

Н. АРЛЕЙ ← К. БУХ

ВВЕДЕНИЕ
В ТЕОРИЮ ВЕРОЯТНОСТЕЙ
И
МАТЕМАТИЧЕСКУЮ
СТАТИСТИКУ



К. БУХ

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

И*Л

Издательство

иностранной

литературы

INTRODUCTION
TO THE THEORY OF
PROBABILITY
AND
STATISTICS

NIELS ARLEY,
K. RANDER BUCH

NEW YORK

1950

НИЛЬС АРЛЕЙ и К. РАНДЕР БУХ

ВВЕДЕНИЕ
В ТЕОРИЮ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКУЮ
СТАТИСТИКУ

Перевод с английского

А. С. МОНИНА и А. А. ПЕТРОВА

Под редакцией

Б. А. СЕВАСТЬЯНОВА

Предисловие

А. Н. КОЛМОГОРОВА

1951

ИЗДАТЕЛЬСТВО
ИНОСТРАННОЙ ЛИТЕРАТУРЫ
Москва

АННОТАЦИЯ

Книга может рассматриваться как практическое руководство для физиков и инженеров, применяющих в своей работе статистические методы. В ней уделяется основное внимание задачам из современной физики и техники, но в то же время освещаются и теоретико-вероятностные основы статистических методов.

ПРЕДИСЛОВИЕ К РУССКОМУ ПЕРЕВОДУ

Предлагаемая вниманию советского читателя книга Арлея и Буха уже выдержала три издания в Дании, а в 1950 г. с небольшими авторскими изменениями переиздана в США. Она задумана как практическое руководство для физиков и достаточно подготовленных в теоретическом отношении инженеров, применяющих в своей работе статистические методы. От большинства англо-американских практических руководств по математической статистике, которые почти без исключения следуют однообразному стандарту, установленному школой Р. А. Фигшера, книга Арлея и Буха выгодно отличается большим вниманием к освещению теоретико-вероятностных основ статистических методов и более реалистическим наглядным подходом к разбору практических задач.

В теоретическом отношении книга Арлея и Буха тесно примыкает к имеющейся в русском переводе книге Крамера „Математические методы статистики“, являющейся наиболее полным выражением тенденций современной скандинавской школы в теории вероятностей и математической статистике. В книге Крамера можно найти математические выводы ряда положений, приведенных у Арлея и Буха без доказательств.

Выбор примеров для иллюстрации теоретических положений в книге Арлея и Буха не всегда удовлетворит советского читателя (см. в § 9.5 о статистическом изучении доходчивости различных видов рекламы и т. п.). Однако внимание, уделяемое авторами преимущественно задачам из современной физики, техники измерений и т. п. вопросам, в которых они, по-видимому, хорошо ориентируются не только с математической, но и с практической стороны, делает их книгу в части примеров и задач в целом более соответствующей интересам советского читателя, чем большинство других зарубежных книг по математической статистике.

В трактовке принципиальных философских вопросов авторы книги непоследовательны. Их определение вероятности $P(A)$ события A при условиях \mathcal{E} как „физической постоянной“, вокруг которой группируются экспериментально наблюдаемые частоты, можно было бы считать материалистическим, если бы далее разъяснялось, что число $P(A)$ характеризует *реальный* характер связи между условиями \mathcal{E} и событием A . Но авторы называют условия \mathcal{E} „категорией наблюдений“,

а само событие A имеют тенденцию трактовать исключительно как „наблюдаемый“ факт, а не как событие, происходящее в реальном мире независимо от нашего сознания. После правильного высказывания — „если эксперимент показывает, что рассматриваемое случайное явление есть явление стохастически случайное, то имеется *нечто*, т. е. некоторое физическое свойство рассматриваемого явления, которое имеет тенденцию оставаться постоянным“ (стр. 17) — авторы торопятся сделать оговорку, что „законность такой интерпретации“ основывается лишь на ее „удобстве“. Такое соскальзывание с первоначальных правильных высказываний к заключительным нравоучениям чисто идеалистического характера, к сожалению, типично даже для лучших в методологическом отношении мест книги.

Некоторые общие махистские декларации авторов, оторванные от конкретного содержания книги, опущены при переводе. Однако классификацию случайных явлений, изложенную авторами в § 1.2 с ошибочных субъективистских позиций, мы сочли целесообразным сохранить в переводе, так как рассмотренные в ней случаи помогают разобраться в способах вхождения случайности в течение того или иного реального процесса:

1) Течение процесса может подчиняться строгим закономерностям, которые его однозначно определяют по начальным условиям, но сами эти начальные условия могут при повторном осуществлении процесса быть случайными.

2) Все временное течение процесса может быть существенно случайным. §

3) Во временном течении процесса может проявляться простая основная закономерность, но она может быть осложнена случайными возмущениями, действующими в течение всего процесса.

Пример рулетки, приведенный авторами книги, ясно иллюстрирует первый тип вхождения случайности. В этом примере дальнейшее детерминированное течение процесса имеет затухающий характер: колесо в конце концов останавливается. Влияние случайного характера начальных условий на временное течение незатухающих процессов изучается при помощи так называемых „эргодических теорем“. Такой подход к обоснованию статистической механики авторы книги упоминают в виде второго примера. Но более отчетливо смысл этого упоминания раскрывается лишь в § 4.16.

Хорошим примером второго типа случайности является броуновское движение взвешенной в жидкости или газе частицы.

Движение несколько более тяжелых частиц в воздухе, которые в основном двигаются по определенному закону падения тел в сопротивляющейся среде, являются уже примером третьего типа случайности: молекулярные толчки, или турбулентные пульсации, в этом случае приводят к случайным возмущениям основной закономерности.

Несколько уклоняясь от мнения авторов книги, мы отнесли бы полет артиллерийского снаряда к смешанному типу, когда случайность

входит *первым* и *третьим* способом. При стрельбе однотипными снарядами при неизменной установке орудия случайность входит в начальные условия через отклонения от нормы массы и баллистического коэффициента отдельных снарядов и в течение всего полета, например, из-за неоднородной турбулентной структуры скорости ветра.

В отношении фактического материала распределение внимания авторов несколько субъективно. Приветствуя желание авторов отойти от стандартов фишеровской школы, следует все же указать, что они, например, напрасно совсем опустили изложение дисперсионного анализа.

Интересной новинкой, заимствованной из оригинальной работы Арлея, является применение γ -распределения к проверке гипотезы нормальности распределений в большом числе малых выборок с меняющимися от выборки к выборке средними и дисперсиями (см. пример 1 § 11.21). Следует, однако, отметить, что авторы, проводя здесь и в § 10.6—10.7 сравнение полигона сумм с теоретическим распределением, напрасно не указывают разработанного советскими авторами метода оценки допустимых при этом отклонений, который изложен в примечании редактора (стр. 239).

Книгу Арлея и Буха нельзя рекомендовать в качестве учебника, но достаточно подготовленные читатели, интересующиеся физическими и техническими приложениями математической статистики, найдут в ней много свежего и интересного материала.

Перевод книги сделан лишь с незначительными сокращениями. Некоторые ошибки математического характера исправлены в тексте. Ряд указаний на необходимые уточнения формулировок авторов сделан в примечаниях редактора. Ссылки на учебную литературу заменены ссылками на соответствующие русские книги.

А. Колмогоров.



ИЗ ПРЕДИСЛОВИЯ АВТОРОВ К ПЕРВОМУ ДАТСКОМУ ИЗДАНИЮ

Цель настоящей книги — дать элементарное введение в теорию вероятностей и математическую статистику, уделяя специальное внимание их практическим приложениям.

В последние годы теория вероятностей и математическая статистика получили чрезвычайно плодотворное развитие как в отношении расширения области их применения, так и в отношении их математического обоснования и математической теории. В этой книге мы пытались осветить современное развитие, все три направления которого имеют, на наш взгляд, существенное значение для всех изучающих физику, биологию, технику и другие ветви современной науки, в которых все больше и больше используется теория вероятностей и математическая статистика в качестве рабочего инструмента при классификации и анализе нашего опыта.

Первые три раздела имеют элементарный характер. В следующих разделах мы предполагаем знание дифференциального и интегрального исчисления, а в последнем разделе — знание теории линейных уравнений. При изложении теории математической статистики мы следовали современной английской школе, основанной, по существу, Фишером, и в нашем изложении теории ошибок используются статистические методы, развитые главным образом этой школой. Эти методы исходят из того, что на практике часто приходится делать выводы на основании небольшого числа наблюдений. При изложении теории выравнивания мы использовали теорию матриц главным образом из-за ее удобства. Хотя теперь все большее число людей приобретает известные знания в теории матриц, мы все же включили приложение, содержащее наиболее важные определения и теоремы. Так как задачи, связанные с практическими вычислениями, достаточно полно представлены в других учебниках, мы в нашем изложении останавливаемся преимущественно только на вопросах принципиального характера.

Мы выбрали примеры из таких различных областей, как теория стрельбы, телефония, теоретическая физика и страховое дело. При этом мы надеялись продемонстрировать фундаментальную роль теории вероятностей в современной науке. В конце книги даны некоторые статистические таблицы и указаны некоторые работы, имеющие значение для практических приложений.

При выборе символов и обозначений мы старались: 1) придерживаться, насколько это возможно, общеупотребительных обозначений, 2) последовательно проводить четкое разграничение между теоретическими и эмпирическими понятиями, обозначая с этой целью первые греческими буквами, а вторые — соответствующими латинскими буквами. К несчастью, в теории вероятностей и математической статистике часто бывают общеупотребительными различные символы и обозначения для одних и тех же понятий. Надо надеяться, что в недалеком будущем будет проведена международная стандартизация, по крайней мере, основных символов и обозначений.

Мы благодарим профессора Стокгольмского университета Г. Крамера за чтение рукописи и за любезное разрешение использовать в различных частях книги его изложение теории вероятностей. Наконец, очень ценной для нас явится критика и замечания читателей, которые мы надеемся использовать впоследствии.

Н. Арлей,

К. Р. Бух.

Копенгаген, 1946.

1. ПОНЯТИЕ ВЕРОЯТНОСТИ ¹⁾

§ 1.1. При изучении проблем самого различного рода, как научных, так и практических, часто встречаются события, которые появляются в *большом числе*. Такие события могут быть классифицированы следующим образом:

1. *Одно и то же событие повторяется* несколько раз, причем условия, при которых совершается это событие, возвращаются к одному и тому же начальному состоянию перед каждым повторением.

Примеры: повторные измерения физической величины; серия бросаний игральной кости.

2. *Одно и то же событие*, меняющееся со временем, наблюдается некоторое количество раз в *последовательные* моменты времени.

Примеры: положение частицы в броуновском движении, наблюдаемое в микроскоп в различные моменты времени; количество требуемый товар в магазине в различные дни.

3. Несколько *различных событий*, которые могут быть рассматриваемы в данной проблеме, принадлежат к *одному и тому же виду* и наблюдаются одновременно.

Примеры: положения нескольких броуновских частиц, наблюдаемых в один и тот же момент времени; количество запросов на некоторый товар в различных магазинах в один и тот же день; рост некоторого числа солдат; размеры урожая пшеницы с некоторого числа различных участков; продолжительность работы некоторого количества электрических лампочек.

Характеристическая особенность статистического описания, в противоположность описанию индивидуальному, заключается в том, что оно применяется не к единичному событию, а лишь ко всей совокупности большого числа событий; таким образом, при статистическом подходе описываются лишь свойства, которыми обладает некоторое количество событий, рассматриваемых не по отдельности, а в совокупности. Так, например, при статистическом изучении рождения

¹⁾ Замечание: в этой книге формулы нумеруются последовательно в каждом параграфе, и ссылка на них производится следующим образом: формула (1) из § 2.5 в этом же самом параграфе упоминается как (1), но формула (2) из § 2.4 упоминается в другом параграфе уже как (2.4.2) и т. д.

Абзацы, отмеченные звездочкой (*), более трудны и могут быть в первом чтении пропущены.

детей исследователь интересуется лишь тем, сколько детей родилось, но не тем, у кого родился ребенок. При таком статистическом изучении можно исследовать, как распределяется количество детей между двумя полами, среди различных групп населения и т. д. Далее, особую роль играют средние значения чисел и отклонения от них.

Пример. Предположим, что некто гуляет на дороге, вдоль которой стоят последовательно пронумерованные телеграфные столбы. и наблюдает последние цифры в номере каждого столба. Наблюдатель может описать свои наблюдения, указав, что числа идут в некоторой правильной последовательности, например 8, 9, 0, 1, 2 и т. д. (индивидуальное описание), или же указав, сколько раз встречается каждое из этих чисел в большом ряду наблюдений (статистическое описание).

Вопрос о том, какое описание, индивидуальное или статистическое, следует применить в каждом данном случае, всецело зависит от того, какой подход более соответствует стоящим перед нами задачам. Как правило, в обоих случаях описание заключается в упрощении и идеализации рассматриваемого явления и построении некоторой его модели — в нашем случае модели математической природы, — наилучшим образом соответствующей тем чертам реального явления, которые мы считаем наиболее существенными.

Примеры: изображение наблюдаемого явления в пространстве с помощью евклидовой геометрии; описание движения тел с помощью аналитической механики; приложение простых математических законов к описанию статистических явлений (см. раздел 10).

Никакая такая идеализированная модель некоторой совокупности экспериментов не может быть отождествлена с самой реальностью. Следовательно, каждая такая модель сама по себе не является „верной“ или „неверной“; это утверждение подкрепляется тем обстоятельством, что мы, как правило, можем построить несколько различных моделей, представляющих одно и то же явление. *Критерием законности использования некоторой модели является то, что эта модель должна давать достаточно удовлетворительное описание реального мира* (см. § 9.7).

§ 1.2. Если некоторая группа событий описывается статистическим образом, то ее трактуют как *случайное* явление. Это означает, что из принципиальных или из практических соображений представляется невозможным предсказать „конечное состояние“ явления, исходя из „начального состояния“ и из известных законов природы.

Случайные явления характеризуются тем, что они обусловлены настолько большим количеством взаимодействующих причин, что практически невозможно проанализировать их с достаточной для составления точных предсказаний аккуратностью. *Первым* основанием к тому, чтобы считать явление случайным, может быть то, что мы не можем фиксировать начальное состояние настолько аккуратно, чтобы можно

было однозначно определить конечное состояние. Таковы следующие случаи:

а) Случай, когда очень небольшое изменение начального состояния может привести к большому изменению конечного состояния, хотя явление имеет простую природу. Например, чрезвычайно маленькое изменение начальной скорости вращения колеса рулетки может сказаться на том, выпадет ли в конечном состоянии „красное” или „черное”. Аналогично в основе всех *азартных игр* лежит такая их конструкция, что даже очень маленькое изменение начального состояния сильно влияет на конечное состояние.

б) Случай, когда начальное состояние настолько сложно, что практически невозможно фиксировать его с достаточной для однозначного определения конечного состояния аккуратностью, хотя явление и здесь может быть очень простым. Например, известно, что в одном грамме водорода содержится $3 \cdot 10^{23}$ молекул, но измерить координаты и скорости всех этих молекул в данный момент (т. е. фиксировать так называемое *микроскопическое* состояние), конечно, невозможно. Можно выбрать для описания начального состояния другие величины, такие, как термодинамическое давление p , объем v и абсолютная температура T ; поскольку такое *макроскопическое* описание не определяет однозначно микроскопического состояния, по одному лишь макроскопическому состоянию нельзя дать однозначного предсказания конечного состояния. Следовательно, хотя движение каждой молекулы однозначно определяется с помощью уравнений классической механики, состояние рассматриваемой массы водорода в зависимости от времени следует трактовать как случайное явление (см. § 4.16).

Такое положение возникает всякий раз, когда имеется противоречие между двумя требованиями, разрешение которых необходимо для определения „состояния” явления. Первое из этих требований заключается в том, что свойства, характеризующие состояние, должны обеспечивать возможность одновременного определения всех других интересующих нас свойств с помощью законов природы; второе требование заключается в том, что свойства, характеризующие состояние, должны быть непосредственно наблюдаемы и могут быть фиксированы.

в) Случай, когда начальное состояние принципиально не может быть измерено, так как сам процесс измерения может вносить неконтролируемые изменения в исследуемое явление. Таков случай атомных явлений, и вследствие этого квантовая механика решающим образом отличается от классической физики, в которой предполагается, что воздействие измерительного прибора на исследуемое явление всегда может быть сделано сколь угодно малым (см. стр. 14).

Второе основание для случайного характера явления может быть отнесено к сложности законов природы. Таковы следующие случаи:

а) Случай, когда относящиеся к делу законы природы настолько сложны, что практически невозможно вычислить характеристики ко-

нечного состояния, хотя принципиальная возможность этого в теории существует. Примером этого может служить серия бросаний игральной кости. Теоретически можно предсказать результат данного броска, если точно известны начальное положение и скорость кости, ее геометрическая форма, масса, момент инерции, упругие свойства кости и стола и т. д. Таким образом, необходим предельно точный анализ всех этих взаимосвязанных причин; но такой анализ был бы столь сложен, что практически он неосуществим.

б) Случай, когда относящиеся к делу законы природы недостаточно хорошо известны, как это имеет место, например, во многих биологических явлениях. Так, если измеряется рост у некоторого числа солдат, то результаты каждого отдельного изменения связаны со многими биологическими процессами, такими, как наследственность и питание, законы которых известны лишь частично.

Третье основание для случайного характера явлений заключается в том, что все законы природы, строго говоря, приложимы лишь к *идеализированным* явлениям. Как уже было указано, чтобы вывести общие законы, мы должны *упростить* и *идеализировать* явление, умышленно пренебрегая многими факторами и рассматривая лишь одну из действующих причин. Примером может служить закон инерции, в котором обычно пренебрегают трением. В действительности явления всегда имеют сложный характер и, более того, подвержены действию возмущающих факторов, как, например, изменений температуры и давления, толчков, люфта в винтах машин, механических, электрических и магнитных воздействий окружающей среды, и часто воздействию самих измерительных аппаратов. Так, например, если мы измеряем силу электрического тока с помощью амперметра, ток изменяется, так как амперметр сам имеет некоторое сопротивление. В этом примере эффект возмущающего фактора может быть вычислен, что позволяет исправить измеренное значение. Однако это исключительный пример. В большинстве случаев величина факторов, возмущающих результат измерения, не может быть вычислена и, как было указано выше об атомных явлениях, эти факторы могут оказаться принципиально неконтролируемыми. Тем не менее эти возмущающие факторы могут иметь такое же решающее значение при определении явления, как и другие причины. Поэтому каждое физическое измерение в меньшей или большей степени является случайным событием, дающим различные результаты при повторных измерениях.

Как правило, при описании явлений мы встречаем одновременно факторы всех трех описанных типов. Рассмотрим в качестве примера бросание игральной кости. Во-первых, очень малые изменения начального состояния могут оказаться решающими для получения результата с цифрой 2 или с цифрой 6. Во-вторых, законы, управляющие этим явлением, настолько сложны, что невозможно довести до конца вычисления конечного состояния, даже если бы начальное состояние было точно известно; наконец, здесь имеют место возмущающие факторы,

как, например, сопротивление воздуха. Другим хорошо известным примером случайного явления, в котором существенны факторы всех трех групп, является артиллерийская стрельба, при которой точки попадания серии выстрелов всегда различны.

§ 1.3. *Наблюдение* определяется как установление определенного результата в случайном явлении. Как правило, результат наблюдения состоит из одного или нескольких чисел, но он может быть охарактеризован и другими способами. В примере с бросанием игральной кости наблюдения заключаются в бросании кости и чтении выпадающих цифр. Результатом наблюдения является одно из чисел 1, 2, 3, 4, 5 или 6. При рассмотрении движения молекулы наблюдение заключается в измерении положения и скорости молекулы в некоторый момент времени, и результат состоит из шести чисел — трех координат и трех компонент скорости; каждое из этих шести чисел может иметь любое значение между $-\infty$ и ∞ . С другой стороны, при исследовании какой-либо статистической популяции результат наблюдения может и не быть числом. Так, например, если наблюдение заключается в определении пола индивидуума из данного списка, то результатом будет или „мужской пол“ или „женский пол“.

§ 1.4. Рассмотрим теперь определенное случайное явление, при исследовании которого была проделана серия в n наблюдений одним из трех способов, описанных в § 1.1. Подсчитаем число n_A результатов, относящихся к некоторому определенному классу. Будем говорить, что в этих случаях произошло *событие* A . Число n_A называется *абсолютной частотой* события A , а дробь

$$f(A) = \frac{n_A}{n}$$

— *относительной частотой* события A в данной серии наблюдений. Опыт показывает, что случайные колебания или, как их иногда называют, статистические флуктуации, как правило, сглаживаются, и процесс сглаживания следует некоторому закону, который мы будем называть *законом случайности*¹⁾, справедливость которого является основным условием для всех приложений теории вероятностей к описанию реальных явлений.

Опыт показывает, что при вычислении относительных частот некоторого события A в различных сериях наблюдений определенной, строго установленной категории \mathcal{S} получаемые числа будут

¹⁾ Этот эмпирический закон часто называется *законом больших чисел*, но это наименование употребляется также для *математического* закона (см. (8.1.14)). Такое двойное употребление термина неудачно, и это привело ко многим недоразумениям при историческом развитии теории вероятностей.

очень мало отличаться друг от друга, если каждая серия состоит из очень большого числа наблюдений.

Другими словами, если мы нанесем эти относительные частоты в виде точек на числовую ось, то точки будут группироваться относительно некоторой средней точки, причем величина рассеяния точек относительно средней будет уменьшаться при возрастании числа наблюдений в каждой серии. Для обозначения этого типа случайности употребляется слово *стохастическая случайность*.

Таблица 1

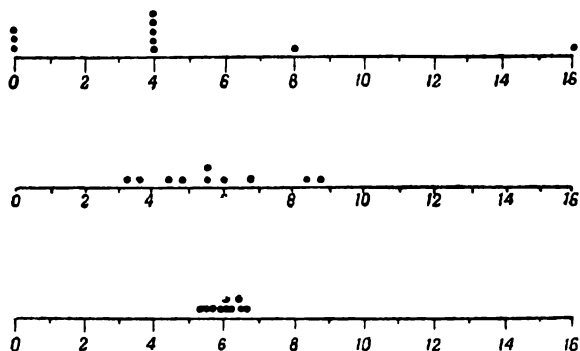
	Число изделий в группе		
	25	250	2500
Число забракованных изделий в группе и соответствующий процент			
1 (4)	12 (4,8)	157 (6,28)	
4 (16)	14 (5,6)	152 (6,08)	
0 (0)	17 (6,8)	157 (6,28)	
0 (0)	11 (4,4)	136 (5,44)	
1 (4)	22 (8,8)	152 (6,08)	
1 (4)	9 (3,6)	135 (5,40)	
2 (8)	15 (6,0)	143 (5,72)	
0 (0)	14 (5,6)	160 (6,40)	
1 (4)	21 (8,4)	149 (5,96)	
1 (4)	8 (3,2)	153 (6,12)	

Пример 1. В качестве примера стохастической случайности рассмотрим процент брака в производстве промышленного изделия. В табл. 1 мы приводим данные о количестве забракованных изделий в каждой из 10 выборочных групп, содержащих сначала по 25, затем по 250 и, наконец, по 2500 изделий соответственно. Цифры в скобках указывают соответствующий процент брака. Колебания процентов брака изображены на фиг. 1, где на числовой оси откладывается число процентов. Каждое из полученных значений изображается точкой над соответствующей точкой числовой оси. Следует заметить, что полученные значения процентов брака (и, следовательно, значения относительных частот) группируются около среднего значения, равного примерно 6% , и что рассеяние нанесенных точек тем меньше, чем больше изделий содержит группа.

Пример 2. На практике встречаются также виды случайности, отличные от стохастической случайности. Рассмотрим еще раз пример из § 1.1 и назовем событием A наблюдение цифры 2 на последнем месте в номере телеграфного столба. Вообще говоря, $n_A/n \sim 1/10$. Однако если отмечать только каждый десятый телеграфный столб, то мы заметим, что $n_A/n = 1$ для всех n , если начинать со столба, номер которого оканчивается цифрой 2, и $n_A/n = 0$ в остальных случаях.

В этом особом случае относительные частоты не группируются относительно какого-нибудь среднего значения. Следовательно, здесь нелепо было бы применять статистическое описание — факт очевидный

так как может быть дано точное индивидуальное описание. При „действительной“ случайности можно ожидать, что относительные частоты n_A/n будут группироваться около некоторого среднего значения, даже если произвести отбор наблюдений каким-нибудь „случайным“ образом, например, если получать числа не тем методом, который описан в начале примера, а выбирая последние цифры номеров



Фиг. 1.

телефонов с какой-нибудь страницы телефонной книги. Мы предоставляем читателю проверить это утверждение на собственном опыте.

Если эксперимент показывает, что рассматриваемое случайное явление есть явление стохастически случайное, то имеется „нечто“, т. е. некоторое физическое свойство рассматриваемого явления, которое имеет тенденцию оставаться *постоянным*.

Преследуя цели математического описания, мы должны, очевидно, идеализировать этот результат опыта, отвлекаясь от различий между относительными частотами в различных сериях наблюдений, и ввести некоторое определенное число, которое будет представлять „истинное“ значение относительной частоты события, так же как мы вводим некоторое определенное число, представляющее „истинное“ значение какой-либо другой физической величины (см. § 11.2). Таким образом, мы интерпретируем относительные частоты как экспериментальные значения одной и той же физической постоянной, которая определяется природой события A и категорией наблюдений Ξ (см. § 1.6). Но, как уже подчеркивалось, законность такой интерпретации основывается лишь на том факте, что эта интерпретация дает возможность удобного описания статистических явлений, встречающихся на практике.

Физическая постоянная, экспериментальными значениями которой являются относительные частоты, называется вероятностью события A в категории наблюдений Ξ и обозначается $P(A)$. Вследствие способа, с помощью которого введена вероятность $P(A)$, она называется также истинной относительной частотой события A .

Следует подчеркнуть, что при этом определении вероятность не определена для событий, которые могут быть осуществлены только один раз, т. е. для событий, которые не могут быть воспроизведены. Например, мы не можем говорить о вероятности того события, что некоторый теннисист выиграет определенную игру, хотя слово „вероятность“ и употребляется в повседневной речи по отношению к таким событиям¹⁾.

Итак, вероятность есть число, которое определяется событием A и категорией наблюдений \mathcal{S} в таком же смысле, как вес бруска железа определяется самим этим бруском, несмотря на то, что повторные измерения этого веса дают более или менее различные значения. Эксперимент показывает, что, как правило, точность измерения можно увеличить, употребляя более чувствительные измерительные приборы. Таким же образом эксперимент показывает, что, как правило, точность эмпирического определения вероятности может быть увеличена, если использовать большее количество наблюдений (см. пример 1). Таким же образом, как нельзя увеличить неограниченно точность какого-либо физического измерения, так и невозможно беспредельное увеличение точности экспериментального определения вероятности. Каждое эмпирическое число почти всегда может быть определено лишь с некоторой ограниченной точностью, и поэтому бессмысленно говорить об осуществлении процесса предельного перехода с помощью „бесконечно большого“ числа наблюдений (см. стр. 21).

Подход к понятию вероятности очевидным образом аналогичен подходу к каждому другому абстрактному понятию, с помощью которого мы описываем и классифицируем наблюдаемые явления (например, прямая линия в евклидовой геометрии, положение как непрерывная функция времени в механике). Подобно тому, как в геометрии мы всегда бессознательно осуществляем, например, переход от „наблюдаемой прямой линии“ к „прямой линии в евклидовой геометрии“, отделяя ее от соответствующей реальности, *в теории вероятностей мы должны привыкнуть, с одной стороны, к интуитивному восприятию вероятности как идеализированной относительной частоты события и, с другой стороны, к разделению в наших рассуждениях двух понятий — частоты и вероятности.* В последующем мы будем всюду употреблять, насколько это окажется возможным, различные буквы для обозначения двух типов величин: латинские буквы для обозначения величин, полученных по опытным данным, и соответствую-

¹⁾ Некоторые авторы старались также определить математическую вероятность невозпроизводимых событий как степень их правдоподобия; но тогда возникают трудности получения численных значений вероятности в конкретных случаях и сравнения данных численных значений с экспериментом. Известными представителями этой школы являются J. M. Keynes, *A Treatise on Probability*, London, 1921, и H. Jeffreys, *Theory of Probability*, London, 2-е изд., 1948 г.

ющие греческие буквы для обозначения соответствующих величин в нашей математической модели.

§ 1.5. В некоторых задачах вероятности рассматриваемых событий можно вычислить теоретически с помощью простых гипотез о природе явления. Такие вероятности называются *априорными*, т. е. определенными до опыта. Наоборот, вероятности, определенные по измерениям соответствующих относительных частот, называются *апостериорными*, т. е. определенными по опытным данным.

Априорные вероятности вычисляются, в частности, в теоретической физике, например в кинетической теории газов и в квантовой теории. Априорные вероятности встречаются также и в теории азартных игр. Это приложение понятия вероятности имеет особый исторический интерес, так как изучение подобных проблем привело к возникновению в XVII в. теории вероятностей, чему мы обязаны, в частности, трудам французских математиков Ферма и Паскаля. Например, при бросании игральной кости можно ожидать, из соображений симметрии, что в длинных сериях бросков кости каждая ее сторона будет выпадать так же часто, как и любая другая. Поскольку здесь имеется всего 6 возможных результатов единичного бросания кости, ясно, что относительная частота каждого из шести отдельных результатов должна быть очень близкой к $1/6$; таким образом, возникает естественное предположение, что каждому из шести результатов следует приписать вероятность $1/6$. Соответственно мы должны приписать событию, заключающемуся в том, что карта, вытасованная наудачу из колоды в 52 карты, окажется, например, тузом треф, вероятность $1/52$, и событию, заключающемуся в том, что эта карта окажется или красной масти или королем, вероятность $28/52$. Это число $28/52$ получается, если заметить, что рассматриваемое событие осуществляется только тогда, когда выбранная карта окажется или одной из 26 карт красной масти или одним из двух королей черной масти. Однако необходимо заметить, что во всех таких априорных определениях вероятности мы принимаем, обычно не формулируя этого, некоторые гипотезы, например, что кость не „фальшивая“, что игра ведется „честно“, что карты хорошо „перетасованы“ и т. д.

Вообще говоря, задачи с азартными играми, поданные упомянутым выше, имеют две общие особенности: (а) имеется лишь конечное число возможных результатов; (б) согласно нашим знаниям об игре, мы „априори“ считаем некоторые результаты „одинаково правдоподобными“, т. е. равновероятными¹⁾.

¹⁾ Подчеркнем, что это является нашим положительным знанием, например, о физических свойствах кости, которое позволяет нам ожидать появления каждого результата как равновероятных событий. Тем не менее иногда считают, что это ожидание есть результат, например, игнорирования поведения кости. Обсуждение этого вопроса см. у H. Poincaré, Handbuch der Physik, т. IV, стр. 67, Berlin, 1929.

Классическое определение вероятности непосредственно использует эти особенности: вероятность события определяется просто как отношение количества результатов, благоприятствующих осуществлению этого события, ко всему количеству возможных (и одинаково правдоподобных) результатов. В сопоставлении с тем обстоятельством, что это определение содержит порочный круг, поскольку „одинаково правдоподобные“ можно определить лишь как „равновероятные“, что само подлежит определению,— ясно, что классическое определение применимо только тогда, когда рассматриваемые наблюдения могут дать лишь конечное число различных результатов. Более того, часто возникают сомнения в том, что следует понимать под фразой „одинаково правдоподобные результаты“, как это будет показано на следующем примере.

Пример. Пусть имеется три одинаковых стола A , B и C , и в каждом из них имеется два ящика. В обоих ящиках стола A лежит по золотой монете, в обоих ящиках стола B — по серебряной монете и, наконец, в столе C в одном ящике лежит золотая монета, а в другом — серебряная монета. Мы выбираем наудачу стол, открываем один ящик и находим золотую монету. Какова вероятность того, что выбранный стол есть стол B ? Во-первых, можно рассуждать так, что поскольку лишь столы A и C содержат золотые монеты, число возможностей равно двум, и лишь одна из них является благоприятной. Следовательно, искомая вероятность равна $1/2$. Но можно рассуждать и так, что имеется вообще 3 золотые монеты, причем стол A содержит 2 из них, а стол C — только 1. Число возможностей равно трем, из них благоприятна лишь одна, и искомая вероятность, таким образом, равна $1/3$. Какое из этих двух рассуждений правильно? Если открытый ящик содержал серебряную монету, мы должны были бы также получить по первому рассуждению ответ $1/2$ и по второму рассуждению — ответ $1/3$. Поэтому вероятность того, что выбранный стол есть стол B , не зависит от того, открывали или не открывали мы ящик. Однако если мы не открывали ящика, то выбор наудачу любого из столов A , C или C одинаково правдоподобен, и исходная вероятность равняется $1/3$. Таким образом, первое рассуждение приводит к противоречию и должно быть забраковано (см. пример 2 § 3.8).

Классическое определение вероятности, созданное в основном Лапласом, было весьма удобным в ранний период развития теории вероятностей, когда она применялась лишь для описания азартных игр. Однако, как уже было указано, это определение при описании общих статистических явлений чрезвычайно стеснительно, так как, во-первых, число возможных результатов может и не быть конечным, и, во-вторых, сведение рассматриваемого явления к некоторому количеству одинаково правдоподобных возможностей может быть неосуществимо. Например, по классическому определению невозможно определить вероятность выпадения цифры 6 при бросании испорченной кости, не считая того, что с испорченной костью можно играть „нечестно“.

В рамках классического определения мы часто не можем говорить о вероятностях некоторых событий. Например, что следует понимать под одинаково правдоподобными результатами, если требуется определить вероятность рождения мальчика? Более того, классическое определение всегда дает для вероятности рациональное число, так что для получения более общего определения вероятности, в котором вероятность могла бы быть и иррациональным числом, требуется рассуждение, аналогичное обобщению понятия числа с области рациональных чисел до области действительных чисел.

Наконец, различные авторы пытались внести понятие вероятности путями, отличными от принятого здесь способа, созданного в основном трудами Фреше, и отличными от классического определения. Известный представитель этой школы Мизес определяет вероятность события A как предел частоты n_A/n при $n \rightarrow \infty$ в определенной бесконечной серии наблюдений (которую Мизес называет *коллективом*)¹⁾ Существование этого предела постулируется как первая аксиома теории. Определение такого типа представляется на первый взгляд очень заманчивым, но более глубокий анализ показывает, что оно приводит к большим математическим трудностям¹⁾. Даже если не считать этих трудностей, такое конструктивное определение содержит смешение реальности и модели, которого избегают во всех других областях прикладной математики в интересах аксиоматического изложения. Например, понятие прямой линии в евклидовой геометрии никогда не вводят как предел линии, проведенной на доске мелом, при неограниченном уменьшении ее толщины, и также понятие о материальной точке никогда не вводят как предел реального тела с неограниченно убывающими размерами.

§ 1.6. Следует подчеркнуть, что по определению вероятности (§ 1.4) мы должны рассматривать не только природу явления, но и природу категории наблюдений. Эта последняя, как правило, в достаточной степени определяется самой задачей. Но часто недооценивают необходимости точного определения того, какие наблюдения следует произвести, а это может привести к ошибочным выводам и парадоксам. Если мы интересуемся вероятностью попадания в данную цель из данного ружья, то не возникает сомнений относительно того, какие действия необходимо здесь произвести. Однако если требуется вычислить вероятность того, что данное лицо умрет в течение года после приобретения страхового полиса, то получить ответ значительно более трудно. Для какой группы лиц должны мы вычислять относительную частоту смертности? Очевидно, будут получены различные значения искомой вероятности, если мы выберем группу лиц одного

¹⁾ См. Fréchet, *The Diverse Definitions of Probability* и *Exposé et discussion de quelques recherches récentes sur les fondements du calcul des probabilités*.

и того же возраста или же группу лиц мужского пола одного и того же возраста, или же если, кроме того, мы учтем профессию, состояние здоровья и т. д.

Указанное обстоятельство еще более наглядно иллюстрируется так называемым „парадоксом Бертрана“, в котором наблюдение заключается в случайном выборе хорды в круге радиуса r и определении ее длины. Пусть требуется найти вероятность того, что длина хорды не превышает стороны вписанного равностороннего треугольника, т. е. не превышает $r\sqrt{3}$. Поскольку процесс „случайного выбора хорды“ не является точно определенной операцией, можно получить много противоречащих друг другу результатов. Приведем лишь два следующих решения:

1. Расстояние от центра круга до хорды есть число, заключенное между 0 и r . Хорда будет иметь длину менее чем $r\sqrt{3}$, если ее расстояние от центра круга больше, чем $r/2$. Если измерять вероятность как отношение благоприятной длины $r/2$ к возможной длине r , то результатом будет $1/2$.

2. Угол, стягиваемый хордой, есть число, заключенное между 0 и π . Хорда будет иметь длину менее чем $r\sqrt{3}$, если она стягивает угол, меньший чем $2\pi/3$. Если измерять вероятность как отношение благоприятного угла $2\pi/3$ к возможному углу π , то результатом будет $2/3$.

2. ОСНОВАНИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

§ 2.1. Исследуя понятия, связанные с описанием статистических явлений, мы установим *основные законы* или *аксиомы*, т. е. такие соотношения между числами, вводимыми нами в качестве вероятностей, которые должны быть выполнены, если наша теория дает правильное описание реальных явлений. Вероятность единственного события мы можем, как правило, произвольно выбирать среди бесконечного множества чисел внутри некоторого интервала (например, между $5,8\%$ и $6,3\%$ в примере 1 § 1.4). Однако при одновременном введении вероятностей нескольких событий мы не можем устанавливать их произвольно, так как существуют некоторые соотношения между соответствующими относительными частотами, а в силу нашего определения вероятности (§ 1.4) мы должны требовать выполнения тех же соотношений и для вероятностей. Таким образом, необходимо установить некоторые основные законы, которые образуют основание для всех дальнейших выводов, но не могут быть сами доказаны математически. Их роль аналогична аксиомам евклидовой геометрии, ньютоновским законам механики и максвелловским уравнениям электродинамики.

Так как относительная частота всегда имеет вид $f = \frac{n_A}{n}$, где $0 \leq n_A \leq n$, то $0 \leq f \leq 1$, и, следовательно, то же самое нужно потребовать и от вероятностей.

Аксиома. *Значение вероятности P есть число, заключенное между 0 и 1 (или совпадающее с одним из этих чисел):*

$$0 \leq P \leq 1. \quad (I)$$

§ 2.2. Под *достоверным событием*, обозначаемым в дальнейшем через E , понимается событие, наступающее при каждом наблюдении. Относительная частота $f(E)$ достоверного события всегда поэтому равна

$$f(E) = \frac{n}{n} \equiv 1,$$

и, следовательно, то же самое нужно потребовать и от его вероятности $P(E)$.

Аксиома. *Вероятность достоверного события равна единице*

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

Подчеркнем, что обратное заключение не всегда было бы правильным. Если событие имеет вероятность 1 (такое событие мы будем называть *стохастически достоверным*), то это означает не достоверность его в обычном смысле этого слова, а только его *практическую достоверность*, т. е. то, что его относительная частота весьма близка к единице, если число наблюдений достаточно велико (ср. § 9.2). В дальнейшем под словами „достоверное событие“ будет подразумеваться стохастически достоверное событие, причем мы не будем каждый раз оговаривать, что это словупотребление несколько отличается от общепринятого.

§ 2.3. Под *невозможным событием*, обозначаемым в дальнейшем через O , понимается событие, которое не может наступить ни при каком наблюдении. Относительная частота $f(O)$ невозможного события всегда равна поэтому

$$f(O) = \frac{0}{n} \equiv 0,$$

и, следовательно, то же самое нужно потребовать и от его вероятности $P(O)$.

III аксиома. *Вероятность невозможного события равна нулю*

$$P(O) = 0. \quad (\text{III})$$

Как и в § 2.2, обратное заключение не всегда будет правильным. Если событие имеет вероятность 0 (такое событие мы будем называть *стохастически невозможным*), то это означает только его *практическую невозможность*, т. е. то, что его относительная частота весьма близка к нулю, если число наблюдений достаточно велико (ср. § 9.2). В дальнейшем, говоря о невозможном событии, мы будем подразумевать стохастически невозможное событие, не оговаривая каждый раз того, что это словупотребление несколько отличается от обычного.

§ 2.4. Рассмотрим два события A и B и серию из n наблюдений. В каждом наблюдении может иметь место одна и только одна из следующих четырех возможностей:

1. Событие A наступает, а событие B нет.
2. Событие B наступает, а событие A нет.
3. Наступают оба события A и B .
4. Ни событие A , ни событие B не наступают. Обозначим через n_1, n_2, n_3, n_4 числа наступлений каждой из соответствующих возможностей в серии из n наблюдений. Тогда

$$n_1 + n_2 + n_3 + n_4 = n. \quad (1)$$

Теперь мы можем образовать следующие относительные частоты:

Относительную частоту события A (независимого от B):

$$f(A) = \frac{n_1 + n_3}{n}. \quad (2)$$

Относительную частоту события B (независимого от A):

$$f(B) = \frac{n_2 + n_3}{n}. \quad (3)$$

Относительную частоту наступления либо события A , либо события B , либо их обоих вместе (соответствующее событие мы будем обозначать через $A + B$):

$$f(A + B) = \frac{n_1 + n_2 + n_3}{n}. \quad (4)$$

Относительную частоту одновременного наступления обоих событий A и B (соответствующее событие мы будем обозначать через AB)¹⁾:

$$f(AB) = \frac{n_3}{n}. \quad (5)$$

Относительную частоту наступления события B при условии наступления события A (соответствующее событие мы будем обозначать через $B|A$):

$$f(B|A) = \frac{n_3}{n_1 + n_3}. \quad (6)$$

Относительную частоту наступления события A при условии наступления события B :

$$f(A|B) = \frac{n_3}{n_2 + n_3}. \quad (7)$$

§ 2.5. Для шести величин (2.4.2) — (2.4.7) имеют место следующие соотношения:

$$f(A + B) = f(A) + f(B) - f(AB), \quad (1)$$

$$f(AB) = f(A)f(B|A) = f(B)f(A|B). \quad (2)$$

Следовательно, как уже отмечалось в § 2.1—2.3, то же самое нужно потребовать и от соответствующих вероятностей. Таким образом, равенство (1) приводит к следующей аксиоме:

IV аксиома. *Вероятность наступления по крайней мере одного из двух событий равна сумме вероятностей каждого события минус вероятность одновременного наступления обоих событий:*

$$P(A + B) = P(A) + P(B) - P(AB). \quad (IV)$$

Эта аксиома называется *законом сложения вероятностей*.

¹⁾ Употребление в этом случае символов $A + B$ и AB оправдывается тем, что к ним в большой степени применимы обычные алгебраические правила.

Пример. Из двух отдельных, хорошо стасованных колод вытаскивается по карте. Вероятность того, что по крайней мере одна из этих карт окажется дамой червей, равна

$$\frac{1}{52} + \frac{1}{52} - \frac{1}{52 \cdot 52} = \frac{103}{2704}.$$

Третий член в формуле (IV) вычисляется здесь следующим образом: вычисляя вероятность того, что обе вытасканные карты окажутся дамами червей, мы имеем $52 \cdot 52$ возможностей (число различных комбинаций двух карт, вынутых из различных колод). Однако только одна из этих $52 \cdot 52$ возможностей дает благоприятный результат — две дамы червей.

§ 2.6. Вероятность $P(B|A)$, которая соответствует относительной частоте $f(B|A)$, определенной равенством (2.4.6), называется *условной вероятностью* события B при условии A . Таким образом, $P(B|A)$ дает вероятность наступления события B при условии, что событие A уже наступило. В отличие от $P(B|A)$ вероятность $P(B)$ называется *абсолютной вероятностью* события B . Вероятность $P(A|B)$ определяется соответствующим образом. Равенство (2.5.2) приводит к последней аксиоме, которую нам нужно установить:

В аксиома. Вероятность одновременного наступления двух событий равна произведению абсолютной вероятности одного из этих событий на условную вероятность другого:

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

Эта аксиома называется *законом умножения* вероятностей.

Пример. Из одной колоды вытаскиваются последовательно две карты, причем вынутая карта уже не возвращается в колоду. Вероятность того, что обе вынутые карты — червонной масти, равна $\frac{13}{52} \frac{12}{51} = \frac{1}{17}$. Если вытащить из колоды карту червонной масти, то в колоде останется 51 карта, из которых 12 — червонной масти. Поэтому условная вероятность второй карте оказаться червонной масти равна $\frac{12}{51}$.

§ 2.7. Математическая теория вероятностей строится на основе пяти аксиом (I — V). Вероятности, встречающиеся в равенствах, которые получаются из этих законов с помощью математической дедукции, являются просто произвольными числами, удовлетворяющими сформулированным аксиомам, независимо от того, каким способом эти числа были получены. Однако в практических приложениях теории при упот-

реблении слова „вероятность“ следует учитывать, каким образом она измеряется на практике, т. е. что она является числом, около которого группируются относительные частоты рассматриваемых событий, — „истинным“ значением этих относительных частот. В этой связи часто бывает удобно представлять себе класс всех результатов, к которым может привести рассматриваемое наблюдение, и затем понимать вероятность как „относительную частоту“ события в этом воображаемом бесконечном классе, который обозначается такими словами, как *популяция, множество, группа, генеральная совокупность* или *коллектив* ¹⁾. Так как, вообще говоря, эти относительные частоты будут иметь вид ∞/∞ , то невозможно без дополнительных уточнений придать этой интерпретации вполне определенный характер, но мы должны относиться к ней как к возможному краткому описанию вводимых вероятностей ²⁾. Если учитывать это, то подобная интерпретация может часто иметь большое значение, как эвристическое, так и мнемоническое, особенно в математической статистике.

Каждая задача в теории вероятностей будет иметь теперь следующий вид. Отправляются от некоторых вероятностей P_1, P_2, \dots , которые заданы некоторыми теоретическими рассмотрениями или гипотезами, т. е. априори, или заданы опытным путем соответствующими относительными частотами f_1, f_2, \dots , т. е. апостериори, или, наконец, заданы как произвольные константы, значения которых должны быть определены позднее. Такие константы называются *параметрами*. Затем с помощью математических теорем, выведенных из аксиом I—V, вычисляются некоторые другие вероятности P'_1, P'_2, \dots как функции от P_1, P_2, \dots . Заданные значения вероятностей P_1, P_2, \dots подставляются в эти функции, и, наконец, проверка теории заключается в сравнении полученных таким образом значений с соответствующими относительными частотами f'_1, f'_2, \dots , являющимися экспериментальными значениями вероятностей P'_1, P'_2, \dots (ср. раздел 9). Если, с другой стороны, P_1, P_2, \dots суть произвольные параметры или величины, зависящие от таких параметров, то мы пытаемся выбрать их численные значения таким образом, чтобы получить возможно лучшее соответствие между теоретическими и экспериментальными значениями вероятностей P'_1, P'_2, \dots .

