* ошибками, а те, которые нельзя обнаружить таким способом, называются *систематическими* ошибками. Чтобы проиллюстрировать различие между этими видами ошибок, рассмотрим несколько примеров. Предположим сначала, что мы измеряем время одного оборота равномерно вращающегося диска. Одним из источников ошибок будет время нашей собственной реакции при запуске и остановке секундомера. Если бы это время реакции всегда было точно одним и тем же, то два запаздывания, обусловленные реакцией, компенсировали бы друг друга. Фактически, однако, время нашей реакции изменяется. Мы можем больше промедлить при запуске и таким образом недооценить время оборота или же больше задержаться при остановке секундомера и в этом случае переоценить время. Так как обе возможности равновероятны, то знак эффекта *случаен*. При многократном повторении измерения мы иногда будем переоценивать время, а иногда — недооценивать. Таким образом, переменное время нашей реакции проявится в различии полученных результатов. Анализируя разброс в результатах методами статистики, мы можем получить очень достоверную оценку ошибки этого типа.

С другой стороны, если наш секундомер постоянно отстает, то все измеренные значения времени будут недооценены и никакое количество повторений (с тем же секундомером) не обнаружит этого источника ошибок. Ошибка такого типа называется *систематической*, поскольку она всегда смещает наш результат в одну сторону. (Если секундомер отстает, мы всегда недооцениваем время, если спешит — всегда переоцениваем.) Систематические ошибки нельзя обнаружить теми статистическими методами, которые мы сейчас будем рассматривать.

В качестве второго примера проявления случайных и систематических ошибок рассмотрим измерение точно определенной длины с помощью линейки. Один из источников погрешности — это необходимость в интерполяции между метками шкалы, и эта погрешность, очевидно, случайна. (При интерполяции мы с равной вероятностью как переоцениваем, так и недооцениваем результат.) Но имеется также вероятность того, что наша линейка дефектна, а этот источник погрешности будет, вероятно, приводить к систематической ошибке. (Если линейка растянута, мы всегда недооцениваем результат, если сжата — всегда переоцениваем.)

Подобно этим двум примерам, почти все измерения подвержены как случайным, так и систематическим погрешностям. Вам не трудно было бы привести и другие примеры. В частности, обратите внимание на то, что типичные источники случайных погрешностей — это небольшие ошибки наблюдателя (как в случае интерполяции), небольшие помехи, воздействующие на аппаратуру (подобные механическим вибрациям), проблемы определения и некоторые другие. Возможно, наиболее очевидная причина систематической ошибки — это раскалибровка приборов (подобно отстающему секундомеру, вытянутой линейке или стрелочному прибору, у которого стрелка до начала измерений не была установлена на нуль).

Различие между случайными и систематическими ошибками не всегда можно ясно определить. Например, при изменении положения вашей головы по отношению к типичному стрелочному прибору (например, обычным часам) результаты считывания будут изменяться. Этот эффект называется *параллаксом*, и он приводит к тому, что правильное считывание со шкалы возможно только в случае, когда ваша голова расположена точно перед стрелкой. Даже если вы очень аккуратный экспериментатор, вы не сможете расположить ваш глаз всегда *точно* перед стрелкой; следовательно, ваши измерения будут содержать малые погрешности, связанные с параллаксом, и эта погрешность будет, вероятно, случайной. С другой стороны, неосторожный экспериментатор, который поставит стрелочный прибор сбоку от себя и забудет о влиянии параллакса, внесет систематическую ошибку во все свои отсчеты. Таким образом, один и тот же эффект, параллакс, может привести к случайным погрешностям в одном случае и систематическим — в другом.

Учет случайных ошибок совершенно отличен от учета систематических ошибок. Статистические методы, описанные в следующем разделе, дают достоверную оценку случайных погрешностей и, как мы увидим, указывают на точно определенный способ их уменьшения. С другой стороны, систематические погрешности трудно оценить и даже обнаружить. Опытный ученый должен уметь предвидеть возможные источники систематических ошибок и обеспечить, чтобы все оставшиеся систематические ошибки были значительно меньше требуемой точности. Для этого потребуется, например, проверка стрелочных приборов по принятым стандартам, их исправление, или даже, если необходимо, приобретение более совершенных приборов. К сожалению, в учебной физической лаборатории такая проверка приборов возможна только в редких случаях, поэтому учет систематических ошибок всегда представляет собой трудное дело. Мы еще вернемся к этому во-

просу в разд. 4.6, а пока будем рассматривать эксперименты, для которых все источники систематических ошибок выявлены и приняты меры, чтобы эти ошибки были намного меньше требуемой точности.

4.2. Среднее и стандартное отклонение

Предположим, что нам надо измерить некоторую величину x и что мы выявили все источники систематической ошибки и уменьшили их влияние до пренебрежимо малого уровня. Поскольку все оставшиеся источники ошибок случайны, мы будем в состоянии обнаружить их, многократно повторяя измерения. Можно было бы, например, выполнить измерения пять раз и получить результаты

$$71, 72, 72, 73, 71 \quad (4.1)$$

(где для удобства мы опустили единицы измерения).

Первый вопрос, который можно задать: если дано пять измеренных значений (4.1), то что мы должны принять за наилучшую оценку $x_{\text{наил}}$ нашей величины x ? Представляется разумным, чтобы нашей наилучшей оценкой было *среднее значение*, или *среднее* $\bar{x}$ пяти найденных значений, и в гл. 5 мы докажем, что это обычно так и будет. Таким образом,

$$x_{\text{наил}} = \bar{x} = \frac{71 + 72 + 72 + 73 + 71}{5} = 71,8. \quad (4.2)$$

В этом выражении дробь после второго знака равенства — просто определение среднего $\bar{x}$ для данных чисел¹⁾.

В более общем случае предположим, что мы производим N измерений величины x (используя одну и ту же аппаратуру и метод измерения) и получаем N значений:

$$x_1, x_2, \dots, x_N. \quad (4.3)$$

И на этот раз наилучшей оценкой величины x обычно будет среднее значение от $x_1, \dots, x_N$, т. е.

$$x_{\text{наил}} = \bar{x}, \quad (4.4)$$

¹⁾ В наш век карманных калькуляторов, возможно, стоит заметить, что усреднение, подобное (4.2), легко произвести в уме. Поскольку все числа находятся в седьмом десятке, то же должно быть верно и для среднего значения. Остается лишь усреднить числа 1, 2, 2, 3, 1 в позиции единиц. Они, очевидно, усредняются как $9/5 = 1,8$, и наш результат будет $\bar{x} = 71,8$.

где

$$\begin{aligned}\bar{x} &= \frac{x_1 + x_2 + \dots + x_N}{N} = \\ &= \frac{\sum x_i}{N}.\end{aligned}\tag{4.5}$$

В последней строке мы ввели полезное обозначение Σ , согласно которому

$$\sum_{i=1}^N x_i = \sum_i x_i = \sum x_i = x_1 + x_2 + \dots + x_N,$$

где второе и третье выражения — обычные сокращения, которые мы будем использовать, когда не возникает опасности путаницы.

Понятие среднего значения, или среднего, конечно, хорошо знакомо большинству читателей. Следующее наше понятие, *стандартное отклонение*, вероятно, менее известно. Стандартное отклонение результатов измерений $x_1, \dots, x_N$ — это оценка *средней погрешности результатов измерений* $x_1, \dots, x_N$, которое определяется следующим образом.

Если принять, что среднее $\bar{x}$ — это наилучшая оценка величины x , то естественно рассмотреть разность $x_i - \bar{x} = d_i$. Эта разность, часто называемая *отклонением* (или остатком) x_i от $\bar{x}$, показывает, *насколько результат i -го измерения x_i отличается от среднего значения $\bar{x}$* . Если отклонения $d_i = x_i - \bar{x}$ очень малы, то результаты наших измерений близки друг к другу и, вероятно, очень точны. Если некоторые из отклонений велики, то наши измерения, очевидно, не очень точны.

Чтобы быть уверенными, что мы усвоили понятие отклонения, вычислим отклонения для набора пяти результатов

Таблица 4.1. Вычисление отклонений

Номер измерения i	Измеренное значение x_i	Отклонение $d_i = x_i - \bar{x}$
1	71	-0,8
2	72	0,2
3	72	0,2
4	73	1,2
5	71	-0,8
	$\bar{x} = 71,8$	$\bar{d} = 0,0$

измерений, приведенных в (4.1). Результаты расчета представлены в табл. 4.1. Обратите внимание, что отклонения (конечно же) не все одинаковы; d_i мало, если в i -м результате измерения x_i оказывается близким к $\bar{x}$, но d_i велико, если x_i далеко отстоит от $\bar{x}$. Заметьте также, что некоторые из d_i положительны, а некоторые — отрицательны, так как некоторые x_i должны быть больше среднего значения $\bar{x}$, а некоторые — меньше.

Чтобы оценить достоверность результатов измерений $x_1, \dots, x_5$ в среднем, мы могли бы, естественно, попытаться усреднить отклонения d_i . К сожалению, как показывает табл. 4.1, среднее значение отклонений равно нулю. На самом деле так будет в случае любого набора результатов измерений $x_1, \dots, x_N$, поскольку уже само определение среднего значения $\bar{x}$ ведет к тому, что $d_i = x_i - \bar{x}$ иногда положительны, а иногда отрицательны таким образом, чтобы $\bar{d}$ было равно нулю (см. задачу 4.3). Очевидно поэтому, что среднее отклонений — это не лучшая характеристика достоверности результатов измерений $x_1, \dots, x_N$.

Лучший способ обойти эту неприятность — возвести в квадрат все отклонения, которые в этом случае будут образовывать набор *положительных* чисел, а затем усреднить эти числа¹⁾. Если мы теперь извлечем квадратный корень из полученного результата, то получим величину, которая измеряется в тех же единицах, что и сама величина x . Это число называется *стандартным отклонением* $x_1, \dots, x_N$ и обозначается σ_x :

$$\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (d_i)^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}. \quad (4.6)$$

Исходя из этого обозначения, стандартное отклонение (или СО) можно описать как *среднеквадратичное* отклонений результатов измерений $x_1, \dots, x_N$ ²⁾. Можно показать, что это полезный способ оценки достоверности измерений. (Как мы увидим вскоре, определение (4.6) иногда модифицируется таким образом, что в знаменателе вместо N стоит $N - 1$.)

¹⁾ Другой возможностью было бы вычисление абсолютных значений $|d_i|$ и последующее их усреднение, но оказывается, что усреднение d_i^2 более полезно. Среднее значение $|d_i|$ иногда называют (ошибочно) *средним отклонением*.

²⁾ В оригинале для этого понятия автор приводит известную аббревиатуру на английском языке — R. M. S. (root mean square). В литературе на русском языке нет эквивалентной общепринятой аббревиатуры, однако в этой книге для удобства мы будем использовать аббревиатуру СО (стандартное отклонение). — Прим. перев.

Таблица 4.2. Вычисление стандартного отклонения

Номер измерения i	Измеренное значение x_i	Отклонение $d_i = x_i - \bar{x}$	d_i^2
1	71	—0,8	0,64
2	72	0,2	0,04
3	72	0,2	0,04
4	73	1,2	1,44
5	71	—0,8	0,64
$\bar{x} = 71,8$			$\sum d_i^2 = 2,80$

Чтобы вычислить стандартное отклонение σ_x в соответствии с (4.6), мы должны сначала рассчитать отклонения d_i , возвести их в квадрат, усреднить эти квадраты и затем извлечь квадратный корень из среднего. Для пяти измерений из табл. 4.1 мы могли бы рассчитать σ_x , как показано в табл. 4.2.

Суммируя числа d_i^2 из четвертого столбца табл. 4.2 и деля сумму на 5, мы получим величину σ_x^2 (часто называемую *дисперсией* измерений):

$$\sigma_x^2 = \frac{1}{N} \sum d_i^2 = \frac{2,80}{5} = 0,56. \quad (4.7)$$

Извлекая квадратный корень, находим стандартное отклонение

$$\sigma_x \approx 0,7. \quad (4.8)$$

Таким образом, средняя погрешность результатов пяти измерений 71, 72, 72, 73, 71 приближенно равна 0,7.

К сожалению, имеется альтернативное определение стандартного отклонения. Существуют теоретические аргументы, согласно которым фактор N в (4.6) следует заменить на $(N - 1)$, и поэтому стандартное отклонение σ_x для $x_1, \dots, x_N$ определяется как

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum d_i^2} = \sqrt{\frac{1}{N-1} \sum (x_i - \bar{x})^2}. \quad (4.9)$$

Здесь мы не будем пытаться доказать, что определение (4.9) для σ_x лучше, чем (4.6), но отметим, что новое «улучшенное» определение, очевидно, приводит к немного большему значению, чем старое (4.6), и это несколько компенсирует недо-

оценку погрешности в результатах измерений $x_1, \dots, x_N$, особенно в случае, когда число измерений N мало. Существование такой тенденции можно понять, если рассмотреть предельный (и абсурдный) случай, когда $N = 1$ (т. е. когда мы сделали только одно измерение). В этом случае среднее значение $\bar{x}$ равно единственному значению x_1 и единственное отклонение автоматически равно нулю. Следовательно, определение (4.6) приводит к абсурдному результату, что $\sigma_x = 0$. С другой стороны, определение (4.9) приводит к неопределенности типа $0/0$, т. е., согласно (4.9), σ_x — неопределенная величина, что корректно отражает нашу полную неосведомленность о погрешности после выполнения только одного измерения. Определение (4.6) иногда называют *стандартным отклонением генеральной совокупности*, а (4.9) — *выборочным стандартным отклонением*¹⁾.

Различие между двумя выражениями (4.6) и (4.9) в численном отношении почти всегда незначительно. Каждый может повторить измерение много раз (по крайней мере пять раз и предпочтительнее много больше). Даже если мы произведем только пять измерений ($N = 5$), разница между $\sqrt{N} = 2,2$ и $\sqrt{N-1} = 2$ для большинства приложений незначительна. Например, если мы пересчитаем стандартное отклонение (4.8) с помощью «улучшенного» определения (4.9), то получим $\sigma_x = 0,8$ вместо $0,7$, т. е. не такую уж значительную разницу. Тем не менее необходимо знать, что существуют два определения. Вероятно, всегда лучше использовать более консервативное (т. е. приводящее к большему значению) определение (4.9), но в любом случае в отчете по лабораторной работе всегда должно быть ясно указано, какое определение использовано, чтобы читатель смог проверить расчеты.

4.3. Стандартное отклонение как погрешность единичного измерения

Мы отметили, что стандартное отклонение σ_x характеризует среднюю погрешность результатов измерений $x_1, \dots, x_N$, по которым оно было вычислено. В гл. 5 мы приведем обоснование этого, доказав следующее более точное утверждение. Если результаты наших измерений распределены нормально и если мы повторим измерение x очень большое число раз (всегда с той же аппаратурой), то приблизительно 70% ре-

¹⁾ В литературе на русском языке выражение (4.6) называют, как правило, смещенной оценкой, а (4.9) — соответственно несмещенной оценкой стандартного отклонения, но обе оценки — выборочные, т. е. зависят от выборки. — *Прим. перев.*

результатов наших измерений¹⁾ будут лежать в пределах σ_x от $\bar{x}$, т. е. 70% результатов наших измерений будут лежать в интервале $\bar{x} \pm \sigma_x$.

Это утверждение можно перефразировать следующим образом. Предположим, как и прежде, что мы получили значения $x_1, \dots, x_N$ и вычислили $\bar{x}$ и σ_x . Если затем мы делаем еще одно измерение (с той же аппаратурой), то *вероятность* того, что результат нового измерения будет лежать в пределах σ_x от $\bar{x}$, равна 70%. Далее, если число измерений N было велико, то $\bar{x}$ должно быть очень надежной оценкой действительного значения x . Следовательно, мы можем сказать, что *вероятность того, что единичное измерение* (полученное с той же аппаратурой) *не будет отличаться более чем на σ_x от действительного значения*, равна 70%. Ясно, что σ_x означает именно то, для чего мы использовали термин «погрешность» в предыдущих главах. Если мы делаем одно измерение величины x , используя данную аппаратуру, то погрешность, связанная с этим измерением, должна быть оценена как $\delta x = \sigma_x$; при таком выборе погрешности мы на 70% уверены, что результат нашего измерения будет лежать в пределах δx от фактического значения.

Чтобы проиллюстрировать использование этих понятий, предположим, что у нас имеется коробка с одинаковыми пружинами и что нас попросили измерить их коэффициенты упругости k . Вы могли бы измерять коэффициенты упругости, подвешивая грузы к каждой пружине и определяя растяжение или, что, пожалуй, лучше, подвешивая какую-то массу к каждой пружине и измеряя время ее колебаний. Какой бы метод мы ни выбрали, нам необходимо узнать k и его погрешность δk для каждой пружины, но было бы бесполезным расточительством времени многократное повторение наших измерений для каждой пружины. Вместо этого можно рассуждать следующим образом. Если мы измерим k для первой пружины несколько раз (скажем, 10 или 20), то среднее этих измерений должно дать хорошую оценку k для первой пружины. Более важно в этом случае то, что стандартное отклонение σ_k этих 10 или 20 измерений дает нам оценку погрешности нашего метода измерений k . При условии, что все наши пружины в достаточной степени одинаковы и что мы используем один и тот же метод измерения, мы могли бы с достаточным основанием ожидать той же самой погрешности в любом другом измерении²⁾. Таким образом, для каждой

¹⁾ Как мы увидим, точное число равно 68,27...%, но, очевидно, абсурдно с такой точностью приводить подобные числа.

²⁾ Если какие-то пружины сильно отличаются от первой, то погрешность измерения для них могла бы быть другой. Таким образом, если

последующей пружины нам нужно провести только одно измерение, и мы можем сразу же сделать вывод, что поскольку δk равна стандартному отклонению σ_k , полученному для первой пружины, то с вероятностью 70% наш результат будет лежать в пределах σ_k от фактического значения.

Чтобы численно проиллюстрировать эти понятия, представим себе, что проведено 10 измерений для первой пружины и получены следующие значения k (в ньютонах на метр):

$$86, 85, 84, 89, 86, 88, 88, 85, 83, 85, \quad (4.10)$$

По этим данным мы вычисляем $\bar{k} = 85,9$ Н/м и по определению (4.9)

$$\sigma_k = 1,9 \text{ Н/м} \approx \quad (4.11)$$

$$\approx 2 \text{ Н/м}. \quad (4.12)$$

Следовательно, погрешность в любом отдельном измерении k составляет примерно 2 Н/м. Если мы теперь сделаем одно измерение для второй пружины и получим результат $k = 71$ Н/м, то можно без дальнейших хлопот принять, что $\delta k = \sigma_k = 2$ Н/м, и утверждать с 70%-ной вероятностью, что

$$k \text{ для второй пружины} = 71 \pm 2 \text{ Н/м}. \quad (4.13)$$

4.4. Стандартное отклонение среднего

Если $x_1, \dots, x_N$ — результаты N измерений одной и той же величины x , то, как мы видели, наша наилучшая оценка величины x есть их среднее $\bar{x}$. Мы также видели, что стандартное отклонение σ_x характеризует среднюю погрешность отдельных измерений $x_1, \dots, x_N$. Однако наш результат $x_{\text{наил}} = \bar{x}$ есть разумная комбинация всех N измерений, и поэтому имеются основания полагать, что он будет более надежным, чем любое из отдельных измерений. В гл. 5 мы докажем это, т. е. получим, что погрешность в нашем результате $x_{\text{наил}} = \bar{x}$ равна стандартному отклонению σ_x , деленному на $\sqrt{N}$. Эта величина называется *стандартным отклонением среднего* и обозначается $\sigma_{\bar{x}}$:

$$\boxed{\sigma_{\bar{x}} = \sigma_x / \sqrt{N}.} \quad (4.14)$$

(Другие названия — *стандартная ошибка* или *стандартная ошибка среднего*.) Таким образом, основываясь на N изме-

пружины существенно различны, мы были бы вынуждены проверить погрешность, выполняя многократные измерения для каждой из небольшого числа различающихся пружин.

ренных значениях $x_1, \dots, x_N$, мы можем сформулировать наш окончательный результат для значения величины x как

$$(\text{значение } x) = x_{\text{наил}} \pm \delta x,$$

где $x_{\text{наил}} = \bar{x}$ ($\bar{x}$ — среднее от $x_1, \dots, x_N$), а δx — стандартное отклонение среднего

$$\delta x = \sigma_{\bar{x}} = \sigma_x / \sqrt{N}. \quad (4.15)$$

В качестве примера рассмотрим десять результатов измерений, приведенных в (4.10). Это результаты десяти измерений коэффициента упругости k одной пружины. Как мы уже знаем, среднее этих значений есть $\bar{k} = 85,9$ Н/м и стандартное отклонение равно $\sigma_k = 1,9$ Н/м. Следовательно, стандартное отклонение среднего равно

$$\sigma_{\bar{k}} = \sigma_k / \sqrt{10} = 0,6 \text{ Н/м}, \quad (4.16)$$

и наш окончательный результат, основанный на этих десяти измерениях, для коэффициента упругости пружины будет равен

$$k = 85,9 \pm 0,6 \text{ Н/м}. \quad (4.17)$$

Когда вы приводите ответ, подобный этому, то важно указать, что означают ваши числа, а именно среднее и стандартное отклонение среднего, — так чтобы читатель был в состоянии судить об их значимости.

Важной величиной в стандартном отклонении среднего $\sigma_{\bar{x}} = \sigma_x / \sqrt{N}$ является множитель $\sqrt{N}$ в знаменателе. Стандартное отклонение σ_x характеризует среднюю погрешность в индивидуальных измерениях $x_1, \dots, x_N$. Поэтому, если бы мы должны были выполнить еще несколько измерений (используя ту же самую аппаратуру), то это не привело бы к заметному изменению стандартного отклонения σ_x . С другой стороны, стандартное отклонение среднего $\sigma_x / \sqrt{N}$ медленно уменьшалось бы с увеличением N . Именно этого мы бы и ожидали. Если бы мы выполнили большее количество измерений перед вычислениями среднего значения, то мы, естественно, ожидали бы, что конечный результат будет более надежным, а именно это и гарантирует знаменатель $\sqrt{N}$ в (4.15). Этот фактор обеспечивает также очевидный способ увеличения точности наших измерений.

К сожалению, $\sqrt{N}$ возрастает довольно медленно с увеличением N . Например, если мы хотим увеличить точность в 10 раз просто за счет увеличения числа измерений N , мы должны были бы увеличить N в 100 раз — перспектива, мягко говоря, пугающая! Более того, мы пока пренебрегали

систематическими ошибками, а они *не* уменьшаются с увеличением числа измерений. Таким образом, на практике, если вы хотите значительно повысить точность, вы, вероятно, поступите лучшим образом, если будете совершенствовать вашу аппаратуру, а не уповать на увеличение числа измерений.

4.5. Примеры

В качестве первого простого примера применения стандартного отклонения среднего представим себе, что мы должны очень точно измерить площадь A прямоугольной пластинки размером примерно $2,5 \text{ см} \times 5 \text{ см}$. Сначала мы подберем подходящий измерительный прибор, которым может быть штангенциркуль, и затем сделаем несколько измерений длины l и ширины b пластинки. Чтобы учесть неоднородности сторон, мы выполним измерения для нескольких различных положений пластинки, а небольшие дефекты прибора учтем, используя несколько различных штангенциркулей (если они имеются). Мы могли бы сделать по десять измерений величин l и b и получить результаты, показанные в табл. 4.3.

По десяти полученным значениям l мы легко можем вычислить среднее $\bar{l}$, стандартное отклонение σ_l и стандартное отклонение среднего $\sigma_{\bar{l}}$, что показано в столбцах, обозначенных как среднее, СО и СОС¹⁾. Аналогично мы можем вычислить $\bar{b}$, σ_b и $\sigma_{\bar{b}}$. Прежде чем производить еще какие-либо расчеты, необходимо проанализировать эти результаты, чтобы убедиться, разумны ли они. Например, два стандартных отклонения σ_l и σ_b — это, как известно, средние погрешности в измерениях l и b . Поскольку l и b измерялись одним и тем же методом, то было бы довольно странно, если бы σ_l и σ_b существенно отличались друг от друга или от той величины, которую мы принимаем за разумную погрешность измерений.

Только убедившись в том, что полученные до сих пор результаты вполне надежны, мы можем быстро закончить расчеты. Наша наилучшая оценка для длины есть среднее $\bar{l}$, а ее погрешность — СОС, $\sigma_{\bar{l}}$; таким образом, окончательный результат для l будет

$$l = 24,245 \pm 0,006 \text{ мм (или } 0,025 \%),$$

где число в скобках есть относительная погрешность в процентах. Аналогично результат для b имеет вид

$$b = 50,368 \pm 0,008 \text{ мм (или } 0,016 \%).$$

¹⁾ Автор использует принятую на английском языке аббревиатуру для понятия стандартного отклонения среднего (SDOM — standard deviation of the mean). Для удобства мы будем использовать в соответствующих местах для понятия стандартного отклонения среднего аббревиатуру СОС. — *Прим. перев.*

Таблица 4.3. Длина и ширина (в миллиметрах)

Измеренные значения	Среднее	CO	COС
l 24,25; 24,26; 24,22; 24,28; 24,24 24,25; 24,22; 24,26; 24,23; 24,24	$\bar{l} = 24,245$	$\sigma_l = 0,019$	$\sigma_{\bar{l}} = 0,006$
b 50,36; 50,35; 50,41; 50,37; 50,36 50,32; 50,39; 50,38; 50,36; 50,38	$\bar{b} = 50,368$	$\sigma_b = 0,024$	$\sigma_{\bar{b}} = 0,008$

Наконец, наша наилучшая оценка для площади $A = lb$ равна произведению этих значений длины и ширины, а ее относительная погрешность определяется квадратичной суммой относительных погрешностей в l и b (в предположении, что ошибки независимы):

$$A = (24,245 \text{ мм} \pm 0,025 \%) \cdot (50,368 \text{ мм} \pm 0,016 \%) = \\ = 1221,17 \text{ мм}^2 \pm 0,03 \% = 1221,2 \pm 0,4 \text{ мм}^2. \quad (4.18)$$

Чтобы получить результат (4.18) для A , мы вычислили средние $\bar{l}$ и $\bar{b}$ с их погрешностями, равными стандартным отклонениям соответствующих средних. Затем мы рассчитали площадь A как произведение $\bar{l}$ и $\bar{b}$ и определили погрешность в A , как в косвенном измерении. Мы могли бы поступить иначе. Например, можно было бы перемножить первые измеренные значения l и b и получить первое значение A . Продолжая тем же способом, мы могли бы получить десять результатов для A и затем, применяя методы статистики к этим 10 результатам, рассчитать $\bar{A}$, σ_A и, наконец, $\sigma_{\bar{A}}$. Однако, если ошибки в l и b независимы и случайны и если мы проделали достаточно измерений, этот альтернативный метод даст (как можно показать) тот же самый результат, что и предыдущий метод¹⁾.

В качестве второго примера рассмотрим случай, когда нельзя применять обычные статистические методы к результатам прямых измерений, но можно — к конечным результатам. Предположим, что мы хотим измерить коэффициент

¹⁾ Во втором методе имеется определенная логическая непоследовательность, поскольку нет каких-либо особых причин связывать результат первого измерения l с первым измерением b . В самом деле, мы могли бы измерить l восемь раз, а b 12 раз, и тогда невозможно было бы образовывать пары величин. Таким образом, с логической точки зрения наш первый метод предпочтителен.

упругости k пружины, определяя период колебаний для некоторой массы m , прикрепленной к ее концу. Из элементарной механики известно, что период таких колебаний равен $T = 2\pi \sqrt{m/k}$. Таким образом, измеряя T и m , мы можем найти k по формуле

$$k = 4\pi^2 m / T^2. \quad (4.19)$$

Простейший способ определить k — это взять одну массу m , известную с заданной точностью, и проделать несколько аккуратных измерений T . Однако по разным причинам более интересным было бы измерение T для *разных* масс m . (Например, в этом случае наряду с измерением k мы могли бы проверить, что $T \sim \sqrt{m}$.) Тогда можно было бы получить ряд значений, подобных приведенным в двух первых строках табл. 4.4.

Очевидно, нет никакого смысла усреднять значения разных масс, приведенных в первой строке (а также и значения периодов во второй строке), так как они *не* являются различными результатами измерений одной и той же величины. Мы также ничего не сможем сказать о погрешности наших измерений, если будем сравнивать эти разные значения m . С другой стороны, каждое значение m можно объединить с соответствующим значением периода T и вычислить k , как показано в последней строке табл. 4.4. Наши данные для k в последней строке *представляют собой* результаты измерений одной и той же величины, и, таким образом, к ним применимы статистические методы. В частности, наилучшей оценкой k будет среднее $\bar{k} = 13,16$ Н/м, а погрешностью — стандартное отклонение среднего $\sigma_{\bar{k}} = 0,06$ Н/м (см. задачу 4.12). Таким образом, итоговый результат, основанный на данных табл. 4.4, имеет вид

$$\text{коэффициент упругости пружины } k = 13,16 \pm 0,06 \text{ Н/м.} \quad (4.20)$$

Если бы мы получили разумные оценки погрешностей для наших исходных измерений m и T , то мы также могли бы оценить погрешность k , используя соответствующие статистические методы расчета ошибок и положив в основу эти

Таблица 4.4. Измерение коэффициента упругости пружины k

Масса m , кг	0,513	0,581	0,634	0,691	0,752	0,834	0,901	0,950
Период T , с	1,24	1,33	1,36	1,44	1,50	1,59	1,65	1,69
$k = 4\pi^2 m / T^2$	13,17	12,97	и т. д.					

оценки для δm и δT ¹⁾. В этом случае неплохо было бы сравнить погрешности в k , полученные двумя методами.

4.6. Систематические ошибки

В последних нескольких разделах мы считали как нечто само собой разумеющееся, что все систематические ошибки сведены до пренебрежимо малого уровня еще до начала измерений. Сейчас мы вновь вернемся к той неприятной ситуации, когда имеются заметные систематические ошибки. В рассмотренном выше примере мы могли бы измерять m на весах, которые систематически давали бы завышенные или заниженные показания, или наши часы могли бы спешить или отставать. Ни одна из этих систематических ошибок не выявилась бы в процедуре сравнения различных результатов для коэффициента упругости пружины k . Отсюда можно сделать вывод, что стандартное отклонение среднего $\sigma_{\bar{k}}$ следует рассматривать как *случайную составляющую* $\delta k_{\text{сл}}$ погрешности δk , но определенно не как полную погрешность δk . Наша задача состоит в том, чтобы решить, как оценить *систематическую составляющую* $\delta k_{\text{сист}}$, а затем — как скомбинировать $\delta k_{\text{сл}}$ и $\delta k_{\text{сист}}$, чтобы получить полную погрешность δk .

Не существует простой теории, которая указала бы, как нам поступать с систематическими ошибками. Фактически единственная теория систематических ошибок состоит в том, что они должны быть выявлены и уменьшены до такой степени, пока не станут намного меньше требуемой точности. Однако в учебной лаборатории это сделать часто невозможно. Как правило, нет возможности поверить стрелочный прибор относительно более надежного, чтобы исправить его, и еще меньше шансов — купить новый прибор, чтобы заменить неисправный старый. По этой причине в некоторых учебных лабораториях принято за правило, что в условиях отсутствия точной информации каждому прибору следует приписать какую-то определенную систематическую погрешность. Например, можно было бы решить, что все секундомеры имеют 0,5%-ную систематическую погрешность, все весы — 1%-ную, все вольтметры и амперметры — 3%-ную и т. д.

Вооружившись правилами такого типа, можно продвигаться дальше разными путями. Выбор каждого из них невозможно строго обосновать, и сейчас мы рассмотрим только

¹⁾ В литературе на русском языке измерения, результаты которых приведены в табл. 4.4, называются *совместными измерениями* двух величин m и T . Адекватный статистический метод обработки результатов таких измерений — это *метод наименьших квадратов* (МНК). Автор рассматривает этот метод в гл. 8. — *Прим. перев.*

один из путей. В случае последнего примера из разд. 4.5 коэффициент упругости пружины $k = 4\pi^2 m/T$ определялся с помощью измерений ряда значений m и соответствующих значений T . Как было отмечено, статистический анализ различных значений k приводит к следующей случайной составляющей δk :

$$\delta k_{\text{сл}} = \sigma_{\bar{k}} = 0,06 \text{ Н/м.} \quad (4.21)$$

Пусть нам известно, что весы, используемые для измерения m , и часы, используемые для измерения T , имеют систематические погрешности 1 и 0,5% соответственно. Теперь мы можем найти систематическую составляющую погрешности δk методом расчета ошибок в косвенных измерениях, причем возникает только один вопрос: складывать ошибки квадратично или непосредственно? Поскольку ошибки в m и в T определены независимо и поэтому возможна некоторая компенсация, то, вероятно, разумно использовать квадратичную сумму, которая дает ¹⁾

$$\frac{\delta k_{\text{сист}}}{k} = \sqrt{\left(\frac{\delta m_{\text{сист}}}{m}\right)^2 + \left(2 \frac{\delta T_{\text{сист}}}{T}\right)^2} = \quad (4.22)$$

$$= \sqrt{1 + 1} \% = 1,4 \% \quad (4.23)$$

и, следовательно,

$$\delta k_{\text{сист}} = (13,16 \text{ Н/м}) \cdot 0,014 = 0,18 \text{ Н/м.} \quad (4.24)$$

Теперь, поскольку мы оценили как случайную, так и систематическую составляющие δk , нам остается только как-то скомбинировать их, чтобы получить саму величину δk . Можно было бы обосновать необходимость их квадратичного сложения, и тогда полная погрешность имела бы вид

$$\delta k = \sqrt{(\delta k_{\text{сл}})^2 + (\delta k_{\text{сист}})^2} = \quad (4.25)$$

$$= \sqrt{(0,06)^2 + (0,18)^2} \approx 0,2 \text{ Н/м.} \quad (4.26)$$

В этом примере систематическая погрешность полностью доминирует над случайной.

Выражение (4.25) для δk не может быть строго доказано. Не ясен также смысл полученного выражения; например,

¹⁾ Должны ли мы использовать квадратичную или обычную сумму, в действительности зависит от того, что подразумевают под утверждением, что весы имеют «1 %-ную систематическую погрешность». Если это утверждение имеет тот смысл, что ошибка *определенно* не превышает 1 % (и аналогично для часов), то адекватным будет прямое сложение и $\delta k_{\text{сист}}$ тогда *определенно* не будет превышать 2 %. С другой стороны, проверка всех весов в лаборатории может показать, что для них справедливо нормальное распределение, и поэтому 70 % из них имеют погрешность меньшую, чем 1 % (и аналогично для часов). В этом случае мы должны были бы использовать квадратичное сложение, как в (4.22), и приписывать 70 %-ную вероятность для получающегося интервала.

мы, по-видимому, не можем утверждать, что с вероятностью 70% истинное значение лежит в интервале $\bar{k} \pm \delta k$. Тем не менее это выражение дает по крайней мере разумную оценку полной погрешности в условиях, когда наши приборы имеют систематические погрешности, которые мы не смогли ликвидировать. В частности, существует один важный аспект, вследствие которого результат (4.25) правдоподобен и поучителен. В разд 4.4 мы видели, что стандартное отклонение среднего $\sigma_{\bar{k}}$ стремится к нулю с увеличением числа измерений N . Это положение означало, что если вы запаслись терпением и выполняете огромное число измерений, то можете бесконечно уменьшать погрешность, не совершенствуя вашего оборудования или метода измерений. Теперь мы можем видеть, что в действительности это не так. Увеличение N приведет к неограниченному уменьшению *случайной* составляющей $\delta k_{сл} = \sigma_{\bar{k}}$. Но любая данная аппаратура характеризуется *некоторыми* систематическими погрешностями, которые *не* уменьшаются, когда мы увеличиваем N . Из (4.25) ясно, что нет большой выгоды в дальнейшем уменьшении $\delta k_{сл}$, когда величина $\delta k_{сл}$ стала меньше $\delta k_{сист.}$. В частности, полная δk никогда не может быть меньше $\delta k_{сист.}$. Все это только подтверждает то, о чем мы уже догадывались, а именно что на практике значительное уменьшение погрешности требует улучшений в методе измерений или в оборудовании, чтобы уменьшить как случайную, так и систематическую ошибки в каждом отдельном измерении.

Задачи

Напоминание: звездочка у номера задачи означает, что задача решается или ее ответ приводится в разделе «Ответы» в конце книги.

***4.1** (разд. 4.2). Студент измеряет величину x пять раз и получает результаты

5, 7, 9, 7, 8.

Вычислите среднее значение $\bar{x}$ и стандартное отклонение σ_x . (Выполните расчеты самостоятельно, не пользуясь калькулятором. Укажите, какое определение σ_x вы используете.)

4.2 (разд. 4.2). Вычислите среднее значение и стандартное отклонение результатов десяти измерений, приведенных в (4.10). (Ответы приведены в тексте книги, но важно, чтобы вы действительно *выполнили* расчеты сами. Вам необходимо подумать об аккуратной, красивой форме представления результатов расчетов; одна из возможностей показана в табл. 4.2.)

***4.3** (разд. 4.2). Среднее $\bar{x}$ от N величин $x_1, \dots, x_N$ определено как их сумма, деленная на N , т. е. $\bar{x} = (\sum x_i)/N$. Отклонение величины x_i есть разность $d_i = x_i - \bar{x}$. Покажите, что среднее отклонений $d_1, \dots, d_N$ автоматически равно нулю.

Если вы не привыкли к обозначению $\sum$, то, возможно, вам полезно будет решить эту задачу как с этим обозначением, так и без него. Например, запишите сумму $\sum (x_i - \bar{x})$ как $(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_N - \bar{x})$ и произведите перегруппировку членов.

*4.4 (разд. 4.2). Чтобы вычислить стандартное отклонение σ_x для N измерений $x_1, \dots, x_N$, необходимо найти сумму $\sum (x_i - \bar{x})^2$. Докажите, что эту сумму можно представить в виде

$$\sum [(x_i - \bar{x})^2] = [\sum (x_i)^2] - N\bar{x}^2. \quad (4.27)$$

(Это хорошее упражнение по использованию обозначения $\sum$. Результат очень полезен на практике, и именно по этой формуле вычисляется σ_x во всех карманных калькуляторах.)

4.5 (разд. 4.2). Вычислите снова стандартное отклонение по данным задачи 4.1, используя тождество (4.27).

4.6 (разд. 4.3). Студент измеряет период колебаний маятника три раза и получает результаты 1,6; 1,8; 1,7 (все в секундах). Чему равны среднее значение и стандартное отклонение? (Используйте определение (4.9) стандартного отклонения.) Если студент произведет четвертое измерение, то какова вероятность того, что результат этого нового измерения будет лежать вне интервала от 1,6 до 1,8 с? (Очевидно, эти числа выбраны не случайно. В гл. 5 мы увидим, как решать такие задачи в случае произвольных границ интервала.)

*4.7 (разд. 4.3).

а. Вычислите среднее $\bar{t}$ и стандартное отклонение σ_t для следующих 30 результатов измерений времени t (все в секундах). Вам потребуется калькулятор, но вы будете нажимать гораздо меньше кнопок, если догадаетесь, что необходимо усреднять, в сущности, только две последние цифры, и если вы перед вычислениями сдвинете десятичную запятую на две позиции вправо. Если у вашего калькулятора нет операции вычисления стандартного отклонения, то вам, вероятно, следует воспользоваться тождеством (4.27)

8,16; 8,14; 8,12; 8,16; 8,18; 8,10; 8,18; 8,18; 8,18; 8,24;

8,16; 8,14; 8,17; 8,18; 8,21; 8,12;; 8,12; 8,17; 8,06; 8,10;

8,12; 8,10; 8,14; 8,09; 8,16; 8,16; 8,21; 8,14; 8,16; 8,13.

б. Мы знаем, что если выполнить большое число измерений, то можно ожидать, что приблизительно 70 % всех результатов окажется в пределах σ_t от $\bar{t}$ (т. е. внутри интервала $\bar{t} \pm \sigma_t$). Как будет показано в гл. 5, мы также можем ожидать, что приблизительно 95 % всех результатов окажется в пределах $2\sigma_t$ от $\bar{t}$ (т. е. внутри интервала $\bar{t} \pm 2\sigma_t$). Для результатов из задания «а» определите, сколько из них, как вы ожидаете, будет лежать *вне* интервала $\bar{t} \pm \sigma_t$? А сколько на самом деле? Ответьте на эти же вопросы в случае интервала $\bar{t} \pm 2\sigma_t$.

4.8 (разд. 4.4). Вычислите стандартное отклонение среднего для пяти измерений из задачи 4.1. Какой итоговый вывод, включая приведение погрешности δx , должен сделать студент относительно x ?

*4.9 (разд. 4.4). Если опираться на данные 30 измерений из задачи 4.7, какова будет ваша наилучшая оценка для времени и его погрешности при условии, что все погрешности случайны?

4.10 (разд. 4.4). После нескольких измерений скорости звука u студент приходит к выводу, что стандартное отклонение измерений равно $\sigma_u = 10$ м/с. Предполагая, что все ошибки случайны, студент может добиться любой желаемой точности, выполняя достаточное число измерений и усредняя. Сколько измерений необходимо сделать, чтобы конечная погрешность результата составляла ± 3 м/с? А сколько, если погрешность будет всего 0,5 м/с?

*4.11 (разд. 4.5). В табл. 4.3 приведены результаты десяти измерений длины l и ширины b прямоугольника, которые используются для вычисления площади $A = lb$. Если бы измерения делались так, что каждый раз

определялась пара значений (одно значение l и одно b), то было бы естественно перемножить результаты в каждой паре и получить значение A : первое l умножается на первое b и получается первое значение A и т. д. Вычислите таким образом десять значений A , затем среднее $\bar{A}$, стандартное отклонение σ_A и стандартное отклонение среднего $\sigma_{\bar{A}}$. Сравните ваши

результаты для $\bar{A}$ и $\sigma_{\bar{A}}$ с ответами (4.18), полученными с помощью вычислений $\bar{l}$ и $\bar{b}$ (когда A определяется как произведение $l\bar{b}$ и погрешности рассчитываются как в косвенных измерениях). (В случае большого числа измерений два метода должны приводить к одинаковым результатам.)

4.12 (разд. 4.5). Закончите вычисления коэффициента упругости пружины k из табл. 4.4. Потом рассчитайте k и его погрешность (т. е. СОС, σ_k).

***4.13** (разд. 4.6).

а. Студент измеряет скорость звука, используя формулу $u = f\lambda$, где f — частота, считываемая с циферблата звукового генератора, а λ — длина волны, измеряемая по расположению нескольких максимумов в резонирующем столбике воздуха. Так как измерения λ выполнялись несколько раз, их можно обработать методами статистики, на основании которых студент делает вывод, что $\lambda = 11,2 \pm 0,5$ см. Измерение частоты $f = 3000$ Гц было произведено только один раз (установка генератора), и у студента нет оснований, чтобы судить о надежности этого значения. Преподаватель говорит, что генератор «определенно надежен с погрешностью 1 %», следовательно, студент указывает 1 %-ную систематическую ошибку в f (но не в λ). Какой ответ даст студент для u и ее погрешности? Важна ли 1 %-ная возможная систематическая ошибка генератора (возникающая вследствие недостаточно точной его калибровки)?

б. Если бы результат измерений студента был $\lambda = 11,2 \pm 0,1$ см, а калибровка генератора была бы произведена с 3 %-ной погрешностью, то каков был ответ? Была бы важна систематическая ошибка?

Глава 5

Нормальное распределение

В этой главе мы продолжим рассмотрение статистических методов обработки многократных измерений. В гл. 4 мы ввели важные понятия среднего, стандартного отклонения и стандартного отклонения среднего; мы обсудили их значение и рассмотрели некоторые примеры их использования. В этой главе будут рассмотрены теоретические обоснования этих статистических понятий и доказаны некоторые из положений, которые принимались без доказательств в предыдущих главах.

Первая проблема, с которой приходится иметь дело при обсуждении многократных измерений, — это поиски методов, которые позволили бы выполнять различные операции со многими полученными значениями и представлять эти значения. Один из удобных методов — использование *распределения*, или *гистограммы*, как описано в разд. 5.1. В разд. 5.2 мы введем понятие *предельного распределения*, т. е. распределения результатов, которое получилось бы, если бы число измерений было бесконечно велико. В разд. 5.3 мы определим *нормальное распределение*, или *распределение Гаусса*, которое описывает предельное распределение результатов любого измерения, подверженного большому числу небольших случайных ошибок.

Как только математические характеристики нормального распределения будут поняты, мы без труда сможем приступить к доказательствам нескольких важных положений. В разд. 5.4 мы докажем, что, как это уже было сказано в гл. 4, примерно 70% результатов всех измерений (одной и той же величины и полученных с помощью одной и той же аппаратуры) должны лежать в пределах одного стандартного отклонения от истинного значения. В разд. 5.5 мы докажем результат, использованный еще в гл. 1, что если произвести N измерений некоторой величины x и получить значения $x_1, x_2, \dots, x_N$, то нашей наилучшей оценкой $x_{\text{наил}}$, основанной на этих значениях, будет среднее $\bar{x} = \sum x_i / N$. В разд. 5.6 приводится обоснование квадратичного сложения при расчете ошибок в косвенных измерениях, когда исходные ошибки не-

зависимы и случайны. В разд. 5.7 доказывается, что погрешность среднего $\bar{x}$, которое используется как наилучшая оценка x , определяется стандартным отклонением среднего $\sigma_{\bar{x}} = \sigma_x / \sqrt{N}$, как было постулировано в гл. 4. Наконец, в разд. 5.8 мы обсудим, как связать степень уверенности в результате, выраженную в виде числовой характеристики, с самими экспериментальными результатами.

Математический аппарат, используемый в этой главе, немного сложнее, чем применяемый до сих пор. Однако читатель, внимательно следящий за обсуждением нормального распределения в разд. 5.3 (и выполняющий все расчеты с помощью карандаша и бумаги, если необходимо), будет в состоянии понять большинство аргументов без большого труда.

5.1. Гистограммы и распределения

Очевидно, что серьезный статистический анализ результатов эксперимента требует выполнения многократных измерений. Таким образом, наша первая задача состоит в том, чтобы найти методы записи и представления большого числа измеренных значений. Предположим, например, что мы должны были сделать десять измерений некоторой длины x . Например, x могло бы быть расстоянием от линзы до изображения, образованного линзой. Мы могли бы получить значения (все в см)

$$26, 24, 26, 28, 23, 24, 25, 24, 26, 25. \quad (5.1)$$

При такой форме записи эти десять чисел сообщают довольно мало информации, и если бы мы должны были таким способом записать существенно больше результатов измерений, то в итоге получили бы беспорядочные джунгли чисел. Очевидно, требуется лучший способ.

В качестве первого шага мы можем расположить числа (5.1) в возрастающем порядке:

$$23, 24, 24, 24, 25, 25, 26, 26, 26, 28. \quad (5.2)$$

Затем, вместо того чтобы делать три записи: 24, 24, 24, можно просто указать, что мы получили значение 24 три раза. Другими словами, мы можем записывать *различные* полученные значения x вместе с *числом*, указывающим, сколько раз получено каждое значение, как показано в табл. 5.1. В этой таблице мы ввели символ x_k ($k = 1, 2, \dots$), чтобы обозначить различные полученные значения: $x_1 = 23$; $x_2 = 24$; $x_3 = 25$ и т. д. Величины n_k ($k = 1, 2, \dots$) обозначают числа, показывающие, сколько раз было получено соответствующее значение x_k : $n_1 = 1$; $n_2 = 3$ и т. д.

Таблица 5.1

Различные значения x_k	23	24	25	26	27	28
Число реализаций n_k	1	3	2	3	0	1

Если записать результаты измерений, как в табл. 5.1, то можно переписать определение среднего $\bar{x}$ таким способом, который окажется более удобным. Из нашего старого определения мы знаем, что

$$\bar{x} = \frac{\sum_i x_i}{N} = \frac{23 + 24 + 24 + 24 + 25 + \dots + 28}{10}. \quad (5.3)$$

Но это то же самое, что и

$$\bar{x} = \frac{23 + (24 \times 3) + (25 \times 2) + \dots + 28}{10},$$

или в общем случае

$$\bar{x} = \frac{\sum_k x_k n_k}{N}. \quad (5.4)$$

В первоначальном варианте (5.3) мы суммировали результаты *всех* сделанных измерений; в (5.4) мы суммируем все *различные* полученные значения, умножая каждое значение на число, показывающее, сколько раз это значение реализовалось. Очевидно, что эти две суммы одинаковы, но оказывается, что представление (5.4) более полезно в случае, когда мы делаем большое число измерений. Сумма, подобная (5.4), иногда называется *суммой с весовыми множителями* или *взвешенной суммой*, поскольку каждое значение x_k умножается на весовой множитель (взвешивается) — число n_k , показывающее, сколько раз это значение реализовалось. Отметим, что если мы сложим все числа n_k , то получим полное число сделанных измерений N , т. е.

$$\sum_k n_k = N. \quad (5.5)$$

(Например, в случае табл. 5.1, согласно этой формуле, сумма чисел в последней строке равна 10.)

Понятия, обсуждаемые выше в этом разделе, можно сформулировать иным способом, который часто более удобен. Вместо того чтобы говорить, что результат $x = 24$ был получен три раза, мы можем сказать, что результат $x = 24$ был получен в 3/10 случаев от полного числа измерений. Другими словами, вместо использования n_k , *числа*, показывающего, сколько раз получился результат x_k , мы введем отношение

$$F_k = \frac{n_k}{N}, \quad (5.6)$$

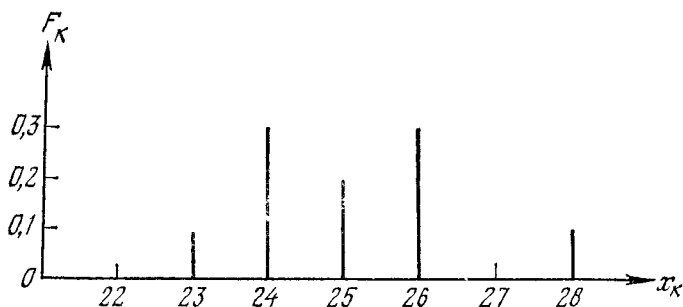


Рис. 5.1. Гистограмма результатов десяти измерений длины x . По вертикальной оси отложена доля числа случаев F_k , когда наблюдается каждое значение x_k .

представляющее собой *долю* полного числа N измерений, с которой реализуется результат x_k , и будем называть его *частотой*. Говорят, что частоты F_k характеризуют *распределение* результатов; они показывают, как результаты наших измерений *распределены* среди различных возможных значений.

В терминах частот F_k мы можем переписать формулу (5.4) для среднего $\bar{x}$ в компактном виде:

$$\bar{x} = \sum_k x_k F_k. \quad (5.7)$$

Таким образом, среднее $\bar{x}$ есть взвешенная сумма всех различных полученных значений x_k , т. е. когда каждое значение x_k умножается на частоту F_k , с которой оно реализуется.

В (5.7) предполагается, что

$$\sum_k F_k = 1, \quad (5.8)$$

т. е. если сложить частоты F_k всех возможных значений x_k , то должна получиться единица. Говорят, что если сумма какого-то набора чисел равна 1, то эти числа *нормированы*, поэтому выражение (5.8) называют *условием нормировки*.

Распределение результатов наших измерений можно представить графически на *гистограмме*, как показано на рис. 5.1. Эта гистограмма есть график зависимости F_k от x_k , на котором по горизонтальной оси отложены различные измеренные значения x_k , а частоты, показывающие долю случаев, когда

реализуется данное значение x_k , определяются высотой вертикальных черточек, проведенных из x_k . (Можно начертить также график зависимости n_k от x_k , но для наших целей более удобен график зависимости F_k от x_k .) Данные, приведенные на гистограмме типа рассмотренной, весьма наглядны, чем широко пользуются авторы газетных и журнальных публикаций.

Гистограмма, подобная приведенной на рис. 5.1, может быть названа *гистограммой для дискретной величины*, так как распределение результатов показано высотой вертикальных черточек над дискретными значениями x_k . Такой тип гистограмм удобен в тех случаях, когда значения x_k далеко отстоят друг от друга и имеют целые значения. (Например, оценки студентов на экзамене — обычно целые числа, и их удобно представить на такой гистограмме.) Однако в большинстве случаев измерения не приводят к точным целым значениям, так как большинство физических величин имеет непрерывный диапазон возможных значений. Например, вместо десяти длин, приведенных в (5.1), вы можете получить следующие существенно более вероятные значения:

26,4; 23,9; 25,1; 24,6; 22,7; 23,8; 25,1; 23,9; 25,3; 25,4. (5.9)

Гистограмма, построенная как для дискретной величины, для этих десяти значений состояла бы из десяти отдельных черточек одинаковой высоты и содержала бы сравнительно мало информации. Если даны результаты измерений, подобные приведенным в (5.9), то лучше всего разбить диапазон возможных значений на подходящее число *интервалов*, или «бинов», и сосчитать, сколько значений попадает в каждый бин. Например, мы могли бы сосчитать число результатов измерений из (5.9) между $x = 22$ и 23 , между $x = 23$ и 24 и т. д. Результаты такого подсчета представлены в табл. 5.2. (Если результат измерения попадет точно на границу между двумя бинами, то необходимо будет решить, куда его поместить. Простой и разумный выход — половину результатов приписывать одному бину и половину — другому.)

Результаты из табл. 5.2 можно нанести на график, который мы назовем *гистограммой для непрерывной величины*, как показано на рис. 5.2. На этом графике доля от полного числа измерений, которая приходится на каждый бин, пока-

Таблица 5.2

Бин	22—23	23—24	24—25	25—26	26—27	27—28
Число отсчетов в бине	1	3	1	4	1	0

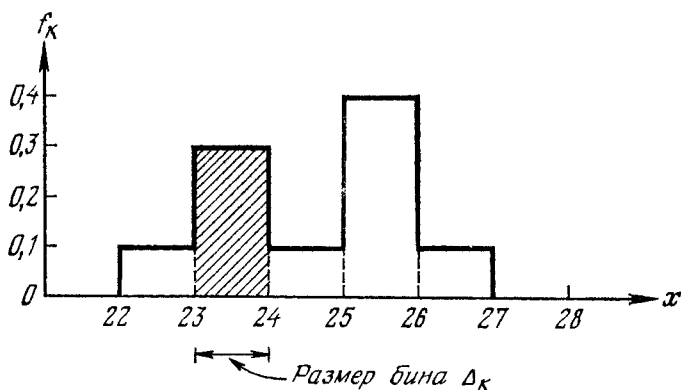


Рис. 5.2. Гистограмма для непрерывной переменной, показывающая долю измерений x , которая попадает в бины от значения 22 до 23, от 23 до 24 и т. д. Площадь прямоугольника над каждым интервалом равна доле измерений, которые попадают в этот интервал. Таким образом, если площадь заштрихованного прямоугольника равна 0,3, то это означает, что 3/10 всех измерений лежат между 23 и 24.

зана как площадь прямоугольника, расположенного над соответствующим бином. Таким образом, заштрихованный прямоугольник над интервалом от $x = 23$ до $x = 24$ имеет площадь $0,3 \times 1 = 0,3$, откуда следует, что $3/10$ результатов измерений попало в этот интервал. В общем случае мы обозначим ширину k -го бина через Δ_k . (Эти ширины обычно одинаковы, хотя это не обязательно.) Высота f_k прямоугольника над этим бином выбирается таким образом, чтобы площадь $f_k \Delta_k$ была равна

$$f_k \Delta_k = \text{доля измерений в } k\text{-м бине.}$$

Другими словами, в случае гистограммы для непрерывной величины площадь $f_k \Delta_k$ k -го прямоугольника имеет тот же смысл, что и высота k -й черточки F_k в случае гистограммы для дискретной величины.

Следует соблюдать некоторую осторожность при выборе ширины бинов Δ_k для гистограммы. Если бины выбраны слишком широкими, то все (или почти все) отсчеты попадут в один бин и гистограмма вырождается в малоинтересный единственный прямоугольник. Если же бины выбраны слишком узкими, то лишь очень небольшое их число будет содержать более чем один отсчет и сама гистограмма будет состоять из большого числа узких прямоугольников, причем почти все эти прямоугольники будут иметь одинаковую высоту. Ясно, что ширину бина следует выбирать таким образом, чтобы в каждом из них содержалось по несколько отсчетов. Следова-

но, если полное число измерений N мало, мы должны выбирать довольно широкие бины, но с увеличением N обычно становится возможным выбрать более узкие бины.

5.2. Предельные распределения

В большинстве экспериментов если увеличивать число измерений, то гистограмма будет принимать некоторую определенную простую форму. Это хорошо видно из рис. 5.3 и 5.4, в которых содержатся результаты 100 и 1000 измерений той же самой величины, что и на рис. 5.2. После 100 измерений на гистограмме вырисовывается единственный пик и она становится более симметричной. После 1000 измерений можно вполтину уменьшить ширину бинов, и гистограмма становится совсем гладкой и регулярной. Эти три схемы иллюстрируют важное свойство большинства измерений. С ростом числа измерений до бесконечности их распределение стремится

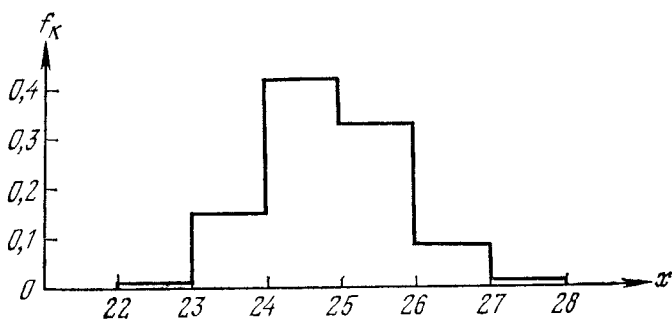


Рис. 5.3. Гистограмма результатов 100 измерений той же величины, что и на рис. 5.2.

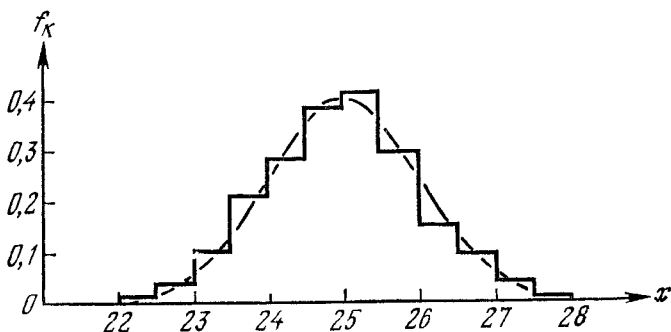


Рис. 5.4. Гистограмма результатов 1000 измерений той же величины, что и на рис. 5.3. Прерывистая кривая — предельное распределение.

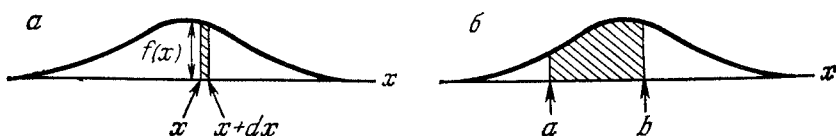


Рис. 5.5. Предельное распределение $f(x)$.

a — после множества измерений доля, которая попадает между x и $x + dx$, равна площади $f(x)dx$ узкой полосы, b — доля, которая попадает между $x = a$ и $x = b$, равна заштрихованной площади.

к некоторой определенной непрерывной кривой. Получающаяся непрерывная кривая называется *предельным распределением*¹⁾. Таким образом, для измерений, представленных на рис. 5.2—5.4, предельное распределение представляет собой симметричную колоколообразную кривую, показанную на рис. 5.4.

Следует подчеркнуть, что предельное распределение — это теоретическая идеализация, к которой никогда нельзя абсолютно точно приблизиться на эксперименте. Чем больше измерений мы делаем, тем ближе становится гистограмма к предельному распределению. Но лишь в случае бесконечного числа измерений и использования бесконечно узких бинов мы смогли бы получить само предельное распределение. Тем не менее имеются надежные предпосылки, что почти для всех измерений *существует* предельное распределение, к которому наша гистограмма все более приближается по мере того, как мы делаем все большее число измерений.

Предельное распределение, подобное гладкой кривой на рис. 5.4, определяет функцию, которую мы обозначим $f(x)$. Значение этой функции пояснено на рис. 5.5. С ростом числа измерений величины x наша гистограмма будет приближаться к предельной кривой $f(x)$. Следовательно, доля измерений, которая попадает в любой малый интервал от x до $x + dx$, будет равна площади $f(x)dx$ заштрихованного участка на рис. 5.5, a :

$f(x)dx =$ доля измерений,

попадающая в интервал от x до $x + dx$. (5.10)

В общем случае доля числа измерений, которая попадает в интервал между любыми двумя значениями a и b , равна площади под кривой между $x = a$ и $x = b$ (рис. 5.5, b). Но

¹⁾ Некоторые из общеупотребительных синонимов (или приблизительных синонимов) понятия предельного распределения: родительское распределение, асимптотическое родительское распределение, генеральное распределение и родительская популяция. (В литературе на русском языке употребляется также термин «генеральная совокупность», — Прим. перев.)

эта площадь есть *определенный интеграл* от $f(x)$. Таким образом, мы получили следующий важный результат:

$$\int_a^b f(x) dx = \text{доля измерений,}$$

попадающая в интервал от $x=a$ до $x=b$. (5.11)

Важно понять значение утверждений (5.10) и (5.11). Они оба определяют долю измерений, которая, как ожидается, будет лежать внутри некоторого интервала в случае *очень большого числа измерений*. То же самое можно выразить другим очень полезным способом: $f(x)dx$ есть *вероятность* того, что единичное измерение x приведет к результату, лежащему в интервале от x до $x+dx$, т. е.

$f(x) dx = \text{вероятность того, что любое единичное}$
 $\text{измерение приведет к результату,}$
 $\text{лежащему в интервале от } x \text{ до } x+dx.$

(5.12)

Аналогично интеграл $\int_a^b f(x) dx$ определяет вероятность того, что результат любого единичного измерения попадет в интервал от $x=a$ до $x=b$. Мы пришли к следующему важному заключению: если бы нам было известно предельное распределение $f(x)$ для результатов измерений данной величины x , полученных с помощью данной аппаратуры, то мы знали бы вероятность получения результата в любом заданном интервале $a \leq x \leq b$.

Так как полная вероятность получения результата, лежащего между $-\infty$ и $+\infty$, должна быть равна единице, то предельное распределение $f(x)$ должно удовлетворять условию

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

(5.13)

Это равенство — полный аналог условия нормировки (5.8), когда $\sum_k F_k = 1$, и поэтому говорят, что функция $f(x)$, удовлетворяющая (5.13), *нормирована*.

Читатель может быть введен в заблуждение пределами $\pm\infty$ в интеграле (5.13). Но эти пределы вовсе не означают, что мы действительно ожидаем получать результаты, изме-

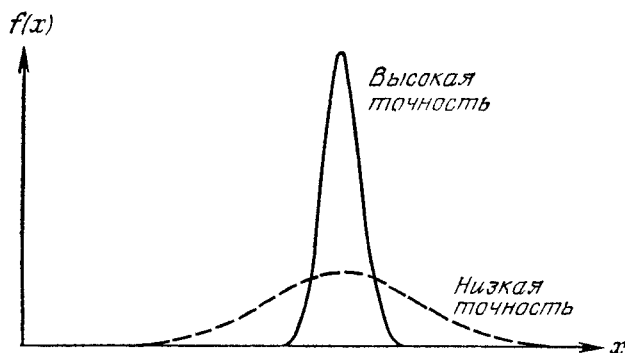


Рис. 5.6. Два предельных распределения: одно для измерений с высокой точностью, а другое — с малой.

няющиеся от $-\infty$ до $+\infty$. Совсем наоборот. В любом реальном эксперименте результаты всех измерений попадают в некоторый достаточно малый конечный интервал. Например, все результаты измерений в случае рис. 5.4 лежат между $x = 21$ и $x = 29$. Даже после бесконечно большого числа измерений их доля, лежащая вне интервала от $x = 21$ до $x = 29$, будет полностью пренебрежимой. Другими словами, $f(x)$ практически равна нулю вне этого интервала, и поэтому нет никакой разницы, равны ли пределы интеграла (5.13) $\pm\infty$ или соответственно 21 и 29. Но поскольку мы в общем случае не знаем этих конечных пределов, то более удобно оставить их как $\pm\infty$.

Если выполняются очень точные измерения, то все полученные значения будут близки к действительному значению x , и поэтому гистограмма результатов и, следовательно, предельное распределение будут выглядеть как отрывистые пики, например как сплошная кривая на рис. 5.6. Если точность измерений мала, то результаты будут сильно различаться и их распределение будет описываться широкой кривой, подобно пунктирной кривой на рис. 5.6.

Предельное распределение $f(x)$ в случае измерений данной величины x с помощью заданной аппаратуры предсказывает, как были бы распределены их результаты после очень многих измерений. Таким образом, зная $f(x)$, мы могли бы рассчитать среднее значение $\bar{x}$, которое было бы найдено после множества измерений. Мы видели [см. (5.7)], что среднее любого числа измерений — это сумма всех различных значений x_k , умноженных на весовой множитель, представляющий долю случаев, когда получается это значение:

$$\bar{x} = \sum_k x_k F_k. \quad (5.14)$$

В данном случае у нас есть бесконечное число измерений с распределением $f(x)$. Если мы разделим весь интервал значений на малые интервалы от x_k до $x_k + dx_k$, то доля значений, попавшая в каждый такой интервал, будет $F_k = f(x_k) dx_k$, и тогда в пределе, когда все интервалы стремятся к нулю, формула (5.14) примет вид

$$\bar{x} = \int_{-\infty}^{+\infty} x f(x) dx. \quad (5.15)$$

Запомните, что эта формула определяет среднее $\bar{x}$, которое соответствует бесконечно большому числу измерений¹⁾.

Аналогично мы можем вычислить стандартное отклонение σ_x , полученное после множества измерений. Поскольку мы рассматриваем случай $N \rightarrow \infty$, то безразлично, какое определение σ_x мы используем, первоначальное (4.6) или же «улучшенное» (4.9) с заменой N на $N-1$. В любом случае при $N \rightarrow \infty$ величина σ_x^2 есть просто среднее квадрата отклонений $(x - \bar{x})^2$. Таким образом, точно те же аргументы, которые привели к (5.15), позволят получить в случае множества измерений

$$\sigma_x^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 f(x) dx. \quad (5.16)$$

5.3. Нормальное распределение

Различные виды измерений имеют разные предельные распределения. Не все предельные распределения имеют вид симметричного колокола, рассмотренного в разд. 5.2. (Например, биномиальное распределение и распределение Пуассона, рассматриваемые в гл. 10 и 11, обычно не симметричны.) Тем не менее оказывается, что огромное множество измерений имеет в качестве предельного распределения симметричную колоколообразную кривую. Действительно, мы покажем в гл. 10, что если на результат измерения оказывает влияние большое число источников небольших случайных ошибок, а систематические ошибки пренебрежимо малы, то измеренные значения

¹⁾ В литературе для более подготовленных читателей среднее, определяемое формулой (5.15), называется *математическим ожиданием*, или просто *средним*, а определяемое формулой (5.14) — *выборочным средним*. Аналогично употребляются понятия *стандартного отклонения* σ_x , определяемого формулой (5.16), и *выборочного стандартного отклонения* [см. (4.6) и (4.9)]. — *Прим. перев.*

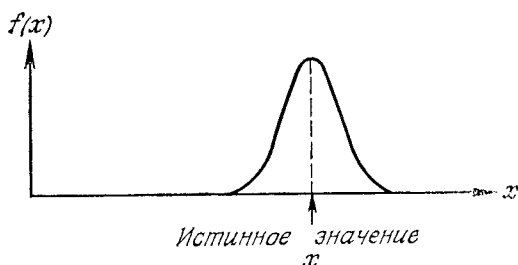


Рис. 5.7. Предельное распределение в случае измерений, результаты которых подвержены влиянию множества небольших случайных ошибок. Распределение имеет колоколообразный вид с центром на истинном значении измеренной величины x .

распределяются по колоколообразной кривой, центр которой будет истинным значением x , как показано на рис. 5.7. В оставшейся части этой главы мы ограничимся рассмотрением измерений только с такими свойствами.

Если наши измерения подвержены заметным систематическим ошибкам, то мы *не* можем ожидать, что предельное распределение будет иметь центр, совпадающий с истинным значением. Случайные ошибки с равной вероятностью смещают наши отсчеты в обе стороны от истинного значения. Если все ошибки случайны, то после многочисленных измерений будет получено одинаково много результатов, как превышающих истинное значение, так и не достигающих его. Однако систематическая ошибка (подобно растянутой рулетке или отстающим часам) смещает все значения в одну сторону и, следовательно, смещает центр распределения полученных значений от истинного значения. В данной главе мы будем предполагать, что центр распределения приходится на истинное значение. Это эквивалентно предположению, что все систематические ошибки уменьшены до пренебрежимо малого уровня.

Теперь настало время обратиться к вопросу, который мы до сих пор избегали обсуждать: что такое «истинное значение» физической величины? Это трудный вопрос, на который нет удовлетворительного простого ответа. Поскольку очевидно, что ни в каком измерении нельзя точно определить истинное значение любой непрерывной переменной (например, длины, времени и т. д.), то не ясно, существует ли вообще истинное значение такой величины. Тем не менее оказывается очень удобным предполагать, что любая физическая величина *имеет* истинное значение, и мы всегда будем исходить из этого предположения.

Истинным значением величины можно считать такое значение, к которому мы приближаемся по мере осуществления

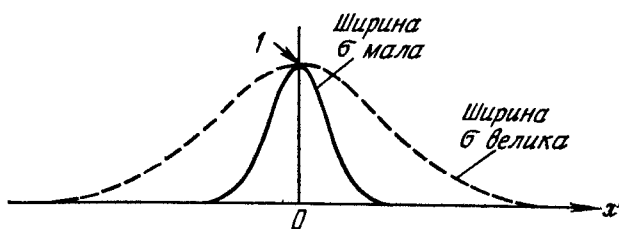


Рис. 5.8. Функция Гаусса (5.17) колоколообразной формы с центром в $x = 0$. Кривая широка, если ширина σ велика, и узка, если σ мала.

все большего числа измерений, выполняемых все более тщательно. Определенное таким образом «истинное значение» есть идеализация, аналогичная понятию математической точки, которая не имеет размеров, или линии, которая не имеет ширины; и, подобно этим двум понятиям, это полезная идеализация. Мы часто будем обозначать истинные значения измеряемых величин $x, y, \dots$ соответствующими прописными буквами $X, Y, \dots$. Если измерения величины x подвержены множеству небольших случайных ошибок и если систематические ошибки пренебрежимо малы, то распределение результатов измерений будет иметь вид симметричной колоколообразной кривой с центром, приходящимся на истинное значение X .

В математике функция, график которой имеет форму колоколообразной кривой, называется *функцией нормального распределения*, или *функцией Гаусса*¹⁾. Основная форма представления этой функции имеет вид

$$e^{-x^2/2\sigma^2}, \quad (5.17)$$

где σ — фиксированный параметр, который мы будем называть *параметром ширины* или *шириной*. Читателю полезно ознакомиться со свойствами этой функции.

Когда $x = 0$, функция Гаусса (5.17) равна единице. Функция симметрична относительно $x = 0$, так как она имеет одни и те же значения как для x , так и для $-x$. При удалении x от нуля в любом направлении функция $x^2/2\sigma^2$ возрастает, причем возрастает быстро, если величина σ мала, и более медленно, если σ велика. Следовательно, по мере удаления x от нуля функция (5.17) стремится к нулю. Таким образом, общий вид функции Гаусса (5.17) будет таким, как показано на рис. 5.8. Этот график объясняет термин «параметр ши-

¹⁾ Другие общепринятые наименования функции Гаусса: гауссова функция (или просто гауссиан), нормальная функция плотности и нормальная функция ошибок. Последнее наименование довольно неудачно, так как обозначение «функция ошибок» часто используется для интеграла от функции Гаусса (который мы рассмотрим в разд. 5.4).

рины» для σ , поскольку при больших значениях σ колокол выглядит широким, а при малых σ — узким.

Функция Гаусса (5.17) представляет собой колоколообразную кривую с центром в $x = 0$. Чтобы получить колоколообразную кривую с центром в какой-то другой точке $x = X$, мы просто заменим x в (5.17) на $x - X$. Таким образом, функция

$$e^{-(x-X)^2/2\sigma^2} \quad (5.18)$$

достигает максимума при $x = X$ и спадает симметрично по обе стороны от $x = X$, как показано на рис. 5.9.

Выражение (5.18) еще не является окончательным, оно не описывает предельное распределение, поскольку любое распределение должно быть *нормировано*, т. е. должно удовлетворять условию

$$\int_{-\infty}^{\infty} f(x) dx = 1. \quad (5.19)$$

Чтобы выполнить это условие, мы введем

$$f(x) = Ne^{-(x-X)^2/2\sigma^2}. \quad (5.20)$$

(Умножение на N не изменяет ни формы, ни положения максимума при $x = X$.) Теперь мы должны выбрать «нормировочный множитель» N таким образом, чтобы $f(x)$ была нормирована в соответствии с (5.19). Для этого необходимо сделать подстановку подынтегрального выражения, которую мы приведем:

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} Ne^{-(x-X)^2/2\sigma^2} dx. \quad (5.21)$$

При оценке интегралов такого вида всегда нужно произвести замену переменных, чтобы упростить подынтегральное

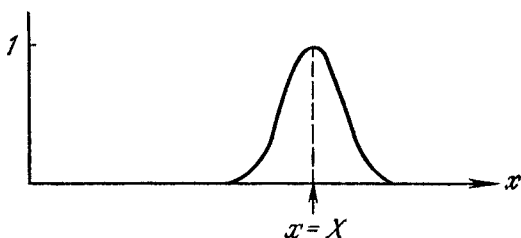


Рис. 5.9. Функция Гаусса (5.18) колоколообразной формы с центром в $x = X$.

выражение. Обозначим $x - X = y$ (в этом случае $dx = dy$) и получим

$$= N \int_{-\infty}^{\infty} e^{-y^2/2\sigma^2} dy. \quad (5.22)$$

Затем мы можем обозначить $y/\sigma = z$ (в этом случае $dy = \sigma dz$), тогда

$$= N\sigma \int_{-\infty}^{\infty} e^{-z^2/2} dz. \quad (5.23)$$

Получившийся интеграл — один из стандартных интегралов математической физики. Его можно рассчитать элементарными методами, но детали вычислений не особенно интересны, поэтому мы просто приведем результат¹⁾

$$\int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi}. \quad (5.24)$$

Возвращаясь к (5.21) и (5.23), находим, что

$$\int_{-\infty}^{\infty} f(x) dx = N\sigma \sqrt{2\pi}.$$

Так как этот интеграл должен быть равен единице, нормировочный множитель N следует выбрать как $N = 1/(\sigma \sqrt{2\pi})$.

Мы приходим к выводу, что нормированная функция Гаусса, или нормированная функция нормального распределения, имеет вид

$$f_{X, \sigma}(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-X)^2/2\sigma^2}.$$

(5.25)

Обратите внимание на то, что мы добавили нижние индексы X, σ , чтобы указать центр и ширину распределения. Функция $f_{X, \sigma}(x)$ описывает предельное распределение результатов измерений величины x , истинное значение которой равно X и на которую оказывают влияние только случайные ошибки. Говорят, что результаты измерений *распределены нормально*, если их предельное распределение описывается функцией Гаусса (5.25).

Вскоре мы выясним значение параметра ширины σ . Однако уже ясно, что малые значения σ приводят к распределе-

¹⁾ Вывод см., например, в книге *Hugh D. Young. Statistical Treatment of Experimental Data*, McGraw-Hill, 1962, приложение D.

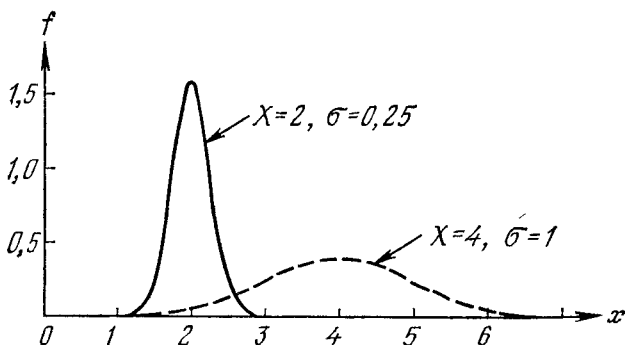


Рис. 5.10. Два нормальных, или гауссовых, распределения.

нию типа острого пика, соответствующего точным измерениям, в то время как большие значения σ дают широкое распределение, соответствующее измерениям с малой точностью. На рис. 5.10 приведены два примера распределения Гаусса с различными центрами X и ширинами σ . Обратите внимание, как фактор σ в знаменателе формулы (5.25) автоматически обеспечивает для более узкого распределения (σ меньше) большую высоту в центре, как это и должно быть, чтобы полная площадь под кривой равнялась 1.

В разд. 5.2 мы видели, что если известно предельное распределение для результатов измерений, то можно вычислить среднее значение $\bar{x}$, ожидаемое в случае очень большого числа измерений. В соответствии с (5.15) ожидаемое среднее значение в случае очень большого числа измерений равно

$$\bar{x} = \int_{-\infty}^{\infty} x f(x) dx. \quad (5.26)$$

Если предельное распределение есть распределение Гаусса $f_{X, \sigma}(x)$ с центром в истинном значении X , то этот интеграл можно вычислить. Прежде чем мы это сделаем, следует заметить, что, как это почти очевидно, среднее значение $\bar{x}$ в случае очень большого числа измерений будет равно X , так как вследствие симметрии функции Гаусса относительно X одинаковое число результатов окажется как больше X на какое-то значение, так и меньше его на то же значение. Таким образом, среднее значение должно равняться X .

Мы можем вычислить интеграл (5.26) для распределения Гаусса следующим образом:

$$\bar{x} = \int_{-\infty}^{\infty} x f_{X, \sigma}(x) dx = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-X)^2/2\sigma^2} dx. \quad (5.27)$$

Если произвести замену переменных $y = x - X$, то получим $dx = dy$ и $x = y + X$. В этом случае интеграл (5.27) можно разбить на два:

$$\bar{x} = \frac{1}{\sigma \sqrt{2\pi}} \left(\int_{-\infty}^{\infty} ye^{-y^2/2\sigma^2} dy + X \int_{-\infty}^{\infty} e^{-y^2/2\sigma^2} dy \right). \quad (5.28)$$

Первый интеграл в этом выражении равен нулю, так как вклад любой точки y точно компенсируется вкладом точки $-y$. Второй интеграл — нормировочный интеграл, встречавшийся нам в (5.22), который равен $\sigma \sqrt{2\pi}$. Это значение сокращается с множителем $\sigma \sqrt{2\pi}$ в знаменателе, и получается ожидаемый результат:

$$\bar{x} = X \quad (5.29)$$

в случае многих измерений. Другими словами, если результаты измерений распределены в соответствии с распределением Гаусса $f_{X, \sigma}(x)$, то в случае очень большого числа измерений среднее значение $\bar{x}$ будет равно истинному значению X , которое соответствует центру функции Гаусса.

Результат (5.29) был бы верен, если бы мы смогли сделать бесконечное число измерений. Его практическая ценность заключается в том, что если мы сделаем большое (но конечное) число измерений, то наше среднее значение будет близко к X .

Другой интересной величиной, которую можно вычислить, является *стандартное отклонение* σ_x в случае многих измерений. В соответствии с (5.16) оно определяется как

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 f_{X, \sigma}(x) dx. \quad (5.30)$$

Этот интеграл легко вычислить. Заменяя $\bar{x}$ на X , производя подстановки $x - X = y$ и $y/\sigma = z$ и, наконец, интегрируя по частям, получаем (см. задачу 5.6)

$$\sigma_x^2 = \sigma^2 \quad (5.31)$$

в случае многих измерений¹⁾. Другими словами, параметр ширины σ функции Гаусса $f_{X, \sigma}(x)$ есть просто стандартное отклонение, которое мы получили бы в результате многих измерений. Это, конечно, объясняет, почему величина σ была использована как параметр ширины и почему σ часто называют стандартным отклонением распределения Гаусса $f_{X, \sigma}(x)$. Однако, строго говоря, σ есть стандартное отклонение, ожидаемое только в случае *бесконечно большого* числа измерений. Если мы сделаем некоторое конечное число измерений

¹⁾ См. примечание переводчика на с. 116. — *Прим. перев.*

(скажем, 10 или 20) величины x , то полученное стандартное отклонение должно быть некоторым приближением к σ , но у нас нет никаких оснований считать, что оно будет *точно* равно σ . В разд. 5.5 мы обсудим дополнительные сведения о среднем и стандартном отклонении, которые можно получить в результате реализации некоторого разумного числа измерений.

5.4. Стандартное отклонение как 68 %-ный доверительный предел

Предельное распределение $f(x)$ результатов измерения некоторой величины x дает возможность вычислить вероятность получения любого данного значения x . Интеграл

$$\int_a^b f(x) dx$$

есть вероятность того, что любое единичное измерение приведет к результату, лежащему внутри интервала $a \leq x \leq b$. Если предельное распределение есть функция Гаусса $f_{x,\sigma}(x)$, то этот интеграл можно вычислить. В частности, мы можем вычислить (как обсуждалось в гл. 4) вероятность того, что результат измерения окажется в пределах одного стандартного отклонения σ от истинного значения X . Эта вероятность равна

$$P \text{ (в пределах } \sigma) = \int_{X-\sigma}^{X+\sigma} f_{x,\sigma}(x) dx = \quad (5.32)$$

$$= \frac{1}{\sigma \sqrt{2\pi}} \int_{X-\sigma}^{X+\sigma} e^{-(x-X)^2/2\sigma^2} dx. \quad (5.33)$$

Смысл этого интеграла проиллюстрирован на рис. 5.11. Этот интеграл можно привести к более простому теперь уже обычной для нас подстановкой $(x - X)/\sigma = z$. В этом случае

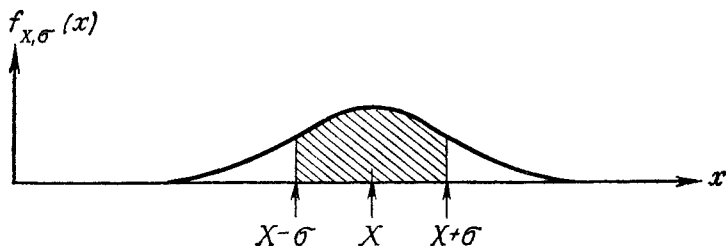


Рис. 5.11. Заштрихованная площадь между $X - \sigma$ и $X + \sigma$ равна вероятности того, что результат измерения будет лежать в пределах одного стандартного отклонения от X .

$dx = \sigma dz$, и пределы интегрирования становятся равными ± 1 . Следовательно,

$$P \text{ (в пределах } \sigma) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{+1} e^{-z^2/2} dz. \quad (5.34)$$

Прежде чем оценивать интеграл (5.34), заметим, что в равной мере мы могли бы найти вероятность того, что результат будет лежать в пределах 2σ от X или $1,5\sigma$ от X . В общем случае мы могли бы вычислить вероятность P (в пределах $t\sigma$), что означает «вероятность того, что результат будет лежать в пределах $t\sigma$ от X », где t — любое положительное число. Эта вероятность показана заштрихованной площадью на рис. 5.12, и расчет, аналогичный приведенному к (5.34), дает (см. задачу 5.7)

$$P \text{ (в пределах } t\sigma) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-z^2/2} dz. \quad (5.35)$$

Интеграл (5.35) — это стандартный интеграл математической физики; он часто называется *функцией ошибок*, обозначаемой $\text{erf}(t)$, или *нормальным интегралом ошибок*. Его нельзя вычислить аналитически, но легко оценить численно с помощью карманного калькулятора. На рис. 5.13 его значения представлены графически и приведены несколько его значений. Более полная таблица значений дана в приложении А в конце книги (см. также приложение Б, где приведены значения для другого, но тесно связанного с рассматриваемым интеграла).

Прежде всего, как можно заметить из рис. 5.13, вероятность того, что результат измерения окажется в пределах одного стандартного отклонения от истинного результата, составляет 68%, как уже принималось в гл. 4 (где говорилось о величине «приблизительно 70%»). Если мы будем рассмат-

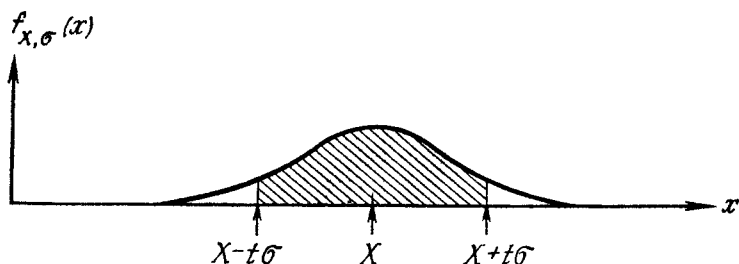


Рис. 5.12. Заштрихованная площадь между $X \pm t\sigma$ и $X - t\sigma$ равна вероятности того, что результат измерения будет лежать в пределах t стандартных отклонений от X .

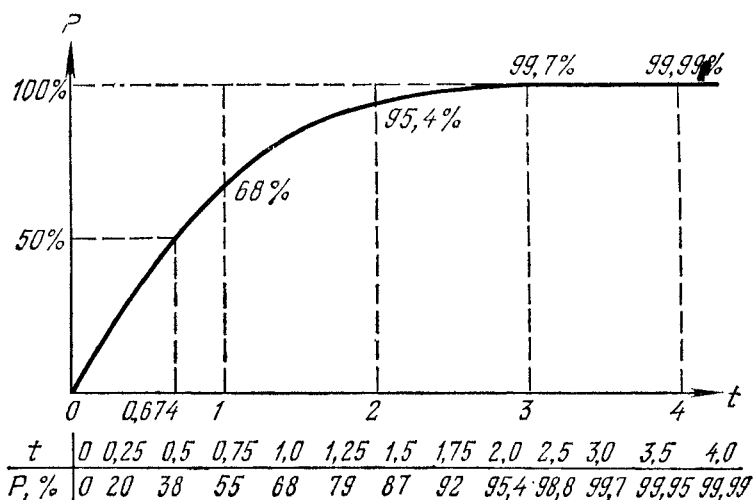


Рис. 5.13. Вероятность P (в пределах $t\sigma$) того, что результат измерения x будет лежать в пределах t стандартных отклонений от истинного значения $x = X$.

ривать стандартное отклонение как нашу погрешность в измерениях (т. е. запишем $x = x_{\text{наил}} \pm \delta x$ и примем $\delta x = \sigma$), то мы можем быть на 68% уверены, что наш результат будет в пределах σ от правильного результата.

Из рис. 5.13 мы также можем видеть, что вероятность P (в пределах $t\sigma$) быстро стремится к 100% с увеличением t . Вероятность того, что результат измерения окажется в пределах 2σ , равна 95,4%; вероятность результата в пределах 3σ — 99,7%. Это же можно выразить и иным способом, а именно: вероятность того, что результат измерения окажется *вне* одного стандартного отклонения, довольно значительна (32%), вероятность того, что он будет лежать вне 2σ — много меньше (4,6%), а того, что он будет лежать за пределами 3σ , исключительно мала (0,3%).

Конечно, нет ничего магического в числе 68%; это просто доверительный уровень, связанный со стандартным отклонением σ . Альтернативой стандартному отклонению может служить так называемая *вероятная ошибка* (ВО)¹⁾, которая определяется как такое отклонение, когда результат измерения с вероятностью 50% окажется внутри интервала $X \pm \text{ВО}$. Из

¹⁾ Автор использует встречающуюся в литературе на английском языке аббревиатуру для вероятной ошибки по первым буквам соответствующих английских слов (Р. Е. — probable error). В литературе на русском языке соответствующая аббревиатура (ВО) для понятия вероятной ошибки не является общеупотребительной. — *Прим. перев.*

рис. 5.13 можно увидеть (для результатов измерений, которые распределены нормально), что вероятная ошибка равна

$$\text{ВО} \approx 0,67\sigma.$$

Некоторые экспериментаторы предпочитают приводить вероятную ошибку ВО в качестве погрешности в своих измерениях. Тем не менее стандартное отклонение σ используется гораздо чаще, поскольку его свойства весьма просты.

5.5. Обоснование среднего как наилучшей оценки

В последних трех разделах мы рассматривали *предельное распределение* $f(x)$, которое получается в случае бесконечного числа измерений величины x . Если бы функция $f(x)$ была известна, то мы могли бы вычислить среднее $\bar{x}$ и стандартное отклонение σ , полученные в случае бесконечно большого числа измерений, и (по крайней мере для нормального распределения) мы могли бы также узнать истинное значение X . К сожалению, мы никогда не знаем предельного распределения. На практике обычно имеется конечное число измеренных значений (5, 10 или, может быть, 50)

$$x_1, x_2, \dots, x_N,$$

и наша задача — найти *наилучшие оценки* для X и σ , основанные на этих N измеренных значениях.

Если бы результаты измерений описывались нормальным распределением $f_{x, \sigma}(x)$ и нам были известны параметры X и σ , то мы могли бы вычислить вероятность получения значений $x_1, \dots, x_N$, которые фактически были получены. Так, вероятность получения отсчета вблизи x_1 в малом интервале dx_1 есть

$$P(x \text{ между } x_1 \text{ и } x_1 + dx_1) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x_1 - X)^2 / 2\sigma^2} dx_1. \quad (5.36)$$

На практике нас не интересуют ни величина интервала dx_1 , ни множитель $\sqrt{2\pi}$, так что можно сократить запись до

$$P(x_1) \sim \frac{1}{\sigma} e^{-(x_1 - X)^2 / 2\sigma^2}. \quad (5.37)$$

Мы будем ссылаться на (5.37) как на вероятность получения значения x_1 , хотя, строго говоря, это есть вероятность получения значения в интервале около x_1 , как в (5.36).

Вероятность получения второго отсчета x_2 есть

$$P(x_2) \sim \frac{1}{\sigma} e^{-(x_2 - X)^2 / 2\sigma^2}, \quad (5.38)$$

и аналогично мы можем записать все остальные вероятности, заканчивая выражением

$$P(x_N) \sim \frac{1}{\sigma} e^{-(x_N - X)^2/2\sigma^2}. \quad (5.39)$$

Формулы (5.37) — (5.39) дают вероятности получения каждого из отсчетов $x_1, \dots, x_N$, рассчитанные в терминах принятого предельного распределения $f_{X, \sigma}(x)$. Вероятность того, что мы будем наблюдать всю совокупность N отсчетов, равна произведению этих отдельных вероятностей¹⁾

$$P_{X, \sigma}(x_1, \dots, x_N) = P(x_1) P(x_2) \cdot \dots \cdot P(x_N),$$

или

$$P_{X, \sigma}(x_1, \dots, x_N) \sim \frac{1}{\sigma^N} e^{-\sum (x_i - X)^2/2\sigma^2}. \quad (5.40)$$

Очень важно понять значение различных величин в (5.40). Числа $x_1, \dots, x_N$ — это фактические результаты N измерений; таким образом, $x_1, \dots, x_N$ — известные фиксированные числа. Величина $P_{X, \sigma}(x_1, \dots, x_N)$ есть вероятность получения N результатов $x_1, \dots, x_N$, вычисленная в терминах X и σ , истинного значения x и ширины его распределения. Числа X и σ *неизвестны*; мы хотим найти наилучшие оценки для X и σ , основываясь на данных наблюдений $x_1, \dots, x_N$. Мы добавили нижние индексы X и σ к обозначению вероятности (5.40), чтобы подчеркнуть, что она зависит от (неизвестных) значений X и σ .

Поскольку действительные значения X и σ неизвестны, мы могли бы зафиксировать некоторые предполагаемые значения X' и σ' и затем с этими предполагаемыми значениями вычислить вероятность $P_{X', \sigma'}(x_1, \dots, x_N)$. Если бы затем мы зафиксировали другую пару предполагаемых значений X'' и σ'' и нашли, что соответствующая вероятность $P_{X'', \sigma''}(x_1, \dots, x_N)$ больше, то мы, естественно, могли бы рассматривать эти новые значения X'' и σ'' как лучшие оценки для X и σ по сравнению с первой парой. Продолжая в том же духе, мы могли бы организовать поиск таких значений X и σ , которые делают вероятность $P_{X, \sigma}(x_1, \dots, x_N)$ максимально большой, а сами эти значения рассматривать как наилучшие оценки для X и σ .

Эта процедура определения наилучших оценок для X и σ называется в статистике *принципом максимального правдопо-*

¹⁾ Мы используем известный результат, а именно что вероятность одновременной реализации нескольких независимых событий равна произведению вероятностей каждого события в отдельности. Например, вероятность того, что при бросании монеты выпадет «орел», равна $1/2$, а вероятность того, что при бросании кости выпадет «шесть», равна $1/6$. Следовательно, вероятность выпадения «орла» и «шести» есть $(1/2) \cdot (1/6) = 1/12$.

добия. Кратко его можно сформулировать следующим образом.

Для данных N наблюдаемых значений $x_1, \dots, x_N$ наилучшими оценками X и σ будут такие значения, для которых эти значения $x_1, \dots, x_N$ наиболее вероятны. Таким образом, наилучшие оценки для X и σ — это такие значения, для которых $P_{X, \sigma}(x_1, \dots, x_N)$ достигает максимума, где

$$P_{X, \sigma}(x_1, \dots, x_N) \sim \frac{1}{\sigma^N} e^{-\sum (x_i - X)^2 / 2\sigma^2}. \quad (5.41)$$

С помощью этого принципа мы легко можем найти наилучшую оценку для истинного значения X . Очевидно, что (5.41) достигает максимума, когда сумма в показателе экспоненты достигает минимума. Таким образом, наилучшая оценка для X — это такое значение X , для которого сумма

$$\sum_{i=1}^N (x_i - X)^2 / \sigma^2 \quad (5.42)$$

достигает минимума. Чтобы найти этот минимум, продифференцируем сумму по X и приравняем производную нулю, что дает

$$\sum_{i=1}^N (x_i - X) = 0,$$

или

$$X = \frac{\sum x_i}{N} \text{ (наилучшая оценка)}. \quad (5.43)$$

Иными словами, наилучшая оценка истинного значения X есть среднее наших N измерений, $\bar{x} = \sum x_i / N$, — результат, который мы принимали без доказательств начиная с гл. 1.

Найти наилучшую оценку σ , ширины предельного распределения, немного труднее, так как вероятность (5.41) представляет собой более сложную функцию σ . Мы должны продифференцировать (5.41) по σ и приравнять производную нулю. (Мы оставляем детали вычислений читателю, см. задачу 5.10.) Эта процедура дает значение σ , при котором (5.41) достигает максимума и которое, следовательно, представляет собой наилучшую оценку σ

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - X)^2} \text{ (наилучшая оценка)}. \quad (5.44)$$

Истинное значение X неизвестно. Таким образом, на практике мы должны заменить X в (5.44) нашей наилучшей оцен-

кой X , а именно средним $\bar{x}$. Это приводит к оценке

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}. \quad (5.45)$$

Другими словами, наша оценка ширины σ предельного распределения есть стандартное отклонение N наблюдаемых значений $x_1, \dots, x_N$, как первоначально было определено в (4.6).

Читатель может быть удивлен тем, что оценка (5.45) совпадает с нашим первоначальным определением стандартного отклонения (4.6), в котором используется N , вместо нашего «улучшенного» определения с делителем $N-1$. В действительности при переходе от наилучшей оценки (5.44) к выражению (5.45) мы умолчали об одной тонкости. Наилучшая оценка (5.44) включает истинное значение X , в то время как в (5.45) мы заменили X на $\bar{x}$ (наилучшую оценку X). Эти числа, вообще говоря, не одинаковы, и легко видеть, что число (5.45) *всегда меньше* или почти равно (5.44)¹⁾. Таким образом, при переходе от (5.44) к (5.45) мы *недооцениваем* ширину σ . Довольно легко оценить, во сколько раз (5.45) меньше (5.44), хотя мы не будем здесь этого делать. В результате получим, что наилучшим приближением к σ будет не (5.45), а то, что получится при умножении (5.45) на множитель $\sqrt{N/(N-1)}$. Таким образом, наш окончательный вывод состоит в том, что наилучшая оценка ширины σ — это «улучшенное» стандартное отклонение для измеренных значений $x_1, \dots, x_N$:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (\text{наилучшая оценка}). \quad (5.46)$$

Теперь, возможно, настало время дать обзор того довольно сложного материала, который мы изложили. Во-первых, если результаты измерения x подвержены только случайным ошибкам, то их предельное распределение есть функция Гаусса $f_{X, \sigma}(x)$ с центром в истинном значении X и с шириной σ . Ширина σ — это 68%-ный доверительный предел, для которого вероятность того, что любое измерение попадает в интервал в пределах σ от истинного значения X , составляет 68%. На практике ни X , ни σ не известны. Вместо них мы знаем наши N измеренных значений $x_1, \dots, x_N$, где N так велико, как позволяют получить наше время и терпение. Основываясь на

¹⁾ Если рассматривать (5.44) как функцию X , то мы видели, что эта функция достигает минимума при $X = \bar{x}$. Таким образом, (5.45) всегда меньше или равно (5.44).

этих N измеренных значениях, мы показали, что наилучшей оценкой истинного значения X будет среднее $\bar{x} = \sum x_i / N$ и наилучшей оценкой ширины σ будет стандартное отклонение σ_x для $x_1, \dots, x_N$, как определено в (5.46). В разд. (5.7) мы обсудим *надежность* $\bar{x}$ как наилучшей оценки X , и аналогично мы могли бы рассмотреть надежность σ_x как наилучшей оценки σ , но здесь мы этого делать не будем.

Все результаты, полученные в последних двух разделах, зависят от предположения, что данные наших измерений распределены нормально¹⁾. Хотя это и разумное допущение, оно относится к разряду допущений, которые трудно проверить на практике, и иногда не совсем верно. Но и в этом случае мы должны подчеркнуть, что, если результаты измерений имеют *не* нормальное распределение, оно почти всегда является *приблизительно* нормальным, и поэтому вполне допустимо использовать понятия этой главы по крайней мере как хорошие приближения.

5.6. Обоснование квадратичного сложения

Теперь мы можем вернуться к проблеме расчета ошибок в косвенных измерениях, которую мы уже рассматривали в гл. 3. Тогда мы постулировали, что если ошибки случайны и независимы, то их можно складывать квадратично в соответствии с определенными стандартными правилами, например либо в соответствии с «простыми правилами» (3.16) и (3.18), либо в соответствии с общим правилом (3.47), которое включает эти «простые» правила как частные случаи. Теперь мы можем обосновать это квадратичное сложение.

Задача расчета ошибок в косвенных измерениях возникает в случае, когда мы измеряем одну или более величин $x, \dots, z$ и определяем их погрешности, а затем используем измеренные значения для расчета некоторой величины $q(x, \dots, z)$. Главная трудность, конечно, состоит в том, чтобы найти погрешность полученного значения q . Если величины $x, \dots, z$ подвержены только случайным ошибкам, то они будут распределены нормально с параметрами ширины²⁾ $\sigma_x, \dots, \sigma_z$, которые мы будем рассматривать как погрешности любого единичного измерения соответствующих величин. Проблема,

¹⁾ И что систематические ошибки уменьшены до пренебрежимо малого уровня.

²⁾ Имея дело с несколькими различными измеренными величинами $x, \dots, z$, мы будем использовать нижний индекс $x, \dots, z$ у параметра ширины соответствующего предельного распределения, чтобы как-то различать эти параметры. Таким образом, σ_x обозначает ширину распределения Гаусса $f_{x, \sigma_x}(x)$ в случае измерений x и т. д.

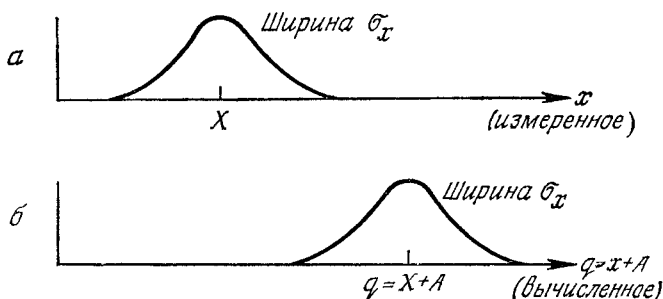


Рис. 5.14. Если измеренные значения x распределены нормально с центром $x = X$ и шириной σ_x , то рассчитанные значения $q = x + A$ (с фиксированным и известным A) будут распределены нормально с центром $q = X + A$ и той же шириной σ_x .

которую теперь надо решить, состоит в следующем: что мы можем сказать о распределении значений q , если известны распределения результатов измерений $x, \dots, z$? И в частности, какова ширина распределения значений q ?

Измеренная величина плюс фиксированное число

Мы начнем с рассмотрения двух очень простых частных случаев. Во-первых, предположим, что мы измеряем некоторую величину x и хотим вычислить значение величины

$$q = x + A, \quad (5.47)$$

где A — некоторое фиксированное число без погрешностей (например, $A = 1$ или π). Предположим, что результаты измерений x распределены нормально около истинного значения X с шириной σ_x , как показано на рис. 5.14, *a*. Тогда вероятность получения любого значения x (в малом интервале dx) равна $f_{X, \sigma_x}(x) dx$ или

$$(\text{вероятность получения значения } x) \sim e^{-(x-X)^2/2\sigma_x^2}. \quad (5.48)$$

Наша цель — найти вероятность получения любого значения величины q , определенной (5.47). Из (5.47) ясно, что $x = q - A$ и, следовательно,

$$\begin{aligned} (\text{вероятность получения значения } q) = \\ = (\text{вероятность получения } x = q - A). \end{aligned}$$

Выражение для второй вероятности дается (5.48), и поэтому (вероятность получения значения q) $\sim$

$$e^{-[(q-A)-X]^2/2\sigma_x^2} = e^{-[q-(X+A)]^2/2\sigma_x^2}. \quad (5.49)$$

Из результата (5.49) следует, что вычисленные значения q распределены нормально с центром в $X + A$ и с шириной σ_x , как показано на рис. 5.14, б. В частности, погрешность в q та же самая (а именно σ_x), что и в x , как дало бы наше правило (3.16).

Измеренная величина умножается на фиксированное число

В качестве второго простого примера рассмотрим случай, когда измеряется x и рассчитывается величина

$$q = Bx,$$

где B — фиксированное число (например, $B = 2$ или $B = \pi$). Если результаты измерений x распределены нормально, то на основании тех же аргументов, что и ранее, мы приходим к выводу, что ¹⁾

(вероятность получения значения q) $\sim$ (вероятность получения $x = q/B$) $\sim$

$$\begin{aligned} &\sim \exp\left[-\left(\frac{q}{B} - X\right)^2 / 2\sigma_x^2\right] = \\ &= \exp\left[-(q - BX)^2 / 2B^2\sigma_x^2\right]. \end{aligned} \quad (5.50)$$

Другими словами, значения $q = Bx$ будут распределены нормально с центром $q = BX$ и с шириной $B\sigma_x$, как показано на рис. 5.15. В частности, погрешность в $q = Bx$ в B раз больше, чем в x , как дало бы наше правило (3.18).

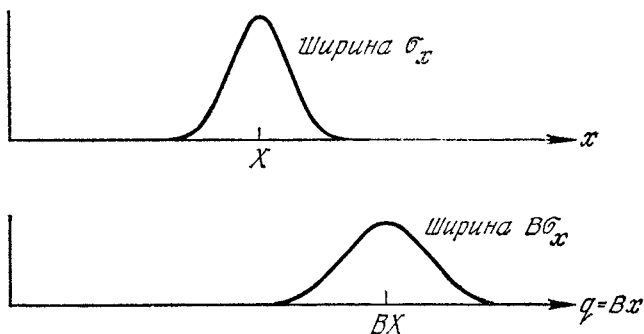


Рис. 5.15. Если измеренные значения x распределены нормально с центром $x = X$ и шириной σ_x , то рассчитанные значения $q = Bx$ (с фиксированным и известным B) будут распределены нормально с центром BX и шириной $B\sigma_x$.

¹⁾ Здесь мы введем альтернативное обозначение $\exp(z)$ для экспоненциальной функции $\exp(z) \equiv e^z$. В случае, когда показатель z становится сложным выражением, обозначение $\exp$ более удобно при написании,

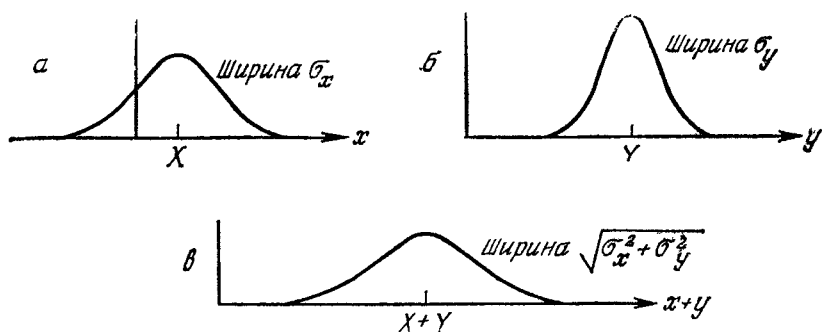


Рис. 5.16. Если результаты измерений x и y независимы и распределены нормально с центрами X и Y и ширинами σ_x и σ_y , то рассчитанные значения $x+y$ будут распределены нормально с центром $X+Y$ и шириной $\sqrt{\sigma_x^2 + \sigma_y^2}$.

Сумма двух измеренных величин

В качестве первого нетривиального примера расчета ошибок в косвенных измерениях рассмотрим случай, когда мы измеряем две независимые величины x и y и вычисляем их сумму $x+y$. Мы будем предполагать, что результаты измерений x и y распределены нормально около соответствующих истинных значений X и Y с ширинами σ_x и σ_y , как показано на рис. 5.16, а и б, и попытаемся определить распределение рассчитанных значений $x+y$. Мы покажем, что значения $x+y$ распределены нормально с центром, равным сумме истинных значений $X+Y$, и с шириной, равной

$$\sqrt{\sigma_x^2 + \sigma_y^2},$$

как показано на рис. 5.16, в. В частности, эта формула обосновывает правило из гл. 3, согласно которому погрешность в $x+y$ равна квадратичной сумме индивидуальных погрешностей в x и y , если x и y подвержены только независимым и случайным погрешностям.

Чтобы упростить наши выкладки, сначала предположим, что оба истинных значения X и Y равны нулю. В этом случае вероятность получения любого частного значения x равна

$$P(x) \sim \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \quad (5.51)$$

и аналогично для y

$$P(y) \sim \exp\left(-\frac{y^2}{2\sigma_y^2}\right). \quad (5.52)$$

Наша задача теперь заключается в том, чтобы определить вероятность получения любого частного значения $x + y$. Сначала отметим, что поскольку x и y измеряются независимо, то вероятность получения любых данных x и y равна произведению вероятностей (5.51) и (5.52):

$$P(x, y) \sim \exp \left[-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right]. \quad (5.53)$$

Зная вероятность получения x и y , мы уже можем рассчитать вероятность любого данного значения $x + y$. Для этого выразим показатель в (5.53) через представляющую интерес переменную $x + y$. Это можно сделать с помощью тождества (которое читатель легко может проверить)

$$\frac{x^2}{A} + \frac{y^2}{B} = \frac{(x+y)^2}{A+B} + \frac{(Bx - Ay)^2}{AB(A+B)} = \quad (5.54)$$

$$= \frac{(x+y)^2}{A+B} + z^2, \quad (5.55)$$

где во второй строке мы ввели сокращенное обозначение z^2 для второго члена в правой части (5.54), поскольку его значение не представляет для нас интереса.

Если подставить (5.55) в (5.53), заменяя A на σ_x^2 и B на σ_y^2 , то получаем

$$P(x, y) \sim \exp \left[-\frac{(x+y)^2}{2(\sigma_x^2 + \sigma_y^2)} - \frac{z^2}{2} \right]. \quad (5.56)$$

Эту вероятность получения данных значений x и y можно также рассматривать как вероятность получения данных значений $x + y$ и z . Таким образом, мы можем переписать (5.56) как

$$P(x + y, z) \sim \exp \left[-\frac{(x+y)^2}{2(\sigma_x^2 + \sigma_y^2)} \right] \exp \left[-\frac{z^2}{2} \right]. \quad (5.57)$$

В конце концов, мы хотим найти вероятность получения данного значения $x + y$ *безотносительно от какого-либо значения z* . Это можно сделать, если просуммировать или, точнее, проинтегрировать (5.57) по всем возможным значениям z , т. е.

$$P(x + y) = \int_{-\infty}^{\infty} P(x + y, z) dz. \quad (5.58)$$

Интегрирование (5.57) по z сводится к интегралу от $\exp(-z^2/2)$, что дает $\sqrt{2\pi}$, и мы получаем

$$P(x + y) \sim \exp \left[-\frac{(x+y)^2}{2(\sigma_x^2 + \sigma_y^2)} \right]. \quad (5.59)$$

Это выражение показывает, что значения $x + y$ распределены нормально с шириной $\sqrt{\sigma_x^2 + \sigma_y^2}$, как и ожидалось.