¹⁾ Отметим, что, к сожалению, те же самые слова употребляются также для обозначения конечного класса наблюдаемых результатов, что приводит к смешению реальности с нашей моделью ее (ср. § 1.4). Поэтому мы будем говорить либо об эмпирической, либо о теоретической популяции, совокупности и т. д.

²⁾ В высшей математике, а именно в абстрактной теории меры, эта интерпретация получает вполне определенное значение. Она позволяет формулировать основы теории вероятностей в весьма общей форме, что впервые было проведено систематически А. Н. Колмогоровым в 1933 г. и изложено в его книге „Основные понятия теории вероятностей“ (Москва, ГТТИ, 1936).

Подчеркнем два следующих факта. Во-первых, теория вероятностей всегда должна исходить из некоторых заданных в рассматриваемой задаче вероятностей, точно так же как механика — из начальных положений и скоростей тел, движение которых она описывает. Во-вторых, теория вероятностей всегда дает результаты в форме вероятностей. Следовательно, теория вероятностей ничего не может сказать относительно исхода одного единственного события.

В разделах 3—8 мы дадим чисто математическую теорию, возвращаясь в разделе 9 к отношению между теорией и опытом, и, наконец, в разделах 10—12 мы рассмотрим различные практические приложения теории.

3. ЭЛЕМЕНТАРНЫЕ ТЕОРЕМЫ

§ 3.1. При некоторых предположениях закон сложения вероятностей, выраженный аксиомой IV, приводится к более простой форме. Если предположить, что события A и B *несовместимы* (т. е. не могут иметь места одновременно, $AB = O$), то в силу аксиомы III $P(AB) = 0$ и аксиома IV примет вид

$$P(A + B) = P(A) + P(B). \quad (1)$$

Если два события несовместимы, то вероятность наступления хотя бы одного из этих событий равна сумме вероятностей каждого события в отдельности.

Пример. При бросании игральной кости вероятность выпадения либо 3, либо 4 равна

$$\frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

§ 3.2. Если $P(B|A) = P(B)$, т. е. если вероятность события B не зависит от наступления или ненаступления события A , то говорят, что событие B *стохастически не зависит* от события A . Из аксиомы V в предположении, что $P(B)$ отлично от нуля, вытекает, что $P(A|B) = P(A)$, т. е. что событие A стохастически не зависит от события B . Поэтому мы можем говорить, что события A и B *стохастически независимы*. В этом случае аксиома V принимает вид

$$P(AB) = P(A)P(B). \quad (1)$$

Если два события стохастически независимы, то вероятность наступления обоих этих событий равна произведению вероятностей каждого события в отдельности.

Пример. Из двух колод вытаскивается по одной карте. Вероятность того, что обе вынутые карты окажутся червонной масти, равна

$$\frac{13}{52} \cdot \frac{13}{52} = \frac{1}{16}$$

(ср. пример § 2.6).

При применении теорем (3.1.1) и (3.2.1) необходимо иметь в виду, что они имеют место только при сформулированных условиях. Это легко упустить из вида и прийти к неверным результатам. В таких простых применениях, как в азартных играх, использованных здесь

в качестве примеров, как правило, легко проверить, выполнены ли нужные условия или нет. Однако в более сложных случаях такая проверка часто оказывается более трудной и тогда предпочтительнее пользоваться непосредственно аксиомами IV и V.

Отметим, что понятия *стохастической* и *причинной независимости* не равнозначны, хотя на практике часто предполагается, что если события причинно независимы, то $P(A|B) = P(A)$, т. е. они и стохастически независимы и обратно. В дальнейшем мы будем для краткости опускать слово „стохастически“ и говорить просто о независимости событий, подразумевая их стохастическую независимость. Таким образом, на практике предположение независимости двух событий является, как правило, гипотезой, справедливость которой устанавливается только опытом.

§ 3.3. Теоремы (3.1.1) и (3.2.1) могут быть легко распространены на случай любого конечного числа событий. Событие, заключающееся в наступлении хотя бы одного из n событий A_1, A_2, \dots, A_n , обозначается символически через $A_1 + A_2 + \dots + A_n$. Пусть эти n событий попарно несовместимы, что мы будем символически записывать так: $A_i A_k = 0$, $i \neq k$. При этих условиях мы в силу (3.1.1) получаем

$$P(A_1 + A_2 + \dots + A_n) = P(A_1) + P(A_2) + \dots + P(A_n). \quad (1)$$

У п р а ж н е н и е. Проверить равенство (1).

В интересах общей теории оказывается необходимым обобщение равенства (1) на случай бесконечного множества попарно несовместимых событий, хотя такой случай, конечно, никогда не может встретиться на практике. Так как это обобщение не следует из равенства (1), то мы сформулируем его в виде особой независимой аксиомы, называемой *аксиомой полной аддитивности*.

VI аксиома. Если A_1, A_2, \dots есть произвольная бесконечная последовательность попарно несовместимых событий, т. е. $A_i A_k = 0$, $i \neq k$, то

$$P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \quad (VI)$$

§ 3.4. Событие, заключающееся в одновременном наступлении n событий A_1, A_2, \dots, A_n , обозначается символически через $A_1 A_2 \dots A_n$. Если каждое из этих n событий независимо от любой комбинации других событий, то в силу (3.2.1)

$$P(A_1 A_2 \dots A_n) = P(A_1) P(A_2) \dots P(A_n). \quad (1)$$

У п р а ж н е н и е. Проверить равенство (1).

§ 3.5. Несомненность наступления по крайней мере одного из n событий A_1, A_2, \dots, A_n записывается символически так:

$$A_1 + A_2 + \dots + A_n = E. \quad (1)$$

В этом случае в силу аксиомы II

$$P(A_1 + A_2 + \dots + A_n) = 1. \quad (2)$$

Далее, если эти n событий попарно несовместимы, т. е. если $A_i A_k = O$, $i \neq k$, то в силу (3.3.1) и (2)

$$P(A_1) + P(A_2) + \dots + P(A_n) = 1. \quad (3)$$

Это важное равенство часто бывает полезно при контроле правильности вычислений и должно применяться там, где это возможно. (См., например, § 3.7.)

Через \bar{A} мы обозначаем событие, заключающееся в ненаступлении события A . Событие \bar{A} называется *противоположным* или *дополнительным* по отношению к событию A . Так как по крайней мере одно из событий A и \bar{A} должно несомненно наступить и так как оба они не могут наступить одновременно, то в наших обозначениях $A + \bar{A} = E$ и $A\bar{A} = O$. В силу (3) имеем

$$P(A) = 1 - P(\bar{A}). \quad (4)$$

§ 3.6. Часто оказывается полезным вычислить вероятность дополнительного события \bar{A} вместо вероятности самого события A . Следующая задача дает знаменитый исторический пример.

В 1654 г. французский игрок де-Мере заявил в письме к Паскалю, что сделал одно наблюдение, результат которого его удивил. Бросая одну игральную кость четыре раза, он заметил, что вероятность выпадения по крайней мере один раз шестерки больше, чем вероятность выпадения двух шестерок при 24 бросаниях двух игральных костей. При бросании одной кости имеется шесть возможных результатов, из которых один считается благоприятствующим; при бросании двух игральных костей возможных результатов в шесть раз больше, а благоприятный результат попрежнему только один. Де-Мере поэтому считал, что при бросании двух костей количество бросков надо увеличить в шесть раз, чтобы сохранить прежнюю вероятность благоприятного результата. Однако Паскаль показал, что это рассуждение ошибочно. В первом случае нам нужно вычислить вероятность P_1 выпадения шестерки по крайней мере один раз при четырех бросках одной кости. Противоположным событием является невыпадение шестерки ни в одном из четырех бросков. Эти четыре броска независимы, и поэтому в силу (3.4.1) вероятность этого события равна $(\frac{5}{6})^4$, так как вероятность невыпадения шестерки при одном броске равна

§/6. Следовательно, в силу (3.5.4)

$$P_1 = 1 - \left(\frac{5}{6}\right)^4 = 0,518.$$

Аналогичным образом Паскаль нашел, что вероятность выпадения по крайней мере один раз двух шестерок при 24 бросках двух костей равна

$$P_2 = 1 - \left(\frac{35}{36}\right)^{24} = 0,491.$$

Эти результаты согласовывались с наблюдениями де-Мере, что P_1 немного больше, а P_2 немного меньше, чем $1/2$.

§ 3.7. Задача Бернулли. Рассмотрим событие A с вероятностью θ . Проведем ν наблюдений и будем интересоваться вероятностью P_r того, что событие A наступит ровно в r из этих наблюдений.

Вычислим сперва вероятность того, что событие A наступает в первых r из ν наблюдений и не наступает в остальных. В силу (3.4.1) она равна

$$\theta^r (1 - \theta)^{\nu - r}. \quad (1)$$

Эта вероятность не зависит от того, в каких именно r наблюдениях происходит наступление события A . Поэтому в силу (3.3.1) искомая вероятность P_r равна вероятности (1), умноженной на число различных способов, которыми можно выбрать r элементов из ν :

$$C_r^\nu = \binom{\nu}{r} = \frac{\nu(\nu-1)\dots(\nu-r+1)}{1 \cdot 2 \dots r} = \frac{\nu!}{r!(\nu-r)!}.$$

Следовательно,

$$P_r = \binom{\nu}{r} \theta^r (1 - \theta)^{\nu - r}. \quad (2)$$

Это выражение называется *биномиальным законом*¹⁾. Оно дает вероятность наступления события A ровно r раз в ν наблюдениях. В силу (3.3.1) вероятность $P_{\geq r}$ наступления события A по крайней мере r раз равна

$$P_{\geq r} = P_r + P_{r+1} + \dots + P_\nu = \sum_{i=r}^{\nu} P_i = \sum_{i=r}^{\nu} \binom{\nu}{i} \theta^i (1 - \theta)^{\nu - i}. \quad (3)$$

Полагая в равенстве (3) $r=0$, получим по биномиальной формуле

$$P_{\geq 0} = \sum_{i=0}^{\nu} \binom{\nu}{i} \theta^i (1 - \theta)^{\nu - i} = [\theta + (1 - \theta)]^\nu = 1. \quad (4)$$

Так как r может принимать только одно из значений $0, 1, 2, \dots, \nu$, то равенство (4) согласуется с (3.5.3).

1) Величины $\nu, \theta, 1 - \theta$ часто обозначаются соответственно через n, p, q

* * * Пример. Формула (1) имеет важное физическое применение. Пусть вероятность распада радиоактивного атома за „бесконечно малое“ время dt равна λdt . Какова вероятность нераспада атома за время t ? Предполагая, что атом возник в момент $t=0$, разделим интервал времени от 0 до t на ν равных частей длины $\Delta t = \frac{t}{\nu}$ и применим (1), полагая $r=0$ и $\theta = \lambda \Delta t$. Переходя к пределу при $\nu \rightarrow \infty$, получим искомую вероятность

$$\lim_{\nu \rightarrow \infty} (1 - \lambda \Delta t)^\nu = e^{-\lambda t}, \quad (5)$$

так как $\lim_{x \rightarrow 0} (1 + x)^{\frac{1}{x}} = e = 2,718 \dots$. Интервал времени T , для которого вероятность распада равна $\frac{1}{2}$, называется периодом полураспада и в силу (5) равен $\frac{\ln 2}{\lambda}$.

Упражнение 1. Орудийная батарея A имеет вероятность поражения одним выстрелом, равную $\frac{3}{5}$; батарея B имеет вероятность поражения, равную $\frac{1}{2}$. Батарея A делает два выстрела, батарея B делает три выстрела. Найти вероятности того, что произошло ровно 0, 1, 2, 3, 4, 5 поражений. Проверить, что их сумма равна единице.

Упражнение 2. Применить равенство (3) к решению задачи де-Мере § 3.6.

* § 3.8. Теорема Байеса. В аксиоме V мы можем считать величины $P(A)$ и $P(AB)$ заданными; тогда при $P(A) \neq 0$ получим

$$P(B|A) = \frac{P(AB)}{P(A)}. \quad (1)$$

Пример 1. Вероятность того, что атом не распадется до момента $t + t'$ (событие B) при условии, что он не распался к моменту t (событие A), определяется равенствами (1) и (3.7.5)

$$P(B|A) = \frac{e^{-\lambda t(t+t')}}{e^{-\lambda t}} = e^{-\lambda t'}.$$

Таким образом, вероятность нераспада радиоактивного атома не зависит от „возраста“ атома, в противоположность тому, что имеет место в биологических процессах.

Подставляя $P(AB) = P(B)P(A|B)$ в равенство (1), получим

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)}. \quad (2)$$

Рассмотрим теперь ν попарно несовместимых событий A_1, A_2, \dots, A_ν , одно из которых обязательно должно наступить. В символической записи $A_i A_k = 0$, $i \neq k$ и $A_1 + A_2 + \dots + A_\nu = E$. Далее, пусть X

*) Места, отмеченные звездочкой, имеют менее элементарный характер и могут быть опущены при первом чтении.

есть произвольное событие (не обязательно являющееся одним из событий A_1, A_2, \dots, A_ν). Очевидно, что вместе с событием X должно наступить одно из событий A_1, A_2, \dots, A_ν , т. е. при наступлении события X наступает одно из событий $A_1X, A_2X, \dots, A_\nu X$, что символически записывается в виде

$$X = A_1X + A_2X + \dots + A_\nu X. \quad (3)$$

Так как события A_i и A_k несовместимы при $i \neq k$, то то же самое имеет место и для событий A_iX и A_kX . Поэтому в силу (3)

$$P(X) = P(A_1X) + P(A_2X) + \dots + P(A_\nu X). \quad (4)$$

Подставляя $P(A_iX) = P(A_i)P(X|A_i)$ в равенство (4), получим

$$P(X) = P(A_1)P(X|A_1) + \dots + P(A_\nu)P(X|A_\nu). \quad (5)$$

Полагая в равенстве (2) $A = X$ и $B = A_i$ и подставляя $P(X)$ из равенства (5), получим, наконец, *теорему Байеса*

$$P(A_i|X) = \frac{P(A_i)P(X|A_i)}{P(X)} = \frac{P(A_i)P(X|A_i)}{P(A_1)P(X|A_1) + \dots + P(A_\nu)P(X|A_\nu)}. \quad (6)$$

Итак, пусть событие X наступило в результате одного из ν парно несовместимых событий A_1, \dots, A_ν . Тогда события A_1, \dots, A_ν называются „гипотезами“ события X , и соотношение (6) дает вероятность „гипотезы“ A_i после наступления события X .

Пример 2. Применим теорему Байеса к примеру § 1.5. Гипотезами A_1, A_2, A_3 являются в этом случае соответственно выбор стола A , выбор стола B и выбор стола B . Событие X заключается в том, что наудачу выбранный ящик стола содержит золотую монету. Таким образом, имеем

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}.$$

Далее, $P(X|A_1) = 1$, так как оба ящика стола A содержат по золотой монете; $P(X|A_2) = \frac{1}{2}$, так как только один ящик стола B содержит золотую монету; и, наконец, $P(X|A_3) = 0$, так как ни один из ящиков стола B не содержит золотой монеты. Подставляя эти числа в (6), получим

$$P(A_2|X) = \frac{\frac{1}{3} \cdot \frac{1}{2}}{\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 0} = \frac{1}{3}$$

в соответствии с результатом, найденным в § 1.5.

Упражнение. Выписать теорему Байеса для случая

$$X = \sum_{i=1}^m A_i, \quad m < \nu.$$

4. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

§ 4.1. В дальнейшем мы будем иметь дело только с такими случайными явлениями, непосредственно наблюдаемый результат которых выражается одним или несколькими числами¹⁾. Величина, которая может быть количественно определена и которая в различных наблюдениях одной и той же категории может принимать различные значения, будет называться случайной величиной²⁾ и обозначаться символически через x . Таким образом, буква x символизирует не только рассматриваемую величину, но также и способ, которым она измеряется, категорию наблюдений (ср. § 1.6). Примерами случайных величин являются: результат бросания игральной кости; вес и рост у некоторой группы солдат; отклонение по дальности, отклонение по высоте и боковое отклонение при стрельбе из орудия. В последнем из упомянутых примеров величина x не определена до тех пор, пока не указаны необходимые данные, как-то: тип и положение орудия, состав пороха, вес снаряда и тому подобное. Будем предполагать сначала, что рассматриваемые случайные величины являются *одномерными*, т. е. могут быть охарактеризованы с помощью одного числа.

§ 4.2. Вероятность того, что величина x принимает значение t , обозначается через $P(x=t)$, а вероятность того, что величина x принимает значение из интервала $a < t \leq b$, — через $P(a < x \leq b)$ или просто через $P(a, b)$. Если при всех a и b известно значение функции $P(a, b)$, то говорят, что известно *распределение* величины x , или известно, как величина x *распределена*. Удобный способ для описания распределения дает так называемая функция распределения. Под *функцией распределения*³⁾ $\Phi(t)$ случайной величины x понимается функция,

1) Даже в тех случаях, когда непосредственно наблюдаемый результат не является числом, мы можем символизировать его посредством числа. Так, в примере, приведенном в конце § 1.3, мы можем связывать результат „мужской пол“ с числом 1, а результат „женский пол“ с числом 0.

2) Случайные величины называются также иногда *статистическими* или *стохастическими*.

3) $\Phi(t)$ называется также иногда полной или интегральной функцией распределения, для того чтобы отличить ее от функции $f(t)$, рассматриваемой ниже (§ 4. 4). Функция распределения часто обозначается через $F(t)$, но мы, в соответствии со сказанным в конце § 1.4, сохраняем последнее обозначение для соответствующей эмпирической функции (см. раздел 10).

которая для всех значений t равна вероятности того, что $x \leq t$:

$$\Phi(t) = P(-\infty < x \leq t) = P(-\infty, t). \quad (1)$$

Если одновременно рассматриваются несколько случайных величин x, y, \dots , то соответствующие функции распределения обозначаются $\Phi_x(t), \Phi_y(t), \dots$. В силу (3. 1. 1) имеем при $t_1 < t_2$

$$P(-\infty, t_2) = P(-\infty, t_1) + P(t_1, t_2)$$

и, подставляя в это равенство (1), получим в силу $P(t_1, t_2) \geq 0$

$$\Phi(t_1) \leq \Phi(t_2). \quad (2)$$

Таким образом, функция распределения $\Phi(t)$ является неубывающей функцией.

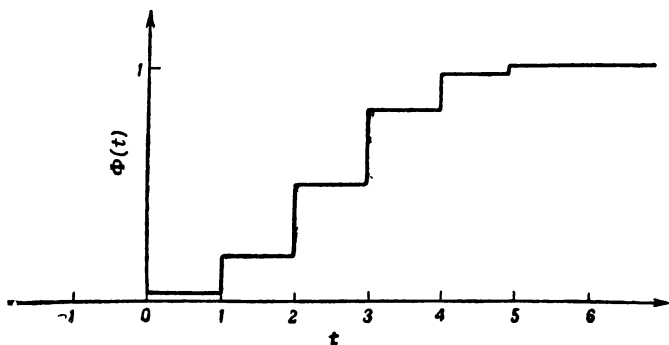
Пример. Вероятность поражения цели при выстреле из винтовки равна $\frac{1}{2}$. Делается пять выстрелов. Случайная величина является числом поражений. Вероятности $P(x=r)$, $r=0, 1, 2, 3, 4, 5$, даются биномиальным законом (3. 7. 2), поэтому $\Phi(t)$ имеет следующий вид:

$$\Phi(t) = \begin{cases} 0 & \text{при } -\infty < t < 0 \\ \frac{1}{32} & \text{при } 0 \leq t < 1 \\ \frac{6}{32} & \text{при } 1 \leq t < 2 \\ \frac{16}{32} & \text{при } 2 \leq t < 3 \\ \frac{26}{32} & \text{при } 3 \leq t < 4 \\ \frac{31}{32} & \text{при } 4 \leq t < 5 \\ 1 & \text{при } 5 \leq t < \infty \end{cases}$$

График функции $u = \Phi(t)$ показан на фиг. 2. Случайная величина, которая может принимать только значения $r=1, 2, \dots$, с вероятностями P_r , заданными равенствами (3. 7. 2), называется *биномиально распределенной*.

Заметим, что в последнем примере функция распределения $\Phi(t)$ непрерывна для всех значений t , для которых вероятность того, что величина x принимает это значение, равна нулю, и разрывна во всех остальных точках, причем скачок в каждой такой точке равен вероятности того, что величина x принимает соответствующее значение t . Заметим, далее, что в точке разрыва функция $\Phi(t)$ равна своему пределу справа. Функция $\Phi(t)$ имеет предел 0 при $t \rightarrow -\infty$ и предел 1 при $t \rightarrow \infty$.

В самом общем случае можно с помощью аксиом I—VI доказать, что любая функция распределения обладает этими свойствами¹⁾:



Ф и г. 2.

$$\lim_{t \rightarrow \infty} \Phi(t) = \lim_{t \rightarrow \infty} P(-\infty < x \leq t) = 1, \quad (3)$$

т. е. функция $\Phi(t)$ стремится к 1 при $t \rightarrow \infty$.

$$\lim_{t \rightarrow -\infty} \Phi(t) = \lim_{t \rightarrow -\infty} P(-\infty < x \leq t) = 0, \quad (4)$$

т. е. функция $\Phi(t)$ стремится к 0 при $t \rightarrow -\infty$.

$$\lim_{t \rightarrow t_0} \Phi(t) = \Phi(t_0), \quad (t > t_0), \quad (5)$$

т. е. функция $\Phi(t)$ стремится к $\Phi(t_0)$, когда t приближается к t_0 справа.

$$\lim_{t \rightarrow t_0} [\Phi(t_0) - \Phi(t)] = P(x = t_0) \quad (t < t_0), \quad (6)$$

т. е. разность $\Phi(t_0) - \Phi(t)$ стремится к $P(x = t_0)$, когда t приближается к t_0 слева. Таким образом, в точке t_0 непрерывности функции $\Phi(t)$ имеет место $P(x = t_0) = 0$.

Функции распределения $\Phi(t)$ случайных величин, встречающихся на практике, принадлежат, как правило, к одному из следующих двух типов.

I. $\Phi(t)$ есть кусочно-постоянная функция (как на фиг. 2).

¹⁾ См., например, Г. Крамер, Случайные величины и распределения вероятностей, гл. 2, 1947.

²⁾ Здесь допускается, что x может принимать только конечные значения. Если x может с положительной вероятностью принимать значения ∞ и $-\infty$, то $\lim_{t \rightarrow \infty} \Phi(t) < \Phi(\infty) = 1$ и $\lim_{t \rightarrow -\infty} \Phi(t) = \Phi(-\infty) > 0$. Это имеет место, например, тогда, когда мы переходим от величины, принимающей с положительной вероятностью значения 0 к ее логарифму, который будет тогда принимать значение $-\infty$ с положительной вероятностью.

II. $\Phi(t)$ есть непрерывная, кусочно-дифференцируемая функция с непрерывной производной.

В этих случаях мы говорим соответственно о дискретных и непрерывных функциях распределения. В обоих случаях аналитические выражения этих функций содержат часто одну или несколько постоянных, называемых *параметрами*, которые мы будем обозначать греческими буквами. Численные значения этих параметров в каждом отдельном случае подбираются по наблюдениям так, чтобы теоретическое распределение описывало эти наблюдения наилучшим образом (см. раздел 10).

§ 4.3. Дискретные функции распределения. Функция распределения $\Phi(t)$ называется *дискретной*, если она кусочно-постоянна, т. е. если существуют некоторые точки

$$\dots < t_{-2} < t_{-1} < t_0 < t_1 < t_2 < \dots,$$

в которых функция $\Phi(t)$ имеет положительные скачки

$$\dots, \varphi_{-2}, \varphi_{-1}, \varphi_0, \varphi_1, \varphi_2, \dots$$

причем функция $\Phi(t)$ постоянна между двумя последовательными значениями t_i . Значения t_i называются часто *спектром* случайной величины x . Если спектр состоит из значений $0, 1, 2, \dots$, то более удобно писать просто i вместо t_i .

Таким образом, функция $u = \Phi(t)$ является ступенчатой функцией, лежащей между прямыми $u = 0$ и $u = 1$, со скачками величины φ_i в точках разрыва t_i . Число скачков может быть как конечным, так и бесконечным. В силу (4. 2. 6) величина скачка φ_i равна вероятности того, что случайная величина x принимает значение t_i . Для всех других значений t имеем $P(x = t) = 0$, в силу непрерывности функции $\Phi(t)$ в этой точке.

Имеем

$$\sum_{i=q}^p \varphi_i = \Phi(t_p) - \Phi(t_{q-1}).$$

Полагая $q \rightarrow -\infty$, получим в силу (4. 2. 4)

$$\Phi(t_p) = \sum_{i=-\infty}^p \varphi_i \quad (1)$$

и, полагая затем $p \rightarrow \infty$, получим в силу (4. 2. 3)

$$\sum_{i=-\infty}^{\infty} \varphi_i = 1. \quad (2)$$

Пример 1. В простейшем случае дискретного распределения случайная величина x принимает единственное значение μ с вероятностью,

равной единице. В этом случае спектр состоит из одной точки и

$$\begin{aligned} t_0 &= \mu \\ \varphi_0 &= P(x = \mu) = 1 \\ \Phi(t) &= P(x \leq t) = \varepsilon(t - \mu) = \begin{cases} 0 & \text{при } t < \mu \\ 1 & \text{при } t \geq \mu. \end{cases} \end{aligned} \quad (3)$$

Здесь имеется один параметр μ . Такое распределение мы будем называть *несобственным*.

Пример 2. Примером дискретного распределения, часто встречающегося на практике, особенно в биологии, является *биномиальное распределение*, уже рассматривавшееся в примере § 4.2. Из соотношений (3.7.2) и (3.7.4) имеем

$$\begin{aligned} i &= 0, 1, 2, \dots, \nu, \\ \varphi_i &= P_i = \binom{\nu}{i} \theta^i (1 - \theta)^{\nu - i}, \\ \sum_{i=0}^{\nu} \varphi_i &= 1. \end{aligned} \quad (4)$$

Здесь θ и ν являются параметрами, но в большинстве случаев только параметр θ остается свободным для выбора, параметр же ν определяется самой задачей.

Примером дискретного распределения может служить также распределение любой случайной величины, принимающей только целочисленные значения. Ниже мы дадим три таких примера, имеющих существенное значение для многих практических приложений.

Пример 3. При *распределении Пуассона* спектр состоит из всех неотрицательных целых чисел. Это распределение задается соотношениями

$$\begin{aligned} i &= 0, 1, 2, \dots, \\ \varphi_i &= e^{-\mu} \frac{\mu^i}{i!}, \quad \mu > 0, \\ \sum_{i=0}^{\infty} \varphi_i &= e^{-\mu} \sum_{i=0}^{\infty} \frac{\mu^i}{i!} = e^{-\mu} e^{\mu} = 1, \end{aligned} \quad (5)$$

где μ является параметром.

Распределение Пуассона часто встречается на практике, как показывают следующие примеры.

Пример 4. Пусть случайная величина x есть число вызовов за время t на телефонном коммутаторе. При предположениях, что а) вероятность вызова за время dt равна λdt , т. е. пропорциональна dt , и б) вызовы являются независимыми событиями, случайная величина x имеет распределение Пуассона.

Обозначая вероятность i вызовов за время t через $P_i(t)$, получим, что при $i > 0$ вероятность $P_i(t + dt)$ равна произведению вероятности

стей i вызовов за время t и 0 вызовов за время dt плюс произведение вероятностей $i-1$ вызовов за время t и одного вызова за время dt . (Так как dt есть дифференциал, то вероятность более чем одного вызова за время dt есть величина высшего порядка малости и ею можно пренебречь.) Вероятность $P_0(t+dt)$ равна произведению вероятностей 0 вызовов за время t и 0 вызовов за время dt . Поэтому, при наших предположениях,

$$P_i(t+dt) = P_i(t)(1-\lambda dt) + P_{i-1}(t)\lambda dt, \quad i > 0,$$

$$P_0(t+dt) = P_0(t)(1-\lambda dt),$$

откуда

$$\left. \begin{aligned} \frac{dP_i(t)}{dt} &= \lambda(P_{i-1}(t) - P_i(t)), \quad i > 0 \\ \frac{dP_0(t)}{dt} &= -\lambda P_0(t). \end{aligned} \right\} \quad (6)$$

Легко видеть, что решение этой бесконечной системы дифференциальных уравнений дается формулой Пуассона с параметром $\mu = \lambda t$:

$$P_i(t) = e^{-\lambda t} \frac{(\lambda t)^i}{i!}, \quad P_i(0) = \begin{cases} 1, & i=0 \\ 0, & i>0 \end{cases} \quad (7)$$

(Полагая $P_i(t) = \frac{(\lambda t)^i}{i!} f(t)$, мы можем из уравнений (6) при начальных условиях (7) найти, что $f(t) = e^{-\lambda t}$.) Формула (7) очень важна, например, при проектировании телефонных станций¹⁾.

Пример 5. Распределение Пуассона, определяемое соотношениями (5), имеет много других применений, так как условия, аналогичные условиям а) и б), оказываются характерными для целого ряда задач. Так, соотношения (7) задают также вероятность распадаения i радиоактивных атомов за время t , вероятность попадания i частиц космических лучей в счетчик Гейгер-Мюллера за время t , вероятность магазину распродать i штук некоторого товара за время t , вероятность прохождения i автомобилей по некоторой улице за время t , вероятность i самоубийств за время t , вероятность наличия i частиц на площади t под микроскопом и т. д.

Пример 6. Распределение Паскаля определяется соотношениями

$$i = 0, 1, 2, \dots$$

$$\varphi_i = \frac{1}{1+\mu} \left(\frac{\mu}{1+\mu} \right)^i, \quad \mu > 0, \quad (8)$$

$$\sum_{i=0}^{\infty} \varphi_i = \frac{1}{1+\mu} \sum_{i=0}^{\infty} \left(\frac{\mu}{1+\mu} \right)^i = \frac{1}{1+\mu} \frac{1}{1 - \frac{\mu}{1+\mu}} \equiv 1,$$

¹⁾ См., например, Т. Фрай, Теория вероятностей для инженеров, ГТТИ, 1934.

где μ есть параметр. (В теории космических лучей это распределение называется также распределением Фарри.)

Пример 7. Распределение Поля определается следующими соотношениями, в которых μ и β являются параметрами:

$$i = 0, 1, 2, \dots,$$

$$\varphi_0 = (1 + \beta\mu)^{-1/\beta}, \quad \beta_i \geq 0, \quad \mu > 0,$$

$$\varphi_i = \left(\frac{\mu}{1 + \beta\mu}\right)^i \frac{1(1 + \beta) \dots (1 + (i-1)\beta)}{i!} \varphi_0, \quad i \geq 1. \quad (9)$$

$$\sum_{i=0}^{\infty} \varphi_i = \varphi_0 \sum_{i=0}^{\infty} \binom{i}{-\frac{1}{\beta}} \left(\frac{-\beta\mu}{1 + \beta\mu}\right)^i = (1 + \beta\mu)^{-1/\beta} \cdot \left(1 - \frac{\beta\mu}{1 + \beta\mu}\right)^{-1/\beta} \equiv 1.$$

Здесь мы использовали общую формулу бинома и то, что

$$\frac{1(1 + \beta) \dots (1 + (i-1)\beta)}{i!} = (-\beta)^i \binom{i}{-\frac{1}{\beta}}.$$

(На основании этого соотношения распределение Поля называется также *отрицательно-биномиальным распределением*.)

Заметим, что распределение Поля содержит два параметра μ и β в противоположность распределениям Пуассона и Паскаля, содержащим один параметр μ . Поэтому оно более сложно, чем эти два распределения, являющиеся, кстати сказать, частными случаями распределения Поля. Действительно, переходя в соотношениях (9) к пределу при $\beta \rightarrow 0$, мы получим распределение Пуассона

$$\varphi_i \xrightarrow{\beta \rightarrow 0} \begin{cases} e^{-\mu} & \text{при } i = 0 \\ e^{-\mu} \frac{\mu^i}{i!} & \text{при } i \geq 1 \end{cases} \quad (10).$$

При $\beta = 1$ мы сразу получаем распределение Паскаля

$$\varphi_i = \frac{1}{1 + \mu} \left(\frac{\mu}{1 + \mu}\right)^i. \quad (11)$$

§ 4.4. Непрерывные распределения. Строго говоря, все случайные величины дискретны, так как на практике мы в результате измерения можем получать только кратные той наименьшей единицы, которую может показать измерительный инструмент. Однако если эта единица достаточно мала по сравнению с наблюдаемыми изменениями величины x , то мы можем с хоршим приближением абстрагироваться от этого факта и считать x непрерывной величиной.

Мы называем распределение *непрерывным*, если функция $\Phi(t)$ всюду непрерывна и кусочно дифференцируема, с непрерывной производной $\Phi'(t) = \varphi(t)$ (т. е. функция $\varphi(t)$ существует и непрерывна всюду, за исключением, может быть, некоторого множества точек, причем на

любом конечном интервале таких точек может быть только конечное число). Таким образом,

$$\int_{t_1}^{t_2} \varphi(t) dt = \Phi(t_2) - \Phi(t_1) = P(t_1 < x \leq t_2), \quad (1)$$

т. е. вероятность того, что величина x принимает значение из некоторого интервала, равна площади, заключенной между кривой $u = \varphi(t)$ и данным интервалом на оси t . Функция $\varphi(t)$ называется *плотностью вероятности* величины x или ее *дифференциальной функцией распределения*¹⁾. Говорят, что все значения t , для которых $\varphi(t) \neq 0$, образуют спектр случайной величины x . Так как $\varphi(t)$ есть производная неубывающей функции, то $\varphi(t) \geq 0$ для всех значений t . Если функция φ имеет только один максимум, то распределение называется *унимодальным* или *односвершичным*, если таких максимумов два, то распределение называется *бимодальным* и т. д. Полагая в (1) $t_2 = t$ и переходя к пределу при $t_1 \rightarrow -\infty$, получаем в силу (4.2.4) соотношение

$$\Phi(t) = \int_{-\infty}^t \varphi(t) dt, \quad (2)$$

аналогичное (4.3.1). Переходя в равенстве (2) к пределу при $t \rightarrow \infty$, получим соотношение

$$\int_{-\infty}^{\infty} \varphi(t) dt = 1, \quad (3)$$

аналогичное равенству (4.3.2).

Полагая в равенстве (1) $t_1 = t$ и $t_2 = t + \Delta t$, получим

$$P(t, t + \Delta t) = \int_t^{t + \Delta t} \varphi(t) dt,$$

откуда в силу теоремы о среднем значении

$$P(t, t + \Delta t) = \varphi(\xi) \Delta t, \quad (4)$$

где ξ заключено между t и $t + \Delta t$.

В силу (4) мы можем сказать, что $\varphi(t) dt$ дает вероятность того, что величина x принимает значение из „бесконечно малого“ интервала от t до $t + dt$. Подчеркнем, что эта вероятность равна $\varphi(t) dt$,

¹⁾ $\varphi(t)$ называется также иногда „функцией частоты“. Это название, однако, может привести к недоразумениям, так как функция $\varphi(t)$ относится к теоретической модели, а не к наблюдаемым частотам. Несогласованность терминологии приводит к необходимости тщательно следить за тем, употребляется ли термин „функция распределения“ в смысле полной или в смысле дифференциальной функции распределения.

а не самой $\varphi(t)$. Поэтому часто дают плотность вероятности в форме

$$d\Phi = \varphi(t) dt, \quad (5)$$

которая называется *дифференциалом вероятности*. Подчеркнем, далее, что при непрерывном распределении нас интересует вероятность того, что величина x принимает значение, лежащее между t и $t + dt$, а не вероятность того, что величина x принимает точно значение t . Дадим теперь некоторые примеры непрерывных распределений.

■ **Пример 1.** Одним из простейших непрерывных распределений является *прямоугольное распределение*, определяемое плотностью вероятности

$$\frac{d\Phi(t)}{dt} = \varphi(t) = \begin{cases} 0 & \text{при } t < \alpha \\ \frac{1}{\beta - \alpha} & \text{при } \alpha < t < \beta \\ 0 & \text{при } \beta < t, \end{cases} \quad (6)$$

где α и β являются двумя параметрами. Так как функция $\varphi(t)$ постоянна, то случайная величина x , как говорят, *разномерно распределена* на интервале (α, β) . Проверьте, что $\int_{-\infty}^{\infty} \varphi(t) dt = 1$, и найдите

$\Phi(t)$. Начертите графики функций $\varphi(t)$ и $\Phi(t)$. Покажите, что при фиксированном α и $\beta \rightarrow \alpha$ имеем $\Phi(t) \rightarrow \varepsilon(t - \alpha)$, где $\varepsilon(t - \alpha)$ есть несобственное распределение, определенное в примере 1 § 4.3, с параметром $\mu = \alpha^*$.

Пример 2. Пусть случайная величина x есть время существования радиоактивного атома. Так как эта величина всегда является неотрицательным числом, то

$$P(x \leq t) = \Phi(t) = 0 \quad \text{при } -\infty < t < 0.]$$

Если $x \leq t$ (при $t \geq 0$), то это означает, что атом должен распасться до истечения времени t . Так как противоположным событием является существование атома по истечении времени t , то в силу (3.5.4) и (3.7.5)

$$P(x \leq t) = \Phi(t) = 1 - e^{-\lambda t} \quad \text{при } 0 \leq t < \infty,$$

т. е.

$$d\Phi(t) = \varphi(t) dt = \begin{cases} 0 \cdot dt & \text{при } t < 0, \\ e^{-\lambda t} \lambda dt & \text{при } 0 \leq t < \infty. \end{cases} \quad (7)$$

Проверьте, что $\int_{-\infty}^{\infty} \varphi(t) dt = 1$, и начертите графики функций $\varphi(t)$ и $\Phi(t)$.

*) Здесь и в следующих примерах имеется в виду сходимость в том смысле, что $\Phi(t) \rightarrow \varepsilon(t - \alpha)$ в точках $t \neq \alpha$. (Прим. ред.)

Результат (7) можно было бы получить и непосредственно, так как он дает вероятность распада атома за интервал времени от t до $t + dt$. Но эта вероятность равна произведению вероятности существования атома по истечении времени t , которая в силу (3.7.5) равна $e^{-\lambda t}$, и вероятности распада атома за время dt , которая, по предположению, равна λdt (см. пример 4 § 4.3). Отсюда видно, что соотношение (7) определяет также распределение промежутка времени между двумя последовательными вызовами на телефонной станции.

Пример 3. Наиболее важным на практике распределением является *нормальное распределение*, определяемое плотностью вероятности

$$\frac{d\Phi(t)}{dt} = \varphi(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right], \quad (8)$$

где μ и σ являются двумя параметрами. Множитель $\frac{1}{\sqrt{2\pi}\sigma}$ называется *нормирующим множителем*, так как он дает $\int_{-\infty}^{\infty} \varphi(t) dt \equiv 1$. (Это

будет доказано при подробном изложении в разделе 7.) Начертите графики функций $\varphi(t)$ и $\Phi(t)$. Покажите, что при $\sigma \rightarrow 0$ функция $\Phi(t)$ стремится к функции $\varepsilon(t-\mu)$, определенной в примере 1 § 4.3.

Пример 4. *Распределение Коши* задается плотностью вероятности

$$\frac{d\Phi(t)}{dt} = \varphi(t) = \frac{1}{\pi a} \cdot \frac{1}{1 + \frac{(t-\mu)^2}{a^2}} \quad (-\infty < t < \infty), \quad (9)$$

где μ и a являются двумя параметрами. Проверьте, что $\int_{-\infty}^{\infty} \varphi(t) dt = 1$, и начертите графики функций $\varphi(t)$ и $\Phi(t)$. Покажите, что при $a \rightarrow 0$ функция $\Phi(t)$, как и в предыдущем примере, стремится к $\varepsilon(t-\mu)$. Это распределение встречается, например, при изучении интенсивности спектральных линий и в теории рассеивания и захвата атомных ядер.

Пример 5. *Распределение Лапласа* задается плотностью вероятности

$$\frac{d\Phi(t)}{dt} = \varphi(t) = \frac{1}{2a} e^{-\frac{|t-\mu|}{a}} \quad (-\infty < t < \infty), \quad (10)$$

где μ и a являются двумя параметрами. Проверьте, что $\int_{-\infty}^{\infty} \varphi(t) dt = 1$, и начертите графики функций $\varphi(t)$ и $\Phi(t)$. Покажите, что при $a \rightarrow 0$, как и в предыдущем случае, $\Phi(t) \rightarrow \varepsilon(t-\mu)$.

Упражнение 1. Пусть ν независимых случайных величин x_1, x_2, \dots, x_ν имеют одну и ту же функцию распределения $\Phi(t)$. Наибольшее из значений

величин x есть снова случайная величина. Показать, что ее функция распределения $\Phi_{\max}(t)$ определяется дифференциалом вероятности

$$d\Phi_{\max}(t) = \sqrt{\Phi}^{-1}(t) \varphi(t) dt = d\Phi^*(t). \quad (11)$$

* Пример 6. Так как при доказательстве общих теорем неудобно повторять одни и те же рассуждения отдельно для случаев дискретного и непрерывного распределения, то часто бывает полезно рассматривать непрерывные распределения как общий случай, содержащий дискретные распределения в качестве специальных предельных случаев. Как показывают приведенные выше примеры, простейшее дискретное распределение — несобственное — $\Phi(t) = \varepsilon(t - \mu)$ может быть получено различными способами как предел непрерывных распределений. Дирак¹⁾ ввел поэтому фиктивную функцию, дающую „плотность вероятности“ $\delta(t - \mu)$ функции распределения $\varepsilon(t - \mu)$. Эта функция должна, очевидно, равняться 0 при $t \neq \mu$ и ∞ при $t = \mu$, и полный интеграл от нее должен равняться единице:

$$\frac{d\varepsilon(t - \mu)}{dt} = \delta(t - \mu) = \delta(\mu - t) = \begin{cases} 0 & \text{при } t \neq \mu \\ \infty & \text{при } t = \mu \end{cases} \quad (12)$$

$$\int_{-\infty}^{\infty} \delta(t - \mu) dt = \int_{\mu - \varepsilon_1}^{\mu + \varepsilon_2} \delta(t - \mu) dt = 1,$$

где ε_1 и ε_2 — произвольные положительные числа. Конечно, такой „функции“ на самом деле не существует, и ее введение не может быть строго математически обосновано. Однако, во-первых, эта функция дает только краткое выражение для результата предельного процесса, и окончательные результаты, полученные с помощью ее, всегда оказываются правильными; во-вторых, этот метод может быть строго обоснован с помощью более сложного понятия — интеграла Стильтьеса (см. § 4.7). Поэтому δ -функция Дирака, будучи значительно более „наглядной“ и удобной в приложениях, чем точный предельный процесс или строгий, но менее известный метод интегралов Стильтьеса, широко применяется, особенно в квантовой теории, в которой „эффективность“ математических методов значительно важнее их „строгости“. Ясно, однако, что применение подобных методов требует известного такта, для того чтобы оно приводило к правильным окончательным результатам.

* У п р а ж н е н и е 2. Показать, что любая дискретная функция распределения может быть записана в виде

$$\Phi(t) = \sum_{i=-\infty}^{\infty} \varphi_i \varepsilon(t - \mu_i). \quad (13)$$

Показать, что с помощью δ -функции Дирака это распределение может быть представлено в виде непрерывного распределения с плотностью вероятности

$$\frac{d\Phi(t)}{dt} = \varphi(t) = \sum_{i=-\infty}^{\infty} \varphi_i \delta(t - \mu_i). \quad (14)$$

* У п р а ж н е н и е 3. Доказать следующие свойства δ -функции:

$$\int_{-\infty}^{\infty} f(t) \delta(t - \mu) dt = f(\mu) \quad (15)$$

¹⁾ P. A. M. Dirac, Principles of Quantum Mechanics, London, 1930.

для любой непрерывной функции $f(t)$;

$$\int_{-\infty}^{\infty} \delta(t - \mu_1) \delta(t - \mu_2) dt = \delta(\mu_1 - \mu_2) \quad (16)$$

и, с помощью интегрирования по частям,

$$\int_{-\infty}^{\infty} f(t) \delta'(t - \mu) dt = -f'(\mu). \quad (17)$$

§ 4.5. Плотность вероятности $\varphi(t)$, если она существует, является более простым и удобным средством для изучения случайной величины, чем ее функция распределения $\Phi(t)$. Однако следует помнить, что функция распределения является общим понятием, применимым во всех случаях.

Большинство случайных величин, встречающихся на практике, имеет или чисто дискретные или чисто непрерывные функции распределения. Однако на практике встречаются также случайные величины, со смешанными функциями распределения, имеющими и точки разрыва и интервалы непрерывности. В атомной теории энергия атома является случайной величиной, спектр которой содержит как непрерывную, так и дискретную части. Можно привести еще следующий пример. 40-летний человек приобретает 25-летний страховой полис, по которому некоторая сумма денег выплачивается данному лицу по достижении им 65-летнего возраста или его наследникам непосредственно после его смерти. Интервал времени между вступлением полиса в силу и днем выплаты по нему есть, очевидно, случайная величина с функцией распределения, равной нулю при $t < 0$, непрерывно возрастающей от $t=0$ до $t=25$ и скачкообразно увеличивающейся до значения 1 при $t=25$.

Укажем, наконец, что теоретически возможен еще другой тип функции распределения, при котором функция $\Phi(t)$ всюду непрерывна, „почти всюду“ дифференцируема с производной $\Phi'(t) \equiv 0$, но, тем не менее, возрастает от 0 до 1). Такие „патологические“ функции распределения, конечно, никогда не встречаются на практике.

§ 4.6. В качестве удобной и часто употребляемой механической интерпретации мы можем считать, что функция распределения задает некоторое распределение массы на оси t . Таким образом, мы можем представлять себе эту ось в виде бесконечно тонкого стержня, различные части которого несут на себе массы различной плотности. Кроме того, некоторые массы сосредоточены в отдельных изолированных точках. Полагая общую сумму всей распределенной массы равной 1 и обозначая через $\Phi(t)$ массу, лежащую в точке t и влево от нее, мы получаем функцию $\Phi(t)$, обладающую всеми свойствами

1) См., например, T i t c h m a r s h, Theory of Functions, § 11.72; London, 1932.