Наше доказательство закончено для случая, когда оба истинных значения x и y равны нулю: $X = Y = 0$. Если X и Y отличны от нуля, мы можем рассуждать следующим образом. Сначала запишем

$$x + y = (x - X) + (y - Y) + (X + Y). \quad (5.60)$$

В этом выражении два первых члена распределены нормально с центрами, равными нулю, с ширинами σ_x и σ_y в соответствии с (5.49). Следовательно, сумма этих двух первых членов распределена нормально с шириной $\sqrt{\sigma_x^2 + \sigma_y^2}$. Третий член в (5.60) — фиксированное число. Следовательно, в соответствии с (5.49) он смещает распределение к $(X + Y)$, но не изменяет его ширину. Другими словами, значения $(x + y)$, представленные формулой (5.60), распределены нормально около $(X + Y)$ с шириной $\sqrt{\sigma_x^2 + \sigma_y^2}$. А это и есть искомый результат.

Общий случай

Доказав формулу для вычисления ошибки в частном случае суммы $x + y$, мы удивительно просто можем получить формулу для расчета ошибки и в общем случае косвенных измерений. Предположим, что мы измеряем две независимые величины x и y , наблюдаемые значения которых распределены нормально, и вычисляем некоторую величину $q(x, y)$ от переменных x и y . Распределение значений $q(x, y)$ легко находится с помощью трех предыдущих результатов.

Во-первых, ширины σ_x и σ_y (погрешности в x и y) должны быть малы. Это означает, что мы имеем дело только с величинами x , близкими к X , и величинами y , близкими к Y , и поэтому можем, используя аппроксимацию (3.42), написать

$$q(x, y) \approx q(X, Y) + \left(\frac{\partial q}{\partial x}\right)_{x, Y} (x - X) + \left(\frac{\partial q}{\partial y}\right)_{x, Y} (y - Y). \quad (5.61)$$

Это — хорошее приближение, поскольку те значения x и y , которые реализуются наиболее часто, близки к X и Y . Мы привели нижние индексы X, Y у частных производных, чтобы подчеркнуть, что эти производные оцениваются в точке X, Y и, следовательно, являются фиксированными числами.

Приближение (5.61) выражает искомую величину $q(x, y)$ в виде суммы трех членов. Первый член $q(X, Y)$ — фиксированное число, поэтому он только смещает распределение. Второй член — фиксированное число $\partial q / \partial x$, умноженное на

$(x - X)$, распределение значений которого имеет ширину σ_x , поэтому значения второго члена распределены с центром в нуле и с шириной

$$\left(\frac{\partial q}{\partial x}\right) \sigma_x.$$

Аналогично значения третьего члена распределены с центром в нуле и с шириной

$$\left(\frac{\partial q}{\partial y}\right) \sigma_y.$$

Рассматривая все три члена в (5.61) и привлекая уже полученные результаты, мы можем сделать вывод, что значения $q(x, y)$ распределены нормально около истинного значения $q(X, Y)$ с шириной

$$\sqrt{\left(\frac{\partial q}{\partial x} \sigma_x\right)^2 + \left(\frac{\partial q}{\partial y} \sigma_y\right)^2}. \quad (5.62)$$

Если рассматривать стандартные отклонения σ_x и σ_y как погрешности в x и y , то результат (5.62) — это правило (3.47) для расчета случайных ошибок в косвенных измерениях в случае, когда q есть функция двух переменных, $q(x, y)$. Если q зависит от нескольких переменных, т. е. $q = q(x, y, \dots, z)$, то предыдущие аргументы можно использовать непосредственно, чтобы вывести общее правило (3.47) для функции нескольких переменных. Так как все правила гл. 3 (касающиеся расчета ошибок в случае косвенных измерений) могут быть получены из (3.47), то теперь все эти правила получили обоснование.

5.7. Стандартное отклонение среднего

Нам осталось доказать еще один важный результат, приведенный в гл. 4. Это касается стандартного отклонения среднего $\sigma_{\bar{x}}$. Мы доказали (в разд. 5.5), что если производится N измерений $x_1, \dots, x_N$ величины x (которая распределена нормально), то наилучшей оценкой истинного значения X будет среднее $\bar{x}$ от $x_1, \dots, x_N$. В гл. 4 мы утверждали, что погрешность в этой оценке есть стандартное отклонение среднего

$$\sigma_{\bar{x}} = \sigma_x / \sqrt{N}. \quad (5.63)$$

Теперь мы можем доказать это утверждение. Доказательство до удивления кратко, и поэтому вы должны будете внимательно проследить за ним.

Предположим, что результаты измерений x распределены нормально около истинного значения X с шириной σ_x . Мы

хотим узнать, какова *надежность среднего значения* N измерений. Чтобы исследовать этот вопрос, естественно представить себе ситуацию, когда эти N измерений повторяются много раз, т. е. представить выполнение целой последовательности экспериментов, в каждом из которых мы делаем по N измерений и вычисляем среднее значение. Затем мы могли бы поинтересоваться распределением этих многих полученных средних значений для N измерений. И это легко осуществить.

В каждом эксперименте мы получаем N измеренных значений $x_1, \dots, x_N$ и затем вычисляем функцию

$$\bar{x} = \frac{x_1 + \dots + x_N}{N}. \quad (5.64)$$

Рассчитанная величина $\bar{x}$ есть простая функция измеренных значений $x_1, \dots, x_N$, и мы легко можем найти распределение наших ответов для $\bar{x}$ с помощью расчета ошибок для косвенных измерений. Единственная непривычная особенность функции (5.64) состоит в том, что все результаты измерений $x_1, \dots, x_N$ — результаты измерений одной и той же величины с тем же самым истинным значением X и с той же самой шириной σ_x .

Сначала мы отметим, что поскольку каждое из измеренных значений $x_1, \dots, x_N$ распределено нормально, то то же самое должно быть справедливо для функции $\bar{x}$, определяемой (5.64). Далее истинное значение для каждого из $x_1, \dots, x_N$ есть X , поэтому истинное значение величины $\bar{x}$, определяемой (5.64), есть

$$\frac{X + \dots + X}{N} = X.$$

Таким образом, после многократного определения среднего значения $\bar{x}$ для N измерений мы найдем, что все наши многочисленные результаты для $\bar{x}$ будут распределены нормально около истинного значения X . Единственный оставшийся (и наиболее важный) вопрос состоит в том, чтобы найти ширину нашего распределения средних. В соответствии с (5.62) применительно к случаю N переменных эта ширина есть

$$\sigma_{\bar{x}} = \sqrt{\left(\frac{\partial \bar{x}}{\partial x_1} \sigma_{x_1}\right)^2 + \dots + \left(\frac{\partial \bar{x}}{\partial x_N} \sigma_{x_N}\right)^2}. \quad (5.65)$$

Так как $x_1, \dots, x_N$ представляют собой результаты измерений одной и той же величины x , то и ширины у них у всех одни и те же и равны σ_x :

$$\sigma_{x_1} = \dots = \sigma_{x_N} = \sigma_x.$$

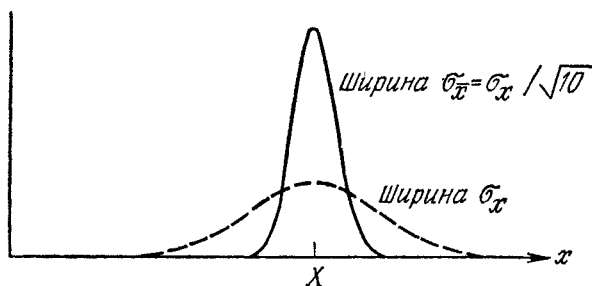


Рис. 5.17. Результаты индивидуальных измерений величины x распределены нормально около X с шириной σ_x (пунктирная кривая). Если мы будем использовать то же самое оборудование для определения многих средних значений 10 измерений, то значения $\bar{x}$ будут распределены нормально около X с шириной $\sigma_{\bar{x}} = \sigma_x / \sqrt{10}$ (сплошная кривая).

Из (5.64) мы также видим, что все частные производные в (5.65) одинаковы:

$$\frac{\partial \bar{x}}{\partial x_1} = \dots = \frac{\partial \bar{x}}{\partial x_N} = \frac{1}{N}.$$

Следовательно, (5.65) сводится к

$$\begin{aligned} \sigma_{\bar{x}} &= \sqrt{\left(\frac{1}{N} \sigma_x\right)^2 + \dots + \left(\frac{1}{N} \sigma_x\right)^2} = \\ &= \sqrt{N \frac{\sigma_x^2}{N^2}} = \\ &= \sigma_x / \sqrt{N}, \end{aligned} \quad (5.66)$$

что и требуется.

К искомому результату (5.66) мы пришли столь быстро, что, вероятно, следует остановиться и понять, что же он означает. Мы представили себе, что выполняется большое число экспериментов, в каждом из которых производилось по N измерений x и затем вычислялось среднее значение $\bar{x}$ этих N измерений. Мы показали, что в результате многократного повторения такого эксперимента наши многочисленные значения $\bar{x}$ будут распределены нормально с центром, равным истинному значению X , и с шириной, которая определяется выражением $\sigma_{\bar{x}} = \sigma_x / \sqrt{N}$, как показано на рис. 5.17 для случая $N = 10$. Эта ширина $\sigma_{\bar{x}}$ есть 68 %-ный доверительный предел для нашего эксперимента. Если мы найдем среднее N измерений *однажды*, то мы можем быть на 68% уверены, что наш результат будет лежать в пределах $\sigma_{\bar{x}}$ от истинного значения X . Именно это мы и хотели бы понимать под *погрешностью среднего*. Этот результат также поясняет,

почему такая погрешность называется стандартным отклонением среднего.

С помощью такого простого и изящного доказательства мы обосновали все результаты относительно случайных погрешностей, приведенные в предыдущих главах.

5.8. Коэффициент доверия

Теперь мы можем вернуться к двум вопросам, затронутым впервые в гл. 2, на которые до сих пор не было дано исчерпывающего ответа. Во-первых, какой смысл мы вкладываем в ставшее уже привычным выражение: «Мы в разумной степени уверены, что некоторая измеренная величина лежит в интервале $x_{\text{наил}} \pm \delta x$ »? Или, выражаясь определеннее, как можно дать количественную характеристику степени нашего доверия к любому экспериментальному результату?

Что касается первого вопроса, то ответ должен быть теперь ясен. Если мы измеряем величину x несколько раз (как это обычно бывает), то наша наилучшая оценка для x есть среднее $\bar{x}$, а его стандартное отклонение $\sigma_{\bar{x}}$ есть наша наилучшая оценка погрешности среднего. Мы могли бы сделать вывод, что

$$(\text{значение } x) = \bar{x} \pm \sigma_{\bar{x}},$$

подразумевая под этим, что согласно нашим наблюдениям можно ожидать, что в 68 % случаев результаты любых последующих измерений $\bar{x}$, сделанных с той же тщательностью, попадут в интервал $\bar{x} \pm \sigma_{\bar{x}}$.

Нашу погрешность можно оценить иначе. Например, мы могли бы предпочесть такую характеристику:

$$(\text{значение } x) = \bar{x} \pm 2\sigma_{\bar{x}}.$$

В этом случае указывался бы интервал, в котором, как мы ожидаем, будет лежать 95% результатов всех однотипных измерений. Ясно, что при представлении любого измеренного значения главное — привести интервал (или погрешность измерения) и *коэффициент доверия, соответствующий этому интервалу*. Наиболее часто приводится стандартное отклонение результата, которое понимается как 68%-ный доверительный предел, т. е. коэффициент доверия в этом случае равен 68%.

Как подчеркивалось в гл. 2, почти все экспериментальные заключения содержат сравнение двух или более чисел. Вооружившись статистической теорией, мы теперь можем дать количественные критерии для многих таких сравнений. Сейчас мы рассмотрим только один тип эксперимента, в котором

получают некоторое число и сравнивают полученный результат с известным ожидаемым ответом. Заметьте, что под этот частный случай попадают многие интересные эксперименты. Например, в эксперименте по проверке закона сохранения импульса мы могли бы измерить начальный и конечный импульсы p и p' , чтобы проверить, что $p = p'$ (в пределах погрешностей), но мы можем в равной мере считать, что ищется значение $(p - p')$, которое сравнивается с ожидаемым ответом, равным нулю. В общем случае, когда мы хотим сравнить результаты любых двух измерений, в которых, по предположению, измеряется одна и та же величина, мы можем образовать их разность и сравнивать ее с ожидаемым ответом, равным нулю. Любой другой эксперимент, в котором измеряется величина (подобная g , ускорению свободного падения), для которой известно точное принятое значение, также попадает под этот тип экспериментов, причем ожидаемым результатом является известное принятое значение.

Предположим, что студент измеряет некоторую величину x (подобную разности двух импульсов, которые предположительно равны) в виде

$$(\text{значение } x) = x_{\text{наил}} \pm \sigma,$$

где σ обозначает стандартное отклонение его результата. (Это будет стандартное отклонение среднего, если $x_{\text{наил}}$ есть среднее нескольких измерений.) Пусть затем он сравнивает свой результат с ожидаемым ответом $x_{\text{ож}}$.

В гл. 2 мы отмечали, что если различие $|x_{\text{наил}} - x_{\text{ож}}|$ меньше (или только незначительно больше), чем σ , то согласие будет удовлетворительным, но если $|x_{\text{наил}} - x_{\text{ож}}|$ много больше, чем σ , то оно неудовлетворительно. Сами по себе эти критерии правильны, но не дают никакой количественной характеристики того, насколько хорошо или плохо это согласие. Они также не говорят нам ничего о границах, которые еще допустимы для согласия. Будет ли различие в $1,5\sigma$ свидетельствовать об удовлетворительном согласии? А в 2σ ?

Теперь мы сможем ответить на эти вопросы, если предположим, что результаты измерений нашего студента подчиняются нормальному распределению (а это определенно разумно). Предположим две рабочие гипотезы относительно этого распределения:

а) центр распределения совпадает с ожидаемым ответом $x_{\text{ож}}$;

б) параметр ширины распределения равен оцененной студентом величине σ .

Гипотеза «а», конечно, заключается в том, что студент при измерении получает правильный ответ. Она добавляет к сделанным допущениям еще то, что систематические ошибки

уменьшены до пренебрежимо малого уровня (вследствие чего распределение имеет центр на истинном значении) и что истинное значение на самом деле равно $x_{ож}$ (т. е. основания, в соответствии с которыми ожидается $x_{ож}$, правильны). Гипотеза «б» представляет собой некоторое приближение, так как σ — лишь оценка стандартного отклонения, но это хорошее приближение, если число измерений, на основании которых определено значение σ , велико¹⁾. Две наши гипотезы, вместе взятые, добавляют к сделанным ранее допущениям еще одно: измерения и вычисления студента в основном правильны.

Теперь мы должны решить, является ли полученное студентом значение $x_{наил}$ разумным, если справедливы наши гипотезы. В случае утвердительного ответа нет оснований сомневаться в гипотезах, и тогда все в порядке, но если ответ отрицателен, то в гипотезах следует усомниться и студент должен проанализировать возможные ошибки в измерениях или в расчетах, найти ранее не обнаруженные систематические ошибки или обнаружить, что ожидаемое значение $x_{ож}$ неверно.

Сначала определим различие $|x_{наил} - x_{ож}|$ и затем величину

$$t = \frac{|x_{наил} - x_{ож}|}{\sigma}, \quad (5.67)$$

число стандартных отклонений, на которое $x_{наил}$ отличается от $x_{ож}$. Затем по таблицам интеграла нормальных ошибок из приложения А можно найти вероятность (которую дают наши гипотезы) получения результата, отличающегося от $x_{ож}$ на t или более стандартных отклонений. Эта вероятность есть

$$P(\text{вне } t\sigma) = 1 - P(\text{в пределах } t\sigma). \quad (5.68)$$

Если эта вероятность велика, то различие $|x_{наил} - x_{ож}|$ вполне разумно и результат $x_{наил}$ приемлем; если же вероятность (5.68) «недопустимо мала», то различие следует рассматривать как *значимое* (т. е. неприемлемое), и в этом случае наш

¹⁾ Мы собираемся судить о надежности измерения величины $x_{наил}$, сравнивая $|x_{наил} - x_{ож}|$ с σ , нашей оценкой ширины нормального распределения. Если число измерений, на основании которых определяется σ , мало, то эта оценка может быть довольно ненадежной и соответствующие доверительные пределы не точны (хотя все еще полезны как грубые оценки). В случае малого числа измерений можно точно вычислить доверительные пределы при помощи так называемого «распределения Стьюдента для t », которое учитывает возможные вариации нашей оценки σ для ширины. См. Alder H. L., Roessler E. B., Introduction to Probability and Statistics, W. H. Freeman. 6th ed., 1977, ch. 10.

незадачливый студент должен попытаться определить, где он допустил ошибку.

Предположим, например, что различие $|x_{\text{наил}} - x_{\text{ож}}|$ равно одному стандартному отклонению. Вероятность такого или большего различия составляет привычные 32%. Ясно, что различие в одно стандартное отклонение вполне допустимо и потому незначимо. В противоположном экстремальном случае вероятность P (вне 3σ) составляет 0,3%, и если наши гипотезы верны, то крайне невероятно, чтобы мы могли получить различие в 3σ . Меняя порядок аргументов, можно сказать, что если различие, полученное студентом, равно 3σ , то крайне невероятно, чтобы наши гипотезы были верны.

Граница между принятием и отвержением гипотезы зависит от уровня, ниже которого мы рассматриваем различие как неразумно маловероятное. Этот уровень зависит от точки зрения экспериментатора. Однако многие считают, что 5% — хорошая граница для «неразумно малых вероятностей». Если принять этот уровень, то различие в 2σ было бы уже неприемлемым, так как P (вне 2σ) = 4,6%. Действительно, из таблицы в приложении А мы видим, что любое различие, превышающее $1,96\sigma$, неприемлемо на этом 5%-ном уровне. На 2%-ном уровне было бы неприемлемым любое различие, превышающее $2,32\sigma$, и т. д.

Итак, у нас до сих пор нет четкого ответа на вопрос, приемлемо или неприемлемо определенное измеренное значение $x_{\text{наил}}$. Однако наша теория нормального распределения дала нам ясную количественную меру разумности любого частного результата. И это лучшее, на что мы могли бы надеяться.

Имеются, конечно, более сложные типы экспериментов, анализ результатов которых требует соответственно более сложных теорий. Однако большинство основных принципов было уже проиллюстрировано нами на простом и важном примере. Читатель, который захочет познакомиться с дополнительными примерами, может найти некоторые из них в ч. II этой книги.

Задачи

Напоминание: звездочка у номера задачи означает, что задача решается или ее ответ приводится в разделе «Ответы» в конце книги.

*5.1 (разд. 5.1). Студент измеряет моменты количества движения L_i и L_f вращающейся системы до и после добавления дополнительной массы. Чтобы проверить закон сохранения момента количества движения, он вычисляет $L_i - L_f$ (ожидаая результата 0). Студент повторяет измерение 50 раз и сортирует полученные данные по бинам, как показано в табл. 5.3, где приведены его результаты (в некоторых условных единицах) после 5, 10 и 50 измерений. Начертите соответствующие гистограммы для каждого из этих трех случаев. (Будьте внимательны в выборе масштабов, площадь

Таблица 5.3

Число результатов	Бин								
	-9	-7	-5	-3	-1	1	3	5	7
	до -7	до -5	до -3	до -1	до 1	до 3	до 5	до 7	до 9
После 5 измерений	0	1	2	0	1	0	1	0	0
После 10 измерений	0	1	2	2	3	1	1	0	0
После 50 измерений	1	3	7	8	10	9	6	4	2

каждого прямоугольника должна равняться доле событий в соответствующем бине.)

*5.2 (разд. 5.2). Предельное распределение результатов в некотором гипотетическом измерении имеет вид

$$f(x) = \begin{cases} C & \text{для } |x| < a, \\ 0 & \text{в остальных точках.} \end{cases}$$

а. Используйте условие нормировки (5.13) и вычислите постоянную C через a .

б. Начертите предельное распределение. В чем смысл постоянной a ?

в. Используя формулы (5.15) и (5.16), вычислите среднее $\bar{x}$ и стандартное отклонение, которые получились бы в результате большого числа измерений.

5.3 (разд. 5.3). Используя подходящую миллиметровую бумагу и хорошо размеченные оси координат, постройте хороший график распределения Гаусса

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

для $X = 2$, $\sigma = 1$ и для $X = 3$, $\sigma = 0.3$. Используйте калькулятор для расчета значений $f_{X,\sigma}(x)$. Если у калькулятора есть регистры памяти для хранения значений $\sigma\sqrt{2\pi}$ и $-2\sigma^2$, то это ускорит ваши расчеты. Если вы будете помнить, что функция симметрична относительно $x = X$, то это сократит вычисления наполовину. Представьте оба графика на одном листе для сравнения.

*5.4 (разд. 5.3). Если вы не построили, то начертите гистограмму для третьего случая задачи 5.1. Студент, выполняя задачу 5.1, решил, что распределение его результатов согласуется с функцией Гаусса $f_{X,\sigma}(x)$, имеющей центр в точке $X = 0$ и ширину $\sigma = 3.4$. Представьте это распределение на том же графике и сравните его с гистограммой. (Прочитайте указания к задаче 5.3. Заметьте, что у вас нет количественного критерия для сравнения степени согласования двух графиков; все, что вы можете, — это посмотреть, хорошо ли функция Гаусса аппроксимирует гистограмму.)

5.5 (разд. 5.3). Ширина распределения Гаусса обычно характеризуется параметром σ . Альтернативный параметр с простой геометрической интерпретацией — *полная ширина на половине высоты*, или ПШПВ¹⁾.

¹⁾ Автор использует принятую в литературе на английском языке аббревиатуру для понятия «полная ширина на половине высоты», составленную из первых букв соответствующих английских слов (full width at half maximum — FWHM). Аналогичная аббревиатура в литературе на русском языке ПШПВ не является общепринятой. — *Прим. перев.*

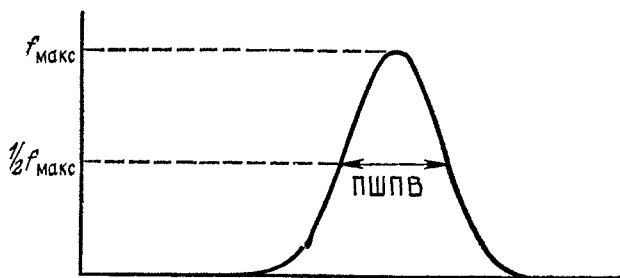


Рис. 5.18. Полная ширина на половине высоты (ПШПВ).

Это — расстояние между двумя точками x , в которых $f_{X,\sigma}(x)$ равна половине своего максимального значения, как показано на рис. 5.18. Докажите, что

$$\text{ПШПВ} = 2\sigma \sqrt{2 \ln 2} = 2,35\sigma.$$

Это означает, что половина максимального значения достигается в точках $X \pm 1,17\sigma$ или, очень приблизительно, $X \pm \sigma$.

***5.6** (разд. 5.3). Выполните детальные выкладки, которые ведут от (5.30) к (5.31), чтобы показать, что стандартное отклонение σ_x в случае большого числа измерений, результаты которых распределены нормально с параметром ширины σ , равно $\sigma_x = \sigma$.

5.7 (разд. 5.4). Если результаты измерений некоторой величины x описываются функцией Гаусса $f_{X,\sigma}(x)$, то вероятность получить значение между $X - t\sigma$ и $X + t\sigma$, очевидно, есть

$$P(\text{в пределах } t\sigma) = \int_{X-t\sigma}^{X+t\sigma} f_{X,\sigma}(x) dx.$$

Докажите подробно, указывая все необходимые замены переменных, что

$$P(\text{в пределах } t\sigma) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{+t} e^{-z^2/2} dz. \quad (5.69)$$

При каждой замене переменных тщательно проверяйте пределы интегрирования. Интеграл (5.69) часто называют функцией ошибок, или интегралом ошибок, и обозначают $\text{erf}(t)$.

***5.8** (разд. 5.4). Студент измеряет некоторую величину y много раз и вычисляет среднее $\bar{y} = 23$ и стандартное отклонение $\sigma_y = 1$. Какую долю отсчетов студента вы ожидали бы найти между

- 22 и 24?
- 22,5 и 23,5?
- 21 и 25?
- 21 и 23?
- 24 и 25?

Наконец, е) в каком интервале (эквидистантном по обе стороны от среднего) вы ожидали бы найти 50 % его отсчетов?

Вся необходимая информация для решения этого вопроса приведена на рис. 5.13. Дополнительная информация относительно вероятностей событий такого рода приведена в приложениях А и Б.

5.9 (разд. 5.4). Массовые обследования показали, что распределение по росту мужчин в некоторой стране нормальное со средним $\bar{h} = 174$ см и стандартным отклонением $\sigma = 5$ см. В случайной выборке из 1000 мужчин сколько из них будут (по вашим ожиданиям) иметь рост

- а) между 169 см и 179 см?
- б) больше, чем 179 см?
- в) больше, чем 189 см?
- г) между 164 и 169 см?

*5.10 (разд. 5.5). Предположим, что мы произвели N измерений $x_1, \dots, x_N$ одной и той же величины x и полагаем, что соответствующее предельное распределение должно быть функцией Гаусса $f_{X, \sigma}(x)$ с неизвестными X и σ . Согласно принципу максимального правдоподобия, наилучшей оценкой для ширины будет такое значение σ , для которого вероятность $P_{X, \sigma}(x_1, \dots, x_N)$ для наблюдаемых значений $x_1, \dots, x_N$ максимальна. Проинтегрируйте $P_{X, \sigma}(x_1, \dots, x_N)$ в (5.41) по σ и покажите, что максимум достигается для значения σ , определяемого (5.44). Как уже обсуждалось после получения (5.44), этот результат означает, что наилучшей оценкой для σ является стандартное отклонение N наблюдаемых значений $x_1, \dots, x_N$.

5.11 (разд. 5.6). Проверьте тождество (5.54), использованное для обоснования квадратичного сложения при расчете ошибок в косвенных измерениях.

*5.12 (разд. 5.7). Ниже приведены результаты сорока измерений $t_1, \dots, t_{40}$ времени падения камня от окна до земли (все в сотых долях секунды):

63	58	74	78	70	74	75	82	68	69
76	62	72	88	65	81	79	77	66	76
86	72	79	77	60	70	65	69	73	77
72	79	65	66	70	74	84	76	80	69

а. Вычислите стандартное отклонение σ_t для этих 40 измерений.

б. Вычислите средние $\bar{t}_1, \dots, \bar{t}_{10}$ по четырем измерениям в каждом из десяти столбцов. Полученные данные можно рассматривать как результаты десяти экспериментов, в каждом из которых определяется *среднее четырех измерений*. Считая известным результат задания а, определите значение для стандартного отклонения, ожидаемое для десяти средних $\bar{t}_1, \dots, \bar{t}_{10}$? Чему оно равно в действительности?

в. Начертите гистограммы для 40 индивидуальных измерений $t_1, \dots, t_{40}$ и для десяти средних $\bar{t}_1, \dots, \bar{t}_{10}$. Используйте одинаковые масштабы и размеры бинов для обоих графиков, чтобы их легко можно было сравнивать. Размеры и границы бинов можно выбирать разными способами, из которых, возможно, простейший состоит в том, чтобы границу одного бина выбрать при значении среднего всех 40 измерений (72,90) и использовать бины, размеры которых равны стандартному отклонению десяти средних значений $\bar{t}_1, \dots, \bar{t}_{10}$.

*5.13 (разд. 5.8). Студент измеряет g , ускорение свободного падения, многократно и тщательно и получает конечный результат $9,5 \text{ м/с}^2$ и стандартное отклонение, равное $0,1$. Если считать, что результаты его измерений распределены нормально с центром, равным принятому значению $9,8$ и с шириной $0,1$, то какова вероятность получения результата, который отличается от $9,8$ так же (или более), как ответ студента? Предполагая, что студент не сделал фактических ошибок, можете ли вы сказать, что, вероятно, его эксперимент подвержен влиянию некоторых невыявленных систематических ошибок?

5.14 (разд. 5.8). Два студента А и Б измеряют одну и ту же величину x и получают результаты $x_A = 13 \pm 1$ и $x_B = 15 \pm 1$, в которых в качестве погрешностей указаны стандартные отклонения.

- а. Предполагая, что все ошибки независимы и случайны, найдите разность $x_A - x_B$ и ее погрешность.
- б. Предполагая, что все результаты в соответствии с ожиданием распределены нормально, что вы можете сказать о вероятности получения такого различия, как у студентов? Считаете ли вы это различие значимым (на 5 %-ном уровне)?

***5.15** (разд. 5.8). Экспериментатор хочет проверить закон сохранения энергии для некоторой ядерной реакции и получает значения начальной и конечной энергий соответственно $E_i = 75 \pm 3$ МэВ и $E_f = 60 \pm 9$ МэВ, где в качестве погрешностей приведены стандартные отклонения результатов. Является ли это различие значимым (на 5 %-ном уровне)? Четко сформулируйте ваши аргументы при ответе на этот вопрос.

ЧАСТЬ II

6. Отбрасывание данных
7. Взвешенные средние
8. Аппроксимация методом наименьших квадратов
9. Смешанный второй момент и корреляция
10. Биномиальное распределение
11. Распределение Пуассона
12. Критерий χ^2 для распределений

Если вы прочитали и поняли содержание гл. 5, то для вас теперь не составит труда изучить ряд более сложных вопросов. Семь глав в ч. II содержат семь таких вопросов, причем одни из них являются приложениями уже развитой статистической теории, а другие — ее дальнейшим развитием. Все эти вопросы важны, и вдумчивый студент обязан разобраться в них рано или поздно. С другой стороны, вовсе не обязательно изучать их все одновременно. По этой причине вопросы излагаются в независимых коротких главах, которые можно изучать в любом порядке в соответствии с вашими нуждами и интересами.

Глава 6

Отбрасывание данных

В этой главе мы обсудим довольно щекотливый вопрос, отбрасывать ли результат измерений, который кажется до такой степени неразумным, что похож скорее на ошибку.

6.1. Проблема отбрасывания данных

Иногда результат одного из серии измерений поразительно расходится со всеми остальными. Когда это происходит, экспериментатор должен решить, является ли такой аномальный результат измерения следствием некоторой ошибки и поэтому должен быть отброшен, или же это законный результат, который должен рассматриваться наряду с другими. Например, представим себе, что мы делаем шесть измерений периода колебаний маятника и получаем результаты (в секундах)

$$3,8; 3,5; 3,9; 3,9; 3,4; 1,8. \quad (6.1)$$

В этом примере значение 1,8 поразительно отличается от остальных, и мы должны решить, что с ним делать.

Из гл. 5 нам известно, что результат нормального измерения *может* значительно отличаться от результатов других измерений той же самой величины. Тем не менее в обычной ситуации столь большое различие, как в случае последнего измерения в (6.1), *очень невероятно*; поэтому мы можем подозревать, что время 1,8 с является результатом какой-то незамеченной ошибки или обусловлено какой-то внешней причиной. Возможно, мы просто ошиблись при считывании этого последнего значения времени, или, может быть, наш электронный секундомер остановился во время последнего измерения из-за внезапного нарушения контакта с блоком питания.

Если бы мы очень тщательно следили за каждым измерением, то иногда могли бы обнаружить какую-то определенную причину аномального результата. Например, наши записи результатов могли бы показать, что в случае последнего измерения в (6.1) использовался другой секундомер, а последующая проверка могла бы показать, что он отстает. В этом случае результат аномального измерения определенно следует отбросить.

К сожалению, обычно не удастся найти какую-то внешнюю причину аномального результата. В этом случае мы должны решить, отбросить этот результат или нет, опираясь только на сами результаты, и тогда наши знания распределения Гаусса оказываются полезными.

Отбрасывание данных — спорный вопрос, по которому у специалистов нет единого мнения. Но это также и *важный* вопрос. В нашем примере наилучшая оценка периода колебаний маятника существенно зависит от того, отбросим ли мы подозреваемое значение 1,8 с. Среднее всех шести измерений равно 3,4 с, в то время как среднее пяти измерений равно 3,7 с, т. е. существенно отличается.

Кроме того, решение отбросить какие-то данные в конечном счете всегда субъективно, и ученого, который принял такое решение, его коллеги могут осудить за такую «подгонку» данных. Однако ситуация осложняется, если учесть вероятность того, что аномальный результат может отражать некоторые важные эффекты. В самом деле, многие важные научные открытия сначала выглядели как аномальные результаты измерений, которые походили скорее на ошибки. Отбрасывая значение времени 1,8 с в примере (6.1), мы, *возможно*, выбрасываем наиболее интересную часть данных.

Действительно, единственно честная реакция при встрече с данными, подобными (6.1), — повторять измерения много раз. Если аномалия повторится снова, мы, вероятно, будем в состоянии выяснить ее причину, будь то ошибка или реальный физический эффект; если же она не повторится, то к тому времени мы сделаем, скажем, 100 измерений, так что не будет существенной разницы для нашего конечного результата, включим мы в расчет аномалию или нет.

Тем не менее в большинстве случаев непрактично (особенно в учебной лаборатории) повторять измерение 100 раз, если только результат покажется подозрительным. Следовательно, нам необходим критерий, согласно которому отвергается подозрительный результат. Имеется несколько таких критериев, причем некоторые из них довольно сложные. Критерий, который мы опишем, называется критерием Шовене; это простой и поучительный случай применения распределения Гаусса.

6.2. Критерий Шовене

Вернемся к шести измерениям примера (6.1):

3,8; 3,5; 3,9; 3,9; 3,4; 1,8.

Если мы предположим на время, что все эти значения — законные результаты измерений величины x , то можно вычис-

лить среднее $\bar{x}$ и стандартное отклонение σ_x :

$$\bar{x} = 3,4 \text{ с} \quad (6.2)$$

и

$$\sigma_x = 0,8 \text{ с.} \quad (6.3)$$

Теперь мы можем установить количественный предел, указывающий, до какой степени подозрительный результат 1,8 аномален. Он отличается от среднего значения 3,4 на 1,6, или на два стандартных отклонения. Если мы предположим, что результаты измерений подчиняются распределению Гаусса с центром и шириной, определяемыми выражениями (6.2) и (6.3), то мы можем вычислить вероятность получения результатов, которые по крайней мере так же сильно отличаются от среднего. В соответствии с данными, приведенными на рис. 5.13, эта вероятность равна

$$\begin{aligned} P(\text{вне } 2\sigma_x) &= 1 - P(\text{в пределах } 2\sigma_x) \\ &= 1 - 0,95 \\ &= 0,05. \end{aligned}$$

Другими словами, предполагая, что значения (6.2) и (6.3) для $\bar{x}$ и σ_x справедливы, мы ожидали бы, что только один результат из 20 отличается от среднего по крайней мере так же сильно, как отличается подозрительное число 1,8 с. Проведя 20 или более измерений, мы действительно должны были бы *ожидать* появления одного или двух результатов настолько плохих, как 1,8 с, и тогда не было бы оснований отбрасывать их. Но мы произвели только шесть измерений, поэтому ожидаемое число результатов, которые были бы так плохи, как 1,8 с, в действительности равно

$$0,05 \cdot 6 = 0,3,$$

т. е. для имеющихся шести результатов измерений вероятность того, что хотя бы одно из этих значений будет настолько плохим, как 1,8 с, составляет $1/3$.

Это число дает нам искомую количественную меру «разумности» подозрительного результата. Если считать, что число $1/3$ «до смешного невероятно», то мы придем к выводу, что значение 1,8 с — ненормальный результат, который должен быть отброшен.

Выбор границы, начиная с которой результат становится «до смешного невероятным», принадлежит экспериментатору. Критерий Шовене в его обычном понимании утверждает, что если ожидаемое число измерений, столь же плохих, как и подозрительный результат, меньше чем $1/2$, то подозритель-

ный результат следует исключить. Очевидно, выбор величины $1/2$ произволен; но он также разумен, и его можно оправдать.

Теперь легко описать применение критерия Шовене к общей задаче. Предположим, мы делаем N измерений

$$x_1, \dots, x_N$$

одной и той же величины x . Учитывая все N измерений, мы вычисляем $\bar{x}$ и σ_x . Если один из результатов измерений отличается от $\bar{x}$ настолько, что представляется подозрительным (обозначим его $x_{\text{под}}$), то мы сначала вычисляем

$$t_{\text{под}} = \frac{x_{\text{под}} - \bar{x}}{\sigma_x}, \quad (6.4)$$

число стандартных отклонений, на которое $x_{\text{под}}$ отличается от $\bar{x}$. Затем мы находим (из рис. 5.13 или из более полной таблицы в приложении А) вероятность P (вне $t_{\text{под}}\sigma_x$) того, что нормальное измерение будет отличаться от $\bar{x}$ на $t_{\text{под}}$ или более стандартных отклонений. Наконец, мы умножаем на N , полное число измерений, чтобы получить

$$n(\text{хуже, чем } x_{\text{под}}) = NP \text{ (вне } t_{\text{под}}\sigma_x).$$

Полученное значение n — число ожидаемых измерений, которые дают столь же плохие результаты, как $x_{\text{под}}$. Если n меньше $1/2$, то $x_{\text{под}}$ не удовлетворяет критерию Шовене и отвергается.

После отбрасывания результата, не удовлетворяющего критерию Шовене, естественно, надо пересчитать $\bar{x}$ и σ_x по оставшимся данным. В этом случае получается значение σ_x , которое будет меньше первоначального, и может случиться так, что с новым значением σ_x некоторые другие результаты измерений не будут удовлетворять критерию Шовене. Однако большинство авторитетных специалистов считает, что критерий Шовене не должен применяться второй раз с использованием пересчитанных значений $\bar{x}$ и σ_x .

Многие ученые полагают, что отбрасывание данных *никогда* не может быть оправдано, пока не найдется *внешнее* свидетельство того, что подозреваемые данные неверны. Может быть, более умеренная позиция состоит в том, что критерий Шовене следует использовать для обнаружения данных, которые могли бы по крайней мере *рассматриваться* как кандидаты на отбрасывание. Добросовестный студент может сделать вычисления дважды: первый раз с учетом данных, которые находятся под вопросом, и второй раз без них, чтобы посмотреть, насколько в действительности подозреваемое значение влияет на окончательное заключение.

6.3. Пример

Студент делает десять измерений одной длины x и получает результаты (все в миллиметрах)

46, 48, 44, 38, 45, 47, 58, 44, 45, 43.

Заметив, что значение 58 кажется аномально большим, он проверяет свои записи, но не находит указаний на то, что этот результат получился по ошибке. Тогда он применяет критерий Шовене. Какой вывод он сделает?

Учитывая временно все десять измерений, он рассчитывает

$$\bar{x} = 45,8 \quad \text{и} \quad \sigma_x = 5,1.$$

Разность между подозрительным значением $x_{\text{под}} = 58$ и средним $\bar{x} = 45,8$ равна 12,2, или 2,4 стандартных отклонений, т. е.

$$\frac{x_{\text{под}} - \bar{x}}{\sigma_x} = \frac{58 - 45,8}{5,1} = 2,4.$$

Из таблицы приложения А он находит: вероятность того, что результат будет отличаться от $\bar{x}$ на $2,4\sigma_x$ или более, равна

$$\begin{aligned} P(\text{вне } 2,4\sigma) &= 1 - P(\text{в пределах } 2,4\sigma) \\ &= 1 - 0,984 \\ &= 0,016. \end{aligned}$$

Для десяти измерений он мог бы, следовательно, получить только 0,16 случаев такого плохого измерения, как подозрительный результат. Так как это число меньше $1/2$, числа, устанавливаемого критерием Шовене, то студент должен по крайней мере рассмотреть возможность отбрасывания этого результата.

Следующий подозрительный результат — это 38, который на 1,5 стандартных отклонений отличается от среднего $\bar{x} = 45,8$. Аналогичные вычисления показывают, что среди десяти результатов он мог бы ожидать 1,3 случаев таких же плохих результатов, как и этот, так что этот результат вполне приемлем. Если же он решит выбросить подозрительный результат 58, то он должен пересчитать $\bar{x}$ и σ_x и получить

$$\bar{x} = 44,4 \quad \text{и} \quad \sigma_x = 2,9.$$

Как можно было ожидать, среднее изменилось немного, а стандартное отклонение заметно уменьшилось.

Задачи

Напоминание: звездочка у номера задачи означает, что задача решается или ее ответ приводится в разделе «Ответы» в конце книги.

6.1 (разд. 6.2). Усердная студентка делает 50 измерений количества теплоты Q , выделяющейся в определенном процессе. Она получает сред-

нее значение и стандартное отклонение, равные соответственно $\bar{Q} = 4,8$ и $\sigma_Q = 0,4$, где оба результата выражены в калориях.

а. Предполагая, что результаты ее измерений подчиняются нормальному распределению, найдите вероятность того, что любое единичное измерение приведет к результату, отличающемуся от $\bar{Q}$ на 0,8 кал или более. Сколько результатов, отличающихся от $\bar{Q}$ на 0,8 кал, ей следует ожидать? Если бы один из результатов ее измерений был равен 4,0 кал и она решила бы использовать критерий Шовене, то отбросила ли бы она этот результат?

б. Если бы один из ее результатов составлял 6,0 кал, то отбросила ли бы она его?

***6.2** (разд. 6.2). Студент измеряет некоторую разность потенциалов V десять раз и получает результаты (в вольтах)

0,86; 0,83; 0,87; 0,84; 0,82; 0,95; 0,83; 0,85; 0,89; 0,88.

а. Вычислите среднее $\bar{V}$ и стандартное отклонение σ_V этих результатов.

б. Если он решит использовать критерий Шовене, то должен ли он отбросить отсчет 0,95 В? Приведите подробно вашу аргументацию.

***6.3** (разд. 6.2). Студент делает 14 измерений периода колебаний генератора и получает результаты (в десятых долях секунды)

7, 3, 9, 3, 6, 9, 8, 7, 8, 12, 5, 9, 9, 3.

Чувствуя, что результат 12 подозрительно велик, он решает использовать критерий Шовене. Отвергнет ли он подозрительный результат? Сколько результатов, так же отличающихся от среднего, как 12, ему следует ожидать?

6.4 (разд. 6.2). Критерий Шовене определяет границу, за пределами которой результат измерения рассматривается как отвергаемый. Если мы делаем десять измерений и один из результатов отличается от среднего более чем на два стандартных отклонения (в любую сторону), то этот результат рассматривается как отвергаемый; в случае 20 измерений соответствующая граница равна 2,2 стандартных отклонений. Составьте таблицу, показывающую «границу отбрасывания» для случаев 5, 10, 15, 20, 50, 100, 200 и 1000 измерений. (Используйте таблицу функции ошибок из приложения А.)

Глава 7

Взвешенные средние

В этой главе рассматривается проблема объединения результатов двух или более отдельных и независимых измерений одной и той же физической величины. Мы покажем, что наилучшей оценкой этой величины, основанной на нескольких измерениях, будет соответствующее *взвешенное среднее* этих измерений.

7.1. Проблема объединения результатов разных измерений

Часто бывает так, что одна физическая величина измеряется несколько раз, возможно даже в нескольких независимых лабораториях, и тогда возникает вопрос, как объединить эти результаты, чтобы получить единственную наилучшую оценку. Предположим, например, что два студента А и Б тщательно измеряют величину x и получают следующие результаты:

$$\text{студент А: } x = x_A \pm \sigma_A \quad (7.1)$$

и

$$\text{студент Б: } x = x_B \pm \sigma_B. \quad (7.2)$$

Каждый из этих результатов, вероятно, и сам по себе является следствием нескольких измерений; например, x_A может быть средним всех измерений студента А и σ_A — стандартным отклонением этого среднего (и аналогично для x_B и σ_B). Вопрос теперь состоит в том, как лучше всего объединить x_A и x_B , чтобы получить единственную наилучшую оценку для x .

Прежде чем ответить на этот вопрос, заметим, что если различие $|x_A - x_B|$ между двумя измерениями намного больше обеих погрешностей σ_A и σ_B , то, по-видимому, что-то не в порядке по крайней мере в случае одного из измерений. В этой ситуации мы сказали бы, что два измерения *противо-*

речивы, и необходимо тщательно проанализировать оба измерения, чтобы проверить, не подвержено ли одно из них (или оба) незамеченным систематическим ошибкам.

Предположим, однако, что два измерения (7.1) и (7.2) *непротиворечивы*, т. е. различие $|x_A - x_B|$ не намного больше, чем любая из погрешностей σ_A и σ_B . В этом случае имеет смысл спросить, какова наилучшая оценка $x_{\text{наил}}$ истинного значения X , основанная на этих двух измерениях. Первой реакцией могло быть вычисление среднего значения $(x_A + x_B)/2$ двух измерений. Однако уже небольшое размышление заставляет предположить, что этот путь не подходит, если две погрешности σ_A и σ_B не равны. Вычисление простого среднего $(x_A + x_B)/2$ делает одинаково важными оба измерения, в то время как более точному отсчету следует приписать больший вес.

7.2. Взвешенное среднее

Мы легко можем решить нашу задачу, используя принцип максимального правдоподобия почти так же, как мы это делали в разд. 5.5. Если предположить, что результаты обоих измерений подчиняются распределению Гаусса, и обозначить неизвестное истинное значение величины x через X , то вероятность того, что студент А получит свое частное значение x_A , равна

$$P_X(x_A) \sim \frac{1}{\sigma_A} e^{-(x_A - X)^2 / 2\sigma_A^2} \quad (7.3)$$

и вероятность того, что студент Б получит значение x_B , равна

$$P_X(x_B) \sim \frac{1}{\sigma_B} e^{-(x_B - X)^2 / 2\sigma_B^2}. \quad (7.4)$$

Введя индекс X , мы указали явным образом, что эти вероятности зависят от неизвестного истинного значения. (Они также зависят от соответствующих ширин σ_A и σ_B , но мы этого не указали.)

Вероятность того, что студент А получит значение x_A и студент Б — значение x_B , равна произведению двух вероятностей (7.3) и (7.4). Теперь должно быть уже привычным, что это произведение будет экспоненциальной функцией с показателем, равным сумме двух показателей в (7.3) и (7.4). Запишем это как

$$\begin{aligned} P_X(x_A, x_B) &= P_X(x_A) P_X(x_B) \sim \\ &\sim \frac{1}{\sigma_A \sigma_B} e^{-\chi^2/2}, \end{aligned} \quad (7.5)$$

где мы ввели удобное краткое обозначение χ^2 для показателя

$$\chi^2 = \left(\frac{x_A - X}{\sigma_A} \right)^2 + \left(\frac{x_B - X}{\sigma_B} \right)^2. \quad (7.6)$$

Эта важная величина представляет собой сумму квадратов отклонений от X результатов двух измерений, деленных на соответствующие погрешности. Ее иногда называют «суммой квадратов».

Принцип максимального правдоподобия утверждает, как уже было отмечено, что наилучшей оценкой для неизвестного истинного значения X будет такое значение, для которого фактически полученные величины x_A и x_B наиболее вероятны. Иными словами, наилучшей оценкой для X будет значение, при котором вероятность (7.5) достигает максимума, или, что эквивалентно, показатель χ^2 минимален. (Поскольку максимизация вероятности влечет за собой минимизацию «суммы квадратов» χ^2 , то этот метод оценки X иногда называют «методом наименьших квадратов».) Таким образом, чтобы определить наилучшую оценку, мы просто продифференцируем (7.6) по X и приравняем производную нулю:

$$2 \frac{x_A - X}{\sigma_A^2} + 2 \frac{x_B - X}{\sigma_B^2} = 0.$$

Решение этого уравнения относительно X есть наилучшая оценка $x_{\text{наил}}$, и, как легко видеть, она равна

$$x_{\text{наил}} = \left(\frac{x_A}{\sigma_A^2} + \frac{x_B}{\sigma_B^2} \right) / \left(\frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2} \right). \quad (7.7)$$

Этот довольно громоздкий результат можно записать компактнее, если определить *веса*

$$w_A = \frac{1}{\sigma_A^2} \quad \text{и} \quad w_B = \frac{1}{\sigma_B^2}. \quad (7.8)$$

Подставляя в (7.7), получим

$$x_{\text{наил}} = \frac{w_A x_A + w_B x_B}{w_A + w_B}. \quad (7.9)$$

Если два исходных измерения одинаково точны ($\sigma_A = \sigma_B$ и, следовательно, $w_A = w_B$), то наш ответ сводится к простому среднему значению $(x_A + x_B)/2$. В общем случае выражение (7.9) дает *взвешенное среднее*; оно аналогично формуле для центра тяжести двух тел, когда w_A и w_B — действительные веса двух тел, а x_A и x_B — их координаты. В данном случае «веса» — обратные значения квадратов погрешностей в исход-

ных измерениях, как видно из (7.8). Если измерения студента А более точны, чем студента Б, то $\sigma_A < \sigma_B$ и, следовательно, $w_A > w_B$; поэтому наилучшая оценка $x_{\text{наил}}$ будет ближе к x_A , чем к x_B , как и должно быть.

Наш анализ можно обобщить на случай, когда объединяются несколько измерений одной и той же величины. Предположим, что у нас есть N отдельных измерений величины x

$$x_1 \pm \sigma_1, \quad x_2 \pm \sigma_2, \quad \dots, \quad x_N \pm \sigma_N$$

с соответствующими погрешностями $\sigma_1, \dots, \sigma_N$. Рассуждая, как и выше, мы получим, что наилучшая оценка, основанная на этих измерениях, равна взвешенному среднему

$$x_{\text{наил}} = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i}, \quad (7.10)$$

где веса w_i — это обратные значения квадратов соответствующих погрешностей

$$w_i = 1/\sigma_i^2 \quad (7.11)$$

для $i = 1, 2, \dots, N$.

Поскольку вес $w_i = 1/\sigma_i^2$, связанный с каждым измерением, содержит квадрат соответствующей погрешности σ_i , то любое измерение, существенно менее точное, чем остальные, внесет много меньший вклад в конечный результат (7.10). Например, если одно измерение в четыре раза менее точно, чем остальные, то его вес в 16 раз меньше, чем другие веса, и во многих случаях это измерение можно просто игнорировать.

Поскольку конечный результат (7.10) для $x_{\text{наил}}$ — это простая функция исходных значений $x_1, \dots, x_N$, то довольно просто вычислить погрешность нашего результата методом расчета ошибок в косвенных измерениях. В задаче 7.5 вам предлагается проверить, что погрешность результата (7.10) для $x_{\text{наил}}$ равна

$$\sigma_{x_{\text{наил}}} = \left(\sum_{i=1}^N w_i \right)^{-1/2}, \quad (7.12)$$

где, как обычно, $w_i = 1/\sigma_i^2$.

7.3. Пример

Три студента измеряют сопротивление несколько раз и получают три следующих ответа (в омах):

(значение первого студента для R) $= 11 \pm 1$;

(значение второго студента для R) $= 12 \pm 1$;

(значение третьего студента для R) $= 10 \pm 3$.

Если даны эти три результата, то какова наилучшая оценка для сопротивления R ?

Три погрешности σ_1 , σ_2 , σ_3 равны соответственно 1, 1 и 3. Следовательно, соответствующие три веса $w_i = 1/\sigma_i^2$ равны

$$w_1 = 1, \quad w_2 = 1, \quad w_3 = \frac{1}{9}.$$

Таким образом, в соответствии с (7.10) наилучшая оценка есть

$$R_{\text{наил}} = \frac{\sum w_i R_i}{\sum w_i} = \frac{(1 \cdot 11) + (1 \cdot 12) + (1/9 \cdot 10)}{1 + 1 + 1/9} = 11,42 \text{ Ом.}$$

Погрешность этого результата определяется (7.12) как

$$\sigma_{R_{\text{наил}}} = \left(\sum w_i \right)^{-1/2} = \left(1 + 1 + \frac{1}{9} \right)^{-1/2} = 0,69.$$

Таким образом, наш конечный вывод имеет вид

$$R = 11,4 \pm 0,7 \text{ Ом.}$$

Интересно проследить, какой ответ мы получили бы, если бы полностью игнорировали результаты измерений третьего студента, которые в три раза менее точны и, следовательно, в девять раз менее важны. В этом случае простое вычисление дает $R_{\text{наил}} = 11,50$ (по сравнению с 11,42) с погрешностью 0,71 (по сравнению с 0,69). Очевидно, третье измерение не имеет большого значения.

Задачи

Напоминание: звездочка у номера задачи означает, что задача решается или ее ответ приводится в разделе «Ответы» в конце книги.

*7.1 (разд. 7.2).

а. Два измерения скорости звука u дают результаты 334 ± 1 и 336 ± 2 (оба в м/с). Считаете ли вы эти измерения непротиворечивыми?

Если да, то вычислите наилучшую оценку для u и ее погрешность.

в. Повторите задание «а» для результатов 334 ± 1 и 336 ± 5 . Стоит ли использовать в расчетах второй результат?

*7.2 (разд. 7.2). Два студента измеряют сопротивление различными методами. Каждый делает по десять измерений и вычисляет среднее и его стандартное отклонение и получает

студент А: $R = 72 \pm 8 \text{ Ом,}$

студент Б: $R = 78 \pm 5 \text{ Ом.}$

а. Рассматривая оба измерения, найдите, чему равны наилучшая оценка R и ее погрешность.

б. Оцените приблизительно, сколько измерений должен сделать студент А (используя свой метод измерений), чтобы вес его результата был бы такой же, как и у студента Б.

7.3 (разд. 7.2). Найдите наилучшую оценку и ее погрешность, основанные на следующих четырех измерениях одной и той же величины:

$$1,4 \pm 0,5; \quad 1,2 \pm 0,2; \quad 1,0 \pm 0,25; \quad 1,3 \pm 0,2.$$

7.4 (разд. 7.2). Предположим, что все N измерений одной и той же величины x имеют одинаковую погрешность. Покажите четко, что в этом случае взвешенное среднее (7.10) сводится к обычному среднему значению, или среднему, $\bar{x} = (\sum x_i)/N$, и что выражение (7.12) для погрешности сводится к знакомому стандартному отклонению среднего.

*7.5 (разд. 7.2). Если даны результаты N измерений $x_1, \dots, x_N$ одной и той же величины x с погрешностями $\sigma_1, \dots, \sigma_N$, то наилучшая оценка для x определяется выражением (7.10), как $x_{\text{наил}} = (\sum w_i x_i) / (\sum w_i)$, где веса $w_i = 1/\sigma_i^2$. Это выражение определяет $x_{\text{наил}}$ как функцию $x_1, \dots, x_N$. Используя формулу (3.47) для расчета ошибок в косвенных измерениях, покажите, что погрешность в $x_{\text{наил}}$ дается (7.12) как

$$\sigma_{x_{\text{наил}}} = (\sum w_i)^{-1/2}.$$

Глава 8

Аппроксимация методом наименьших квадратов

Наше обсуждение статистической обработки данных до сих пор было сосредоточено исключительно на многократных измерениях одной и той же величины не потому, что такой анализ многократных измерений одной величины является наиболее важной задачей статистики, а потому, что эта простая задача должна быть хорошо понята, прежде чем мы сможем перейти к более общим проблемам. Теперь мы наконец обсудим нашу первую, очень важную более общую проблему.

8.1. Данные, которые должны ложиться на прямую линию

Один из наиболее общих и интересных типов экспериментов состоит в измерении нескольких значений двух различных физических переменных для исследования математической связи этих двух переменных¹⁾. Например, можно бросать камень с разных высот $h_1, \dots, h_N$ и измерять соответствующие времена падения $t_1, \dots, t_N$, чтобы проверить, связаны ли эти значения высот и времен ожидаемым соотношением $h = \frac{1}{2}gt^2$.

Вероятно, наиболее важными экспериментами такого типа являются те, где ожидаемая связь *линейна*, как это реализуется в случае, который мы рассмотрим первым. Например, если мы допускаем, что тело падает с постоянным ускорением свободного падения g , то его скорость v должна быть линейной функцией времени t :

$$v = v_0 + gt.$$

В общем случае мы будем рассматривать любые две физические переменные x и y , которые, как мы считаем, связаны

¹⁾ В литературе на русском языке такие измерения двух или более разных величин принято называть *совместными*. — Прим. перев.

линейной зависимостью вида

$$y = A + Bx, \quad (8.1)$$

где A и B — постоянные. К сожалению, для линейной зависимости используется много разных обозначений; остерегайтесь спутать (8.1) со столь же часто встречающейся записью $y = ax + b$.

Если две переменные y и x связаны линейной зависимостью вида (8.1), то график y от x должен быть прямой линией с наклоном B , которая пересекает ось y в точке $y = A$. Если бы мы измерили N различных значений $x_1, \dots, x_N$ и соответствующих значений $y_1, \dots, y_N$ и если бы результаты наших измерений не содержали погрешностей, то каждая точка (x_i, y_i) легла бы точно на линию $y = A + Bx$, как показано на рис. 8.1, а. На практике же всегда имеются погрешности, и большее, что мы можем ожидать, — это то, что расстояние каждой точки (x_i, y_i) от линии должно быть сравнимо в разумных пределах с погрешностями, как показано на рис. 8.1, б.

При выполнении ряда измерений описанного типа возникают два вопроса. Во-первых, если мы примем как факт, что y и x действительно *связаны* линейно, то придем к интересной задаче определения прямой линии $y = A + Bx$, которая наилучшим образом аппроксимирует результаты измерений, т. е. к задаче определения наилучших оценок постоянных A и B , основанных на данных $(x_1, y_1), \dots, (x_N, y_N)$. Эта задача может быть решена графически, как кратко рассматривалось в разд. 2.6. Ее можно решать также аналитически при помощи метода максимального правдоподобия. Этот аналитический

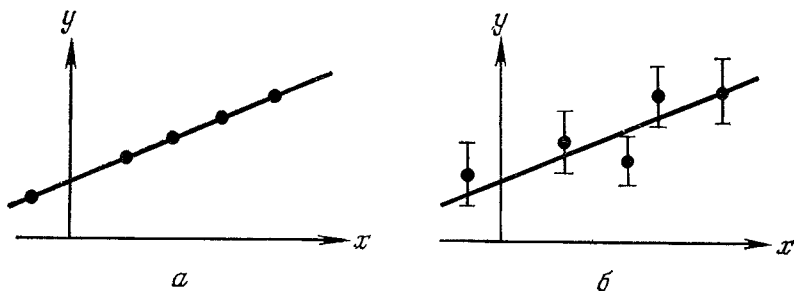


Рис. 8.1. а — если две переменные y и x связаны линейной зависимостью, как в формуле (8.1), и если нет экспериментальных погрешностей, то измеренные точки (x_i, y_i) точно лягут на линию $y = A + Bx$. б — на практике всегда имеются погрешности, которые можно изобразить черточками ошибок, и точки (x_i, y_i) в этом случае, как следует ожидать, будут располагаться на разумно близких расстояниях от линии. В данном случае показано, что только значения y подвержены влиянию заметных погрешностей.

метод определения наилучшей прямой линии, которая аппроксимирует серию экспериментальных точек, называется *линейной регрессией* или *аппроксимацией прямой методом наименьших квадратов* и является главной темой этой главы.

Второй вопрос, который можно задать, — это действительно ли измеренные значения $(x_1, y_1), \dots, (x_N, y_N)$ оправдывают наши ожидания, что функция y линейна по x . Сначала мы можем определить линию, которая наилучшим образом аппроксимирует данные, но затем мы должны предложить некоторую меру, которая показала бы, *насколько хорошо* эта линия аппроксимирует данные. Этот вопрос мы рассмотрим в гл. 9.

8.2. Расчет постоянных A и B

Вернемся теперь к задаче определения наилучшей прямой линии $y = A + Bx$, аппроксимирующей набор экспериментальных точек $(x_1, y_1), \dots, (x_N, y_N)$. Для упрощения будем предполагать, что, хотя результаты наших измерений x и y содержат некоторые погрешности, погрешность в измерениях x пренебрежимо мала. Это — разумное допущение, так как погрешности в одной переменной часто намного больше, чем погрешности в другой, которые мы, следовательно, можем без всякого опасения игнорировать. Далее мы будем предполагать, что все погрешности в y одинаковы по величине. (Это также разумное допущение для многих экспериментов, но если погрешности различны, то наше рассмотрение может быть обобщено с помощью соответствующего учета весов этих измерений; см. задачу 8.4). Если говорить точнее, мы предположим, что результат измерения каждого y_i подчиняется распределению Гаусса с одинаковой шириной σ_y во всех измерениях.

Зная постоянные A и B , мы могли бы для любого данного значения x_i (которое, по нашим предположениям, не содержит погрешности) вычислить истинное значение соответствующей величины y_i :

$$(\text{истинное значение } y_i) = A + Bx_i. \quad (8.2)$$

Результат измерения y_i подчиняется нормальному распределению с центром на истинном значении и с шириной σ_y . Следовательно, вероятность получения значения y_i равна

$$P_{A,B}(y_i) \sim \frac{1}{\sigma_y} e^{-(y_i - A - Bx_i)^2 / 2\sigma_y^2}, \quad (8.3)$$

где нижние индексы A и B указывают, что эта вероятность зависит от (неизвестных) значений A и B . Вероятность полу-

чения всего набора результатов измерений $y_1, \dots, y_N$ равна произведению

$$P_{A,B}(y_1, \dots, y_N) = P_{A,B}(y_1) \cdot \dots \cdot P_{A,B}(y_N) \sim \frac{1}{\sigma_y^N} e^{-\chi^2/2}, \quad (8.4)$$

где показатель дается формулой

$$\chi^2 = \sum_{i=1}^N \frac{(y_i - A - Bx_i)^2}{\sigma_y^2}. \quad (8.5)$$

Теперь уже нам известно, что наилучшие оценки для неизвестных постоянных A и B , основанные на данных измерениях, — это такие значения A и B , для которых вероятность $P_{A,B}(y_1, \dots, y_N)$ максимальна или для которых сумма квадратов χ^2 (8.5) минимальна (вот почему этот метод известен как аппроксимация методом наименьших квадратов). Чтобы найти эти значения, продифференцируем χ^2 по A и B и приравняем эти производные нулю:

$$\frac{\partial \chi^2}{\partial A} = (-2/\sigma_y^2) \sum_{i=1}^N (y_i - A - Bx_i) = 0, \quad (8.6)$$

и

$$\frac{\partial \chi^2}{\partial B} = (-2\sigma_y^2) \sum_{i=1}^N x_i (y_i - A - Bx_i) = 0. \quad (8.7)$$

Эти два выражения можно переписать как систему уравнений для A и B :

$$AN + B \sum x_i = \sum y_i \quad (8.8)$$

и

$$A \sum x_i + B \sum x_i^2 = \sum x_i y_i. \quad (8.9)$$

(В дальнейшем мы опускаем границы суммирования от $i=1$ до N у знака суммирования $\sum$.) Эти два уравнения, известные как *нормальные уравнения*, легко решаются и дают *оценки метода наименьших квадратов для постоянных A и B* :

$$A = \frac{(\sum x_i^2)(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{\Delta} \quad (8.10)$$

и

$$B = \frac{N(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\Delta}, \quad (8.11)$$

где мы использовали принятое обозначение

$$\Delta = N(\sum x_i^2) - (\sum x_i)^2. \quad (8.12)$$

Формулы (8.10) и (8.11) дают наилучшие оценки постоянных A и B для прямой линии $y = A + Bx$, основанные на измеренных точках $(x_1, y_1), \dots, (x_N, y_N)$. Получившаяся линия называется *линией аппроксимации методом наименьших квадратов* этих данных, или *линией регрессии* y от x . Теперь естественно спросить, каковы погрешности в наших оценках A и B . Оказывается, что, прежде чем мы сможем ответить на этот вопрос, необходимо рассмотреть погрешности σ_y в наших исходных измерениях $y_1, \dots, y_N$, что мы и сделаем.

8.3. Погрешность в измерениях y

В процессе измерения значений $y_1, \dots, y_N$ у вас, вероятно, уже сложилось некоторое представление об их погрешности. Тем не менее важно знать, как вычислить погрешность только из анализа самих данных. Нужно помнить, что числа $y_1, \dots, y_N$ не представляют собой N результатов измерений одной и той же величины. (Они могли бы быть, например, временами, за которые камень падает с N различных высот.) Таким образом, мы не получим никакого представления о надежности этих результатов, исследуя разброс в их значениях.

Тем не менее можно легко оценить погрешность σ_y в числах $y_1, \dots, y_N$. Результат измерения каждого y_i (как мы предполагаем) распределен нормально около истинного значения $A + Bx_i$ с параметром ширины σ_y . Таким образом, отклонения $y_i - A - Bx_i$ распределены нормально, причем все с одним и тем же центральным значением 0 и одной и той же шириной σ_y . Это сразу же ведет к предположению, что хорошая оценка σ_y могла бы определяться суммой квадратов в уже известном нам виде

$$\sigma_y^2 = \frac{1}{N} \sum (y_i - A - Bx_i)^2. \quad (8.13)$$

Действительно, этот результат может быть подтвержден при помощи принципа максимального правдоподобия. Как обычно, наилучшая оценка для представляющего интерес параметра (в данном случае σ_y) есть значение, для которого вероятность (8.4) получения значений $y_1, \dots, y_N$ максимальна. Как легко проверить, дифференцируя (8.4) по σ_y и приравнивая производную нулю, этой наилучшей оценкой будет точно выражение (8.13).

К сожалению, как вы могли бы уже догадаться, оценка (8.13) для σ_y^2 вовсе не окончательная. Числа A и B в (8.13) — неизвестные истинные значения постоянных A и B . На практике они должны быть заменены нашими *наилучшими оценками* для A и B , а именно выражениями (8.10) и (8.11), и такая замена немного изменяет значение (8.13). Можно показать, что это изменение компенсируется заменой фактора N в знаменателе на $(N-2)$. Таким образом, наш конечный ответ для погрешности в измерениях $y_1, \dots, y_N$ есть

$$\sigma_y^2 = \frac{1}{N-2} \sum_{i=1}^N (y_i - A - Bx_i)^2, \quad (8.14)$$

где A и B даются выражениями (8.10) и (8.11). Если бы у нас была независимая оценка погрешности в $y_1, \dots, y_N$, то мы могли бы сравнить ее с величиной σ_y , рассчитанной из (8.14).

Мы не пытались обосновать фактор $(N-2)$ в (8.14), но можем сделать некоторые комментарии. Во-первых, пока N умеренно велико, различие между N и $(N-2)$ практически неважно. Во-вторых, то, что фактор $(N-2)$ *отражает реальную ситуацию*, становится очевидным, если рассмотреть случай измерения только двух пар значений: (x_1, y_1) и (x_2, y_2) . Если имеются только две точки, мы всегда можем провести прямую, проходящую *точно* через эти точки, и аппроксимация методом наименьших квадратов дает эту линию. Таким образом, имея только две пары данных, мы не можем, вероятно, сделать никакого заключения о надежности наших измерений. Далее, поскольку обе точки лежат точно на наилучшей прямой, то оба слагаемых в суммах (8.13) и (8.14) равны нулю. Таким образом, формула (8.13) (с $N=2$ в знаменателе) дает абсурдный результат $\sigma_y = 0$, в то время как (8.14) с $N-2=0$ в знаменателе дает $\sigma_y = 0/0$, конкретно указывая, что после проведения всего лишь двух измерений величина σ_y не определена.

Наличие фактора $(N-2)$ в (8.14) напоминает множитель $(N-1)$ в знаменателе выражения (5.46) для оценки стандартного отклонения N измерений одной и той же величины x . Там было выполнено N измерений $x_1, \dots, x_N$ одной величины x . Прежде чем рассчитать σ_x , мы должны были использовать данные для определения $\bar{x}$. Это оставляло только $(N-1)$ независимых измеренных значений, поэтому мы говорим, что, вычислив $\bar{x}$, мы оставили только $(N-1)$ *степеней свободы*. Сейчас мы тоже сделали N измерений, но перед расчетом σ_y мы должны были вычислить *две* величины: A и B . Сделав это, мы оставили только $(N-2)$ степеней свободы. В общем случае мы определим *число степеней свободы* на любом этапе

статистических вычислений как число независимых измерений *минус* число параметров, рассчитанных из этих измерений. Можно показать (хотя мы не будем здесь этого делать), что именно число степеней свободы, а *не* число измерений должно появляться в знаменателе формул, подобных (8.14) и (5.46). Так объясняется, почему выражение (8.14) содержит фактор $(N-2)$, а (5.46)— фактор $(N-1)$.

8.4. Погрешность в постоянных A и B

Найдя погрешность σ_y в полученных числах $y_1, \dots, y_N$, мы легко можем вернуться к нашим оценкам постоянных A и B и рассчитать их погрешности. Суть здесь в том, что оценки (8.10) и (8.11) для A и B — точно определенные функции измеренных значений $y_1, \dots, y_N$. Следовательно, погрешности в A и B определяют простым расчетом ошибок в косвенных измерениях, исходя из погрешностей в $y_1, \dots, y_N$. Мы предоставляем читателю проверить (задача 8.8), что

$$\sigma_A^2 = \sigma_y^2 \sum x_i^2 / \Delta \quad (8.15)$$

и

$$\sigma_B^2 = N \sigma_y^2 / \Delta, \quad (8.16)$$

где Δ определяется (8.12), как обычно.

8.5. Пример

Если объем некоторого количества идеального газа поддерживать постоянным, то его температура T будет линейной функцией давления в газе P

$$T = A + BP. \quad (8.17)$$

Здесь постоянная A — температура, при которой давление P падает до нуля (если газ не сконденсируется сначала в жидкость); она называется *абсолютным нулем температуры* и имеет принятое значение

$$A = -273,15 \text{ градусов Цельсия.} \quad (8.18)$$

Постоянная B зависит от природы газа, его массы и объема¹⁾. Измеряя ряд значений T и P , мы можем определить наилуч-

¹⁾ Разность $T - A$ называется *абсолютной температурой*. Перепишав (8.17) с учетом этого обозначения, мы получаем, что абсолютная температура пропорциональна давлению (при постоянном объеме).

Таблица 8.1. Измерение давления и температуры

Номер опыта i	Давление P_i , мм рт. ст.	Температура T_i , °C	$A + BP_i$
1	65	-20	-22,2
2	75	17	14,9
3	85	42	52,0
4	95	94	89,1
5	105	127	126,2

шие оценки для постоянных A и B . В частности, постоянная A дает абсолютный нуль температуры.

Одна система пяти измерений P и T , полученная студентом, приведена в трех первых столбцах табл. 8.1. Студент считает, что погрешность в измерениях P пренебрежимо мала, а погрешности в T все одинаковы и равны «нескольким градусам». Предполагая, что его точки должны ложиться на прямую линию типа (8.17), он вычисляет наилучшую оценку постоянной A (абсолютного нуля) и ее погрешность. Какими должны быть его выводы?

Все, что нам необходимо сейчас сделать, — это использовать формулы (8.10) и (8.15), заменяя x_i на P_i и y_i на T_i , чтобы рассчитать все величины, представляющие интерес. Для этого потребуется вычисление сумм $\sum P_i$, $\sum P_i^2$, $\sum T_i$, $\sum P_i T_i$. Многие карманные калькуляторы могут вычислять все эти суммы автоматически; но даже и без таких калькуляторов мы легко сможем выполнить все необходимые расчеты, если надлежащим образом представим данные. Согласно табл. 8.1, мы можем вычислить

$$\begin{aligned}\sum P_i &= 425, \\ \sum P_i^2 &= 37\,125, \\ \sum T_i &= 260, \\ \sum P_i T_i &= 25\,810, \\ \Delta &= 5\,000,\end{aligned}$$

где $\Delta = N(\sum P_i^2) - (\sum P_i)^2$. В расчетах такого рода важно сохранять много значащих цифр, поскольку нам придется вычислять разности этих больших чисел. Зная эти суммы, мы сразу же можем вычислить наилучшие оценки постоянных A и B :

$$A = \frac{(\sum P_i^2)(\sum T_i) - (\sum P_i)(\sum P_i T_i)}{\Delta} = -263,35$$

и

$$B = \frac{N(\sum P_i T_i) - (\sum P_i)(\sum T_i)}{\Delta} = 3,71.$$

Это даст наилучшую оценку студента для абсолютного нуля $A = -263^\circ\text{C}$.

Зная постоянные A и B , мы можем рассчитать числа $A + BP_i$, т. е. температуры, «ожидаемые» на основе нашей наилучшей аппроксимации соотношения $T = A + BP$. Они показаны в правом столбце таблицы, и, как мы и надеялись, все они согласуются в разумных пределах с наблюдаемыми температурами. Теперь можно рассчитать разности между числами в двух последних столбцах и получить

$$\sigma_T^2 = \frac{1}{N-2} \sum (T_i - A - BP_i)^2 = 44,6,$$

и, следовательно, стандартное отклонение

$$\sigma_T = 6,7.$$

Это значение хорошо согласуется с оценкой студента, согласно которой его измерения температуры неопределенны с точностью «нескольких градусов».

Наконец, мы можем рассчитать погрешность в A , используя (8.15):

$$\sigma_A^2 = \sigma_T^2 (\sum P_i^2) / \Delta = 331,$$

или

$$\sigma_A = 18.$$

Итак, соответствующим образом округленный итоговый вывод студента должен иметь вид

$$\text{абсолютный нуль, } A = -260 \pm 20^\circ\text{C},$$

что удовлетворительно согласуется с принятым значением -273°C .

Как часто бывает, эти результаты становятся намного яснее, если по ним построить график, как на рис. 8.2. Пять экспериментальных точек вместе с их погрешностями $\pm 7^\circ$ по шкале T показаны справа вверху. Наилучшая прямая линия проходит через четыре черточки ошибок и близко от пятой.

Чтобы найти значение абсолютного нуля, линию следует продолжить за пределы области, где лежат все экспериментальные точки, до ее пересечения с осью T . Этот процесс *экстраполяции* (продолжение кривой за пределы точек, по которым она определяется) может привести к большим погрешностям, как это ясно из рисунка. Очень небольшое изме-

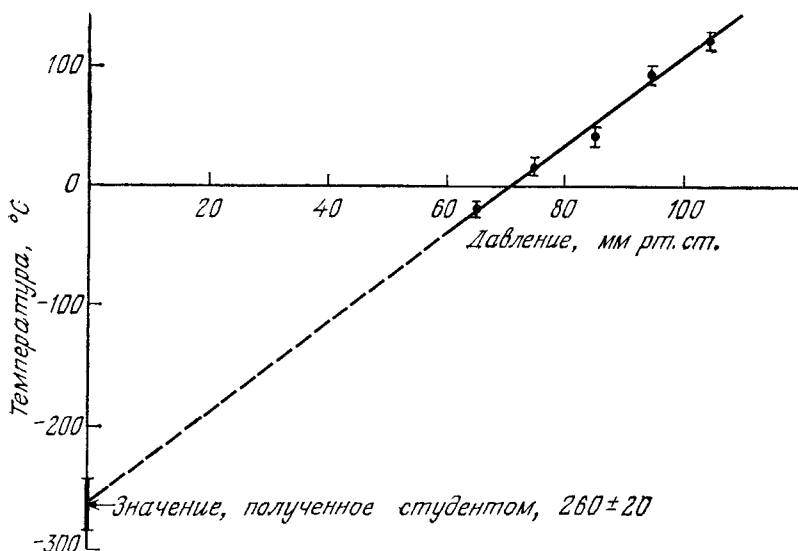


Рис. 8.2. График зависимости температуры T от давления P для газа при постоянном объеме.

Черточки ошибок имеют длину в одно стандартное отклонение σ_T по каждую сторону от каждой из пяти экспериментальных точек, а линия — это наилучшая аппроксимация, полученная методом наименьших квадратов. Абсолютный нуль температуры можно найти, если экстраполировать линию до ее пересечения с осью T .

Изменение в наклоне линии приведет к большим изменениям в положении точки пересечения этой линии с осью T , поскольку эта ось сильно удалена от экспериментальных точек. Таким образом, любая погрешность в данных значительно возрастает, если мы вынуждены экстраполировать на большие расстояния. Это объясняет, почему погрешность в значении абсолютного нуля ($\pm 18^\circ$) намного больше, чем в исходных измерениях температуры ($\pm 7^\circ$).

8.6. Аппроксимация другими кривыми методом наименьших квадратов

До сих пор в этой главе мы рассматривали случай двух переменных, удовлетворяющих линейной зависимости $y = A + Bx$, и обсудили, как вычислять постоянные A и B . Эта важная задача — только частный случай широкого класса задач по аппроксимации кривыми, многие из которых могут быть решены подобным образом. В этом последнем разделе мы кратко рассмотрим несколько таких задач.

Аппроксимация полиномом

Часто одна переменная y выражается через полином от второй переменной x :

$$y = A + Bx + Cx^2 + \dots + Hx^n. \quad (8.19)$$

Например, можно ожидать, что высота y , которую пролетает падающее тело, квадратично зависит от времени t :

$$y = y_0 + v_0 t - \frac{1}{2} g t^2,$$

где y_0 и v_0 — начальные высота и скорость, а g — ускорение свободного падения. Если дана совокупность наблюдений двух переменных, то можно определить наилучшие оценки постоянных $A, B, \dots, H$ в (8.19) с помощью выкладок, которые производятся, как в разд. 8.2, что мы теперь и сделаем.

Для упрощения предположим, что полином (8.19) в действительности только квадратичный:

$$y = A + Bx + Cx^2, \quad (8.20)$$

(Заинтересованный читатель легко распространит доказательство на общий случай.) Предположим, как и прежде, что у нас есть серия измерений (x_i, y_i) , $i = 1, \dots, N$, где у всех y_i одинаковые погрешности и все x_i точны. Для каждого x_i соответствующее истинное значение y_i дается (8.20), где параметры A, B и C пока неизвестны. Мы примем, что результаты измерений y_i подчиняются нормальным распределениям, каждое с центром на соответствующем истинном значении и все с одной и той же шириной σ_y . Это позволяет представить вероятность получения наблюдаемых значений $y_1, \dots, y_N$ в уже привычном виде:

$$P(y_1, \dots, y_N) \sim e^{-\chi^2/2}, \quad (8.21)$$

где теперь

$$\chi^2 = \sum_{i=1}^N \frac{(y_i - A - Bx_i - Cx_i^2)^2}{\sigma_y^2}. \quad (8.22)$$

(Это соответствует выражению (8.5) в линейном случае.) Наилучшие оценки A, B и C — такие, для которых вероятность $P(y_1, \dots, y_N)$ максимальна или величина χ^2 минимальна. Дифференцируя χ^2 по A, B и C и приравнявая эти производные нулю, мы получим три уравнения (которые вам следует проверить):

$$\begin{aligned} AN + B \sum x_i + C \sum x_i^2 &= \sum y_i, \\ A \sum x_i + B \sum x_i^2 + C \sum x_i^3 &= \sum x_i y_i, \\ A \sum x_i^2 + B \sum x_i^3 + C \sum x_i^4 &= \sum x_i^2 y_i. \end{aligned} \quad (8.23)$$

Для любого данного набора результатов измерений (x_i, y_i) эта система уравнений для A , B и C (известная как система *нормальных уравнений*) может быть решена и, таким образом, могут быть найдены наилучшие оценки для A , B и C . С найденными таким путем значениями A , B , C выражение $y = A + Bx + Cx^2$ называется полиномиальной аппроксимацией, полученной методом наименьших квадратов, или полиномиальной регрессией для данных результатов измерений.

Метод полиномиальной регрессии легко обобщается для полиномов любой степени, хотя получающиеся нормальные уравнения становятся очень громоздкими в случае полиномов высокой степени. В принципе аналогичный метод может быть применен к *любой* функции $y = f(x)$, которая зависит от различных неизвестных параметров A , B К сожалению, получающиеся нормальные уравнения, которые определяют наилучшие оценки для A , B , ..., трудно, а порой и невозможно решить. Однако имеется один класс задач, которые всегда *можно* решить, а именно задачи, в которых функция $y = f(x)$ линейно зависит от параметров A , B , ... Этот класс включает все полиномы [очевидно, что полином (8.19) линейно зависит от коэффициентов A , B ...], а также многие другие функции. Например, в случае некоторых задач y представляется суммой тригонометрических функций:

$$y = A \sin x + B \cos x. \quad (8.24)$$

Для этой функции и фактически для любой функции, которая линейна относительно параметров A , B , ..., нормальные уравнения, которые определяют наилучшие оценки для A , B ..., — это система линейных уравнений, которая всегда может быть решена (см. задачи 8.12 и 8.13).

Экспоненциальные функции

Одна из наиболее важных функций в физике — экспоненциальная функция

$$y = Ae^{Bx}, \quad (8.25)$$

где A и B — постоянные. Интенсивность I излучения после прохождения расстояния x через защиту спадает экспоненциально:

$$I = I_0 e^{-\mu x},$$

где I_0 — начальная интенсивность и μ характеризует поглощение в защите. Заряд на замкнутом через сопротивление конденсаторе спадает экспоненциально:

$$Q = Q_0 e^{-\lambda t},$$

где Q_0 — начальный заряд и $\lambda = 1/(RC)$, а R и C — сопротивление и емкость.

Если постоянные A и B в (8.25) неизвестны, то естественно искать их оценки, основываясь на измерениях x и y . К сожалению, прямое применение изложенных выше методов приводит к таким уравнениям для A и B , которые не имеют простого решения. Однако можно преобразовать нелинейную связь (8.25) между y и x в линейное соотношение, к которому мы уже можем применить наш способ аппроксимации методом наименьших квадратов.

Достигнуть желаемой «линеаризации» можно, если просто взять логарифмы от обеих частей (8.25):

$$\ln y = \ln A + Bx. \quad (8.26)$$

Мы видим, что, хотя y нелинейно зависит от x , $\ln y$ зависит линейно. Это преобразование нелинейного соотношения (8.25) в линейное (8.26) полезно во многих случаях помимо аппроксимации методом наименьших квадратов. Если бы мы захотели проверить соотношение (8.25) графически, то не смогли бы этого сделать, так как на обычном графике y от x получается кривая, которую трудно идентифицировать визуально. С другой стороны, график зависимости $\ln y$ от x (или $\log y$ от x) — это прямая линия, которую легко можно идентифицировать. (Такой график особенно легко построить, если использовать «полулогарифмическую» миллиметровую бумагу, у которой деления на одной из осей расположены в соответствии со значениями логарифмов. Такая бумага позволяет строить графики $\log y$ непосредственно, т. е. даже не вычисляя значений логарифма переменной.)

Польза в линейном представлении (8.26) для аппроксимации методом наименьших квадратов несомненна. Если мы полагаем, что y и x должны быть связаны зависимостью $y = Ae^{Bx}$, то в новых переменных $z = \ln y$ и x должны быть связаны соотношением

$$z = \ln A + Bx. \quad (8.27)$$

Если имеется серия измерений (x_i, y_i) , то для каждого y_i мы можем вычислить $z_i = \ln y_i$. Тогда точки (x_i, z_i) должны лежать на линии (8.27). Эта линия может быть рассчитана методом наименьших квадратов, и таким образом будут получены наилучшие оценки для постоянных $\ln A$ (по которой мы можем найти A) и B .

Пример

Многие популяции (людей, бактерий, радиоактивных ядер и т. д.) изменяются со временем экспоненциально. Если ка-

Таблица 8.2. Популяции бактерий

Время t_i , дни	Численность популяции N_i	$z_i = \ln N_i$
0	153 000	11,94
1	137 000	11,83
2	128 000	11,76

кая-то популяция N уменьшается со временем экспоненциально, мы записываем

$$N = N_0 e^{-t/\tau}, \quad (8.28)$$

где τ называется *средним временем жизни* популяции (которое тесно связано со временем *полураспада* $t_{1/2}$; действительно, $t_{1/2} = 0,693\tau$). Некоторый биолог считает, что популяция бактерий уменьшается со временем экспоненциально согласно (8.28); он измеряет численность популяции на каждый из трех последовательных дней и получает результаты, представленные в табл. 8.2. Какую наилучшую оценку среднего времени жизни τ он получит на основании этих данных?

Если N изменяется согласно (8.28), то переменная $z = \ln N$ должна линейно зависеть от t :

$$z = \ln N = \ln N_0 - \frac{t}{\tau}. \quad (8.29)$$

Следовательно, наш биолог рассчитывает три числа $z_i = \ln N_i$ ($i = 0, 1, 2$), приведенные в третьем столбце табл. 8.2. Используя эти три числа, он рассчитывает прямую линию (8.29) методом наименьших квадратов и получает следующие наилучшие оценки для коэффициентов $\ln N_0$ и $(-1/\tau)$:

$$\ln N_0 = 11,93 \text{ и } (-1/\tau) = -0,089 \text{ день}^{-1}.$$

Из второго значения следует, что его наилучшая оценка для среднего времени жизни равна

$$\tau = 11,2 \text{ дней.}$$

Описанный выше метод привлекательно прост (особенно при наличии калькулятора, который автоматически выполняет операцию линейной регрессии) и часто используется. Тем не менее этот метод не вполне обоснован с логической точки зрения. Наш вывод аппроксимации прямой линией $y = A + Bx$ методом наименьших квадратов основывался на допущении, что все измеренные значения $y_1, \dots, y_N$ имеют одинаковые погрешности. В данном же случае мы реализуем эту аппроксимацию методом наименьших квадратов по отношению к переменной $z = \ln y$. Если все измеренные значения y_i имеют

одинаковую погрешность, то значения $z_i = \ln y_i$ — определенно не одинаковую. Действительно, из простого расчета ошибок в случае косвенных измерений мы знаем, что

$$\sigma_z = \left| \frac{dz}{dy} \right| \sigma_y = \frac{\sigma_y}{y}. \quad (8.30)$$

Таким образом, если σ_y одинаковы для всех измерений, то σ_z отличаются (σ_z больше для меньших y). Очевидно, что переменная $z = \ln y$ не удовлетворяет требованию, согласно которому погрешности должны быть одинаковы для всех результатов измерений, если переменная y удовлетворяет этому требованию.

Выход из этого положения очевиден. Необходимо обобщить метод наименьших квадратов на случай, когда погрешности в результатах измерений неодинаковы, при условии что все эти различные погрешности известны. (Этот метод *наименьших квадратов с весами* описан в задаче 8.4.) Если известно, что результаты измерений $y_1, \dots, y_N$ действительно имеют одинаковую погрешность, то формула (8.30) показывает, как изменяются погрешности в $z_1, \dots, z_N$, и поэтому мы можем применить метод наименьших квадратов с весами к уравнению $z = \ln A + Bx$.

На практике, однако, часто нельзя быть уверенным в том, что погрешности в $y_1, \dots, y_N$ действительно одинаковы, поэтому можно рассматривать предположение, согласно которому одинаковы погрешности в $z_1, \dots, z_N$, и использовать обычный метод наименьших квадратов без весов. Часто изменения в погрешностях бывают малы, и поэтому практически безразлично, какой метод использовать, как это и было в предыдущем примере. В любом случае непосредственное применение обычной (без весов) аппроксимации методом наименьших квадратов — это недвусмысленный и простой способ получения *разумных* (если не *наилучших*) оценок для постоянных A и B из соотношения $y = Ae^{Bx}$, поэтому он часто используется именно в этом виде.

Множественная регрессия

До сих пор мы рассматривали случай, когда наблюдаются только *две* переменные x и y , и обсуждали их связь. Во многих реальных задачах необходимо рассматривать большее число переменных. Например, измеряя давление газа P , можно обнаружить, что оно зависит от объема V и температуры T , и поэтому нужно исследовать P как функцию V и T . Простейшим примером может служить задача, когда одна переменная z зависит линейно от двух других x и y :

$$z = A + Bx + Cy. \quad (8.31)$$

Эта задача может быть рассмотрена методом, являющимся прямым обобщением случая двух переменных. Если у нас есть серия измерений (x_i, y_i, z_i) , $i = 1, \dots, N$ (где погрешности всех z_i одинаковы, а x_i и y_i точны), то мы можем использовать принцип максимального правдоподобия, как и в разд. 8.2, и получить, что наилучшие оценки постоянных A, B, C определяются нормальными уравнениями следующего вида:

$$\begin{aligned} AN + B \sum x_i + C \sum y_i &= \sum z_i, \\ A \sum x_i + B \sum x_i^2 + C \sum x_i y_i &= \sum x_i z_i, \\ A \sum y_i + B \sum x_i y_i + C \sum y_i^2 &= \sum y_i z_i. \end{aligned} \quad (8.32)$$

Эти уравнения можно решить относительно A, B и C и получить наилучшую аппроксимацию для соотношения (8.31). Этот случай называется *множественной регрессией* («множественной», так как имеется более двух переменных), но мы его сейчас рассматривать не будем.

Задачи

Напоминание: звездочка у номера задачи означает, что задача решается или ее ответ приводится в разделе «Ответы» в конце книги.

***8.1** (разд. 8.2). Используйте метод наименьших квадратов, чтобы определить линию $y = A + Bx$, которая наилучшим образом аппроксимирует четыре точки: (1; 12), (2; 13), (3; 18), (4; 19). Нарисуйте точки и линию.

8.2 (разд. 8.2). Чтобы определить коэффициент упругости k пружины, студент подвешивает к ней различные массы m и измеряет соответствующие длины l . Его результаты приведены в табл. 8.3. Так как сила mg равна $k(l - l_0)$, где l_0 — длина пружины в нерастянутом состоянии, то эти данные должны ложиться на линию $l = l_0 + (g/k)m$. Найдите аппроксимацию этих данных методом наименьших квадратов и получите наилучшие оценки для собственной длины пружины l_0 и коэффициента упругости пружины k .

***8.3** (разд. 8.3). Пусть переменные x и y связаны соотношением $y = Bx$, т. е. известно, что они лежат на прямой линии, проходящей через начало координат. Предположим, что вы сделали N измерений (x_i, y_i) , причем погрешности в x пренебрежимо малы, а в y — одинаковы. Используя метод доказательства, аналогичный приведенному в разд. 8.2, покажите, что, согласно методу наименьших квадратов, наилучшая оценка для B равна

$$B = \sum x_i y_i / \sum x_i^2.$$

***8.4** (разд. 8.2). Предположим, что мы измеряем N пар значений (x_i, y_i) двух переменных x и y , которые, как мы полагаем, связаны линейной зависимостью $y = A + Bx$. Предположим, что погрешность изме-

Таблица 8.3

Масса груза m , г	200	300	400	500	600	700	800	900
Длина l , см	5,1	5,5	5,9	6,8	7,4	7,5	8,6	9,4

Таблица 8.4

Расстояние, м	0	1000	2000	3000
Время, с	17,6	40,4	67,7	90,1

рений всех x_i пренебрежимо мала, а все y_i имеют разные погрешности σ_i (т. е. y_1 имеет погрешность σ_1 ; $y_2 - \sigma_2$ и т. д.). Просмотрите еще раз вывод аппроксимации методом наименьших квадратов в разд. 8.2 и обобщите его на случай, когда погрешности в y_i неодинаковы. Покажите, что наилучшие оценки для A и B равны

$$A = [(\sum w_i x_i^2)(\sum w_i y_i) - (\sum w_i x_i)(\sum w_i x_i y_i)] / \Delta \quad (8.33)$$

и

$$B = [(\sum w_i)(\sum w_i x_i y_i) - (\sum w_i x_i)(\sum w_i y_i)] / \Delta, \quad (8.34)$$

где веса $w_i = 1/\sigma_i^2$ и

$$\Delta = (\sum w_i)(\sum w_i x_i^2) - (\sum w_i x_i)^2. \quad (8.35)$$

Этот метод наименьших квадратов с весами можно использовать, когда погрешности σ_i (или по крайней мере их относительные величины) известны. Возможно, наиболее типичной ситуацией, когда все это реализуется, является пример эксперимента с подсчетом событий, подобным подсчету распадов радиоактивных ядер. Как обсуждалось в разд. 3.1 (и доказывалось в гл. 11), погрешность, соответствующая любому подсчету v , равна $\sqrt{v}$.

8.5 (разд. 8.2). Пусть известно, что y линейно зависит от x , так что $y = A + Bx$, и предположим, что у нас есть три измерения (x, y) : $(1; 2 \pm 0,5)$; $(2; 3 \pm 0,5)$; $(3; 2 \pm 1,5)$, где погрешности в x пренебрежимо малы. Используйте метод наименьших квадратов с весами, т. е. формулы (8.33) — (8.35), и найдите A и B . Сравните свои результаты со значениями, которые вы получили бы, если бы пренебрегли изменениями в погрешности, т. е. использовали бы аппроксимацию без весов в соответствии с (8.10) — (8.12). Представьте данные и обе линии на графике и попытайтесь понять разницу.

*8.6 (разд. 8.4). Поезд, движущийся, как полагают, с постоянной скоростью, проезжает четыре различных пункта, в каждом из которых делаются замеры времени; результаты приведены в табл. 8.4. Примените аппроксимацию методом наименьших квадратов в виде $d = d_0 + vt$ и найдите наилучшую оценку скорости поезда v . Какова погрешность в v ?

8.7 (разд. 8.4). Студент измеряет давление газа P для пяти различных значений температуры T , поддерживая его объем V постоянным. Его результаты приведены в табл. 8.5. Все данные должны удовлетворять линейной зависимости вида $T = A + BP$, где A — абсолютный нуль температуры (принятое значение которой равно -273°C , как уже обсуждалось в разд. 8.5). Найдите наилучшую аппроксимацию данных студента

Таблица 8.5

Давление P_i , мм рт. ст.	79	82	85	88	90
Температура T_i , $^\circ\text{C}$	8	17	30	37	52

и, следовательно, его наилучшую оценку для абсолютного нуля и ее погрешность.

*8.8 (разд. 8.4).

а. Используйте метод максимального правдоподобия в том виде, в каком он применялся при выводе (8.13), и покажите, что (8.13) определяет погрешность σ_y в y для серии измерений $(x_1, y_1), \dots, (x_N, y_N)$, которые должны ложиться на прямую линию.

б. Используйте расчет ошибок в случае косвенных измерений и покажите, что погрешности σ_A и σ_B в параметрах прямой линии $y = A + Bx$ определяются выражениями (8.15) и (8.16).

*8.9 (разд. 8.4). В аппроксимации методом наименьших квадратов, примененной к набору точек $(x_1, y_1), \dots, (x_N, y_N)$, переменные x и y рассматриваются несимметричным образом. A именно определяется наилучшая аппроксимация линией $y = A + Bx$ в предположении, что все погрешности в $y_1, \dots, y_N$ одинаковы, а погрешности в $x_1, \dots, x_N$ пренебрежимо малы. Если обратить ситуацию, то надо было бы поменять местами x и y и аппроксимировать линией $x = A' + B'y$. Две линии $y = A + Bx$ и $x = A' + B'y$ совпадали бы между собой, если бы все N точек *точно* ложились на линию; но в общем случае эти две линии будут слегка различаться. Определите, как данные задачи 8.1 аппроксимируются линией $x = A' + B'y$ (рассматривая все x_i как одинаково неточные, а y_i — как точные). Найдите A' и B' и их погрешности $\sigma_{A'}$ и $\sigma_{B'}$. Каковы будут значения A' и B' , полученные из ответов задачи 8.1? Сравните линии, найденные двумя методами. Существенна ли разница?

8.10 (разд. 8.6). Рассмотрите задачу аппроксимации набора результатов измерений (x_i, y_i) , $i = 1, \dots, N$, полиномом $y = A + Bx + Cx^2$. Используйте метод максимального правдоподобия и покажите, что наилучшие оценки для A, B, C , основанные на приведенных данных, даются формулами (8.23). Следуйте аргументам, приведенным между формулами (8.20) и (8.23).

*8.11 (разд. 8.6). Один из методов определения ускорения свободно падающего тела состоит в измерении высот y_i , на которых оно находилось в последовательные равные промежутки времени t_i (например, с помощью кино съемки), и в последующем определении наилучшей аппроксимации ожидаемым полиномом

$$y = y_0 + v_0 t - \frac{1}{2} g t^2. \quad (8.36)$$

Используйте формулы (8.23) для вычисления наилучших оценок для трех коэффициентов в (8.36) и, следовательно, наилучшей оценки для g , основанных на результатах пяти измерений, приведенных в табл. 8.6.

Обратите внимание на то, что мы можем приводить времена в каком угодно виде. Может показаться, что более естественным был бы выбор $t = 0, 1, \dots, 4$. Однако когда вы будете решать задачу, то увидите, что использование времен, симметрично расположенных вокруг нуля, приводит к тому, что примерно половина всех подлежащих вычислению сумм оказывается равной нулю, что сильно упрощает расчеты. Этот прием можно использовать всегда, когда значения независимой переменной разделены равными промежутками.

Таблица 8.6

Время t , десятые доли секунды	—2	—1	0	1	2
Высота y , см	131	113	89	51	7

Таблица 8.7

t , десятичные доли секунды	-4	-2	0	2	4
y , см	3	-16	6	9	-8

8.12 (разд. 8.6). Предположим, что, как ожидают, y имеет зависимость от x вида $y = Af(x) + Bg(x)$, где A и B — неизвестные параметры, а f и g — фиксированные известные функции (например, такие, как $f = x$ и $g = x^2$ или $f = \cos x$ и $g = \sin x$). Используйте принцип максимального правдоподобия и покажите, что наилучшие оценки для A и B , основанные на данных (x_i, y_i) , $i = 1, \dots, N$, должны определяться из

$$\begin{aligned} A \sum [f(x_i)]^2 + B \sum f(x_i)g(x_i) &= \sum y_i f(x_i), \\ A \sum f(x_i)g(x_i) + B \sum [g(x_i)]^2 &= \sum y_i g(x_i). \end{aligned} \quad (8.37)$$

*8.13 (разд. 8.6). Груз, подвешенный к вертикально расположенной пружине, совершает колебания, которые описываются высотой его положения y , определяемой формулой

$$y = A \cos \omega t + B \sin \omega t.$$

Студент измеряет ω и находит, что $\omega = 1$ рад/с с пренебрежимо малой погрешностью. При помощи киносъемки он определяет затем y для пяти равноотстоящих моментов времени, как показано в табл. 8.7.

Используйте формулы (8.37) и найдите наилучшие оценки A и B . Нарисуйте на графике данные и вашу наилучшую аппроксимацию. (Если вы сначала нарисуете данные, то убедитесь, как трудно было бы подобрать наилучшую аппроксимацию без помощи метода наименьших квадратов.) Если студент пришел к выводу, что его измеренные значения y были неточны в пределах «пары сантиметров», могли бы вы сказать, что такие данные хорошо аппроксимируются ожидаемой кривой?

*8.14 (разд. 8.6). Число распадов R в единицу времени для некоторого образца радиоактивного материала уменьшается экспоненциально со временем по мере истощения образца:

$$R = R_0 e^{-t/\tau},$$

где τ — среднее время жизни образца. Студент проводит наблюдение за образцом радиоактивного материала в течение трех часов и получает результаты, приведенные в табл. 8.8. Определите аппроксимацию $\ln R = \ln R_0 - t/\tau$ методом наименьших квадратов и найдите наилучшую оценку для среднего времени жизни τ .

Таблица 8.8

Время t	0	1	2	3
Темп R , произв. ед.	13,8	7,9	6,1	2,9

Глава 9

Смешанный второй момент и корреляция

В этой главе мы введем важное понятие смешанного второго момента¹⁾. Понятие смешанного второго момента естественно возникает в связи с проблемой расчета ошибок в косвенных измерениях, поэтому мы введем его в разд. 9.2 после краткого обзора расчета ошибок в косвенных измерениях в разд. 9.1. В разд. 9.3 мы используем это понятие второго смешанного момента для определения коэффициента линейной корреляции для случая N измеренных точек $(x_1, y_1), \dots, (x_N, y_N)$. Этот коэффициент, обозначаемый r , является мерой того, насколько хорошо полученные точки (x_i, y_i) ложатся на прямую линию вида $y = A + Bx$. Использование этого коэффициента рассматривается в разд. 9.4 и 9.5.

9.1. Обзор расчета ошибок в косвенных измерениях

В этом и следующем разделах мы в последний раз вернемся к важному вопросу расчета ошибок в косвенных измерениях. Впервые мы рассмотрели вычисление ошибок в косвенных измерениях в гл. 3, где было сделано несколько выводов. Мы считали, что измеряются две величины x и y и вычисляется некоторая функция $q(x, y)$, такая, например, как $q = x + y$ или $q = x^2 \sin y$. (Фактически мы рассматривали также функцию $q(x, \dots, z)$ произвольного числа переменных $x, \dots, z$; для простоты мы будем обсуждать здесь только случай двух переменных.) Простое рассмотрение показывает, что погрешность значения q определяется как

$$\delta q \approx \left| \frac{\partial q}{\partial x} \right| \delta x + \left| \frac{\partial q}{\partial y} \right| \delta y. \quad (9.1)$$

Сначала мы получили этот результат для простых частных случаев сумм, разностей, произведений и частных. Например,

¹⁾ В литературе на русском языке для этого понятия используются также термины *ковариация* и *корреляционный момент*. — Прим. перев.

если q есть сумма $q = x + y$, то (9.1) сводится к известному выражению $\delta q \approx \delta x + \delta y$. Общий результат (9.1) был получен в (3.43).

Затем мы поняли, что соотношение (9.1), вероятно, часто переоценивает δq , поскольку возможна частичная компенсация ошибок в x и y . Мы утверждали без доказательства, что когда ошибки в x и y независимы и случайны, то лучшая оценка погрешности в рассчитанном значении $q(x, y)$ есть

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x} \delta x\right)^2 + \left(\frac{\partial q}{\partial y} \delta y\right)^2}. \quad (9.2)$$

Мы также без доказательства утверждали, что вне зависимости от того, являются ли ошибки случайными и независимыми, более простая формула (9.1) всегда дает верхний предел для δq , т. е. погрешность δq никогда не больше того значения, которое определяется (9.1).

В гл. 5 мы дали надлежащее определение и доказательство соотношения (9.2). Во-первых, мы увидели, что хорошей мерой погрешности δx в измерениях является стандартное отклонение σ_x ; в частности, мы показали, что если результаты измерений x распределены нормально, то с 68%-ной вероятностью можно быть уверенным в том, что измеренное значение лежит в пределах σ_x от истинного значения. Во-вторых, мы видели, что если результаты измерений x и y подчиняются независимым нормальным распределениям со стандартными отклонениями σ_x и σ_y , то значения $q(x, y)$ также распределены нормально со стандартным отклонением

$$\sigma_q = \sqrt{\left(\frac{\partial q}{\partial x} \sigma_x\right)^2 + \left(\frac{\partial q}{\partial y} \sigma_y\right)^2}. \quad (9.3)$$

Этот результат обосновывает нашу оценку ошибки (9.2).

В разд. 9.2 мы выведем точную формулу для погрешности в q , которая применима вне зависимости от того, независимы ли x и y и имеют ли они нормальное распределение. В частности, мы докажем, что (9.1) всегда дает верхнюю границу погрешности в q .

Прежде чем получить эти результаты, обсудим еще раз определение стандартного отклонения. Стандартное отклонение σ_x для N результатов измерений $x_1, \dots, x_N$ первоначально было определено формулой

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2. \quad (9.4)$$

Если результаты измерений x распределены нормально, то в пределе, когда число N велико, определение (9.4) эквива-

лентно определению σ_x как параметра ширины, появляющегося в функции Гаусса

$$\frac{1}{\sigma_x \sqrt{2\pi}} e^{-(x-\bar{x})^2/2\sigma_x^2},$$

в соответствии с которой распределяются результаты измерений x . Поскольку теперь мы будем рассматривать случай, когда ошибки в x могут не быть нормально распределенными, то это второе определение для нас больше неприменимо. Однако мы можем и будем определять σ_x , как в (9.4). Вне зависимости от того, нормально или нет распределены ошибки, это определение σ_x дает разумную меру случайных погрешностей в нашем измерении x . (Как и в гл. 5, мы будем предполагать, что все систематические ошибки выявлены и уменьшены до пренебрежимо малого уровня, так что все оставшиеся погрешности случайны.)

Остается обычная неопределенность: будем ли мы использовать для σ_x определение (9.4) или же «улучшенное» определение, когда N в знаменателе заменяется на $(N-1)$. К счастью, все рассуждения, которые мы будем приводить, одинаково применимы к любому определению, пока мы последовательно придерживаемся одного из них. Для простоты мы будем использовать определение (9.4) с N в знаменателе.

9.2. Смешанный второй момент при расчете ошибок в косвенных измерениях

Предположим, что для нахождения функции $q(x, y)$ мы измеряем две величины x и y несколько раз и получаем N пар данных $(x_1, y_1), \dots, (x_N, y_N)$. По результатам N измерений $x_1, \dots, x_N$ мы можем вычислить среднее $\bar{x}$ и стандартное отклонение σ_x обычным образом; аналогично по данным $y_1, \dots, y_N$ мы можем вычислить $\bar{y}$ и σ_y . Затем, используя N пар результатов измерений, мы можем рассчитать N значений величины, представляющей интерес:

$$q_i = q(x_i, y_i), \quad (i = 1, \dots, N).$$

Получив $q_1, \dots, q_N$, мы сможем вычислить их среднее $\bar{q}$, которое, как мы предполагаем, дает наилучшую оценку для q , и стандартное отклонение σ_q , являющееся мерой случайной погрешности в значениях q_i .

Будем предполагать, что все наши погрешности малы и что, следовательно, все числа $x_1, \dots, x_N$ близки к $\bar{x}$, а все $y_1, \dots, y_N$ — к $\bar{y}$. В этом случае справедливо приближение

$$\begin{aligned} q_i &= q(x_i, y_i) \approx \\ &\approx q(\bar{x}, \bar{y}) + \frac{\partial q}{\partial x}(x_i - \bar{x}) + \frac{\partial q}{\partial y}(y_i - \bar{y}). \end{aligned} \quad (9.5)$$

В этом выражении частные производные $\partial q/\partial x$ и $\partial q/\partial y$ вычисляются в точке $x = \bar{x}$, $y = \bar{y}$, и, следовательно, они одинаковы для всех $i = 1, \dots, N$. В рамках этого приближения среднее принимает вид

$$\bar{q} = \frac{1}{N} \sum_{i=1}^N q_i = \frac{1}{N} \sum_{i=1}^N \left[q(\bar{x}, \bar{y}) + \frac{\partial q}{\partial x} (x_i - \bar{x}) + \frac{\partial q}{\partial y} (y_i - \bar{y}) \right].$$

Это выражение для $\bar{q}$ есть сумма трех членов. Первый член — $q(\bar{x}, \bar{y})$, а два других точно равны нулю. (Например, из определения среднего $\bar{x}$ следует, что $\sum (x_i - \bar{x}) = 0$.) Таким образом, мы получили замечательно простой результат

$$\bar{q} = q(\bar{x}, \bar{y}), \quad (9.6)$$

т. е. чтобы найти среднее $\bar{q}$, мы просто должны вычислить значение функции $q(x, y)$ в точке $x = \bar{x}$ и $y = \bar{y}$.

Стандартное отклонение для N значений $q_1, \dots, q_N$ есть

$$\sigma_q^2 = \frac{1}{N} \sum (q_i - \bar{q})^2.$$

Подставляя в это выражение (9.5) и (9.6), мы находим, что

$$\begin{aligned} \sigma_q^2 &= \frac{1}{N} \sum \left[\frac{\partial q}{\partial x} (x_i - \bar{x}) + \frac{\partial q}{\partial y} (y_i - \bar{y}) \right]^2 = \\ &= \left(\frac{\partial q}{\partial x} \right)^2 \frac{1}{N} \sum (x_i - \bar{x})^2 + \left(\frac{\partial q}{\partial y} \right)^2 \frac{1}{N} \sum (y_i - \bar{y})^2 + \\ &+ 2 \frac{\partial q}{\partial x} \frac{\partial q}{\partial y} \frac{1}{N} \sum (x_i - \bar{x}) (y_i - \bar{y}). \end{aligned} \quad (9.7)$$

Суммы в первых двух членах — это определения стандартных отклонений σ_x и σ_y . Последняя сумма имеет вид, с которым мы еще не встречались прежде. Она называется *смешанным вторым моментом*¹⁾ x и y и обозначается

$$\sigma_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}) (y_i - \bar{y}). \quad (9.8)$$

¹⁾ Термин *смешанный второй момент* для σ_{xy} (для переменных x, y) введен, чтобы провести параллель с термином *второй момент*²⁾ для σ_x^2 (для одной переменной x). Чтобы подчеркнуть эту параллель, смешанный второй момент иногда обозначается σ_{xy}^2 , что не особенно удачно, поскольку смешанный второй момент может быть отрицательным. Удобная особенность нашего определения (9.8) заключается в том, что σ_{xy} имеет размерность xy , так же как σ_x имеет размерность x .

²⁾ Чтобы не нарушать логику употребления терминов, мы перевели в данном случае термин, обозначающий *дисперсию*, как *второй момент*, который широко встречается в литературе на русском языке. — Прим. перев.