функции распределения. Точки разрыва функции $\Phi(t)$ являются точками сосредоточения массы. В точках непрерывности этой функции плотность массы равна $\Phi'(t) = \varphi(t)$. Из этой механической аналогии и происходит термин „плотность вероятности“. Часто говорят также о „массе вероятности“, подразумевая величину вероятности.

* § 4.7. Как уже было указано, полезно рассматривать дискретные и непрерывные распределения единообразно. Удобный, но нестрогий метод Дирака, указанный в § 4.4, может быть обоснован с помощью так называемого интеграла Стильтьеса, являющегося простым обобщением обычного интеграла¹⁾.

Пусть $F(t)$ есть неубывающая действительная функция, заданная на сегменте $a \leq t \leq b$, а действительная функция $g(t)$ непрерывна на том же сегменте. Разобьем этот сегмент на n частей точками разбиения

$$a = t_0 < t_1 < \dots < t_{n-1} < t_n = b$$

и выберем в каждом из полученных интервалов по произвольной точке

$$\xi_v, \quad t_v \leq \xi_v \leq t_{v+1}.$$

Точно так же, как и в случае обычного интеграла, [можно показать, что при достаточно мелком разбиении сумма

$$\sum_{v=0}^{n-1} g(\xi_v) (F(t_{v+1}) - F(t_v))$$

отличается от некоторого предела на произвольно малую величину. Этот предел называется интегралом Стильтьеса и обозначается через

$$\int_a^b g(t) dF(t). \quad (1)$$

У п р а ж н е н и е 1. Показать, что если функция $F(t)$ имеет непрерывную производную, то интеграл Стильтьеса приводится к обычному интегралу

$$\int_a^b g(t) dF(t) = \int_a^b g(t) F'(t) dt. \quad (2)$$

В частности, если $F(t) = t$, то (1) приводится к обычному интегралу от функции $g(t)$.

У п р а ж н е н и е 2. Показать, что если ступенчатая функция $F(t)$ имеет скачки величины F_i в точках t_i (см. § 4.3), то интеграл Стильтьеса приводится к сумме

$$\int_a^b g(t) dF(t) = \sum g(t_i) F_i. \quad (3)$$

¹⁾ Более подробное изложение см., например, в книге Г. Крамер, „Математические методы статистики, § 7.5, или D. V. Widder, The Laplace Transform, гл. 1, 1946.

Легко показать, что элементарные свойства обычных интегралов имеют место и для интегралов Стильтьеса. Интеграл

$$\int_{-\infty}^{\infty} g(t) dF(t) \quad (4)$$

определяется обычным образом. Если в (4) положить $g(t) = 1$ и $F(t) = \Phi(t)$, то в силу (4.2.3) и (4.2.4)

$$\int_{-\infty}^{\infty} dF(t) = 1. \quad (5)$$

В этой формуле, очевидно, содержатся соотношения (4.3.2) и (4.4.3).

Однако, хотя интеграл Стильтьеса является идеальным инструментом теории вероятностей, мы не будем пользоваться им в этой книге, так как большинство изучающих прикладную математику еще не достаточно хорошо знакомы с ним.

* § 4.8. Под функцией $y = f(x)$ понимается новая случайная величина y , принимающая значение $f(t)$, когда x принимает значение t . Если при всех t выполнено одно из условий: $\frac{df}{dt} > 0$ или $\frac{df}{dt} < 0$, то

$$\Phi_x(t) = P(x \leq t) = \begin{cases} P(y = f(x) \leq f(t) = u) = \Phi_y(u) & \text{при } \frac{df}{dt} > 0 \\ P(y = f(x) \geq f(t) = u) = 1 - \Phi_y(u) & \text{при } \frac{df}{dt} < 0. \end{cases}$$

Дифференцирование обеих частей равенства дает в этих случаях

$$d\Phi_x = \varphi_x(t) dt = \varphi_x(t(u)) \left| \frac{dt}{du} \right| du = \varphi_y(u) du = d\Phi_y. \quad (1)$$

Эта формула играет очень важную роль во многих практических приложениях теории вероятностей.

Пример. Пусть случайная величина v есть абсолютное значение скорости молекулы массы m газа при абсолютной температуре T . Распределение случайной величины v задается *законом Максвелла — Больцмана*

$$d\Phi_v = \varphi_v(t) dt = at^2 e^{-\beta t^2} dt, \quad \beta = \frac{m}{2kT} \quad (0 \leq t < \infty), \quad (2)$$

где k есть физическая константа, называемая *константой Больцмана*, и a есть нормирующий множитель (см. упражнение 3 § 7.4).

В силу (1) распределение кинетической энергии $E = \frac{1}{2} mv^2 = \gamma v^2$ определяется дифференциалом вероятности

$$d\Phi_E = \varphi_E(u) du = \frac{au}{\gamma} e^{-\frac{\beta}{\gamma} u} \frac{1}{2\sqrt{\gamma u}} du = a' u^{1/2} e^{-\beta' u} du \quad (0 \leq u < \infty), \quad (3)$$

где $a' = \frac{a}{2\sqrt{\gamma^3}}$ и $\beta' = \frac{\beta}{\gamma} = \frac{1}{kT}$ являются двумя новыми константами.

У п р а ж н е н и е. Показать с помощью приложения 1, что

$$\alpha = \frac{4}{\sqrt{\pi}} \beta^{\alpha_2} = \sqrt{\frac{2}{\pi}} \left(\frac{m}{kT} \right)^{\alpha_2},$$

и, таким образом,

$$\alpha' = \frac{2}{\sqrt{\pi}} \left(\frac{1}{kT} \right)^{\alpha_2}.$$

§ 4.9. Распространим теперь наши формулы на двумерные и многомерные случайные величины, т. е. на случайные величины, которые определяются заданием двух или нескольких чисел. Так как все понятия и формулы могут быть непосредственно обобщены со случая двух измерений на случай ν измерений, то мы будем рассматривать только случай двух измерений, обозначая случайную величину через $z = (x, y)$.

П р и м е р. Если при бросании монеты обозначать через 1 выпадение решетки и через 0 выпадение герба, то при бросании двух монет случайная величина z может принимать значения (1,1), (0,1), (1,0) и (0,0).

Если, как это имеет место в последнем примере, нас интересует распределение двух одномерных величин, то естественно интерпретировать их как компоненты одной двумерной величины. Можно интерпретировать x и y как декартовы координаты на плоскости и интересоваться вероятностью того, что точка (x, y) лежит в заданной области этой плоскости. Механическая интерпретация § 4.6 может быть непосредственно обобщена на этот случай, причем масса, равная 1, либо непрерывно распределяется по плоскости, либо сосредоточивается в некоторых точках или на некоторых линиях.

Под *совместной функцией распределения* $\Phi(t, u)$, или, просто, функцией распределения двумерной случайной величины $z = (x, y)$, понимается функция, которая для всех значений t и u равна вероятности того, что

$$x \leq t \quad \text{и} \quad y \leq u:$$

$$\Phi(t, u) = P(-\infty < x \leq t, -\infty < y \leq u). \quad (1)$$

Как и в одномерном случае, можно с помощью аксиом I—VI показать, что имеют место следующие естественные обобщения соотношений (4. 2. 3) и (4. 2. 4)¹⁾:

$$\lim_{\substack{t \rightarrow \infty \\ u \rightarrow \infty}} \Phi(t, u) = 1, \quad (2)$$

$$\lim_{t \rightarrow -\infty} \Phi(t, u_0) = 0, \quad (3)$$

$$\lim_{u \rightarrow -\infty} \Phi(t_0, u) = 0 \quad (4)$$

при любых фиксированных t_0 и u_0 .

¹⁾ При этом предполагается, что случайные величины x и y могут принимать только конечные значения (см. примечание на стр. 37).

Упражнение 1. Показать, что при $a > 0$ и $b > 0$ имеем $\Phi(t, u+b) \geq \Phi(t, u)$; $\Phi(t+a, u) \geq \Phi(t, u)$; $\Phi(t+a, u+b) \geq \Phi(t, u+b)$; $\Phi(t+a, u+b) \geq \Phi(t+a, u)$; $\Phi(t+a, u+b) \geq \Phi(t, u)$; $\Phi(t+a, u+b) - \Phi(t+a, u) \geq \Phi(t, u+b) - \Phi(t, u)$.
Показать затем, что $P(a_1 < x \leq a_2, b_1 < y \leq b_2) = \Phi(a_2, b_2) + \Phi(a_1, b_1) - \Phi(a_1, b_2) - \Phi(a_2, b_1)$.

Упражнение 2. Доказать, что

$$P(x+y > a+b) \leq P(x > a) + P(y > b)$$

и

$$P(|x+y| > a+b) \leq P(|x| > a) + P(|y| > b).$$

Как и в случае одномерных величин, мы отдельно рассматриваем два типа распределений — дискретные и непрерывные.

§ 4.10. Пусть

$$\begin{aligned} \dots < t_{-2} < t_{-1} < t_0 < t_1 < t_2 < \dots \\ \dots < u_{-2} < u_{-1} < u_0 < u_1 < u_2 < \dots \end{aligned}$$

— две последовательности чисел. Будем говорить, что двумерная случайная величина z имеет дискретное распределение, если все возможные значения величины z имеют вид (t_i, u_j) . Обозначим соответствующие вероятности через φ_{ij} . В каждом интервале $(t_p \leq t < t_{p+1}, u_q \leq u < u_{q+1})$ функция распределения случайной величины z постоянна. Эта функция распределения равна

$$\Phi(t_p, u_q) = \sum_{i=-\infty}^p \sum_{j=-\infty}^q \varphi_{ij}, \quad (1)$$

откуда в силу (4.9.2)

$$\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \varphi_{ij} = 1. \quad (2)$$

Пример. Предположим, что случайная величина z принимает только значения $(0,0)$, $(2,1)$, $(1,2)$, $(3,3)$ соответственно с вероятностями $1/4$, $1/8$, $3/8$, $1/4$. Графическое изображение этого распределения и значения соответствующей функции распределения показаны на фиг. 3 и 4.

Упражнение. Найти функцию распределения случайной величины определенной в примере § 4.9.

§ 4.11. Вероятность того, что величина x принимает значение t_i , независимо от возможных значений y , обозначается через φ_i и равна

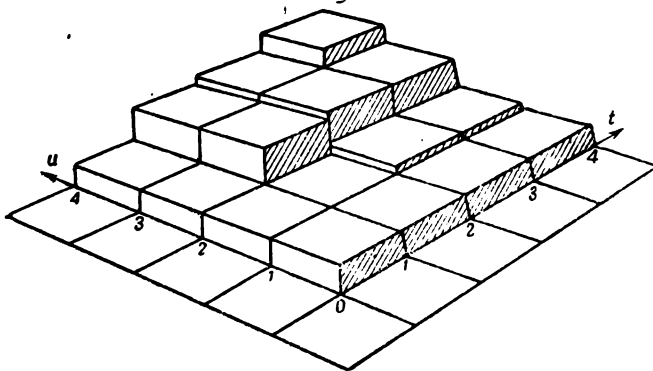
$$\varphi_i = \sum_{j=-\infty}^{\infty} \varphi_{ij}. \quad (1)$$

Аналогично, вероятность того, что величина y принимает значение u_j , независимо от возможных значений x , обозначается через φ_j

равна

$$\varphi_{\cdot j} = \sum_{i=-\infty}^{\infty} \varphi_{ij}. \quad (2)$$

Распределения φ_i и $\varphi_{\cdot j}$ называются соответственно *частными распределениями* величин x и y .



Фиг. 3.

u					
3	1/4	5/8	3/4	1	
0	1/4	5/8	3/4	3/4	
2	1/4	1/4	5/8	3/8	
1	0	1/4	1/4	1/4	
0	0	0	0	0	
	t	0	1	2	3

Фиг. 4.

Условная вероятность того, что величина x принимает значение t_i в предположении, что величина y принимает значение u_j , обозначается через $\varphi_{i|j}$. Аналогично, через $\varphi_{j|i}$ обозначается условная вероятность того, что величина y принимает значение u_j в предположении, что величина x принимает значение t_i . В силу закона умножения (аксиома V) имеем

$$\varphi_{ij} = \varphi_i \cdot \varphi_{j|i} = \varphi_{\cdot j} \varphi_{i|\cdot j}. \quad (3)$$

Упражнение 1. Показать, что

$$\sum_{i=-\infty}^{\infty} \varphi_{i|j} = \sum_{j=-\infty}^{\infty} \varphi_{j|i} = 1. \quad (4)$$

Если $\varphi_{i|j} = \varphi_i$ при всех значениях i и j , откуда в силу (3) следует, что $\varphi_{j|i} = \varphi_j$, то мы говорим, что случайные величины x и y *стохастически независимы*, или просто *независимы* (см. § 3.2). В этом случае соотношение (3) приводится к

$$\varphi_{ij} = \varphi_i \cdot \varphi_j. \quad (5)$$

Упражнение 2. Показать, что, обратно, если выполнено соотношение (5), то величины x и y независимы.

Упражнение 3. Показать, что если величины x и y независимы, то случайная величина $z = x + y$ имеет распределение, задаваемое соотношением

$$P(z = x + y = v_k) = \sum_{t_i + u_j = v_k} \varphi_i \cdot \varphi_j. \quad (6)$$

Упражнение 4. Показать с помощью (6), что если величины x и y биномиально распределены соответственно с параметрами ν_1, θ и ν_2, θ (см. пример 2 § 4.3) и независимы, то случайная величина $z = x + y$ также биномиально распределена с параметрами $\nu = \nu_1 + \nu_2, \theta$.

Упражнение 5. Показать с помощью (6), что если величины x и y имеют распределения Пуассона соответственно с параметрами μ_1 и μ_2 и независимы, то случайная величина $z = x + y$ также имеет распределение Пуассона с параметром $\mu = \mu_1 + \mu_2$.

§ 4.12. Будем говорить, что двумерная случайная величина z имеет непрерывную функцию распределения, если существует кусочно-непрерывная функция $\varphi(t, u) \geq 0$ (т. е. функция φ существует и непрерывна всюду, кроме, может быть, некоторых точек, лежащих на конечном числе кривых) такая, что вероятность того, что величина z принимает значение из области ω плоскости t, u , равна*)

$$P(z \in \omega) = \iint_{\omega} \varphi^2(t, u) dt du. \quad (1)$$

Функция $\varphi(t, u)$ называется *плотностью вероятности* величин x и y . В анализе доказывается, что

$$\varphi(t, u) = \frac{\partial^2 \Phi(t, u)}{\partial t \partial u} \quad (2)$$

во всех точках, в которых функция $\varphi(t, u)$ существует и непрерывна.

Таким образом, вероятность того, что величина z принимает значение из некоторой области, равна объему, заключенному между поверхностью $v = \varphi(t, u)$ и заданной областью на плоскости t, u (ср. одномерный случай). Говорят, что $\varphi(t, u) dt du$ задает вероятность того, что случайная величина z принимает значение из „бесконечно

*) $z \in \omega$ означает „ z входит в ω “. (Прим. ред.)

малого" интервала $t \leq x \leq t + dt$, $u = y \leq u + du$. Подчеркнем, что эта вероятность дается именно $\varphi(t, u) dt du$, а не самой $\varphi(t, u)$. По аналогии с (4.4.5) распределение часто задается в виде *дифференциала вероятности*

$$d\Phi = \varphi(t, u) dt du. \quad (3)$$

Из соотношения (1) или (2) вытекает, что

$$\Phi(t, u) = \int_{-\infty}^t \int_{-\infty}^u \varphi(t, u) dt du \quad (4)$$

и в силу (4.9.2)

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(t, u) dt du = 1. \quad (5)$$

Пример 1. При стрельбе из орудий обычно считают, что координаты x и y точки падения снаряда распределены нормально с некоторыми параметрами μ_x , σ_x и μ_y , σ_y (см. пример 3 § 4.4). Если x есть дальность, а y — боковое отклонение от некоторого заданного направления, то считают x и y независимыми; поэтому в силу закона умножения распределение случайной величины $z = (x, y)$ задается дифференциалом вероятности

$$d\Phi = \varphi(t, u) = \frac{1}{2\pi \sigma_x \sigma_y} \exp \left[- \left\{ \frac{(t - \mu_x)^2}{2\sigma_x^2} + \frac{(u - \mu_y)^2}{2\sigma_y^2} \right\} \right] dt du$$

$$(-\infty < t < \infty, -\infty < u < \infty). \quad (6)$$

Общее двумерное нормальное распределение может быть получено из (6) введением дополнительного члена в выражение, стоящее под знаком экспоненциала, и соответствующим изменением нормирующего множителя:

$$d\Phi = \varphi(t, u) dt du = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} \times$$

$$\times \exp \left[- \frac{1}{1 - \rho^2} \left\{ \frac{(t - \mu_x)^2}{2\sigma_x^2} - \rho \frac{(t - \mu_x)(u - \mu_y)}{\sigma_x \sigma_y} + \frac{(u - \mu_y)^2}{2\sigma_y^2} \right\} \right] dt du$$

$$(-\infty < t < \infty, -\infty < u < \infty). \quad (7)$$

Здесь ρ — новый параметр, удовлетворяющий неравенствам $-1 \leq \rho \leq 1$. При $\rho = \pm 1$ величины x и y подчинены линейной зависимости, так как в этом случае полная масса вероятности сосредоточена на прямой $t - \mu_x = \pm \frac{\sigma_x}{\sigma_y} (u - \mu_y)$. (Позднее в разделе 7 мы исследуем это распределение с большей подробностью.)

Пример 2. Другое непрерывное распределение задается дифференциалом вероятности

$$d\Phi = \varphi(t, u) dt du = \frac{1}{\pi} \frac{1}{(1+t^2+u^2)^2} dt du. \quad (8)$$

Проверьте, что в этом случае выполнено соотношение (5).

§ 4.13. Вероятность того, что величина x принимает значение между $-\infty$ и t независимо от возможных значений величины y , равна

$$\Phi_x(t) = \int_{-\infty}^t dt \int_{-\infty}^{\infty} \varphi(t, u) du. \quad (1)$$

Функция $\Phi_x(t)$, или, короче, $\Phi(t)$, называется *частным распределением* величины x . В анализе доказывается, что в предположении, что функция

$$\varphi_x(t) = \int_{-\infty}^{\infty} \varphi(t, u) du \quad (2)$$

имеет не более конечного числа точек разрыва, функция $\Phi_x(t)$ соответствует непрерывному распределению с плотностью вероятности $\varphi_x(t)$, определяемой соотношением (2). Аналогично, вероятность того, что величина y принимает значение между $-\infty$ и u независимо от возможных значений величины x , равна

$$\Phi_y(u) = \int_{-\infty}^u du \int_{-\infty}^{\infty} \varphi(t, u) dt. \quad (3)$$

Функция $\Phi_y(u)$ называется *частным распределением* величины y и соответствует непрерывному распределению с плотностью вероятности

$$\varphi_y(u) = \int_{-\infty}^{\infty} \varphi(t, u) dt \quad (4)$$

в предположении, что функция $\varphi_y(u)$ имеет не более конечного числа точек разрыва. Подчеркнем, что, хотя частные распределения однозначно определяются двумерным распределением, обратное места не имеет. Для любых заданных функций $\varphi_x(t) \geq 0$ и $\varphi_y(u) \geq 0$, удовлетворяющих только соотношению (4.4.3), всегда можно построить двумерное распределение с плотностью вероятности $\varphi(t, u) = \varphi_x(t) \varphi_y(u)$. Однако, как показывает следующий пример, может существовать бесконечно много двумерных распределений, имеющих одни и те же частные распределения.

Пример. Для нормального распределения (4.12.7) частные распределения, независимо от ρ , также нормальны и имеют соответственно

параметры μ_x , σ_x и μ_y , σ_y . Для того чтобы доказать это, удобно положить

$$x = \frac{t - \mu_x}{\sigma_x}, \quad y = \frac{u - \mu_y}{\sigma_y}. \quad (5)$$

Подставляя (5) в (4.12.7), получим в силу (2)

$$\begin{aligned} d\Phi_x = \varphi_x(t) dt &= \frac{dx}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right] dy = \\ &= \frac{dx}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{x^2}{2}\right] \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2(1-\rho^2)}(y - \rho x)^2\right] dy = \\ &= \frac{dx}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right] \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{v^2}{2}\right] dv, \end{aligned}$$

если ввести $v = \frac{y - \rho x}{\sqrt{1-\rho^2}}$ в качестве новой переменной. Так как по-

следний интеграл равен 1 (это доказано в разделе 7), то, подставляя с помощью (5) t вместо x , получим

$$d\Phi_x = \varphi_x(t) dt = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{(t - \mu_x)^2}{2\sigma_x^2}\right] dt, \quad (6)$$

что и требовалось доказать. Аналогично доказывается, что

$$d\Phi_y = \varphi_y(u) du = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{(u - \mu_y)^2}{2\sigma_y^2}\right] du. \quad (7)$$

Соотношения (6) и (7) показывают, что нормирующий множитель в соотношении (4.12.7) выбран правильно, т. е. соотношение (4.12.5) выполнено.

Определим функцию $\varphi_{x|y}(t|u)$, или, короче, $\varphi(t|u)$, положив

$$\varphi_{x|y}(t|u) = \frac{\varphi(t, u)}{\varphi_y(u)}, \quad (8)$$

и функцию $\varphi_{y|x}(u|t)$, или, короче, $\varphi(u|t)$, положив

$$\varphi_{y|x}(u|t) = \frac{\varphi(t, u)}{\varphi_x(t)}. \quad (9)$$

Тогда из тождества

$$\varphi(t, u) dt du = \varphi_x(t) dt \varphi_{y|x}(u|t) du = \varphi_y(u) du \varphi_{x|y}(t|u) dt \quad (10)$$

и закона умножения (аксиома V) следует, что $d\Phi_{x|y} = \varphi_{x|y}(t|u) dt$ есть условная вероятность того, что величина x принимает значение между t и $t + dt$ в предположении, что y принимает значение между u

и $u + du$. Соответствующую интерпретацию имеет и дифференциал вероятности $d\Phi_{y|x} = \varphi_{y|x}(u|t) du$. Оба соответствующих распределения называются *условными распределениями*.

У п р а ж н е н и е 1. Показать, что

$$\int_{-\infty}^{\infty} \varphi_{x|y}(t|u) dt = \int_{-\infty}^{\infty} \varphi_{y|x}(u|t) du = 1. \quad (11)$$

Если $\varphi_{x|y}(t|u) = \varphi_x(t)$ для всех значений t и u , откуда в силу (10) вытекает, что $\varphi_{y|x}(u|t) = \varphi_y(u)$, то говорят, что случайные величины x и y *стохастически независимы*, или просто *независимы*. В этом случае

$$\varphi(t, u) = \varphi_x(t) \varphi_y(u). \quad (12)$$

У п р а ж н е н и е 2. Показать, что если имеет место соотношение (12), то величины x и y независимы.

У п р а ж н е н и е 3. Используя результаты, полученные в примере этого параграфа, показать, что при нормальном распределении условные распределения также нормальны и имеют соответственно параметры $\mu = \mu_x + \rho \frac{\sigma_x}{\sigma_y} (u - \mu_y)$, $\sigma = \sqrt{1 - \rho^2} \sigma_x$ и $\mu = \mu_y + \rho \frac{\sigma_y}{\sigma_x} (t - \mu_x)$, $\sigma = \sqrt{1 - \rho^2} \sigma_y$. Следовательно,

$$\begin{aligned} d\Phi_{x|y} &= \varphi_{x|y}(t|u) dt = \\ &= \frac{1}{\sqrt{2\pi} \sqrt{1 - \rho^2} \sigma_x} \exp \left[-\frac{1}{2(1 - \rho^2) \sigma_x^2} \left(t - \mu_x - \rho \frac{\sigma_x}{\sigma_y} (u - \mu_y) \right)^2 \right] dt \end{aligned} \quad (13)$$

и

$$\begin{aligned} d\Phi_{y|x} &= \varphi_{y|x}(u|t) du = \\ &= \frac{1}{\sqrt{2\pi} \sqrt{1 - \rho^2} \sigma_y} \exp \left[-\frac{1}{2(1 - \rho^2) \sigma_y^2} \left(u - \mu_y - \rho \frac{\sigma_y}{\sigma_x} (t - \mu_x) \right)^2 \right] du. \end{aligned} \quad (14)$$

* § 4.14. Под функцией $\mathbf{w} = f(\mathbf{z}) = f(x, y)$ понимается новая одномерная случайная величина \mathbf{w} , принимающая значение $f(t, u)$, когда x принимает значение t , а y — значение u . Функция распределения $\Phi_{\mathbf{w}}(s)$ случайной величины \mathbf{w} , равная вероятности $P(\mathbf{w} \leq s)$ того, что \mathbf{w} принимает значение, меньшее или равное s , определяется в силу (4.12.1) интегралом функции $\varphi(t, u)$ по той области в плоскости t, u , для которой $\mathbf{w} \leq s$:

$$\Phi_{\mathbf{w}}(s) = P(\mathbf{w} \leq s) = \iint_{\mathbf{w} \leq s} \varphi(t, u) dt du. \quad (1)$$

Часто (например, при вычислении интеграла (1)) бывает удобно перейти к новым переменным t', u' , связанным со старыми переменными t, u соотношениями $t = t(t', u')$, $u = u(t', u')$. Можно показать¹⁾,

¹⁾ См. любой учебник анализа, например, Р. Курант, Курс дифференциального и интегрального исчисления, ч. II, ГТТИ, 1931.

что если так называемый *якобиан*

$$\frac{\partial(t, u)}{\partial(t', u')} = \begin{vmatrix} \frac{\partial t}{\partial t'} & \frac{\partial t}{\partial u'} \\ \frac{\partial u}{\partial t'} & \frac{\partial u}{\partial u'} \end{vmatrix} \neq 0 \quad (2)$$

не равен нулю во всей плоскости t, u , то

$$\begin{aligned} d\Phi &= \varphi(t, u) dt du = \varphi(t(t', u'), u(t', u')) \cdot \left| \frac{\partial(t, u)}{\partial(t', u')} \right| dt' du' = \\ &= \varphi'(t', u') dt' du' = d\Phi', \end{aligned} \quad (3)$$

что является непосредственным обобщением (4.8.1).

Вводя в соотношение (1) новые переменные $s = f(t, u)$ и $r = g(t, u)$ (r может быть произвольной переменной, скажем $r = t$, для которой выполнено соотношение (2)), получим

$$\Phi_w(s) = \int_{-\infty}^s ds \int_{-\infty}^{\infty} \varphi(t(r, s), u(r, s)) \left| \frac{\partial(t, u)}{\partial(r, s)} \right| dr. \quad (4)$$

Дифференцируя (4) по s , получим плотность вероятности случайной величины w

$$\varphi_w(s) = \int_{-\infty}^{\infty} \varphi(t(r, s), u(r, s)) \left| \frac{\partial(t, u)}{\partial(r, s)} \right| dr. \quad (5)$$

Пример. Пусть случайные величины x и y независимы и имеют соответственно плотности вероятности $\varphi_x(t)$ и $\varphi_y(u)$. Для случайной величины $w = x + y$ мы получим, полагая $s = t + u$ и, скажем, $r = t$, в силу (5)

$$\varphi_w(s) = \int_{t+u=s} \varphi_x(t) \varphi_y(u) \left| \frac{\partial(t, u)}{\partial(r, s)} \right| dr = \int_{-\infty}^{\infty} \varphi_x(r) \varphi_y(s-r) dr, \quad (6)$$

так как

$$\frac{\partial(t, u)}{\partial(r, s)} = 1.$$

Читателю предоставляется обобщить формулы § 4.9 — 4.14 на случай величин более чем двух измерений.

***§ 4.15. Стохастические процессы.** Часто возникают проблемы, в которых распределение случайной величины зависит от неслучайной величины, являющейся непрерывно изменяющимся параметром, как, например, время. В таких случаях говорят о *стохастическом* или *случайном процессе*¹⁾. В примере 4

1) Общую теорию стохастических процессов читатель найдет, например, в книге Хинчина, Асимптотические законы теории вероятностей, и в работе Феллера, К теории стохастических процессов, Успехи матем. наук, вып. V, 1938.

§ 4.3 был рассмотрен частный случай *дискретного случайного процесса*, т. е. процесса, в котором случайная величина имеет дискретное распределение. В теории броуновского движения, в которой положение (x, y, z) броуновской частицы рассматривается как трехмерная случайная величина, мы встречаемся с примером *непрерывного случайного процесса*, т. е. процесса, в котором случайная величина имеет непрерывное распределение. В этом примере полная „масса вероятности“ $P(V, t) = \iiint_V \varphi(x, y, z, t) dx dy dz$, приходящаяся

на данный объем V , является функцией времени. (В этом и в двух последующих параграфах мы употребляем одну и ту же букву как для обозначения случайной величины, так и для обозначения соответствующего аргумента в функции распределения этой случайной величины.) Поскольку вероятность следует считать субстанцией, которая не может ни возникнуть, ни уничтожиться, всякое изменение величины $P(V, t)$ должно соответствовать потоку „массы вероятности“ через поверхность F , ограничивающую объем V , „потоку вероятности“ в механической интерпретации § 4.6.

Если вектор \vec{s} обозначает *плотность потока вероятности*, т. е. массу вероятности, протекающую в единицу времени через единичную площадку, перпендикулярную направлению вектора \vec{s} , то согласно закону сохранения массы вероятности

$$-\frac{dP}{dt} = -\frac{d}{dt} \iiint_V \varphi dv = \iint_F s_n df, \quad (1)$$

где dv есть элемент объема, df — элемент поверхности с нормалью \vec{n} , положительное направление которой — изнутри наружу, и, наконец, s_n — компонента вектора \vec{s} по направлению \vec{n} . С помощью теоремы Гаусса из выражения (1) получается так называемое *уравнение неразрывности*

$$\operatorname{div} \vec{s} + \frac{\partial \varphi}{\partial t} = \frac{\partial s_x}{\partial x} + \frac{\partial s_y}{\partial y} + \frac{\partial s_z}{\partial z} + \frac{\partial \varphi}{\partial t} = 0, \quad (2)$$

аналогичное соответствующим уравнениям в теории электричества, теории теплопроводности и в квантовой механике.

Часто можно получить хорошее приближение к действительности, считая, что \vec{s} пропорционально $\operatorname{grad} \varphi$, как это обычно делается, например, в теории теплопроводности. Тогда

$$\vec{s} = -D \operatorname{grad} \varphi = -D \left(\frac{\partial \varphi}{\partial x}, \frac{\partial \varphi}{\partial y}, \frac{\partial \varphi}{\partial z} \right), \quad (3)$$

где $D (> 0)$ есть постоянная, называемая *коэффициентом диффузии*. Тогда из (2) получаем

$$D \operatorname{div} \operatorname{grad} \varphi = D \Delta \varphi = \frac{\partial \varphi}{\partial t}, \quad (4)$$

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Упражнение. Покажите, что нормальное распределение, заданное формулой

$$d\Phi = \varphi(x, y, z, t) dv = \frac{1}{(2\sqrt{\pi Dt})^3} \exp \left[-\frac{(x - \mu_x)^2 + (y - \mu_y)^2 + (z - \mu_z)^2}{4Dt} \right] dx dy dz, \quad (5)$$

и являющееся трехмерным обобщением распределения (4.12.6) при $\sigma_x = \sigma_y = \sigma_z = \sqrt{2Dt}$, является решением уравнения (4). Очевидно, что при $t \rightarrow 0$ имеем $\Phi \rightarrow \varepsilon(x - \mu_x)\varepsilon(y - \mu_y)\varepsilon(z - \mu_z)$ (см. пример 1 § 4.3), так что в момент $t = 0$ броуновская частица находится в точке (μ_x, μ_y, μ_z) . Используя δ -функцию Дирака (см. конец § 4.4), мы сможем выразить это обстоятельство следующим образом:

$$\varphi(x, y, z; t = 0) dv = \delta(x - \mu_x) \delta(y - \mu_y) \delta(z - \mu_z) dv.$$

Если в выражении (3), кроме „потока диффузии“, нужно учесть также и „поток конвекции“, т. е. перенос субстанции частицами с некоторой скоростью \vec{v} , то в правой части выражения (3) следует добавить слагаемое $\vec{v}\varphi$. Если, кроме того, коэффициент диффузии D считать функцией координат (x, y, z) , то из (2) мы получим вместо уравнения (4) более общее уравнение

$$\Delta(D\varphi) - \operatorname{div}(\vec{v}\varphi) = \frac{\partial \varphi}{\partial t}. \quad (6)$$

Это так называемое *уравнение Фоккера — Планка* есть общее уравнение трехмерного непрерывного стохастического процесса. Оно позволяет однозначно определить величину φ для всех моментов времени, если известно значение φ для какого-нибудь момента времени $t = t_0$. (Это уравнение может быть обобщено на случай любого количества измерений.)

Стохастические процессы, как непрерывные, так и дискретные, играют все возрастающую роль во многих прикладных областях, как, например, в физике, инженерном деле (особенно в телефони), биологии и в страховом деле¹⁾.

***§ 4.15. Статистическая механика.** Специальные случаи непрерывных процессов и многомерных распределений встречаются в *статистической механике*, области механики, в которой движение тел изучается при том ограничении, что начальные условия известны не настолько хорошо, чтобы можно было применить причинное описание (в классической механике это в принципе всегда возможно²⁾).

Точная характеристика состояния механической системы с f степенями свободы заключается в указании „точки“ $\vec{r} = (q_1, \dots, q_f, p_1, \dots, p_f)$ в $2f$ -мерном фазовом пространстве, т. е. в указании численных значений f обобщенных координат q_1, \dots, q_f и соответствующих f обобщенных импульсов p_1, \dots, p_f (являющихся обобщением декартовых координат и импульсов соответственно). Поскольку в достаточно сложных случаях, как, например, при описании состояния грамма водорода (см. § 1.2), или практически невозможно, или затруд-

¹⁾ Из многочисленных приложений укажем лишь на следующие: Arley, *Stochastic Processes and Cosmic Radiation* (теория дискретных процессов с приложением к теории космических лучей); Чандрасекар, *Stochastic problems in physics and astronomy*; Lundberg, *Random Processes and Sickness and Accident Statistics* (теория дискретных процессов с приложением к страховому делу); Bartlett, *Stochastic Processes* (общая теория с иллюстрациями из различных областей).

²⁾ См., например, Гиббс, *Основные принципы статистической механики*, Гостехиздат, 1946 (общая теория); Tolman, *The Principles of Statistical Mechanics*, Oxford, 1938 (общая теория); Lindsay, *Introduction to Physical Statistics*, Wiley, 1941 (со многими приложениями). Изложение предмета с точки зрения современной теории вероятностей см. А. Я. Хинчин, *Математические основания статистической механики*, Гостехиздат, 1943, и в работе Moyal, *Stochastic Processes and Statistical Physics*, J. Roy. Stat. Soc., B, т. 11, 1949.

нительно измерить и фиксировать все эти $2f$ величин, мы применяем статистическое описание, т. е. рассматриваем вектор \vec{r} как $2f$ -мерную случайную величину с непрерывной функцией распределения, зависящей от времени. Таким образом,

$$d\Phi = \varphi(\vec{r}, t) dv = \varphi(q_1, \dots, q_f; p_1, \dots, p_f; t) dq_1 \dots dq_f dp_1 \dots dp_f, \quad (1)$$

где dv — элемент объема фазового пространства. Можно показать, что φ удовлетворяет уравнению Фоккера — Планка (4.15.6), в котором $\vec{v} = \left(\frac{dq_1}{dt}, \dots, \frac{dp_f}{dt} \right)$ есть „скорость“ фазовой точки \vec{r} :

$$\text{div}(\varphi \vec{v}) + \frac{\partial \varphi}{\partial t} = \sum_{k=1}^f \left[\frac{\partial}{\partial q_k} \left(\varphi \frac{dq_k}{dt} \right) + \frac{\partial}{\partial p_k} \left(\varphi \frac{dp_k}{dt} \right) \right] + \frac{\partial \varphi}{\partial t} = 0. \quad (2)$$

Это уравнение позволяет однозначно определить φ для всех моментов времени, если φ известна для какого-нибудь момента времени $t = t_0$.

Заметим, что разница между общим стохастическим процессом, описываемым уравнением (4.15.6), и частным случаем, описываемым уравнением (2), заключается в том, что элементы случайности возникают здесь двумя различными путями. В первом случае переход от начального состояния $\vec{r}_1 = (x_1, y_1, z_1)$ системы в момент времени t_1 к конечному состоянию $\vec{r}_2 = (x_2, y_2, z_2)$ системы в момент времени $t_2 (> t_1)$ происходит случайным образом; во втором же случае случайным является само начальное состояние, переход же к конечному состоянию может быть причинно (однозначно) определен по законам ньютоновской механики, если начальное состояние $\vec{r}' = (q_1', \dots, p_f')$ известно точно, т. е. если начальное распределение несобственное: $\Phi = \varepsilon(q_1 - q_1') \dots \varepsilon(p_f - p_f')$.

Если температура системы поддерживается постоянной, то, как показывает опыт, при $t \rightarrow \infty$ распределение φ стремится к некоторому *стационарному распределению*, не зависящему от начального состояния; в этом случае говорят, что система переходит в состояние *термического* или *статистического равновесия*. Подобные состояния равновесия являются главным объектом, изучаемым в приложениях статистической механики. Однако в последнее время подверглась изучению также проблема о том, каким образом достигаются равновесные состояния (так называемые *явления переноса*). Опыт показывает, что стационарное распределение Φ_{stat} зависит лишь от полной энергии системы $E = E(q_1, \dots, p_f)$ и что Φ_{stat} удовлетворительно описывается *каноническим распределением* Гиббса

$$d\Phi_{\text{stat}} = \varphi_{\text{stat}}(\vec{r}) dv = c \exp \left[-\frac{E(q_1, \dots, p_f)}{kT} \right] dq_1 \dots dq_f dp_1 \dots dp_f. \quad (3)$$

Здесь c есть нормирующий множитель, часто записываемый в виде $c = \exp \left[\frac{\psi}{kT} \right]$, где T — абсолютная температура, а k — физическая постоянная, называемая *постоянной Больцмана*.

Таким образом, $\theta = kT (> 0)$ есть произвольный параметр; он называется *модулем* распределения. (В частном случае, если рассматривать лишь такие изменения q_1, \dots, p_f в фазовом пространстве, при которых $E = \text{const}$, распределение, определяемое формулой (3), называется *микрочаноническим распределением*.)

У п р а ж н е н и е. Пусть система состоит из N частиц и является консервативной, так что в декартовых координатах $E = T + V = \sum_{i=1}^N \frac{1}{2m_i} (p_{x_i}^2 + p_{y_i}^2 + p_{z_i}^2) + V(x_1, \dots, z_N)$, где T — кинетическая энергия, V — потенциальная энергия, $p_{x_i} = m_i v_{x_i}$ и т. д. Покажите, что компоненты скоростей не зависят от координат и взаимно независимы и что частное распределение каждой из компонент скорости есть нормальное распределение с параметрами $\mu = 0$, $\sigma = \sqrt{kT/m_i}$. (о распределении для $v_i = \sqrt{v_{x_i}^2 + v_{y_i}^2 + v_{z_i}^2}$ см. упражнение 3 § 7.4).

П р и м е р 1. Напомним читателю, как следует сравнивать с опытом формулу (3). Для этой цели следует рассмотреть большое число n идентичных копий рассматриваемой механической системы, находящихся в термическом равновесии при одной и той же абсолютной температуре T . Предположим, что в один и тот же момент времени мы определили состояние каждой из этих n систем, т. е. их фазовые точки, и получили в результате

$$\vec{r}_1 = (q_{11}, \dots, p_{f1}), \dots, \vec{r}_n = (q_{1n}, \dots, p_{fn}).$$

Пусть V — произвольная область фазового пространства и n_V — количество измеренных фазовых точек, которые лежат в области V . Тогда относительная частота n_V/n будет экспериментальным значением вероятности $p = \int \dots \int_V \varphi d\nu$

того, что \vec{r} принадлежит области V ¹⁾.

Эмпирическая совокупность n измеренных фазовых точек по классификации § 1.1 имеет, очевидно, тип (3); мы будем называть ее эмпирической *пространственной* совокупностью. Далее, можно построить другую эмпирическую совокупность фазовых точек, рассматривая лишь одну из копий нашей механической системы, про которую предполагается, что она находится в состоянии термического равновесия при абсолютной температуре T , но измеряя ее состояние, т. е. фазовую точку \vec{r} , в большом количестве последовательных моментов времени $t_1 < t_2 < \dots < t_n$ в течение большого промежутка времени. Эта вторая эмпирическая совокупность измеренных фазовых точек имеет, очевидно, тип (2) § 1.1; будем называть ее эмпирической *временной* совокупностью. Если вычислить относительную частоту n'_V/n в этой второй эмпирической совокупности, то, как показывает опыт, n'_V/n можно считать эмпирическим значением той же самой вероятности $p = \int \dots \int_V \varphi d\nu$. Этот приме-

чательный эмпирический факт, показывающий, что и пространственная и временная совокупности приводят к одному и тому же результату, является характеристической чертой систем, находящихся в равновесии. Размышление показывает, однако, что этот факт не является неожиданным; если система находится в равновесии, то естественно, что за достаточно большой промежу-

1) Следует заметить, что в действительности измерить n_V/n непосредственно, конечно, невозможно, так как даже единственную фазовую точку нельзя фактически измерить вследствие большого значения $f \sim 10^{23}$. Однако другие свойства системы измерить можно, так что проверить (3) можно косвенным образом.

ток времени изображающая систему фазовая точка $\vec{r} = \vec{r}(t)$ будет двигаться по фазовому пространству таким образом, что пройдет сколь угодно близко от каждой из фазовых точек, и время ее пребывания в любой области V фазового пространства будет приблизительно пропорционально вероятности $p = \int \dots \int_V \varphi dv$. Действительно, эта так называемая *эргодическая теорема*

в теории вероятностей может быть доказана при весьма общих условиях¹⁾.

Пример 2. Из канонического распределения (3) фазовой точки механической системы можно вывести распределение вероятностей для любой другой физической характеристики системы. Таким образом, если система состоит из N идентичных частиц (как, например, некоторое количество газа, состоящее из N идентичных молекул), так что число степеней свободы системы равно $2f = 6N$, то можно, например, исследовать следующий вопрос. Вместо того, чтобы рассматривать $\vec{r} = (q_1, \dots, q_f, p_1, \dots, p_f)$ как случайную величину $2f (= 6N)$ измерений, т. е. иметь дело с фазовой точкой в $6N$ -мерном фазовом пространстве, которое будем называть γ -пространством, можно рассматривать \vec{r} как совокупность N , вообще говоря, зависимых шестимерных случайных величин; тогда мы будем иметь дело с N фазовыми точками в шестимерном фазовом пространстве, которое будем называть μ -пространством. Если разбить μ -пространство на конечное число k областей $\omega_1, \dots, \omega_k$, которые назовем *ячейками*, то можно задаться вопросом об определении вероятности некоторого *распределения по ячейкам*, т. е. того, что из N фазовых точек ровно N_1 лежат в области ω_1 , N_2 в области ω_2, \dots, N_k в области ω_k . Поскольку числа N_1, \dots, N_k при разных измерениях (получаемых или по пространственной, или по временной совокупности) могут иметь различные значения, эти числа образуют k -мерную случайную величину. Распределение по ячейкам само может быть описано функцией распределения $F(p)$, равной доле всего количества фазовых точек, приходящейся на ячейки с номерами, не превосходящими $p = 1, 2, \dots, k$. Формально функция $F(p)$ является ступенчатой функцией, имеющей точно те же свойства, что и дискретная функция распределения (§ 4.3). Таким образом, мы видим, что в некоторых случаях полную функцию распределения можно рассматривать как случайную величину (с числом измерений, равным числу разрывов, в нашем случае равным k). Если выбрать очень большое k , то $F(p)$ можно аппроксимировать непрерывной функцией распределения. Таким образом, мы приходим к необходимости рассматривать всю непрерывную функцию распределения $F(t)$, т. е. каждое значение $F(t)$, как случайную величину и, таким образом, говорить о распределении распределения. Однако размерность этой случайной величины, являющейся элементом „функционального пространства“, очевидно, бесконечна. Поэтому мы оказываемся вне сферы классической теории вероятностей; но в современной, наиболее общей, теории Колмогорова можно рассматривать также и такие бесконечномерные вероятности.

Распределение по ячейкам в μ -пространстве не следует путать с каноническим распределением в γ -пространстве; первое всегда, и в классической, и в квантовой статистике, можно вывести из второго; дальнейшие сведения по вопросам, затронутым в настоящем параграфе, и по дальнейшим приложениям теории вероятностей читатель найдет в курсах статистической механики.

***§ 4.17. Квантовая теория.** В квантовой теории, являющейся, по существу, подробной теорией некоторого специального вида стохастических процессов, эти стохастические процессы описываются методами, отличными от методов,

¹⁾ См., например, Э. Хопф, Эргодическая теория, Успехи матем. наук, т. IV, вып. 1, 1949.