С этим определением выражение (9.7) для стандартного отклонения σ_q переходит в

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2 + 2 \frac{\partial q}{\partial x} \frac{\partial q}{\partial y} \sigma_{xy}. \quad (9.9)$$

Это выражение определяет стандартное отклонение σ_q вне зависимости от того, независимы ли измерения x и y и нормально ли они распределены.

Если измерения x и y *независимы*, то легко видеть, что после многих измерений смешанный второй момент σ_{xy} должен стремиться к нулю¹⁾. Вне зависимости от значения y_i величина $x_i - \bar{x}$ с равной вероятностью может быть как положительной, так и отрицательной. Таким образом, после многих измерений положительные и отрицательные члены в (9.8) должны примерно скомпенсироваться, и в пределе бесконечного числа измерений множитель $1/N$ в (9.8) обеспечит равенство σ_{xy} нулю. (После конечного числа измерений величина σ_{xy} не будет точно равна нулю, но должна быть *мала*, если ошибки в x и y действительно независимы и случайны.) При $\sigma_{xy} = 0$ выражение для σ_q сводится к

$$\sigma_q^2 = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2, \quad (9.10)$$

знакомому результату для независимых и случайных погрешностей.

Если измерения x и y *не* независимы, то смешанный второй момент σ_{xy} не должен равняться нулю. Например, легко представить себе ситуацию, когда переоценка x всегда влечет за собой переоценку y и наоборот. В этом случае числа $(x_i - \bar{x})$ и $(y_i - \bar{y})$ будут всегда иметь один и тот же знак (положительный или отрицательный), а их произведение будет всегда положительно. Поскольку все члены в сумме (9.8) положительны, то σ_{xy} не должна исчезать даже в пределе, когда мы делаем бесконечно много измерений.

Когда смешанный второй момент σ_{xy} отличен от нуля (даже в пределе бесконечно большого числа измерений), мы говорим, что ошибки в x и y *коррелированы*. В этом случае погрешность σ_q в $q(x, y)$, определяемая (9.9), это *не* то же самое, что мы получили бы в соответствии с формулой (9.10) для независимых и случайных ошибок.

¹⁾ В случае конечного числа измерений N в литературе на русском языке для σ_{xy} используется понятие *выборочный смешанный второй момент* по аналогии с выборочным средним, выборочным стандартным отклонением и т. д. — *Прим. перев.*

С помощью формулы (9.9) мы можем получить верхний предел для σ_q , который всегда справедлив. С помощью простых алгебраических преобразований (задача 9.1) можно показать, что смешанный второй момент σ_{xy} удовлетворяет так называемому *неравенству Шварца*

$$|\sigma_{xy}| \leq \sigma_x \sigma_y. \quad (9.11)$$

Если подставить (9.11) в выражение (9.9) для погрешности σ_q , то получим, что

$$\begin{aligned} \sigma_q^2 &\leq \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^2 + 2 \left| \frac{\partial q}{\partial x} \frac{\partial q}{\partial y} \right| \sigma_x \sigma_y = \\ &= \left[\left| \frac{\partial q}{\partial x} \right| \sigma_x + \left| \frac{\partial q}{\partial y} \right| \sigma_y \right]^2, \end{aligned}$$

т. е.

$$\boxed{\sigma_q \leq \left| \frac{\partial q}{\partial x} \right| \sigma_x + \left| \frac{\partial q}{\partial y} \right| \sigma_y.} \quad (9.12)$$

Этим результатом мы окончательно установили точный смысл нашего первоначального простого выражения

$$\delta q \approx \left| \frac{\partial q}{\partial x} \right| \delta x + \left| \frac{\partial q}{\partial y} \right| \delta y \quad (9.13)$$

для погрешности δq . Если рассматривать стандартное отклонение σ_q как нашу меру погрешности в q , то (9.12) показывает, что старое выражение (9.13) в действительности есть *верхний предел* погрешности. Вне зависимости от того, независимы ли ошибки в x и y и нормально ли они распределены, погрешность в q никогда не больше правой части (9.13). Если измерения x и y коррелированы таким образом, что $|\sigma_{xy}| = \sigma_x \sigma_y$, т. е. $|\sigma_{xy}|$ достигает максимально возможного значения в соответствии с (9.11), то погрешность в q может быть действительно такой, как она дается (9.13), но она никогда не может быть больше.

Роль смешанного второго момента σ_{xy} в нашем обсуждении расчета ошибок в косвенных измерениях является чисто теоретической, и практически понятие смешанного второго момента обычно не имеет существенного значения при расчете ошибок в косвенных измерениях. Теперь мы обсудим проблему, в которой это понятие действительно играет центральную роль и имеет практическое значение.

9.3. Коэффициент линейной корреляции

Понятие смешанного второго момента σ_{xy} , введенное в разд. 9.2, позволяет нам ответить на вопрос, поставленный в гл. 8 о том, насколько хорошо набор результатов измерений

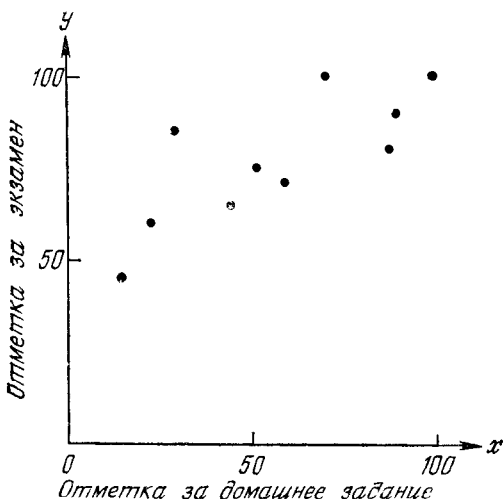


Рис. 9.1. «График рассеяния», показывающий отметки студентов за экзамены и домашние задания. Каждая из десяти точек (x_i, y_i) указывает отметку студента за домашнее задание x_i и за экзамен y_i .

$(x_1, y_1), \dots, (x_N, y_N)$ для двух переменных подтверждает гипотезу о линейной зависимости x и y .

Предположим, что мы получили N пар измеренных значений $(x_1, y_1), \dots, (x_N, y_N)$ двух переменных, которые, как мы ожидаем, должны быть связаны линейной зависимостью вида

$$y = A + Bx.$$

Важно заметить, что $x_1, \dots, x_N$ в данном случае не результаты измерений лишь одной величины, как это было в случае двух последних разделов; на самом деле это результаты измерений N различных значений одной переменной (например, N различных высот, с которых мы бросали камень). То же самое относится и к $y_1, \dots, y_N$.

С помощью метода наименьших квадратов мы можем найти значения A и B для линии, которая наилучшим образом аппроксимирует точки $(x_1, y_1), \dots, (x_N, y_N)$. Если у нас есть надежные оценки погрешностей в измерениях, то мы можем видеть, действительно ли измеренные точки лежат разумно близко к линии (по сравнению с известными погрешностями). Если это так, то измерения подтверждают наше предположение, что x и y связаны линейно.

К сожалению, во многих экспериментах трудно определить надежные оценки погрешностей заранее, и поэтому мы должны использовать исходные данные, чтобы судить, связаны ли две переменные линейно. В частности, имеется такой при-

мер эксперимента, когда *невозможно* определить величину погрешностей заранее. Этот эксперимент, который более подходит к социальным, чем к физическим, наукам, лучше пояснить на примере.

Представим себе, что профессор, желающий убедить своих студентов в том, что выполнение домашних заданий поможет им хорошо сдать экзамены, собирает сведения об их отметках за домашнее задание и за экзамен и изображает их на «графике разбросов», как показано на рис. 9.1. На этом графике отметки за домашнее задание отложены по горизонтальной оси, а за экзамен — по вертикальной¹⁾. Каждая точка (x_i, y_i) показывает оценку одного студента за домашнее задание x_i и за экзамен y_i . Профессор надеется показать, что высокие оценки за экзамен *коррелируют* с высокими отметками за домашнее задание и наоборот (и его график разбросов определенно подтверждает, что это приблизительно так). В этом примере эксперимента нет никаких погрешностей в точках; две отметки каждого студента известны точно²⁾. Погрешность будет скорее в степени, до которой *коррелированы* отметки, и именно это должно быть определено из данных.

Две переменные x и y (в случае любого типичного физического эксперимента или такого, как описанный выше) могут быть, конечно, связаны и более сложной зависимостью, чем простая линейная связь вида $y = A + Bx$. Например, множество физических законов приводит к квадратичной зависимости типа $y = A + Bx + Cx^2$. Тем не менее мы ограничим наше рассмотрение случаем более простой задачи, когда надо решить, подтверждает ли данный набор точек гипотезу о *линейной* связи $y = A + Bx$.

Степень, до которой набор точек $(x_1, y_1), \dots, (x_N, y_N)$ подтверждает линейную зависимость между x и y , измеряется *коэффициентом линейной корреляции*, или просто *коэффициентом корреляции*

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}, \quad (9.14)$$

где смешанный второй момент σ_{xy} и стандартные отклонения σ_x и σ_y определяются точно так же, как и ранее, формулами (9.8) и (9.4)³⁾. Подставляя эти определения в (9.14), мы

¹⁾ На рис. 9.1 приведены отметки по 100-балльной системе. — *Прим. перев.*

²⁾ В принципе, ввиду того что трудно абсолютно точно и объективно выставить оценку (например, отметки между тройкой и четверкой в случае нашей системы оценок), можно ввести в рассмотрение погрешность оценки. — *Прим. перев.*

³⁾ Заметьте, однако, что их значения слегка различны. Например, в разд. 9.2 $x_1, \dots, x_N$ были результатами измерений *одного* числа, и если бы эти измерения были точными, то величина σ_x была бы малой. В дан-

можем переписать выражение для коэффициента корреляции в виде

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2]^{1/2}}. \quad (9.15)$$

Как мы скоро увидим, число r показывает, насколько хорошо точки (x_i, y_i) аппроксимируются прямой линией. Это число принимает значения между -1 и 1 . Если r близко к ± 1 , то точки лежат вблизи некоторой прямой линии; если r близко к 0 , то точки не коррелированы и либо незначительно, либо совсем не группируются около прямой линии.

Чтобы доказать эти утверждения, сначала заметим, что из неравенства Шварца (9.11) $|\sigma_{xy}| \leq \sigma_x \sigma_y$ сразу же следует, что $|r| \leq 1$ или

$$-1 \leq r \leq 1,$$

как и утверждалось. Далее, предположим, что все точки (x_i, y_i) лежат *точно* на линии $y = A + Bx$. В этом случае $y_i = A + Bx_i$ для всех i и, следовательно, $\bar{y} = A + B\bar{x}$. Вычитая эти два равенства, мы видим, что

$$y_i - \bar{y} = B(x_i - \bar{x})$$

для каждого i . Подставляя полученное выражение в (9.15), находим

$$r = \frac{B \sum (x_i - \bar{x})^2}{[\sum (x_i - \bar{x})^2 B^2 \sum (x_i - \bar{x})^2]^{1/2}} = \frac{B}{|B|} = \pm 1, \quad (9.16)$$

т. е. если точки $(x_1, y_1), \dots, (x_N, y_N)$ лежат точно на прямой, то $r = \pm 1$, причем знак r определяется наклоном линии ($r = 1$ для положительного B и $r = -1$ для отрицательного B)¹⁾. Даже если переменные x и y действительно связаны линейной зависимостью, мы не должны ожидать, что экспериментальные точки будут лежать *точно* на линии. Таким образом, не следует ожидать, что r точно равно ± 1 . С другой стороны, мы действительно должны ожидать, что r *близко* к ± 1 , если считаем, что x и y связаны линейно.

ном случае $x_1, \dots, x_N$ — результаты измерений *различных* значений переменной, и даже если бы измерения были точными, то все равно не было бы основания думать, что величина σ_x будет малой. Заметим также, что некоторые авторы используют число r^2 называя его *коэффициентом точности измерений*.

¹⁾ Если линия строго горизонтальна, то $B = 0$, и (9.16) дает $r = 0/0$, т. е. r — неопределенная величина. К счастью, этот частный случай не важен для практики, так как он соответствует ситуации, когда y — постоянная, не зависящая от x .

Предположим, что нет никакой связи между переменными x и y . Тогда вне зависимости от значения y_i каждое x_i с одинаковой вероятностью может быть как больше, так и меньше $\bar{x}$. Таким образом, члены в сумме

$$\sum (x_i - \bar{x})(y_i - \bar{y})$$

в числителе выражения (9.15) для r с одинаковой вероятностью могут быть как положительными, так и отрицательными. В то же время члены в знаменателе выражения для r всегда положительны. Таким образом, в пределе, когда число измерений N стремится к бесконечности, коэффициент корреляции r будет равен нулю. В случае конечного числа экспериментальных точек мы не должны ожидать, что коэффициент r будет точно равен нулю, но мы действительно ожидаем, что он должен быть *мал* (если две переменные действительно не связаны линейной зависимостью).

Если две переменные x и y таковы, что в пределе бесконечно большого числа измерений их смешанный второй момент равен нулю (и, следовательно, $r = 0$), то мы говорим, что переменные *некоррелированы*. Если после конечного числа измерений коэффициент корреляции $r = \sigma_{xy}/\sigma_x\sigma_y$ мал, это будет подтверждением гипотезы о том, что x и y не коррелированы.

В качестве иллюстрации можно рассмотреть пример с отметками за экзамен и домашнее задание, показанный на рис. 9.1. Эти отметки приведены также в табл. 9.1. Простой расчет (задача 9.4) дает, что коэффициент корреляции для этих десяти пар отметок равен $r = 0,8$. Профессор делает вывод, что это значение «разумно близко» к 1, и поэтому может объявить студентам в следующем году, что поскольку имеется хорошая корреляция между отметками за домашнее задание и за экзамен, то очень важно выполнять домашнее задание.

Если бы профессор получил, что коэффициент корреляции r близок к нулю, то он оказался бы в затруднительном положении, поскольку он обнаружил бы, что отметки за домашнее задание никак не связаны с отметками за экзамен. Если бы получилось, что величина r близка к -1 , то ему пришлось бы сделать еще более смущающее открытие, что

Таблица 9.1. Отметки студентов

Студент, i	1	2	3	4	5	6	7	8	9	10
Домашнее задание, x_i	90	60	45	100	15	23	52	30	71	88
Экзамен, y_i	90	71	65	100	45	60	75	85	100	80

отметки за домашнее задание и за экзамен подчинены *отрицательной* корреляции, т. е. что студенты, которые хорошо выполняют домашнее задание, плохо сдают экзамены.

9.4. Количественный критерий значимости r

Из рассмотренного примера должно быть ясно, что у нас все еще нет полного ответа на первоначальный вопрос о том, насколько хорошо экспериментальные точки подтверждают линейную связь между x и y . Наш профессор получил коэффициент корреляции $r = 0,8$, и сделал вывод о том, что это значение «разумно близко» к 1. Но как можно объективно решить, что значит «разумно близко к 1»? Будет ли $r = 0,6$ разумно близко? Или 0,4? Мы можем ответить на эти вопросы следующим образом.

Предположим, что две переменные x и y в действительности *некоррелированы*, т. е. в пределе бесконечно большого числа измерений коэффициент корреляции был бы равен нулю. После конечного числа измерений очень маловероятно, чтобы r был точно равен нулю¹⁾. Оказывается, можно вычислить вероятность того, что r будет не меньше, чем любое заданное значение. Обозначим через

$$P_N(|r| \geq r_0)$$

вероятность того, что N измерений двух некоррелированных переменных x и y приведут к значению коэффициента r , не меньшему²⁾, чем любое частное значение r_0 . Например, мы могли бы рассчитать вероятность

$$P_N(|r| \geq 0,8)$$

того, что после N измерений некоррелированных переменных x и y коэффициент корреляции будет по крайней мере не меньше, чем полученное профессором значение 0,8. Расчет таких вероятностей довольно сложен, и мы его здесь не будем приводить. Однако результаты таких вычислений для небольшого числа значений параметров представлены в табл. 9.2, а более полные таблицы приведены в приложении В.

Хотя мы не показали, как вычисляются вероятности, приведенные в табл. 9.2, можно понять в общих чертах их поведение и научиться их использовать. Левый столбец показывает число экспериментальных точек N . (В нашем примере

¹⁾ В случае конечного числа измерений в литературе на русском языке употребляется термин «выборочный коэффициент корреляции». — *Прим. перев.*

²⁾ Так как корреляция означает, что r близко к $+1$, или -1 , то мы рассматриваем вероятность получения *абсолютного значения* $|r| \geq r_0$.

Таблица 9.2. Вероятность $P_N (|r| \geq r_0)$ того, что N измерений двух некоррелированных переменных x и y дадут коэффициент корреляции $|r| \geq r_0$. Значения вероятностей приведены в процентах; прочерками отмечены значения, меньшие 0,05%

N	r_0										
	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
3	100	94	87	81	74	67	59	51	41	29	0
6	100	85	70	56	43	31	21	12	6	1	0
10	100	78	58	40	25	14	7	2	0,5	—	0
20	100	67	40	20	8	2	0,5	0,1	—	—	0
50	100	49	16	3	0,4	—	—	—	—	—	0

профессор собрал отметки десяти студентов, так что $N = 10$.) Числа в каждом последующем столбце — вероятности того, что N измерений двух некоррелированных переменных дадут коэффициент r , который по крайней мере не меньше, чем самое верхнее число в столбце. Например, как мы видим, вероятность того, что десять некоррелированных точек дадут $|r| \geq 0,8$, невелика, составляет только 0,5%. Следовательно, наш профессор может сказать, что *весьма невероятно*, чтобы некоррелированные отметки дали для коэффициента корреляции значение $|r|$, большее или равное величине 0,8, которую он получил. Другими словами, очень *вероятно*, что отметки за домашнее задание и за экзамен действительно коррелированы.

Несколько особенностей табл. 9.2 нуждаются в комментариях. Все значения в первом столбце равны 100%, потому что $|r|$ всегда больше или равно нулю, так что вероятность получения $|r| \geq 0$ всегда равна 100%. Аналогично все значения в последнем столбце равны нулю, так как вероятность получения $|r| \geq 1$ равна нулю¹⁾. Числа в промежуточных столбцах изменяются с числом экспериментальных точек N . Это также легко понять. Если мы сделаем только три измерения, то вероятность получить коэффициент корреляции, скажем $|r| \geq 0,5$, очевидно, довольно велика (фактически 67%). Но если мы сделаем 20 измерений и если две переменные действительно не коррелированы, то вероятность получить $|r| \geq 0,5$, очевидно, очень мала (фактически 2%).

Получив значения вероятности из табл. 9.2 (или из более полных таблиц приложения В), мы теперь можем дать наи-

¹⁾ Хотя невозможно получить $|r| > 1$, но $|r| = 1$ в принципе возможно. Однако r — непрерывная переменная, и вероятность получения значения $|r|$, точно равного единице, равна нулю. Таким образом, $P_N(|r| \geq 1) = 0$.

более полный возможный ответ на вопрос о том, насколько хорошо N пар значений (x_i, y_i) подтверждают линейную связь между x и y . По измеренным точкам можно сначала вычислить значение коэффициента корреляции r_o . Затем, используя одну из этих таблиц, мы можем найти вероятность $P_N(|r| \geq |r_o|)$ того, что N некоррелированных точек дадут для коэффициента значение не меньшее, чем полученный коэффициент r_o . Если эта вероятность «достаточно мала», то мы можем заключить, что очень *невероятно*, чтобы x и y были не коррелированы, и, следовательно, очень *вероятно*, что они в действительности коррелированы.

Мы еще должны выбрать значение вероятности, которое будем рассматривать как «достаточно малое». Один довольно распространенный выбор состоит в том, чтобы рассматривать наблюдаемую корреляцию r_o как «значимую», если вероятность получения коэффициента r , такого, что $|r| \geq |r_o|$, для некоррелированных переменных меньше 5%. Корреляцию иногда называют «высокозначимой», если соответствующая вероятность меньше 1%. Какой бы выбор мы ни сделали, мы *не* получим точно определенного ответа, какие данные коррелированы, а какие нет; вместо этого у нас есть количественная мера, показывающая, насколько *невероятно* предположение о том, что они не коррелированы.

9.5. Примеры

Предположим, что мы измеряем три пары значений (x_i, y_i) и находим, что коэффициент корреляции равен 0,7 (или -0,7). Подтверждает ли это значение гипотезу, что x и y связаны линейно?

Обращаясь к табл. 9.2, мы видим, что даже если переменные x и y совсем не коррелированы, то вероятность получения $|r| \geq 0,7$ при $N=3$ составляет 51%. Другими словами, вполне возможно, что x и y не коррелированы; таким образом, у нас нет надежного доказательства корреляции. Действительно, в случае только трех измерений было бы очень трудно получить убедительное подтверждение корреляции. Даже наблюдаемое значение коэффициента 0,9 недостаточно для утверждения корреляции, поскольку вероятность получения $|r| \geq 0,9$ в случае трех измерений некоррелированных переменных равна 29%.

Если бы мы нашли значение коэффициента 0,7 по шести измерениям, то ситуация была бы несколько лучше, но все же еще недостаточно хорошей. С $N=6$ вероятность получения $|r| \geq 0,7$ для некоррелированных переменных равна 12%. Эта цифра не так мала, чтобы исключить возможность того, что x и y не коррелированы.

С другой стороны, если бы мы получили $r = 0,7$ после 20 измерений, то у нас было бы сильное подтверждение корреляции, так как при $N = 20$ вероятность получения $|r| \geq 0,7$ для двух некоррелированных переменных равна только 0,1%. По любым критериям это очень неправдоподобно, и мы могли бы уверенно утверждать, что корреляция обнаружена. В частности, эта корреляция могла быть названа «высокозначимой», так как соответствующая вероятность меньше 1%.

Задачи

Напоминание: звездочка у номера задачи означает, что задача решается или ее ответ приводится в разделе «Ответы» в конце книги.

***9.1** (разд. 9.2). Докажите, что смешанный второй момент σ_{xy} , определенный в (9.8), удовлетворяет неравенству Шварца (9.11)

$$|\sigma_{xy}| \leq \sigma_x \sigma_y. \quad (9.17)$$

Указание: введите произвольное число t и рассмотрите функцию

$$A(t) = \frac{1}{N} \sum [(x_i - \bar{x}) + t(y_i - \bar{y})]^2 \geq 0. \quad (9.18)$$

Так как $A(t)$ — положительная величина независимо от значения t , то вы можете найти минимальное значение $A_{\min}$, приравняв производную dA/dt нулю, и это минимальное значение будет все же больше или равно нулю. Покажите, что $A_{\min} = \sigma_x^2 - (\sigma_{xy}^2/\sigma_y^2)$, и получите (9.17).

9.2 (разд. 9.2).

а. Представьте себе серию N измерений двух фиксированных длин x и y , выполненных с целью вычислить значение некоторой функции $q(x, y)$. Предположите, что использовалось несколько различных типов рулеток, но при этом каждая пара точек (x_i, y_i) измерялась одной и той же рулеткой, т.е. пара (x_1, y_1) измерялась одной рулеткой, (x_2, y_2) — другой рулеткой и т. д. Предполагая, что главным источником ошибок является некоторая укороченность одних и некоторая удлиненность других рулеток, четко покажите, что смешанный второй момент σ_{xy} должен быть положительным.

б. Покажите далее, что при тех же предположениях $\sigma_{xy} = \sigma_x \sigma_y$, т.е. что значение σ_{xy} настолько велико, насколько это позволяет неравенство Шварца (9.17).

Указание: сделайте допущение, что i -я рулетка короче на фактор α_i (где α_i близко к 1), тогда длина, которая в действительности равна X , будет измерена как $x_i = \alpha_i X$. Смысл этой задачи заключается в том, чтобы показать, что существуют ситуации, когда смешанным вторым моментом определенно нельзя пренебречь.

***9.3** (разд. 9.3).

а. Докажите тождество

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - N \bar{x} \bar{y}.$$

б. Следовательно, покажите, что выражение для коэффициента корреляции r (9.15) можно переписать как

$$r = \frac{\sum x_i y_i - N \bar{x} \bar{y}}{[(\sum x_i^2 - N \bar{x}^2)(\sum y_i^2 - N \bar{y}^2)]^{1/2}}. \quad (9.19)$$

Таблица 9.3

x_i	74	83	85	96	98	100	106	107	120	124
y_i	76	103	99	109	111	107	91	101	120	119

Часто это более удобный способ вычислять r , так как он не приводит к необходимости рассчитывать индивидуальные отклонения $x_i - \bar{x}$ и $y_i - \bar{y}$.

9.4 (разд. 9.4).

а. Проверьте, что коэффициент корреляции r для десяти пар отметок из табл. 9.1 равен $r \approx 0,8$.

б. Используя таблицу вероятностей из приложения В, найдите вероятность того, что кто-то получил бы коэффициент корреляции r с $|r| \geq 0,8$, если бы две отметки и в самом деле были некоррелированы.

*9.5 (разд. 9.4). В случае фотоэффекта кинетическая энергия K испускаемых электронов, как полагают, есть линейная функция частоты f используемого света:

$$K = hf - \phi, \quad (9.20)$$

где h и ϕ — постоянные. Чтобы проверить это соотношение, студент измеряет K для N различных значений f и вычисляет коэффициент корреляции r по своим результатам.

а. Если он делает пять измерений ($N = 5$) и получает $r = 0,7$, то подтверждает ли он линейную зависимость (9.20) на уровне принятого коэффициента значимости 5 %?

б. А что будет, если $N = 20$ и $r = 0,5$?

*9.6 (разд. 9.4).

а. Начертите график разбросов для следующих пяти пар измерений:

$$\begin{array}{cccccc} x & = & 1 & 2 & 3 & 4 & 5 \\ y & = & 4 & 4 & 3 & 2 & 1. \end{array}$$

Вычислите по этим данным коэффициент корреляции r .

Для этого, вероятно, проще использовать выражение (9.19).

Показывают ли данные, что имеется значимая корреляция?

Нужные вероятности можно найти в приложении В.

б. Повторите задание «а» для следующих данных:

$$\begin{array}{cccccc} x & = & 1 & 2 & 3 & 4 & 5 \\ y & = & 3 & 1 & 2 & 2 & 1. \end{array}$$

9.7 (разд. 9.4). Психолог, исследуя связь между умственными способностями отцов и детей, измеряет некоторый коэффициент интеллектуальности, который мы обозначим для краткости КИ, для десяти отцов и их сыновей и получает результаты, представленные в табл. 9.3, где $x_i = \text{КИ отца}$, $y_i = \text{КИ сына}$.

Подтверждают ли эти данные корреляцию между умственными способностями отцов и сыновей?

Глава 10

Биномиальное распределение

Распределение Гаусса, или нормальное распределение, — пока единственный пример изученного нами распределения. Теперь мы рассмотрим два других важных примера, а именно биномиальное распределение (в гл. 10) и распределение Пуассона (в гл. 11).

10.1. Распределения

В гл. 5 мы ввели понятие *распределения* как функции, которая определяет долю случаев реализации каждого из всех возможных результатов при многократных измерениях. Например, мы могли бы сделать N измерений периода T маятника и найти распределение различных измеренных значений T , или мы могли бы измерять рост h у N американцев и найти распределение различных измеренных значений роста h .

Затем мы ввели понятие *предельного распределения* как распределения, которое мы получили бы в пределе, когда число измерений N становится очень большим. Предельное распределение можно рассматривать также как такое, которое определяет *вероятность* того, что единственное измерение приведет к любому из возможных значений: вероятность того, что одно измерение периода даст какое-то определенное значение T ; вероятность того, что какой-то американец (взятый наудачу) будет иметь некоторый определенный рост h . По этой причине предельное распределение иногда называют также *распределением вероятности*.

Из множества возможных предельных распределений мы рассмотрели только одно распределение Гаусса, или нормальное распределение, которое описывает распределение результатов любых измерений в случае, если эти измерения подвержены действию множества небольших и случайных ошибок. Поэтому распределение Гаусса является для физиков наиболее важным из всех предельных распределений,

и именно поэтому мы уделили ему столько внимания и места. Тем не менее имеется несколько других распределений, которые также важны с теоретической и практической точек зрения. Два примера таких распределений обсуждаются в настоящей и следующих главах.

В данной главе мы рассмотрим биномиальное распределение. Это распределение не имеет слишком большого значения для экспериментальной физики, однако из-за простоты оно может служить отличным введением к пониманию многих характеристик распределений. Оно также важно и с теоретической точки зрения, так как из него мы можем получить более важное распределение Гаусса.

10.2. Вероятности при бросании игральных костей ¹⁾

Биномиальное распределение лучше всего может быть описано на примере. Предположим, что мы предприняли такой «эксперимент», как бросание трех игральных костей, причем каждый раз мы подсчитываем число выпавших граней с одним очком, т. е. число единичек. Возможные результаты такого эксперимента — это 0, 1, 2 или 3 единички. Если мы повторим эксперимент огромное число раз, то найдем предельное распределение, из которого можно будет найти вероятность того, что при любой одной попытке бросания (всех трех костей) мы получим v единичек, где $v = 0, 1, 2$ или 3.

Этот эксперимент достаточно прост, и мы легко можем рассчитать вероятности четырех возможных исходов. Сначала, предполагая, что кости подлинные, отметим, что вероятность выпадения *одного* очка при бросании одной кости равна $1/6$. Теперь будем бросать все три кости и найдем сначала вероятность выпадения трех единичек ($v = 3$). Так как в случае каждой отдельной кости вероятность выпадения одного очка равна $1/6$ и так как все кости вращаются при бросании независимо, то вероятность выпадения трех единичек равна

$$P(3 \text{ единички в } 3 \text{ бросаниях}) = \left(\frac{1}{6}\right)^3 \approx 0,5 \, \%.$$

Рассчитать вероятность выпадения двух единичек ($v = 2$) несколько труднее, так как мы можем получить этот результат различными способами. Первая и вторая кости могли бы выпасть на единички, а третья — нет ($A, A, \text{ не } A$); или пер-

¹⁾ Игральная кость — маленький кубик, на гранях которого имеется маркировка — от одного до шести кружков, которые считаются как очки, т. е. как «единички», «двойки» и т. д. При игре в кости бросают кубик на плоскую поверхность и смотрят, какая грань оказалась наверху, т. е. какие очки «выпали». — *Прим. перев.*

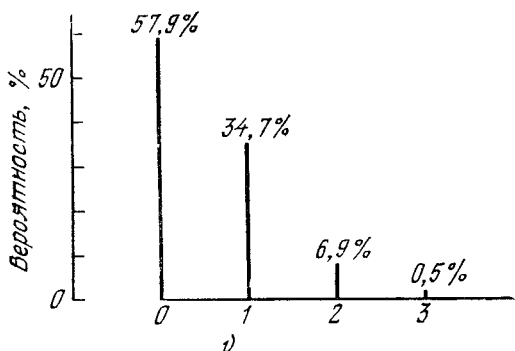


Рис. 10.1. Вероятность получения v единиц при бросании трех игральных костей. Эта функция представляет собой биномиальное распределение $b_{n,p}(v)$ с $n = 3$ и $p = 1/6$.

вая и третья могли бы выпасть на единички, а вторая — нет (A , не A , A) и т. д. Будем вычислять вероятность в два приема. Сначала оценим вероятность выпадения двух единиц при какой-то заданной определенной комбинации, например такой, как (A , A , не A). Вероятность того, что первая кость выпадет на единичку, равна $1/6$ и аналогично для второй. С другой стороны, вероятность того, что последняя кость не выпадет на единичку, равна $5/6$. Таким образом, вероятность выпадения двух единиц для заданной частной комбинации равна

$$P(A, A, \text{не } A) = \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right).$$

Вероятность выпадения двух единиц для любой другой определенной комбинации та же самая. Существуют только три различные комбинации, когда мы могли бы получить две единички: (A , A , не A), или (A , не A , A), или (не A , A , A). Таким образом, полная вероятность получения двух единиц (в какой угодно комбинации) равна

$$P(2 \text{ единички в } 3 \text{ бросаниях}) = 3 \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right) \approx 6,9 \%. \quad (10.1)$$

Подобные расчеты дают значения вероятности выпадения одной единички в трех бросаниях (34,7%) и ни одной единички (57,9%). Полученные численные значения могут быть представлены в виде графика распределения вероятностей для единиц, полученных в одной попытке бросания трех костей, как показано на рис. 10.1. Этот график — пример биномиального распределения, к описанию общего вида которого мы сейчас приступаем.

10.3. Определение биномиального распределения

Чтобы определить в общем виде биномиальное распределение, нам необходимо ввести некоторые понятия. Во-первых, представим, что мы делаем n независимых *испытаний*, таких, как бросание n костей, бросание n монет или опробование n хлопушек. Каждое испытание может иметь различные исходы: кость может выпасть на любую грань от 1 до 6, монета может выпасть на орла или решку, хлопушка может хлопнуть или «прошипеть». Будем называть исход, в котором мы заинтересованы, как *успех* или *выигрыш*. Таким образом, «успехом» могли бы быть выпадения очка при бросании кости, или орла при бросании монеты, или же взрыв хлопушки. Обозначим через p вероятность успеха в любом одном испытании и через $q = 1 - p$ вероятность «проигрыша» (т. е. получения любого исхода, кроме представляющего интерес). Таким образом, при бросании кости вероятность выпадения одного очка $p = 1/6$; вероятность выпадения орла при подбрасывании монеты $p = 1/2$, и вероятность взрыва p могла бы достигать 95% для данного сорта хлопушек.

С помощью этих определений мы теперь можем найти вероятность получения v успехов в n испытаниях. Вычисления, которые мы вскоре опишем, показывают, что эта вероятность дается так называемым *биномиальным распределением*:

$$P(v \text{ успехов в } n \text{ испытаниях}) = b_{n,p}(v) = \frac{n(n-1)\dots(n-v+1)}{1 \cdot 2 \cdot \dots \cdot v} p^v q^{n-v}. \quad (10.2)$$

Буква b в формуле означает «биномиальное»; нижние индексы n и p в $b_{n,p}(v)$ указывают, что распределение зависит от n , числа сделанных испытаний, и p , вероятности успеха в одном испытании.

Распределение (10.2) называется биномиальным из-за его тесной связи с хорошо известным разложением бинома в ряд. А именно дробь в (10.2) — *биномиальный коэффициент*, часто обозначаемый $\binom{n}{v}$

$$\binom{n}{v} = \frac{n(n-1)\dots(n-v+1)}{1 \cdot 2 \cdot \dots \cdot v} = \quad (10.3)$$

$$= \frac{n!}{v!(n-v)!}, \quad (10.4)$$

где мы ввели полезное понятие *факториала*

$$n! = 1 \cdot 2 \cdot \dots \cdot n.$$

Биномиальный коэффициент появляется в разложении бинома

$$\begin{aligned}(p+q)^n &= p^n + np^{n-1}q + \dots + q^n = \\ &= \sum_{v=0}^n \binom{n}{v} p^v q^{n-v},\end{aligned}\quad (10.5)$$

которое справедливо для любых двух чисел p и q и любого положительного целого n (см. задачу 10.4).

Используя обозначение (10.3), мы можем переписать выражение для биномиального распределения в более компактном виде

$$\begin{aligned}P(v \text{ успехов в } n \text{ испытаниях}) &= b_{n,p}(v) = \\ &= \binom{n}{v} p^v q^{n-v},\end{aligned}$$

(10.6)

где, как обычно, p обозначает вероятность успеха в одном испытании и $q = 1 - p$.

Вывод выражения (10.6) аналогичен получению формулы (10.1) в примере с игральными костями:

$$P(2 \text{ единички в } 3 \text{ бросаниях}) = 3 \cdot \left(\frac{1}{6}\right)^2 \cdot \left(\frac{5}{6}\right). \quad (10.7)$$

Действительно, если подставить $v = 2$, $n = 3$, $p = 1/6$ и $q = 5/6$ в (10.6), то получим точно (10.7), что вы должны проверить. Более того, смысл каждого множителя в (10.6) тот же самый, что и смысл соответствующего множителя в (10.7). Множитель p^v — вероятность получения только успехов в любых определенных v испытаниях, а q^{n-v} — вероятность проигрыша в оставшихся $n - v$ испытаниях. Как легко показать, биномиальный коэффициент $\binom{n}{v}$ — число различных комбинаций, в которых получается v успехов в n испытаниях. Приведенное рассуждение показывает, что биномиальное распределение (10.6) на самом деле определяет вероятность, как было сказано выше.

Пример

Предположим, что мы бросаем монету четыре раза ($n=4$) и подсчитываем число выпадений орла v . Какова вероятность получения различных возможных значений $v = 0, 1, 2, 3, 4$?

Поскольку вероятность выпадения орла при одном бросании равна $p = 1/2$, то искомая вероятность — просто биномиальное распределение $b_{n,p}(v)$ с $n = 4$ и $p = q = 1/2$,

$$P(v \text{ орлов в } 4 \text{ бросаниях}) = \binom{4}{v} \left(\frac{1}{2}\right)^4.$$

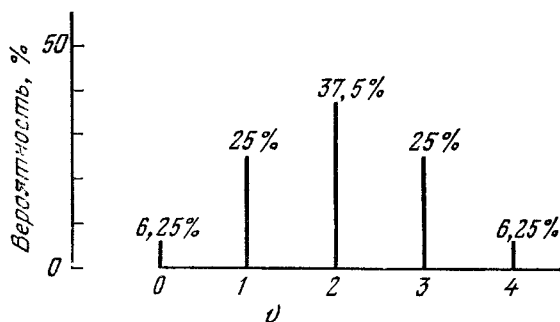


Рис. 10.2. Биномиальное распределение $b_{n,p}(v)$ с $n = 4$, $p = 1/2$, определяющее вероятность выпадения v орлов при подбрасывании четырех монет.

Эти числа легко рассчитываются (задача 10.5) и приводят к распределению, показанному на рис. 10.2.

Мы видим, что наиболее вероятное число выпадений орла $v = 2$, как можно было бы ожидать. В данном случае вероятности симметричны относительно этого наиболее вероятного значения. Иными словами, вероятность выпадения трех орлов та же, что и одного, а вероятность выпадения четырех орлов равна вероятности не выпадения ни одного. Как мы увидим, такая симметрия существует, только если $p = 1/2$.

10.4. Свойства биномиального распределения

Биномиальное распределение $b_{n,p}(v)$ определяет вероятность реализации v «успехов» в n испытаниях в случае, когда p есть вероятность успеха в единственном испытании. Если бы мы повторили полностью весь эксперимент (состоящий из n испытаний) много раз, то было бы естественно спросить: а каково среднее число успехов $\bar{v}$? Это число определяется как

$$\bar{v} = \sum_{v=0}^n v b_{n,p}(v) \quad (10.8)$$

и легко вычисляется (задача 10.8) как

$$\boxed{\bar{v} = np.} \quad (10.9)$$

Таким образом, если мы повторим серию n испытаний много раз, то среднее число успехов будет равно вероятности успеха в одном испытании (p), умноженной на n , как можно было

бы ожидать. Аналогично можно рассчитать и стандартное отклонение σ_v для нашего числа успехов (задача 10.10).

Результат равен

$$\sigma_v = \sqrt{np(1-p)}. \quad (10.10)$$

Когда $p = 1/2$ (как при бросании монеты), то среднее число успехов равно просто $n/2$. Более того, легко показать, что в случае $p = 1/2$

$$b_{n, 1/2}(v) = b_{n, 1/2}(n - v) \quad (10.11)$$

(см. задачу 10.11), т. е. биномиальное распределение в случае $p = 1/2$ симметрично относительно среднего значения $n/2$, как мы заметили на рис. 10.2.

В общем случае, когда $p \neq 1/2$, биномиальное распределение несимметрично. Например, график на рис. 10.1 явно несимметричен; наиболее вероятное число успехов равно $v=0$, и вероятности монотонно уменьшаются для $v = 1, 2$ и 3 . Кроме того, среднее число успехов ($\bar{v} = 0,5$) в данном случае не совпадает с наиболее вероятным числом успехов ($v = 0$).

Интересно сравнить биномиальное распределение $b_{n,p}(v)$ с более привычным распределением Гаусса $f_{x,\sigma}(x)$. Возможно, наибольшее различие состоит в том, что результаты эксперимента, описываемые первым распределением, — это *дискретные*¹⁾ значения $v = 0, 1, 2, \dots, n$, в то время как последнее описывает *непрерывные* значения измерений величины x . Распределение Гаусса представляет собой симметричный пик с центром на среднем значении $x = X$; это означает, что среднее значение X есть также и наиболее вероятное значение (такое, для которого $f_{x,\sigma}(x)$ максимальна). Как мы видели, биномиальное распределение симметрично, только когда $p = 1/2$, а в общем случае среднее значение не совпадает с наиболее вероятным значением.

Аппроксимация биномиального распределения гауссовым

Несмотря на все различия, имеется важная связь между биномиальным и гауссовым распределениями. Если рассмотреть биномиальное распределение $b_{n,p}(v)$ для любого фиксированного значения p в случае, когда n велико, то $b_{n,p}(v)$ хорошо аппроксимируется распределением Гаусса $f_{x,\sigma}(v)$

¹⁾ «Дискретный» означает «отделенный от другого» и противоположно по понятию *непрерывный*.

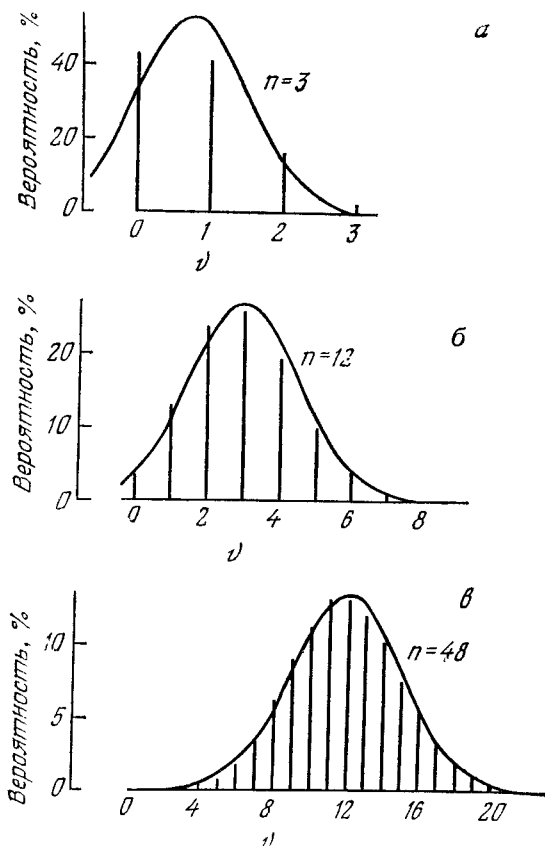


Рис. 10.3. Биномиальные распределения с $p = 1/4$ и $n = 3, 12$ и 48 . Непрерывная кривая на каждом графике есть функция Гаусса с тем же средним и тем же стандартным отклонением

с тем же средним значением и тем же стандартным отклонением, т. е.

$$b_{n,p}(v) \approx f_{X,\sigma}(v) \quad (n \text{ велико}) \quad (10.12)$$

с

$$X = np \quad \text{и} \quad \sigma = \sqrt{np(1-p)}. \quad (10.13)$$

Мы будем ссылаться на (10.12) как на гауссову аппроксимацию биномиального распределения. Мы не будем доказывать здесь этот результат¹⁾, но его справедливость хорошо иллюстрируется рис. 10.3, где показано биномиальное распре-

¹⁾ Доказательство см. в книгах: Meyer S. L. Data Analysis for Scientists and Engineers, John Wiley, 1975, p. 226, или Young H. D. Statistical Treatment of Experimental Data, McGraw-Hill, 1962, приложение С.

деление для $p = 1/4$ и для трех последовательно увеличивающихся значений n ($n = 3, 12, 48$). На каждое биномиальное распределение наложено распределение Гаусса с тем же средним и с тем же стандартным отклонением. В случае только трех испытаний ($n = 3$) биномиальное распределение сильно отличается от соответствующего гауссова. В частности, биномиальное распределение явно асимметрично, в то время как гауссово, конечно, идеально симметрично относительно среднего. В случае $n = 12$ асимметрия биномиального распределения гораздо менее выражена и два распределения близки друг к другу. Когда $n = 48$, различие между биномиальным и соответствующим гауссовым распределениями так незначительно, что почти незаметно в масштабе рис. 10.3, в.

Тот факт, что биномиальное распределение может быть аппроксимировано функцией Гаусса при больших n , очень полезен для практики. Расчеты биномиальной функции для значений n , которые больше или порядка 20, чрезвычайно утомительны, в то время как вычисления функции Гаусса всегда просты для любых значений X и σ . Чтобы проиллюстрировать это, предположим, что мы захотели узнать вероятность выпадения 23 орлов в 36 бросаниях монеты. Эта вероятность дается биномиальным распределением $b_{36, 1/2}(v)$, поскольку вероятность выпадения орла в одном бросании $p = 1/2$. Таким образом,

$$P(23 \text{ орла в } 36 \text{ бросаниях}) = b_{36, 1/2}(23) = \quad (10.14)$$

$$= \frac{36!}{23!13!} \left(\frac{1}{2}\right)^{36}, \quad (10.15)$$

что после довольно утомительных вычислений¹⁾ оказывается равным

$$P(23 \text{ орла}) = 3,36 \, \%.$$

С другой стороны, поскольку среднее распределения равно $np = 18$, а стандартное отклонение $\sigma = \sqrt{np(1-p)} = 3$, мы можем аппроксимировать (10.14) функцией Гаусса $f_{18,3}(23)$ и после тривиальных вычислений получить

$$P(23 \text{ орла}) \approx f_{18,3}(23) = 3,32 \, \%.$$

Практически для всех случаев это — отличное приближение.

Польза от гауссовой аппроксимации еще более очевидна, если надо вычислить вероятности нескольких исходов. На-

¹⁾ Некоторые карманные калькуляторы позволяют вычислять функцию $n!$ автоматически, и с помощью таких калькуляторов расчет (10.15) прост. Однако большинство таких калькуляторов может выполнять операцию $n!$ только при $n < 70$, а при $n \geq 70$ происходит переполнение, и этой функцией нельзя воспользоваться.

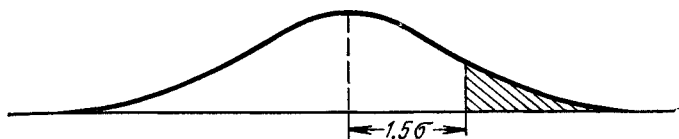


Рис. 10.4. Вероятность получения результата, более чем на $1,5\sigma$ превышающего среднее, равна заштрихованной площади под кривой Гаусса.

пример, вероятность получения 23 или более орлов в 36 бросаниях есть

$$P(23 \text{ или более орлов}) = P(23 \text{ орла}) + P(24 \text{ орла}) + \dots + P(36 \text{ орлов}),$$

сумма, которую чрезвычайно утомительно рассчитать. Однако если аппроксимировать биномиальное распределение гауссовым, то эту вероятность найти легко. Так как в случае распределения Гаусса v — непрерывная переменная, то вероятность $v = 23, 24, \dots$ рассчитывается лучше всего как $P_{\text{Гаусс}}(v \geq 22,5)$, т. е. как вероятность любого $v \geq 22,5$. В случае распределения Гаусса $v = 22,5$ и на $1,5$ стандартных отклонений превышает среднее значение 18. (Помните, что $\sigma = 3$, так что $4,5 = 1,5\sigma$.) Вероятность значений, на $1,5\sigma$ и более превышающих среднее значение, показана как заштрихованная площадь под гауссовой кривой на рис. 10.4. Она легко рассчитывается с помощью таблиц приложения Б, и мы находим

$$P(23 \text{ или более орлов}) \approx P_{\text{Гаусс}}(v \geq X + 1,5\sigma) = 6,7 \, \%.$$

Эта цифра хорошо согласуется с точным результатом $6,6 \, \%$ (представленным двумя значащими цифрами).

10.5. Распределение Гаусса случайных ошибок

В гл. 5 утверждалось, что если результаты измерений подвержены множеству небольших случайных ошибок, то они будут распределены нормально. Сейчас мы в состоянии доказать это утверждение с помощью простой модели для измерений такого типа.

Предположим, что мы измеряем величину x , истинное значение которой равно X . Допустим, что в случае наших измерений систематические ошибки пренебрежимо малы и что имеется n независимых источников случайных ошибок (эффекты параллакса, времени реакции и т. д.). Чтобы упростить наше доказательство, предположим еще, что все эти источники характеризуются случайными ошибками одной и той же

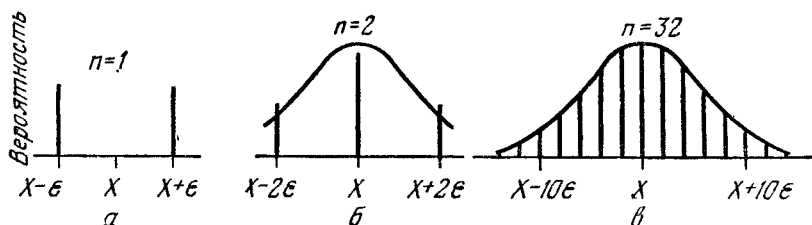


Рис. 10.5. Распределения результатов измерений, подверженных влиянию n случайных ошибок величиной ε для $n = 1, 2$ и 32 .

Непрерывные кривые на графиках б и в — гауссовы функции с теми же центрами и шириной. (Вертикальный масштаб различен для трех графиков.)

фиксированной величины ε , т. е. каждый источник ошибок увеличивает или уменьшает результат на ε с равной вероятностью $p = 1/2$. Например, если истинное значение есть X и имеется только один источник ошибок, то с равной вероятностью возможны результаты $x = X - \varepsilon$ и $x = X + \varepsilon$. Если имеются два источника ошибок, то результат измерения мог бы быть $x = X - 2\varepsilon$ (если окажется, что обе ошибки отрицательны), или $x = X$ (если одна ошибка отрицательна, а другая положительна), или $x = X + 2\varepsilon$ (если окажется, что обе ошибки положительны). Эти возможности проиллюстрированы на рис. 10.5, а и б.

В общем случае если имеется n источников ошибок, наш результат мог бы изменяться от $x = X - n\varepsilon$ до $x = X + n\varepsilon$. Для данного измерения, если окажется, что v источников дают положительные ошибки, а $(n - v)$ — отрицательные, результат будет равен

$$\begin{aligned} x &= X + v\varepsilon - (n - v)\varepsilon = \\ &= X + (2v - n)\varepsilon. \end{aligned} \quad (10.16)$$

Вероятность того, что это случится, равна биномиальной вероятности

$$P(v \text{ положительных ошибок}) = b_{n, 1/2}(v). \quad (10.17)$$

Таким образом, возможные результаты наших измерений будут распределены симметрично около истинного значения X с вероятностями, даваемыми биномиальной функцией (10.17). Все сказанное проиллюстрировано на рис. 10.5 для $n = 1, 2$ и 32 .

Теперь мы утверждаем, что если число источников ошибок n велико и если индивидуальные ошибки ε малы, то результаты измерений будут распределены нормально. Если быть более точным, то заметим, что стандартное отклонение биномиального распределения есть $\sigma_v = \sqrt{np(1 - p)} = \sqrt{n/4}$. Следовательно, в соответствии с (10.16) стандартное отклонение

результатов измерений x будет равно $\sigma_x = 2\epsilon\sigma_v = \epsilon\sqrt{n}$. Устремим $n \rightarrow \infty$ и $\epsilon \rightarrow 0$ таким образом, чтобы величина $\sigma_x = \epsilon\sqrt{n}$ оставалась постоянной. Тогда произойдет следующее. Во-первых, как уже обсуждалось в предыдущем разделе, биномиальное распределение будет стремиться к распределению Гаусса с центром X и с шириной σ_x . Это отчетливо заметно на рис. 10.5, б и в, где приведены соответствующие функции Гаусса. Во-вторых, если $\epsilon \rightarrow 0$, то возможные результаты измерений становятся ближе друг к другу (что также ясно видно на рис. 10.5), так что распределение дискретной величины переходит в непрерывное распределение, которое представляет собой ожидаемое распределение Гаусса.

10.6. Применения. Испытание гипотез

Поскольку нам известно, как должны быть распределены результаты измерений, мы можем спросить: а действительно ли результаты эксперимента *распределены* так, как ожидалось? Такой вид проверки распределений — важный элемент в физических науках, но, возможно, он еще более важен в биологической и социальных науках. Один важный и общий тест — критерий χ^2 — обсуждается в гл. 12. Сейчас мы рассмотрим два примера более простого критерия, который можно использовать при решении некоторых задач, связанных с биномиальным распределением.

Испытание новой лыжной мази

Предположим, что производитель лыжных мазей заявил, что он изобрел новую мазь, которая существенно уменьшает трение между лыжами и снегом. Чтобы проверить это заявление, мы могли бы взять десять пар лыж и смазать одну лыжу из каждой пары этой мазью. Затем мы могли бы спускаться по подходящему снежному склону каждую пару лыж и смотреть, какая из них в каждой паре, смазанная или несмазанная, скатится быстрее.

Если бы смазанная лыжа скатывалась быстрее во всех десяти случаях, то у нас было бы, очевидно, убедительное доказательство того, что мазь эффективна. К сожалению, такие однозначные результаты весьма редки, но даже и в этом случае мы бы предпочли иметь количественную меру убедительности доказательства. Таким образом, мы должны рассмотреть два вопроса. Во-первых, как можно определить количественно меру эффективности мази (или неэффективности)? Во-вторых, где должна быть граница? Если бы смазанные лыжи

скатились быстрее в девяти случаях, было бы это убедительно? А что было бы, если бы в восьми? Или в семи?

Точно такие же вопросы возникают во многих подобных статистических испытаниях. Если бы мы захотели проверить эффективность какого-то удобрения, мы могли бы сравнить какие-то параметры растений, под которые вносились и не вносились удобрения. Чтобы предсказать, кто из кандидатов может победить на выборах, мы могли бы взять случайную выборку избирателей и проверить шансы кандидатов на примере этой выборки.

Чтобы ответить на поставленные вопросы, нам необходимо решить более точно, что же мы должны ожидать от наших испытаний. Или, употребляя принятую терминологию, мы должны сформулировать *статистическую гипотезу*. В примере с лыжной мазью простейшая гипотеза — это *нулевая гипотеза*, согласно которой новая мазь в действительности не оказывает никакого эффекта. В рамках этой гипотезы мы можем вычислить вероятности различных возможных исходов нашего испытания и затем судить о правдоподобности какого-то частного результата.