изложенных в § 4.15 и § 4.16). В то время как в явлении диффузии элемент случайности возникает лишь при переходе от одного состояния системы к другому, а в статистической механике этот элемент имеется лишь в начальном состоянии, а сам переход от одного состояния к другому является причинным, в квантовой теории элемент случайности имеется как в начальном состоянии, так и при переходе от одного состояния к другому. Дальнейшая разница заключается в том, что в статистической механике начальное состояние рассматривается как случайное лишь из практических соображений, в принципе же точное измерение начального состояния возможно; в квантовой теории, наоборот, начальное состояние является случайным в принципе вследствие того, что для атомных явлений воздействие измерительных приборов на измеряемую систему нельзя считать пренебрежимо малым (см. § 1.2). Таким образом, начальное состояние не может быть измерено с достаточной для использования причинного описания точностью. Каждое состояние системы может быть вследствие указанного описано лишь с помощью некоторой функции распределения, отличной, вообще говоря, от несобственного распределения $\epsilon(t - \mu)$. Переход от классической теории к квантовой как в механике, так и в электродинамике, производится с помощью замены несобственных распределений подпадающих точному измерению физических величин классической теории более общими распределениями (так называемое *квантование*). Эти общие распределения включают в себя классические несобственные распределения как частные предельные случаи, получающиеся при $\hbar \rightarrow 0$, где \hbar — новая фундаментальная постоянная, называемая *постоянной Планка*. Вместо непосредственного задания рассматриваемых распределений вероятностей с помощью их функций распределения в квантовой теории вводятся некоторые комплексные функции, называемые *амплитудами вероятности*, квадраты модулей которых равны плотности вероятности. Так, например, состояние одиночной частицы массы m , движущейся в поле консервативных сил с потенциальной энергией $V(x, y, z)$, характеризуется так называемой *волновой функцией Шредингера*, или амплитудой вероятности положения $\psi(x, y, z, t)$, удовлетворяющей волновому уравнению Шредингера

$$\Delta \psi - \frac{8\pi^2 m}{\hbar^2} \psi = -i \frac{4\pi m}{\hbar} \frac{\partial \psi}{\partial t}. \quad (1)$$

Это уравнение аналогично уравнению диффузии (4.15.4); в нем Δ определяется так же, как в (4.15.4), \hbar есть постоянная Планка, $i = \sqrt{-1}$. $|\psi|^2$ равно плотности вероятности для распределения значений (x, y, z) положения частицы в момент времени t , так что вероятность нахождения частицы в момент времени t в фиксированной области V обычного трехмерного пространства равна

$$p(V, t) = \iiint_V |\psi(x, y, z, t)|^2 dv; \quad \iiint_{-\infty}^{\infty} |\psi|^2 dv = 1. \quad (2)$$

У п р а ж н е н и е. Выведите из (1) уравнение неразрывности, соответствующее уравнению (4.15.2):

$$\operatorname{div} \vec{s} + \frac{\partial}{\partial t} |\psi|^2 = 0, \quad \vec{s} = \frac{\hbar}{4\pi m i} (\psi^* \operatorname{grad} \psi - \psi \operatorname{grad} \psi^*). \quad (3)$$

Здесь ψ^* есть величина, комплексно сопряженная с ψ , а $\operatorname{grad} \psi$ определяется

1) Основные эксперименты, теоретические принципы и математическая символика квантовой теории изложены, например, в книге В. Гейзенберга, *Физические принципы квантовой теории*, ГТТИ, 1932.

так же, как и в (4.15.3). (Умножьте (1) на ψ^* , соответствующее уравнение для ψ^* на ψ , и образуйте разность полученных выражений.)

То обстоятельство, что в математическом описании как первичные используются амплитуды вероятности, а не сами функции распределения, приводит к затруднениям, в результате которых возникает необходимость некоторого видоизменения обычных законов теории вероятностей (см., например, Гейзенберг, Физические принципы квантовой теории, ГТТИ, 1932, гл. IV, § 2). По ψ -функции, характеризующей некоторое состояние рассматриваемой механической системы, можно вычислить распределение любой физической величины. Дальнейшее изложение этого приложения теории вероятностей читатель найдет в курсах квантовой механики.

5. СРЕДНЕЕ ЗНАЧЕНИЕ И ДИСПЕРСИЯ

§ 5.1. Для полной характеристики распределения необходимо, конечно, знать всю функцию распределения $\Phi(t)$ или, в случае непрерывного распределения, всю плотность вероятности $\varphi(t)$ случайной величины x . Однако во многих практических приложениях масса вероятности случайной величины x сосредоточена в большей своей части внутри некоторого относительно узкого интервала, и поэтому часто указание на положение этого интервала посредством некоторой *характеристики расположения*, представляющей типичное значение случайной величины x , дает грубое представление о распределении в целом. Хотя любая такая характеристика, как правило, однозначно определяется функцией распределения $\Phi(t)$, обратное места не имеет.

С Можно построить бесконечно много таких характеристик расположения, но на практике чаще всего употребляются три следующие характеристики: *мода*, которая определяется как наиболее вероятное значение случайной величины x , т. е. такое значение t , для которого φ_i (в случае дискретного распределения) или $\varphi(t)$ (в случае непрерывного распределения) принимает максимальное значение; *медиана*, которая определяется как значение t , для которого $\Phi(t) = 1/2$, т. е. для которого вероятность того, что x принимает значение, меньшее t , равна вероятности того, что x принимает значение, большее t^* ; *среднее значение* или, короче, *среднее*, являющееся в механической интерпретации § 4.6 центром тяжести полной массы вероятности. Выбор между этими или какими-либо другими характеристиками расположения совершенно произволен и определяется исключительно соображениями удобства. Так как в большинстве случаев правила для вычисления среднего значения особенно просты, то эта характеристика расположения наиболее общеупотребительна, хотя в отдельных частных задачах могут оказаться более удобными другие характеристики.

Упражнение 1. Показать, что если $\varphi(t)$ существует и симметрична относительно $t=a$, то медиана и среднее значение (если оно существует) равны a , и если, кроме того, $\varphi(t)$ имеет только один максимум, то и мода также равна a .

*) Это определение годится только в том случае, когда распределение $\Phi(t)$ непрерывно. (Прим. ред.)

Из механической интерпретации § 4.6 и определения центра тяжести следует, что среднее значение случайной величины x , которое мы будем обозначать через $\mathfrak{M}\{x\}$ или μ_x , или, просто, через μ^1) в случае, если оно существует, равно

$$\mu = \mathfrak{M}\{x\} = \begin{cases} \frac{\sum_{i=-\infty}^{\infty} t_i \varphi_i}{\sum_{i=-\infty}^{\infty} \varphi_i} & \text{для дискретных распределений} \\ \frac{\int_{-\infty}^{\infty} t \varphi(t) dt}{\int_{-\infty}^{\infty} \varphi(t) dt} & \text{для непрерывных распределений} \end{cases} \quad (1)$$

(определение t_i и φ_i см. в § 4.3, определение $\varphi(t)$ — в § 4.4) Однако, так как мы предположили, что соотношения (4.3.2) и (4.4.3) всегда выполнены, для вычисления среднего значения мы не нуждаемся в нормировке. Поэтому формулы (1) записываются проще:

$$\mu = \mathfrak{M}\{x\} = \begin{cases} \sum_{i=-\infty}^{\infty} t_i \varphi_i & \text{для дискретных распределений} \\ \int_{-\infty}^{\infty} t \varphi(t) dt & \text{для непрерывных распределений} \end{cases} \quad (2)$$

(Кроме того предполагается, что сумма или интеграл в формуле (2) абсолютно сходятся, если μ вообще существует.)

Пример 1. Если случайная величина x ограничена, т. е. может принимать только значения, лежащие в конечном сегменте $g \leq t \leq G$, то в дискретном случае

$$g = g \sum_i \varphi_i \leq \sum_i t_i \varphi_i = \mu \leq G \sum_i \varphi_i = G$$

и в непрерывном случае

$$g = g \int_g^G \varphi(t) dt < \int_g^G t \varphi(t) dt = \mu < G \int_g^G \varphi(t) dt = G.$$

¹⁾ Другие обозначения для среднего значения: $\langle x \rangle$, $M\{x\}$, $E\{x\}$. Среднее значение называется также *математическим ожиданием* — термин, происходящий из приложений теории вероятностей к азартным играм. Символ \bar{x} , может быть, наиболее удобен для обозначения среднего значения и часто используется, например, в физике. Однако в статистике символ \bar{x} имеет несколько иное значение (см. § 6.4) и поэтому не будет употребляться нами для обозначения среднего. Мы пишем $\mathfrak{M}\{x\}$, а не $\mathfrak{M}(x)$, так как среднее является не функцией от x , а „функционалом“, т. е. числом, соответствующим всей функции распределения $\Phi(t)$ величины x .

В обоих случаях μ лежит между наибольшим и наименьшим значениями x , что оправдывает нашу терминологию.

Пример 2. Если случайная величина x принимает только два значения 0 и 1 с вероятностями, равными соответственно θ и $1 - \theta$, то

$$\mu = \mathfrak{M}\{x\} = 1 \cdot \theta + 0 \cdot (1 - \theta) = \theta, \quad (3)$$

т. е. среднее значение равно вероятности того, что x принимает значение, равное 1.

Пример 3. Если случайная величина x принимает с равными вероятностями значения $1, 2, \dots, \nu$, то

$$\mu = \mathfrak{M}\{x\} = 1 \cdot \frac{1}{\nu} + 2 \cdot \frac{1}{\nu} + \dots + \nu \cdot \frac{1}{\nu} = \frac{\nu(\nu+1)}{2} \cdot \frac{1}{\nu} = \frac{\nu+1}{2}. \quad (4)$$

Таким образом, если x есть результат броска игральной кости, то $\nu = 6$ и $\mu = 3,5$.

Пример 4. Для биномиального распределения (пример 2 § 4.3) в силу формулы бинома

$$\begin{aligned} \mu &= \mathfrak{M}\{x\} = \sum_{i=0}^{\nu} i \binom{\nu}{i} \theta^i (1-\theta)^{\nu-i} = \\ &= \nu \theta \sum_{i=1}^{\nu} \binom{\nu-1}{i-1} \theta^{i-1} (1-\theta)^{(\nu-1)-(i-1)} = \\ &= \nu \theta \sum_{i'=0}^{\nu-1} \binom{\nu-1}{i'} \theta^{i'} (1-\theta)^{(\nu-1)-i'} = \nu \theta. \end{aligned} \quad (5)$$

Позднее мы дадим более простое доказательство этой формулы (см. пример 1 § 6.4).

Пример 5. Для распределения Пуассона (пример 3 § 4.3)

$$\mathfrak{M}\{x\} = \sum_{i=0}^{\infty} i e^{-\mu} \frac{\mu^i}{i!} = e^{-\mu} \mu \sum_{i=1}^{\infty} \frac{\mu^{i-1}}{(i-1)!} = \mu, \quad (6)$$

что оправдывает обозначение параметра этого распределения через μ . Таким образом, в примере 4 § 4.3 $\mu = \lambda t$, т. е. λ есть среднее число вызовов за единицу времени.

* **Пример 6.** Результат предыдущего примера имеет важное применение в теории счетчиков Гейгера—Мюллера (см. пример 5 § 4.3). Пусть n счетчиков Гейгера—Мюллера имеют одно и то же время разрешения τ , являющееся минимальным интервалом времени между двумя последовательными попаданиями частиц, которые счетчик может фиксировать как два отдельных события. Пусть среднее число фикса-

рованных попаданий за единицу времени для каждого из n счетчиков равно соответственно N_1, \dots, N_n . Нас интересует среднее число фиктивных, или случайных, совпадений между показаниями n счетчиков за единицу времени. Условная вероятность фиксации каждым из оставшихся $n-1$ счетчиков хотя бы одного попадания за интервал времени τ , в предположении, что один из счетчиков, скажем 1, уже фиксировал попадание, равна

$$(1 - e^{-N_2\tau}) \dots (1 - e^{-N_n\tau})$$

(почему?). Следовательно, среднее число случайных совпадений, в которых счетчик 1 фиксирует попадание, равно этой вероятности, умноженной на N_1 . Любой из счетчиков 1, 2, \dots , n может оказаться первым, фиксировавшим попадание. Поэтому полное среднее число случайных совпадений за единицу времени в предположении, что $N_1\tau \ll 1, \dots, N_n\tau \ll 1$, равно

$$nN_1 \dots N_n \tau^{n-1}. \quad (7)$$

Пример 7. Для распределения Паскаля (пример 6 § 4.3) параметр μ также дает среднее значение:

$$\mathfrak{M}\{x\} = \sum_{i=0}^{\infty} i \frac{1}{1+\mu} \left(\frac{\mu}{1+\mu}\right)^i = \mu. \quad (8)$$

Это можно показать следующим, часто используемым приемом. Положив $\frac{\mu}{1+\mu} = x (< 1)$, получим

$$\sum_{i=0}^{\infty} ix^i = x \sum_{i=0}^{\infty} ix^{i-1} = x \frac{d}{dx} \sum_{i=0}^{\infty} x^i = x \frac{d}{dx} \frac{1}{1-x} = \frac{x}{(1-x)^2}.$$

Подставляя полученный результат в (8) и заменяя x его выражением через μ , сразу получаем требуемый результат.

Упражнение 2. Показать, что для несобственного распределения (пример 1 § 4.3), распределения Пуля (пример 7 § 4.3) и распределения Лапласа (пример 5 § 4.4) параметр μ также равен среднему значению.

Пример 8. Как мы увидим позднее (§ 7.3), для нормального распределения (пример 3 § 4.4) параметр μ также равен среднему значению, что следует также из результатов упражнения 1.

Пример 9. Для распределения Коши (пример 4 § 4.4) имеем

$$\begin{aligned} \mathfrak{M}\{x\} &= \frac{1}{\pi a} \int_{-\infty}^{\infty} t \frac{dt}{1 + \frac{(t-\mu)^2}{a^2}} = \int_{-\infty}^{\infty} ((t-\mu) + \mu) \varphi(t) dt = \\ &= \frac{a}{2\pi} \left| \ln \left(1 + \frac{(t-\mu)^2}{a^2} \right) \right|_{-\infty}^{\infty} + \mu = \infty - \infty + \mu. \end{aligned}$$

В этом случае выражение для среднего значения не определено. Однако в силу симметрии функции $\varphi(t)$ относительно $t = \mu$ естественно определить условно сходящийся интеграл так, что

$$\mathfrak{M}\{x\} = \mu \quad (9)$$

(так называемое главное значение Коши). Мода и медиана этого распределения существуют и равны параметру μ (см. упражнение 1).

Упражнение 3. Показать, что среднее значение времени существования радиоактивного атома (соответствующее распределение дано в примере 2 § 4.4) равно (см. приложение 1)

$$\mathfrak{M}\{x\} = \int_0^{\infty} t e^{-\lambda t} dt = \frac{1}{\lambda} \left(= \frac{T}{\ln 2} \right), \quad (10)$$

где T есть период полураспада (см. пример § 3.7). Таким образом в примере 4 § 4.3 $\frac{1}{\lambda}$ дает среднее время между двумя последовательными вызовами (см. пример 2 § 4.4), которое поэтому является обратным к среднему числу вызовов за единицу времени (см. пример 5).

§ 5.2. В самом общем случае мы определим среднее значение $\mathfrak{M}\{y\}$, где $y = f(x)$ (см. § 4.8), положив

$$\mathfrak{M}\{y\} = \mathfrak{M}\{f(x)\} = \begin{cases} \sum_{i=-\infty}^{\infty} f(t_i) \varphi_i & \text{при дискретном рас-} \\ & \text{пределении } x \\ \int_{-\infty}^{\infty} f(t) \varphi(t) dt & \text{при непрерывном рас-} \\ & \text{пределении } x \end{cases} \quad (1)$$

(если эти выражения существуют и абсолютно сходятся). Можно доказать¹⁾, что (1) дает то же самое число, которое получается из (5.1.2), если предварительно по распределению величины x найти распределение случайной величины y .

Упражнение 1. Доказать это при специальных предположениях § 4.8, используя (4.8.1).

Упражнение 2. Пусть x есть результат бросания игральной кости и

$$y = f(x) = \begin{cases} 0 & \text{при } x \text{ четном} \\ 1 & \text{при } x \text{ нечетном} \end{cases}$$

Найти распределение случайной величины y и показать, что (1) и (5.1.2) дают одно и то же значение для среднего, равное $1/2$.

Упражнение 3. Показать, что для $y = a$, где a — константа,

$$\mathfrak{M}\{a\} = a. \quad (2)$$

¹⁾ См., например, Крамер, Случайные величины и распределения вероятностей, стр. 29—30, 1947.

Показать, что для $y = ax$ имеем

$$\mathfrak{M}\{ax\} = a\mathfrak{M}\{x\}. \quad (3)$$

Если положить $y = y_1 + y_2$, где y_1 и y_2 — две произвольные функции одной и той же случайной величины x , то соотношение (1) дает

$$\mathfrak{M}\{y_1 + y_2\} = \mathfrak{M}\{y_1\} + \mathfrak{M}\{y_2\}. \quad (4)$$

В § 6.1 мы докажем, что это имеет место и в общем случае.

Из соотношений (4) и (2) получим

$$\mathfrak{M}\{x + a\} = \mathfrak{M}\{x\} + a. \quad (5)$$

Подчеркнем, что если $f(x)$ не является линейной функцией, то среднее значение неинвариантно относительно преобразования $y = f(x)$, т. е.

$$\mathfrak{M}\{f(x)\} \neq f(\mathfrak{M}\{x\}). \quad (6)$$

Например, в упражнении 2 выражение $f(\mathfrak{M}\{x\})$ даже не определено (ср. § 6.5).

*Пример 1. При распределении скоростей Максвелла—Больцмана (пример § 4.8) можно с помощью приложения 1 найти

$$\begin{aligned} \mathfrak{M}\{v\} &= \frac{\int_0^{\infty} at^3 \exp[-\beta t^2] dt}{\int_0^{\infty} at^2 \exp[-\beta t^2] dt} = \frac{\beta^{3/2} \int_0^{\infty} x^3 \exp[-x^2] dx}{\beta^{3/2} \int_0^{\infty} x^2 \exp[-x^2] dx} = \\ &= \frac{2}{\sqrt{\pi\beta}} = 2 \sqrt{\frac{2kT}{\pi m}} \end{aligned} \quad (7)$$

и для соответствующего распределения энергии

$$\begin{aligned} \mathfrak{M}\{E\} &= \frac{\int_0^{\infty} a'u^{3/2} \exp[-\beta'u] du}{\int_0^{\infty} a'u^{1/2} \exp[-\beta'u] du} = \frac{\beta'^{3/2} \int_0^{\infty} x^{3/2} e^{-x} dx}{\beta'^{3/2} \int_0^{\infty} x^{1/2} e^{-x} dx} = \\ &= \frac{3}{2\beta'} = \frac{3}{2} kT. \end{aligned} \quad (8)$$

Таким образом, $\mathfrak{M}\{E\} = \frac{3}{2} kT$ не равно $\frac{1}{2} m(\mathfrak{M}\{v\})^2 = \frac{4}{\pi} kT$. Заметим, что этот пример показывает, что не всегда (например, тогда, когда нас интересуют только средние значения) бывает необходимо вычислять нормирующий множитель a .

Последний пример демонстрирует серьезный недостаток среднего значения в качестве характеристики расположения, а именно его инвариантность относительно нелинейных преобразований. Поэтому медиана иногда бывает предпочтительнее среднего значения, так как при условиях § 4.8 медиана обладает указанным свойством инвариантности.

У п р а ж н е н и е 4. Проверить это.

*Пример 2. В литературе часто используются средние значения некоторых простых функций случайных величин (для простоты мы рассматриваем здесь только непрерывные распределения и обозначаем через x переменную интегриации). Для функции $y = x^k$, где k — действительное неотрицательное число, получим из (1)

$$\mathfrak{M}\{x^k\} = \int_{-\infty}^{\infty} x^k \varphi(x) dx = \mu_k. \quad (9)$$

Это выражение, если оно существует, называется *моментом порядка k* случайной величины x (относительно точки $x=0$). Аналогично, $\mathfrak{M}\{|x|^k\}$ называется *абсолютным моментом порядка k* , а $\mathfrak{M}\{(x-\mu)^k\}$ — *центральным моментом порядка k* случайной величины x . Мы видим, что μ_k есть обобщение $\mu_1 = \mu$ и что если μ_k существует, то оно однозначно определяется функцией $\varphi(x)$. Обратно, при некоторых специальных условиях функция $\varphi(x)$ однозначно определяется своими моментами μ_1, μ_2, \dots целых неотрицательных порядков¹⁾.

Для функции $y = x^{(k)} = x(x-1)\dots(x-k+1)$, $k=1, 2, 3, \dots$ получим из (1)

$$\mathfrak{M}\{x^{(k)}\} = \int_{-\infty}^{\infty} x(x-1)\dots(x-k+1)\varphi(x) dx = \mu_{(k)}. \quad (10)$$

Это выражение, если оно существует, называется *факториальным моментом порядка k* случайной величины x .

Для функции $y = t^x$, где t — действительное неотрицательное число, получим из (1)

$$\mathfrak{M}\{t^x\} = \int_{-\infty}^{\infty} t^x \varphi(x) dx = \gamma_x(t). \quad (11)$$

Это выражение является функцией от t и в случае, если оно существует, называется *производящей функцией* случайной величины x .

¹⁾ См., например, Крамер, Математические методы статистики стр. 199, 1948.

Для функции $y = e^{tx}$, где t есть действительное число, получим из (1)

$$\mathfrak{M}\{e^{tx}\} = \int_{-\infty}^{\infty} e^{tx} \varphi(x) dx = \mu_x(t). \quad (12)$$

Это выражение является функцией от t и в случае, если оно существует, называется *производящей функцией моментов* случайной величины x . Если функция $\mu_x(t)$ может быть разложена в ряд по степеням t , то k -ый коэффициент этого разложения будет равен, очевидно, $\mu_k/k!$

Для функции $y = e^{itx}$, где t — действительное число, $i = \sqrt{-1}$, получим из (1)

$$\mathfrak{M}\{e^{itx}\} = \int_{-\infty}^{\infty} e^{itx} \varphi(x) dx = \chi_x(t). \quad (13)$$

Эта функция от t существует для всех значений t и для всех распределений. Она называется *характеристической функцией* случайной величины x и играет большую роль в современной теории вероятностей¹⁾. Если эта функция может быть разложена в ряд по степеням it , то k -ый коэффициент этого разложения равен, очевидно, $\mu_k/k!$ Функция $\chi_x(t)$ не только однозначно определяется функцией распределения $\Phi(t)$, но и сама однозначно определяет ее. Для непрерывного распределения из теории интегралов Фурье вытекает, что

$$\varphi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \chi(t) dt, \quad (14)$$

а для дискретного распределения из теории так называемых [почти] периодических функций следует, что

$$\varphi_j = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T e^{itx_j} \chi(t) dt, \quad (15)$$

где x_j есть j -ая точка разрыва функции $\Phi(t)$, а φ_j — скачок функции $\Phi(t)$ в этой точке (см. § 4.3; если x_j не является точкой разрыва функции $\Phi(t)$, то выражение (15) равно нулю). Далее, имеет место важная теорема, утверждающая, что если функции распределения $\Phi_1(t)$, $\Phi_2(t)$, ... имеют соответственно характеристические функции $\chi_1(t)$, $\chi_2(t)$, ..., то для того, чтобы последовательность функций

¹⁾ Доказательства следующих ниже утверждений и их приложения читатель может найти, например, в книгах Крамера Математические методы статистики, 1948, или Случайные величины и распределения вероятностей, гл. IV, 1947.

$\Phi_n(t)$ сходилась к функции распределения*) $\Phi(t)$, необходимо и достаточно, чтобы при любом t последовательность $\chi_n(t)$ сходилась к пределу $\chi(t)$, непрерывному при $t=0$. В этом случае $\chi(t)$ является характеристической функцией функции распределения $\Phi(t)$.

* У п р а ж н е н и е 5. Выписать выражения, соответствующие (9) — (13), для дискретных распределений.

* У п р а ж н е н и е 6. Показать, что если случайная величина x имеет характеристическую функцию $\chi_x(t)$, то $y = ax + b$, где a и b — произвольные константы, имеет характеристическую функцию $\chi_y(t) = \exp[bit] \chi_x(at)$.

* У п р а ж н е н и е 7. Показать, что для следующих распределений характеристическими функциями являются:

$$\chi(t) = e^{i\mu t} \text{ для несобственного распределения,} \quad (16)$$

$$\chi(t) = (1 + \theta(e^{it} - 1))^\nu \text{ для биномиального распределения,} \quad (17)$$

$$\chi(t) = \exp[\mu(e^{it} - 1)] \text{ для распределения Пуассона,} \quad (18)$$

$$\chi(t) = (1 + \mu(1 - e^{it}))^{-1} \text{ для распределения Паскаля,} \quad (19)$$

$$\chi(t) = (1 + \beta\mu(1 - e^{it}))^{-1/\beta} \text{ для распределения Поля,} \quad (20)$$

$$\chi(t) = \exp\left[i\mu t - \frac{\sigma^2}{2} t^2\right] \text{ для нормального распределения,} \quad (21)$$

$$\chi(t) = \exp[i\mu t - \alpha |t|] \text{ для распределения Коши,} \quad (22)$$

$$\chi(t) = \frac{\exp[i\mu t]}{1 + \alpha^2 t^2} \text{ для распределения Лапласа.} \quad (23)$$

Найти соответствующие производящие функции и производящие функции моментов.

* У п р а ж н е н и е 8. Разлагая натуральный логарифм функции $\chi(t)$ в ряд по степеням it , будем записывать коэффициенты этого разложения в виде $\frac{x_k}{k!}$. Величины x_k называются *семи-инвариантами* или *кумулянтами* случайной величины x . Показать (без исследования вопросов, связанных с существованием и сходимостью), что

$$\begin{aligned} x_1 &= \mu_1, & \mu_1 &= x_1, \\ x_2 &= \mu_2 - \mu_1^2, & \mu_2 &= x_2 + x_1^2, \\ x_3 &= \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3, & \mu_3 &= x_3 + 3x_1x_2 + x_1^3, \\ & \dots & & \dots \end{aligned} \quad (24)$$

* У п р а ж н е н и е 9. Показать, что

$$\chi_x(1) = 1, \quad \mathfrak{M}\{x\} = \left[\frac{\partial \chi_x(t)}{\partial t} \right]_{t=1}, \quad \mathfrak{M}\{x^2\} = \left[\frac{\partial^2 \chi_x(t)}{\partial t^2} + \frac{\partial \chi_x(t)}{\partial t} \right]_{t=1}, \quad (25)$$

$$\mu_x(0) = 1, \quad \mathfrak{M}\{x\} = \left[\frac{\partial \mu_x(t)}{\partial t} \right]_{t=0}, \quad \mathfrak{M}\{x^2\} = \left[\frac{\partial^2 \mu_x(t)}{\partial t^2} \right]_{t=0} \quad (26)$$

$$\chi_x(0) = 1, \quad \mathfrak{M}\{x\} = -t \left[\frac{\partial \chi_x(t)}{\partial t} \right]_{t=0}, \quad \mathfrak{M}\{x^2\} = - \left[\frac{\partial^2 \chi_x(t)}{\partial t^2} \right]_{t=0}. \quad (27)$$

Проверить эти формулы для распределений, указанных в упражнении 7.

*) Строго говоря, сходимость здесь понимается как сходимость в каждой точке непрерывности предельной функции $\Phi(t)$. (Прим. ред.)

* У п р а ж н е н и е 10. Решение вероятностной задачи часто облегчается, если найти предварительно производящую функцию, производящую функцию моментов или характеристическую функцию. Вывести из (4.3.6) дифференциальное уравнение для производящей функции $\varphi(u)$, решить его и получить (4.3.7), разлагая решение в ряд по степеням u .

§ 5.3. Как уже подчеркивалось, любая характеристика расположения не определяет распределение однозначно, а только дает грубое представление о нем. Однако для целей практики часто бывает достаточно знать одну из таких характеристик (скажем, среднее значение) и характеристику рассеяния, показывающую, как разбросана масса вероятности около выбранной характеристики расположения (хотя, конечно, эти два числа, представляющие значения выбранных характеристик расположения и рассеяния, не будут однозначно определять функцию распределения $\Phi(t)$). На практике используются следующие характеристики рассеяния: *среднее отклонение* $\mathbb{M}\{|x - \mu|\}$ (где μ — среднее значение); *дисперсия* $\sigma^2\{x\}$, являющаяся в механической интерпретации § 4.6 моментом инерции полной массы вероятностей относительно $x = \mu$, или квадратный корень из нее, $\sigma\{x\}$, называемый *стандартным* (или *средним квадратическим*) *отклонением*; *половина полушироты*, т. е. $\gamma = \frac{1}{2}(t_2 - t_1)$, где t_1, t_2 — корни уравнения $\varphi(t) = \frac{1}{2}\varphi_{\max}$ (φ_{\max} — максимальное значение функции $\varphi(t)$), причем предполагается, что указанное уравнение имеет только два корня.

Кроме этого, иногда используется *семи-интерквартильная широта* $\frac{1}{2}(t_{3/4} - t_{1/4})$, где $t_{3/4}$ и $t_{1/4}$, являющиеся соответственно корнями уравнений $\Phi(t) = \frac{3}{4}$ и $\Phi(t) = \frac{1}{4}$, называются *квартилями*, или, вообще, *семи-интерквантильная широта* $\frac{1}{2}(t_1 - \theta - t_0)$, где t_0 есть корень уравнения $\Phi(t) = \theta < 1$, называемый *квантилью* порядка θ (при $\theta = p/10$; $p = 1, 2, \dots, 9$ величины t_θ называются также *децилями*). Понятие квантили является очевидным обобщением медианы, которая равна $t_{1/2}$ *). Если случайная величина x ограничена (см. пример 1 § 5.1), то иногда удобной характеристикой рассеяния оказывается половина *широты*, т. е. $\frac{1}{2}(G - g)$.

Так как правила для вычисления дисперсии $\sigma^2\{x\}$ особенно просты, эта характеристика рассеяния наиболее употребительна, хотя иногда применяются и другие характеристики. Так, например, в физике часто в качестве характеристики рассеяния применяется полуширота (см.

* Здесь, подобно сказанному на стр. 65, имеются в виду непрерывные распределения $\Phi(t)$. В случае дискретных распределений определение t_θ должно быть изменено (см. Крамер, Математические методы статистики, стр. 201, 1948). (Прим. ред.)

пример 10). В силу определения момента инерции дисперсия $\sigma^2 \{x\}$, обозначаемая также через σ_x^2 или, короче, через σ^2 , равна¹⁾

$$\sigma^2 \{x\} = \begin{cases} \sum_{i=-\infty}^{\infty} (t_i - \mu)^2 \varphi_i & \text{для дискретных распределений} \\ \int_{-\infty}^{\infty} (t - \mu)^2 \varphi(t) dt & \text{для непрерывных распределений} \end{cases} \quad (1)$$

(μ — среднее значение, t_i , φ_i определены в § 4.3, $\varphi(t)$ в § 4.4). Величина $\delta = \frac{\sigma}{\mu}$ называется *коэффициентом изменчивости*.

Пример 1. Для непрерывных распределений из (1) в силу кусочной непрерывности функции $\varphi(t) \geq 0$ вытекает, что $\sigma^2 > 0$. Для дискретного распределения может иметь место $\sigma^2 = 0$, но тогда из (1) вытекает, что это распределение несобственное, т. е. $\Phi = \varepsilon(t - \mu)$ (см. (4.3.3)).

В силу (5.2.1)

$$\sigma^2 \{x\} = \mathfrak{M} \{(x - \mu)^2\}. \quad (2)$$

Таким образом, σ^2 является центральным моментом второго порядка (см. пример 2 § 5.2).

Упражнение 1. Показать, что если a — константа, то

$$\sigma^2 \{ax\} = a^2 \sigma^2 \{x\}, \quad (3)$$

$$\sigma^2 \{x + a\} = \sigma^2 \{x\}. \quad (4)$$

Упражнение 2. Пусть случайная величина x имеет среднее значение μ и дисперсию σ^2 . Показать, что для случайной величины $y = \frac{x - \mu}{\sigma}$ имеем $\mathfrak{M} \{y\} = 0$ и $\sigma^2(y) = 1$. Такие случайные величины называются *нормированными*. Мы видим, что нормирование случайной величины x означает просто определенный удобный выбор начала отсчета и масштаба на оси x .

Упражнение 3. Пусть нормированная величина y имеет функцию распределения $\Phi(t)$ и плотность вероятности $\varphi(t)$. Показать, что случайная величина $x = \sigma y + \mu$ имеет функцию распределения $\Phi\left(\frac{t - \mu}{\sigma}\right)$, плотность вероятности

$$\frac{1}{\sigma} \varphi\left(\frac{t - \mu}{\sigma}\right), \quad \mathfrak{M} \{x\} = \mu, \quad \sigma \{x\} = \sigma.$$

* Упражнение 4. Показать, что если нормированная величина y имеет характеристическую функцию $\chi_y(t)$, то $x = \sigma y + \mu$ имеет характеристическую функцию $\chi_x(t) = e^{i\mu t} \chi_y(\sigma t)$.

¹⁾ Для $\sigma^2 \{x\}$ часто применяется также обозначение $D \{x\}$.

Для произвольной константы a имеет место тождество

$$(x - a)^2 = ((x - \mu) + (\mu - a))^2 = (x - \mu)^2 + 2(x - \mu)(\mu - a) + (\mu - a)^2.$$

Находя среднее значение от обеих частей равенства, получим

$$\mathfrak{M}\{(x - a)^2\} = \sigma^2\{x\} + (\mu - a)^2. \quad (5)$$

У п р а ж н е н и е 5. Проверить это. Как формулируется соответствующая теорема для моментов инерции (теорема Штейнера)?

Если, в частности, в равенстве (5) положить $a = 0$, то мы получим важное соотношение, часто употребляемое для вычисления σ^2 :

$$\sigma^2\{x\} = \mathfrak{M}\{x^2\} - \mathfrak{M}^2\{x\}. \quad (6)$$

П р и м е р 2. Из примера 2 § 5.2 видно, что последнее равенство может быть записано в виде $\sigma^2 = \mu_2 - \mu_1^2$ или, из упражнения 8 § 5.2, $\sigma^2 = x_2$.

У п р а ж н е н и е 6. Показать, что

$$\sigma^2\{x\} = \mathfrak{M}\{x(x - 1)\} - \mu(\mu - 1). \quad (7)$$

Эта формула часто более удобна, чем (6), особенно в случае дискретных распределений.

*П р и м е р 3. Из примера 2 § 5.2 видно, что равенство (7) может быть записано также в виде $\sigma^2 = \mu_{(2)} - \mu_1(\mu_1 - 1)$.

П р и м е р 4. Для случайной величины x из примера 2 § 5.1 имеем в силу (5.1.3)

$$\sigma^2\{x\} = (1 - \theta)^2 \theta + (0 - \theta)^2 (1 - \theta) = \theta(1 - \theta). \quad (8)$$

П р и м е р 5. Для случайной величины x из примера 3 § 5.1 имеем в силу хорошо известной алгебраической теоремы

$$\mathfrak{M}\{x^2\} = \sum_{i=1}^{\nu} i^2 \frac{1}{\nu} = \frac{\nu(\nu+1)(2\nu+1)}{6} \cdot \frac{1}{\nu} = \frac{(\nu+1)(2\nu+1)}{6},$$

т. е. в силу (6) и (5.1.4)

$$\sigma^2\{x\} = \frac{(\nu+1)(2\nu+1)}{6} - \left(\frac{\nu+1}{2}\right)^2 = \frac{\nu^2 - 1}{12}. \quad (9)$$

П р и м е р 6. Для биномиального распределения (см. пример 2 § 4.3) имеем в силу формулы бинома

$$\begin{aligned} \mathfrak{M}\{x(x-1)\} &= \sum_{i=0}^{\nu} i(i-1) \binom{i}{\nu} \theta^i (1-\theta)^{\nu-i} = \\ &= \nu(\nu-1) \theta^2 \sum_{i=2}^{\nu} \binom{i-2}{\nu-2} \theta^{i-2} (1-\theta)^{(\nu-2)-(i-2)} = \\ &= \nu(\nu-1) \theta^2, \end{aligned}$$

т. е. в силу (7) и (5.1.5)

$$\sigma^2 \{x\} = \nu(\nu - 1) \theta^2 - (\nu \theta)(\nu \theta - 1) = \nu \theta (1 - \theta). \quad (10)$$

Позднее (см. пример 1 § 6.4) мы дадим более простое доказательство этой формулы.

Пример 7. Для распределения Пуассона (см. пример 3 § 4.3)

$$\mathfrak{M} \{x(x-1)\} = \sum_{i=0}^{\infty} i(i-1) e^{-\mu} \frac{\mu^i}{i!} = e^{-\mu} \mu^2 \sum_{i=2}^{\infty} \frac{\mu^{i-2}}{(i-2)!} = \mu^2,$$

т. е. в силу (7) и (5.1.6)

$$\sigma^2 \{x\} = \mu^2 - \mu(\mu - 1) = \mu. \quad (11)$$

Таким образом коэффициент изменчивости равен $\delta = \frac{\sigma}{\mu} = \frac{1}{\sqrt{\mu}}$, т. е. δ убывает с возрастанием μ , что означает все большую и большую концентрацию массы вероятности около μ .

Пример 8. Для распределения Паскаля (пример 6 § 4.3), полагая $\frac{\mu}{1+\mu} = x$, как мы это делали в примере 7 § 5.1 при вычислении $\mathfrak{M} \{x\}$, находим, что

$$\begin{aligned} \mathfrak{M} \{x(x-1)\} &= \sum_{i=0}^{\infty} i(i-1) \frac{1}{1+\mu} \left(\frac{\mu}{1+\mu}\right)^i = \\ &= \frac{x^2}{1+\mu} \frac{d^2}{dx^2} \sum_{i=0}^{\infty} x^i = \frac{x^2}{1+\mu} \frac{d^2}{dx^2} \frac{1}{1-x} = 2\mu^2, \end{aligned}$$

т. е. в силу (7) и (5.1.8)

$$\sigma^2 \{x\} = 2\mu^2 - \mu(\mu - 1) = \mu^2 + \mu. \quad (12)$$

Пример 9. Позднее мы увидим (§ 7.3), что для нормального распределения (см. пример 3 § 4.4) дисперсия равна квадрату параметра σ .

Пример 10. Для распределения Коши (пример 4 § 4.4) имеем

$$\begin{aligned} \sigma^2 \{x\} &= \frac{1}{\pi \alpha} \int_{-\infty}^{\infty} (t - \mu)^2 \frac{dt}{1 + \frac{(t - \mu)^2}{\alpha^2}} = \frac{\alpha^2}{\pi} \int_{-\infty}^{\infty} \frac{x^2 dx}{1 + x^2} = \\ &= \frac{\alpha^2}{\pi} \left[x - \arctg x \right]_{-\infty}^{\infty} = 2\alpha - \alpha^2 = \infty. \quad (13) \end{aligned}$$

Таким образом, в этом случае дисперсия не существует. С другой стороны, половина полушироты вычисляется легко и оказывается равной α . Поэтому она является удобной характеристикой рассеяния для этого распределения, часто употребляемой в физике.

У п р а ж н е н и е 7. Показать с помощью формулы бинома, что для распределения Пуассона (см. пример 7 § 4.3)

$$\mathfrak{M}\{x(x-1)\} = (1 + \beta)\mu^2,$$

т. е. в силу (7) и упражнения 2 § 5.1

$$\sigma^2\{x\} = \mu(1 + \beta\mu). \quad (14)$$

Последнее равенство содержит (11) и (12) в качестве частных случаев.

У п р а ж н е н и е 8. Показать с помощью приложения 1 и упражнения 2 § 5.1, что для распределения Лапласа (см. пример 5 § 4.4)

$$\sigma^2\{x\} = 2\alpha^2. \quad (15)$$

У п р а ж н е н и е 9. Показать с помощью приложения 1 и (5.1.10), что для распределения, приведенного в примере 2 § 4.4,

$$\sigma^2\{x\} = \frac{1}{\lambda^2}. \quad (16)$$

* § 5.4. С помощью интеграла Стильтьеса, определенного в § 4.7, для произвольного распределения определяется

$$\mu = \mathfrak{M}\{x\} = \int_{-\infty}^{\infty} t d\Phi(t). \quad (1)$$

Тогда для произвольной функции $y = f(x)$ имеем

$$\mathfrak{M}\{y\} = \mathfrak{M}\{f(x)\} = \int_{-\infty}^{\infty} f(t) d\Phi(t). \quad (2)$$

Полагая $y = (x - \mu)^2$, получим

$$\sigma^2\{x\} = \int_{-\infty}^{\infty} (t - \mu)^2 d\Phi(t). \quad (3)$$

Для $y = x^k$ получим

$$\mathfrak{M}\{x^k\} = \int_{-\infty}^{\infty} t^k d\Phi(t) = \mu_k. \quad (4)$$

Для $y = e^{itx}$ получим

$$\mathfrak{M}\{e^{itx}\} = \int_{-\infty}^{\infty} e^{itx} d\Phi(x) = \chi_x(t) \quad (5)$$

и т. д. Читателю предоставляется проверить, что в случае дискретных или непрерывных распределений эти формулы сводятся к установленным ранее.

§ 5.5. Для двумерной случайной величины $z = (x, y)$ среднее значение (в том случае, если оно существует) определяется как центр тяжести полной массы вероятности, т. е. посредством ее „статиче-

ских моментов" относительно осей t и u :

$$\mathfrak{M}\{z\} = (\mu_x, \mu_y) = \begin{cases} \left(\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} t_i \varphi_{ij}, \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} u_j \varphi_{ij} \right) = \\ = \left(\sum_{i=-\infty}^{\infty} t_i \varphi_{i \cdot}, \sum_{j=-\infty}^{\infty} u_j \varphi_{\cdot j} \right) \\ \text{для дискретных распределений.} \\ \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} t \varphi(t, u) dt du, \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u \varphi(t, u) dt du \right) = \\ = \left(\int_{-\infty}^{\infty} t \varphi_x(t) dt, \int_{-\infty}^{\infty} u \varphi_y(u) du \right) \\ \text{для непрерывных распределений.} \end{cases} \quad (1)$$

(Относительно обозначений см. § 4.10—4.13; сходимость, вообще говоря, предполагается абсолютной). Таким образом, для вычисления $\mathfrak{M}\{z\}$ нужно знать только два частных распределения, так как в силу (1)

$$\mathfrak{M}\{z\} = (\mu_x, \mu_y) = (\mathfrak{M}\{x\}, \mathfrak{M}\{y\}). \quad (2)$$

Пример 1. Для двумерного нормального распределения (4.12.7) частные распределения (4.13.6) и (4.13.7) показывают, что параметры μ_x и μ_y дают две компоненты среднего значения $\mathfrak{M}\{z\}$.

* Пример 2. Вычисляя средние значения условных распределений (для простоты мы рассматриваем только непрерывный случай, § 4.13), получаем две функции, которые в предположении, что они существуют, называются соответственно *регрессией x на y* и *регрессией y на x* :

$$\begin{aligned} \mathfrak{M}\{x|u\} &= \int_{-\infty}^{\infty} t \varphi(t|u) dt = \mu_x(u), \\ \mathfrak{M}\{y|t\} &= \int_{-\infty}^{\infty} u \varphi(u|t) du = \mu_y(t). \end{aligned} \quad (3)$$

Заметим, что функции $t = \mu_x(u)$ и $u = \mu_y(t)$ не будут, вообще говоря, взаимно обратными. Графики этих двух функций называются *кривыми регрессии*. Для нормального распределения из соотношений (4.13.13) и (4.13.14) вытекает, что кривые регрессии являются прямыми линиями, пересекающимися в точке (μ_x, μ_y) :

$$\begin{aligned} \mu_x(u) &= \mu_x + \rho \frac{\sigma_x}{\sigma_y} (u - \mu_y), \\ \mu_y(t) &= \mu_y + \rho \frac{\sigma_y}{\sigma_x} (t - \mu_x). \end{aligned} \quad (4)$$

Эти прямые существуют и для других распределений (см. § 5.6) и называются *прямыми средней квадратической регрессии*. Угловые коэффициенты этих прямых называются *коэффициентами регрессии*. Заметим, что прямые средней квадратической регрессии совпадают тогда и только тогда, когда $\rho = \pm 1$. Однако в этом случае вся масса вероятности сосредоточена на прямой регрессии, т. е. величины x и y подчинены линейной зависимости, и распределение поэтому сводится к одномерному распределению (см. пример 1 § 4.12). Понятие регрессии играет важную роль в статистике. За дальнейшими сведениями, однако, мы отсылаем читателя к учебникам статистики, например, к неоднократно уже цитированной книге Крамера.

Для функции $w = f(z) = f(x, y)$ случайной величины z , где величина w может быть либо одномерной, либо двумерной, среднее значение величины w , если оно существует, определяется соотношениями

$$\mathfrak{M}\{w\} = \begin{cases} \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} f(t_i, u_j) \varphi_{ij} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t, u) \varphi(t, u) dt du, \end{cases} \quad (5)$$

являющимися обобщением соотношений (5.2.1). Можно показать, что этот результат совпадает с тем, который получается, если сперва вычислить распределение величины w (например, так как это делалось в § 4.14), а затем применить (5.1.2)¹⁾. Подчеркнем, что, вообще говоря (ср. (5.2.6)),

$$\mathfrak{M}\{f(z)\} \neq f(\mathfrak{M}\{z\}). \quad (6)$$

В следующей главе мы рассмотрим две простые функции $w = x + y$ и $w = xy$.

* Пример 3. Величины, определенные в примере 2 § 5.2, могут быть легко обобщены на случай двумерных распределений. Так, например, для непрерывных распределений моменты k -го порядка (относительно точки $(0, 0)$) определяются, если они существуют, соотношением

$$\mathfrak{M}\{x^l y^m\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^l y^m \varphi(x, y) dx dy, \quad l + m = k. \quad (7)$$

Характеристическая функция определяется выражением

$$\begin{aligned} \mathfrak{M}\{\exp[i(tx + uy)]\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[i(tx + uy)] \varphi(x, y) dx dy = \\ &= \chi_{x, y}(t, u), \end{aligned} \quad (8)$$

¹⁾ См., например, Крамер, Случайные величины и распределения вероятностей, стр. 29—30.

являющимися функцией от t и u , существующей для всех значений t и u и всех распределений. Если она может быть разложена в ряд по степеням it и iu , то коэффициент при $(it)^l (iu)^m$ равен, как легко видеть, соответствующему моменту $\mathfrak{M}\{x^l y^m\}$. В двумерном случае, так же как и в одномерном, функция распределения Φ однозначно определяется характеристической функцией $\chi(t, u)$ и получается из нее при помощи двумерных обобщений соотношений (5.2.14) и (5.2.15).

* У п р а ж н е н и е 1. Показать, что характеристические функции двух частных распределений равны соответственно $\chi_x(t) = \chi_{x,y}(t, 0)$ и $\chi_y(u) = \chi_{x,y}(0, u)$.