Предположим, что мы приняли гипотезу, согласно которой лыжная мазь не оказывает никакого эффекта. Тогда в любом испытании смазанная и несмазанная лыжи с равной вероятностью могут скатиться одна быстрее другой, т. е. вероятность того, что смазанная лыжа скатится быстрее, равна $p = 1/2$. Вероятность того, что смазанные лыжи скатятся быстрее в v испытаниях из десяти, определяется биномиальным распределением

$$\begin{aligned} P(v \text{ выигрышей в } 10 \text{ испытаниях}) &= b_{10, 1/2}(v) = \\ &= \frac{10!}{v!(10-v)!} \left(\frac{1}{2}\right)^{10}. \end{aligned} \quad (10.18)$$

В соответствии с (10.18) вероятность того, что смазанные лыжи скатятся быстрее во всех десяти случаях, равна

$$P(10 \text{ выигрышей в } 10 \text{ испытаниях}) = \left(\frac{1}{2}\right)^{10} \approx 0,1 \%, \quad (10.19)$$

т. е. если наша нулевая гипотеза верна, то *очень* невероятно, чтобы смазанные лыжи скатились быстрее во всех десяти случаях. И наоборот, если смазанные лыжи *действительно* скатились быстрее во всех десяти испытаниях, то очень невероятно, чтобы нулевая гипотеза была верна. Действительно, вероятность (10.19) так мала, что мы могли бы сказать, что доказательство эффективности мази «высокозначимо», как мы вскоре будем говорить.

С другой стороны, предположим, что смазанные лыжи скатились быстрее в восьми случаях из десяти испытаний. Тогда мы могли бы вычислить вероятность восьми или более выигрышей:

$$P(8 \text{ или более выигрышей в } 10 \text{ испытаниях}) = \\ = P(8 \text{ выигрышей}) + P(9 \text{ выигрышей}) + \\ + P(10 \text{ выигрышей}) \approx 5,5 \%. \quad (10.20)$$

То, что смазанные лыжи скатятся быстрее в восьми или более испытаниях, все еще невероятно, но не до такой степени невероятно, как в случае всех десяти выигрышей.

Чтобы решить, какой вывод можно сделать из случая восьми выигрышей, мы должны осознать, что в действительности имеются только две альтернативы: либо

а) нулевая гипотеза верна (мазь не оказывает никакого эффекта), но по воле случая произошло редкое событие (смазанные лыжи скатились быстрее в восьми испытаниях);

б) нулевая гипотеза неверна, и мазь эффективна.

В статистических испытаниях обычно выбирают некоторую определенную вероятность (например, 5 %) и рассматривают ее как определяющую границу, ниже которой вероятность события считается неприемлемо низкой. Если вероятность реального исхода (в нашем случае восемь или более выигрышей) ниже этой границы, то мы выбираем альтернативу «б», отвергаем гипотезу и говорим, что результаты эксперимента *значимы*.

Обычно говорят, что результат значим, если соответствующая вероятность не превышает 5%, и «высокозначим», если соответствующая вероятность не превышает 1%. Так как вероятность (10.20) равна 5,5 %, то восьми выигрышей в случае смазанных лыж *недостаточно*, чтобы обеспечить «значимое» доказательство того, что мазь эффективна. С другой стороны, мы видели, что вероятность десяти выигрышей составляет 0,1%. Поскольку это меньше 1%, то мы можем сказать, что десять выигрышей представляли бы «высокозначимое» доказательство эффективности мази¹⁾.

¹⁾ Вероятно, следует подчеркнуть исключительную простоту описанного выше испытания. Мы могли бы измерять различные дополнительные параметры, такие, как время, за которое скатывается каждая лыжа, максимальная скорость каждой лыжи и т. д. Вместо этого мы просто определяли, какая лыжа скатывалась раньше. Испытания, которые не включают такие дополнительные параметры, называются *непараметрическими* испытаниями. Такие испытания обладают большими преимуществами простоты и широкой применимости.

Общая процедура

Методы рассмотренного выше примера можно применить к любой системе n подобных и независимых испытаний, каждое из которых может иметь два возможных исхода: «выигрыш» или «проигрыш». Сначала формулируется гипотеза, которая в данном случае состоит просто в том, что выбирается вероятность p выигрыша в любом отдельном испытании. Это принятое значение p определяет среднее ожидаемое число выигрышей $\bar{v} = np$ в n испытаниях¹⁾. Если действительное число выигрышей v в n испытаниях близко к np , то против гипотезы нет возражений. (Если смазанные лыжи выигрывают пять из десяти испытаний, то нет свидетельства того, что мазь оказывает какой-то эффект.) Если v заметно больше, чем np , то мы рассчитываем вероятность (в рамках данной гипотезы) получения v или более выигрышей. Если эта вероятность меньше, чем наш выбранный «уровень значимости» (т. е. 5 или 1%), то мы заключаем, что наблюдаемое нами число неприемлемо невероятно (если наша гипотеза верна) и, следовательно, что нашу гипотезу следует отвергнуть. Таким же образом, если полученное нами число выигрышей v заметно меньше, чем np , мы можем сделать аналогичное заключение, только нам пришлось бы рассчитать вероятность получения v или менее выигрышей²⁾.

Как и следовало ожидать, описанная процедура не дает простого ответа, что наша гипотеза безусловно верна или безусловно неверна. В действительности она дает количественную меру достоверности наших результатов в рамках гипотезы. Таким образом, мы можем выбрать объективный, хотя и неоднозначный, критерий, по которому мы отвергаем гипотезу. Когда экспериментатор делает вывод, основанный на подобных рассуждениях, то важно, чтобы он четко указал, какой критерий использовался и чему была равна вычисленная вероятность, чтобы читатель сам мог судить о достоверности выводов.

Опрос избирателей

В качестве второго примера рассмотрим выборы, когда баллотируются два кандидата A и B . Пусть кандидат A за-

¹⁾ Как обычно, $\bar{v} = np$ есть среднее число успехов, ожидаемое в том случае, если бы мы были в состоянии повторить всю систему n испытаний много раз.

²⁾ Как мы увидим ниже, в некоторых экспериментах соответствующая вероятность — сумма двух «хвостов», т. е. вероятность получения значений v , которые отличаются в *любую сторону* от np так же, как фактически полученное значение, или более.

явил, что, согласно тщательному обследованию, в его пользу высказываются 60% избирателей, и предположим, что кандидат B попросил нас проверить это заявление (в надежде, конечно, показать, что число избирателей, поддерживающих A , значительно меньше 60%).

В данном случае наша статистическая гипотеза состоит в том, что 60% избирателей поддерживают A , так что вероятность того, что случайно выбранный избиратель будет поддерживать A , равна $p = 0,6$. Сознывая, что мы не можем опросить всех избирателей, мы тщательно подбираем случайную выборку из 600 человек и спрашиваем об их симпатиях. Если действительно 60% поддерживают A , то ожидаемое число людей, поддерживающих A , в нашей выборке равно $np = 600 \cdot 0,6 = 360$. Если фактически только 330 предпочтут A , то можем ли мы заявить, что гипотеза о 60%-ной поддержке A на значимом уровне сомнительна?

Чтобы ответить на этот вопрос, заметим, что (в соответствии с гипотезой) вероятность того, что v избирателей высказутся в поддержку A , определяется биномиальным распределением

$$P(v \text{ избирателей за } A) = b_{n,p}(v), \quad (10.21)$$

где $n = 600$ и $p = 0,6$. Так как n очень велико, то замена биномиальной функции распределением Гаусса с центром $np = 360$ и стандартным отклонением $\sigma_v = \sqrt{np(1-p)} = 12$ будет отличным приближением:

$$P(v \text{ избирателей за } A) \approx f_{360,12}(v). \quad (10.22)$$

Среднее ожидаемое число избирателей, отдающих предпочтение A , равно 360. Таким образом, число тех, кто действительно поддерживает A по нашей выборке (а именно 330), на 30 меньше ожидаемого. Так как стандартное отклонение равно 12, то полученный результат на 2,5 стандартных отклонений меньше предполагаемого среднего. Вероятность такого или меньшего значения (согласно таблице приложения Б) равна 0,6%¹⁾. Таким образом, наш результат «высокозначим» и на 1%-ном уровне значимости мы можем отвергнуть гипотезу, согласно которой 60% избирателей поддерживают A .

Этот пример иллюстрирует две общие особенности испытаний такого рода. Во-первых, обнаружив, что 330 избирателей поддерживают A (на 30 меньше, чем ожидалось), мы вычислили вероятность того, что число лиц, поддерживающих

¹⁾ Строго говоря, мы должны были бы рассчитать вероятность для $v \leq 330,5$, так как в случае распределения Гаусса v — непрерывная переменная. Это значение на 2,46σ меньше среднего, поэтому правильное значение вероятности равно в действительности 0,7%; но столь малое отличие никак не влияет на наш вывод.



Рис. 10.6. *а* — «однохвостовая» вероятность получения результата, который более чем на 30 меньше среднего. *б* — «двуххвостовая» вероятность получения результата, который отличался бы на 30 и более в любую сторону. (Не в масштабе.)

А, будет 330 или меньше. На первый взгляд можно было бы рассматривать вероятность того, что число лиц, поддерживающих *А*, равно $v = 330$. Однако эта вероятность чрезвычайно мала (в действительности 0,15%), и даже наиболее вероятный результат ($v = 360$) имеет невысокую вероятность (3,3%). Чтобы получить надлежащую меру того, насколько неожиданным является результат $v = 330$, мы должны рассмотреть $v = 330$ и любой результат, который еще меньше среднего.

Наш результат $v = 330$ на 30 меньше, чем ожидаемый 360. Вероятность результата, который на 30 или более меньше среднего, иногда называют «однохвостовой вероятностью», поскольку она определяется площадью под одним хвостом кривой распределения, как показано на рис. 10.6, *а*. В некоторых испытаниях соответствующая вероятность есть «двуххвостовая вероятность» получения результата, который отличается от ожидаемого среднего на 30 или более в *любую сторону*, т. е. вероятность получения $v \leq 330$ и $v \geq 390$, как показано на рис. 10.6, *б*. Вопрос о том, использовать ли в статистическом тесте однохвостовую или двуххвостовую вероятность, зависит от того, что считается альтернативой исходной гипотезе. В данном случае мы были заинтересованы в том, чтобы показать, что кандидат *А* пользуется поддержкой *меньшей*, чем объявленные 60%, так что однохвостовая вероятность соответствовала сути дела. Если бы мы были заинтересованы в том, чтобы показать, что число лиц, поддерживающих *А*, *отличается* от 60% (в любую сторону), то двуххвостовая вероятность соответствовала бы сути дела. На практике обычно довольно ясно, какую вероятность надо использовать. В любом случае экспериментатор всегда должен четко указать, какая вероятность и какой уровень значимости были выбраны и чему равна вычисленная вероятность. Располагая этой информацией, читатель может судить о значимости результатов самостоятельно.

Задачи

Напоминание: звездочка у номера задачи означает, что задача решается или ее ответ приводится в разделе «Ответы» в конце книги.

10.1 (разд. 10.2). Рассмотрите эксперимент из разд. 10.2, в котором бросаются три игральные кости. Получите вероятности для выпадения случая, когда нет ни одной единички, и для выпадения одной единички. Проверьте все четыре вероятности, приведенные на рис. 10.1.

***10.2** (разд. 10.2).

а. Рассчитайте вероятности P (v единичек в двух бросаниях) для всех возможных v при бросании двух костей. Нанесите их на гистограмме.

б. Сделайте то же самое для случая бросания четырех костей.

10.3 (разд. 10.3).

а. Вычислите $5!, 6!, 25!/23!$

б. Используйте соотношение $n! = (n+1)!/(n+1)$, чтобы показать, что $0!$ должен быть определен как 1.

в. Докажите, что биномиальные коэффициенты, определяемые формулой (10.3), равны

$$\binom{n}{v} = \frac{n!}{v!(n-v)!}.$$

***10.4** (разд. 10.3). Рассчитайте биномиальные коэффициенты $\binom{3}{v}$ для

$v = 0, 1, 2, 3$ и $\binom{4}{v}$ для $v = 0, \dots, 4$. Потом запишите биномиальное разложение (10.5) выражения $(p+q)^n$ для $n = 3$ и 4 .

10.5 (разд. 10.3).

а. Рассчитайте и начертите гистограмму биномиальной функции распределения $b_{n,p}(v)$ для $n = 4$, $p = 1/2$ и всех возможных v .

б. Повторите задание «а» для $n = 4$ и $p = 1/5$.

***10.6** (разд. 10.3). Больница принимает четырех пациентов, страдающих от болезни, смертность от которой составляет 80 %. Используйте результаты задачи 10.5, б и определите вероятности следующих исходов:

а) ни один из пациентов не выживет;

б) выживет только один;

в) выживут двое или более.

***10.7** (разд. 10.3). Определите вероятности получения v единичек при бросании пяти костей для $v = 0, 1, \dots, 5$.

10.8 (разд. 10.4). Докажите, что среднее число выигрышей

$$\bar{v} = \sum_{v=0}^n v b_{n,p}(v)$$

для биномиального распределения равно np .

Существует много способов доказать это, и один из лучших состоит в следующем: запишите биномиальное разложение (10.5) для $(p+q)^n$. Поскольку оно верно для любых p и q , вы можете продифференцировать его по p . Если вы затем положите $p+q=1$ и умножите почленно на p , то получите требуемый результат.

***10.9** (разд. 10.4). Стандартное отклонение для любого распределения $f(v)$ определяется как

$$\sigma_v^2 = \overline{(v - \bar{v})^2}.$$

Докажите, что это то же самое, что и $\bar{v}^2 - (\bar{v})^2$.

***10.10** (разд. 10.4). Используйте результаты задачи 10.9 и докажите, что для биномиального распределения $b_{n,p}(v)$

$$\sigma_v^2 = np(1-p).$$

(Используйте тот же прием, что и в задаче 10.8, но продифференцируйте по p дважды.)

10.11 (разд. 10.4). Докажите, что в случае $p = 1/2$ для биномиального распределения верно соотношение

$$b_{n, 1/2}(v) = b_{n, 1/2}(n - v),$$

т. е. что распределение симметрично относительно $v = n/2$.

10.12 (разд. 10.4). Гауссова аппроксимация (10.12) биномиального распределения превосходна для больших n и, что удивительно, хороша для малых n (особенно если p близко к $1/2$). Чтобы показать это, вычислите $b_{n, 1/2}(v)$ (для $v = 0, 1, \dots, 4$) по точным формулам и с использованием гауссовой аппроксимации. Сравните ваши результаты.

***10.13** (разд. 10.4). Используйте гауссову аппроксимацию для определения вероятности выпадения точно 15 орлов в случае, когда вы бросаете монету 25 раз. Вычислите точное значение той же вероятности и сравните результаты.

***10.14** (разд. 10.4). Используйте гауссову аппроксимацию и определите вероятность выпадения 18 или более орлов при 25 бросаниях монеты. (При использовании распределения Гаусса вы должны определить вероятность для $v \geq 17.5$.) Сравните с точным результатом 2,16%.

10.15 (разд. 10.6). В испытаниях лыжной мази, описанных в разд. 10.6, предположим, что смазанные лыжи скатились быстрее в девяти из десяти испытаний. Предполагая, что мазь не оказывает никакого эффекта, рассчитайте вероятность девяти или более выигрышей. Обеспечивают ли девять выигрышей «значимое» доказательство эффективности мази (на уровне 5%)? Является ли доказательство «высокозначимым» (на 1%-ном уровне)?

***10.16** (разд. 10.6). Чтобы испытать новое удобрение, огородник выбирает 14 пар одинаковых растений и вносит удобрение под одно растение каждой пары. Спустя два месяца 12 обработанных растений выглядят более развитыми, чем их необработанные партнеры (два оставшихся выглядят менее развитыми). Если фактически удобрение не оказывает никакого эффекта, какова будет вероятность, что просто случайно у огородника получится 12 и более выигрышей? Представляют ли 12 выигрышей значимое доказательство того, что удобрение полезно (на 5%-ном уровне)? Является ли доказательство «высокозначимым» (на 1%-ном уровне)?

10.17 (разд. 20.6). Известно, что всхожесть семян определенного сорта равна 25%. Чтобы испытать новый «стимулятор всхожести», 100 таких семян высаживаются и обрабатываются стимулятором. Если 32 из них всходят, то можно ли заключить (на 5%-ном уровне значимости), что стимулятор действует?

***10.18** (разд. 10.6). В некоторой школе 420 из 600 учеников выдержали испытания на стандартный математический тест, который в целом по стране выдерживают 60% учеников. Если студенты не имеют никакого специального отношения к этому тесту, то сколько из них, по вашим ожиданиям, выдержат это испытание и какова вероятность, что его выдержат 420 или более человек? Может ли школа утверждать, что ее ученики значительно лучше подготовлены к испытанию?

Глава 11

Распределение Пуассона

В этой главе мы будем изучать наш третий пример предельного распределения, а именно распределение Пуассона. Оно описывает результаты экспериментов, в которых считают события, происходящие случайно, но в определенном среднем темпе. Это распределение особенно важно в атомной и ядерной физике, где подсчитывают числа распадов нестабильных атомов и ядер.

11.1. Определение распределения Пуассона

В качестве примера распределения Пуассона рассмотрим случай, когда мы имеем дело с образцом радиоактивного материала. При помощи счетчика Гейгера мы можем сосчитать число ν электронов, испущенных в радиоактивных распадах за одну минуту. Если счетчик исправен, то в значении ν не будет погрешности. Тем не менее если мы будем повторять эксперимент, то обязательно получим разные значения ν . Эта вариация в значениях ν не является погрешностью непосредственно в подсчете, а скорее отражает характерные свойства процесса радиоактивного распада.

Каждое радиоактивное ядро характеризуется определенной вероятностью распада за любой одномоментный интервал. Если бы мы знали эту вероятность и число ядер в нашем образце, то могли бы рассчитать *ожидаемое среднее число распадов* за минуту. Тем не менее каждое ядро распадается в случайный момент времени, и число распадов за любую одну минуту может отличаться от ожидаемого среднего числа.

Очевидно, вопрос, который нам следует задать, состоит в следующем: если бы мы повторяли наш эксперимент много раз (восполняя образец, если он существенно истощается), то какое распределение для числа распадов ν , наблюдаемых за одномоментные интервалы, мы должны были бы получить? Если вы изучили гл. 10, то поймете, что искомое распределение является биномиальным. Если имеется n ядер и вероят-

ность того, что любое одно ядро распадется, равна p , то вероятность v распадов — это просто вероятность v «успехов» в n «испытаниях», или $b_{n,p}(v)$. Однако в эксперименте такого рода, который мы сейчас рассматриваем, имеются особенности, позволяющие сделать важное упрощение. Число «испытаний» (т. е. ядер) огромно (вероятно, $n \sim 10^{20}$), и вероятность «успеха» (т. е. распада) для любого одного ядра ничтожна (часто $p \sim 10^{-20}$). При этих условиях (n велико, p мало) можно показать, что биномиальное распределение неотличимо от более простой функции, называемой распределением Пуассона, т. е. что

$$P(v \text{ отсчетов за любой определенный интервал}) = p_\mu(v), \quad (11.1)$$

где *распределение Пуассона* $p_\mu(v)$ определяется как

$$p_\mu(v) = e^{-\mu} \frac{\mu^v}{v!}. \quad (11.2)$$

В этом определении μ — положительный параметр ($\mu > 0$), который, как мы скоро увидим, представляет собой ожидаемое среднее число отсчетов за рассматриваемый интервал времени, а $v!$ — обычная факториальная функция (и $0! = 1$).

Мы не будем сейчас выводить распределение Пуассона (11.2), но просто примем, что оно является соответствующим распределением для эксперимента рассматриваемого типа¹⁾. Чтобы установить роль параметра μ в (11.2), мы должны вычислить среднее число ожидаемых отсчетов $\bar{v}$ в случае, если бы мы повторяли наш эксперимент с подсчетами много раз. Это среднее число есть

$$\bar{v} = \sum_{v=0}^{\infty} v p_\mu(v) = \sum_{v=0}^{\infty} v e^{-\mu} \frac{\mu^v}{v!}. \quad (11.3)$$

Первый член в этой сумме может быть опущен (так как он равен нулю), а $v/v!$ можно заменить на $1/(v-1)!$. Если вынести за знак суммы общий множитель $\mu e^{-\mu}$, то получим

$$\bar{v} = \mu e^{-\mu} \sum_{v=1}^{\infty} \frac{\mu^{v-1}}{(v-1)!}. \quad (11.4)$$

Остающаяся бесконечная сумма

$$1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \dots = e^\mu \quad (11.5)$$

¹⁾ Вывод см., например, в книге *Young H. D. Statistical Treatment of Experimental Data*, McGraw-Hill, 1962, sect. 8 или *Meyer S. L., Data Analysis for Scientists and Engineers*, John Wiley, 1975, p. 207.

есть просто экспоненциальная функция e^μ (как указано). Таким образом, экспонента $e^{-\mu}$ в (11.4) сокращается с этой суммой, и мы приходим к простому заключению, что

$$\bar{v} = \mu. \quad (11.6)$$

Таким образом, параметр μ , который характеризует распределение Пуассона $p_\mu(v)$, — это просто *среднее число отсчетов, которое мы ожидаем в случае многократного повторения счетного эксперимента*.

11.2. Свойства распределения Пуассона

На рис. 11.1 приведены распределения Пуассона для случаев $\mu = 0,8$ и 3. Из рис. 11.1, а для $\mu = 0,8$ видно, что наиболее вероятные числа отсчетов равны $v=0$ или 1 (причем $v=0$ несколько более вероятно) и что имеется заметная вероятность получения $v=2$ или 3. Из рис. 11.1, б для $\mu = 3$ видно, что наиболее вероятные отсчеты — это 2 и 3 и с заметной вероятностью встречаются отсчеты в интервале от $v=0$ до $v=7$. На обоих рисунках распределения заметно асимметричны.

Если рассмотреть эксперимент с большим средним значением отсчетов, например с $\mu = 9$, как показано на рис. 11.2, то мы увидим, что распределение будет приблизительно симметричным относительно среднего. Действительно, можно доказать, что при $\mu \rightarrow \infty$ распределение Пуассона становится все более симметричным и стремится к распределению Гаусса с тем же средним и стандартным отклонением¹⁾. На рис. 11.2

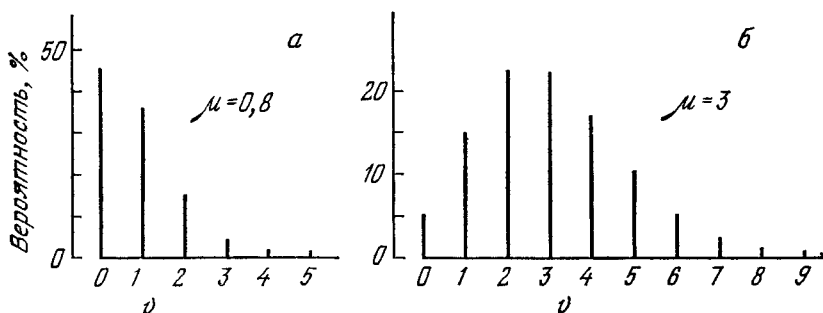


Рис. 11.1. Распределения Пуассона со средними числами отсчетов $\mu = 0,8$ и 3.

¹⁾ См. ссылку на работу Мейера на с. 214

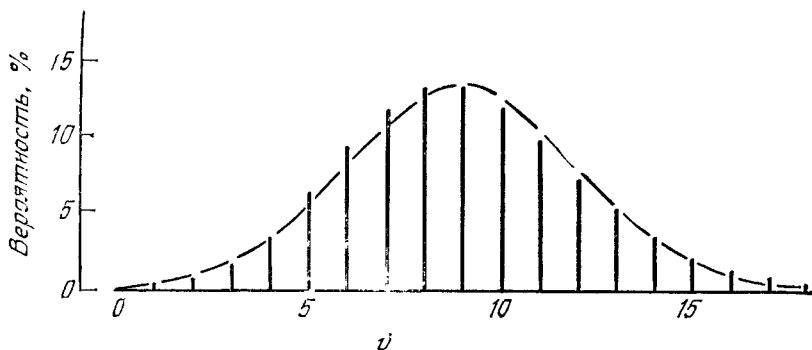


Рис. 11.2. Распределение Пуассона с $\mu = 9$. Прерывистая линия есть функция Гаусса с тем же центром и стандартным отклонением.

пунктирной кривой представлено распределение Гаусса с центром при $\mu = 9$ и с тем же стандартным отклонением, что и у распределения Пуассона. Можно видеть, что, даже когда μ равно только 9, распределение Пуассона очень близко к соответствующей функции Гаусса, а небольшое отличие отражает остающуюся асимметрию в функции Пуассона. Как мы вскоре увидим, для практики очень удобно, что в случае больших μ распределение Пуассона можно аппроксимировать соответствующим гауссовым.

Другое интересное свойство распределения Пуассона обнаруживается, если вычислить его стандартное отклонение σ_v . Как мы видели в гл. 4, σ_v^2 — это среднее квадратов отклонений $(v - \bar{v})^2$. Таким образом,

$$\sigma_v^2 = \overline{(v - \bar{v})^2}$$

или (используя результаты задачи 10.9)

$$\sigma_v^2 = \bar{v}^2 - (\bar{v})^2. \quad (11.7)$$

Мы уже вычислили, что $\bar{v} = \mu$, и аналогичные расчеты дают $\bar{v}^2 = \mu^2 + \mu$ (см. задачу 11.6). Таким образом, $\sigma_v^2 = \mu$, или

$$\sigma_v = \sqrt{\mu}. \quad (11.8)$$

Распределение Пуассона со средним числом отсчетов μ имеет стандартное отклонение $\sqrt{\mu}$.

Результат (11.8) чрезвычайно полезен на практике. Если мы провели один счетный эксперимент и получили в итоге v отсчетов, то, как легко видеть (используя принцип максимального правдоподобия, как, например, в задаче 11.9), наилуч-

шая оценка для ожидаемого среднего числа отсчетов μ будет равна $\mu_{\text{наил}} = v$. Из (11.8) немедленно следует, что наилучшая оценка для стандартного отклонения будет равна $\sqrt{v}$. Другими словами, если мы проделаем одно измерение числа отсчетов за некоторый временной интервал и получим в результате число v , то наш итоговый вывод для ожидаемого среднего числа отсчетов за этот же временной интервал будет

$$v \pm \sqrt{v}. \quad (11.9)$$

Это и есть результат, приводившийся без доказательства в соотношении (3.2). Если бы нам пришлось вести измерения в течение более длительного времени, то мы получили бы большее значение v . В соответствии с (11.9) это означало бы большую погрешность $\sqrt{v}$. Однако относительная погрешность, которая определяется как

$$\text{относительная погрешность} = \frac{\sqrt{v}}{v} = \frac{1}{\sqrt{v}},$$

уменьшалась бы, если бы счет продолжался более длительное время.

Интересно сравнить распределения Пуассона и Гаусса. В-первых, распределение Гаусса $f_{x, \sigma}(x)$ является *непрерывным*, так как x — непрерывная переменная, а распределение Пуассона $p_{\mu}(v)$ *дискретно* (подобно биномиальному), поскольку $v = 0, 1, 2, \dots$. Во-вторых, гауссово распределение $f_{x, \sigma}(x)$ определяется *двумя* параметрами: средним X и шириной σ , а распределение Пуассона $p_{\mu}(v)$ определяется единственным параметром μ , поскольку, как мы только что видели, ширина σ_v распределения Пуассона автоматически определяется средним μ (а именно $\sigma_v = \sqrt{\mu}$). Наконец, если мы рассмотрим распределение Пуассона, для которого среднее число отсчетов μ велико, то дискретная природа v становится менее существенной и, как уже обсуждалось в связи с рис. 11.2, распределение Пуассона (подобно биномиальному) хорошо аппроксимируется функцией Гаусса $f_{x, \sigma}(x)$ с теми же средним и шириной, т. е.

$$p_{\mu}(v) \approx f_{x, \sigma}(v) \quad (\mu \text{ велико}), \quad (11.10)$$

где

$$X = \mu \quad \text{и} \quad \sigma = \sqrt{\mu}.$$

Приближение (11.10) называется гауссовой аппроксимацией распределения Пуассона. Оно аналогично соответствующей аппроксимации биномиального распределения (рассмотренной в разд. 10.4) и полезно при выполнении аналогичных условий, а именно когда рассматриваемые параметры велики. Чтобы проиллюстрировать это, предположим, что мы хотим

рассчитать распределение Пуассона для $\mu = 64$. Вероятность, например, 72 отсчетов есть

$$P(72 \text{ отсчета}) = p_{64}(72) = e^{-64} \frac{(64)^{72}}{72!}, \quad (11.11)$$

и в результате трудоемких вычислений получаем для нее численное значение

$$P(72 \text{ отсчета}) = 2,9\%.$$

Однако в соответствии с (11.10) вероятность (11.11) хорошо представляется как

$$P(72 \text{ отсчета}) \approx f_{64,8}(72),$$

откуда после простых вычислений имеем

$$P(72 \text{ отсчета}) \approx 3,0\%.$$

Если бы мы захотели вычислить непосредственно вероятность 72 или более отсчетов в том же эксперименте, то чрезвычайно громоздкий расчет дал бы

$$\begin{aligned} P(v \geq 72) &= p_{64}(72) + p_{64}(73) + \dots = \\ &= 17,3\%. \end{aligned}$$

Если бы мы использовали аппроксимацию (11.10), то должны были бы рассчитать только вероятность получения $v \geq 71,5$ (так как в случае распределения Гаусса v считается непрерывной переменной). Поскольку 71,5 на 7,5 или на $0,94\sigma$ превышает среднее, то искомая вероятность может быть определена по таблице приложения Б как

$$\begin{aligned} P(v \geq 72) &\approx P_{\text{Гаусс}}(v \geq 71,5) = P_{\text{Гаусс}}(v \geq X + 0,94\sigma) = \\ &= 17,4\%, \end{aligned}$$

что почти по любому критерию является отличным приближением.

11.3. Примеры

Как мы уже подчеркивали, распределение Пуассона описывает распределение результатов в эксперименте, когда ведется счет событий, происходящих случайно, но в определенном ожидаемом среднем темпе. В лаборатории вводного курса физики два наиболее известных примера — это подсчет распадов радиоактивных ядер и подсчет частиц космических лучей.

Другой очень важный пример — эксперимент по изучению ожидаемого предельного распределения, подобного распределению Гаусса, биномиальному распределению или распределению Пуассона. С помощью любого предельного распреде-

ления можно узнать, сколько событий любого частного типа ожидается, если эксперимент повторяется несколько раз. (Например, с помощью функции Гаусса $f_{\chi, a}(x)$ можно узнать, каково ожидаемое число результатов измерений x , которое попадает в произвольный интервал от $x = a$ до $x = b$.) На практике наблюдаемое число редко в точности совпадает с ожидаемым. На самом деле оно флуктуирует в соответствии с распределением Пуассона. В частности, если ожидаемое число событий некоторого типа равно n , то полученное число может отличаться от n на число порядка $\sqrt{n}$.

Во многих ситуациях разумно предполагать, что некоторые числа распределены приблизительно по закону Пуассона. Следует ожидать, что числа яиц, откладываемых домашними птицами на птицеферме за час, числа рождений в день в родильном доме будут следовать распределению Пуассона, по крайней мере приблизительно (хотя, вероятно, они также будут отражать и сезонные изменения). Чтобы проверить это предположение, вы должны регистрировать нужные числа много раз. Отложив на графике полученное распределение, вы могли бы сравнить его с распределением Пуассона и получить некоторое представление о том, насколько хорошо они совпадают. Если желательно применить количественный тест, то вы могли бы использовать критерий χ^2 , описанный в гл. 12.

Подсчет частиц космического излучения

В качестве конкретного примера распределения Пуассона рассмотрим эксперимент с космическими лучами. Эти «лучи» в действительности представляют собой заряженные частицы, такие, как протоны, и α -частицы, которые попадают в атмосферу Земли из космического пространства. Некоторые из них проходят все расстояние до поверхности Земли и могут быть зарегистрированы (например, с помощью счетчика Гейгера) в лаборатории¹⁾. В последующей задаче мы будем использовать тот факт, что число частиц космических лучей, попадающих на любую заданную площадь за данное время, должно быть распределено в соответствии с законом Пуассона.

Пусть студент А утверждает, что он измерил число частиц космических лучей, упавших на счетчик Гейгера за одну минуту. Он заявляет, что проделал измерения многократно и

¹⁾ На самом деле первичные космические лучи, попадающие в атмосферу из космического пространства, взаимодействуют с ядрами атомов воздуха и генерируют вторичные частицы, которые в свою очередь тоже испытывают взаимодействия или распадаются, и т. д. В лаборатории на Земле можно детектировать практически только вторичные частицы (мюоны, электроны). Автор для простоты изложения опускает эти детали, которые несущественны в данном случае. — *Прим. перев.*

тщательно и нашел, что в среднем девять частиц проходят через счетчик за одну минуту, а погрешность «пренебрежимо» мала. Чтобы проверить это утверждение, студент Б подсчитывает, сколько частиц регистрируется за одну минуту, и получает в результате цифру 12. Вызывает ли этот результат серьезные сомнения в правильности утверждения студента А о том, что ожидаемое число отсчетов равно девяти?

Чтобы сделать более тщательную проверку, студентка В считает число упавших частиц за десять минут. По результатам студента А она ожидает получить цифру 90, но в действительности регистрирует 120. Вызывает ли этот результат существенное сомнение в правильности утверждения студента А?

Рассмотрим сначала результат студента Б. Если А прав, то ожидаемое среднее число отсчетов равно 9. Поскольку распределение отсчетов должно следовать закону Пуассона, то стандартное отклонение должно быть равно $\sqrt{9} = 3$. Результат студента Б есть 12, следовательно, он только на одно стандартное отклонение отличается от среднего, равного 9. Это определенно не столь большое отличие, чтобы можно было говорить о противоречии с результатом А. Более строго, зная, что вероятность любого результата v должна быть $p_9(v)$, мы можем рассчитать полную вероятность получения результата, который отличается от 9 на 3 и более. Оказывается, что эта вероятность составляет 40% (см. задачу 11.11). Очевидно, что результат студента Б вовсе не удивителен, и у студента А нет причин для беспокойства.

Совсем другое дело — результат студентки В. Если А прав, то В должна была ожидать 90 отсчетов за 10 минут. Поскольку распределение должно быть пуассоновым, то стандартное отклонение должно составлять $\sqrt{90} = 9,5$. Таким образом, результат В, равный 120, более чем на *три стандартных отклонения* отличается от предсказания студента А, равного 90. В случае таких больших чисел распределение Пуассона неотличимо от функции Гаусса, и мы сразу можем найти по таблице приложения А, что вероятность отсчета, отличающегося более чем на три стандартных отклонения от среднего, равна 0,3%, т. е. если А прав, то чрезвычайно невероятно, чтобы В наблюдала 120 отсчетов. Иначе говоря, мы можем почти наверное утверждать, что где-то допущен просчет. Возможно, А не был столь аккуратен, как он утверждал. Может быть, счетчик плохо работал в случае А или В, что приводило к систематическим ошибкам в одном из результатов. Или, может быть, А делал измерения в момент времени, когда поток космических лучей был действительно меньше нормального.

Задачи

Напоминание: звездочка у номера задачи означает, что задача решается или ее ответ приводится в разделе «Ответы» в конце книги.

*11.1 (разд. 11.1).

а. Рассчитайте распределение Пуассона $p_\mu(v)$ для $\mu = 0,5$ и $v = 0, 1, \dots, 6$ и постройте гистограмму $p_\mu(v)$ для этих дискретных значений v .

б. Выполните задание «а» для $\mu = 1$.

в. Выполните задание «а» для $\mu = 2$.

*11.2 (разд. 11.1).

а. Распределение Пуассона, подобно всем функциям распределения, должно удовлетворять «условию нормировки»

$$\sum_{v=0}^{\infty} p_\mu(v) = 1. \quad (11.12)$$

Согласно этому условию, полная вероятность наблюдения *всех* возможных значений v должна быть равна 1. Докажите это. [Вспомните о бесконечном ряде (11.5) для e^μ .]

б. Продифференцируйте (11.12) по μ , затем умножьте на μ и получите таким образом альтернативное доказательство того, что $\bar{v} = \mu$, как в случае формулы (11.6).

11.3 (разд. 11.1). По данным за 28 дней фермер на птицеферме определил, что между 10.00 и 10.30 утра все его курицы откладывают в среднем 2,5 яйца. Предполагая, что число откладываемых яиц подчиняется распределению Пуассона с $\mu = 2,5$, подсчитайте, сколько за это время было дней, когда между 10.00 и 10.30 утра не оказывалось ни одного отложенного яйца? А сколько дней, когда было 2 яйца *или меньше*? 3 или больше?

*11.4 (разд. 11.1). Некоторый образец радиоактивного вещества содержит $1,5 \cdot 10^{20}$ ядер, каждое из которых с вероятностью $p = 10^{-20}$ может распасться за любую фиксированную минуту.

а. Чему равно ожидаемое среднее число μ распадов в образце за одну минуту?

б. Вычислите вероятность $p_\mu(v)$ наблюдения v распадов в минуту для $v = 0, 1, 2, 3$.

в. Какова вероятность наблюдения четырех или более распадов в одну минуту?

*11.5 (разд. 11.1). Ожидается, что в некотором образце радиоактивного вещества происходят три распада в минуту. Студент наблюдает число v распадов в 100 отдельных одиominутных интервалах и получает результаты, показанные в табл. 11.1.

а. Постройте гистограмму этих результатов, откладывая f_v (долю случаев, когда реализуется результат v) в зависимости от v .

б. На этом же графике постройте ожидаемое распределение $p_3(v)$. Согласуются ли данные с ожидаемым распределением?

11.6 (разд. 11.2).

а. Докажите, что среднее значение $\bar{v}^2$ для распределения Пуассона $p_\mu(v)$ равно $\bar{v}^2 = \mu^2 + \mu$. [Простейший способ показать это со-

Таблица 11.1

Число распадов v	0	1	2	3	4	5	6	7	8	9
Число наблюдений	5	19	23	21	14	12	3	2	1	0

стоит, вероятно, в том, чтобы дважды продифференцировать тождество (11.12) по μ .]

б. Докажите, следовательно, что стандартное отклонение для v равно $\sigma_v = \sqrt{\mu}$. [Используйте тождество (11.7).]

*11.7 (разд. 11.2). Вычислите среднее $\bar{v}$ и стандартное отклонение σ_v данных из задачи 11.5. Сравните полученные вами значения с ожидаемыми результатами 3 и $\sqrt{3}$.

11.8 (разд. 11.2). Известно, что средний темп распада ядер в некотором образце равен примерно 20 в минуту. Если вы хотите измерить этот темп с 4%-ной точностью, то сколько времени вам потребуется для счета?

*11.9 (разд. 11.2).

а. Предположим, что мы считаем число частиц космических лучей, упавших на счетчик за одну минуту, и получаем результат v_0 . Предполагая, что это число подчиняется распределению Пуассона $p_\mu(v)$, где μ — неизвестный ожидаемый средний темп, запишите вероятность получения значения v_0 . Используйте принцип максимального правдоподобия и докажите, что наилучшая оценка для μ равна

$$\mu_{\text{наил}} = v_0.$$

(Помните, что наилучшая оценка для μ — это такое значение, для которого вероятность получения v_0 максимальна.)

б. Предположим, что мы сделали N отдельных измерений $v_1, \dots, v_N$; рассуждая, как и выше, покажите, что в этом случае $\mu_{\text{наил}}$ равно среднему

$$\mu_{\text{наил}} = \frac{1}{N} \sum_{i=1}^N v_i.$$

*11.10 (разд. 11.2). Ожидаемое среднее число отсчетов в некотором эксперименте равно $\mu = 16$.

а. Используйте аппроксимацию Гаусса (11.10) для оценки вероятности получения десяти отсчетов. Сравните полученное значение с точным результатом $p_{16}(10)$.

б. Используйте аппроксимацию Гаусса для оценки вероятности получения десяти или менее отсчетов. [Помните, что надо вычислять $P_{\text{Гаусс}}(v \leq 10,5)$, чтобы учесть тот факт, что в случае распределения Гаусса v является непрерывной переменной. Нужная вероятность может быть вычислена с помощью таблицы приложения Б.] Рассчитайте точный результат и сравните его с приближенным.

Заметьте, как даже в случае таких малых значений μ , когда оно равно только 16, аппроксимация Гаусса дает вполне приемлемые результаты, и по крайней мере для задания «б» требует существенно менее трудоемких расчетов, чем точные вычисления.

*11.11 (разд. 11.3).

а. Вычислите вероятности $p_9(v)$ получения v отсчетов для $v = 7, 8, 9, 10$ и 11 в эксперименте, в котором ожидаемое среднее число отсчетов равно 9.

б. Найдите по этим данным полную вероятность получения отсчета, который отличался бы от среднего 9 на 3 или более. Заставит ли вас отсчет 12 подозревать, что ожидаемое среднее в действительности не равно 9?

Глава 12

Критерий χ^2 для распределений

Теперь у нас уже имеется некоторый опыт обращения с предельными распределениями. Это такие функции, которые описывают ожидаемое распределение результатов в случае, когда эксперимент повторяется большое число раз. Существует много различных предельных распределений, соответствующих множеству различных типов возможных экспериментов. Вероятно, три наиболее важных в физике распределения — как раз те, которые мы уже рассмотрели: гауссово (или нормальное), биномиальное и распределение Пуассона.

В этой последней главе мы будем рассматривать вопрос о том, как решить, подчиняются ли результаты данного эксперимента ожидаемому предельному распределению. Или подробнее: предположим, что мы выполняем эксперимент, для которого, как мы полагаем, нам известно ожидаемое распределение результатов. Предположим далее, что мы повторяем эксперимент несколько раз и регистрируем результаты наблюдений. Вопрос, который мы можем задать теперь, состоит в следующем: как определить, согласуется ли наблюдаемое распределение с ожидаемым теоретическим распределением? Мы увидим, что на этот вопрос можно ответить, используя простую процедуру, называемую *критерием χ^2* .

12.1. Введение в критерий χ^2

Начнем с конкретного примера. Предположим, что мы сделали 40 измерений $x_1, \dots, x_{40}$ длины траектории x пули, вылетающей из некоторого ружья, и получили значения, при-

Таблица 12.1. Измеренные значения x (в сантиметрах)

731	772	771	681	722	688	653	757	733	742
739	780	709	676	760	748	672	687	766	645
678	748	689	810	805	778	764	753	709	675
698	770	754	830	725	710	738	638	787	712

веденные в табл. 12.1. Пусть у нас имеются основания полагать, что результаты этих измерений распределены в соответствии с законом Гаусса $f_{X, \sigma}(x)$, что вполне естественно. В такого рода экспериментах обычно не известны заранее ни центр X , ни ширина σ ожидаемого распределения. Следовательно, нашим первым шагом будет расчет наилучших оценок для этих величин по 40 результатам наших измерений:

$$\begin{aligned} (\text{наилучшая оценка } X) = \bar{x} &= \sum_{i=1}^{40} x_i / 40 = \\ &= 730,1 \text{ см} \end{aligned} \quad (12.1)$$

и

$$\begin{aligned} (\text{наилучшая оценка } \sigma) &= \sqrt{\frac{\sum (x_i - \bar{x})^2}{39}} = \\ &= 46,8 \text{ см.} \end{aligned} \quad (12.2)$$

Теперь мы можем спросить, удовлетворяет ли фактическое распределение наших результатов $x_1, \dots, x_{40}$ гипотезе, что эти результаты распределены в соответствии с законом Гаусса $f_{X, \sigma}(x)$ с оцененными выше значениями X и σ . Чтобы ответить на этот вопрос, мы должны рассчитать, как, согласно нашим ожиданиям, были бы распределены наши 40 результатов, если бы гипотеза была верна, и сравнить это ожидаемое распределение с нашим фактически полученным распределением. Первая трудность заключается в том, что x есть непрерывная переменная, поэтому мы не можем говорить об ожидаемом числе измерений для какого-то одного значения x . Вместо этого мы должны говорить об ожидаемом числе в некотором интервале $a < x < b$, т. е. мы должны поделить весь интервал возможных значений на *бины*. В случае 40 измерений мы могли бы выбрать границы бинов при $X - \sigma$, X и $X + \sigma$, определяя таким образом четыре бина, как показано в табл. 12.2.

Таблица 12.2. Возможный выбор бинов для данных табл. 12.1
Последняя строка показывает число данных, которые попали в соответствующий бин

Номер бина k	1	2	3	4
Значения x в бине	$x < X - \sigma$ или $x < 683,3$	$X - \sigma < x < X$ или $683,3 < x < 730,1$	$X < x < X + \sigma$ или $730,1 < x < 776,9$	$X + \sigma < x$ или $776,9 < x$
Числа наблюдений O_k в бине	8	10	16	6

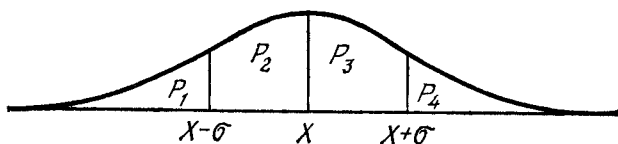


Рис. 12.1. Вероятности $P_1, \dots, P_4$ того, что результаты попадут в каждый из четырех бинов ($k = 1, \dots, 4$), равны соответствующим четырем площадям, показанным под функцией Гаусса.

Позже мы обсудим критерии для выбора бинов. В частности, их следует выбирать таким образом, чтобы все бины содержали по несколько измеренных значений x_i . Обычно мы будем обозначать число бинов через n ; в данном случае, например, с четырьмя бинами, $n = 4$.

Поделив весь интервал возможных измеренных значений на бины, мы можем сформулировать наш вопрос более точно. Во-первых, мы можем сосчитать число результатов измерений, которые попадают в каждый бин k ¹⁾. Обозначим это число через O_k . Для данных нашего примера наблюдаемые числа O_1, O_2, O_3, O_4 показаны в последней строке табл. 12.2. Далее, предполагая, что результаты наших измерений распределены нормально (с X и σ , как мы оценили), мы можем рассчитать *ожидаемое* число E_k результатов измерений для каждого бина k . Затем необходимо будет решить, насколько хорошо наблюдаемые числа O_k согласуются с ожидаемыми числами E_k .

Расчет ожидаемых чисел E_k очевиден. *Вероятность* того, что результат любого одного измерения попадает в интервал $a < x < b$, равен площади под функцией Гаусса между $x = a$ и $x = b$. В нашем примере вероятности P_1, P_2, P_3, P_4 попадания результата измерения в каждый из четырех бинов — это четыре площади, показанные на рис. 12.1. Две равные площади P_2 и P_3 вместе дают хорошо известное значение 68%, так что вероятность попадания в один из двух центральных бинов составляет 34%, т. е. $P_2 = P_3 = 0,34$. Две внешние площади представляют оставшиеся 32%; таким образом, $P_1 = P_4 = 0,16$. Чтобы найти ожидаемые числа E_k , мы просто умножим эти вероятности на полное число измерений $N = 40$. Эти ожидаемые числа приведены в табл. 12.3. Тот факт, что числа E_k не целые, служит нам напоминанием о том, что «ожидаемое число» — это не то, которое мы действительно ожидаем в любом отдельном эксперименте; это скорее среднее ожидаемое число, которое получится, когда мы повторим всю нашу серию измерений много раз.

¹⁾ Если результат измерения попадает на границу между двумя бинами, то вы можете половину результатов приписать одному бину и половину — другому.

Таблица 12.3. Ожидаемые числа E_k и наблюдаемые числа O_k для 40 измерений, результаты которых приведены в табл. 12.1

Бин k	1	2	3	4
Вероятность P_k , %	16	34	34	16
Ожидаемое число $E_k = NP_k$	6,4	13,6	13,6	6,4
Наблюдаемое число O_k	8	10	16	6

Теперь наша задача — решить, насколько хорошо ожидаемые числа E_k согласуются с соответствующими наблюдаемыми числами O_k (в нижней строке табл. 12.3). Мы, естественно, не ожидаем *идеального* согласия между E_k и O_k после любого конечного числа измерений. Но с другой стороны, если наша гипотеза о том, что результаты измерений распределены нормально, правильна, то мы ожидали бы, что в некотором смысле отклонения

$$O_k - E_k \quad (12.3)$$

будут *малы*. И наоборот, если бы оказалось, что отклонения $O_k - E_k$ *велики*, то мы стали бы подозревать, что наша гипотеза неверна.

Чтобы придать точный смысл утверждениям, что отклонения $O_k - E_k$ «малы» или «велики», мы должны решить, насколько большими мы *ожидали* бы значения $O_k - E_k$, если бы результаты наших измерений действительно были распределены нормально. К счастью, это легко сделать. Если представить себе, что вся наша серия из 40 измерений повторяется много раз, то числа O_k результатов измерений в любом одном бине k можно рассматривать как результат счетного эксперимента типа описанного в гл. 11. Множество наших различных результатов для O_k должно было бы иметь среднее значение E_k и флуктуировать относительно E_k со стандартным отклонением порядка $\sqrt{E_k}$. Таким образом, два числа, которые нужно сравнивать, это отклонение $O_k - E_k$ и ожидаемая величина его флуктуаций $\sqrt{E_k}$.

Итак, мы должны рассмотреть отношение

$$\frac{O_k - E_k}{\sqrt{E_k}}. \quad (12.4)$$

Для некоторых бинов k это отношение будет положительным, для других — отрицательным; для малого числа бинов k оно может существенно превышать единицу, но для большинства бинов оно должно быть порядка единицы или меньше. Чтобы проверить нашу гипотезу (о том, что результаты измерений распределены нормально), естественно число (12.4) возвести

Таблица 12.4. Расчет χ^2 для данных табл. 12.1

Номер бина k	1	2	3	4
Бин	$x < X - \sigma$	$X - \sigma < x < X$	$X < x < X + \sigma$	$X + \sigma < x$
Наблюдаемое число O_k	8	10	16	6
Ожидаемое число E_k	6,4	13,6	13,6	6,4
$O_k - E_k$	1,6	-3,6	2,4	-0,4

в квадрат для каждого k и затем просуммировать по всем бином $k = 1, \dots, n$ (в данном случае $n = 4$). Эта процедура определяет число, называемое χ^2

$$\chi^2 = \sum_{k=1}^n \frac{(O_k - E_k)^2}{E_k}. \quad (12.5)$$

Должно быть ясно, что это число χ^2 служит показателем того, насколько хорошо согласуются наблюдаемое и ожидаемое распределения. Если $\chi^2 = 0$, то согласие идеальное, т. е. $O_k = E_k$ для всех бинов k , что в высшей степени невероятно. В общем случае отдельные члены в сумме (12.5) должны быть порядка 1, а в сумме всего n членов. Итак, если

$$\chi^2 \leq n$$

(χ^2 порядка n или меньше), то наблюдаемое и ожидаемое распределения согласуются настолько хорошо, насколько можно было бы ожидать. Другими словами, если $\chi^2 \leq n$, то у нас нет оснований сомневаться в том, что результаты наших измерений были распределены так, как ожидалось. С другой стороны, если

$$\chi^2 \gg n$$

(χ^2 значительно больше, чем число бинов), то наблюдаемые и ожидаемые числа значительно различаются и есть все основания подозревать, что результаты наших измерений не распределены в соответствии с ожидаемым законом.

В нашем примере наблюдаемые и ожидаемые числа для четырех бинов еще раз приведены в табл. 12.4, и простой расчет с их использованием дает

$$\begin{aligned} \chi^2 &= \sum_{k=1}^4 \frac{(O_k - E_k)^2}{E_k} = \\ &= \frac{(1,6)^2}{6,4} + \frac{(-3,6)^2}{13,6} + \frac{(2,4)^2}{13,6} + \frac{(-0,4)^2}{6,4} = \\ &= 1,80. \end{aligned} \quad (12.6)$$

Так как величина 1,80 для χ^2 меньше, чем число членов в сумме (а именно 4), то у нас нет оснований сомневаться в гипотезе, что результаты наших измерений распределены нормально.

12.2. Общее определение χ^2

До сих пор мы ограничивались рассмотрением одного частного примера, когда 40 раз измеряется непрерывная переменная x — расстояние, которое пролетает пуля, вылетевшая из некоторого ружья. Мы определили число χ^2 и увидели, что оно может служить по крайней мере грубой мерой, характеризующей согласие наблюдаемого распределения результатов измерений и распределения Гаусса, в соответствии с которым, как мы ожидаем, распределены эти наши результаты. Сейчас мы увидим, что можно определять и использовать χ^2 аналогичным образом и в случае многих других экспериментов.

Рассмотрим любой эксперимент, в котором мы измеряем какое-то число x и для которого у нас есть основания ожидать некоторое определенное распределение результатов. Мы можем представить себе, что эксперимент повторяется много раз (N), и, поделив интервал возможных значений x на n бинов, $k = 1, \dots, n$, можем подсчитать числа O_k наблюдений, когда результат попадает в каждый бин k . Предполагая, что результаты измерений действительно распределены в соответствии с ожидаемым распределением, мы затем вычисляем ожидаемые числа E_k измерений для k -го бина. Наконец, вычисляем χ^2 точно так же, как и в случае (12.5):

$$\chi^2 = \sum_{k=1}^n \frac{(O_k - E_k)^2}{E_k}. \quad (12.7)$$

Роль χ^2 всегда примерно такая же, как и в предыдущем примере, т. е. если $\chi^2 \leq n$, то согласие между наблюдаемым и ожидаемым распределениями приемлемое, если $\chi^2 \gg n$, то имеется существенное расхождение.

Процедура выбора бинов для слагаемых, по которым вычисляется χ^2 , в некотором отношении зависит от характера эксперимента. В частности, она зависит от того, непрерывна или дискретна измеряемая величина x . Сейчас мы рассмотрим эти две ситуации.

Измерения непрерывной переменной

Пример, рассмотренный в разд. 12.1, относится к непрерывной переменной x , и лишь немного можно добавить к

тому, что уже было сказано. Единственное предельное распределение, которое мы рассмотрели в случае непрерывной переменной, — это распределение Гаусса, но существует, конечно, множество различных распределений, которые можно было бы ожидать. Например, в случае многих атомных и ядерных экспериментов ожидаемое распределение измеряемой переменной x (энергии) — это распределение Лоренца

$$f(x) \sim \frac{1}{(x - X)^2 + \gamma^2},$$

где X и γ — некоторые постоянные.