У п р а ж н е н и е 2. Показать, что характеристическая функция двумерного нормального распределения является простым обобщением (5.2.21)

$$\chi_{x,y}(t, u) = \exp\left[i(\mu_x t + \mu_y u) - \frac{1}{2}(\sigma_x^2 t^2 + 2\rho\sigma_x\sigma_y tu + \sigma_y^2 u^2)\right]. \quad (9)$$

§ 5.6. При механической интерпретации § 4.6 мы можем говорить также о четырех „кватрических моментах“ массы вероятности относительно точки (μ_x, μ_y) , представляющих естественное обобщение дисперсии одномерного распределения¹⁾:

$$\begin{aligned} \mathfrak{M}\{(x - \mu_x)^2\} &= \mu_{xx} = \sigma_x^2, \\ \mathfrak{M}\{(y - \mu_y)^2\} &= \mu_{yy} = \sigma_y^2, \\ \mathfrak{M}\{(x - \mu_x)(y - \mu_y)\} &= \mu_{xy} = \mu_{yx}. \end{aligned} \quad (1)$$

Затем мы вводим коэффициент корреляции $\rho\{x, y\} = \rho_{xy}$, обозначаемый короче через ρ , между случайными величинами x и y , полагая

$$\rho\{x, y\} = \rho = \frac{\mu_{xy}}{\sigma_x\sigma_y}. \quad (2)$$

Как показано в примере § 6.1,

$$-1 \leq \rho \leq 1. \quad (3)$$

У п р а ж н е н и е. Показать, что если величины x и y пропорциональны, то

$$\begin{aligned} \rho &= 1 \quad \text{при } x = a^2 y, \\ \rho &= -1 \quad \text{при } x = -a^2 y, \end{aligned} \quad (4)$$

где a — произвольное действительное число, не равное нулю.

В обычных случаях $|\rho| < 1$. В § 6.2 мы увидим, что если величины x и y независимы, то $\rho = 0$. Таким образом, величина ρ может служить характеристикой степени зависимости или корреляции случайных величин x и y , что и оправдывает ее название. Однако если, наоборот, $\rho\{x, y\} = 0$, то величины x и y не обязательно независимы

¹⁾ Моменты (1) называются *центральными моментами* второго порядка. Величина $\mu_{xy} = \mu_{yx}$ называется *смешанным моментом* второго порядка.

см. пример 1 § 6.2). В случае $\rho=0$ величины x и y называются иногда *некоррелированными*.

Пример. Для двумерного нормального распределения (4.12.7) из выражений (4.13.6) и (4.13.7) для частных распределений непосредственно видно, что параметры σ_x и σ_y являются квадратическими моментами, определенными соотношениями (1). Позднее мы увидим (§ 7.6), что параметр ρ является коэффициентом корреляции, определенным соотношением (2).

Читателю предоставляется обобщить содержание §§ 5.5—5.6 случая двух измерений на случай многих измерений.

6. СРЕДНЕЕ ЗНАЧЕНИЕ И ДИСПЕРСИЯ СУММЫ, ПРОИЗВЕДЕНИЯ И ДРУГИХ ФУНКЦИЙ СЛУЧАЙНЫХ ВЕЛИЧИН

§ 6.1. Если в (5.5.5) положить $\mathbf{w} = \mathbf{x} + \mathbf{y}$, то (в предположении абсолютной сходимости) получим

$$\begin{aligned} \mathfrak{M} \{ \mathbf{x} + \mathbf{y} \} &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} (t_i + u_j) \varphi_{ij} = \\ &= \sum_{i=-\infty}^{\infty} t_i \varphi_{i.} + \sum_{j=-\infty}^{\infty} u_j \varphi_{.j} = \mathfrak{M} \{ \mathbf{x} \} + \mathfrak{M} \{ \mathbf{y} \} \end{aligned} \quad (1)$$

для дискретных и

$$\begin{aligned} \mathfrak{M} \{ \mathbf{x} + \mathbf{y} \} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (t + u) \varphi(t, u) dt du = \\ &= \int_{-\infty}^{\infty} t \varphi_x(t) dt + \int_{-\infty}^{\infty} u \varphi_y(u) du = \mathfrak{M} \{ \mathbf{x} \} + \mathfrak{M} \{ \mathbf{y} \} \end{aligned} \quad (2)$$

для непрерывных распределений.

Среднее значение суммы двух (необязательно независимых) случайных величин \mathbf{x} и \mathbf{y} равно сумме соответствующих средних значений

$$\mathfrak{M} \{ \mathbf{x} + \mathbf{y} \} = \mathfrak{M} \{ \mathbf{x} \} + \mathfrak{M} \{ \mathbf{y} \}. \quad (3)$$

У п р а ж н е н и е. Показать, что величина μ_{xy} , определенная соотношением (5.6.1), равна

$$\mu_{xy} = \mu_{yx} = \mathfrak{M} \{ \mathbf{x} \mathbf{y} \} - \mathfrak{M} \{ \mathbf{x} \} \mathfrak{M} \{ \mathbf{y} \}. \quad (4)$$

Пример. Теперь мы можем доказать неравенство (5.6.3). Рассмотрим функцию

$$F = [(\mathbf{x} - \mu_x) a + (\mathbf{y} - \mu_y) b]^2 \geq 0,$$

где a и b — произвольные действительные числа. Для среднего значения получим

$$\begin{aligned} \mathfrak{M} \{ F \} &= \mathfrak{M} \{ (\mathbf{x} - \mu_x)^2 \} a^2 + 2 \mathfrak{M} \{ (\mathbf{x} - \mu_x) (\mathbf{y} - \mu_y) \} ab + \\ &+ \mathfrak{M} \{ (\mathbf{y} - \mu_y)^2 \} b^2 = \mu_{xx} a^2 + 2 \mu_{xy} ab + \mu_{yy} b^2 \geq 0. \end{aligned}$$

Условием неотрицательности этой однородной квадратичной формы переменных a и b является

$$\mu_{xx}\mu_{yy} - \mu_{xy}^2 \geq 0, \text{ т. е. } \frac{\mu_{xy}^2}{\mu_{xx}\mu_{yy}} = \rho^2 \leq 1, \quad (5)$$

что и требовалось доказать.

§ 6.2. Положив в соотношении (5.5.5) $w = xy$, где величины x и y независимы, получим

$$\begin{aligned} \mathfrak{M}\{xy\} &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} t_i u_j \varphi_{ij} = \\ &= \sum_{i=-\infty}^{\infty} t_i \varphi_i \cdot \sum_{j=-\infty}^{\infty} u_j \varphi_j = \mathfrak{M}\{x\} \mathfrak{M}\{y\} \end{aligned} \quad (1)$$

для дискретных и

$$\begin{aligned} \mathfrak{M}\{xy\} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} tu \varphi(t, u) dt du = \\ &= \int_{-\infty}^{\infty} t \varphi_x(t) dt \int_{-\infty}^{\infty} u \varphi_y(u) du = \mathfrak{M}\{x\} \mathfrak{M}\{y\} \end{aligned} \quad (2)$$

для непрерывных распределений.

Среднее значение произведения двух независимых случайных величин x и y равно произведению соответствующих средних значений сомножителей

$$\mathfrak{M}\{xy\} = \mathfrak{M}\{x\} \mathfrak{M}\{y\}. \quad (3)$$

Заметим, что соотношения, аналогичные (6.1.3) и (3), не имеют места для моды или медианы. По этой причине среднее значение является, как правило, наиболее удобной характеристикой расположения. В силу (6.1.4) для выполнения равенства (3) необходимо и достаточно, чтобы величины x и y были некоррелированы, т. е. $\rho\{x, y\} = 0$. Таким образом, если величины x и y независимы, то они некоррелированы. Обратное, вообще говоря, неверно, т. е. из $\rho\{x, y\} = 0$ не вытекает независимость величин x и y .

Пример 1. Для двумерного распределения, определенного соотношением (4.12.8), получаем (в силу симметрии $\mu_x = \mu_y = 0$)

$$\mu_{xy} = \frac{1}{\pi} \int_{-\infty}^{\infty} t dt \int_{-\infty}^{\infty} \frac{u du}{(1+t^2+u^2)^2} = 0, \quad (4)$$

так как подинтегральная функция в интеграле по u является нечетной. Так как σ_x и σ_y существуют (соответствующие интегралы сходятся), то ρ существует и равно нулю. Однако в этом случае вели-

чины x и y , очевидно, не являются независимыми, так как функция $\varphi(t, u)$ не может быть представлена в виде $\varphi_x(t)\varphi_y(u)$ (см. § 4.13).

В некоторых частных случаях, например, если величина $z = (x, y)$ имеет нормальное распределение, из $\rho = 0$ вытекает независимость величин x и y (см. пример § 12.7). Действительно, в этом случае соотношение (4.12.7) показывает, что при $\rho = 0$ мы имеем $\varphi(t, u) = \varphi_x(t)\varphi_y(u)$, т. е. величины x и y независимы. Так как на практике часто бывает возможно предполагать величину (x, y) нормально распределенной, то величины x и y часто считают независимыми, если только $\rho = 0$. Поэтому коэффициент корреляции играет важную роль во многих статистических исследованиях (см. § 9.5 и § 11.13).

Упражнение 1. Показать, что для произвольных констант a_1, a_2 и $b_1, b_2 > 0$

$$\begin{aligned} \mathbb{M}\{xy\} - \mathbb{M}\{x\}\mathbb{M}\{y\} &= \mathbb{M}\{(x - a_1)(y - a_2)\} - \\ &- \mathbb{M}\{x - a_1\}\mathbb{M}\{y - a_2\} \end{aligned} \quad (5)$$

и

$$\rho \left\{ \frac{x - a_1}{b_1}, \frac{y - a_2}{b_2} \right\} = \rho \{x, y\}. \quad (6)$$

* Пример 2. Из соотношения (3) следует одно важное свойство характеристических функций (пример 2 § 5.2), которое подчеркивает важность этого понятия. Для независимых случайных величин x и y

$$\begin{aligned} \chi_{x+y}(t) &= \mathbb{M}\{e^{i(x+y)t}\} = \mathbb{M}\{e^{ixt}e^{iyt}\} = \mathbb{M}\{e^{ixt}\}\mathbb{M}\{e^{iyt}\} = \\ &= \chi_x(t)\chi_y(t), \end{aligned} \quad (7)$$

т. е. характеристическая функция суммы независимых случайных величин равна произведению соответствующих характеристических функций. Сравнивая это с (4.14.6), мы видим, что значительно проще образовать характеристическую функцию суммы, чем ее функцию распределения. (То же имеет место, очевидно, и для производящей функции, а также и для производящей функции моментов.) Отметим, что логарифм характеристической функции суммы равен сумме логарифмов соответствующих характеристических функций. Это же имеет место для всех коэффициентов разложения логарифма характеристической функции в степенной ряд, т. е. для семи-инвариантов или кумулянтов (см. упражнение 8 § 5.2).

* Упражнение 2. Пусть $\chi_1(t), \dots, \chi_n(t)$ — характеристические функции n независимых случайных величин x_1, \dots, x_n . Показать, что величина $z = a_1x_1 + \dots + a_nx_n$, где a_1, \dots, a_n — произвольные константы, имеет характеристическую функцию (см. упражнение 6 § 5.2)

$$\chi_z(t) = \chi_1(a_1t) \dots \chi_n(a_nt). \quad (8)$$

* Упражнение 3. Для произведения двух независимых случайных величин мы не имеем такого простого выражения для характеристической функции, как для суммы. Однако во многих физических явлениях, как-то:

цепное деление в атомной бомбе, поток электронов в счетчике Гейгер-Мюллера, каскадные ливни в космических лучах, встречаются мультипликативные процессы. В этих явлениях первичная „частица“ (например, нейтрон) вызывает процесс (например, цепное деление), в котором первичная частица исчезает, но вызывает возникновение некоторого числа вторичных „частиц“ (например, нейтронов, возникающих при цепной реакции). Пусть случайная величина x дает число первичных частиц, случайная величина y — число вторичных частиц, возникших от одной первичной частицы, и случайная величина z — общее число возникших частиц, и пусть $\gamma_x(t)$, $\gamma_y(t)$, $\gamma_z(t)$ — соответствующие производящие функции. Показать, что

$$\gamma_z(t) = \gamma_x(\gamma_y(t)). \quad (9)$$

Затем с помощью (5.2.25) показать, что

$$\mathfrak{M}\{z\} = \mathfrak{M}\{x\} \mathfrak{M}\{y\}, \quad (10)$$

$$\sigma^2\{z\} = \sigma^2\{x\} \mathfrak{M}^2\{y\} + \sigma^2\{y\} \mathfrak{M}\{x\}. \quad (11)$$

§ 6.3. В силу предыдущих результатов мы получаем для дисперсии суммы и разности двух случайных величин x и y выражение

$$\sigma^2\{x \pm y\} = \sigma^2\{x\} + \sigma^2\{y\} \pm 2\rho\{x, y\} \sigma\{x\} \sigma\{y\}. \quad (1)$$

Это соотношение мы будем называть „законом сложения дисперсий“.

У п р а ж н е н и е 1. Проверить соотношение (1). Найти соответствующую формулу для $\sigma^2\{ax + by\}$, где a и b — константы.

В частности, если величины x и y независимы, или хотя бы некоррелированы, то в силу (1)

$$\sigma^2\{x \pm y\} = \sigma^2\{x\} + \sigma^2\{y\}. \quad (2)$$

Для двух некоррелированных случайных величин дисперсия их суммы равна сумме их дисперсий. Заметим, что в правой части соотношения (2) стоит знак $+$ независимо от того, какой знак стоит в левой части.

У п р а ж н е н и е 2. Рассмотрим одно бросание двух игральных костей и обозначим через x и y результаты, показанные соответственно первой и второй костями. Тогда в силу (5.1.4) $\mathfrak{M}\{x\} = \mathfrak{M}\{y\} = 7/2$ и в силу (5.3.9) $\sigma^2\{x\} = \sigma^2\{y\} = 35/12$. Найти возможные значения величин $x + y$ и $x - y$ и их вероятности. На основании полученного результата вычислить $\mathfrak{M}\{x + y\}$, $\mathfrak{M}\{x - y\}$ и $\sigma^2\{x + y\}$. Проверить, что эти результаты согласуются с (6.1.3), (6.2.3) и (6.3.2).

§ 6.4. Предыдущие формулы могут быть непосредственно обобщены на случай более чем двух случайных величин. Пусть y случайных величин x_1, x_2, \dots, x_n имеют средние значения $\mu_1, \mu_2, \dots, \mu_n$ и стандартные отклонения $\sigma_1, \sigma_2, \dots, \sigma_n$ и пусть a_1, a_2, \dots, a_n — произвольные постоянные. Тогда для случайной величины

$$z = a_1x_1 + a_2x_2 + \dots + a_nx_n, \quad (1)$$

имеем

$$\mu_z = \mathfrak{M}\{z\} = a_1\mu_1 + a_2\mu_2 + \dots + a_n\mu_n, \quad (2)$$

и общий закон сложения дисперсий

$$\begin{aligned} \sigma_z^2 = \sigma^2 \{z\} = & a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_v^2 \sigma_v^2 + \\ & + 2a_1 a_2 \rho_{12} \sigma_1 \sigma_2 + \dots + 2a_1 a_v \rho_{1v} \sigma_1 \sigma_v + \dots \\ & \dots + 2a_{v-1} a_v \rho_{v-1, v} \sigma_{v-1} \sigma_v, \end{aligned} \quad (3)$$

где $\rho_{ij} = \rho \{x_i, x_j\}$.

Упражнение 1. Проверить соотношение (3).

В частности, если величины x_1, \dots, x_v некоррелированы (это означает, что они попарно некоррелированы, т. е. $\rho_{ij} = 0$ при $i \neq j$), то соотношение (3) дает

$$\sigma_z^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_v^2 \sigma_v^2. \quad (4)$$

Часто встречаются случаи, когда для случайной величины (1)

$$\mu_1 \approx \mu_2 \approx \dots, \sigma_1 \approx \sigma_2 \approx \dots \text{ и } a_1 \approx a_2 \approx \dots$$

Тогда $\mathbb{M}\{z\} \approx \text{const } \nu$ и $\sigma^2\{z\} \approx \text{const } \nu$. Таким образом, для коэффициента изменчивости случайной величины z имеем $\frac{\sigma}{\mu} \approx \frac{\text{const}}{\sqrt{\nu}} \rightarrow 0$. На этом основании отклонениями от среднего значения часто пренебрегают (например, в приложениях к статистической механике, см. § 4.16).

Если положить $a_1 = a_2 = \dots = a_v = \frac{1}{\nu}$, то z становится *средним арифметическим* случайных величин x_1, \dots, x_v и обозначается через \bar{x} :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_v}{\nu}. \quad (5)$$

в этом случае из соотношения (2) вытекает

$$\mathbb{M}\{\bar{x}\} = \frac{\mu_1 + \mu_2 + \dots + \mu_v}{\nu} = \overline{\mathbb{M}\{x\}}. \quad (6)$$

Если величины x_1, \dots, x_v некоррелированы, то из соотношения (4) вытекает, кроме того,

$$\sigma^2\{\bar{x}\} = \frac{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_v^2}{\nu^2} = \frac{1}{\nu} \overline{\sigma^2\{x\}}. \quad (7)$$

Эти формулы особенно интересны, когда величины x_1, x_2, \dots, x_v интерпретируются как ν независимых наблюдений одной и той же случайной величины x со средним значением μ и стандартным отклонением σ . Так как в этом случае $\mu_1 = \mu_2 = \dots = \mu_v = \mu$ и $\sigma_1 = \sigma_2 = \dots = \sigma_v = \sigma$, то соотношения (6) и (7) дают

$$\mathbb{M}\{\bar{x}\} = \mu, \quad (8)$$

$$\sigma^2\{\bar{x}\} = \frac{\sigma^2}{\nu}. \quad (9)$$

Эти формулы очень важны в теории ошибок и в других вопросах (см. раздел 11).

Упражнение 2. Часто образуется не простое среднее арифметическое, а *взвешенное среднее*

$$\bar{x}^{(p)} = \frac{p_1 x_1 + \dots + p_v x_v}{p_1 + \dots + p_v}, \quad (10)$$

где p_1, \dots, p_v — произвольные неотрицательные константы, называемые *весеами*. Найти $\mathfrak{M}\{\bar{x}^{(p)}\}$ и $\sigma^2\{\bar{x}^{(p)}\}$ для общего случая и специального случая $\mu_1 = \mu_2 = \dots = \mu$ и $\sigma_1 = \sigma_2 = \dots = \sigma$.

Пример 1. Приведенные выше формулы часто более удобны для вычисления $\mathfrak{M}\{x\}$ и $\sigma^2\{x\}$, чем определяющие эти величины соотношения, например в случае биномиального распределения (см. примеры 4 § 5.1 и 6 § 5.3). Пусть все величины x_1, \dots, x_v принадлежат к типу, рассмотренному в примере 2 § 5.1, тогда $\mu_1 = \dots = \mu_v = \theta$ в силу (5.1.3) и $\sigma_1 = \dots = \sigma_v = \theta(1 - \theta)$ в силу (5.3.8). Случайная величина x в задаче Бернулли (§ 3.7), имеющая биномиальное распределение, может быть, очевидно, записана в виде

$$x = x_1 + x_2 + \dots + x_v. \quad (11)$$

Поэтому в силу (2)

$$\mathfrak{M}\{x\} = v\theta \quad (12)$$

в соответствии с (5.1.5) и в силу (4)

$$\sigma^2\{x\} = v\theta(1 - \theta) \quad (13)$$

в соответствии с (5.3.10). Таким путем, $\mathfrak{M}\{x\}$ и $\sigma^2\{x\}$ получаются значительно проще, чем непосредственным вычислением. Для относительной частоты $f = \frac{x}{v}$ имеем

$$\mathfrak{M}\{f\} = \theta \quad (14)$$

$$\text{и} \\ \sigma^2\{f\} = \frac{\theta(1 - \theta)}{v}. \quad (15)$$

Так как функция $y = \theta(1 - \theta)$ на интервале $0 \leq \theta \leq 1$ принимает свое максимальное значение $1/4$ в точке $\theta = 1/2$, то

$$\sigma^2\{f\} \leq \frac{1}{4v}. \quad (16)$$

* Пример 2. Часто удобно ввести *матрицу вторых моментов* случайных величин x_1, \dots, x_v , т. е. квадратную симметричную матрицу порядка v (см. приложение 2), определяемую соотношением $M_{rs} = \{m_{rs}\} = \{\mathfrak{M}\{(x_r - \mu_r)(x_s - \mu_s)\}\} = \{\mathfrak{M}\{x_r x_s\} - \mathfrak{M}\{x_r\}\mathfrak{M}\{x_s\}\}$, (17)

т. е.

$$\mu_{ii} = \sigma^2\{x_i\} = \sigma_i^2, \quad \mu_{ij} = \sigma_i \sigma_j \rho_{ij}. \quad (18)$$

Тогда для того, чтобы величины x_1, \dots, x_v были некоррелированы, необходимо и достаточно, чтобы матрица M была диагональной. Иногда вводится также квадратная симметричная *корреляционная матрица* случайных величин x_1, \dots, x_v , определяемая соотношением

$$C = \left\{ \frac{\mu_{rs}}{\sigma_r \sigma_s} \right\} = \{\rho_{rs}\}, \quad (19)$$

где $\rho_{ii} = \frac{\mu_{ii}}{\sigma_i \sigma_i} = 1$. Если Σ — диагональная матрица с диагональными элементами $\sigma_1, \sigma_2, \dots, \sigma_v$, то $C = \Sigma^{-1} M \Sigma^{-1}$ (проверить). Поэтому другим необходимым и достаточным условием некоррелированности величин x_1, \dots, x_v является $C = E$.

* Упражнение 3. Часто бывает полезно проделать линейное преобразование случайных величин x_1, \dots, x_v в другие величины y_1, \dots, y_λ :

$$y_i = f_{i0} + \sum_{j=1}^v f_{ij} x_j, \quad i = 1, 2, \dots, \lambda, \quad (20)$$

где λ не обязательно равно v , а f_{0i} и f_{ij} — константы. Показать, что вторые моменты величин y определяются по вторым моментам величин x соотношениями

$$\mu_{rs}^{(y)} = \sum_{i=1}^v \sum_{j=1}^v f_{ri} f_{sj} \mu_{ij}^{(x)}; \quad r, s = 1, 2, \dots, \lambda. \quad (21)$$

Затем показать, что с помощью матричной символики (см. приложение 2) соотношения (20) и (21) могут быть записаны в удобной форме

$$Y_{\lambda 1} = F_{\lambda 1}^0 + F_{\lambda v} X_{v1}, \quad (22)$$

$$M_{\lambda \lambda}^{(y)} = F_{\lambda v} M_{vv}^{(x)} F_{v \lambda}^*. \quad (23)$$

§ 6.5. В § 6.4 мы вывели простые формулы для вычисления $M\{z\}$ и $\sigma\{z\}$, если z является линейной функцией некоторого числа случайных величин. Однако на практике часто бывает важно провести эти вычисления и тогда, когда z не является линейной функцией. Например, в физике большинство величин измеряются косвенным образом и определяются как нелинейные функции других, непосредственно измеряемых величин. Например, объем сферы равен $V = \frac{\pi}{6} d^3$, где d — диаметр; удельное сопротивление провода равно $\rho = \frac{\pi R d^2}{4l}$, где R — полное сопротивление, l — длина и d — диаметр провода.

В принципе мы можем исследовать и случай, когда $z = f(x_1, \dots, x_v)$ является нелинейной функцией, так как распределения величин x_1, \dots, x_v определяют распределение величины z , а следовательно, и $M\{z\}$ и $\sigma\{z\}$, но, вообще говоря, эти величины не являются простыми функциями средних значений и дисперсий величин x_1, \dots, x_v . Однако часто масса вероятности в совместном распределении сосредоточена в относительно узкой области около точки (μ_1, \dots, μ_v) , и z является

такой мало изменяющейся функцией, что внутри этой области она с хорошим приближением может рассматриваться как линейная функция.

Будем рассматривать только непрерывные распределения, причем сперва ограничимся одномерным случаем $\nu=1$, т. е. $z=f(x)$. В любом случае большая часть массы вероятности случайной величины x будет лежать внутри интервала от $\mu - a\sigma$ до $\mu + a\sigma$ при подходящим образом выбранном значении константы a (см. (8.1.4)). Предположим, что, как это часто бывает на практике, a является небольшим целым числом, скажем $a \approx 1$. Так как вероятность значений величины x , лежащих вне интервала $|x - \mu| \approx \sigma$, предположена, таким образом, малой, то нужно рассмотреть только значения x , лежащие внутри этой области. Разлагая $z=f(x)$ в ряд Тейлора около точки $x=\mu$, получим

$$z=f(x)=f(\mu)+(x-\mu)f'(\mu)+\frac{(x-\mu)^2}{2}f''(\xi), \quad (1)$$

где ξ есть некоторое значение между μ и x . Рассматривая только те значения x , для которых $|x-\mu|<\sigma$, и предполагая, кроме того, что z есть слабо меняющаяся функция в этой области, т. е. что

$$\frac{\sigma^2}{2}|f''(\mu)| \ll \sigma |f'(\mu)| \text{ или } |f''(\mu)| \ll |f'(\mu)|, \quad (2)$$

можно пренебречь членом второго порядка в (1) и написать

$$z \approx f(\mu) + (x - \mu)f'(\mu). \quad (3)$$

Из соотношения (3) получаем

$$\mathfrak{M}\{z\} \approx f(\mathfrak{M}\{x\}) = f(\mu), \quad (4)$$

$$\sigma\{z\} \approx |f'(\mu)| \sigma\{x\}. \quad (5)$$

Упражнение 1. Показать с помощью соотношения (3) и упражнения 3 § 5.3, что распределение величины z приближенно задается соотношениями

$$\Phi_z(u) \approx \Phi_x\left(\frac{u-f(\mu)}{f'(\mu)} + \mu\right), \quad \varphi_z(u) \approx \frac{1}{|f'(\mu)|} \varphi_x\left(\frac{u-f(\mu)}{f'(\mu)} + \mu\right). \quad (6)$$

Упражнение 2. Показать, что для $z = \frac{1}{x}$ имеем

$$\mathfrak{M}\{z\} \approx \frac{1}{\mu}, \quad \sigma\{z\} \approx \frac{\sigma}{\mu^2}, \quad \text{т. е. } \frac{\sigma_z}{\mu_z} \approx \frac{\sigma}{\mu}, \quad (7)$$

и обобщить этот результат на $z = x^k$; $k = 0, \pm 1, \pm 2, \dots$

Пусть $z=f(x_1, \dots, x_n)$ есть функция, мало изменяющаяся внутри той области, в которой сосредоточена большая часть массы вероятности, и z может быть аппроксимирована своей касательной плоскостью. Мы предполагаем, что эта область определяется неравенствами $|x_i - \mu_i| < \sigma_i$, и, обобщая (3), получаем

$$z \approx f(\mu_1, \dots, \mu_n) + \frac{\partial f}{\partial x_1}(x_1 - \mu_1) + \dots + \frac{\partial f}{\partial x_n}(x_n - \mu_n), \quad (8)$$

где частные производные берутся в точке (μ_1, \dots, μ_n) . В предположении некоррелированности величин x_1, \dots, x_n соотношение (8) дает

$$\mathbb{M}\{z\} \approx f(\mu_1, \dots, \mu_n), \quad (9)$$

$$\sigma^2\{z\} \approx \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_1^2 + \dots + \left(\frac{\partial f}{\partial x_n}\right)^2 \sigma_n^2. \quad (10)$$

У п р а ж н е н и е 3. Выписать соотношение, соответствующее (10), в случае, когда некоррелированность величин x_1, \dots, x_n не предполагается.

7. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

§ 7.1. Как в теории, так и на практике наиболее важную роль играет *нормальное распределение*, которое, как уже было сказано в примере 3 § 4.4, является непрерывным распределением с плотностью вероятности

$$\frac{d\Phi}{dt} = \varphi(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] = \left(\frac{h}{\sqrt{\pi}} \exp\left[-h^2(t-\mu)^2\right]\right). \quad (1)$$

В § 7.2 мы покажем, что соотношение (1) уже нормировано, т. е.

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] dt = 1. \quad (2)$$

График функции $\varphi(t)$ называется *гауссовской кривой ошибок*, а параметр

$$h = \frac{1}{\sqrt{2}\sigma} \quad (3)$$

— *мерой точности* (так как чем больше h , тем круче кривая $\varphi(t)$ падает к нулю). Легко видеть, что $\varphi(\mu+t) = \varphi(\mu-t)$, т. е. функция $\varphi(t)$ симметрична относительно прямой $t = \mu$. Так как

$$\varphi'(t) = \frac{-1}{\sqrt{2\pi}\sigma^3} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] (t-\mu) \quad (4)$$

и

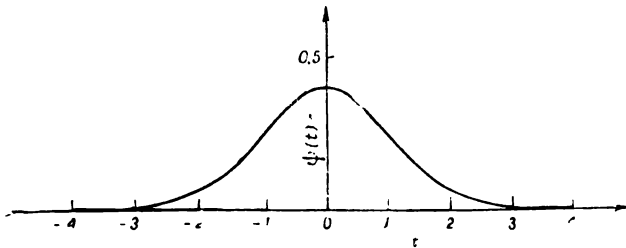
$$\varphi''(t) = \frac{1}{\sqrt{2\pi}\sigma^3} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] \left(\frac{(t-\mu)^2}{\sigma^2} - 1\right), \quad (5)$$

то кривая $\varphi(t)$ имеет один и только один максимум при $t = \mu$ и две точки перегиба при $t = \mu \pm \sigma$.

Если, в частности, $\mu = 0$ и $\sigma = 1$, то распределение называется *нормированным* (см. упражнения 1 и 2 § 5.3). В этом случае плотность вероятности обозначается так:

$$\frac{d\Psi}{dt} = \phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}; \quad (6)$$

ее график дан на фиг. 5, а ее значения в табл. 1.



Фиг. 5.

Нормальная функция распределения получается из (1)

$$\Phi(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^t \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] dt \quad (7)$$

и имеет в силу (4) одну точку перегиба при $t = \mu$. Так как $\varphi(t)$ симметрична относительно прямой $t = \mu$, то в силу (2)

$$\int_{-\infty}^{\mu+t} \varphi(t) dt + \int_{\mu+t}^{\infty} \varphi(t) dt = \int_{-\infty}^{\mu+t} + \int_{-\infty}^{\mu-t} = \Phi(\mu+t) + \Phi(\mu-t) = 1, \quad (8)$$

т. е., в частности,

$$\Phi(\mu) = \frac{1}{2}. \quad (9)$$

Таким образом, функция $\Phi(t)$ симметрична относительно точки $(\mu, \frac{1}{2})$ ¹⁾. Нормированная функция распределения, получаемая при $\mu = 0$ и $\sigma = 1$, обозначается так:

$$\Psi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{t^2}{2}} dt. \quad (10)$$

Ее график дан на фиг. 6, а ее значения — в табл. I в конце книги. Подчеркнем, что при $t \rightarrow \pm\infty$ функции $\varphi(t)$ и $\Psi(t)$ очень быстро стремятся к своим предельным значениям.

Часто табулируется не функция $\Psi(t)$, а так называемый *интеграл ошибок* (называемый также *функцией ошибок* или *интегралом вероятностей*)

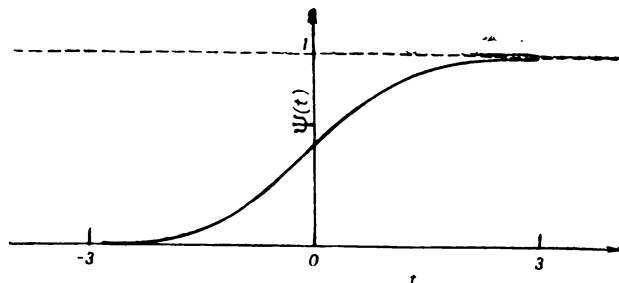
$$\theta(t) = \operatorname{erf} t = \frac{2}{\sqrt{\pi}} \int_0^t e^{-t^2} dt = 2\Psi(\sqrt{2}t) - 1. \quad (11)$$

¹⁾ Заметим, что это имеет место для любого распределения, симметричного относительно $t = \mu$, например для распределения Коши (4.4.9).

У п р а ж н е н и е 1. Проверить соотношение (11).

У п р а ж н е н и е 2. Показать, что если величина x нормирована и нормально распределена, то $|x|$ имеет функцию распределения, равную

$\Theta\left(\frac{t}{\sqrt{2}}\right)$ при $t \geq 0$ и равную нулю при $t < 0$.



Ф и г. 6.

У п р а ж н е н и е 3. Показать, что для произвольного нормального распределения

$$\Phi(t) = \Psi\left(\frac{t - \mu}{\sigma}\right) \text{ и } \varphi(t) = \frac{1}{\sigma} \phi\left(\frac{t - \mu}{\sigma}\right) \quad (12)$$

(см. упражнение 3 § 5.3). Таким образом, графики функций $\varphi(t)$ и $\Phi(t)$ легко получаются из графиков, данных на фиг. 5 и 6.

У п р а ж н е н и е 4. Показать, что случайная величина $y = ax + b$ распределена нормально, если величина x распределена нормально, а $a \neq 0$ и b — произвольные константы.

У п р а ж н е н и е 5. Показать, что если величина x распределена нормально, то произвольная функция $y = f(x)$ распределена приблизительно нормально с параметрами $\mu_y = f(\mu_x)$ и $\sigma_y = |f'(\mu_x)| \sigma_x$, если только f удовлетворяет условию $\sigma |f''(\mu_x)| \ll |f'(\mu_x)|$ (см. § 6.5; разложить $f(x)$ с помощью теоремы Тейлора).

§ 7.2. Докажем теперь соотношение (7.1.2). Интеграл, стоящий в левой части этого соотношения, не может быть вычислен непосредственно, однако с помощью теории двойных и плоских интегралов его вычисление сводится к определению объема некоторого тела. С этой целью мы образуем квадрат выражения, стоящего в левой части соотношения (7.1.2) и представляющего собой неотрицательное число. Полагая в первом множителе $\frac{t - \mu}{\sigma} = x$ и во втором

$\frac{t - \mu}{\sigma} = y$, найдем, что

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

Таким образом, нам нужно определить объем тела, лежащего между плоскостью x, y и поверхностью $z = \frac{1}{2\pi} \exp \left[-\frac{x^2 + y^2}{2} \right] = \frac{1}{2\pi} \exp \left[-\frac{r^2}{2} \right]$, где r есть расстояние между точками $(0, 0)$ и (x, y) . Этот объем мы находим благодаря его симметричности относительно оси z , разбивая его на „бесконечно тонкие“ цилиндры, основаниями которых являются кольца с радиусами r и $r + dr$, имеющие, следовательно, площадь $2\pi r dr$, а высоты равны $\frac{1}{2\pi} \exp \left[-\frac{r^2}{2} \right]$, т. е. объем каждого такого цилиндра равен $r \exp \left[-\frac{r^2}{2} \right] dr$. Поэтому полный объем равен

$$\int_0^{\infty} r \exp \left[-\frac{r^2}{2} \right] dr = \int_0^{\infty} e^{-u} du = 1, \quad (2)$$

где мы положили $\frac{r^2}{2} = u$ в качестве [новой переменной]. Этим наше доказательство завершено.

§ 7.3. Среднее значение нормально распределенной случайной величины x получим из (7.1.1)

$$\begin{aligned} \mathfrak{M} \{x\} &= \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} t \exp \left[-\frac{(t-\mu)^2}{2\sigma^2} \right] dt = \\ &= \int_{-\infty}^{\infty} ((t-\mu) + \mu) \varphi(t) dt = 0 + \mu \cdot 1 = \mu \end{aligned} \quad (1)$$

в силу (7.1.2) и симметричности $\varphi(t)$ относительно $t = \mu$. Дисперсию этой величины получим с помощью интегрирования по частям, полагая $\frac{t-\mu}{\sigma} = u$ в качестве новой переменной,

$$\begin{aligned} \sigma^2 \{x\} &= \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} (t-\mu)^2 \exp \left[-\frac{(t-\mu)^2}{2\sigma^2} \right] dt = \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u d \left(-\exp \left[-\frac{u^2}{2} \right] \right) = \sigma^2 \left(-\frac{u \exp \left[-\frac{u^2}{2} \right]}{\sqrt{2\pi}} \right) \Big|_{-\infty}^{\infty} + \\ &+ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[-\frac{u^2}{2} \right] du = \sigma^2 (0 + 1) = \sigma^2 \end{aligned} \quad (2)$$

в силу (7.1.2) и того, что $u \exp \left[-\frac{u^2}{2} \right] \rightarrow 0$ при $u \rightarrow \pm \infty$.

Таким образом, параметры μ и σ равны соответственно среднему значению и стандартному отклонению. Это и является основанием для введения множителя 2 в выражении (7.1.1) под знаком экспоненциала.

§ 7.4. Вероятность того, что величина x принимает значение в интервале между t_1 и t_2 , равна

$$P(t_1 \leq x \leq t_2) = \frac{1}{\sqrt{2\pi}\sigma} \int_{t_1}^{t_2} \exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right] dt. \quad (1)$$

Эта формула особенно интересна, когда $t_1 = \mu - a\sigma$, $t_2 = \mu + a\sigma$, где a — произвольная положительная константа. Полагая в этом случае $\frac{t-\mu}{\sigma} = u$ в качестве новой переменной, получим из (1)

$$\begin{aligned} P(|x - \mu| \leq a\sigma) &= \frac{1}{\sqrt{2\pi}} \int_{-a}^a \exp\left[-\frac{u^2}{2}\right] du = \\ &= \int_{-\infty}^a - \int_{-\infty}^{-a} = 2\Phi(a) - 1 = \theta\left(\frac{a}{\sqrt{2}}\right), \end{aligned} \quad (2)$$

где мы использовали (7.1.8) и (7.1.11). Эта функция табулирована как функция от a в табл. I в конце книги. В частности, из этой таблицы находим, что

$$P(|x - \mu| \leq \sigma) = 0,6827 \approx 2/3. \quad (3)$$

Вероятность того, что x отклоняется от μ меньше чем на σ , равна приблизительно $2/3$.

Обратно, из соотношения (2) можно определить значения a , соответствующие заданным значениям P , которые мы будем называть *доверительными пределами* случайной величины x , т. е. пределами, в которых величина x лежит с заданной вероятностью P (предполагая этот интервал симметричным относительно μ). Для целей практики, однако, более удобно искать a , соответствующие заданной вероятности дополнительного события, а именно $P(|x - \mu| \geq a\sigma)$. В табл. II мы даем $a = a(P)$. Наиболее важными значениями являются

P	a	
0,001	3,29	(4)
0,01	2,58	
0,05	1,96	

которые называются соответственно 0,1, 1 и 5% пределами.

Упражнение 1. Предположим, что при стрельбе из ружья условия примера 1 § 4. 12 выполнены. Найти вероятность поражения вертикальной квадратной цели с центром в точке $(x, y) = (\mu_x, \mu_y)$, называемой в теории стрельбы центром рассеивания, и сторонами $2a = 2\sigma_x = 2\sigma_y$.

Упражнение 2. Пусть имеют место условия, указанные в предыдущем упражнении. Расстояние r от центра рассеивания до точки попадания является тогда случайной величиной. Показать, что при $\sigma_x = \sigma_y = \sigma$ распределение этой величины задается соотношением

$$\frac{d\Phi}{dt} = \frac{1}{\sigma^2} t \exp \left[-\frac{t^2}{2\sigma^2} \right]. \quad (5)$$

Найти вероятность попадания в круг с центром в центре рассеивания и радиусом $R = \sigma$.

Упражнение 3. В упражнении § 4.16 мы показали, что для консервативной системы из N частиц три компоненты скорости v_x, v_y, v_z каждой частицы взаимно независимы и что все они нормально распределены с параметрами $\mu_x = \mu_y = \mu_z = 0$ и $\sigma_x = \sigma_y = \sigma_z = \sigma = \sqrt{\frac{kT}{m}}$. Показать, что случай-

ная величина $v = \sqrt{v_x^2 + v_y^2 + v_z^2}$ имеет распределение Максвелла — Больцмана (4. 8. 2) с параметрами

$$\alpha = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma^3} = \sqrt{\frac{2}{\pi}} \left(\frac{m}{kT} \right)^{3/2}, \quad \beta = \frac{1}{2\sigma^2} = \frac{m}{2kT}.$$

Пример 1. *Вероятным отклонением* ρ называют 50% предел $\alpha = \alpha \left(\frac{1}{2} \right)$, определяемый соотношением (2), т. е. ρ определяется соотношением

$$P(|x - \mu| \geq \rho) = P(|x - \mu| \leq \rho) = \frac{1}{2}. \quad (6)$$

Из соотношения (2) и табл. II получим

$$\Psi \left(\frac{\rho}{\sigma} \right) = \frac{3}{4}, \quad \text{т. е. } \rho = 0,67449 \sigma \approx \frac{2}{3} \sigma. \quad (7)$$

Интервал, симметричный относительно μ и такой, что вероятность величине x попасть в этот интервал, так же как и вероятность попасть вне этого интервала, равна $\frac{1}{2}$, определяется приблизительно точками $\mu \pm \frac{2}{3} \sigma$.

В старой литературе ρ часто использовалось вместо σ в качестве дисперсионного параметра нормального распределения. В современной литературе параметр ρ употребляется все реже и реже.

Пример 2. Иногда в старой литературе в качестве дисперсионного параметра использовалось *среднее отклонение* $\theta = \mathfrak{M} \{ |x - \mu| \}$.

Из соотношения (7. 1. 1) получим, вводя новое переменное $\frac{1}{2} \left(\frac{t-\mu}{\sigma} \right)^2 = u$,

$$\begin{aligned} 0 &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} |t-\mu| \exp \left[-\frac{(t-\mu)^2}{2\sigma^2} \right] dt = \\ &= \frac{2\sigma}{\sqrt{2\pi}} \int_0^{\infty} e^{-u} du = \sqrt{\frac{2}{\pi}} \sigma = 0,79788\sigma \approx \frac{4}{5} \sigma. \end{aligned} \quad (8)$$

Пример 3. Часто употребляемой в физике характеристикой рассеяния является половина полушироты γ . Из $\varphi(t) = \frac{1}{2} \varphi(\mu)$ мы получим $\frac{(t-\mu)^2}{2\sigma^2} = \ln 2$, т. е.

$$\gamma = \frac{1}{2} (t_2 - t_1) = \sqrt{2 \ln 2} \sigma = 1,1774 \sigma. \quad (9)$$

§ 7.5. Нормальное распределение имеет несколько важных свойств. Мы приведем здесь некоторые из них (см. также § 8.2 и § 8.3).

Сумма двух независимых случайных величин x и y , распределенных нормально соответственно с параметрами μ_x, σ_x и μ_y, σ_y , распределена нормально с параметрами $\mu_{x+y} = \mu_x + \mu_y$ и $\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2$.

Последняя часть этой теоремы следует из (6.1.3) и (6.3.2) и того, что параметры представляют соответственно среднее значение и дисперсию. Первая часть теоремы следует из (4.14.6). Подставляя в эту формулу нормальные плотности вероятности вместо φ_x и φ_y , получим

$$\varphi_{x+y}(s) = \frac{1}{2\pi\sigma_x\sigma_y} \int_{-\infty}^{\infty} \exp \left[-\frac{(r-\mu_x)^2}{2\sigma_x^2} - \frac{(s-r-\mu_y)^2}{2\sigma_y^2} \right] dr. \quad (1)$$

Используя выражения для μ_{x+y} и σ_{x+y} , можно записать выражение, стоящее под знаком экспоненциала в формуле (1), в виде

$$\begin{aligned} &\frac{(r-\mu_x)^2}{2\sigma_x^2} + \frac{(s-r-\mu_y)^2}{2\sigma_y^2} = \\ &= \frac{\sigma_{x+y}^2}{2\sigma_x^2\sigma_y^2} \left(r - \frac{\sigma_y^2\mu_x + \sigma_x^2(s-\mu_y)}{\sigma_{x+y}^2} \right)^2 + \frac{(s-\mu_{x+y})^2}{2\sigma_{x+y}^2}. \end{aligned}$$

Вводя последнее выражение в соотношение (1), полагая в качестве новой переменной $t = \frac{\sigma_{x+y}}{\sigma_x\sigma_y} \left(r - \frac{\sigma_y^2\mu_x + \sigma_x^2(s-\mu_y)}{\sigma_{x+y}^2} \right)$ и используя соотно-

шение (7.1.2), получим

$$\begin{aligned}\varphi_{x+y}(s) &= \frac{1}{2\pi\sigma_{x+y}} \int_{-\infty}^{\infty} \exp \left[-\frac{t^2}{2} - \frac{(s - \mu_{x+y})^2}{2\sigma_{x+y}^2} \right] dt = \\ &= \frac{1}{\sqrt{2\pi}\sigma_{x+y}} \exp \left[-\frac{(s - \mu_{x+y})^2}{2\sigma_{x+y}^2} \right],\end{aligned}\quad (2)$$

что и доказывает наше утверждение.

Пример 1. Следует отметить, что это свойство суммы иметь распределение того же вида, что и распределения слагаемых, не является присущим одному только нормальному распределению. Например, мы уже доказали выполнение этого свойства для биномиального распределения (упражнение 4 § 4.11) и распределения Пуассона (упражнение 5 § 4.11). Оно может быть доказано также для распределения Коши (см. задачу 40) и так называемого χ^2 -распределения (см. § 7.8, упражнение 2).

* **Пример 2.** Из выражений для характеристических функций биномиального распределения, распределения Пуассона, нормального распределения и распределения Коши (упражнение 7 § 5.2) и теоремы, сформулированной в примере 2 § 6.2, сразу следует, что если две независимые случайные величины имеют одно и то же из перечисленных распределений, то их сумма также имеет это распределение. Это еще раз подчеркивает полезность понятия характеристической функции.

Наша теорема может быть, очевидно, обобщена следующим образом: *если ν случайных величин x_1, \dots, x_ν распределены нормально и a_1, \dots, a_ν — произвольные константы, то величина*

$$z = a_1 x_1 + \dots + a_\nu x_\nu, \quad (3)$$

также распределена нормально с параметрами, определяемыми соотношениями (6.4.2) и (6.4.4).

Упражнение 1. Проверить последнее утверждение.

Упражнение 2. Доказать следующее обобщение результата упражнения 5 § 7.1. Если независимые случайные величины x_1, \dots, x_ν распределены нормально соответственно с параметрами $\mu_1, \sigma_1; \dots; \mu_\nu, \sigma_\nu$, то произвольная функция $y = f(x_1, \dots, x_\nu)$ распределена приближенно нормально с параметрами

$$\mu_y = f(\mu_1, \dots, \mu_\nu) \quad (4)$$

и

$$\sigma_y^2 = \left(\frac{\partial f(\mu_1, \dots, \mu_\nu)}{\partial x_1} \right)^2 \sigma_1^2 + \dots + \left(\frac{\partial f(\mu_1, \dots, \mu_\nu)}{\partial x_\nu} \right)^2 \sigma_\nu^2, \quad (5)$$

если только f мало изменяется в области $|x_i - \mu_i| < \sigma_i$, в которой сосредоточена основная масса вероятности ν -мерной случайной величины (x_1, \dots, x_ν) (см. § 6.5; разложить f с помощью теоремы Тейлора).