Каким бы ни было ожидаемое распределение $f(x)$, полная площадь под графиком $f(x)$ всегда равна 1, и вероятность того, что результат измерения попадет между $x = a$ и $x = b$, — это площадь под графиком между a и b :

$$P(a < x < b) = \int_a^b f(x) dx.$$

Таким образом, если k -й бин имеет размеры от $x = a_k$ до $x = a_{k+1}$, то ожидаемое число измерений в k -м бине (после выполнения всех N измерений) будет равно

$$\begin{aligned} E_k &= N \cdot P(a_k < x < a_{k+1}) = \\ &= N \int_{a_k}^{a_{k+1}} f(x) dx. \end{aligned} \quad (12.8)$$

Когда мы будем рассматривать использование критерия χ^2 с количественной стороны в разд. 12.4, то увидим, что ожидаемые числа E_k не должны быть слишком малы. Хотя не существует определенной нижней границы, однако E_k должны быть, вероятно, примерно равны или больше 5,

$$E_k \geq 5. \quad (12.9)$$

Следовательно, мы должны выбирать бины таким образом, чтобы E_k , определяемые (12.8), удовлетворяли этому условию. Мы увидим также, что число бинов не должно быть слишком мало. Так, в случае примера из разд. 12.1, когда ожидаемым было распределение Гаусса, у которого центр X и ширина σ не были известны заранее, критерий χ^2 не действует (как мы увидим), если число бинов меньше четырех, т. е. в случае этого примера нам необходимо иметь

$$n \geq 4. \quad (12.10)$$

Рассматривая вместе (12.9) и (12.10), мы увидим, что нельзя использовать критерий χ^2 в такого рода экспериментах, если наше полное число наблюдений меньше 20.

Измерение дискретной переменной

Предположим, что теперь мы измеряем дискретную переменную, такую, как уже знакомое число единичек при бросании нескольких игральных костей. На практике наиболее часто встречающаяся дискретная переменная — целое число (подобно числу единичек), и мы будем обозначать дискретную переменную через v вместо x (x мы будем использовать для непрерывной переменной). Если мы бросаем пять костей, то возможные значения v — это $v = 0, 1, \dots, 5$, и фактически нет необходимости разбивать возможные результаты по бинам. Просто можно сосчитать, сколько раз мы получили каждое из шести возможных значений. Мы можем выразить это и иначе, сказав, что мы выбрали шесть бинов, причем каждый бин содержит только одно возможное значение.

Тем не менее часто бывает желательно сгруппировать несколько различных результатов в один бин. Например, если мы бросали пять наших костей 200 раз, то (в соответствии с вероятностями, определяемыми в задаче 10.7) ожидаемое распределение результатов будет таким, как приведено в двух первых столбцах табл. 12.5. Мы видим, что в данном случае ожидаемые числа бросаний, в которых реализуются четыре или пять единичек, равны соответственно 0,6 и 0,03, т. е. оба много меньше, чем примерно пять, что требуется для каждого бина, если мы хотим применять критерий χ^2 . Эту трудность легко устранить, сгруппировав результаты для $v = 3, 4$ и 5 в один бин. Это приводит к четырем бинам $k = 1, 2, 3, 4$, которые показаны вместе с соответствующими ожидаемыми числами E_k в последних двух столбцах табл. 12.5.

Выбрав бины, как описано, мы могли бы сосчитать наблюдаемые числа выпадений O_k для каждого бина. Затем мы могли бы вычислить χ^2 и посмотреть, согласуются ли наблюдаемое и ожидаемое распределения. В этом эксперименте нам

Таблица 12.5. Ожидаемые числа выпадений v единичек ($v = 0, 1, \dots, 5$) при бросании пяти игральных костей 200 раз

Результат	Ожидаемое число выпадений	Номер бина k	Ожидаемое число E_k
Ни одной единички	80,4	1	80,4
Одна единичка	80,4	2	80,4
Две	32,2	3	32,2
Три	6,4	4	7,0
Четыре	0,6		
Пять	0,03		

известно, что ожидаемое распределение — это, конечно, биномиальное распределение $b_{5, 1/6}(v)$ при условии, что все игральные кости подлинные (т. е. p действительно равно $1/6$). Таким образом, наше испытание распределения в данном случае есть просто проверка, подлинные ли игральные кости мы используем или с утяжелениями у каких-то граней.

В любом эксперименте с дискретной переменной бины можно выбирать так, чтобы каждый содержал только один результат при условии, что ожидаемое число реализаций в каждом бине составляет по крайней мере необходимые 5 или около того. В противном случае несколько различных результатов следует сгруппировать в один бин большего размера, который действительно включает достаточное ожидаемое число реализаций.

Другие формы χ^2

Мы использовали обозначение χ^2 еще раньше в этой книге. Оно применялось в формуле (7.6) и вновь в (8.5) и могло быть использовано для обозначения суммы квадратов в (5.42). Во всех этих случаях χ^2 — это сумма квадратов, в общем виде записываемая как

$$\chi^2 = \sum_1^n \left(\frac{\text{наблюдаемое значение} - \text{ожидаемое значение}}{\text{стандартное отклонение}} \right)^2. \quad (12.11)$$

Во всех случаях χ^2 служит показателем согласия между наблюдаемыми и ожидаемыми значениями некоторой переменной. При хорошем согласии χ^2 будет порядка n , при плохом χ^2 будет много больше, чем n .

К сожалению, мы можем использовать критерий χ^2 для проверки этого согласия только в том случае, если нам известны ожидаемые значения и стандартное отклонение, что позволит нам вычислить (12.11). По-видимому, наиболее типичные случаи, когда эти величины достаточно хорошо известны, рассмотрены в примерах данной главы, в которых предельное распределение определяет E_k и стандартное отклонение $\sqrt{E_k}$. Тем не менее критерий χ^2 очень широко используется. Обсудим, например, рассмотренную в гл. 8 задачу об измерении двух переменных x и y , где ожидается, что y есть некоторая определенная функция x :

$$y = f(x)$$

(подобная $y = A + Bx$). Предположим, что у нас есть N измеренных пар значений (x_i, y_i) , причем погрешность в x_i пренебрежимо мала, а погрешность σ_i в y_i известна. В данном

случае ожидаемое значение y_i есть $f(x_i)$, и мы могли бы проверить, насколько хорошо y аппроксимирует функцию $f(x)$ с помощью вычисления

$$\chi^2 = \sum_1^N \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2.$$

Все наши предыдущие замечания об ожидаемом значении χ^2 применимы к этому числу, и можно использовать количественные критерии, описанные в последующих разделах. Мы не будем рассматривать здесь этого важного применения, так как оно довольно редко встречается в лаборатории вводного курса физики, поскольку погрешности σ_i в этом случае должны быть достаточно хорошо известны (см., однако, задачу. 12.14).

12.3. Степени свободы и приведенное значение χ^2

Мы утверждали, что можем проверять согласие между наблюдаемым и ожидаемым распределениями, вычисляя χ^2 и сравнивая его значение с числом бинов, используемых для представления данных. Оказывается, что более строгой процедурой было бы сравнение χ^2 не с числом бинов n , а с *числом степеней свободы*, обозначаемым через d . Мы кратко упоминали о понятии числа степеней свободы в разд. 8.3. Теперь необходимо обсудить это понятие более подробно.

В общем случае число степеней свободы d в статистических расчетах определяется как число полученных данных *минус* число параметров, вычисленных по этим данным и используемых в расчетах. Для задач, рассматриваемых в этой главе, полученные данные — числа наблюдений O_k в n бинах, где $k = 1, \dots, n$. Таким образом, число полученных данных равно n , числу бинов. Следовательно, для рассматриваемых здесь задач

$$d = n - c,$$

где n — число бинов, c — число параметров, которые были вычислены по данным для расчета ожидаемых чисел E_k . Число c часто называют числом *связей*, что мы кратко поясним.

Число связей c изменяется в зависимости от обсуждаемой задачи. Рассмотрим сначала эксперимент с бросанием игральных костей из разд. 12.2. Если мы бросаем пять костей и проверяем гипотезу, что кости подлинны, то ожидаемое распределение чисел выпадения единичек — биномиальное

распределение $b_{5, 1/6}(v)$, где $v = 0, \dots, 5$ — числа единичек в любом одном бросании. Оба параметра этой функции — число игральные кости (пять) и вероятность выпадения единички ($1/6$) — известны заранее, и их не надо вычислять из данных эксперимента. Рассчитывая ожидаемое число выпадений любого частного значения v , мы должны умножить биномиальную вероятность на полное число бросков N (в случае нашего примера $N = 200$). Этот параметр и в самом деле зависит от данных. В частности, N есть просто сумма чисел O_k :

$$N = \sum_{k=1}^n O_k. \quad (12.12)$$

Таким образом, вычисляя ожидаемые результаты в нашем эксперименте с игральными костями, мы должны рассчитать только один параметр (N) из данных. Число связей равно, следовательно,

$$c = 1,$$

и число степеней свободы определяется как

$$d = n - 1.$$

В табл. 12.5 результаты эксперимента с костями были сгруппированы в четыре бина (т. е. $n = 4$), поэтому в таком эксперименте было три степени свободы.

Формула (12.12) хорошо иллюстрирует термины «связи» и «степени свободы». Когда число N определено, то (12.12) можно рассматривать как уравнение, которое «связывает» значения $O_1, \dots, O_n$. Точнее, можно сказать, что из-за связи (12.12) только $n - 1$ чисел $O_1, \dots, O_{n-1}$ могут принимать любое значение (в определенных границах), но зато последнее число O_n полностью определяется формулой (12.12). В этом смысле только $n - 1$ данных *могут* принимать независимые значения; поэтому мы говорим, что имеется только $n - 1$ независимых степеней свободы.

В первом примере этой главы дальность полета x пули была измерена 40 раз ($N = 40$). Результаты были собраны в четыре бина ($n = 4$) и сравнивались с теми значениями, которые мы ожидали бы от распределения Гаусса $f_{\chi, \sigma}(x)$. В этом случае имелось *три* связи и, следовательно, только одна степень свободы

$$d = n - c = 4 - 3 = 1.$$

Первая связь такая же, как и (12.12): полное число наблюдений N равно сумме чисел наблюдений O_k во всех бинах. Но в данном случае имеются еще другие связи, поскольку

(как обычно бывает в такого рода экспериментах) нам заранее не известны параметры X и σ ожидаемого распределения Гаусса $f_{X, \sigma}(x)$. Таким образом, прежде чем мы смогли бы рассчитать ожидаемые числа E_k , мы должны были оценить X и σ , используя данные. Следовательно, всего было три связи, поэтому в данном примере

$$d = n - 3. \quad (12.13)$$

Отсюда ясно, почему мы должны были использовать по крайней мере четыре бина в этом эксперименте. Мы увидим, что число степеней свободы всегда должно быть не меньше единицы, поэтому из (12.13) очевидно, что нам пришлось выбирать $n \geq 4$.

В рассмотренных здесь примерах всегда имеется по крайней мере одна связь (а именно связь $N = \sum O_k$, включающая полное число измерений), но могут существовать еще одна или две. Таким образом, число степеней свободы d будет изменяться от $n - 1$ до $n - 3$ (в наших примерах). Когда число n велико, различие между n и d практически не имеет значения, но если n мало (как это часто бывает, к сожалению), имеется, очевидно, существенная разница.

Теперь, используя понятие числа степеней свободы, мы можем приступить к выработке более точного критерия χ^2 . Можно показать (хотя мы и не будем этого делать), что ожидаемое значение χ^2 точно равно d , числу степеней свободы

$$(\text{ожидаемое среднее значение } \chi^2) = d. \quad (12.14)$$

Эта важная формула не означает, что мы действительно ожидаем получить $\chi^2 = d$ после одной серии измерений. Но она имеет тот смысл, что если бы мы могли повторить всю нашу серию измерений бесконечное число раз и каждый раз вычислять χ^2 , то среднее этих значений χ^2 было бы равно d . Тем не менее даже после *одной* серии измерений сравнение χ^2 и d служит показателем согласия. В частности, если наше ожидаемое распределение было *правильно*, то крайне невероятно, чтобы значение χ^2 было значительно больше, чем d . И наоборот, можно сказать: если мы нашли $\chi^2 \gg d$, то можем утверждать, что крайне невероятно, чтобы наше ожидаемое распределение было верным.

Мы не доказали формулу (12.14), но можем видеть, что некоторые ее следствия очень разумны. Например, так как $d = n - c$, формулу (12.14) можно переписать как

$$(\text{ожидаемое среднее значение } \chi^2) = n - c. \quad (12.15)$$

Иными словами, для любого данного n ожидаемое значение χ^2 будет меньше, когда c будет больше (т. е. если мы вычислим больше параметров по данным). Это как раз то, что

нам следует ожидать. В случае примера из разд. 12.1 мы использовали данные для расчетов положения центра X и ширины σ ожидаемого распределения $f_{X, \sigma}(x)$. Естественно, поскольку X и σ были выбраны таким образом, чтобы аппроксимировать данные, то нам следует ожидать несколько лучшего согласия между наблюдаемым и ожидаемым распределениями, т. е. следует ожидать, что эти две лишние связи уменьшат значение χ^2 . А это как раз то, что предполагает (12.15).

Из формулы (12.14) следует, что существует несколько более удобный способ понимания критерия χ^2 . Мы введем понятие *приведенного значения χ^2* (или χ^2 на одну степень свободы), которое мы обозначим через $\tilde{\chi}^2$ и определим как

$$\tilde{\chi}^2 = \chi^2/d. \quad (12.16)$$

Поскольку ожидаемое значение χ^2 равно d , то мы получаем

$$(\text{ожидаемое среднее значение } \tilde{\chi}^2) = 1. \quad (12.17)$$

Таким образом, каким бы ни было число степеней свободы, наш критерий можно сформулировать следующим образом: если мы получаем значение $\tilde{\chi}^2$ порядка 1 или меньше, то у нас нет оснований сомневаться в нашем ожидаемом распределении; если мы получаем значение $\tilde{\chi}^2$ много большее, чем единица, то невероятно, чтобы наше ожидаемое распределение было верным.

12.4. Вероятности для χ^2

Наш критерий проверки согласия между полученными данными и их ожидаемым распределением все еще остается только качественным. Нам хотелось бы иметь *количественную* меру согласия. В частности, нам нужно некоторое указание на то, где проводить границу между согласием и несогласием. Например, в случае эксперимента из разд. 12.1 мы сделали 40 измерений некоторых расстояний x , распределение которых, как мы полагаем, должно быть гауссовым. Мы распределили наши данные по четырем бинам и нашли, что $\chi^2 = 1,80$. При трех связях оставалась только одна степень свободы ($d = 1$), поэтому приведенное значение χ^2 , $\tilde{\chi}^2 = \chi^2/d$, также равно 1,80:

$$\tilde{\chi}^2 = 1,80.$$

Теперь возникает вопрос: достаточно ли велико значение $\tilde{\chi}^2 = 1,80$ по сравнению с единицей, чтобы отвергнуть ожидаемое распределение Гаусса?

Чтобы ответить на этот вопрос, будем исходить из предположения, что результаты наших измерений действительно распределены в соответствии с ожидаемым распределением (например, гауссовым). В рамках этого предположения можно вычислить вероятность получения значения $\tilde{\chi}^2$, которое будет не меньше, чем наше значение 1,80. В данном случае оказывается, что эта вероятность равна

$$P(\tilde{\chi}^2 \geq 1,80) \approx 18 \%,$$

как мы вскоре увидим. Следовательно, если бы наши результаты подчинялись ожидаемому распределению, то существовала бы 18%-ная вероятность получения значения $\tilde{\chi}^2$, большего или равного 1,80, которое мы фактически получили. Другими словами, в этом эксперименте значение $\tilde{\chi}^2$, равное 1,80, вовсе не является неразумным, поэтому у нас нет причин (основанных на этих данных) отвергнуть наше ожидаемое распределение.

В общем случае наша процедура должна быть теперь довольно ясна. После завершения любой серии измерений мы вычисляем приведенное значение $\tilde{\chi}^2$, которое теперь мы будем обозначать через $\tilde{\chi}_o^2$ [где нижний индекс «о» означает «полученное» (obtained), так как $\tilde{\chi}_o^2$ есть значение, которое действительно получено]. Затем, предполагая, что результаты наших измерений распределены в соответствии с ожидаемым распределением, мы вычисляем вероятность

$$P(\tilde{\chi}^2 > \tilde{\chi}_o^2) \quad (12.18)$$

получения значения $\tilde{\chi}^2$, большего или равного значению $\tilde{\chi}_o^2$, фактически полученному. Если эта вероятность велика, то наше значение $\tilde{\chi}_o^2$ вполне приемлемо и нет оснований отвергать наше ожидаемое распределение. Если эта вероятность неразумно мала, то значение $\tilde{\chi}^2$, столь же большое, как наше полученное $\tilde{\chi}_o^2$, очень невероятно (если результаты наших измерений были распределены, как ожидалось), и соответственно невероятно, чтобы наше ожидаемое распределение было правильным.

Как всегда бывает со статистическими критериями, мы должны решить, где проходит граница между тем, что «разумно вероятно», и тем, что нет. Два обычных выбора уже упоминались в связи с корреляциями. Выбрав границу в 5%, мы бы сказали, что наше полученное значение $\tilde{\chi}_o^2$ обнаруживает «значимое разногласие», если

$$P(\tilde{\chi}^2 \geq \tilde{\chi}_o^2) < 5 \%,$$

Таблица 12.6. Выраженная в процентах вероятность $P_d(\tilde{\chi}^2 \geq \tilde{\chi}_0^2)$ получения значения $\tilde{\chi}^2$, большего или равного любому частному значению $\tilde{\chi}_0^2$, в предположении, что результаты рассматриваемых измерений распределены в соответствии с ожидаемым распределением. Прочерки соответствуют вероятностям, меньшим 0,05%

d	$\tilde{\chi}_0^2$												
	0	0,25	0,5	0,75	1,0	1,25	1,5	1,75	2	3	4	5	6
1	100	62	48	39	32	26	22	19	16	8	5	3	1
2	100	78	61	47	37	29	22	17	14	5	2	0,7	0,2
3	100	86	68	52	39	29	21	15	11	3	0,7	0,2	—
5	100	94	78	59	42	28	19	12	8	1	0,1	—	—
10	100	99	89	68	44	25	13	6	3	0,1	—	—	—
15	100	100	94	73	45	23	10	4	1	—	—	—	—

и мы отвергли бы наше ожидаемое распределение на «5%-ном уровне значимости». Если бы мы установили границу в 1%, то могли бы сказать, что расхождение «высокозначимо», если $P(\tilde{\chi}^2 \geq \tilde{\chi}_0^2) < 1\%$, и отвергнуть ожидаемое распределение на «1%-ном уровне значимости».

Какой бы уровень мы ни выбрали в качестве границы, когда мы отвергаем гипотезу, выбранный уровень следует указать. Может быть, даже более важно привести и вероятность $P(\tilde{\chi}^2 \geq \tilde{\chi}_0^2)$, так чтобы читатель смог самостоятельно судить о ее разумности.

Расчеты вероятностей $P(\tilde{\chi}^2 \geq \tilde{\chi}_0^2)$ слишком сложны, чтобы их описывать в этой книге. Однако их результаты можно легко свести в таблицы, как, например, табл. 12.6 или более полная таблица в приложении Г. Оказывается, что вероятность получения любого частного значения $\tilde{\chi}^2$ зависит от числа степеней свободы. Поэтому мы будем записывать интересующую нас вероятность как $P_d(\tilde{\chi}^2 \geq \tilde{\chi}_0^2)$, чтобы подчеркнуть ее зависимость от d .

В обычных расчетах вероятностей $P_d(\tilde{\chi}^2 \geq \tilde{\chi}_0^2)$ полученные числа O_k рассматриваются как непрерывные переменные, которые распределены относительно их ожидаемых значений E_k в соответствии с распределением Гаусса. В задачах, которые мы здесь рассматриваем, O_k — дискретная переменная, распределенная в соответствии с законом Пуассона¹⁾. При усло-

¹⁾ Мы показали, что определение чисел O_k происходит в счетном эксперименте и что, следовательно, O_k должны быть распределены по закону Пуассона. Если размер бина k слишком велик, то эта аргументация

вии что все рассматриваемые числа достаточно велики, дискретный характер O_k не важен и распределение Пуассона хорошо аппроксимируется функцией Гаусса. При этих условиях можно использовать табулированные значения вероятностей $P_d(\tilde{\chi}^2 \geq \tilde{\chi}_o^2)$ на достаточно законных основаниях. Именно по этой причине мы говорили, что бины должны выбираться таким образом, чтобы ожидаемые отсчеты E_k в каждом бине были достаточно велики (по крайней мере порядка 5). По той же причине число бинов не должно быть слишком малым.

С этими предупреждениями мы приводим значения рассчитанных вероятностей $P_d(\tilde{\chi}^2 \geq \tilde{\chi}_o^2)$ для небольшого числа выбранных значений d и $\tilde{\chi}_o^2$ в табл. 12.6. Числа в левом столбце предоставляют выбор из шести значений d , числа степеней свободы ($d = 1, 2, 3, 5, 10, 15$). Числа сверху над другими столбцами — возможные значения полученных $\tilde{\chi}_o^2$. Каждое число в таблице — выраженная в процентах вероятность $P_d(\tilde{\chi}^2 \geq \tilde{\chi}_o^2)$ как функция d и $\tilde{\chi}_o^2$. Например, для десяти степеней свободы ($d = 10$) мы находим, что вероятность получения $\tilde{\chi}^2 \geq 2$ равна 3%:

$$P_{10}(\tilde{\chi}^2 \geq 2) = 3\%.$$

Таким образом, если бы мы получили приведенное значение $\tilde{\chi}^2$, равное 2, в эксперименте с десятью степенями свободы, мы могли бы заключить, что наши наблюдения «значимо» отличаются от ожидаемого распределения, и отвергнуть ожидаемое распределение на 5%-ном уровне значимости (но не на 1%-ном уровне).

Все вероятности во втором столбце табл. 12.6 равны 100%, так как получение $\tilde{\chi}^2 \geq 0$ всегда достоверно. С ростом $\tilde{\chi}_o^2$ вероятность получения $\tilde{\chi}^2 \geq \tilde{\chi}_o^2$ уменьшается, но характер этого уменьшения зависит от d . Например, для двух степеней свободы ($d = 2$) вероятность $P_d(\tilde{\chi}^2 \geq 1)$ составляет 37%, в то время как для $d = 15$ $P_d(\tilde{\chi}^2 \geq 1)$ составляет 45%. Заметьте, что $P_d(\tilde{\chi}^2 \geq 1)$ всегда велика (по крайней мере не менее 32%), поэтому значение $\tilde{\chi}_o^2$, равное 1 или менее, вполне разумно и никогда не приводит к требованию отвергнуть ожидаемое распределение.

Минимальная величина $\tilde{\chi}_o^2$, которая ставит под вопрос ожидаемое распределение, зависит от d . Для одной степени свободы мы видим, что $\tilde{\chi}_o^2$ может быть и большим, напри-

не совсем верна, так как вероятность для такого бина не мала по сравнению с 1 (что является одним из условий справедливости распределения Пуассона, как упомянуто в разд. 11.1); поэтому мы должны иметь разумное число бинов.

мер 4, прежде чем расхождение станет «значимым» (на 5%-ном уровне). В случае двух степеней свободы соответствующая граница равна $\tilde{\chi}_0^2 = 3$; для $d = 5$ она близка к 2 (фактически $\tilde{\chi}_0^2 = 2,2$) и так далее.

Используя вероятности из табл. 12.6 (или из приложения Г), мы теперь можем приписывать количественный критерий значимости для значения $\tilde{\chi}_0^2$, полученного в любом конкретном эксперименте. В разд. 12.5 приведено несколько примеров.

12.5. Примеры

Мы уже довольно подробно проанализировали пример из разд. 12.1. В этом разделе мы рассмотрим еще три примера, чтобы проиллюстрировать применение критерия χ^2 .

Другой пример распределения Гаусса

Пример из разд. 12.1 относился к измерениям, результаты которых, как ожидалось, были распределены нормально. Нормальное, или гауссово, распределение встречается настолько часто, что мы кратко рассмотрим еще один пример. Предположим, что некий антрополог интересуется ростом аборигенов на некотором острове. Он полагает, что взрослые мужчины по росту должны быть распределены нормально, и измеряет рост для выборки из 200 мужчин. Используя эти результаты, он вычисляет среднее значение и стандартное отклонение и использует эти числа как наилучшие оценки центра X и параметра ширины σ ожидаемого нормального распределения $f_{X, \sigma}(x)$. Затем он выбирает восемь бинов, как показано в двух левых столбцах табл. 12.7, распределяет по

Таблица 12.7. Измерения роста 200 взрослых мужчин

Номер бина k	Рост в бине	Наблюдаемое число O_k	Ожидаемое число E_k
1	менее $X - 1,5\sigma$	14	13,4
2	между $X - 1,5\sigma$ и $X - \sigma$	29	18,3
3	между $X - \sigma$ и $X - 0,5\sigma$	30	30,0
4	между $X - 0,5\sigma$ и X	27	38,3
5	между X и $X + 0,5\sigma$	28	38,3
6	между $X + 0,5\sigma$ и $X + \sigma$	31	30,0
7	между $X + \sigma$ и $X + 1,5\sigma$	28	18,3
8	более $X + 1,5\sigma$	13	13,4

ним свои наблюдения и получает результаты, показанные в третьем столбце.

Далее наш антрополог желает проверить, согласуются ли эти результаты с ожидаемым нормальным распределением $f_{X,\sigma}(x)$. С этой целью он сначала рассчитывает вероятность P_k , того, что любой один мужчина имеет рост в любом заданном бине k (предполагая нормальное распределение). Она равна интегралу от $f_{X,\sigma}(x)$ в пределах границ бина и легко находится по таблице интеграла ошибок приложения Б. Затем ожидаемое число E_k в каждом бине находится умножением P_k на полное число выбранных мужчин (200). Эти числа приведены в последнем столбце табл. 12.7.

Чтобы вычислить ожидаемые числа E_k , антропологу пришлось использовать три параметра, которые были рассчитаны по его данным (полный объем выборки и оценки для X и σ). Таким образом, хотя имеется восемь бинов, но было три связи, поэтому число степеней свободы равно $d = 8 - 3 = 5$. Простой расчет с использованием данных табл. 12.7 дает для приведенного значения $\tilde{\chi}^2$

$$\tilde{\chi}^2 = \frac{1}{d} \sum_{i=1}^8 \frac{(O_k - E_k)^2}{E_k} = 3,5.$$

Так как эта величина заметно больше, чем 1, мы сразу же начинаем подозревать, что рост островитян не подчиняется нормальному распределению. Точнее, из табл. 12.6 мы видим, что если бы по росту островитяне были распределены, как ожидалось, то вероятность $P_5(\tilde{\chi}^2 \geq 3,5)$ получения значения $\tilde{\chi}^2$, не меньшего чем 3,5, составляет примерно 0,5 %. По любым стандартам это практически невероятно, и мы заключаем, что практически невероятно, чтобы островитяне были распределены по росту нормально. В частности, на 1 %-ном (или «высокозначимом») уровне мы можем отвергнуть гипотезу о нормальном распределении по росту.

Вновь игральные кости

В разд. 12.2 мы рассмотрели эксперимент, в котором пять костей бросалось много раз и в каждом таком бросании подсчитывалось число выпавших единичек. Предположим, что мы сделали 200 бросков и распределили результаты по бинам, как было обсуждено раньше. Предполагая, что кости подлинные, мы можем рассчитать ожидаемые числа E_k , как и прежде. Они приведены в третьем столбце табл. 12.8.

В фактическом испытании пять костей бросалось 200 раз, и в последнем столбце табл. 12.8 представлены числа, кото-

Таблица 12.8. Распределение чисел единичек в 200 бросаниях 5 игральных костей

Номер бина k	Результаты в бине	Ожидаемое число E_k	Наблюдаемое число O_k
1	Ни одной единички	80,4	60
2	Одна единичка	80,4	88
3	Две единички	32,2	39
4	3, 4 или 5 единичек	7,0	13

рые в действительности наблюдались. Чтобы проверить согласие между полученным и ожидаемым распределениями, мы просто заметим, что имеется три степени свободы (четыре бина минус одна связь), и вычислим

$$\tilde{\chi}^2 = \frac{1}{3} \sum_{k=1}^4 \frac{(O_k - E_k)^2}{E_k} = 4,16.$$

Используя опять табл. 12.6, мы видим, что для трех степеней свободы вероятность получения столь большого или большего значения $\tilde{\chi}^2$ составляет примерно 0,7%, если кости подлинны. Мы заключаем, что почти наверное кости фальшивые. Сравнение чисел E_k и O_k в табл. 12.8 заставляет предположить, что по крайней мере одна кость утяжелена со стороны единички.

Пример распределения Пуассона

В качестве последнего примера по использованию критерия χ^2 рассмотрим эксперимент, в котором ожидаемое распределение есть закон Пуассона. Предположим, что мы используем счетчик Гейгера для счета частиц космических лучей в некотором месте. Предположим далее, что мы считаем частицы, пришедшие в 100 отдельных одноминутных интервалах, и получаем результаты, приведенные в двух первых столбцах табл. 12.9.

Анализ чисел во втором столбце сразу же заставляет нас объединить все отсчеты с $v \geq 5$ в один бин. Такой выбор шести бинов ($k = 1, \dots, 6$) показан в третьем столбце, а соответствующие числа O_k приведены в четвертом.

Гипотеза, которую мы хотим проверить, состоит в том, что числа v распределены в соответствии с законом Пуассона $p_\mu(v)$. Поскольку ожидаемый средний отсчет μ неизвестен, мы должны сначала вычислить среднее значение для нашей сотни отсчетов. Оно легко находится и оказывается равным

Таблица 12.9. Числа частиц космических лучей, полученных в 100 отдельных одноминутных интервалах

Числа ν отсчетов за одну минуту	Числа наблюдений	Номер бина k	Числа наблюдений O_k в бине k	Ожидаемые числа E_k
Ни одного	7	1	7	7,5
1	17	2	17	19,4
2	29	3	29	25,2
3	20	4	20	21,7
4	16	5	16	14,1
5	8	6	11	12,1
6	1			
7	2			
8 или более	0			
Полное число	100			

$\bar{\nu} = 2,59$, что дает нам нашу наилучшую оценку для μ . Используя это значение $\mu = 2,59$, мы можем рассчитать вероятность $p_\mu(\nu)$ любого частного отсчета ν и, следовательно, ожидаемые числа E_k , которые приведены в последнем столбце.

При расчете чисел E_k мы использовали два параметра, определенные из данных, а именно полное число наблюдений (100) и нашу оценку μ ($\mu = 2,59$). (Заметьте, что поскольку закон Пуассона полностью определяется заданием μ , нам не пришлось оценивать стандартное отклонение σ . В самом деле, так как $\sigma = \sqrt{\mu}$, то наша оценка μ автоматически дает нам оценку σ .) Следовательно, имеются две связи, которые для шести бинов приводят к четырем степеням свободы $d = 4$.

Простой расчет, в котором используются числа из двух последних столбцов табл. 12.9, дает для приведенного значения $\tilde{\chi}^2$

$$\tilde{\chi}^2 = \frac{1}{d} \sum_{k=1}^6 \frac{(O_k - E_k)^2}{E_k} = 0,35.$$

Так как это значение меньше единицы, то мы можем сразу же сделать вывод, что согласие между нашими наблюдениями и ожидаемым распределением Пуассона удовлетворительно. Более точно, из таблицы приложения Г видно, что значение $\tilde{\chi}^2$, равное 0,35, весьма вероятно; действительно, $P_4 (\tilde{\chi}^2 \geq 0,35) \approx 85\%$. Таким образом, наш эксперимент не дает абсолютно никаких оснований сомневаться в ожидаемом распределении Пуассона.

Значение $\tilde{\chi}^2 = 0,35$, полученное в этом эксперименте, действительно значительно меньше, чем 1; это свидетельствует

о том, что наши наблюдения очень хорошо удовлетворяют распределению Пуассона. Однако это малое значение *не* может служить более строгим свидетельством того, что наши результаты измерений распределены по ожидаемому закону, по сравнению с тем, что давало бы значение $\tilde{\chi}^2 \approx 1$. Если результаты действительно распределены по ожидаемому закону и если бы мы могли повторить нашу серию измерений много раз, то мы должны были бы получить множество различных значений $\tilde{\chi}^2$, флуктуирующих относительно среднего значения 1. Таким образом, если результаты распределены в соответствии с ожидаемым распределением, то значение $\tilde{\chi}^2 = 0,35$ — результат очень вероятной флуктуации относительно ожидаемого среднего значения. Оно никоим образом не дает лишних подтверждений для нашего вывода о том, что результаты наших измерений действительно распределены по ожидаемому закону.

Разобравшись в этих трех примерах, вы не должны испытывать затруднений в применении критерия χ^2 к любой задаче, с которой можно встретиться в лаборатории вводного курса физики. Некоторые дополнительные примеры можно найти среди предлагаемых ниже задач. Вам, безусловно, следует проверить уровень вашего понимания, пробуя решать некоторые из них.

Задачи

Напоминание: звездочка у номера задачи означает, что задача решается или ее ответ приводится в разделе «Ответы» в конце книги.

*12.1 (разд. 12.1). Каждому студенту в группе из 50 человек дали по кусочку одного и того же металла (по крайней мере, им было сказано, что это один и тот же металл) и попросили определить его плотность. По 50 результатам были рассчитаны среднее $\bar{\rho}$ и стандартное отклонение σ_{ρ} , и затем было решено проверить, подчиняются ли результаты измерений нормальному распределению. С этой целью данные измерений были распределены по четырем бинам с границами при $\bar{\rho} - \sigma_{\rho}$, $\bar{\rho}$ и $\bar{\rho} + \sigma_{\rho}$. Полученные результаты приведены в табл. 12.10.

Предполагая, что результаты измерений распределены нормально с центром $\bar{\rho}$ и шириной σ_{ρ} , рассчитайте числа E_k результатов измерений, ожидаемых в каждом бине. Затем вычислите χ^2 . Подчиняются ли результаты измерений нормальному распределению?

Таблица 12.10

Бин k	Значения ρ в бине	Числа наблюдений O_k в бине
1	Меньше $\bar{\rho} - \sigma_{\rho}$	12
2	Между $\bar{\rho} - \sigma_{\rho}$ и $\bar{\rho}$	13
3	Между $\bar{\rho}$ и $\bar{\rho} + \sigma_{\rho}$	11
4	Больше $\bar{\rho} + \sigma_{\rho}$	14

Таблица 12.11

Номер выпавшей грани k	1	2	3	4	5	6
Числа наблюдений O_k	20	46	35	45	42	52

12.2 (разд. 12.1). В задаче 4.7 приведены 30 результатов измерений времени t , их среднее $\bar{t} = 8,15$ с и стандартное отклонение $\sigma_t = 0,04$ с. Распределите данные по четырем бинам с границами при $\bar{t} - \sigma_t$, $\bar{t}$ и $\bar{t} + \sigma_t$ и определите числа наблюдений O_k для каждого бина $k = 1, 2, 3, 4$. Предполагая, что результаты измерений распределены нормально с центром $\bar{t}$ и шириной σ_t , найдите ожидаемые числа E_k для каждого бина. Вычислите χ^2 . Имеются ли основания сомневаться в том, что результаты измерений распределены нормально?

12.3 (разд. 12.2). Игрок решает проверить игральную кость, бросая ее 240 раз. В каждом броске может реализоваться один из шести возможных исходов ($k = 1, 2, \dots, 6$, где k — номер выпавшей грани); распределение результатов его бросаний приведено в табл. 12.11.

Чему равны ожидаемые числа наблюдений E_k в предположении, что игральная кость подлинная? Рассматривая результат для каждой грани как отдельный бин, вычислите χ^2 . Представляется ли вероятным, что кость утяжелена?

*12.4 (разд. 12.2). Три игральные кости бросают 400 раз, наблюдают число шестерок, выпавших в каждом броске, и получают результаты, приведенные в табл. 12.12. Предполагая, что кости подлинные, рассчитайте ожидаемые числа E_k для каждого из трех бинов. (Необходимые вероятности — биномиальные вероятности, рассмотренные в разд. 10.2). Вычислите χ^2 . Имеются ли основания подозревать, что кости утяжелены?

*12.5 (разд. 12.3).

а. Для каждой из задач от 12.1 по 12.4 найдите число связей s и число степеней свободы d .

б. Предположим, что в задаче 12.1 принятое значение $\rho_{\text{пр}}$ плотности известно и мы решили проверить гипотезу о том, что результаты подчиняются нормальному распределению с центром $\rho_{\text{пр}}$. В случае этого испытания сколько имеется связей и каково число степеней свободы?

*12.6 (разд. 12.4). По данным задачи 12.1 вычислите приведенное значение $\tilde{\chi}^2$. Если бы результаты измерений были распределены нормально, какова была бы вероятность получения значения $\tilde{\chi}^2$, столь же большого или даже большего? На 5 %-ном уровне значимости можете ли вы отвергнуть гипотезу о том, что результаты измерений были распределены нормально? А на 1 %-ном уровне? (Нужные вероятности найдите в приложении Г.)

12.7 (разд. 12.4). В задаче 12.2 можете ли вы отвергнуть предположение о нормальном распределении на 5 %-ном или на 1 %-ном уровне значимости? (Нужные вероятности найдите в приложении Г.)

Таблица 12.12

Результат	Бин k	Числа наблюдений O_k
Ни одной шестерки	1	217
Одна шестерка	2	148
Две или три шестерки	3	35

Таблица 12.13

Сумма	2	3	4	5	6	7	8	9	10	11	12
Числа наблюдений	6	14	23	35	57	50	44	49	39	27	16

*12.8 (разд. 12.5). Пара игральных костей бросается 360 раз, и для каждого броска регистрируется суммарное число выпавших очков. Возможные значения суммы равны 2, 3, ..., 12, и их числа наблюдений приведены в табл. 12.13.

Рассчитайте вероятность для каждой суммы и, следовательно, ожидаемые числа ее появления (предполагая, что кости подлинные). Вычислите χ^2 , d и $\tilde{\chi}^2 = \chi^2/d$. Предполагая, что кости подлинные, найдите, какова вероятность получения значения $\tilde{\chi}^2$, равного найденному, или большего. На 5 %-ном уровне значимости можете ли вы отвергнуть гипотезу о том, что кости подлинные? А на 1 %-ном уровне? (Нужные вероятности найдите в приложении Г.)

12.9 (разд. 12.5). Для задачи 12.3 найдите значение $\tilde{\chi}^2$. Можем ли мы сделать вывод о том, что игральная кость утяжелена на 5 %-ном уровне значимости? А на 1 %-ном уровне? (Нужные вероятности найдите в приложении Г.)

*12.10 (разд. 12.5). Чему равно значение $\tilde{\chi}^2$ в задаче 12.4? Если игральные кости действительно подлинные, то какова вероятность получения значения $\tilde{\chi}^2$, равного найденному, или большего? Объясните, позволяют ли результаты предположить, что кости утяжелены? (Нужные вероятности найдите в приложении Г.)

12.11 (разд. 12.5). Вычислите χ^2 по данным задачи 11.5, предполагая, что результаты наблюдений должны быть распределены по закону Пуассона со средним числом отсчетов $\mu = 3$. (Сгруппируйте все данные для $v \geq 6$ в один бин.) Сколько имеется степеней свободы? (Не забудьте, что число μ было дано заранее, а не вычислялось по данным.) Чему равно $\tilde{\chi}^2$? Согласуются ли данные с ожидаемым распределением Пуассона? (Нужные вероятности найдите в приложении Г.)

*12.12 (разд. 12.5).

а. Утверждается, что в некотором образце радиоактивного вещества происходят в среднем два распада в минуту. Чтобы проверить это, студент измеряет число распадов в 40 отдельных одноминутных интервалах и получает результаты, приведенные в табл. 12.14.

Если распады происходят в соответствии с распределением Пуассона с $\mu = 2$, то какие ожидаемые числа наблюдений обнаружит студент? (Сгруппируйте все данные с $v \geq 3$ в один бин.) Вычислите χ^2 , d и $\tilde{\chi}^2 = \chi^2/d$ (не забудьте, что число μ не вычислялось из данных). На 5 %-ном уровне значимости отвергаете ли вы гипотезу о том, что в образце происходят распады в соответствии с законом Пуассона с $\mu = 2$?

б. Студент замечает, что фактически среднее значение результатов равно $\bar{v} = 1,35$, и поэтому решает проверить, подчиняются ли его данные распределению Пуассона с $\mu = 1,35$. Каковы в этом случае значения d и $\tilde{\chi}^2$? Согласуются ли данные с этой новой гипотезой?

Таблица 12.14

Число распадов v	0	1	2	3	4	5 или больше
Числа наблюдений	11	12	11	4	2	0

***12.13** (разд. 12.5). В гл. 10 мы обсудили метод испытания гипотез на основе биномиального распределения. Мы рассматривали n попыток, каждую с двумя исходами: выигрышем (с вероятностью p) и проигрышем (с вероятностью $1 - p$). Затем мы проверяли, совместимо ли полученное число выигрышей v с некоторым принятым значением p . Если рассматриваемые числа достаточно велики, мы можем рассматривать эту же самую задачу с помощью критерия χ^2 , используя два бина: $k = 1$ в случае выигрыша и $k = 2$ в случае проигрыша и одну степень свободы. В последующем мы будем использовать оба метода и сравним их результаты. Если числа велики, то вы получите отличное согласие, если же числа малы, то согласие будет менее удовлетворительным, но все же достаточно хорошим, чтобы утверждать, что метод χ^2 очень полезен и чувствителен.

- а. Производитель супов считает, что он может вводить различные добавки из теста в свои порошковые куриные супы, не оказывая влияния на их спрос. Чтобы проверить эту гипотезу, он изготавливает 16 пакетиков, помеченных «рецепт X» и содержащих новую смесь, и 16 пакетиков, помеченных «рецепт Y» и содержащих старый состав. Затем он рассылает по одному пакету каждого состава 16 дегустаторам и спрашивает их, какой состав предпочтительнее. Если его гипотеза верна, то мы должны были бы ожидать, что восемь дегустаторов предпочтут X и восемь — Y. Фактически число дегустаторов, которые предпочли X, оказалось равным $v = 11$. Вычислите χ^2 и вероятность получения значения не меньшего, чем вычисленное. Указывает ли испытание на значимую разницу между двумя видами смесей? Затем вычислите соответствующие вероятности точно с помощью биномиального распределения и сравните ваши результаты. Заметьте, что критерий χ^2 использует отклонения в обе стороны от ожидаемых чисел. Следовательно, для такого сравнения вы должны вычислять «двухстороннюю» вероятность для значений v , отличающихся от восьми на три или более в любую сторону, т. е. как для $v = 11, 12, \dots, 16$, так и для $v = 5, 4, \dots, 1$.
- б. Повторите задание «а» для следующего испытания, когда производитель изготавливает по 400 пакетиков каждой смеси, а число предпочитающих X равно 225. (При расчете биномиальных вероятностей используйте гауссову аппроксимацию.)
- в. В случае «а» числа довольно малы, так что критерий χ^2 оказался весьма приближенным. (Он давал вероятность 14 % по сравнению с правильным значением 21,0 %.) В случае одной степени свободы мы можем немного улучшить критерий χ^2 , заменяя его «модифицированным χ^2 », определяемым как

$$\text{модифицированный } \chi^2 = \sum_{k=1}^2 \frac{(|O_k - E_k| - 1/2)^2}{E_k}.$$

Вычислите «модифицированный χ^2 » для данных задания «а» и покажите, что использование его значения (вместо обычного χ^2) при нахождении вероятности в таблице приложения Г даст более точное приближение¹⁾.

12.14 (разд. 12.5). Критерий χ^2 может быть использован для проверки того, насколько хорошо набор измерений (x_i, y_i) двух переменных аппроксимируется ожидаемой зависимостью $y = f(x)$ при условии, что по-

¹⁾ Мы не обосновали здесь понятие «модифицированный χ^2 », но этот пример действительно показывает его преимущества. Подробности см в книге Alder H. L., Roessler E. B., Introduction to Probability and Statistics, Freeman, 1977, p. 263.

Таблица 12.15

x (погрешность пренебрежимо мала)	1	2	3	4	5
y (для всех значений ± 4)	60	56	71	66	86

грешности хорошо известны. Предположим, что y и x связаны линейной зависимостью

$$y = f(x) = A + Bx. \quad (12.19)$$

(Например, y может быть длиной металлического прута, а x — его температурой.) Предположим, что, согласно теоретическим предсказаниям, A и B имеют значения, равные $A = 50$ и $B = 6$, и что пять измерений x и y дали результаты, приведенные в табл. 12.15.

Погрешность, приведенная для y , — стандартное отклонение, т.е. все пять результатов измерений y имеют одно и то же стандартное отклонение $\sigma = 4$. Составьте таблицу полученных и ожидаемых значений y_i и вычислите χ^2 по формуле

$$\chi^2 = \sum_i^5 \left(\frac{y_i - f(x_i)}{\sigma} \right)^2.$$

Поскольку ни один из параметров не оценивается по данным, связей нет и, следовательно, будет пять степеней свободы. Вычислите $\tilde{\chi}^2$ и по таблице приложения Г определите вероятность получения значения $\tilde{\chi}^2$ такой величины, предполагая, что y действительно удовлетворяет зависимости (12.19). На 5 %-ном уровне отвергнете ли вы ожидаемую зависимость (12.19)? (Если бы постоянные A и B не были известны заранее, то пришлось бы рассчитать их по данным методом наименьших квадратов. Затем можно было бы продолжать, как и прежде, но в этом случае было бы уже только три степени свободы.)

Приложение А

Интеграл ошибок. I

Если результат измерения непрерывной переменной x подвержен влиянию множества небольших случайных ошибок, то ожидаемое распределение результатов будет нормальным, или гауссовым, распределением

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

где X есть истинное значение x , а σ — стандартное отклонение.

Интеграл от функции нормального распределения $\int_a^b f_{X, \sigma}(x) dx$ называется *интегралом ошибок* и определяет вероятность того, что результат измерения окажется между $x = a$ и $x = b$,

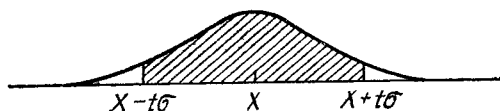
$$P(a \leq x \leq b) = \int_a^b f_{X, \sigma}(x) dx.$$

В табл. А приведены значения этого интеграла для $a = X - t\sigma$ и $b = X + t\sigma$. Они определяют вероятность того, что результат окажется в пределах t стандартных отклонений в любую сторону от X :

$$\begin{aligned} P(\text{в пределах } t\sigma) &= P(X - t\sigma \leq x \leq X + t\sigma) = \\ &= \int_{X-t\sigma}^{X+t\sigma} f_{X, \sigma}(x) dx = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-t}^t e^{-z^2/2} dz. \end{aligned}$$

Эту функцию иногда обозначают $\text{erf}(t)$, но такое обозначение используется также и для несколько отличающейся функции.

Таблица А.
Вероятность,
выраженная
в процентах,
 P (в пределах $t\sigma$) =
 $\int_{x-t\sigma}^{x+t\sigma} f_{X, \sigma}(x) dx$,
как функция t



t	0,09	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,00	0,80	1,60	2,39	3,19	3,99	4,78	5,58	6,38	7,17
0,1	7,97	8,76	9,55	10,34	11,13	11,92	12,71	13,50	14,28	15,07
0,2	15,85	16,63	17,41	18,19	18,97	19,74	20,51	21,28	22,05	22,82
0,3	23,58	24,34	25,10	25,86	26,61	27,37	28,12	28,86	29,61	30,35
0,4	31,08	31,82	32,55	33,28	34,01	34,73	35,45	36,16	36,88	37,59
0,5	38,29	38,99	39,69	40,39	41,08	41,77	42,45	43,13	43,81	44,48
0,6	45,15	45,81	46,47	47,13	47,78	48,43	49,07	49,71	50,35	50,98
0,7	51,61	52,23	52,85	53,46	54,07	54,67	55,27	55,87	56,46	57,05
0,8	57,63	58,21	58,78	59,35	59,91	60,47	61,02	61,57	62,11	62,65
0,9	63,19	63,72	64,24	64,76	65,28	65,79	66,29	66,80	67,29	67,78
1,0	68,27	68,75	69,23	69,70	70,17	70,63	71,09	71,54	71,99	72,43
1,1	72,87	73,30	73,73	74,15	74,57	74,99	75,40	75,80	76,20	76,60
1,2	76,99	77,37	77,75	78,13	78,50	78,87	79,23	79,59	79,95	80,29
1,3	80,64	80,98	81,32	81,65	81,98	82,30	82,62	82,93	83,24	83,55
1,4	83,85	84,15	84,44	84,73	85,01	85,29	85,57	85,84	86,11	86,38
1,5	86,64	86,90	87,15	87,40	87,64	87,89	88,12	88,36	88,59	88,82
1,6	89,04	89,26	89,48	89,69	89,90	90,11	90,31	90,51	90,70	90,90
1,7	91,09	91,27	91,46	91,64	91,81	91,99	92,16	92,33	92,49	92,65
1,8	92,81	92,97	93,12	93,28	93,42	93,57	93,71	93,85	93,99	94,12
1,9	94,26	94,39	94,51	94,64	94,76	94,88	95,00	95,12	95,23	95,34
2,0	95,45	95,56	95,66	95,76	95,86	95,96	96,06	96,15	96,25	96,34
2,1	96,43	96,51	96,60	96,68	96,76	96,84	96,92	97,00	97,07	97,15
2,2	97,22	97,29	97,36	97,43	97,49	97,56	97,62	97,68	97,74	97,80
2,3	97,86	97,91	97,97	98,02	98,07	98,12	98,17	98,22	98,27	98,32
2,4	98,36	98,40	98,45	98,49	98,53	98,57	98,61	98,65	98,69	98,72
2,5	98,76	98,79	98,83	98,86	98,89	98,92	98,95	98,98	99,01	99,04
2,6	99,07	99,09	99,12	99,15	99,17	99,20	99,22	99,24	99,26	99,29
2,7	99,31	99,33	99,35	99,37	99,39	99,40	99,42	99,44	99,46	99,47
2,8	99,49	99,50	99,52	99,53	99,55	99,56	99,58	99,59	99,60	99,61
2,9	99,63	99,64	99,65	99,66	99,67	99,68	99,69	99,70	99,71	99,72

Продолжение табл. А

t	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
3,0	99,73	—	—	—	—	—	—	—	—	—
3,5	99,95	—	—	—	—	—	—	—	—	—
4,0	99,994	—	—	—	—	—	—	—	—	—
4,5	99,9993	—	—	—	—	—	—	—	—	—
5,0	99,99994	—	—	—	—	—	—	—	—	—

Вероятность того, что результат измерения окажется *вне* этого же интервала, можно найти с помощью вычитания:

$$P(\text{вне } t\sigma) = 100 \% - P(\text{в пределах } t\sigma).$$

Дополнительные сведения вы найдете в разд. 5.4 и в приложении Б.

Приложение Б

Интеграл ошибок. II

В некоторых расчетах удобной формой интеграла ошибок является следующая:

$$Q(t) = \int_X^{X+t\sigma} f_{X,\sigma}(x) dx = \\ = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-z^2/2} dz.$$



(Этот интеграл, конечно, равен половине значения интеграла, табулированного в приложении А.) Вероятность $P(a \leq x \leq b)$ того, что результат измерения окажется в любом интервале $a \leq x \leq b$, можно найти по значению $Q(t)$ с помощью одного вычитания или сложения. Например,

$$P(X + \sigma \leq x \leq X + 2\sigma) = \\ = Q(2) - Q(1).$$



Аналогично

$$P(X - 2\sigma \leq x \leq X + \sigma) = \\ = Q(2) + Q(1).$$



Вероятность того, что результат измерения окажется больше, чем любое $X + t\sigma$, равна $0,5 - Q(t)$. Например,

$$P(x \geq X + \sigma) = \\ = 50\% - Q(1).$$



Приложение В

Вероятности коэффициентов корреляции

Степень, с которой N точек $(x_1, y_1), \dots, (x_N, y_N)$ аппроксимируются прямой линией, определяется коэффициентом линейной корреляции:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2]^{1/2}},$$

который всегда лежит в интервале $-1 \leq r \leq 1$. Значения r , близкие к ± 1 , означают высокую степень линейной корреляции; значения, близкие к 0, указывают на небольшую корреляцию или на ее отсутствие.

Количественная мера аппроксимации может быть получена с помощью табл. В. Для любого определенного r_0 $P_N(|r| \geq |r_0|)$ есть вероятность того, что результаты N измерений двух некоррелированных переменных будут иметь коэффициент корреляции r , не меньший чем r_0 . Таким образом, если мы получим коэффициент r_0 , для которого вероятность $P_N(|r| \geq |r_0|)$ мала, то невероятно, чтобы наши переменные были некоррелированными, т. е. корреляция существует. В частности, если $P_N(|r| \geq |r_0|) \leq 5\%$, корреляция называется *значимой*, если эта вероятность меньше 1%, то корреляция называется *высокозначимой*.

Например, вероятность того, что результаты 20 измерений ($N = 20$) двух некоррелированных переменных дадут $|r| \geq 0,5$, определяется таблицей в 2,5%. Таким образом, если бы результаты 20 измерений дали $r = 0,5$, то у нас было бы *значимое* доказательство линейной корреляции между двумя переменными. Дополнительные сведения вы найдете в разд. 9.3—9.5.

Значения, приведенные в табл. В, были вычислены по формуле

$$P_N(|r| \geq |r_0|) = \frac{2\Gamma[(N-1)/2]}{\sqrt{\pi} \Gamma[(N-2)/2]} \int_{r_0}^1 (1-r^2)^{(N-4)/4} dr.$$

Для справок, см., например, Pugh E. M. Winslow G. H., The Analysis of Physical Measurements, 9, Addison-Wesley, 1966, Sect. 12.8.

Таблица В. Выраженная в процентах вероятность $P_N(|r| \geq |r_0|)$ того, что результаты N измерений двух некоррелированных переменных дадут коэффициент корреляции $|r| \geq |r_0|$, как функция N и r_0 . (Прочерки указывают на вероятности, которые меньше 0,05%.)