Если, в частности, в соотношении (3) положить $a_1 = a_2 = \dots = a_n = \frac{1}{\nu}$, т. е. $\mathbf{z} = \bar{\mathbf{x}}$, и интерпретировать x_1, \dots, x_n как ν независимых наблюдений одной и той же случайной величины x , то мы получим теорему: *величина \bar{x} распределена нормально с параметрами μ и $\frac{\sigma}{\sqrt{\nu}}$.*

Пример 3. Отметим без доказательства, что имеет место следующая теорема¹⁾:

Если сумма двух независимых случайных величин x и y распределена нормально, то и x и y сами распределены нормально.

§ 7.6. Двумерное нормальное распределение определяется соотношением (4.12.7), т. е.

$$d\Phi = \varphi(t, u) dt du = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left[-\frac{1}{1-\rho^2} \left\{ \frac{(t-\mu_x)^2}{2\sigma_x^2} - \rho \frac{(t-\mu_x)(u-\mu_y)}{\sigma_x\sigma_y} + \frac{(u-\mu_y)^2}{2\sigma_y^2} \right\} \right] dt du. \quad (1)$$

Как показано в примере § 4.13, соотношение (1) нормировано, т. е.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(t, u) dt du = \int_{-\infty}^{\infty} \varphi_x(t) dt = \int_{-\infty}^{\infty} \varphi_y(u) du = 1. \quad (2)$$

В примере 1 § 5.5 и примере § 5.6 мы уже показали, что параметры μ_x , μ_y и σ_x , σ_y выбраны так, что они обозначают соответственно средние значения и стандартные отклонения. Наконец, можно показать, что параметр ρ выбран так, что он обозначает коэффициент корреляции, определенный соотношением (5.6.2).

* Пример 1. Для того чтобы показать это, возьмем в качестве новых переменных $x = \frac{t-\mu_x}{\sigma_x}$ и $y = \frac{u-\mu_y}{\sigma_y}$ и получим

$$\begin{aligned} \mu_{xy} &= \mathfrak{M} \{ (x - \mu_x)(y - \mu_y) \} = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (t - \mu_x)(u - \mu_y) \varphi(t, u) dt du = \frac{\sigma_x \sigma_y}{2\pi \sqrt{1-\rho^2}} \times \\ &\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \exp \left[-\frac{1}{2(1-\rho^2)} (x^2 - 2\rho xy + y^2) \right] dx dy. \quad (3) \end{aligned}$$

¹⁾ См. Г. Крамер, Случайные величины и распределения вероятностей, стр. 66, 1947.

Для того чтобы вычислить этот довольно сложный двойной интеграл, введем новые переменные v и w , такие, что

$$x = \frac{v-w}{\sqrt{2}}, \quad y = \frac{v+w}{\sqrt{2}}, \quad \text{т. е. } xy = \frac{v^2-w^2}{2},$$

$$x^2 - 2\rho xy + y^2 = v^2(1-\rho) + w^2(1+\rho). \quad (4)$$

Так как в нашем случае якобиан (4.14.2) равен

$$\frac{\partial(x, y)}{\partial(v, w)} = \begin{vmatrix} \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{vmatrix} = 1 \quad (5)$$

и $1-\rho^2 = (1-\rho)(1+\rho)$, то из (4.14.3) получим

$$\begin{aligned} \mu_{xy} &= \frac{\sigma_x \sigma_y}{4\pi \sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (v^2 - w^2) \exp \left[-\frac{1}{2} \left(\frac{v^2}{1+\rho} + \frac{w^2}{1-\rho} \right) \right] dv dw = \\ &= \frac{\sigma_x \sigma_y}{2} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v^2 \exp \left[-\frac{1}{2} \left(\frac{v^2}{1+\rho} + \frac{w^2}{1-\rho} \right) \right] \frac{dv}{\sqrt{1+\rho}} \frac{dw}{\sqrt{1-\rho}} - \right. \\ &\quad \left. - \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w^2 \exp \left[-\frac{1}{2} \left(\frac{v^2}{1+\rho} + \frac{w^2}{1-\rho} \right) \right] \frac{dv}{\sqrt{1+\rho}} \frac{dw}{\sqrt{1-\rho}} \right\} = \\ &= \frac{\sigma_x \sigma_y}{2} \{(1+\rho) - (1-\rho)\} = \sigma_x \sigma_y \rho, \end{aligned} \quad (6)$$

что доказывает наше утверждение. (При этом мы просто замечаем, что два интеграла, стоящие в фигурных скобках в соотношении (6), дают стандартные отклонения двумерной случайной величины, распределенной нормально с параметрами $(\mu_v, \mu_w, \sigma_v, \sigma_w, \rho_{vw}) = (0, 0, \sqrt{1+\rho}, \sqrt{1-\rho}, 0)$.

* Упражнение 1. (См. приложение 2.) Заметим, что выражение, стоящее под знаком экспоненциала в соотношении (1), является однородной квадратической формой переменных $(t_1 - \mu_1) = (t - \mu_x)$ и $(t_2 - \mu_2) = (t - \mu_y)$:

$$-\frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^2 a_{ij} (t_i - \mu_i) (t_j - \mu_j).$$

Сравнение с (1) показывает, что квадратная симметричная матрица $A = \{a_{rs}\}$ этой формы определяется равенством

$$A = \frac{1}{1 - \rho^2} \left\{ \begin{array}{cc} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x \sigma_y} \\ -\frac{\rho}{\sigma_x \sigma_y} & \frac{1}{\sigma_y^2} \end{array} \right\}. \quad (7)$$

Найти детерминант $|A| \neq 0$ и показать, что Φ в соотношении (1) может быть записано в виде

$$\begin{aligned} d\Phi &= \varphi(t, u) dt du = \frac{\sqrt{|A|}}{(\sqrt{2\pi})^2} \exp \left[-\frac{1}{2} \sum_{i=1}^2 \sum_{j=1}^2 a_{ij} (t_i - \mu_i) (t_j - \mu_j) \right] dt_i dt_j = \\ &= \frac{\sqrt{|A|}}{(\sqrt{2\pi})^2} \exp \left[-\frac{1}{2} (T - \mathfrak{M})^* \cdot A \cdot (T - \mathfrak{M}) \right] dt_1 dt_2, \end{aligned} \quad (8)$$

где T_{21} и \mathfrak{M}_{21} являются матрицами-столбцами соответственно с элементами t_1, t_2 и μ_1, μ_2 . Наконец, показать, что

$$A^{-1} = M, \quad (9)$$

где M есть матрица вторых моментов, определенная соотношением (6.4.17).

* Пример 2. (См. приложение 2.) В качестве очевидного обобщения соотношения (8) ν -мерная нормальная случайная величина (x_1, \dots, x_ν) определяется как величина, имеющая распределение

$$\begin{aligned} d\Phi &= \varphi(t_1, \dots, t_\nu) dt_1 \dots dt_\nu = \\ &= \frac{\sqrt{|A|}}{(\sqrt{2\pi})^\nu} \exp \left[-\frac{1}{2} \sum_{i=1}^\nu \sum_{j=1}^\nu a_{ij} (t_i - \mu_i) (t_j - \mu_j) \right] dt_1 \dots dt_\nu = \\ &= \frac{\sqrt{|A|}}{(\sqrt{2\pi})^\nu} \exp \left[-\frac{1}{2} (T - \mathfrak{M})^* \cdot A \cdot (T - \mathfrak{M}) \right] dt_1 \dots dt_\nu, \end{aligned} \quad (10)$$

где A есть ν -мерная квадратная симметричная матрица, такая, что квадратичная форма под знаком экспоненциала неотрицательна. Действительно, можно показать, что

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(t_1, \dots, t_\nu) dt_1 \dots dt_\nu = 1, \quad (11)$$

$$\mathfrak{M} \{(x_1, \dots, x_\nu)\} = (\mu_1, \dots, \mu_\nu), \quad (12)$$

$$M = \{\mathfrak{M} \{(x_r - \mu_r)(x_s - \mu_s)\}\} = A^{-1}. \quad (13)$$

(Заметим, что соотношение (13) показывает, что если величины x_1, \dots, x_ν попарно некоррелированы, то они и независимы.) Для того чтобы доказать это, нужно только воспользоваться тем хорошо известным фактом, что симметричная матрица может быть преобразована в диагональную с помощью ортогонального преобразования, т. е. можно

ввести у новых координат y_1, \dots, y_ν , с помощью ортогонального преобразования (см. упражнение 3 § 6.4)

$$X_{\nu,1} - \mathfrak{M}_{\nu,1} = F_{\nu\nu} Y_{\nu,1}, \quad (14)$$

где матрица F ортогональна, т. е.

$$F \cdot F^* = E, \text{ откуда } |F| = \pm 1 \text{ и } F^* = F^{-1}, \quad (15)$$

и определена так, что

$$F^* A F = F^{-1} A F = D, \text{ откуда } |D| = |F^{-1}| |F| |A| = |A|. \quad (16)$$

Здесь D есть диагональная матрица, диагональные элементы которой мы будем обозначать $\frac{1}{d_1^2}, \dots, \frac{1}{d_\nu^2}$.

Так как в силу (14) якобиан (4.14.2) равен $\frac{\partial(t_1, \dots, t_\nu)}{\partial(u_1, \dots, u_\nu)} = |F| = \pm 1$ и в силу (16) $V|A| = V|D| = \frac{1}{d_1} \dots \frac{1}{d_\nu}$, то из (4.14.3) и (7.1.2) получим

$$\begin{aligned} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(t_1, \dots, t_\nu) dt_1 \dots dt_\nu &= \\ &= \frac{V|D|}{(V2\pi)^\nu} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} U^* \cdot D \cdot U \right] du_1 \dots du_\nu = \\ &= \prod_{i=1}^{\nu} \left(\frac{1}{V2\pi d_i} \int_{-\infty}^{\infty} \exp \left[-\frac{u_i^2}{2d_i^2} \right] du_i \right) = 1, \end{aligned} \quad (17)$$

что доказывает (11). Чтобы доказать (12), заметим, что из (17) следует, что

$$\begin{aligned} \mathfrak{M}\{y_r\} &= \frac{1}{(V2\pi)^\nu d_1 \dots d_\nu} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} u_r \times \\ &\times \exp \left[-\frac{1}{2} \sum_{i=1}^{\nu} \frac{u_i^2}{d_i^2} \right] du_1 \dots du_\nu = 0; \quad r = 1, 2, \dots, \nu, \end{aligned} \quad (18)$$

т. е. $\mathfrak{M}\{Y\} = 0$. Из соотношений (18) и (14) сразу следует результат (12). Далее, соотношение (17) показывает, что $M^{(\nu)}$ есть диагональная матрица с диагональными элементами d_1^2, \dots, d_ν^2 , т. е. в силу (15) и (16)

$$M^{(\nu)} = D^{-1} = F^{-1} A^{-1} F. \quad (19)$$

Из соотношений (19), (14), (15) и (6.4.23), где Y и X переставлены местами, следует сразу результат (13), так как

$$M^{(X)} = FM^{(Y)}F^* = FF^{-1}A^{-1}FF^{-1} = A^{-1}. \quad (20)$$

* Упражнение 2. (См. приложение 2.) Показать, что характеристической функцией распределения (10) является

$$\begin{aligned} \lambda_{x_1, \dots, x_\nu}(t_1, \dots, t_\nu) &= \exp \left[i \sum_{k=1}^{\nu} \mu_k t_k - \frac{1}{2} \sum_{k=1}^{\nu} \sum_{l=1}^{\nu} (A^{-1})_{kl} t_k t_l \right] = \\ &= \exp \left[i M^* T - \frac{1}{2} T^* A^{-1} T \right] \end{aligned} \quad (21)$$

(см. пример 3 § 5.5). Показать, что из выражения (21) для характеристической функции следуют соотношения (11) — (13) (см. (5.2.27)). Из выражения (21) следует далее, что частное распределение величины x_k (при $k = 1, 2, \dots, \nu$) нормально с параметрами μ_k, σ_k (см. упражнение 1 § 5.5). Используя понятие характеристической функции, мы, таким образом, избегаем необходимости проводить непосредственное вычисление сложных интегралов для доказательства этого результата.

За более подробными сведениями относительно многомерного нормального распределения мы отсылаем читателя к учебникам статистики, например к книге Крамера.

* § 7.7. В этом и четырех следующих параграфах мы рассмотрим некоторые распределения, связанные с нормальным распределением, которые очень важны для приложений к практике (см. §§ 11.8 — 11.17). Для дальнейших целей удобно несколько обобщить понятие нормального распределения*) (см. пример 2 § 10.3). Пусть x есть случайная величина и $y = G(x)$ — функция, для которой $-\infty < y < \infty$, $a \leq x \leq b$ ($-\infty \leq a, b \leq \infty$) и $G'(x) \geq 0$ для всех x в области определения. Предположим, что величина y распределена нормально, тогда (см. § 4.8) величина x имеет распределение¹⁾

$$d\Phi_x = \varphi(y) dx = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} (G(x) - \mu)^2 \right] \left| \frac{dG(x)}{dx} \right| dx. \quad (1)$$

Если $G(x) = x$, то распределение (1) сводится к обычному нормальному распределению.

Упражнение 1. Показать, что

$$M\{G(x)\} = \mu, \quad \sigma^2\{G(x)\} = \sigma^2. \quad (2)$$

*) Даваемое авторами обобщение нормального распределения является совершенно излишним для содержания §§ 7.7—7.11, в которых идет речь об известных в статистике распределениях χ^2 , Стьюдента и т. п. (Прим. ред.)

¹⁾ Для удобства мы будем в § 7.7—7.11 обозначать случайную величину и переменную функции распределения одной и той же буквой.

Рассмотрим затем n независимых величин x_1, \dots, x_n , имеющих одно и то же распределение (1) (в дальнейшем x_1, \dots, x_n будут обозначать n независимых наблюдений одной и той же величины x). Совместное распределение величин x_1, \dots, x_n определяется тогда соотношением

$$d\Phi = \varphi(x_1, \dots, x_n) dx_1 \dots dx_n = \\ = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (G(x_i) - \mu)^2 \right] \prod_{i=1}^n |G'(x_i)| dx_1 \dots dx_n. \quad (3)$$

Вместо x_1, \dots, x_n введем n новых случайных величин, определяемых соотношениями

$$m = \overline{G(x)} = \frac{G(x_1) + \dots + G(x_n)}{n} \quad (-\infty < m < \infty), \quad (4)$$

$$v_i = G(x_i) - m, \quad (5)$$

$$q = \sqrt{\sum_{i=1}^n v_i^2} \quad (0 \leq q < \infty),$$

$$U_i = \frac{G(x_i) - m}{q} \quad (-1 < u_i < 1), \quad i = 1, 2, \dots, n. \quad (6)$$

У п р а ж н е н и е 2. Показать, что

$$\mathfrak{M} \{m\} = \mu, \quad \sigma \{m\} = \frac{\sigma}{\sqrt{n}} \quad (7)$$

и

$$\sum_{i=1}^n u_i = 0, \quad \sum_{i=1}^n u_i^2 = 1 \quad (8)$$

и что отсюда вытекает, что $|u_i| < 1$, т. е. случай $|u_i| = 1$ исключается. В силу двух алгебраических соотношений (8) только $n-2$ переменные u являются свободными. Будем рассматривать как свободные переменные u_1, u_2, \dots, u_{n-2} . Тогда вместе с m и q мы ввели всего n новых переменных. Показать, что для якобиана (4.14.2) имеем

$$\frac{\partial(x_1, \dots, x_n)}{\partial(m, q, u_1, \dots, u_{n-2})} = q^{n-2} f(u_1, u_2, \dots, u_{n-2}; n) \frac{1}{\prod_{i=1}^n |G'(x_i)|}, \quad (9)$$

где f есть некоторая функция от одних u , которую нет необходимости выписывать в явном виде. Наконец, показать, что

$$\sum_{i=1}^n (G(x_i) - \mu)^2 = n(m - \mu)^2 + q^2. \quad (10)$$

В силу результатов упражнения 2 и соотношения (4.14.3) наше распределение запишется в новых переменных в виде

$$d\Phi = \left\{ \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp \left[-\frac{n(m-\mu)^2}{2\sigma^2} \right] dm \right\} \times \\ \times \left\{ \frac{1}{\left(\frac{n-3}{2}\right)! 2^{\frac{n-3}{2}}} \left(\frac{q}{\sigma}\right)^{n-2} \exp \left[-\frac{q^2}{2\sigma^2} \right] \frac{dq}{\sigma} \right\} \times \\ \times \{ \text{const } f(u_1, \dots, u_{n-2}; n) du_1, \dots, du_{n-2} \}, \quad (11)$$

где постоянные коэффициенты в фигурных скобках выбраны так, чтобы каждая скобка в отдельности оказалась нормированной (проверить с помощью приложения 1). Соотношение (11) показывает, что m и q являются независимыми случайными величинами, что величина m распределена нормально с параметрами μ и $\frac{\sigma}{\sqrt{n}}$, а q имеет распределение

$$d\Phi = \varphi(q) dq = \frac{1}{\left(\frac{f-2}{2}\right)! 2^{\frac{f-2}{2}}} \left(\frac{q}{\sigma}\right)^{f-1} \exp \left[-\frac{q^2}{2\sigma^2} \right] \frac{dq}{\sigma} \quad (12) \\ (0 \leq q < \infty), \quad f = n - 1,$$

найденное впервые Гельмэртом. В выражении (12) мы вместо n ввели так называемое число *степеней свободы* $f = n - 1$. Это объясняется тем, что q^2 есть сумма квадратов n величин v_1, \dots, v_n , подчиненных одной связи, определяемой соотношением (8), так что только $n - 1$ из величин v являются свободными переменными.

Упражнение 3. Показать с помощью приложения 1, что

$$\mathfrak{M}\{q^2\} = f\sigma^2, \quad \mathfrak{M}\{q^4\} = (f+2)f\sigma^4, \quad \text{т. е. } \sigma^2\{q^2\} = 2f\sigma^4. \quad (13)$$

Показать затем, что

$$\mathfrak{M}\{q\} = \frac{\left(\frac{f-1}{2}\right)!}{\left(\frac{f-2}{2}\right)!} \sqrt{2}\sigma \sim \sqrt{f - \frac{1}{2}} \sigma + O\left(\frac{1}{f^2}\right), \quad (14)$$

т. е.

$$\sigma^2\{q\} = \left(f - 2 \left(\frac{\left(\frac{f-1}{2}\right)!}{\left(\frac{f-2}{2}\right)!} \right)^2 \right) \sigma^2 \sim \frac{\sigma^2}{2} \quad (15)$$

(для доказательства приближенного равенства в выражении (15) использовать формулу Стирлинга)*).

У п р а ж н е н и е 4. Показать, что частными случаями распределения (12) величины q являются: 1) при $f=1$ распределение величины $|x|$, где величина x распределена нормально с параметрами $0, \sigma$; 2) при $f=2$ распределение (7.4.5); 3) при $f=3$ распределение Максвелла — Больцмана (4.8.2) (см. упражнение 3 § 7.4).

У п р а ж н е н и е 5. Вместо величины q часто удобно рассматривать величину

$$s = \frac{q}{\sqrt{f}}, \quad f = n - 1. \quad (16)$$

Показать, что

$$\mathfrak{M}\{s^2\} = \sigma^2, \quad \sigma^2\{s^2\} = \frac{2\sigma^4}{f} \quad (17)$$

и

$$\mathfrak{M}\{s\} = \frac{\left(\frac{f-1}{2}\right)!}{\left(\frac{f-2}{2}\right)!} \sqrt{\frac{2}{f}} \sigma \sim \sqrt{1 - \frac{1}{2f}} \sigma \quad (18)$$

$$\sigma^2\{s\} = \left(1 - \frac{2}{f} \left(\frac{\left(\frac{f-1}{2}\right)!}{\left(\frac{f-2}{2}\right)!}\right)^2\right) \sigma^2 \sim \frac{\sigma^2}{2f}. \quad (19)$$

* § 7.8. Во многих приложениях удобнее иметь дело с распределением величины q^2 , а не с распределением величины q . Более того, удобно рассматривать величину $\frac{q^2}{\sigma^2}$, обозначаемую через χ^2)

$$\chi^2 = \frac{q^2}{\sigma^2}. \quad (1)$$

У п р а ж н е н и е 1. Показать, что величина χ^2 имеет распределение

$$d\Phi = \varphi(\chi^2) d(\chi^2) = \frac{1}{\left(\frac{f-2}{2}\right)! 2^{\frac{f-2}{2}}} (\chi^2)^{\frac{f-2}{2}} \exp\left[-\frac{1}{2}\chi^2\right] d(\chi^2) \quad (0 \leq \chi^2 < \infty) \quad (2)$$

и

$$\mathfrak{M}\{\chi^2\} = f, \quad \mathfrak{M}\{\chi^4\} = (f+2)f, \quad \text{т. е. } \sigma^2\{\chi^2\} = 2f. \quad (3)$$

*) Для вывода асимптотического выражения (14) формула Стирлинга, приведенная в приложении I, должна быть уточнена добавлением еще одного члена

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left[1 + \frac{1}{12n} + O\left(\frac{1}{n^2}\right)\right].$$

(Прим. ред.)

1) Хотя применение греческой буквы для обозначения случайной величины, а не параметра, противоречит нашему основному принципу обозначений и хотя обозначение χ^2 менее удобно, чем χ , мы не решаемся применять другой символ, так как обозначение χ^2 , повидимому, общепринято.

Это распределение табулировано. Можно показать, что для больших значений f величина $V\sqrt{2\chi^2}$ распределена приблизительно нормально со средним значением $V2f-1$ и дисперсией 1. Распределение χ^2 обладает тем важным свойством, что если χ_1^2 и χ_2^2 являются независимыми величинами, имеющими распределение (2) соответственно со степенями свободы f_1 и f_2 , то величина $\chi^2 = \chi_1^2 + \chi_2^2$ имеет распределение (2), причем $f = f_1 + f_2$.

У п р а ж н е н и е 2. Проверить это. (Ввести в совместное распределение величин χ_1^2 и χ_2^2 новые переменные χ^2 и p , определяемые соотношениями $\chi_1^2 = p\chi^2$, $\chi_2^2 = (1-p)\chi^2$, и проинтегрировать по p от 0 до 1 (см. (4.14.6).)

У п р а ж н е н и е 3. Показать, что если величины y_1, \dots, y_n независимы и распределены нормально с параметрами $\mu = 0$ и $\sigma = 1$, то величина

$$\chi^2 = y_1^2 + y_2^2 + \dots + y_n^2 \quad (4)$$

имеет распределение (2) с числом степеней свободы $f = n$. Затем показать, что при предположениях § 7.7 величина

$$\chi_0^2 = \frac{q_0^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (G(x_i) - \mu)^2 \quad (5)$$

имеет то же самое распределение, что и величина

$$\chi_1^2 = \frac{q^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (G(x_i) - m)^2, \quad (6)$$

с той только разницей, что для первой величины $f = n$, а для второй величины $f = n - 1$.

*** § 7.9.** Пусть случайные величины y и q независимы, величина y распределена нормально, причем $M\{y\} = 0$ и $\sigma\{y\} = a\sigma$ (где a — константа), а величина q имеет распределение (7.7.12) с числом степеней свободы f . Нормированной величиной, соответствующей величине y , является $\frac{y}{a\sigma}$ (см. упражнение 2 § 5.3). Заменим теперь σ величиной s , определяемой соотношением (7.6.16), и обозначим результат через t :

$$t = \frac{y}{as} = \frac{V\bar{f}}{a} \frac{y}{q}. \quad (1)$$

Чтобы получить распределение величины t , выпишем совместное распределение величин y и q . Затем введем новые переменные t и q вместо y и q и получим из (4.14.3)

$$d\Phi = \varphi(t, q) dt dq = \frac{1}{V2\pi V\bar{f} \left(\frac{f-2}{2}\right)! 2^{\frac{f-2}{2}}} \left(\frac{q}{\sigma}\right)^f \times \\ \times \exp \left[- \left(1 + \frac{t^2}{f}\right) \frac{q^2}{2\sigma^2} \right] dt \frac{dq}{\sigma}. \quad (2)$$

{проверить). Интегрируя (2) по всем значениям q , мы получим частное распределение величины t (для этого надо ввести $\left(1 + \frac{t^2}{f}\right)^{\frac{1}{2}}$ в качестве новой переменной интегрирования)

$$d\Phi = \varphi(t) dt = \frac{\left(\frac{f-1}{2}\right)! dt}{\sqrt{\pi} \sqrt{f} \left(\frac{f-2}{2}\right)! \left(1 + \frac{t^2}{f}\right)^{\frac{f+1}{2}}} \quad (-\infty < t < \infty) \quad (3)$$

(проверить этот результат с помощью приложения 1). Это распределение называется *распределением Стьюдента*. Подчеркнем, что оно не зависит от параметров μ и σ , а зависит только от числа степеней свободы f . Из предельного соотношения

$$\left(1 + \frac{t^2}{f}\right)^{-\frac{f+1}{2}} \xrightarrow{f \rightarrow \infty} \exp\left[-\frac{t^2}{2}\right]$$

нетрудно видеть, что при больших f распределение величины t будет весьма близко к нормальному распределению (7.1.6).

У п р а ж н е н и е 1. Проверить это с помощью формулы Стирлинга.

Вероятность $P(t)$ того, что $|t| \geq t$, равна

$$P(t) = 2 \int_t^{\infty} \varphi_t(t) dt. \quad (4)$$

В табл. III в конце книги приведены значения t в зависимости от f при $P=10, 5, 1$ и $0,1\%$. Можно убедиться, что при $f \rightarrow \infty$ мы получим соответствующие доверительные пределы для нормального распределения (см. табл. II).

У п р а ж н е н и е 2. Пусть величина m определена соотношением (7.7.4), величина q — соотношением (7.7.5) и величина s — соотношением (7.7.16). Показать, что величина

$$t = \frac{m - \mu}{\frac{s}{\sqrt{n}}} = \sqrt{\frac{n(n-1)}{n}} \frac{m - \mu}{q} \quad (5)$$

имеет распределение Стьюдента (3) с $f = n - 1$. Пусть затем случайная величина x независима от m и q и имеет распределение (7.7.1). Показать, что величина

$$t = \frac{m - G(x)}{\sqrt{1 + \frac{1}{n}} s} = \sqrt{\frac{n(n-1)}{n+1}} \frac{m - G(x)}{q} \quad (6)$$

также имеет распределение Стьюдента (3) с $f = n - 1$.

У п р а ж н е н и е 3. Пусть $x_{11}, x_{12}, \dots, x_{1n_1}$ и $x_{21}, x_{22}, \dots, x_{2n_2}$ — два взаимно независимых множества независимых наблюдений величины x , имею-

в качестве новой переменной интегрирования),

$$d\Phi = \varphi(w) dw = \frac{2 \left(\frac{f-2}{2}\right)!}{\left(\frac{f_1-2}{2}\right)! \left(\frac{f_2-2}{2}\right)!} \left(\frac{f_1}{f_2}\right)^{\frac{f_1}{2}} \frac{w^{f_1-1}}{\left(1 + \frac{f_1}{f_2} w^2\right)^{\frac{f}{2}}} dw \quad (3)$$

(проверить этот результат с помощью приложения 1).

Часто вместо w удобнее рассматривать величину w^2 .

У п р а ж н е н и е 1. Показать, что распределение величины w^2 определяется соотношением

$$d\Phi = \varphi(w^2) d(w^2) = \frac{\left(\frac{f-2}{2}\right)!}{\left(\frac{f_1-2}{2}\right)! \left(\frac{f_2-2}{2}\right)!} \left(\frac{f_1}{f_2}\right)^{\frac{f_1}{2}} \frac{(w^2)^{\frac{f_1-2}{2}}}{\left(1 + \frac{f_1}{f_2} w^2\right)^{\frac{f}{2}}} d(w^2), \quad (4)$$

$$f = f_1 + f_2 \quad (0 \leq w^2 < \infty).$$

Подчеркнем, что это распределение, так же как и предыдущие, не зависит от параметров μ и σ . Вероятность $P(w^2)$ того, что $w^2 \geq w^2$, равна

$$P(w^2) = \int_{w^2}^{\infty} \varphi_{w^2}(w^2) d(w^2). \quad (5)$$

В табл. V в конце книги приведены значения w^2 в зависимости от f_1 и f_2 при $P = 5\%$. Здесь предполагается, что $s_1 \geq s_2$, т. е. $w^2 \geq 1$. Заметим, что w^2 -распределение содержит в качестве частных случаев χ^2 -распределение, t -распределение Стьюдента, а следовательно, также и r -распределение. При $w = \chi/\sqrt{f_1}$ и $f_2 \rightarrow \infty$ распределение (4) сводится к (7.8.2) с $f = f_1$. При $w = t$ и $f_1 = 1$ распределение (4) сводится к (7.9.3) с $f = f_2$.

У п р а ж н е н и е 2. Проверить эти утверждения с помощью таблиц.

8. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ

§ 8.1. Пусть $u = f(t)$ есть неотрицательная функция, для которой

$$u = f(t) \geq a > 0 \quad (1)$$

(где a — некоторая константа) для всех значений t внутри некоторой области ω . Для случайной величины x с непрерывным распределением имеем тогда в силу (5.2.1)

$$\mathfrak{M}\{f(x)\} = \int_{-\infty}^{\infty} f(t) \varphi(t) dt \geq a \int_{\omega} \varphi(t) dt = aP(x \in \omega), \quad (2)$$

т. е.

$$P(x \in \omega) \leq \frac{\mathfrak{M}\{f(x)\}}{a}. \quad (3)$$

Упражнение 1. Показать, что (3) имеет место и для дискретных распределений.

Если в соотношении (3) положить $f(x) = (x - \mathfrak{M}\{x\})^2$ и $\sqrt{a} = k$, то в силу (5.3.1) получим *неравенство Чебышева*

$$P(|x - \mathfrak{M}\{x\}| \geq k) \leq \frac{\sigma^2\{x\}}{k^2}. \quad (4)$$

Упражнение 2. Найти соответствующее неравенство для $f(x) = |x - \mathfrak{M}\{x\}|^n$ и $\sqrt[n]{a} = k$.

Если мы имеем последовательность случайных величин x_1, x_2, \dots , средние значения μ_1, μ_2, \dots и дисперсии $\sigma_1^2, \sigma_2^2, \dots$ которых существуют, то в силу (4), если

$$\mathfrak{M}\{(x_v - \mu_v)^2\} = \sigma_v^2 \rightarrow 0, \quad (5)$$

$v \rightarrow \infty$

для произвольного, но фиксированного $\varepsilon > 0$

$$P(|x_v - \mu_v| \geq \varepsilon) \leq \frac{\sigma_v^2}{\varepsilon^2} \rightarrow 0. \quad (6)$$

$v \rightarrow \infty$

В этом случае мы говорим, что последовательность случайных величин $x_v - \mu_v$ *сходится по вероятности* к 0¹⁾. Вообще, если суще-

1) Или, как говорят, *сходится стохастически*.

щей распределение (7.7.1). Пусть величины m_1 и m_2 соответствуют двум этим множествам согласно определению, даваемому соотношением (7.7.4), и $s_1 = \frac{q_1}{\sqrt{f_1}}$ и $s_2 = \frac{q_2}{\sqrt{f_2}}$ — соответствующие величины, определенные соотношением (7.7.16). Показать, что

$$\mathfrak{M}\{m_1 - m_2\} = 0, \quad \sigma\{m_1 - m_2\} = \sigma \sqrt{\frac{1}{n} + \frac{1}{n_2}}. \quad (7)$$

Затем показать, что „нормированная“ величина

$$t = \frac{m_1 - m_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad s = \frac{q}{\sqrt{f}} = \frac{\sqrt{q_1^2 + q_2^2}}{\sqrt{f_1 + f_2}} \quad \begin{matrix} f_1 = n_1 - 1 \\ f_2 = n_2 - 1 \end{matrix} \quad (8)$$

имеет распределение Стьюдента (3) с $f = f_1 + f_2$.

* § 7.10. Рассмотрим „относительное отклонение“ величины $G(x_i)$, заменяя μ на m и σ на $s = \frac{q}{\sqrt{n-1}}$,

$$r_i = \frac{v_i}{\sqrt{\frac{n-1}{n}} s} = \sqrt{n} \frac{G(x_i) - m}{q} = \sqrt{n} u_i, \quad i = 1, 2, \dots, n, \quad (1)$$

где m , v_i , u_i , q и s определены в § 7.7. Из соотношения (7.7.10) можно вывести, что частное распределение любой величины r_i определяется соотношением¹⁾

$$d\Phi = \varphi(r) dr = \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{f+1}} \frac{\left(\frac{f-1}{2}\right)!}{\left(\frac{f-2}{2}\right)!} \left(1 - \frac{r^2}{f+1}\right)^{\frac{f-2}{2}} dr, \\ |r| \leq \sqrt{f+1}, \quad f = n - 2. \quad (2)$$

Упражнение 1. Показать, что величины r связаны двумя соотношениями

$$\sum_{i=1}^n r_i = 0, \quad \sum_{i=1}^n r_i^2 = n, \quad (3)$$

т. е. имеют на одну связь больше, чем величины v_i . Поэтому мы и положили здесь $f = n - 2$.

Распределение (2) может быть получено из распределения Стьюдента, как это утверждается в следующем упражнении.

¹⁾ См. Arley, *Danske Vid. Selsk. Mat.-fys. Medd.*, т. XVIII, № 3, 1940. Там дано совместное распределение r_1, \dots, r_{n-2} и более подробная, чем здесь, таблица r -распределения.

Упражнение 2. Показать, что величина

$$t' = \frac{\sqrt{n-2} r_1}{\sqrt{n-1-r_1^2}} = \frac{\sigma_1}{\sqrt{\frac{1}{n-2} \sum_{i=2}^n \sigma_i^2}} \quad (4)$$

имеет распределение Стьюдента (7.9.3) с $f=n-2$. Показать далее, что если r_1 возрастает от $-\sqrt{n-1}$ до $\sqrt{n-1}$, то t' возрастает от $-\infty$ до ∞ , таким образом, $P(r_1 \leq r_1) = P\left(t' \leq \frac{\sqrt{n-2} r_1}{\sqrt{n-1-r_1^2}}\right)$. Наконец, подставляя распределение величины t' и дифференцируя, получить (2).

Заметим, что распределение величины r подобно распределению величины t не зависит от параметров μ и σ и весьма близко к нормальному распределению для больших значений f . Вероятность $P(r)$ того, что $|r| \geq r$, равна

$$P(r) = 2 \int_r^{\sqrt{f+1}} \varphi_r(r) dr. \quad (5)$$

В табл. IV в конце книги приведены значения r в зависимости от f при $P=10, 5, 1$ и $0,10/0$. Можно убедиться, что при $f \rightarrow \infty$ мы получим как раз соответствующие доверительные пределы для нормального распределения (см. табл. II в конце книги).

* § 7.11. Рассмотрим, наконец, две независимые случайные величины q_1 и q_2 , имеющие q -распределение (7.7.12) соответственно со степенями свободы f_1 и f_2 . Вводя в их совместное распределение вместо q_1 величину w , определенную соотношением

$$w = \frac{s_1}{s_2} = \sqrt{\frac{f_2}{f_1}} \frac{q_1}{q_2} \quad (0 \leq w < \infty), \quad (1)$$

получим из (4.14.3)

$$\frac{1}{\left(\frac{f_1-2}{2}\right)! \left(\frac{f_2-2}{2}\right)! 2^{\frac{f-4}{2}}} \left(\frac{f_1}{f_2}\right)^{\frac{f_1}{2}} w^{f_1-1} \left(\frac{q_2}{\sigma}\right)^{f-1} \times \\ \times \exp\left[-\left(1 + \frac{f_1}{f_2} w^2\right) \frac{q_2^2}{2\sigma^2}\right] d w \frac{d q_2}{\sigma}, \quad f = f_1 + f_2. \quad (2)$$

Интегрируя по q_2 от 0 до ∞ , получим частное распределение величины w (для этого надо ввести

$$\left(1 + \frac{f_1}{f_2} w^2\right)^{\frac{1}{2}} \frac{q_2}{\sqrt{2}\sigma}$$

ствует такая случайная величина x , что

$$P(|x_v - x| \geq \varepsilon) \xrightarrow{v \rightarrow \infty} 0, \quad (7)$$

то мы говорим, что x_v сходится по вероятности к x . В современной теории вероятностей это понятие играет важную роль. Подчеркнем, что сходимость по вероятности не нужно путать с обычной сходимостью $x_v \rightarrow x$, где x_v и x представляют соответственно наблюдаемые значения случайных величин x_v и x . Действительно, очевидно, что мы не можем ничего доказать математически относительно поведения эмпирических значений. Доказаны могут быть только утверждения, относящиеся к теоретическим понятиям нашей модели, а не к реальности (см. § 1.1 и § 9.2). Поэтому для „сходимости по вероятности“ вводится специальный символ

$$x_v \xrightarrow[\text{по вер.}]{v \rightarrow \infty} x \quad (8)$$

или, более употребительное обозначение,

$$x_v \xrightarrow[\text{по вер.}]{v \rightarrow \infty} x. \quad (9)$$

* У п р а ж н е н и е 3. Доказать, что для сходимости по вероятности необходимо и достаточно, чтобы

$$\Phi_{x_v - x}(t) \xrightarrow{v \rightarrow \infty} \varepsilon(t) \quad (10)$$

(где $\varepsilon(t)$ определено соотношением (4.3.3)) для всех фиксированных $t \neq 0$. Доказать, что для сходимости по вероятности необходимо, чтобы

$$\Phi_{x_v}(t) \xrightarrow{v \rightarrow \infty} \Phi_x(t) \quad (11)$$

для любого фиксированного значения t , являющегося точкой непрерывности функции $\Phi(t)$. (Если $\Phi_x(t) = \varepsilon(x - \mu)$, то условие (11) является также достаточным.)

* У п р а ж н е н и е 4. Пусть случайная величина x со средним значением μ и дисперсией σ^2 такова, что

$$M\{(x_v - x)^2\} = \sigma^2 \{x_v - x\} \xrightarrow{v \rightarrow \infty} 0. \quad (12)$$

(В этом случае мы говорим, что x_v сходится в среднем к x .) Показать, что

$$\text{тогда} \quad x_v \xrightarrow[\text{по вер.}]{v \rightarrow \infty} x.$$

Пример 1. Легко видеть, что может иметь место $x_v \xrightarrow[\text{по вер.}]{v \rightarrow \infty} x$, хотя ни одно из μ_1, μ_2, \dots и $\sigma_1, \sigma_2, \dots$ не существует. Рассмотрим, например, последовательность случайных величин x_1, x_2, \dots , каждая из которых имеет распределение Коши с параметрами $\mu_1 = \mu_2 = \dots = 0$ и $\alpha_v = 1/v$, т. е. $\alpha_v \xrightarrow{v \rightarrow \infty} 0$. В этом случае определенные интегралы для

средних значений не существуют и $\sigma_1 = \sigma_2 = \dots = \infty$, однако, как показано в примере 4 § 4.4, $\Phi_\nu(t) \xrightarrow{\nu \rightarrow \infty} \varepsilon(t)$ для всех значений $t \neq 0$,

откуда в силу упражнения 3 следует, что $x_\nu \xrightarrow[\nu \rightarrow \infty]{\text{по вер.}} x$, где случайная величина x имеет распределение $\varepsilon(t)$, определенное соотношением (4.3.3), т. е. x есть константа, равная нулю. (Кроме того, мы видим, что $\sigma^2 \{x_\nu - x\} = \mathbb{M} \{ (x_\nu - x)^2 \} = \mathbb{M} \{ x_\nu^2 \} = \infty$, т. е. x_ν не сходится в среднем к пределу x . Таким образом, сходимость по вероятности является более общим понятием и поэтому более широко применимо.)

В частности, положим в соотношении (6) $x_\nu = \frac{r}{\nu} = f$, где r есть абсолютная, а f — относительная частота среди ν независимых наблюдений некоторого события с вероятностью наступления в каждом отдельном наблюдении, равной θ . Как показано в § 3.7, величина r имеет биномиальное распределение, и в силу (6.4.14) и (6.4.15) соотношение (6) дает для произвольного, но фиксированного $\varepsilon > 0$

$$P(|f - \theta| \geq \varepsilon) \xrightarrow{\nu \rightarrow \infty} 0, \quad \text{т. е. } f \xrightarrow[\nu \rightarrow \infty]{\text{по вер.}} \theta. \quad (13)$$

Последнее соотношение выражает *теорему Бернулли*:

Относительная частота f среди ν независимых наблюдений некоторого события сходится по вероятности при $\nu \rightarrow \infty$ к вероятности этого события θ , т. е. вероятность того, что f отклоняется от θ больше чем на ε , становится произвольно малой при неограниченном возрастании ν .

Подставим теперь вместо x_ν в соотношении (6) $\bar{x} = \frac{1}{\nu}(x_1 + \dots + x_\nu)$, где x_1, \dots, x_ν суть ν независимых наблюдений случайной величины x со средним значением μ и дисперсией σ^2 . Тогда в силу (6.4.8) и (6.4.9) из соотношения (6) вытекает, что для произвольного, но фиксированного $\varepsilon > 0$ мы имеем

$$P(|\bar{x} - \mu| \geq \varepsilon) \xrightarrow{\nu \rightarrow \infty} 0, \quad \text{т. е. } \bar{x} \xrightarrow[\nu \rightarrow \infty]{\text{по вер.}} \mu. \quad (14)$$

Последнее соотношение выражает *закон больших чисел*¹⁾:

Среднее арифметическое \bar{x} из ν независимых наблюдений случайной величины x с конечными μ и σ сходится по вероятности к μ при $\nu \rightarrow \infty$, т. е. вероятность отклонения \bar{x} от μ больше чем на ε становится произвольно малой при неограниченном возрастании ν .

Пример 2. При сравнении теоремы Бернулли или закона больших чисел с опытом следует иметь в виду, что величины f и \bar{x} яв-

¹⁾ См. примечание на стр. 15.

ляются случайными, и представлять себе всю серию из ν наблюдений, дающую только одно, хотя и ν -мерное, наблюдение величины f или \bar{x} . При этом, как было сказано в разделе 1, мы должны рассматривать большое число n таких „единичных“ наблюдений величины f или \bar{x} , и каждое из этих n наблюдений состоит само из ν наблюдений, дающих, однако, только одно численное значение f величины f или значение \bar{x} величины \bar{x} . Относительное число $\frac{n'}{n}$ тех из n значений величины f или величины \bar{x} , для которых $|f - \theta| \geq \varepsilon$ или $|\bar{x} - \mu| \geq \varepsilon$ будет тем ближе к теоретическому значению — вероятности, чем больше значение n при фиксированном значении ν . (То есть если ν само велико, то это теоретическое значение практически равно 0, как это показывают соотношения (13) и (14).) Для того чтобы подчеркнуть различный смысл величин n и ν , мы обозначили их разными буквами.

Важно знать, что для того, чтобы закон больших чисел имел место, существенно, чтобы среднее значение μ существовало в строгом смысле, т. е. сумма или интеграл, определяющие μ , сходились абсолютно. В этом можно убедиться на следующем примере.

Пример 3. Пусть величины x_1, \dots, x_ν имеют распределения Коши с одинаковыми параметрами μ и α (см. (4.4.9)). В этом случае интеграл, определяющий среднее значение, не сходится абсолютно (см. пример 9 § 5.1). Как показано в задаче 40 или примере 2 § 7.5, величина $x_1 + \dots + x_\nu$ также имеет распределение Коши, но уже с параметрами $\nu\mu$ и $\nu\alpha$. Отсюда непосредственно следует, что величина \bar{x} имеет распределение Коши с теми же параметрами, что и исходные величины x_1, \dots, x_ν , т. е. μ и α (проверить). Следовательно, вероятность

$$P(|\bar{x} - \mu| \geq \varepsilon) = \frac{2}{\pi\alpha} \int_{\mu+\varepsilon}^{\infty} \frac{dt}{1 + \frac{(t-\mu)^2}{\alpha^2}} \quad (15)$$

не зависит от ν и поэтому \bar{x} не сходится по вероятности к μ .

Можно было бы думать, что условие $\sigma < \infty$ также является необходимым для закона больших чисел. Однако это не так, и закон больших чисел может быть доказан при одном только предположении, что μ существует в строгом смысле (теорема Хинчина).

* Пример 4. Последнее утверждение легко доказать с помощью характеристических функций. Из определения $\chi_x(t)$ следует, что если μ существует, то при $t \rightarrow 0$ мы имеем $\chi_x(t) = 1 + \mu it + o(t)$, где $\frac{o(t)}{t} \rightarrow 0$. Тогда в силу (6.2.8) для любого фиксированного t

$$\chi_{\bar{x}}(t) = \left(\chi_x\left(\frac{t}{\nu}\right) \right)^\nu = \left(1 + \mu i \frac{t}{\nu} + o\left(\frac{t}{\nu}\right) \right)^\nu \xrightarrow{\nu \rightarrow \infty} e^{i\mu t}, \quad (16)$$

где справа стоит характеристическая функция распределения $\epsilon(t - \mu)$ (см. (5.2.16)). В силу теоремы сходимости из § 5.2 и результата упражнения 3 имеем $\overset{\text{по вер.}}{x} \xrightarrow{\nu \rightarrow \infty} \mu$, что и требовалось доказать.

§ 8.2. Одно из важных свойств нормального распределения состоит в том, что оно является пределом многих других распределений в некоторых предельных процессах (см. § 7.8, 7.9, 7.10 и 10.10). Здесь мы покажем это только для одного специального, но важного случая, а именно для случая биномиального распределения (см. пример 2 § 4.3). Можно показать, что для больших значений ν

$$\varphi_i = \binom{i}{\nu} \theta^i (1 - \theta)^{\nu - i} = \frac{1}{\sqrt{2\pi\nu\theta(1-\theta)}} \exp \left[-\frac{(i - \nu\theta)^2}{2\nu\theta(1-\theta)} \right] + \frac{R_\nu}{\nu}, \quad (1)$$

где R_ν по абсолютной величине меньше некоторого числа, не зависящего от ν . Эта формула называется иногда *формулой Лапласа*, хотя на самом деле она была выведена Бернулли. Она имеет место в предположении, что $\sigma\{x\} = \sqrt{\nu\theta(1-\theta)} \gg 1$, и для таких значений i , для которых нормированная величина

$$l = \frac{i - \nu\theta}{\sqrt{\nu\theta(1-\theta)}} \quad (2)$$

ограничена. Из соотношения (1) следует, что вероятность того, что $l_1 \leq l \leq l_2$ (где l_1 и l_2 не зависят от ν), равна

$$P(l_1 \leq l \leq l_2) = \frac{1}{\sqrt{2\pi\nu\theta(1-\theta)}} \sum \exp \left[-\frac{l^2}{2} \right] + \sum \frac{R_\nu}{\nu}, \quad (3)$$

где суммирование распространяется на те значения i , для которых $l_1 \leq l \leq l_2$. Число таких значений i в любом случае меньше, чем $(l_2 - l_1) \sqrt{\nu\theta(1-\theta)} + 1$ (второй член появляется оттого, что обе конечные точки l_1 и l_2 могут соответствовать целым значениям i). Таким образом,

$$\sum \frac{R_\nu}{\nu} \xrightarrow{\nu \rightarrow \infty} 0.$$

Первый член правой части равенства (3) дает сумму значений функции $\exp \left[-\frac{l^2}{2} \right]$, для которых в силу (2) расстояние между двумя последовательными значениями l равно $\frac{1}{\sqrt{\nu\theta(1-\theta)}}$.

Из определения определенного интеграла следует, что

$$P(l_1 \leq l \leq l_2) \xrightarrow{\nu \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{l_1}^{l_2} e^{-\frac{t^2}{2}} dt. \quad (4)$$

Это соотношение выражает *теорему Муазра*. Итак, для больших значений ν величина l , а следовательно, и величины x и $f = \frac{x}{\nu}$ распределены приблизительно нормально.

Упражнение 1. Показать, что в последнем результате содержится теорема Бернулли (8.1.13).

* Пример. Формулу Лапласа можно доказать с помощью формулы Стирлинга (см. приложение 1). Подставляя значения, даваемые этой формулой для $\nu!$, $i!$ и $(\nu - i)!$, в левую часть соотношения (1), найдем после некоторых преобразований

$$\begin{aligned} & \binom{\nu}{i} \theta^i (1 - \theta)^{\nu - i} = \\ &= \frac{1}{\sqrt{2\pi\nu\theta(1-\theta)}} \left(\frac{\nu\theta}{i}\right)^{i+\frac{1}{2}} \left(\frac{\nu(1-\theta)}{\nu-i}\right)^{\nu-i+\frac{1}{2}} \exp \left[\frac{1}{12} \left(\frac{\theta_1}{\nu} - \frac{\theta_2}{i} - \frac{\theta_3}{\nu-i} \right) \right] = \\ &= \frac{1}{\sqrt{2\pi\nu\theta(1-\theta)}} e^{-z}, \end{aligned} \quad (5)$$

где z , выраженное через l (см. (2)), равно

$$\begin{aligned} z = & \left(i + \frac{1}{2}\right) \ln \left(\frac{i}{\nu\theta}\right) + \left(\nu - i + \frac{1}{2}\right) \ln \left(\frac{\nu - i}{\nu(1-\theta)}\right) - \frac{1}{12} \left(\frac{\theta_1}{\nu} - \frac{\theta_2}{i} - \frac{\theta_3}{\nu-i}\right) = \\ & \left(\nu\theta + l\sqrt{\nu\theta(1-\theta)} + \frac{1}{2}\right) \ln \left(1 + \frac{l(1-\theta)}{\sqrt{\nu\theta(1-\theta)}}\right) + \\ & \left(\nu(1-\theta) - l\sqrt{\nu\theta(1-\theta)} + \frac{1}{2}\right) \ln \left(1 - \frac{l\theta}{\sqrt{\nu\theta(1-\theta)}}\right) - \\ & - \frac{1}{12} \left(\frac{\theta_1}{\nu} - \frac{\theta_2}{i} - \frac{\theta_3}{\nu-i}\right) \end{aligned} \quad (6)$$

(проверить). Для достаточно больших значений ν и l , лежащего между пределами l_1 и l_2 , не зависящими от ν , $\frac{l(1-\theta)}{\sqrt{\nu\theta(1-\theta)}}$ и $\frac{l\theta}{\sqrt{\nu\theta(1-\theta)}}$ по абсолютной величине меньше, чем $1/4$. В этом случае в силу теоремы Тейлора $\ln(1+x) = x - \frac{x^2}{2} + \theta_4 x^3$, $|\theta_4| < 1$. Используя это в равенстве (6), получим после некоторых преобразований

$$z = \frac{l^2}{2} + \frac{r_\nu}{\sqrt{\nu}}, \quad (7)$$

где r_ν есть число, зависящее от ν , θ и l , но меньшее по абсолютной величине некоторого числа, не зависящего от ν и l . Отсюда без труда получается соотношение (1).

* Упражнение 2. Показать с помощью интегрирования по частям что имеют место следующие точные выражения:

$$\sum_{i=r}^{\nu} \binom{i}{\nu} \theta^i (1-\theta)^{\nu-i} = \frac{\nu!}{(r-1)! (\nu-r)!} \int_0^{\theta} t^{r-1} (1-t)^{\nu-r} dt$$

$$(r=1, 2, \dots, \nu) \quad (8)$$

$$\sum_{i=0}^r \binom{i}{\nu} \theta^i (1-\theta)^{\nu-i} = \frac{\nu!}{r! (\nu-r-1)!} \int_{\theta}^1 t^r (1-t)^{\nu-r-1} dt$$

$$(r=0, 1, \dots, \nu-1). \quad (9)$$

Интегралы в правых частях этих соотношений называются неполными бета-функциями, и их значения приводятся в специальных таблицах ¹⁾.

* § 8.3. Результат (8.2.4) является частным случаем более общей предельной теоремы. В § 7.5 мы видели, что сумма ν независимых нормально распределенных случайных величин сама распределена нормально. Однако при достаточно большом ν эта сумма будет, вообще говоря, распределена приблизительно нормально даже тогда, когда сами величины x не являются нормально распределенными. Этот факт утверждается *центральной предельной теоремой*:

Пусть x_1, x_2, \dots есть последовательность независимых случайных величин с произвольными распределениями, для которых средние значения μ_1, μ_2, \dots и дисперсии $\sigma_1^2, \sigma_2^2, \dots$ существуют. Образует новые

случайные величины $\sum_{i=1}^{\nu} x_i$ и пронормируем их, положив

$$y_{\nu} = \frac{\sum_{i=1}^{\nu} x_i - \sum_{i=1}^{\nu} \mu_i}{\left(\sum_{i=1}^{\nu} \sigma_i^2 \right)^{\frac{1}{2}}}; \quad \nu = 1, 2, \dots$$

Легко видеть, что

$$\mathbb{M}\{y_{\nu}\} = 0, \quad \sigma\{y_{\nu}\} = 1 \quad (1)$$

(проверить). Тогда при весьма общих условиях

$$\Phi_{y_{\nu}}(t) \xrightarrow{\nu \rightarrow \infty} \Psi(t), \quad (2)$$

где $\Psi(t)$ есть нормированная нормальная функция распределения. Таким образом, для достаточно больших значений ν величины y_{ν} , а следовательно, и $x_1 + \dots + x_{\nu}$ распределены приблизительно нормально. Укажем на следующее условие, достаточное для выполнения центральной предельной теоремы: соотношение (2) имеет место, если су-

¹⁾ K. Pearson, Tables of Incomplete B-function, London, 1934.

существуют такие числа m и M , что

$$\sigma_i^2 = \mathfrak{M} \{ (x_i - \mu_i)^2 \} > m > 0 \text{ и } \mathfrak{M} \{ |x_i - \mu_i|^3 \} < M$$

для всех $i = 1, 2, \dots$ (3)

Мы видим, что из условия 3 вытекает, что

$$\sum_{i=1}^{\nu} \sigma_i^2 \xrightarrow{\nu \rightarrow \infty} \infty. \quad (4)$$

Про последнее условие говорят, что оно выражает тот факт, что ни одна из величин x_i не является преобладающей.

Если условие (4) (а следовательно, и условие (3)) не выполнено, то центральная предельная теорема может иметь место только в том случае, когда все величины x_i распределены нормально (т. е. когда $\Phi_{\nu, \nu}(t) = \Psi(t)$ для всех ν).

Очевидно, что для выполнения условия (3) в случае, когда все величины x_1, x_2, \dots имеют одно и то же распределение, достаточно потребовать только существования третьего центрального момента. Поэтому, так как биномиальное распределение может быть представ-

лено в виде $x = \sum_{i=1}^{\nu} x_i$, где все величины x_i имеют одно и то же распределение, теорема Муавра (8.2.4) содержится в центральной предельной теореме. Эта теорема имеет место даже при более слабых условиях, чем (3), но мы здесь не будем давать ни их точных формулировок, ни доказательств¹⁾. Теорема эта имеет место также для некоторых классов зависимых величин²⁾.

Пример 1. Для случая одинаково распределенных случайных величин x_1, x_2, \dots центральная предельная теорема может быть легко доказана с помощью характеристических функций^{*}). Пусть $\Phi(t)$ есть

функция распределения нормированной величины $z_i = \frac{x_i - \mu_i}{\sigma_i}$, а $\chi(t)$ есть соответствующая характеристическая функция. В силу определения характеристической функции $\chi(t)$ и того, что для $\Phi(t)$ $\mu = 0$ и $\sigma = 1$,

имеем $\chi(t) = 1 - \frac{1}{2} t^2 + o(t^2)$, где $\frac{o(t^2)}{t^2} \rightarrow 0$. Поэтому

$\chi\left(\frac{t}{\sqrt{\nu}}\right) = 1 - \frac{t^2}{2\nu} + o\left(\frac{t^2}{\nu}\right)$. Для характеристической функции вели-

чины $y_{\nu} = \frac{z_1 + \dots + z_{\nu}}{\sqrt{\nu}}$ имеем в силу (6.2.8)

$$\chi_{y_{\nu}}(t) = \left(\chi\left(\frac{t}{\sqrt{\nu}}\right) \right)^{\nu} = \left(1 - \frac{t^2}{2\nu} + o\left(\frac{t^2}{\nu}\right) \right)^{\nu} \xrightarrow{\nu \rightarrow \infty} \exp\left[-\frac{t^2}{2}\right], \quad (5)$$

¹⁾ См., например, Крамер, Математические методы статистики, 1948 или Случайные величины и распределения вероятностей, 1947 и Хинчин, Асимптотические законы теории вероятностей, 1936.

²⁾ См. Lévy, Théorie de l'addition des variables aléatoires, 1937.

^{*}

где справа стоит характеристическая функция распределения $\Psi(x)$ (см. (5.2.21)). Из теоремы сходимости, данной в § 5.2, сразу следует теперь (2).

Пример 2. Интересен следующий пример, показывающий, что если условие (4) не выполнено, то центральная предельная теорема может не иметь места. Рассмотрим случайную величину x , имеющую распределение Лапласа (4.4.10) с параметрами $\mu=0$ и $\alpha=1$. Тогда в силу упражнения 2 § 5.1 $\mathfrak{M}\{x\}=\mu=0$ и в силу (5.3.15) $\sigma\{x\}=\sqrt{2\alpha}=\sqrt{2}$. Положим $x_1=\frac{2}{\pi}\frac{1}{1}x$, $x_2=\frac{2}{\pi}\frac{1}{3}x$, $x_3=$

$=\frac{2}{\pi}\frac{1}{5}x$, ... Тогда $\sum_{i=1}^{\nu} \mu_i=0$ и

$$\sum_{i=1}^{\nu} \sigma_i^2 = \frac{8}{\pi^2} \sum_{i=1}^{\nu} \frac{1}{(2i-1)^2} \rightarrow \frac{8}{\pi^2} \sum_{i=1}^{\infty} \frac{1}{(2i-1)^2} = \frac{8}{\pi^2} \zeta(2)(1-2^{-2})=1,$$

где $\zeta(z)$ есть дзета-функция Римана¹⁾. Таким образом, условие (4) не

выполнено. Для вычисления $\lim_{\nu \rightarrow \infty} \Phi_{y_{\nu}}(t) = \Phi_y(t)$, $y = \lim_{\nu \rightarrow \infty} y_{\nu}$, $y_{\nu} = \sum_{i=1}^{\nu} x_i$, $\mathfrak{M}\{y\}=0$ и $\sigma\{y\}=1$ используем свойства характеристических функций. В силу (5.2.23) $\chi_x(t) = \frac{1}{1+t^2}$. Следовательно, в силу упражнения 2 § 6.2 и теоремы сходимости из примера 2 § 5.2 имеем

$$\begin{aligned} \chi_y(t) &= \lim_{\nu \rightarrow \infty} \chi_{x_1}(t) \dots \chi_{x_{\nu}}(t) = \prod_{i=1}^{\infty} \frac{1}{1 + \left(\frac{2}{\pi} \frac{t}{2i-1}\right)^2} = \\ &= \frac{1}{\operatorname{ch} t} = \frac{2}{e^t + e^{-t}}. \end{aligned}$$

В силу (5.2.14) получим

$$\begin{aligned} \varphi_y(u) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iut} \frac{2}{e^t + e^{-t}} dt = \\ &= \frac{1}{\exp\left[\frac{1}{2}\pi u\right] + \exp\left[-\frac{1}{2}\pi u\right]} = \frac{1}{2\operatorname{ch}\left(\frac{1}{2}\pi u\right)}, \end{aligned} \quad (6)$$

где интеграл может быть найден с помощью интегрирования функции комплексного переменного по прямоугольнику высоты π и бесконечно-большой ширины, одну из сторон которого образует действительная ось t . Очевидно, что распределение, определяемое плотностью вероятности (6), отлично от нормального.

¹⁾ См. Уиттекер и Ватсон, Курс современного анализа, т. II, 1934.

§ 8.4. В § 8.2 мы рассмотрели предел биномиального распределения при постоянном θ и $\nu \rightarrow \infty$. Пусть теперь при $\nu \rightarrow \infty$ остается постоянным не θ , а $\mathfrak{M}\{x\} = \mu = \nu\theta$ (и, значит, $\theta \rightarrow 0$ при $\nu \rightarrow \infty$). Тогда в силу (4.3.4) получим, вводя μ вместо θ в выражение для φ_i ,

$$\begin{aligned} \lim_{\nu \rightarrow \infty} \varphi_i &= \lim_{\nu \rightarrow \infty} \binom{i}{\nu} \left(\frac{\mu}{\nu}\right)^i \left(1 - \frac{\mu}{\nu}\right)^{\nu-i} = \\ &= \frac{\mu^i}{i!} \lim_{\nu \rightarrow \infty} 1 \cdot \left(1 - \frac{1}{\nu}\right) \dots \left(1 - \frac{i-1}{\nu}\right) \left(1 - \frac{\mu}{\nu}\right)^{-i} \left(1 - \frac{\mu}{\nu}\right)^\nu = \\ &= \frac{\mu^i}{i!} e^{-\mu}, \end{aligned} \quad (1)$$

так как $\lim_{x \rightarrow 0} (1+x)^{\frac{1}{x}} = e$. Таким образом, при постоянном μ и $\nu \rightarrow \infty$ биномиальное распределение сходится к распределению Пуассона. При условии, что $\nu \gg i$ и $\nu \gg \mu = \nu\theta$, т. е. $\theta \ll 1$, мы получаем из соотношения (1) формулу, соответствующую (8.2.1),

$$\varphi_i = \binom{i}{\nu} \theta^i (1-\theta)^{\nu-i} = \frac{\mu^i e^{-\mu}}{i!} + \frac{R_\nu}{\nu}, \quad \mu = \nu\theta, \quad (2)$$

где R_ν по абсолютной величине меньше некоторого числа, не зависящего от ν . Равенство (2) называется *формулой Пуассона* или, на основании условия $\theta \ll 1$, *законом малых чисел*. Мы уже встречались с примерами случайных величин, имеющих распределение Пуассона (см. примеры 4 и 5 § 4.3).

Пример. Если μ очень велико, то естественно ожидать, что формула Пуассона перейдет в формулу Лапласа с $\nu\theta = \mu$ и $\theta = 0$, а именно

$$\varphi_i = \frac{1}{\sqrt{2\pi\mu}} \exp\left[-\frac{(i-\mu)^2}{2\mu}\right] + \frac{R_\nu}{\nu}, \quad (3)$$

что можно получить непосредственно из равенства (2) с помощью формулы Стирлинга.

9. ОТНОШЕНИЕ ТЕОРИИ ВЕРОЯТНОСТЕЙ К ОПЫТУ И ЕЕ ПРАКТИЧЕСКОЕ ЗНАЧЕНИЕ

§ 9.1. Заканчивая изложение математической теории и переходя к обсуждению некоторых приложений этой теории, мы должны сначала выяснить вкратце следующие два вопроса:

(I) Как можно проверить теорию на опыте? (II) Где и каким образом применяется теория вероятностей?

I. Отношение к опыту

§ 9.2. При сравнении между теорией и опытом мы должны требовать выполнения следующего основного критерия согласия:

А. Если вероятность осуществления события пренебрежимо мала, то это событие „практически“ не должно осуществляться в единичном опыте.

При этом мы не требуем, чтобы событие вообще никогда не происходило, так как даже наиболее невероятное событие все же может осуществиться. Так, например, если нужно приготовить годовой отчет большого банка, то можно, конечно, постараться сократить количество работы и просто выдумать все числа, и в принципе возможно, что все выдуманные числа окажутся правильными, но никто не стал бы отрицать, что такое событие настолько неправдоподобно, что для всех практических целей возможностью его осуществления можно пренебречь.

Пример 1. При использовании критерия А следует соблюдать некоторую осторожность, если имеется лишь конечное число равновероятных возможностей. Например, поскольку все возможные комбинации карточной игры в бридж равновероятны, можно ожидать, что такая комбинация, в которой хотя бы один из партнеров получит при сдаче 13 карт одной масти, столь же возможна, как и любая другая комбинация, несмотря на то, что вероятность первого из этих событий равна всего лишь $1/(4 \cdot 10^{10})$ (см. задачу 7). Тем не менее, если кажется непонятным, что такие, несомненно, редкие события, как осуществление определенной комбинации сдачи карт, все время происходят, то причиной этого непонимания является бессознательное разделение возможных исходов события всего лишь на две группы. Если вместо рассмотрения обычной колоды карт перенумеровать 52 карты случайным образом номерами от 1 до 52, то рассматриваемый результат будет соответ-

говать некоторому количеству перестановок чисел от 1 до 52, ни одну из которых мы не считаем чем-либо выделяющейся. Таким образом, совершенно неважно, какое событие мы выделяем, так как бессознательно при рассмотрении проблемы мы сравниваем лишь две возможности: или 13 карт имеют одну и ту же масть, или карт одной и той же масти меньше, чем 13. Однако в этом случае очевидно, что вероятность осуществления первой из этих возможностей пренебрежимо мала по сравнению с вероятностью осуществления второго исхода.

Далее, подчеркнем, что мы не требуем, чтобы событие не могло осуществиться в *серии* наблюдений. Например, если событие имеет вероятность $1/n$, где n очень велико, то вероятность осуществления события хотя бы один раз в серии из n наблюдений равна

$$1 - \left(1 - \frac{1}{n}\right)^n \sim 1 - e^{-1} \approx 0,63$$

(почему?), и эта вероятность никоим образом не пренебрежима. Действительно, событие осуществляется в среднем один раз в каждом n наблюдениях (почему?).

§ 9.3. Следует заметить, что, строго говоря, основной критерий согласия **A** не может быть проверен непосредственно. Действительно, если событие имеет пренебрежимо малую вероятность и если оно в единичном опыте не произошло, теория, несомненно, оказалась в согласии с опытом. Но если, тем не менее, событие произошло, находится ли теория в согласии с опытом или нет? Однако такое затруднение вытекает из самой природы проблемы, в силу которой единичное наблюдение никогда не может ни доказать, ни опровергнуть *статистическую* теорию. Доказательство или опровержение возможно лишь при осуществлении многих наблюдений, как это будет сформулировано в нижеследующих критериях I—VI, выведенных из основного критерия **A** и из математически доказанных теорем теории вероятностей, в частности, тех, которые связаны с понятием *сходимости по вероятности* (§ 8.1).

По теореме Бернулли (8.1.13), относительная частота осуществления события сходится по вероятности к вероятности этого события. Таким образом, согласно этой теореме и критерию **A** мы должны требовать выполнения следующего критерия согласия между теорией и опытом:

I. Если количество наблюдений в данной серии велико, то наблюдаемая в ней относительная частота должна быть очень близкой к вычисленной вероятности.

Таким образом, критерий I имеет значение лишь тогда, когда вероятность события вычислена с помощью других вероятностей; в противном случае критерий I был бы просто эквивалентен определению понятия вероятности (см. § 1.4).

Если для случайной величины выведена некоторая функция распределения, то мы должны потребовать следующее:

II. *Большое количество наблюдаемых значений случайной величины должно быть распределено в соответствии с теоретической функцией распределения этой величины.*

В статистике (см. § 10.4—10.8) обсуждается, как можно осуществить практически такое сравнение.

Согласно **A** и (8.1.6) мы должны требовать следующее:

III. *Функция, зависящая от результатов n наблюдений и имеющая дисперсию, стремящуюся к нулю при неограниченном возрастании числа наблюдений n , при больших значениях n должна быть практически равна своему среднему значению.*

В частности, поскольку согласно закону больших чисел среднее арифметическое сходится по вероятности к математическому ожиданию, мы должны выставить также следующее требование:

IV. *Среднее арифметическое большого числа независимых измерений случайной величины должно быть практически равным математическому ожиданию этой величины.*

Таким образом, математическое ожидание указывает средний порядок значений рассматриваемой случайной величины.

Согласно критерию **A** и неравенству Чебышева (8.1.4) мы должны потребовать следующее:

V. *Наблюдаемое значение случайной величины практически не должно отклоняться от математического ожидания этой величины более чем на среднее квадратическое отклонение, взятое небольшое количество раз.*

Пример. В § 8.2 с помощью теоремы Бернулли было получено утверждение о распределении ряда относительных частот около соответствующей вероятности.

Согласно (8.2.4) и критерию **A** мы должны выставить следующее требование:

VI. *Относительные частоты некоторого события в большом количестве серий наблюдений, каждая из которых состоит из большого числа наблюдений, должны быть нормально распределены со средним значением, равным вероятности этого события.*

II. Практическое значение теории вероятностей

§ 9.4. Как мы уже видели в разделе 1, теория вероятностей применяется, когда явления описываются и анализируются с помощью статистического описания. Но, прежде всего, в чем заключается использование теории вероятностей?

Во-первых, теория вероятностей используется в *статистике* для *цели* *описательных* целей, для выражения статистического материала в кратком и сжатом виде (см. § 10.1). Важные приложения статистики мы находим в *экономических науках* и в *статистике народонаселения*, а также в *теории ошибок* и в *теории уравнивания* (см. разделы 11 и 12).

§ 9.5. *Во-вторых*, теория вероятностей используется в *статистике* для *цели* *анализа*. Произведя множество экспериментальных исследований, мы стараемся проанализировать некоторое явление, сравнивая наблюдения, полученные при разных условиях. Основной принцип такого анализа заключается в том, чтобы сохранить, насколько это возможно, постоянными все факторы, кроме одного, меняющегося, чтобы изучить влияние этого фактора на рассматриваемое явление. Однако если рассматриваемые явления случайны и обнаруживают *статистические флуктуации* (§ 1.1), т. е. при повторении экспериментов при одних и тех же условиях могут давать более или менее различные результаты, то необходимо изучать эти явления *статистическими методами*. Так, например, если при артиллерийской стрельбе химический состав пороха меняется и нужно выяснить возможность стрельбы на большую дистанцию с новым сортом пороха, то благодаря статистическим флуктуациям недостаточно выстрелить один раз со старым и один раз с новым сортом пороха. Чтобы измерить степень статистических флуктуаций и таким образом решить, обеспечивает ли новый состав пороха получение желаемых результатов, следует произвести целую серию выстрелов как со старым, так и с новым сортом пороха. После этого возникает проблема: является ли полученная разница *значимой*, или она лишь такова, какой следовало ожидать вследствие статистических флуктуаций (см. § 11.15)?

Такой статистический анализ играет в современной науке все возрастающую роль¹⁾. Так, например, если в *медицине* нужно сравнить действие различных снотворных средств, то нужно испытать их действие на ряде пациентов, после чего возникает проблема: имеется ли какая-нибудь *значимая* разница в действии этих средств, помимо той разницы, которую следует ожидать из-за статистических флуктуаций? Или, например, в *агрономической науке* бывает нужно сравнить различные удобрения. Тогда следует разбить некоторое количество опытных земельных участков на делянки и удобрить все делянки на каждом участке различными удобрениями. Опять возникает вопрос: имеется ли какая-нибудь *значимая* разница в эффективности различных удобрений, кроме той разницы, которую следует ожидать из-за статистических флуктуаций? Или, например, в *торговом деле* может потребоваться выяснить, какой вид рекламы более доходчив до потребителя;

¹⁾ Ряд таких примеров можно найти в книге Fisher, *Statistical Methods for Research Workers*, особенно в гл. V, 1938.

для этого используют в различных группах населения различные виды рекламы, и вновь возникает вопрос: имеется ли какая-либо *значимая* разница помимо вызванной статистическими флуктуациями? Подобных примеров можно привести множество.

Другой вид *статистического анализа* применяется в проблемах (а) *корреляции* и (б) *регрессии*. Здесь исследуется вопрос, (а) зависят ли не зависят друг от друга некоторые явления, или коррелированы ли они, и (б) если зависимость между явлениями есть, то какова она (см. § 11.18 и § 12.15)¹⁾. Такие вопросы встречаются в *агрономической науке* при исследовании связи между количеством осадков и урожаем; или в *артиллерии* при исследовании того, имеется ли какая-нибудь связь между боковыми отклонениями и отклонениями по высоте попаданий (см. пример § 11.18), и т. д.

Третий вид *статистического анализа* применяется в *технологии* при исследовании того, удовлетворяют ли стандартизированные изделия определенным условиям²⁾. Так, например, если дирекция электротехнического завода хочет быть уверенной, что продолжительность работы выпускаемых заводом электроламп не будет ниже определенной нормы, то, конечно, дирекция не может организовать испытание каждой из выпускаемых электроламп. Вместо этого из всей продукции извлекаются выборки, и на основании результатов испытания этих выборок оценивается степень надежности гарантированной нормы для всей продукции. Очевидно, что такой анализ может иметь большое экономическое значение.

§ 9.6. В-третьих, теория вероятностей используется для *предсказания* будущего хода *случайного явления*. Старейшим приложением теории вероятностей для этих целей является область *азартных игр*. Практически чрезвычайно важное приложение мы находим в *страховом деле*, в котором на основании статистических наблюдений над смертностью, заболеваемостью, пожарами, кражами и т. д. необходимо предсказать количество этих событий на заданный период будущего времени. Третье приложение такого рода мы находим в различных областях *теоретической физики*³⁾, где бывает нужно предсказать ход экспериментов, связанных с кинетической теорией газов (см. пример § 4.8; упражнение 3 § 7.4; § 4.16), радиоактивным распадом (см. пример § 3.7; примеры: 1 § 3.8, 5 § 4.3, 2 § 4.4) или с другими явлениями в теории атома (см. § 4.15, § 4.17, пример 6 § 5.1) и т. д. В *технологии* теория вероятностей также используется для целей предсказания, например в телефонии при проектировании коммутаторов, когда нужно выяснить, сколько нужно иметь телефонисток на центральной станции при заданном числе абонентов, чтобы время

¹⁾ Ряд таких примеров можно найти в книге Fisher, *Statistical Methods for Research Workers*, особенно в гл. V и VI.

²⁾ См., например, Фрай, *Теория вероятностей для инженеров*, 1934.

³⁾ См., например, F ü r t h, *Theoretische Physik*.

ожидания при вызове центральной станции не превосходило определенного предела (см. примеры 4 § 4.3, 2 § 4.4), и т. д. В *артиллерии* теория вероятностей используется для вычисления доли выстрелов, когда снаряды лягут перед целью, если орудие отрегулировано для максимального количества попаданий, или для оценки того, в каких случаях попадание можно считать случайным (см. § 11.17)¹⁾, и т. д. В *торговом деле* теория вероятностей может быть использована для оценки того, насколько большие склады нужно приготовить для хранения запасов некоторого товара, сколько продавцов нужно иметь в различных отделениях магазина (см. пример 5 § 4.3) и т. д.

1) См. например, Hayes, Elements of Ordnance.

10. ПРИЛОЖЕНИЕ ТЕОРИИ ВЕРОЯТНОСТЕЙ К СТАТИСТИКЕ

§ 10.1. *Статистика может быть охарактеризована вкратце, как наука о сокращении и анализе наблюдаемого материала* ¹⁾. Как правило, „сырой“ статистический материал, состоящий из некоторого числа наблюдаемых значений x_1, x_2, \dots, x_n случайной величины x (см. § 4.1), расположенных в том порядке, в котором они были наблюдаемы, представляет трудности как для обзора данных, так и для их воспроизведения. Поэтому такой материал неудобен для непосредственного получения каких-либо сведений относительно исследуемой величины x . Целью статистики является замена наблюдаемого материала сравнительно небольшим количеством чисел, представляющих весь материал, или, иными словами, содержащих как можно больше сведений относительно величины x .

Все наблюдения, содержащиеся в статистическом материале, лишь очень редко могут быть получены из теоретических соображений. Следовательно, данный статистический материал, как правило, надо считать *случайной выборкой*, которая сама по себе подвержена статистическим флуктуациям, так как если произвести новые n наблюдений, то получатся новые значения x'_1, x'_2, \dots, x'_n .

Таким образом, даже и в простейших случаях естественно использовать понятие вероятности и вводить для относительных частот различных возможных результатов эксперимента некоторые идеализированные значения — вероятности (§ 1.4), т. е. связывать с x некоторую функцию распределения $\Phi_x(t)$ (§ 4.2).

Как было указано в § 2.7, можно удобно и коротко выразить отмеченное обстоятельство, говоря, что наблюдаемая выборка извлечена случайным образом из гипотетического бесконечного множества теоретической *генеральной совокупности*. Этот способ выражения широко используется в статистике, хотя, строго говоря, он имеет точный смысл лишь при весьма абстрактной формулировке понятия вероятности, данной Колмогоровым (см. § 2.7).

Заметим, что, как правило, мы не интересуемся фактическими *эмпирическими* числами x_1, \dots, x_n , но лишь теоретическим понятием—

¹⁾ Эта формулировка цели статистики принадлежит английскому статистiku Р. А. Фишеру. См. Fisher, *Statistical Methods for Research Workers*, 1938.

функцией распределения $\Phi_x(t)$, т. е. *моделью*, с помощью которой мы описываем имеющиеся данные (см. § 1.1). Аналогично этому, например в физике, при исследовании, скажем, закона свободного падения тел $s = \frac{1}{2}gt^2$ мы не интересуемся фактическими измеренными значениями (s, t) , которые благодаря неизбежным ошибкам измерений более или менее разбросаны относительно теоретической кривой. При этом исследовании мы интересуемся лишь тем, как с помощью полученных наблюдений получить численное значение параметра g , его „истинное“ значение, и, конечно, тем, дает ли предложенный теоретический закон удовлетворительное описание наблюдаемого материала.

§ 10.2. Теоретическая функция распределения $\Phi_x(t)$ содержит, как правило, одну или несколько постоянных или *параметров*, как, например, μ и σ в нормальном распределении. Устанавливая численные значения этих параметров, мы полностью характеризуем исследуемую случайную величину. Таким образом, нашей целью является *оценить* или построить *оценки* для численных значений параметров функции распределения $\Phi_x(t)$ на основе имеющихся результатов наблюдений. Такие эмпирические оценки теоретических параметров сами являются, конечно, случайными величинами, подверженными статистическим колебаниям. Чтобы получить меру ожидаемой величины этих колебаний и тем самым меру надежности, с которой можно судить о значениях параметров по их значениям, вычисленным по выборке, мы должны, пользуясь функцией распределения $\Phi_x(t)$, вывести функции распределения для предложенных оценок. Сделав это, мы решим свою задачу и сократим данный наблюдаемый материал настолько, насколько это возможно.

Таким образом, в каждой статистической задаче требуется получить ответы на следующие четыре вопроса:

I. Какая функция распределения соответствует рассматриваемой случайной величине?

II. Каким образом можно выяснить, насколько хорошо выбранная функция распределения описывает имеющиеся данные наблюдений?

III. Как вычислить по данной выборке, т. е. по данным наблюдаемым значениям x_1, \dots, x_n , наилучшие возможные оценки неизвестных параметров функции распределения?

IV. Как вывести, пользуясь функцией распределения $\Phi_x(t)$, распределения этих оценок, т. е. получить меру их надежности?

§ 10.3. Итак, прежде чем решать данную статистическую проблему, мы должны сначала выбрать *гипотезу* относительно математического вида функции распределения. Иногда из предыдущего опыта бывает известно, что можно использовать некоторую определенную форму распределения, например *нормальное* распределение (7.1.1). Или же можно *вывести* функцию распределения с помощью правил теории вероятностей, исходя из некоторых упрощающих предположе-

ний о рассматриваемом явлении. Как правило, этот последний случай имеет место в теоретической физике и в теории азартных игр, что было показано на ряде примеров в разделах 1—9.

Пример 1. Так же обстоит дело и в теории ошибок (см. § 11.4), в которой можно предполагать, что ошибки измерений являются результатом действия большого числа независимых причин, и поэтому наблюдаемую ошибку можно представлять как сумму так называемых „элементарных ошибок“. Вследствие центральной предельной теоремы (см. § 8.3) наблюдаемую ошибку с высокой степенью приближения можно считать распределенной нормально.

Пример 2. Обобщая „схемы элементарных ошибок“ предыдущего примера, Кэптейн¹⁾ вывел из нормального распределения целый класс новых распределений. Допустим, что случайная величина x непосредственно не является суммой большого количества величин, но что ее значения являются результатом действия большого числа причин, производящих последовательные „импульсы“, эффект которых зависит частично от интенсивности самих импульсов и частично от величины x , созданной действием предыдущих импульсов. Пусть z_1, z_2, \dots, z_n суть независимые импульсы, производимые n причинами, и пусть x_i есть результат действия первых i этих импульсов. Допустим, наконец, что x_{i+1} зависит лишь от значения x_i , но не зависит от предистории, т. е. от значений x_k при $k < i$. Другими словами, мы предполагаем существование такой функции $g(x)$, что

$$x_{i+1} = x_i + z_{i+1}g(x_i), \quad g(x) \neq 0.$$

Тогда если n велико и каждый импульс невелик, то

$$z_1 + z_2 + \dots + z_n = \sum_{i=0}^{n-1} \frac{x_{i+1} - x_i}{g(x_i)} \sim \int_{x_0}^x \frac{dx}{g(x)} = G(x). \quad (1)$$

При весьма общих условиях сумма $z_1 + \dots + z_n$, при больших n будет распределена приблизительно нормально, в соответствии с центральной предельной теоремой (§ 8.3). Таким образом, в этом случае не сама величина x , но некоторая функция $G(x)$, заданная формулой (1), имеет нормальное распределение: мы получаем класс распределений Кэптейна, который мы уже рассматривали в (7.7.1):

$$d\Phi_x(t) = d\Psi \left[\frac{G(t) - \mu}{\sigma} \right] = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(G(t) - \mu)^2}{2\sigma^2} \right] \left| \frac{dG(t)}{dt} \right| dt, \quad (2)$$

$$\mathfrak{M}\{G(x)\} = \mu, \quad \sigma\{G(x)\} = \sigma.$$

В частности, если $g(x) = x$, т. е. эффект действия данной причины пропорционален уже достигнутому значению случайной величины, то

¹⁾ Каптейн и van Uyven, Skew Frequency Curves.

в силу $\int \frac{dx}{x} = \ln x$ мы получаем следующее распределение, называемое *логарифмически-нормальным*:

$$dF_x(t) = dF\left(\frac{\ln t - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\ln t - \mu)^2}{2\sigma^2}\right] \frac{dt}{t} \quad (0 \leq t < \infty), \quad (3)$$

$$M\{\ln x\} = \mu, \quad \sigma\{\ln x\} = \sigma.$$

Эта формула дает, например, удовлетворительное описание распределения прироста веса школьников за период каникул.

Вообще говоря, мы не можем вывести функцию распределения теоретически, а должны стараться получить ее непосредственно из наблюдений. Далее, поскольку всегда возможно осуществить лишь конечное число наблюдений, к данному эмпирическому материалу можно подобрать целое множество теоретических формул, ни одна из которых не является ни „истинной“, ни „ложной“ (см. § 1.1). Но эти формулы могут описывать наблюдения более или менее удовлетворительно, и те, которые дают лишь грубое совпадение, могут быть исключены с помощью различных критериев (см. § 10.4). Вообще говоря, мы должны выбрать математическую форму для функции распределения $F_x(t)$, или известной нам аналитически или протабулированной численно¹⁾. В современной статистике класс распределений Кэптейна, определенный формулой (2), используется все чаще и чаще, так как в этом случае можно получить распределения для оценок. Однако также употребительны и некоторые ранее известные классы распределений. Примеры: класс распределений Пирсона, получаемых как решения дифференциального уравнения

$$\frac{d\varphi(t)}{dt} = \frac{(t - \alpha)\varphi(t)}{\beta_0 + \beta_1 t + \beta_2 t^2},$$

где α , β_0 , β_1 и β_2 суть четыре параметра; класс распределений Грама—Шарлье, состоящий из разложений по нормальной функции распределения $\Psi\left(\frac{t - \mu}{\sigma}\right)$ и ее производным¹⁾.

§ 10.4. Предположим теперь, что вопрос I уже решен, и гипотеза относительно функции распределения $F_x(t)$ высказана. Далее нам надлежит ответить на вопрос II, т. е. выработать методы для проверки того, насколько хорошо теоретическое распределение удовлетворяет имеющимся экспериментальным данным. Простейший критерий состоит в нанесении эмпирических данных на график и в проведении на этом же графике для сравнения теоретической кривой.

Исследуем сначала случай *дискретных распределений* (§ 4.3).

¹⁾ О распределениях общего вида см. Kendall, The Advanced Theory of Statistics.

Нанесем на график по оси абсцисс возможные значения случайной величины t_i и по оси ординат — соответствующие относительные частоты

$$f_i = \frac{n_i}{n}, \quad (1)$$

причем n_i есть число тех из данных n наблюдений, в которых был получен результат t_i . Говорят, что такой график дает *эмпирическое* или *наблюденное распределение*, которое называют также *распределением выборки*. Для сравнения нанесем на тот же график *теоретическое* или *истинное* распределение, т. е. вероятности φ_i результатов t_i . Поскольку вероятности φ_i определены лишь при $t = t_i$, может оказаться удобным соединить *теоретические* точки гладкой кривой, как показано, например, на фиг. 7. Также можно нанести на график в качестве ординат абсолютные частоты n_i , причем если n_i суть лишь небольшие числа, то может оказаться наиболее удобным нанести над каждым значением t_i количество точек, равное числу n_i , как это сделано, например, на фиг. 1 стр. 17. Для сравнения мы нанесем теоретические значения n_i , т. е. числа $\nu_i = n\varphi_i$.

Поскольку f_i есть относительная частота события $x = t_i$, имеющего вероятность φ_i , при n независимых наблюдениях величины x , эта частота при каждом фиксированном значении i есть случайная величина, для которой по (6.4.14) и (6.4.15) имеем

$$\mathbb{M}\{f_i\} = \varphi_i \quad (2)$$

$$\sigma^2\{f_i\} = \frac{\varphi_i(1-\varphi_i)}{n} \xrightarrow[n \rightarrow \infty]{} 0. \quad (3)$$

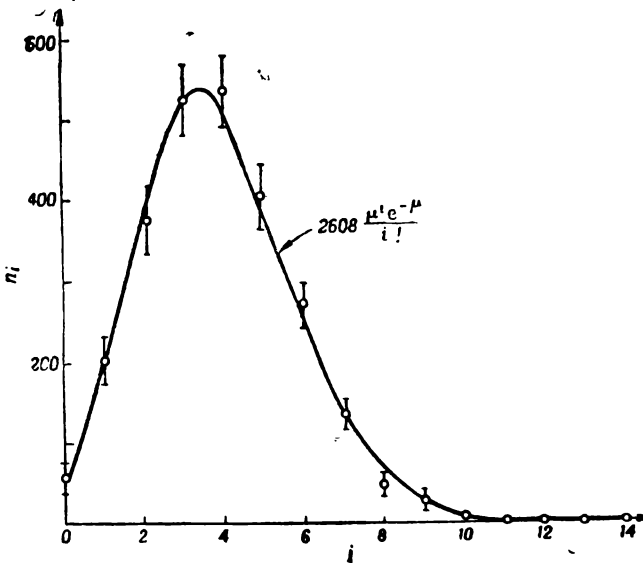
Таким образом, согласно теореме Бернулли (8.1.13), $f_i \xrightarrow[n \rightarrow \infty]{\text{по вер.}} \varphi_i$, т. е. следует ожидать, что при каждом фиксированном значении t_i наблюдаемое значение f_i при больших n должно быть очень близким к вероятности φ_i .

Пример. Пусть случайная величина есть число x радиоактивных атомов, распадающихся в течение некоторого промежутка времени. В примере 5 § 4.3 показано, что x имеет распределение Пуассона. При измерении за $n = 2608$ промежутков времени (каждый в $1/8$ минуты) были получены значения x , приведенные в табл. 2¹⁾. В этой таблице n_i означает количество промежутков времени, в течение которых согласно наблюдениям распалось i атомов; $\nu_i = n\varphi_i$ есть соответствующее теоретическое значение, причем вероятности φ_i вычислены по формуле (4.3.5) с $\mu = 3,87$ (см. пример 3 § 10.10). На фиг. 7 нанесены при каждом i значения n_i и ν_i . Как из таблицы, так и из графика видно, что согласие между теорией и экспериментом, т. е. между теоретическим распределением и распределением выборки, вполне удовлетворительное (см. пример 1 § 10.8).

¹⁾ См. Rutherford, Chadwick и Ellis, Radiations from Radioactive Substances, стр. 172, London, 1930.

Таблица 2

Количество атомов, расплавшихся за один интервал	Наблюдаемое количество интервалов	Теоретическое количество интервалов
i	n_i	ν_i
0	57	54
1	203	210
2	383	407
3	525	525
4	532	508
5	408	394
6	273	255
7	139	140
8	45	68
9	27	29
10	10	11
11	4	4
12	0	1
13	1	1
14	1	1
$n = \sum_i n_i = 2608$		$\nu = \sum_i \nu_i = 2608$



Ф и г. 7.

§ 10.5. Рассмотрим теперь *непрерывные распределения* (§ 4.4). Произведем *группировку* данных, т. е. разделим весь интервал, на котором лежат наблюдаемые значения x_1, x_2, \dots, x_n , на некоторое количество $m < n$ частей длин Δt_i , называемых *интервалами группировки*, которые не обязаны быть одинаковой длины. Средние точки этих интервалов обозначим t_1, t_2, \dots, t_m . Для каждого t_i подсчитаем число наблюдений n_i , результаты которых суть числа, принадлежащие i -му интервалу группировки, т. е. для которых

$$t_i - \frac{\Delta t_i}{2} < x \leq t_i + \frac{\Delta t_i}{2}. \quad (1)$$

Образует соответствующие относительные частоты, деленные на Δt_i ,

$$f(t_i) = \frac{1}{\Delta t_i} \frac{n_i}{n} \quad (2)$$

и нанесем на график в системе координат (t, u) ступенчатую функцию

$$u = f(t) = f(t_i) \text{ при } t_i - \frac{\Delta t_i}{2} < t \leq t_i + \frac{\Delta t_i}{2}. \quad (3)$$

Такой график называется *гистограммой*. Мы видим, что относительное число наблюдений, приходящихся на некоторый интервал группировки, равно *площади* между кривой $u = f(t)$ и интервалом на оси t . График функции $u = f(t)$ дает удобное представление результатов наблюдений, если n не слишком мало, и длины интервалов группировки Δt_i выбраны подходящим образом (см. ниже). Говорят, что такой график дает *эмпирическое*, или *наблюденное*, *распределение*, которое называют также *распределением выборки*. Далее для сравнения нанесем на тот же график соответствующее *теоретическое*, или *истинное*, *распределение*, т. е. плотность вероятности $u = \varphi(t)$. Также может оказаться удобным нанести на график в качестве ординат абсолютные частоты $\frac{n_i}{\Delta t_i}$ и для сравнения соответствующую теоретическую кривую $u = n\varphi(t)$. Если интервалы группировки имеют одинаковую длину Δt , то может оказаться удобным не делить на Δt и наносить на график числа n_i , а для сравнения — числа $n\Delta t\varphi(t)$.

Иногда, нанеся на график точки $\left(t_i, \frac{1}{\Delta t_i} \frac{n_i}{n}\right)$, $\left(t_i, \frac{n_i}{\Delta t_i}\right)$ или (t_i, n_i) , соединяют их прямыми линиями, получая, таким образом, непрерывную кривую, называемую *полигоном частот*. Однако такой график менее удобен, чем гистограмма, так как мы оставляем здесь без внимания то обстоятельство, что, в противоположность теоретическим распределениям, каждое эмпирическое распределение дискретно, поскольку возможно сделать лишь конечное число наблюдений, результаты

которых суть числа, выраженные в единицах наименьшего доступного измерению приращения рассматриваемой величины.

Аналогично формулам (10.4.2) и (10.4.3) частота $f(t_i)$ при каждом фиксированном значении i есть случайная величина, для которой

$$\mathfrak{M} \{ f(t_i) \} = \frac{1}{\Delta t_i} \mathfrak{M} \left\{ \frac{n_i}{n} \right\} = \frac{1}{\Delta t_i} \int_{t_i - \frac{\Delta t_i}{2}}^{t_i + \frac{\Delta t_i}{2}} \varphi(t) dt = \varphi(\xi_i), \quad (4)$$

$$t_i - \frac{\Delta t_i}{2} < \xi_i < t_i + \frac{\Delta t_i}{2},$$

$$\sigma^2 \{ f(t_i) \} = \frac{1}{(\Delta t_i)^2} \sigma^2 \left\{ \frac{n_i}{n} \right\} = \frac{\varphi(\xi_i)[1 - \varphi(\xi_i) \Delta t_i]}{n \Delta t_i}. \quad (5)$$

Таким образом, согласно теореме Бернулли (8.1.13), $f(t_i) \xrightarrow[n \rightarrow \infty]{\text{по вер.}}$ $\varphi(\xi_i)$,

т. е. следует ожидать, что при каждом фиксированном значении t_i и при больших значениях n наблюдаемое значение $f(t_i)$ будет очень близким к $\varphi(\xi_i)$, причем $\varphi(\xi_i)$ есть некоторое среднее значение функции $\varphi(t)$ на интервале группировки Δt_i . При небольших Δt_i величину $\varphi(\xi_i)$ практически можно считать равной $\varphi(t_i)$.