N	r_0										
	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
3	100	94	87	81	74	67	59	51	41	29	0
4	100	90	80	70	60	50	40	30	20	10	0
5	100	87	75	62	50	39	28	19	10	3,7	0
6	100	85	70	56	43	31	21	12	5,6	1,4	0
7	100	83	67	51	37	25	15	8,0	3,1	0,6	0
8	100	81	63	47	33	21	12	5,3	1,7	0,2	0
9	100	80	61	43	29	17	8,8	3,6	1,0	0,1	0
10	100	78	58	40	25	14	6,7	2,4	0,5	—	0
11	100	77	56	37	22	12	5,1	1,6	0,3	—	0
12	100	76	53	34	20	9,8	3,9	1,1	0,2	—	0
13	100	75	51	32	18	8,2	3,0	0,8	0,1	—	0
14	100	73	49	30	16	6,9	2,3	0,5	0,1	—	0
15	100	72	47	28	14	5,8	1,8	0,4	—	—	0
16	100	71	46	26	12	4,9	1,4	0,3	—	—	0
17	100	70	44	24	11	4,1	1,1	0,2	—	—	0
18	100	69	43	23	10	3,5	0,8	0,1	—	—	0
19	100	68	41	21	9,0	2,9	0,7	0,1	—	—	0
20	100	67	40	20	8,1	2,5	0,5	0,1	—	—	0
25	100	63	34	15	4,8	1,1	0,2	—	—	—	0
30	100	60	29	11	2,9	0,5	—	—	—	—	0
35	100	57	25	8,0	1,7	0,2	—	—	—	—	0
40	100	54	22	6,0	1,1	0,1	—	—	—	—	0
45	100	51	19	4,5	0,6	—	—	—	—	—	0
	0	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	
50	100	73	49	30	16	8,0	3,4	1,3	0,4	0,1	
60	100	70	45	25	13	5,4	2,0	0,6	0,2	—	
70	100	68	41	22	9,7	3,7	1,2	0,3	0,1	—	
80	100	66	38	18	7,5	2,5	0,7	0,1	—	—	
90	100	64	35	16	5,9	1,7	0,4	0,1	—	—	
100	100	62	32	14	4,6	1,2	0,2	—	—	—	

Приложение Г

Вероятности для χ^2

Если результаты серии измерений распределены по k бинам, где $k = 1, \dots, n$, то через O_k мы обозначим число наблюдений, попавших в бин k . *Ожидаемое* число (полученное на основе некоторого предположенного или ожидаемого распределения) результатов в бине k обозначается E_k . Степень, с которой результаты наблюдений аппроксимируются предполагаемым распределением, характеризуется приведенным значением $\tilde{\chi}^2$, определяемым как

$$\tilde{\chi}^2 = \frac{1}{d} \sum_{k=1}^n \frac{(O_k - E_k)^2}{E_k},$$

где d есть число степеней свободы, равное $d = n - c$, а c — число связей (см. разд. 12.3). Ожидаемое среднее значение $\tilde{\chi}^2$ равно 1. Если $\tilde{\chi}^2 \gg 1$, то результаты наблюдений не согласуются с предполагаемым распределением; если $\tilde{\chi}^2 \leq 1$, то согласие удовлетворительное.

Этот критерий становится количественным благодаря вероятностям, приведенным в табл. Г. Пусть $\tilde{\chi}_o^2$ обозначает значение $\tilde{\chi}^2$, фактически полученное в эксперименте с d степенями свободы. Число $P_d(\tilde{\chi}^2 \geq \tilde{\chi}_o^2)$ есть вероятность получения значения $\tilde{\chi}^2$ не меньшего, чем полученное $\tilde{\chi}_o^2$, если результаты измерений действительно распределены в соответствии с предположенным законом. Таким образом, если $P_d(\tilde{\chi}^2 \geq \tilde{\chi}_o^2)$ велика, то полученное и ожидаемое распределения согласуются; если эта вероятность мала, то они, по-видимому, различаются. В частности, если $P_d(\tilde{\chi}^2 \geq \tilde{\chi}_o^2)$ меньше 5%, мы говорим, что расхождение *значимо*, и отвергаем предположенное распределение на 5%-ном уровне. Если эта вероятность меньше 1%, расхождение называется *высокозначимым* и мы отвергаем предположенное распределение на 1%-ном уровне.

Например, предположим, что мы получили для приведенного значения $\tilde{\chi}^2 2,6$ (т. е. $\tilde{\chi}_o^2 = 2,6$) в эксперименте с шестью степенями свободы ($d = 6$). В соответствии с табл. Г

Таблица Г. Выраженная в процентах вероятность $P_d(\tilde{\chi}^2 \geq \tilde{\chi}_o^2)$ получения значения $\tilde{\chi}^2 \geq \tilde{\chi}_o^2$ в эксперименте с d степенями свободы как функция d и $\tilde{\chi}_o^2$. (Прочерки указывают на значения вероятности, которые меньше 0,05%.)

	χ^2_o															
d	0	0,5	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0	5,5	6,0	8,0	10,0	
1	100	48	32	22	16	11	8,3	6,1	4,6	3,4	2,5	1,9	1,4	0,5	0,2	
2	100	61	37	22	14	8,2	5,0	3,0	1,8	1,1	0,7	0,4	0,2	—	—	
3	100	68	39	21	11	5,8	2,9	1,5	0,7	0,4	0,2	0,1	—	—	—	
4	100	74	41	20	9,2	4,0	1,7	0,7	0,3	0,1	0,1	—	—	—	—	
5	100	78	42	19	7,5	2,9	1,0	0,4	0,1	—	—	—	—	—	—	
	0	0,2	0,4	0,6	0,8	1,0	1,2	1,4	1,6	1,8	2,0	2,2	2,4	2,6	2,8	3,0
1	100	65	53	44	37	32	27	24	21	18	16	14	12	11	9,4	8,3
2	100	82	67	55	45	37	30	25	20	17	14	11	9,1	7,4	6,1	5,0
3	100	90	75	61	49	39	31	24	19	14	11	8,6	6,6	5,0	3,8	2,9
4	100	94	81	66	52	41	31	23	17	13	9,2	6,6	4,8	3,4	2,4	1,7
5	100	96	85	70	55	42	31	22	16	11	7,5	5,1	3,5	2,3	1,6	1,0
6	100	98	88	73	57	42	30	21	14	9,5	6,2	4,0	2,5	1,6	1,0	0,6
7	100	99	90	76	59	43	30	20	13	8,2	5,1	3,1	1,9	1,1	0,7	0,4
8	100	99	92	78	60	43	29	19	12	7,2	4,2	2,4	1,4	0,8	0,4	0,2
9	100	99	94	80	62	44	29	18	11	6,3	3,5	1,9	1,0	0,5	0,3	0,1
10	100	100	95	82	63	44	29	17	10	5,5	2,9	1,5	0,8	0,4	0,2	0,1
11	100	100	96	83	64	44	28	16	9,1	4,8	2,4	1,2	0,6	0,3	0,1	0,1
12	100	100	96	84	65	45	28	16	8,4	4,2	2,0	0,9	0,4	0,2	0,1	—
13	100	100	97	86	66	45	27	15	7,7	3,7	1,7	0,7	0,3	0,1	0,1	—
14	100	100	98	87	67	45	27	14	7,1	3,3	1,4	0,6	0,2	0,1	—	—
15	100	100	98	88	68	45	26	14	6,5	2,9	1,2	0,5	0,2	0,1	—	—
16	100	100	98	89	69	45	26	13	6,0	2,5	1,0	0,4	0,1	—	—	—
17	100	100	99	90	70	45	25	12	5,5	2,2	0,8	0,3	0,1	—	—	—
18	100	100	99	90	70	46	25	12	5,1	2,0	0,7	0,2	0,1	—	—	—
19	100	100	99	91	71	46	25	11	4,7	1,7	0,6	0,2	0,1	—	—	—
20	100	100	99	92	72	46	24	11	4,3	1,5	0,5	0,1	—	—	—	—
22	100	100	99	93	73	46	23	10	3,7	1,2	0,4	0,1	—	—	—	—
24	100	100	100	94	74	46	23	9,2	3,2	0,9	0,3	0,1	—	—	—	—
26	100	100	100	95	75	46	22	8,5	2,7	0,7	0,2	—	—	—	—	—
28	100	100	100	95	76	46	21	7,8	2,3	0,6	0,1	—	—	—	—	—
30	100	100	100	96	77	47	21	7,2	2,0	0,5	0,1	—	—	—	—	—

вероятность получения $\tilde{\chi}^2 \geq 2,6$ составляет 1,6%, если результаты измерений распределены в соответствии с предположенным законом. Таким образом, на 5%-ном уровне (но не на 1%-ном) мы отвергли бы предположенное распределение. Дополнительные сведения вы найдете в гл. 12.

Значения, представленные в табл. Г, были рассчитаны по формуле

$$P_d(\tilde{\chi}^2 \geq \tilde{\chi}_0^2) = \frac{2}{2^{d/2} \Gamma(d/2)} \int_{\tilde{\chi}_0}^{\infty} x^{d-1} e^{-x/2} dx.$$

Для справок см., например, *Pugh E. M., Winslow G. H., The Analysis of Physical Measurements, Addison-Wesley, 1966. Sect 12.5.*

БИБЛИОГРАФИЯ

Следующие книги я счел полезными. Они приводятся в порядке, приблизительно соответствующем их математическому уровню и степени охвата материала.

Исключительно ясное введение в статистические методы, в котором автор сумел обойтись без помощи математического анализа, дано в книге *Lacy O. L.*, *Statistical Methods in Experimentation*, MacMillan, 1953.

Книга по статистике более высокого уровня, но также очень понятная и в которой тоже не используются методы математического анализа, *Altder H. L.*, *Roessler E. B.*, *Introduction to Probability and Statistics*, Freeman, 1977.

Следующие три книги написаны приблизительно на том же уровне, что и данная книга, и охватывают в основном тот же материал:

Baird D. C., *Experimentation; An Introduction to Measurement Theory and Experiment Design*, Prentice Hall, 1962.

Barford N. C., *Experimental Measurements; Precision, Error, and Truth*, Addison-Wesley, 1967.

Young H. D., *Statistical Treatment of Experimental Data*, McGraw-Hill, 1962.

Многочисленные другие вопросы и выводы формул можно найти в следующих книгах более высокого уровня:

Bevington P. R., *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, 1969.

Meyer S. L., *Statistical Treatment of Experimental Data*, John Wiley, 1975.

Pugh E. M., *Winslow G. H.*, *The Analysis of Physical Measurements*, Addison-Wesley, 1966.

ЛИТЕРАТУРА, ДОБАВЛЕННАЯ ПРИ ПЕРЕВОДЕ

Следующие книги предназначены для студентов младших курсов.

1. Деденко Л. Г., Керженцев В. В. Математическая обработка и оформление результатов эксперимента. — М.: Изд-во МГУ, 1977.

2. Соловьев В. А., Яхонтова В. Е. Элементарные методы обработки результатов. — Л.: Изд-во ЛГУ, 1977.

3. Зайдель А. Н. Ошибки измерений физических величин. — Л.: Наука, 1974.

4. Кассандрова О. Н., Лебедев В. В. Обработка результатов измерений. — М.: Наука, 1970.

5. Свейников А. А. Основы теории ошибок. — Л.: Изд-во ЛГУ, 1972.

6. Сквайрс Дж. Практическая физика. — М.: Мир, 1971.

Следующие книги полезны для старшекурсников

7. Пытьев Ю. П., Шишмарев И. А. Курс теорий вероятностей и математической статистики для физиков. — М.: Изд-во МГУ, 1983.

8. Гришин В. К. Статистические методы анализа и планирования экспериментов. — М.: Изд-во МГУ, 1975.

ОТВЕТЫ К ИЗБРАННЫМ ЗАДАЧАМ

Замечание о числе значащих цифр: небольшие различия в последней значащей цифре могут быть обусловлены использованием различных способов округления, и они (различия) обычно не важны. В случае задач из гл. 2 и 3 приведенные ниже погрешности были получены наиболее грубым возможным способом, когда производится округление до одной значащей цифры на каждом этапе вычислений. В небольшом числе случаев, когда более строгие вычисления дают другой ответ, точный ответ, соответствующим образом округленный после всех вычислений, указывается в скобках. В случае задач из гл. 4—12 все ответы были получены с помощью калькулятора (с десятью значащими цифрами) и округлялись после всех вычислений.

Глава 2

- 2.2. а. $5,03 \pm 0,04$ м. б. В этом случае имеется настоятельная необходимость удерживать в вычислениях лишнюю цифру, и тогда получается $19,5 \pm 1$ с. в. $(-3,2 \pm 0,3) \cdot 10^{-19}$ Кл. г. $(0,56 \pm 0,07) \cdot 10^{-6}$ м. д. $(3,27 \pm 0,04) \cdot 10^3$ г·см/с.
- 2.3. а. Вероятно, единственный надежный вывод на этом этапе — это ответ $1,9 \pm 0,1$ г/см³. б. Различие равно $0,05$ г/см³, и оно не является значимым.
- 2.5. В столбце $(L - L')$ должны стоять цифры: $0,3 \pm 0,9$; $-0,6 \pm 1,5$; $-2,2 \pm 2$ (следует округлить до -2 ± 2); 1 ± 4 ; 1 ± 4 ; -4 ± 4 . Разность $(L - L')$ теоретически должна равняться нулю. Во всех случаях, кроме одного, измеренное значение меньше соответствующей погрешности, и даже в этом единственном случае $(-2,2 \pm 2)$ измеренное значение только незначительно больше. Таким образом, полученные значения согласуются с ожидаемым нулевым значением.
- 2.8. а. Поскольку может быть найдена прямая линия (как на рис. А2.8), которая проходит через 0 и через все черточки ошибок, то данные согласуются с предсказанием $v^2 \sim h$. б. Наклон наилучшей аппроксимирующей линии $\approx 18,4$; наклон наиболее крутой возможной линии $\approx 20,4$; наклон наиболее пологой возможной линии $\approx 16,4$. Таким образом, наклон равен $18,4 \pm 2$ м/с² (или, возможно, 18 ± 2), что согласуется с ожидаемым значением $19,6$ м/с². Вопрос: следует ли настаивать, чтобы такие линии проводились через 0, или нет? Ответ зависит от методов измерений. В данном случае мы не провели все линии через 0 и, следовательно, с избытком оценили погрешность.
- 2.9. а. На рис. А2.9, а, где изображены нулевые отметки обеих шкал, не видно, изменяется ли T с A . На рис. А2.9, б, где масштаб вертикальной шкалы сильно увеличен, ясно видно, что T на самом деле изменяется с A . Очевидно, следует тщательно анализировать, какой выбор осей координат будет наиболее подходящим для данного конкретного случая.

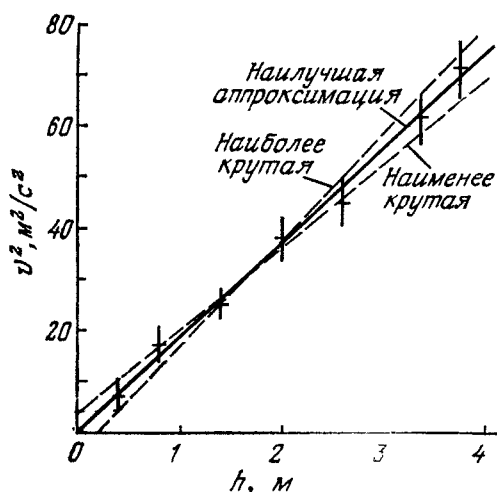


Рис. А2.8.

б. Если перечертить оба графика для случая, когда ошибки равны 0,3 с (вверх и вниз), то было бы ясно, что T не зависит от A .

2.12. а. Ответы для $v_1 - v_i$ равны $4,0 \pm 0,3$ и $0,6 \pm 0,4$ см/с. б. Погрешности, выраженные в процентах, равны 8 и 70 %.

2.14. Ответ	Погрешность, выраженная в %	Абсолютная погрешность
а) 292 см ²	3	9 см ² (или 7 "точно")
б) 270 см · с	10	30 см · с
в) 1,2 кг · м	10	0,1 кг · м

2.15. а. $q_{\text{наил}} = 10 \cdot 20 = 200$;
(наибольшее вероятное значение q) = $11 \cdot 21 = 231$; (наименьшее вероятное значение q) = $9 \cdot 19 = 171$. Правило (2.27) дает $q = 200 \pm 30$, что хорошо согласуется с полученными значениями.

б. (наибольшее вероятное значение) = $18 \cdot 35 = 630$; (наименьшее вероятное значение) = $2 \cdot 5 = 10$. Правило (2.27) дает $q = 200 \pm 300$ (т. е. $q_{\text{макс}} = 500$; $q_{\text{мин}} = -100$). Причина столь сильного расхождения состоит в том, что правило (2.27) применимо только тогда, когда относительные погрешности малы по сравнению с единицей. Это условие (которое обычно реализуется на практике) в данном случае нарушено.

Глава 3

- 3.1. а. $32 \pm \sqrt{32} \approx 32 \pm 6$. б. $786 \pm \sqrt{786} \approx 790 \pm 30$.
в. 16 ± 3 для А; $13,1 \pm 0,5$ для Б. Заметьте, что ответы А и Б согласуются, но ответ Б более ценен из-за меньшей погрешности.
- 3.3. а. 3 ± 7 . б. 40 ± 20 . в. $0,5 \pm 0,1$. г. 63 ± 6 .
- 3.4. а. $0,48 \pm 0,02$ с (или 4 %). б. $0,470 \pm 0,005$ с (или 1 %). в. Нет. Во-первых, в конце концов маятник остановится, если его не подталкивать. Если даже его и подталкивать, другие эффекты рано или поздно станут важными и обрекут на неудачу наше стремление ко

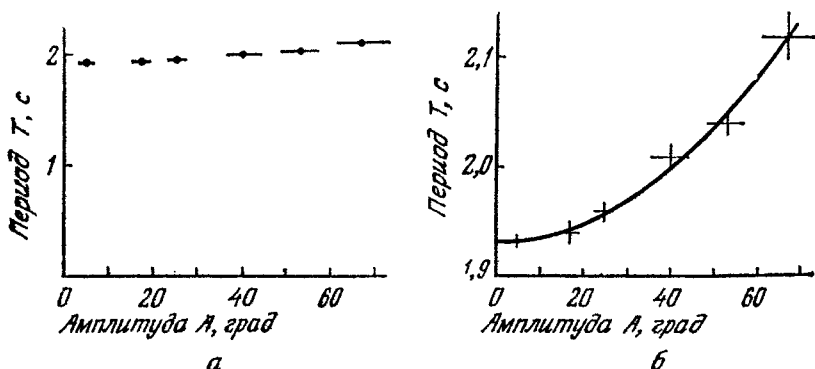


Рис. А2.9.

все большей и большей точности. Например, если мы будем проводить измерения в течение нескольких часов, погрешность секундомера может стать ограничивающим фактором; наконец, период t может изменяться в зависимости от температуры, влажности и т. д.

- 3.6. Глубина = 40 ± 10 м. (Более точный расчет дает 44 ± 15 , что можно не округлять.)

3.8. “ошибки складываются” “ошибки складываются квадратично”

$a + b$	80 ± 8	80 ± 6
$a + c$	90 ± 6	90 ± 5
$a + d$	58 ± 5	58 ± 5

- 3.10. а. $0,70 \pm 0,05$ МэВ. б. $0,40 \pm 0,02$ МэВ.

- 3.11. а. $\sin \theta = 0,82 \pm 0,02$. (Не забудьте, что $\delta \theta$ необходимо выразить в радианах при использовании $\delta(\sin \theta) = |\cos \theta| \delta \theta$.)

б. $f_{\text{наил}} = e^{a_{\text{наил}}}$; $\delta f = f_{\text{наил}} \delta a$; $e^a = 20 \pm 2$.

в. $f_{\text{наил}} = \ln a_{\text{наил}}$; $\delta f = \delta a / a_{\text{наил}}$; $\ln a = 1,10 \pm 0,03$.

- 3.14. $n = 1,66 \pm 20\%$, $1,52 \pm 9\%$; $1,54 \pm 6\%$; $1,58 \pm 3\%$; $1,53 \pm 2\%$. С увеличением угла величина $\delta n/n$ уменьшается главным образом из-за того, что абсолютная погрешность постоянна, поэтому относительная погрешность будет меньше, когда угол будет больше.

- 3.16. а. 1 и 1. б. y и x . в. $2xy^3$ и $3x^2y^2$.

- 3.17. в.

Левая сторона (ЛС) = $(x + u)^2 (y + v)^3 = (x^2 + 2xu + u^2) (y^3 + 3y^2v + 3yv^2 + v^3) = x^2y^3 + 2xy^3u + 3x^2y^2v +$
 + (члены, содержащие u^2 , uv , v^2 и более высокие степени).

Правая сторона (ПС) = $x^2y^3 + 2xy^3u + 3x^2y^2v$. Следовательно, ЛС $\approx$ ПС, когда u и v малы.

- 3.19. а. Правильный ответ равен $\delta q = 0,005$, но вычисления методом «шаг за шагом» дают $\delta q = 0,1$. б. $\delta q = 0,1$ при любом методе. В случае «а» числа таковы, что небольшая ошибка в x изменяет $x + y$ и $x + z$ почти на одинаковый множитель, и поэтому в отношении $(x + y)/(x + z)$ она пропадает; метод «шаг за шагом» не учитывает

этого сокращения ошибок. В случае «б» ошибка в x увеличивает $(x + y)$, но уменьшает $(x + z)$ или наоборот, и поэтому она не компенсируется в q .

Глава 4

4.1. $\bar{x} = 7,2$; $\sigma_x = 1,5$ по определению (4.9) или 1,3 — по (4.6).

$$4.3. \quad d = (1/N) \sum d_i = (1/N) \sum (x_i - \bar{x}) = (1/N) \sum x_i - \\ - (1/N) N\bar{x} = \bar{x} - \bar{x} = 0.$$

Если какой-то из этих этапов не ясен, распишите суммы в явном виде, например как $\sum d_i = d_1 + d_2 + \dots + d_N$ и т. д.

$$4.4. \quad \sum (x_i - \bar{x})^2 = \sum (x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \\ = \sum x_i^2 - 2\bar{x} \sum x_i + N\bar{x}^2 = \\ = \sum x_i^2 - 2\bar{x} N\bar{x} + N\bar{x}^2 = \sum x_i^2 - N\bar{x}^2.$$

(И в этом случае, если у вас есть какие-либо сомнения, распишите суммы в явном виде.)

4.7. а. $\bar{t} = 8,149$ с, $\sigma_t = 0,039$ с. б. Вне $\bar{t} \pm \sigma_t$ мы ожидаем 30 %, или 9 результатов измерений, а получили 8. Вне $\bar{t} \pm 2\sigma_t$ мы ожидаем 5 %, или 1,5, а получили 2.

4.9. (Итоговый результат для t) $= \bar{t} \pm \sigma_t = 8,149 \pm 0,007$ с.

4.11. $\bar{A} = 1221,2$ мм². $\sigma_{\bar{A}} = 0,3$ мм². Эти значения хорошо совпадают с приведенным в тексте результатом $1221,2 \pm 0,4$ мм².

4.13. а. 336 ± 15 м/с. Систематической погрешностью в 1 % в f можно пренебречь по сравнению с погрешностью 4,5 % в λ . б. 336 ± 11 м/с. В данном случае систематическая ошибка преобладает.

Глава 5

5.1. См. рис. А5.1. Пунктирная кривая на рис. А5.1, в — функция Гаусса для задачи 5.4.

5.2. а. $C = 1/(2a)$. б. Все значения между $-a$ и a равновероятны; ни один из результатов измерений не попадает вне интервала от $-a$ до a . в. $\bar{x} = 0$; $\sigma_x = a/\sqrt{3}$.

5.4. См. ответы к задаче 5.1.

5.6. Интеграл $\int z^2 e^{-z^2/2} dz$ можно переписать в стандартном виде $\int u dv$, где $u = z$ и $v = e^{-z^2/2}$. При интегрировании по частям граничные значения члена $[uv]_{-\infty}^{\infty}$ равны нулю.

5.8. а. 68 %. б. 38 %. в. 95 %. г. 48 %. д. 14 %. е. $22,3 \leq y \leq 23,7$.

5.10. Будьте внимательны при дифференцировании P , вы должны получить $\partial P / \partial \sigma = \sigma^{-(N+3)} \left[\sum (x_i - X)^2 - N\sigma^2 \right] \exp \left[- \sum (x_i - X)^2 / 2\sigma^2 \right]$; P максимальна, когда $\partial P / \partial \sigma = 0$, откуда следует искомый ответ.

5.12. а. $\sigma_t = 7,04$, б. $\bar{t}_1 = 74,25$, $\bar{t}_2 = 67,75$ и т. д. Если $\bar{t}$ обозначает среднее значение любой группы из четырех измерений, то мы ожидали бы $\sigma_{\bar{t}} = \sigma_t / \sqrt{4} = 3,52$; фактически СО десяти средних равно 3,56. в. См. рис. А5.12.

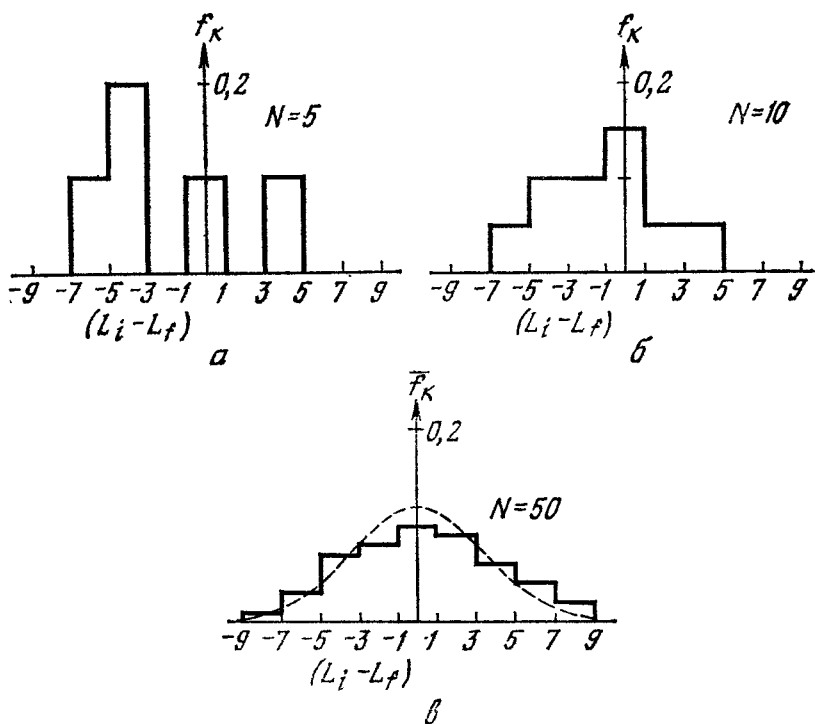


Рис. А5.1.

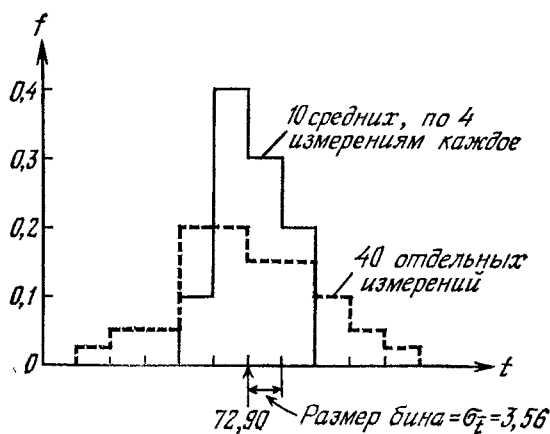


Рис. А5.12.

- 5.13. Ответ студента (9,5) отличается от предполагаемого центра распределения (9,8) на 0,3 или на три стандартных отклонения. Вероятность того, что результат отличается на три или более стандартных отклонения от центра распределения, равна P (вне 3σ) = 0,3 %. Это настолько невероятно, что мы должны заподозрить, что результаты его измерений не распределены нормально относительно 9,8 с $\sigma = 0,1$, т.е. либо он допустил явную ошибку, либо результаты его измерений подвержены влиянию некоторой систематической ошибки.
- 5.15. $E_f - E_i = 15$ МэВ, стандартное отклонение равно 9,5 МэВ. Если результаты измерений были распределены нормально с центром $E_f - E_i = 0$ и $\sigma = 9,5$ МэВ, то полученное значение отличается от истинного на 15/9,5, или на 1,6 стандартных отклонений. Так как P (вне $1,6\sigma$) = 11 %, то результат вполне разумен и нет оснований сомневаться в законе сохранения энергии.

Глава 6

- 6.2. а. $\bar{V} = 0,862$ В, $\sigma_V = 0,039$ В. б. Он отвергнет этот результат. Значение 0,95 отличается от $\bar{V}$ на 0,088 или на $2,3\sigma$. Так как P (вне $2,3\sigma$) = 2,1 %, то для десяти измерений следует ожидать только 0,21 случая, когда результат настолько или более отличался бы от $\bar{V}$. Согласно критерию Шовене, такой результат должен быть отвергнут.
- 6.3. Он не отвергнет результат 12. В данном случае $\bar{T} = 7,00$ и $\sigma_T = 2,72$; поэтому 12 отличается от $\bar{T}$ на 5 или $1,84\sigma$. Так как вероятность P (вне $1,84\sigma$) = 6,6 %, то для 14 измерений следует ожидать 0,92 случая, когда результат настолько или более отличался бы от $\bar{T}$.

Глава 7

- 7.1. а. Результаты обоих измерений согласуются друг с другом; наилучшая оценка, основанная на обоих измерениях, равна $334,4 \pm 0,9$ м/с. б. Эти результаты также согласуются (фактически в большей степени). В данном случае наилучшая оценка равна $334,08 \pm 0,98$, что, конечно, следует округлить до 334 ± 1 м/с. Очевидно, что второй результат характеризуется значительно большей погрешностью и его не следует включать.
- 7.2. а. 76 ± 4 Ом. б. Приблизительно 26 измерений.
- 7.5. В соответствии с (3.47)

$$(\sigma_{x_{\text{наил}}})^2 = \sum_i \left(\frac{\partial x_{\text{наил}}}{\partial x_i} \sigma_{x_i} \right)^2.$$

Искомая производная равна $\partial x_{\text{наил}} / \partial x_i = w_i / (\sum w_i)$. Если вы затрудняетесь в получении этого результата, то распишите сумму $\sum w_i x_i$ как $w_1 x_1 + \dots + w_N x_N$, а затем продифференцируйте по x_1, x_2 и т. д. Следовательно,

$$(\sigma_{x_{\text{наил}}})^2 = \frac{1}{(\sum w_i)^2} \sum (w_i \sigma_{x_i})^2,$$

или, так как $\sigma_{x_i} = 1/\sqrt{w_i}$, $(\sigma_{x_{\text{наил}}})^2 = 1/\sum w_i$.

Глава 8

- 8.1. $A = 9,00$; $B = 2,60$. Эти значения дают сплошную линию на рис. А8.1. (Пунктирная прямая получается для задачи 8.9.)
- 8.3. Аргументы аналогичны рассуждениям в тексте при выводе с (8.2) по (8.12); единственная разница состоит в том, что в данном случае

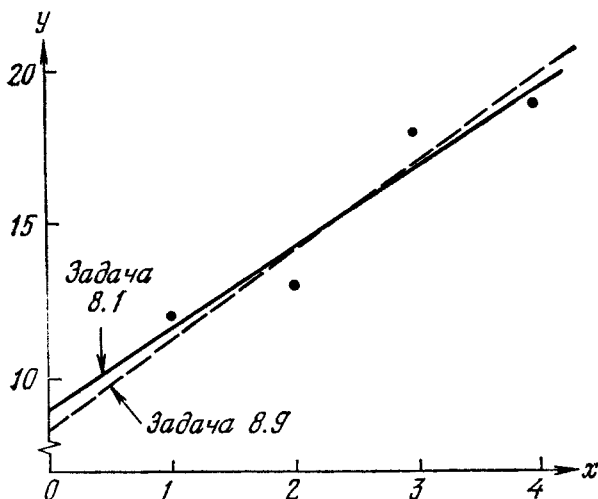


Рис. А8.1.

$A = 0$. Таким образом, $P(y_1, \dots, y_N) \sim \exp(-\chi^2/2)$, как в (8.4), а χ^2 определяется (8.5), за исключением того, что $A = 0$. Дифференцирование по B приводит к (8.7) (вновь с $A = 0$), и решение равно $B = (\sum x_i y_i) / (\sum x_i^2)$.

- 8.4. Как и в случае задачи 8.3, аргументы аналогичны рассуждениям в тексте при выводе выражений (8.2)–(8.12). Как и в (8.4), $P(y_1, \dots, y_N) \sim \exp(-\chi^2/2)$, но из-за того, что результаты измерений характеризуются разными погрешностями, в данном случае $\chi^2 = \sum w_i (y_i - A - Bx_i)^2$. (Помните, что $w_i = 1/\sigma_i^2$.) Далее те же рассуждения, что и прежде.

- 8.6. $v = 41 \pm 1$ м/с.

- 8.8. а. Вы должны определить значение σ , для которого $P(y_1, \dots, y_N)$ в (8.4) максимальна. Производная $\partial P / \partial \sigma$ есть

$$\sigma^{-(N+3)} \left[\sum (y_i - A - Bx_i)^2 - N\sigma^2 \right] \exp(-\chi^2/2);$$

приравнявая это выражение нулю, находим искомое значение σ . б. Постоянные A и B — определенные функции $x_1, \dots, x_N$ и $y_1, \dots, y_N$. Так как у x_i нет погрешностей, то формула расчета ошибок в случае косвенных измерений (3.47) дает, например,

$$\sigma_A^2 = \sum_i \left(\frac{\partial A}{\partial y_i} \sigma_{y_i} \right)^2.$$

Подставляя $\partial A / \partial y_i = [(\sum x_i^2) - x_i(\sum x_i)] / \Delta$, после небольших алгебраических преобразований получаем (8.15). Аналогичный расчет дает σ_B .

- 8.9. $A' = -2,9 \pm 1,2$; $B' = 0,35 \pm 0,08$. При использовании постоянных A и B из задачи 8.1 мы получили бы $A' = -3,5$ и $B' = 0,38$, что

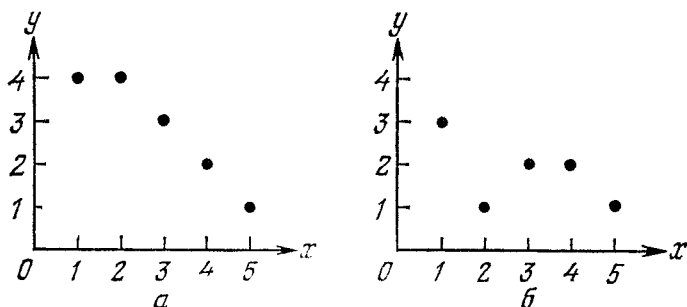


Рис. А9.6.

в пределах погрешностей совпадает с новыми значениями. Таким образом, хотя два метода определяют действительно разные линии (см. рис. А8.1), различие фактически незначимо.

8.11. Наилучшая оценка для g равна $9,4 \text{ м/с}^2$.

8.13. $A = 5,5 \text{ см}$, $B = 11,1 \text{ см}$.

8.14. $\tau = 2,02 \text{ ч}$.

Глава 9

9.1. Это доказательство будет наиболее простым, если вы заметите, что функцию $A(t)$ можно представить в виде

$$A(t) = \sigma_x^2 + 2t\sigma_{xy} + t^2\sigma_y^2.$$

$$\begin{aligned} 9.3. \text{ а. } \sum (x_i - \bar{x})(y_i - \bar{y}) &= \sum (x_i y_i - \bar{x} y_i - \bar{y} x_i + \bar{x} \bar{y}) = \\ &= (\sum x_i y_i) - \bar{x} (\sum y_i) - \bar{y} (\sum x_i) + N \bar{x} \bar{y} = \\ &= (\sum x_i y_i) - N \bar{x} \bar{y}. \end{aligned}$$

9.5. а. $P_5(|r| \geq 0,7) = 19 \%$. Таким образом, после пяти измерений значение $r = 0,7$ весьма вероятно, даже если K и f не коррелированы. В частности, это значение не подтверждает «значимым» образом линейную связь.

б. $P_{20}(|r| \geq 0,5) = 2 \%$. Так как это значение меньше 5% , то оно является «значимым» доказательством линейной корреляции.

9.6. а. $r = -0,97$. Так как $P_5(|r| \geq 0,97) \approx 1,2 \%$, то наблюдается «значимая» корреляция. б. $r = -0,57$. Так как $P_5(|r| \geq 0,57) \approx 31 \%$, то это значение незначимо.

Глава 10

10.2. а. В случае двух бросаний вероятности выпадения 0, 1 и 2 единичек равны соответственно 69,44; 27,78 и 2,78 %. б. В случае четырех бросаний вероятности выпадения 0, 1, ..., 4 единичек равны 48,23; 38,58; 11,57; 1,54 и 0,08 % соответственно.

$$10.4. (p + q)^3 = \sum_{v=0}^3 \binom{3}{v} p^v q^{3-v} = q^3 + 3pq^2 + 3p^2q + p^3.$$

10.6. Вероятность выжить для любого пациента равна $p = 0,2$; таким образом, $P(v \text{ пациентов выживают}) = b_{4; 0,2}(v)$.

а. 41 %. б. 41 %. в. 18 %.

10.7. 40,2 %; 40,2 %; 16,1 %; 3,2 %; 0,32 %; 0,01 %.

$$10.9. \text{ а. } \sigma_v^2 = (v - \bar{v})^2 = \sum_v f(v) (v - \bar{v})^2 = \sum f(v) (v^2 - 2\bar{v}v + \bar{v}^2) = \\ = \left[\sum f(v) v^2 \right] - 2\bar{v} \sum f(v) v + \bar{v}^2 \sum f(v) = \overline{v^2} - \bar{v}^2.$$

Заметьте, что если бы мы заменили суммы интегралами, то доказали бы то же утверждение и для непрерывного распределения, подобного гауссову.

10.10. Для любых p и q имеем $(p + q)^n = \sum \binom{n}{v} p^v q^{n-v}$. Дифференцируя дважды по p , находим

$$n(n-1)(p+q)^{n-2} = \sum v(v-1) \binom{n}{v} p^{n-2} q^{n-v}$$

Умножая на p^2 и подставляя $q = 1 - p$, получаем

$$n(n-1)p^2 = \sum (v^2 - v) b_{n,p}(v) = \overline{v^2} - \bar{v}.$$

Так как $\bar{v} = np$, то отсюда следует, что $\overline{v^2} = n(n-1)p^2 + np$.

Подставляя это выражение в формулу задачи 10.9 ($\sigma_v^2 = \overline{v^2} - \bar{v}^2$) (где $\bar{v} = np$), мы получаем искомый результат.

10.13. 9,68 % (аппроксимация Гаусса); 9,74 % (точное значение).

10.14. $P(v \geq 18) \approx P_{\text{Гаусс}}(v \geq 17,5) = P_{\text{Гаусс}}(v \geq \bar{v} + 2\sigma) = 2,28\%$.

10.16. $P(v \geq 12) = 0,65\%$ (если удобрение не эффективно). Таким образом, 12 успехов «значимы» и даже «высокозначимы».

10.18. Ожидается 360 успешных испытаний; $P(420 \text{ успешных исходов или более}) \approx P_{\text{Гаусс}}(v \geq 360 + 5\sigma) = 0,00003\%$, что представляет собой высокосзначимый результат.

Глава 11

11.1. а. Для $v = 0, 1, \dots, 6$ $p_{1/2}(v) = 60,7; 30,3; 7,6; 1,3; 0,2; 0,02; 0,001\%$.

11.2. а. $\sum p_\mu(v) = e^{-\mu} \sum \mu^v / v! = e^{-\mu} e^\mu = 1$

б. Дифференцирование (11.12) по μ дает

$$\sum e^{-\mu} (v\mu^{v-1} - \mu^v) / v! = 0,$$

или, используя повторно (11.12), имеем

$$\sum v e^{-\mu} \mu^{v-1} / v! = 1.$$

Умножая на μ , получаем $\sum v p_\mu(v) = \mu$, что является искомым результатом.

11.4. а. $\mu = (\text{число ядер}) \cdot p = 1,5$. б. Вероятности v распадов, где $v = 0, 1, 2, 3$, равны 22,3; 33,5; 25,1 и 12,6 % соответственно. в. $P(v \geq 4) = 6,5\%$.

11.5. Вертикальные линии на рис. А11.5 представляют полученное распределение. Ожидаемое распределение Пуассона $p_3(v)$ представлено непрерывной кривой, чтобы легко было различать распределения. Согласие хорошее.

11.7. $\bar{v} = 2,84$; $\sigma_v = 1,70$. Эти величины хорошо согласуются с ожидаемыми значениями 3 и $\sqrt{3} = 1,73$.

11.9. а. Вероятность получения отсчета v_0 равна $p_\mu(v_0) = e^{-\mu} \mu^{v_0} / v_0!$. Значение μ , для которого эта вероятность максимальна, можно найти,

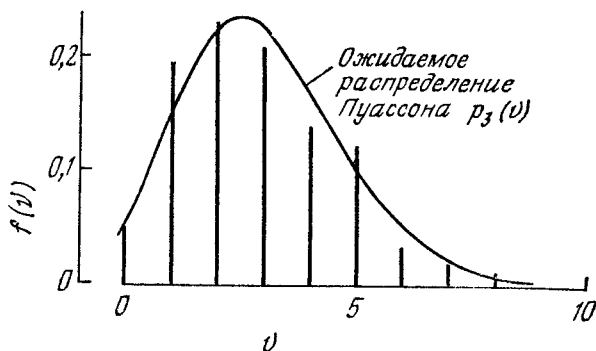


Рис. А11.5.

если продифференцировать выражение для вероятности по μ и приравнять производную нулю. Эта производная равна

$$e^{-\mu} (\nu_0 \mu^{\nu_0-1} - \mu^{\nu_0}) / \nu_0!,$$

и она обращается в нуль, когда $\mu = \nu_0$.

- 11.10. а. 3,2% (аппроксимация Гаусса); 3,4% (точное значение).
 б. 8,5% (аппроксимация Гаусса); 7,7% (точное значение).
 11.11. а. $p_9(7) = 11,7\%$ и т. д. б. $P(\nu \leq 6) + P(\nu \geq 12) = 40,3\%$. Отсчет, равный 12, неудивителен, поэтому нет оснований сомневаться в том, что $\mu = 9$.

Глава 12

- 12.1 Ожидаемые числа = 7,9; 17,1; 17,1; 7,9 и $\chi^2 = 10$; данные очень плохо аппроксимируются нормальным распределением.
 12.4. Ожидаемые числа = 231,5; 138,9; 29,6; $\chi^2 = 2,5$; в случае 3 бинов значение $\chi^2 = 2,5$ вполне разумно, и поэтому нет оснований подозревать, что кости фальшивые.
 12.5. а. В задачах 12.1 и 12.2 $c = 3$ и $d = 1$; в задаче 12.3 $c = 1$ и $d = 5$; в задаче 12.4 $c = 1$ и $d = 2$. б. Если $\rho_{\text{пр}}$ известно заранее, то $c = 2$ и $d = 2$.
 12.6. Так как $d = 1$, $\tilde{\chi}^2 = \chi^2 = 10$; $P_1(\tilde{\chi}^2 \geq 10) = 0,2\%$, поэтому мы можем отвергнуть нормальное распределение на 5- и 1%-ном уровнях.
 12.8. Вероятности сумм 2, 3, ... 12 равны $1/36, 2/36, \dots, 6/36, \dots, 1/36$. Ожидаемые числа E_k равны 10, 20, ..., 60, ..., 10; $\chi^2 = 19,8$; $d = 10$ и $\tilde{\chi}^2 = 1,98$. $P_{10}(\tilde{\chi}^2 \geq 1,98) = 3,2\%$. На 5%-ном уровне мы сказали бы, что кости утяжелены, но этого нельзя утверждать на 1%-ном уровне.
 12.10. $\tilde{\chi}^2 = 1,2$. $P_2(\tilde{\chi}^2 \geq 1,2) \approx 30\%$. Так как значение $\tilde{\chi}^2 \geq 1,2$ весьма вероятно, то нет оснований подозревать, что кости фальшивы.
 12.12. а. $E(\nu = 0) = 5,4$; $E(\nu = 1) = E(\nu = 2) = 10,8$; $E(\nu \geq 3) = 13,0$. $\chi^2 = 9,7$, $d = 3$; $\tilde{\chi}^2 = 3,2$, $P_3(\tilde{\chi}^2 \geq 3,2) \approx 2,5\%$, поэтому на 5%-ном уровне мы отвергли бы распределение Пуассона с $\mu = 2$. б. $d = 2$, $\tilde{\chi}^2 = 0,3$, и данные согласуются с распределением Пуассона с $\mu = 1,35$.
 12.13. а. $\chi^2 = 2,25$; $P_1(\chi^2 \geq 2,25) \approx 14\%$, значимой разницы нет. $P(\nu \geq 11) + P(\nu \leq 5) = 21,0\%$. б. $\chi^2 = 6,25$; $P_1(\chi^2 \geq 6,25) \approx 1,2\%$; $P(\nu \geq 224,5) + P(\nu \leq 175,5) = 1,4\%$. На 5%-ном уровне имеется значимая разница. в. «Модифицированный χ^2 » = 1,56. $P_1(\chi^2 \geq 1,56) \approx 21,1\%$, что находится в отличном согласии с точным результатом, равным 21,0%.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Архимеда проблема 14

БИН 110

- выбор 111, 224, 228, 236
- Биномиальное разложение 198
- Биномиальная теорема 56, 82
- Биномиальное распределение 197
 - — определение 197
 - — симметрия при $p = 1/2$ 200
 - — сравнение с гауссовым 200
 - — среднее 200, 211
 - — стандартное отклонение 200
- Биномиальный коэффициент

Вероятность «двуххвостовая» 210

- Взвешенная сумма 108
Взвешенное среднее 155
— -- определение 157

Гипотеза 206

- гипотеза 205
 - испытание 205
 - нулевая 206
 - статистическая 206
- Гистограмма 109
- для дискретной величины 110
 - — непрерывной величины 110
- Графики 34
- разбросов 186

Дисперсия измерений 93

Значащие цифры 26

- в произведениях 48
- и относительные погрешности 40
- Значение χ^2 227, 228, 232
 - — — таблица 255
 - — — модифицированное 246
 - — — на одну степень свободы 235
 - — — приведенное 235
- Значимости уровень 208, 237

Измерения непротиворечивые 155

- противоречивые 154, 155
- Измеренных значений разность 32, 47, 54, 65
- — сравнение 31
- — сумма 46, 54, 65, 133
- — произведение 41, 57, 65
- — частное 55, 57, 65
- Интерполяция 20
- Испытания непараметрические 207

Квадратичное сложение 61, 62

- независимых случайных погрешностей 61
- обоснование 130
- уменьшение погрешности 67
- Корреляции коэффициент 186, 188
- — вероятность 190
- — таблицы 253
- Критерий χ^2 223
- — для распределения Гаусса 224, 239
- — — Пуассона 241

Линеаризация 172

Максимального правдоподобия принцип
127. 222

Наилучшая оценка 24

Наименьших квадратов метод 156

- — — аппроксимация кривыми 169
- — — — полиномом 170
- — — — прямой 162
- — — — экспоненциальной функцией 171
- — — — оценка для A и B 163
- — — — погрешность 166
- — — с весами 176
- Некоррелированные переменные 188
- Нормально распределенные результаты 120
- Нормальные уравнения 163, 171, 175
- Нормировки условие 109

Определения проблема 14. 51

Отбрасывание данных 148

Отклонение 90

- среднее 90, 104
- стандартное 91, 94, 96, 104
- выборочное 94
- генеральной совокупности 94
- как 68 % -ный доверительный предел 123
- — погрешность единичного измерения 94
- определеение 93
- — «улучшенное» 93, 129
- распределения 129
- биномиального 201
- — гауссова 122
- предельного 116
- — *Пуассона* 216
- Ошибка 12, 25
- вероятная 125
- истинная 29
- компенсирующая 79
- систематическая 22, 87, 101
- случайная 87, 101
- стандартная среднего 96
- Ошибка нормальная функция 118, 124
- нормальный интеграл 124
- — таблицы 248
- черточки 37

Параллакс 89

Погрешность 24

- абсолютная 39
- взвешенное среднее 156
- в косвенных измерениях 49, 179
- — — верхний предел 184
- — — в произведениях и частных
53, 55, 57, 65
- — — суммах и разностях 54, 65
- — — для функций нескольких пе-
ременных 78
- — — — одной переменной 67
- — — метод «шаг за шагом» 71
- — — общая формула 78
- — — определение из графиков 68,
84
- — — при возведении в степень 59
- — — смешанный второй момент
181
- — — процентах 39

- — прямых измерениях 50
- — случае многократных измерений 21
- — счетных экспериментах 53
- коэффициент доверия 139
- независимая 61, 64
- относительная 40, 217
- при считывании со шкалы 18, 20
- — умножении измеренной величины на точное число 58
- размерность 27, 40
- смысл понятия 25, 139
- среднего 138
- Принятое значение 29
- ПШПВ 143, 144

Различие 28, 140

— значимое 141

Разность измеренных величин 32, 47, 54
65

Распределение 109

— биномиальное 194

— — определение 197

— — симметрия при $p = 1/2$ 200

— — сравнение с гауссовым 200

— — среднее значение 199, 211

— — стандартное отклонение 201

— вероятности 194

— гауссово 200, 229 (см. Нормальное распределение)

— дискретное 200

— лоренцево 229

— нормальное 116

— — критерий χ^2 224, 239

— — нормировка 119

— — определение 120

— — параметр ширины 118

— — сравнение с биномиальным 200

— — — распределением Пуассона 217

— — среднее значение 121

— — стандартное отклонение 123

— предельное 112, 126, 194

— — как вероятность 114

— — нормировка 114

— Пуассона 213

— — критерий χ^2 241

— — нормировка 221

— — определение 214

— — сравнение с гауссовым 217

— — среднее значение 215

— — стандартное отклонение 216

— — родительское 113

Регрессионная линия 164

Регрессия линейная 162

— множественная 174, 175

— полиномиальная 171

Результат высокозначимый (1 %) 207, 237

— значимый (5 %) 207, 236

Связь 232

Смешанный второй момент 182

— — — и коэффициент корреляции 186

— — — при расчете ошибок 181

Сравнение двух измеренных значений 31, 139

— измеренного и принятого значений 29, 139

Среднее как наилучшая оценка 90, 126

— распределения биномиального 200, 211

— — нормального 121

— — предельного 116

— — Пуассона 215

Степени свободы 165, 232

Факториал 197

Цифровое табло 52

Частные производные 80, 85

«Шаг за шагом» метод 71

Шварца неравенство 184, 192

Шовене критерий 149

Экстраполяция 168

ОГЛАВЛЕНИЕ

Предисловие переводчика	5
Предисловие	7

ЧАСТЬ I

Глава 1. Предварительное знакомство с теорией ошибок	12
1.1. Ошибки как погрешности	12
1.2. Неизбежность погрешностей	12
1.3. Как важно знать погрешности	14
1.4. Другие примеры	16
1.5. Оценка погрешностей при считывании со шкалы	18
1.6. Оценка погрешностей в случае многократных измерений	21
Глава 2. Как приводить и использовать погрешности	24
2.1. Наилучшая оценка $\pm$ погрешность	24
2.2. Значащие цифры	26
2.3. Различие	28
2.4. Сравнение измеренного и принятого значений	29
2.5. Сравнение двух измеренных значений	31
2.6. Проверка пропорциональности с помощью графика	34
2.7. Относительные погрешности	39
2.8. Значащие цифры и относительные погрешности	40
2.9. Умножение двух измеренных значений	41
Глава 3. Погрешности в косвенных измерениях	49
3.1. Погрешности в прямых измерениях	50
3.2. Суммы и разности; произведения и частные	53
3.3. Независимые погрешности в сумме	61
3.4. Еще о независимых погрешностях	64
3.5. Произвольная функция одной переменной	67
3.6. Метод «шаг за шагом»	71
3.7. Примеры	72
3.8. Более сложный пример	75
3.9. Общая формула для вычисления ошибок в косвенных измерениях	78
Глава 4. Статистический анализ случайных погрешностей	87
4.1. Случайные и систематические ошибки	88
4.2. Среднее и стандартное отклонение	90
4.3. Стандартное отклонение как погрешность единичного измерения	94
4.4. Стандартное отклонение среднего	96
4.5. Примеры	98
4.6. Систематические ошибки	101
Глава 5. Нормальное распределение	106
5.1. Гистограммы и распределения	107
5.2. Предельные распределения	112
5.3. Нормальное распределение	116
5.4. Стандартное отклонение как 68%-ный доверительный предел	123
5.5. Обоснование среднего как наилучшей оценки	126
5.6. Обоснование квадратичного сложения	130
5.7. Стандартное отклонение среднего	136
5.8. Коэффициент доверия	139

ЧАСТЬ II

Глава 6. Отбрасывание данных	148
6.1. Проблема отбрасывания данных	148
6.2. Критерий Шовене	149
6.3. Пример	152
Глава 7. Взвешенные средние	154
7.1. Проблема объединения результатов разных измерений	154
7.2. Взвешенное среднее	155
7.3. Пример	158
Глава 8. Аппроксимация методом наименьших квадратов	160
8.1. Данные, которые должны ложиться на прямую линию	160
8.2. Расчет постоянных A и B	162
8.3. Погрешность в измерениях y	164
8.4. Погрешность в постоянных A и B	166
8.5. Пример	163
8.6. Аппроксимация другими кривыми методом наименьших квадратов	169
Глава 9. Смешанный второй момент и корреляция	179
9.1. Обзор расчета ошибок в косвенных измерениях	179
9.2. Смешанный второй момент при расчете ошибок в косвенных измерениях	181
9.3. Коэффициент линейной корреляции	184
9.4. Количественный критерий значимости r	189
9.5. Примеры	191
Глава 10. Биномиальное распределение	194
10.1. Распределения	194
10.2. Вероятности при бросании игральных костей	195
10.3. Определение биномиального распределения	197
10.4. Свойства биномиального распределения	199
10.5. Распределение Гаусса случайных ошибок	203
10.6. Применения. Испытание гипотез	205
Глава 11. Распределение Пуассона	2 3
11.1. Определение распределения Пуассона	213
11.2. Свойства распределения Пуассона	215
11.3. Примеры	218
Глава 12. Критерий χ^2 для распределений	223
12.1. Введение в критерий χ^2	223
12.2. Общее определение χ^2	228
12.3. Степени свободы и приведенное значение χ^2	232
12.4. Вероятности для χ^2	235
12.5. Примеры	239
Приложения	248
Приложение А. Интеграл ошибок. I	248
Приложение Б. Интеграл ошибок. II	251
Приложение В. Вероятности коэффициентов корреляции	253
Приложение Г. Вероятности для χ^2	255
Библиография	258
Литература, добавленная при переводе	258
Ответы к избранным задачам	259
Предметный указатель	269