Если на график наносятся значения n_i , то следует вычислить соответствующие теоретические значения

$$\gamma_i = n \int_{t_i - \frac{\Delta t_i}{2}}^{t_i + \frac{\Delta t_i}{2}} \varphi(t) dt \quad (6)$$

для каждого значения i и сравнить n_i с γ_i , что наиболее удобно осуществить с помощью таблицы (см. пример).

Как легко видеть, для того чтобы произвести сравнение теоретических и эмпирических значений насколько возможно детальнее, надо выбрать интервалы группировки возможно меньшей длины. Однако формула (5) показывает, что чем меньше длины интервалов Δt_i , тем больше статистические флуктуации частот $f(t_i)$, что приводит к менее регулярной форме гистограммы (см. фиг. 10 и 11). Поэтому n следует выбирать так, чтобы можно было выбрать интервалы группировки малых длин Δt_i и чтобы в то же время числа $n \Delta t_i$ были велики.

На практике обычно стараются выбрать число наблюдений n и длины интервалов группировки Δt_i так, чтобы каждый интервал группировки содержал хотя бы пять наблюдаемых значений (за исключением, возможно, лишь крайних интервалов группировки). Если длины интервалов Δt_i слишком малы, то будут доминировать стати-

стические флуктуации, если же длины Δt_i слишком велики, то могут быть утеряны детали функции распределения. Таким образом, за исключением случаев, когда n очень велико, гистограмма может употребляться лишь для грубого изображения, в то время как интегральная кривая (§ 10.6) может быть использована для всякого детального сравнения с теоретическим распределением.

Пример. Из 8-миллиметрового пулемета было произведено 96 выстрелов по цели, находящейся на расстоянии 300 м, после чего были измерены боковые отклонения попаданий и отклонения по высоте от центра цели, принятого за начало координат. Данные измерений представлены в следующих таблицах:

Таблица 3

Боковое отклонение в см		
t_i	n_i	v_i
-30	0	0,17
-20	2	1,80
-10	9	9,73
0	28	25,43
10	30	32,22
20	21	19,83
30	5	5,91
40	1	0,85
50	0	0,06
$\sum_i n_i =$ $= n = 96$		$\sum_i v_i =$ $= v = 96,00$

Таблица 4

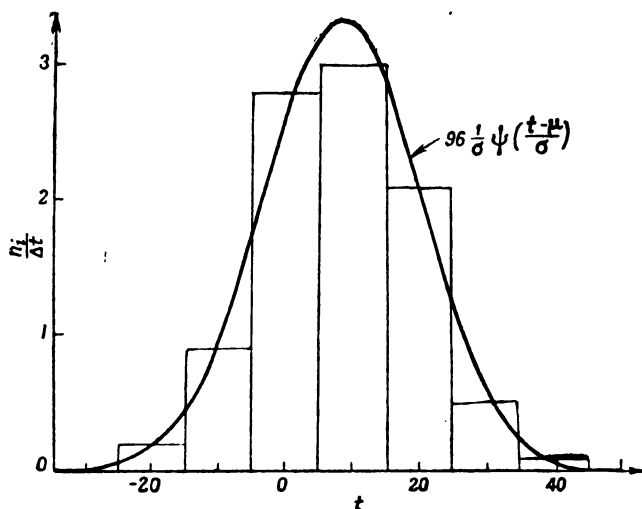
Отклонение по высоте в см		
t_i	n_i	v_i
-60	0	0,54
-50	3	1,95
-40	5	5,92
-30	13	12,86
-20	18	19,94
-10	21	22,07
0	21	17,46
10	10	9,86
20	5	3,98
30	0	1,42
$\sum_i n_i =$ $= n = 96$		$\sum_i v_i =$ $= v = 96,00$

Здесь n_i есть число выстрелов, при которых боковое отклонение или отклонение по высоте) оказалось принадлежащим i -тому интервалу группировки с серединой в точке t_i и длиной $\Delta t = 10$ см, а v_i даны формулой (6) для нормального распределения, т. е.

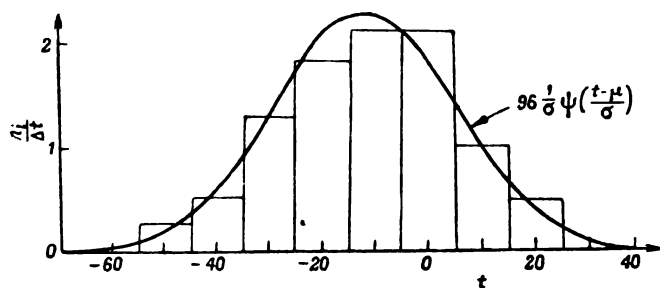
$$= 96 \int_{t_i - \frac{\Delta t}{2}}^{t_i + \frac{\Delta t}{2}} \frac{1}{\sigma} \phi\left(\frac{t - \mu}{\sigma}\right) dt = 96 \left[\Psi\left(\frac{t_i + \frac{\Delta t}{2} - \mu}{\sigma}\right) - \Psi\left(\frac{t_i - \frac{\Delta t}{2} - \mu}{\sigma}\right) \right], \quad (7)$$

где Ψ — функция, значения которой приведены в табл. I в конце книги, а параметры имеют численные значения $\mu = 8,3$ см и $\sigma = 11,4$ см для бокового отклонения и $\mu = -12,0$ см, $\sigma = 17,0$ см для отклонения по высоте (см. пример в § 11.12).

На фиг. 8 и 9 приведены две гистограммы для боковых отклонений и для отклонений по высоте (причем деление на $n=96$ не производилось). Для сравнения нанесены соответствующие теоретические



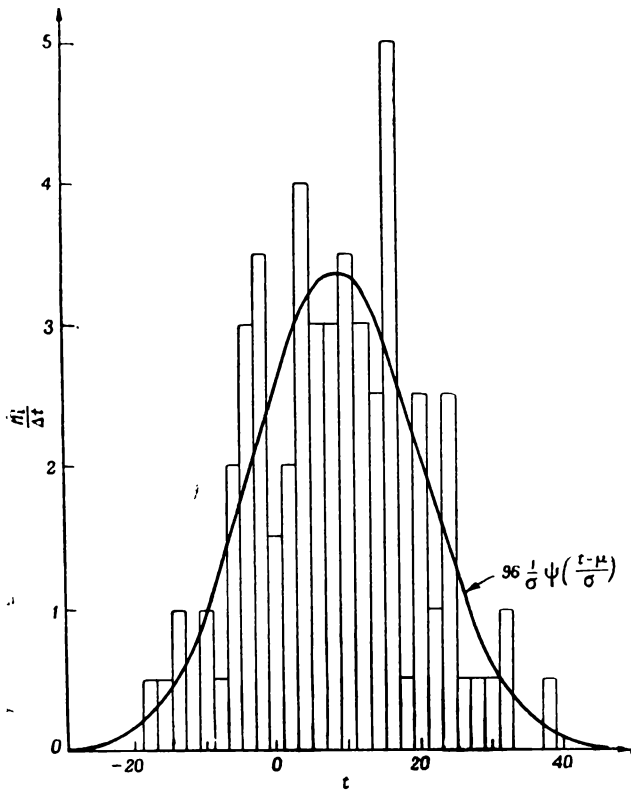
Фиг. 8. Боковое отклонение. $\Delta t = 10$ см.



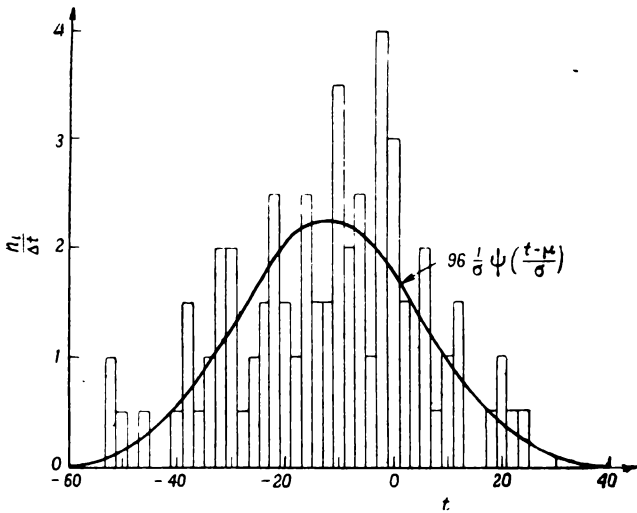
Фиг. 9. Отклонение по высоте. $\Delta t = 10$ см.

кривые $96 \frac{1}{\sigma} \psi \left(\frac{t-\mu}{\sigma} \right)$, полученные по табл. I при указанных выше значениях параметров μ и σ . Как из таблиц, так и из графиков видно, что согласие между теорией и наблюдениями вполне удовлетворительное. Действительно, в следующих двух параграфах будет показано, что согласие в этом случае лучше, чем это можно усмотреть из фиг. 8 и 9.

На фиг. 10 и 11 приведены гистограммы для тех же самых наблюдений, но при интервалах группировки длины $\Delta t = 2$ см. Как показывают чертежи, при этом статистические флуктуации возросли настолько, что, исходя из одних лишь этих кривых, мы остаемся



Фиг. 10. Боковое отклонение. $\Delta t = 2$ см.



Фиг. 11. Огклонение по высоте. $\Delta t = 2$ см.

в большом сомнении относительно того, согласуются результаты наблюдений с теоретическими кривыми или нет.

§ 10.6. Как для дискретных, так и для непрерывных распределений всегда может быть применен метод, заключающийся в нанесении на график полигона сумм

$$u = F(t) = \frac{N(t)}{n}, \quad (1)$$

где $N(t)$ есть число наблюдений, результаты которых суть числа, не превосходящие t . $F(t)$ называется также *накопленной гистограммой*, *эмпирической* или *наблюденной функцией распределения* или *функцией распределения выборки*. Для сравнения на тот же график наносят соответствующее *теоретическое* или *истинное распределение*, т. е. значения функции распределения $u = \Phi(t)$.

Здесь может оказаться удобным опустить деление на n в формуле (1) и наносить для сравнения соответствующую теоретическую кривую $y = n\Phi(t)$. Поскольку

$$\mathfrak{M}\{F(t)\} = \Phi(t), \quad (2)$$

$$\sigma^2\{F(t)\} = \frac{\Phi(t)[1 - \Phi(t)]}{n} \quad (3)$$

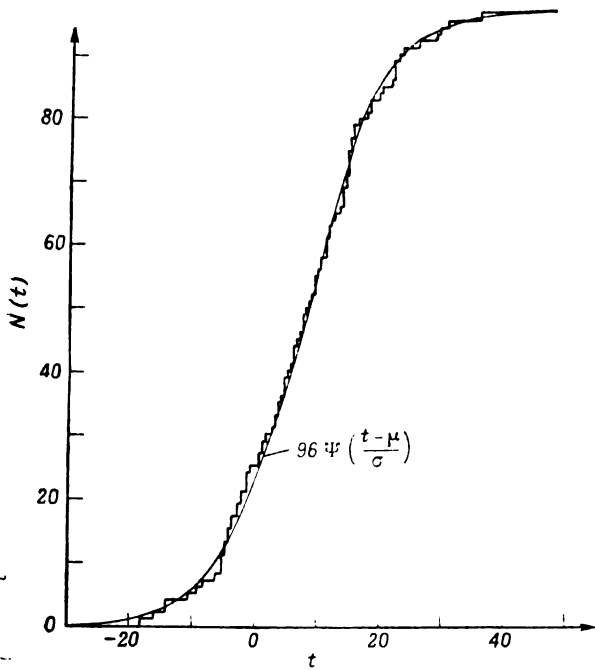
(проверьте), то *из теоремы Бернулли (8.1.13) следует, что $F(t) \xrightarrow{\text{по вер.}} \Phi(t)$, т. е. следует ожидать, что при каждом фиксированном*

значении t и при больших значениях n наблюдаемое значение $F(t)$ будет очень близким к значению $\Phi(t)$. Этот метод позволяет наиболее детально сравнить теорию с результатами наблюдений. Однако при больших значениях n вычерчивать интегральную кривую затруднительно; но в таких случаях можно группировать наблюдаемые данные по интервалам малой длины Δt и затем вычислять полигон сумм.

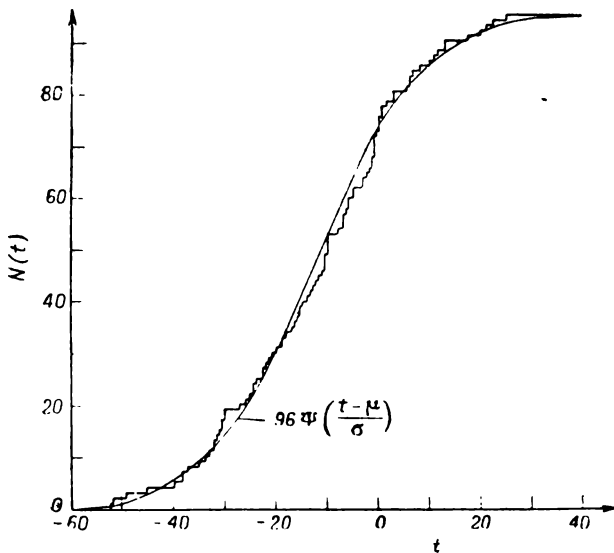
Пример. На фиг. 12 и 13 приведены два полигона сумм для боковых отклонений и отклонений по высоте 96 попаданий в цель по данным примера § 10.5 (деление на n вновь опущено). Для сравнения на графиках нанесены соответствующие теоретические кривые, т. е.

$96\Psi\left(\frac{t-\mu}{\sigma}\right)$, с указанными выше значениями параметров μ и σ . Легко видеть, что согласие очень хорошее и много лучше, чем это можно было усмотреть на ранее приведенных фигурах, особенно фиг. 10 и 11.

§ 10.7. Для распределений Кэптейна, определенных формулой (10.3.2) и являющихся обобщениями нормального распределения, существует особенно простой метод сравнения данных наблюдений с теорией. Поскольку $u = \Psi(t)$ есть функция, всюду возрастающая



Ф и г. 12. Боковое отклонение.



Ф и г. 13. Отклонение по высоте.

при $-\infty < t < \infty$, можно определить t как функцию от u , $t = \Psi^{-1}(u)$ ¹⁾. Так, например, для нормального распределения общего вида $u = \Psi\left(\frac{t - \mu}{\sigma}\right)$ мы получим $\frac{t - \mu}{\sigma} = \Psi^{-1}(u)$. Подставляя сюда $u = \Psi(v_1)$, мы получим функцию $\frac{t - \mu}{\sigma} = \Psi^{-1}[\Psi(v_1)]$, равную $v_1 = \frac{t - \mu}{\sigma}$. Если, далее, образовать $v_2 = \Psi^{-1}[F(t)]$, где $F(t)$ определена формулой (10.6.1), то можно ожидать, что эта функция будет очень близка к прямой $v_1 = \frac{t - \mu}{\sigma}$, проходящей через точку $(\mu, 0)$ и имеющей наклон $1/\sigma$. Этот метод, который можно назвать методом *выпрямленной диаграммы*, очень удобен на практике. Поскольку нормальное распределение протабулировано лишь для значений $\mu = 0$ и $\sigma = 1$, описанный метод значительно сокращает вычисления. (Существует также так называемая вероятностная бумага, на которой указанное преобразование осуществляется графически).

Если нормально распределена не сама величина x , а некоторая функция этой величины $G(x)$, то можно ожидать, что значения функции $v_2 = \Psi^{-1}[F(t)]$ будут лежать очень близко к кривой $v_1 = \frac{G(t) - \mu}{\sigma}$. Таким образом, выпрямленная диаграмма непосредственно показывает, какую следует выбрать функцию от x . Для самого нормального распределения кривая v_1 , как уже было указано, есть прямая, что особенно удобно, так как глаз человека крайне чувствителен даже к очень маленьким отклонениям от прямой линии. Другим большим преимуществом выпрямленной диаграммы в этом последнем случае является то, что и без знания численных значений параметров прямая диаграмма позволяет проверить нормальный *характер* функции распределения, для чего достаточно посмотреть, близка ли полученная кривая к прямой линии. Более того, выпрямленная диаграмма может быть использована для графического получения численных значений параметров μ и σ .

Пример 1. На фиг. 14 и 15 изображены выпрямленные диаграммы для боковых отклонений и отклонений по высоте 96 попаданий в цель при стрельбе из пулемета (см. данные из примера § 10.5). Для сравнения на графики нанесены соответствующие прямые линии $v = \frac{t - \mu}{\sigma}$. Вновь видно, что согласие эмпирических и теоретических графиков очень хорошее и много лучше, чем это можно усмотреть по фиг. 8 и 9 или 10 и 11.

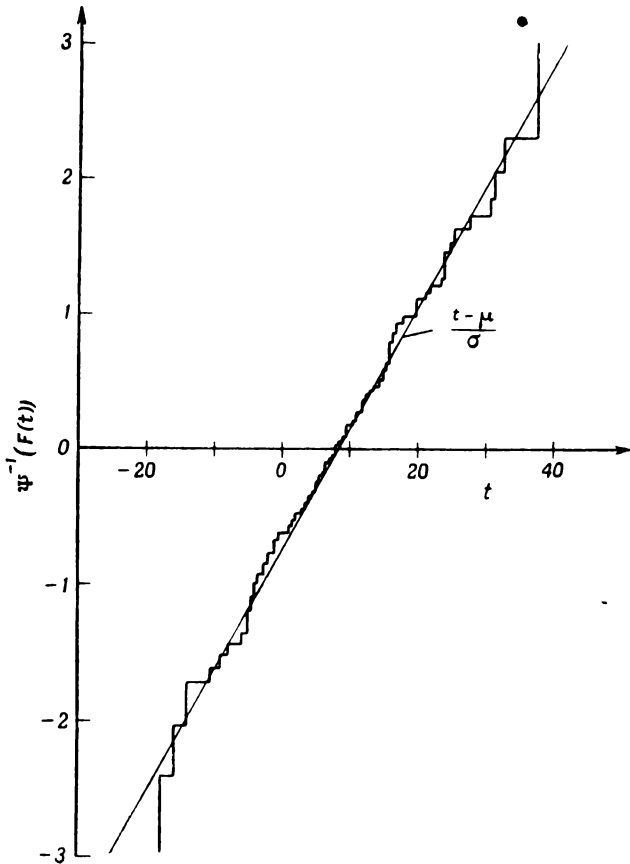
У п р а ж н е н и е. Отметим, что на фиг. 14 и 15 статистические флуктуации на концах нанесенных ломаных заметно больше, чем в окрестности значения

¹⁾ Значения этой функции протабулированы в книге Fisher и Yates, *Statistical Tables*, Table IX.

$t = \mu$. Покажите, что этого следовало ожидать, исходя из следующих соотношений:

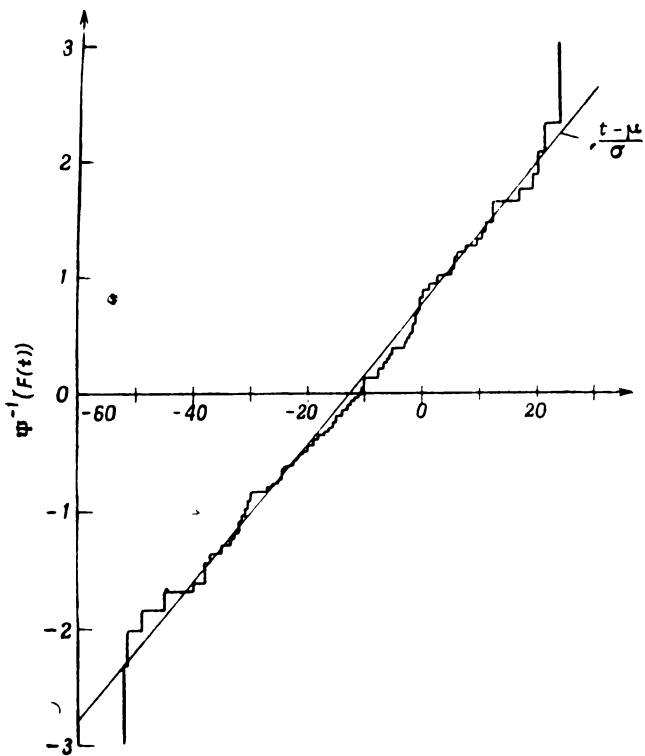
$$\sigma^2 \{ \Psi^{-1} [F(t)] \} \sim \frac{1}{\psi^2(v)} \frac{\Psi(v) [1 - \Psi(v)]}{n} \xrightarrow[|v| \rightarrow \infty]{} \infty, \quad v = \frac{t - \mu}{\sigma}. \quad (1)$$

Пример 2. Гистограмма на фиг. 16 изображает распределение величины расходования электроэнергии, измеряемой в часах работы электроприборов при максимальной нагрузке, 750-ю потребителями (на графике $\Delta t = 100$ часов).

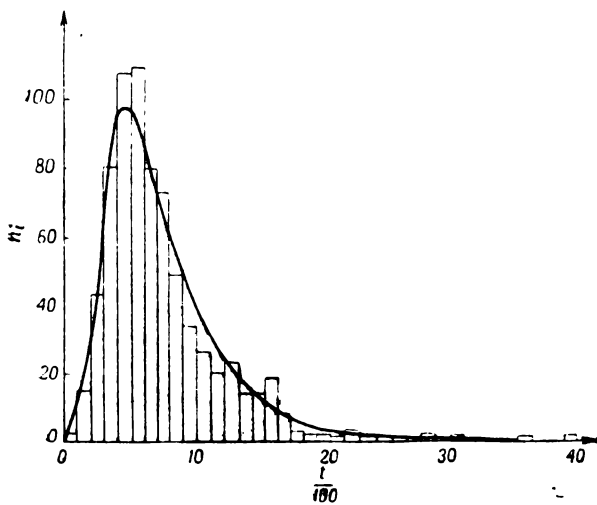


Ф и г. 14. Боковое отклонение.

Соответствующая выпрямленная диаграмма, приведенная на фиг. 17, очевидно, не близка к прямой линии, а более напоминает логарифмическую кривую. Действительно, если начертить выпрямленную диаграмму, приняв на оси t логарифмический масштаб, то диаграмма будет весьма близка к прямой, как это показано на фиг. 18. Определяя по этому

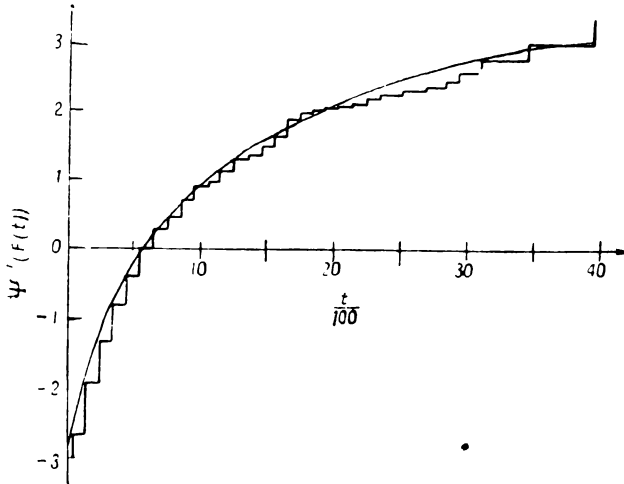


Ф и г. 15. Отклонение по высоте.

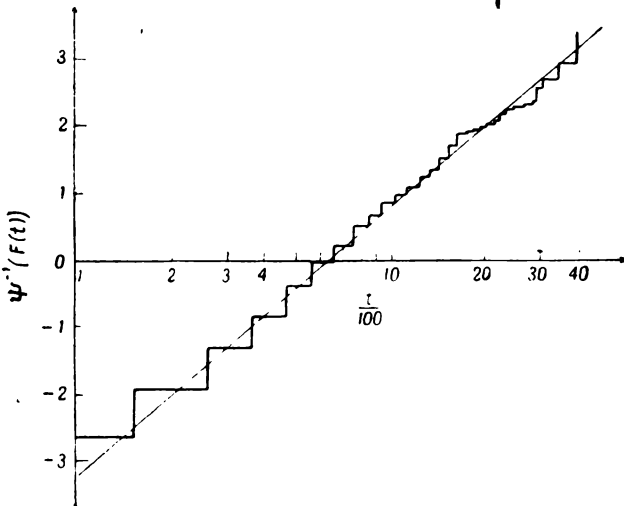


Ф и г. 16.

последнему графику значений μ и σ и вычисляя $n\Delta t\varphi(t)$ по формуле (10.3.3), мы получим теоретическую кривую, которая нанесена на



Фиг. 17.



Фиг. 18.

фиг. 16. Как фиг. 16, так и фиг. 18 показывают, что рассматриваемое распределение можно считать логарифмически нормальным.

§ 10.8. По графикам гистограммы и соответствующей теоретической кривой (или f_i и φ_i в случае дискретных распределений) можно оценить согласие между теорией и данными наблюдений. При этом

всякое достаточно большое несоответствие обнаруживается легко, но по одному такому графику трудно оценить количественно, отклоняются ли данные наблюдений от теоретической кривой более, чем этого следовало бы ожидать вследствие статистических колебаний¹⁾. Далее, на основании формулы Лапласа (8.2.1) различные частоты n_i и $N(t)$ при каждом фиксированном значении i и t и при больших значениях n распределены приблизительно нормально. Исходя из выражений (10.4.3), (10.5.5) и (10.6.3) для дисперсий с помощью табл. II (в конце книги) можно вычислить доверительные пределы для частот при любом заданном доверительном уровне P и затем убедиться, попадают ли наблюдаемые частоты в эти пределы. В статистической практике принято использовать пятипроцентные пределы, которые согласно (7.4.4) весьма близки к пределам ($\mu - 2\sigma$, $\mu + 2\sigma$). Эти доверительные пределы могут быть нанесены непосредственно на сами графики, причем появляется хорошая возможность оценки степени согласия между теорией и наблюдаемыми данными. Заметим, что выбор доверительного уровня 50%, конечно, совершенно произволен и основан исключительно на том, что такой доверительный уровень на практике оказался наиболее удобным.

Как правило, φ_i , $\varphi(\xi_i)\Delta t_i$ и $\Phi(t)$ суть малые числа, и поскольку они входят в формулы для стандартных отклонений, вытекающие из формул (10.4.3), (10.5.5) и (10.6.3), только под знаком квадратного корня, то для грубой оценки доверительных пределов можно полагать

$$\sigma\{n_i\} = \sqrt{n\varphi_i} \sqrt{1 - \varphi_i} = \sqrt{\mathfrak{M}\{n_i\}} \sqrt{1 - \varphi_i} \sim \sqrt{n_i}, \quad (1)$$

$$\begin{aligned} \sigma\{n_i\} &= \sqrt{n\varphi(\xi_i)\Delta t_i} \sqrt{1 - \varphi(\xi_i)\Delta t_i} = \\ &= \sqrt{\mathfrak{M}\{n_i\}} \sqrt{1 - \varphi(\xi_i)\Delta t_i} \sim \sqrt{n_i}, \end{aligned} \quad (2)$$

$$\sigma\{N(t)\} = \sqrt{n\Phi(t)} \sqrt{1 - \Phi(t)} = \sqrt{\mathfrak{M}\{N(t)\}} \sqrt{1 - \Phi(t)} \sim \sqrt{N(t)}. \quad (3)$$

При этом предполагается, что фактически наблюдаемые величины лишь мало отклоняются от соответствующих математических ожиданий. Это означает, что стандартные отклонения можно оценивать непосредственно по наблюдениям, что очень удобно, особенно тогда, когда теоретические распределения заранее неизвестны.

Пример 1. На фиг. 7 вертикальные отрезки обозначают интервалы ($n_i - 2\sqrt{n_i}$, $n_i + 2\sqrt{n_i}$). Видно, что теоретическая кривая проходит между этими пределами.

Пример 2. Часто результаты наблюдений пишут в виде

$$n_i \pm \sqrt{n_i}, \quad (4)$$

но следует заметить, что такая запись может вводить в заблуждение, создавая впечатление, что вне указанного интервала не может ока-

¹⁾ Действительно, поскольку невероятно, чтобы все наблюдаемые точки попали точно на теоретическую кривую, иногда считают, что слишком хорошее согласие также означает, что теоретическая кривая не дает приемлемого описания наблюдаемых данных.

заться ни одного значения, хотя на самом деле этот интервал является лишь оценкой доверительного интервала с определенным доверительным уровнем, выбранным совершенно произвольно и приблизительно равным $\frac{2}{3}$ (см. (7.4.3)).

Пример 3. Вместо пятипроцентных доверительных пределов, равных приблизительно $\pm 2\sqrt{n_i}$, некоторые авторы используют также, более по сложившейся традиции, „вероятные“ пределы, т. е. пятидесятипроцентные доверительные пределы, равные приблизительно $\pm \frac{2}{3}\sqrt{n_i}$. Однако такой выбор пределов очень неудобен, так как при этом выборе допустимые статистические флуктуации переоцениваются, и в результате наблюдения считаются более точными, чем они суть на самом деле. Это может быть опасно, например, если результаты наблюдений используются для оценки различия между двумя теоретическими распределениями.

В рассмотренных до сих пор методах каждая относительная частота изучалась независимо. Очевидно, эмпирическое распределение в целом есть случайная величина, хотя и многомерная. С помощью рассмотрения всего эмпирического распределения в целом были выработаны другие методы для количественной оценки степени согласия теоретического и эмпирического распределений. Однако обсуждение этих методов, дающих возможность получить объективный критерий для бракования теоретических распределений, когда их несогласованность с данными наблюдений столь велика, что их нельзя считать правдоподобными, выходит за рамки настоящей книги ¹⁾.

§ 10.9. Рассмотрим теперь третий вопрос из § 10.2, касающийся оценки параметров. Эта проблема является в математической статистике основной. Рассмотрим случайную величину x , одномерную или многомерную, и предположим, что с этой величиной x связана гипотетическая функция распределения $\Phi_x(x)$ определенного математического вида, но содержащая один или несколько параметров $\theta_1, \dots, \theta_k$ ²⁾. Тогда возникает задача получения оценки для „наилучших“ численных значений параметров $\theta_1, \dots, \theta_k$ по выборке, состоящей из некоторого количества n результатов наблюдений x_1, x_2, \dots, x_n . Прежде

¹⁾ См., например, Fisher, *Statistical Methods for Research Workers*, гл. IV, и Г. Крамер, *Математические методы статистики*, гл. 30, 1948.

²⁾ В целях удобства в остальной части книги мы используем одну и ту же букву для случайной величины x и для аргумента в соответствующей функции распределения x, \dots (Для многомерных случайных величин x и т. д. суть векторы, имеющие компоненты (x_1, x_2, \dots) и т. д. Для простоты в этой главе мы ограничимся рассмотрением лишь одномерных случайных величин, но обобщение на случай многомерных величин производится непосредственно.) Далее, мы будем обозначать наблюдаемые значения буквами x_1, x_2, \dots . Напоминаем также, что все теоретические понятия обозначаются греческими буквами, а соответствующие эмпирические величины — соответствующими латинскими буквами, например θ и t .

всего, очевидно, что любая оценка t для какого-либо параметра θ будет некоторой функцией $t = t(x_1, \dots, x_n)$ от выборочных значений x_1, \dots, x_n , но не от параметров $\theta_1, \dots, \theta_k$. Таким образом, оценка t сама есть случайная величина, так как вследствие статистических флуктуаций новая выборка даст новое значение оценки t . Такая функция называется *статистикой*¹⁾, и мы будем писать

$$\theta \approx t, \quad (1)$$

что означает: оценкой для параметра θ является t . Если задана функция распределения величины x , то однозначно определена и функция распределения для любой статистики, причем эта последняя функция распределения будет, как правило, также зависеть от параметров $\theta_1, \dots, \theta_k$. Идеальной оценкой для какого-либо из параметров θ была бы, конечно, такая статистика t , которая при всех значениях n имела бы несобственную функцию распределения $\varepsilon(t - \theta)$, определенную формулой (4.3.3), так как в этом случае каждое измерение величины t будет с несомненностью давать в результате искомое значение θ . Однако на практике при любом *конечном* значении n такое идеальное положение недостижимо, и приблизиться к идеальному положению можно лишь при $n \rightarrow \infty$. Поэтому естественно считать статистику t подходящей оценкой для θ , если выполняется лишь следующее требование: $t \xrightarrow[n \rightarrow \infty]{\text{по вер.}}$ θ . Статистики, удовлетворяющие этому требованию, называются *состоятельными* оценками; из критерия А § 9.2 следует, что *наблюдяное значение состоятельной оценки t при больших значениях n практически будет равно „истинному“ значению параметра θ* . Из неравенства Чебышева (8.1.4) следует, что достаточным (но не необходимым) условием для того, чтобы статистика t была состоятельной оценкой для параметра θ , является выполнение соотношений $\mathbb{M}\{t\} \xrightarrow[n \rightarrow \infty]{} \theta$ и $\sigma^2\{t\} \xrightarrow[n \rightarrow \infty]{} 0$. В частности, если при всех n имеем $\mathbb{M}\{t\} = \theta$, то статистика t называется *несмещенной* оценкой для параметра θ .

Но состоятельность оценки не дает каких-либо сведений о свойствах статистики t при сравнительно малых значениях n , обычно встречающихся на практике. Более того, при одной и той же функции распределения $\Phi_x(t)$ для одного и того же параметра θ существует бесконечное множество состоятельных оценок; например, если t есть состоятельная оценка для параметра θ , то, очевидно, состоятельной оценкой будет и $g_n t$, где g_n — такая величина, что $g_n \xrightarrow[n \rightarrow \infty]{} 1$. Чтобы исключить эту тривиальную многозначность, естественно „нормировать“ оценки*), требуя, чтобы они были несмещенными. Более того,

¹⁾ Большинство из употребляемых здесь терминов введено Фишером. См. его работы: *Mathematical Foundations of Theoretical statistics and Theory of Statistical Estimation*.

*) В том случае, конечно, когда это возможно. (*Прим. ред.*)

состоятельную несмещенную статистику t_1 естественно считать более хорошей оценкой для θ по сравнению с другой состоятельной несмещенной статистикой t_2 , если в распределении вероятностей для t_1 масса вероятности более концентрирована около значения θ , чем в распределении вероятностей для t_2 ; при этом степень концентрации можно описывать любой мерой рассеяния. Рассмотрим в качестве меры рассеяния лишь стандартное отклонение σ , определяемое с помощью формулы $\sigma^2 \{t\} = \mathfrak{M} \{(t - \mathfrak{M} \{t\})^2\} = \mathfrak{M} \{t - \theta\}^2$. При весьма общих условиях можно показать ¹⁾, что при любых фиксированных значениях $\theta_1, \dots, \theta_k$ и n дисперсия $\sigma^2 \{t\}$ всегда превосходит некоторое фиксированное число. В частности, если имеется лишь один неизвестный параметр, то для любой *несмещенной* оценки

$$\sigma^2 \{t\} \geq \sigma_0^2 = -\frac{1}{n \mathfrak{M} \left\{ \frac{\partial^2 \ln \varphi}{\partial \theta^2} \right\}} > 0, \quad (2)$$

где φ в случае непрерывного распределения есть плотность вероятности $\varphi = \varphi(x, \theta)$, а в случае дискретного распределения — вероятность $\varphi = \varphi_i(\theta)$. Если существует такая (несмещенная) оценка, для которой в формуле (2) достигается знак равенства, то эта оценка, очевидно, является наилучшей возможной оценкой; она называется тогда *эффективной* оценкой $t_{\text{эфф}}$ (из формулы (2) и условия $\mathfrak{M} \{t_{\text{эфф}}\} = \theta$ следует, что эффективная оценка $t_{\text{эфф}}$ является также состоятельной). Для других несмещенных оценок имеем

$$0 \leq \frac{\sigma_0^2}{\sigma^2 \{t\}} = e(t) \leq 1, \quad (3)$$

где e — так называемая *эффективность* оценки t .

Иногда использование эффективной оценки может оказаться неудобным, так как оно может приводить к усложнению расчетов; но дисперсии других оценок будут больше, так что точность их будет меньше, хотя эта излишняя неточность и уменьшается при возрастании числа используемых наблюдений.

Пример. В распределении Кэптейна (10.3.2) будем считать параметр σ известным, а параметр μ неизвестным, подлежащим оценке. Тогда по формуле (2) получаем

$$\begin{aligned} \sigma_0^2 &= \frac{-1}{n \mathfrak{M} \left\{ \frac{\partial^2}{\partial \mu^2} \left[-\ln(\sqrt{2\pi}\sigma) - \frac{(G(x) - \mu)^2}{2\sigma^2} + \ln G'(x) \right] \right\}} = \\ &= \frac{-1}{n \mathfrak{M} \left\{ \frac{-1}{\sigma^2} \right\}} = \frac{\sigma^2}{n}. \end{aligned} \quad (4)$$

¹⁾ Условия, при которых справедливы теоремы, формулируемые в этом параграфе, а также доказательство этих теорем читатель найдет в книге: Крамер, Математические методы статистики, гл. 32—34, 1948.

Используя в качестве оценки для μ среднее значение (7.7.4), т. е. $m = \overline{G(x)}$, мы увидим, сравнив формулы (7.7.7) и (4), что m есть несмещенная эффективная оценка для μ . Другие возможные оценки для μ будут иметь больше дисперсии, чем оценка m . Это можно проверить непосредственно в каждом конкретном случае.

* Упражнение 1. Покажите, что если в распределении Кэптейна (10.3.2) считать μ известным параметром, а $\theta = \sigma^2$ неизвестным параметром, подлежащим оценке, то формула (2) дает

$$\sigma^2 \{t\} \geq \sigma_0^2 = \frac{2\sigma^4}{n}. \quad (5)$$

Пусть

$$s_0^2 = \overline{[G(x) - \mu]^2} = \frac{[G(x_1) - \mu]^2 + \dots + [G(x_n) - \mu]^2}{n}. \quad (6)$$

Покажите, применив приложение 1 и приняв $y = G(x)$ за новую переменную, что

$$\mathfrak{M}\{s_0^2\} = \sigma^2, \quad \sigma^2 \{s_0^2\} = \frac{2\sigma^4}{n}. \quad (7)$$

Таким образом, s_0^2 есть несмещенная эффективная оценка для параметра $\theta = \sigma^2$. Пусть, далее,

$$s_1^2 = \overline{[G(x) - m]^2} = \frac{[G(x_1) - m]^2 + \dots + [G(x_n) - m]^2}{n}, \quad (8)$$

где $m = G(x)$. Положив $y = G(x)$, покажите, что $\frac{ns_1^2}{\sigma^2}$ имеет распределение (7.8.2) с $f = n - 1$ (см. упражнение 3 § 7.8). Используя (7.8.3), покажите, что

$$\mathfrak{M}\{s_1^2\} = \frac{n-1}{n} \sigma^2, \quad \sigma^2 \{s_1^2\} = \frac{n-1}{n} \frac{2\sigma^4}{n}. \quad (9)$$

Таким образом, $s^2 = \frac{n}{n-1} s_1^2$ (величина, введенная в (7.7.16)) есть несмещенная оценка для σ^2 . Покажите, что эффективность оценки s^2 равна $\frac{n-1}{n}$.

* Упражнение 2. Покажите, что если в распределении Кэптейна (10.3.2) считать μ известным параметром, а неизвестным и подлежащим оценке параметром считать не σ^2 , а $\theta = \sigma$, то формула (2) дает

$$\sigma^2 \{t\} \geq \sigma_0^2 = \frac{\sigma^2}{2n}. \quad (10)$$

Полагая $y = G(x)$, покажите, что величина $\sqrt{n} s_0$, где s_0 определено формулой (6), имеет распределение (7.7.12) с $f = n$ (см. упражнение 3 § 7.8). С помощью формул (7.7.14) и (7.7.15) покажите, что

$$s'_0 = \sqrt{\frac{n}{2} \frac{\left(\frac{n-2}{2}\right)!}{\left(\frac{n-1}{2}\right)!}} s_0 \sim \sqrt{\frac{n}{n-\frac{1}{2}}} s_0 \quad (11)$$

есть несмещенная оценка для σ с дисперсией

$$\sigma^2 \{s'_0\} = \left\{ \frac{n}{2} \left[\frac{\left(\frac{n-2}{2}\right)!}{\left(\frac{n-1}{2}\right)!} \right]^2 - 1 \right\} \sigma^2 = \frac{\sigma^2}{2n} + O\left(\frac{1}{n^2}\right). \quad (12)$$

Таким образом, эффективность $e\{s'_0\}$ при $n \rightarrow \infty$ стремится к единице, и при том довольно быстро. Покажите, что при $n=2$ формула (12) дает $e\{s'_0\} = \frac{\pi}{4(4-\pi)} = 0,915$, а при $n=3$ $e\{s'_0\} = \frac{4}{3(3\pi-8)} = 0,936$. Покажите, что величина $\sqrt{ns_1}$, определяемая формулой (8), имеет распределение (7.7.12) с $f = n-1$. Далее покажите, используя формулы (7.7.14) и (7.7.15), что

$$s'_1 = \sqrt{\frac{n}{2} \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{n-2}{2}\right)!}} s_1 \sim \sqrt{\frac{n}{n-\frac{3}{2}}} s_1 = \sqrt{\frac{n-1}{n-\frac{3}{2}}} s \quad (13)$$

есть также несмещенная оценка для σ , дисперсия которой равна

$$\sigma^2 \{s'_1\} = \left\{ \frac{n}{2} \left[\frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{n-2}{2}\right)!} \right]^2 - 1 \right\} \sigma^2 = \frac{\sigma^2}{2n} + O\left(\frac{1}{n^2}\right). \quad (14)$$

Таким образом, эффективность $e\{s'_1\}$ также стремится к единице при $n \rightarrow \infty$, но медленнее, чем $e\{s'_0\}$. Покажите, что при $n=2$ формула (14) дает $e\{s'_1\} = \frac{1}{2(\pi-2)} = 0,438$, а при $n=3$ $e\{s'_1\} = \frac{\pi}{6(4-\pi)} = 0,610$.

Покажите, наконец, что для величины

$$s_2 = \sqrt{\frac{\pi}{2} |G(x) - \mu|} = \sqrt{\frac{\pi}{2} \frac{|G(x_1) - \mu| + \dots + |G(x_n) - \mu|}{n}} \quad (15)$$

имеем

$$\mathcal{M}\{s_2\} = \sigma, \quad \sigma^2 \{s_2\} = (\pi-2) \frac{\sigma^2}{2n}. \quad (16)$$

Таким образом, s_2 также является несмещенной оценкой для σ , но с эффективностью $1/(\pi-2) = 0,876$.

Упражнение 3. В упражнениях 1 и 2 мы видели, что если t есть несмещенная эффективная оценка для параметра θ , то $f(t)$ не обязательно будет несмещенной эффективной оценкой для $f(\theta)$. Покажите, однако, что если t

есть состоятельная оценка для θ , т. е. $t \xrightarrow[n \rightarrow \infty]{\text{по вер.}} \theta$, то $f(t)$ есть состоятельная

оценка для $f(\theta)$, т. е. $f(t) \xrightarrow[n \rightarrow \infty]{\text{по вер.}} f(\theta)$ *).

*) Это верно в случае, когда $f(t)$ непрерывна. (Прим. ред.)

У п р а ж н е н и е 4. Покажите, что если в биномиальном распределении (4.3.4) считать ν известным параметром, а θ — неизвестным параметром, подлежащим оценке, то формула (2) дает

$$\sigma^2 \{t\} \geq \sigma_0^2 = \frac{\theta(1-\theta)}{n\nu}. \quad (17)$$

Пусть

$$t = \frac{\bar{x}}{\nu} = \frac{x_1 + \dots + x_n}{\nu n}. \quad (18)$$

Покажите, что t есть несмещенная эффективная оценка для θ .

У п р а ж н е н и е 5. Покажите, что если в распределении Пуассона (4.3.5) считать μ неизвестным параметром, подлежащим оценке, то формула (2) дает

$$\sigma^2 \{t\} \geq \sigma_0^2 = \frac{\mu}{n}. \quad (19)$$

Пусть

$$m = \bar{x} = \frac{x_1 + \dots + x_n}{n}. \quad (20)$$

Покажите, что m есть несмещенная эффективная оценка для θ .

*** § 10.10.** На практике для получения состоятельных оценок используются два метода, хотя в отдельных случаях могут быть удобными также и другие оценки, для получения которых эти методы непригодны. В настоящем параграфе мы рассмотрим *метод максимума правдоподобия*, развитый в основном Р. А. Фишером, хотя до Фишера этот метод использовался в частных случаях еще Гауссом. В § 10.12 будет рассмотрен *метод моментов*, введенный Пирсоном.

Образум вначале вероятность появления данной выборки, предполагая, что n выборочных значений x_1, \dots, x_n независимы. В дискретном случае эта вероятность равна

$$P(x_1, \dots, x_n) = \varphi_{x_1} \varphi_{x_2} \dots \varphi_{x_n}; \quad (1)$$

в непрерывном случае имеем

$$P(x_1, \dots, x_n) dx_1 \dots dx_n = \varphi(x_1) \varphi(x_2) \dots \varphi(x_n) dx_1 \dots dx_n. \quad (2)$$

В общих случаях P есть функция от x_1, \dots, x_n и также от параметров $\theta_1, \dots, \theta_k$. Будем считать выборочные значения x_1, \dots, x_n фиксированными, а $\theta_1, \dots, \theta_k$ — переменными величинами. Тогда функция

$$P(x_1, \dots, x_n; \theta_1, \dots, \theta_k) \geq 0 \quad (3)$$

в обоих случаях будет называться *функцией правдоподобия*. Очевидно, что если значения величин $\theta_1, \dots, \theta_k$ таковы, что P очень мало, то согласно критерию А § 9.2 нельзя ожидать фактического появления данной выборки.

Такая гипотеза о значениях параметров $\theta_1, \dots, \theta_k$ должна быть поэтому отвергнута, как неправдоподобная. *Метод максимума правдоподобия заключается в том, что наилучшие возможные оценки*

для параметров $\theta_1, \dots, \theta_k$ определяются, как те значения переменных аргументов функции правдоподобия; при которых она достигает максимального значения. Другими словами, наилучшие оценки $t_1(x_1, \dots, x_n), \dots, t_k(x_1, \dots, x_n)$ для параметров $\theta_1, \dots, \theta_k$ получаются из условия $P = \max$ или, что то же самое, из условия

$$\ln P = L(x_1, \dots, x_n; \theta_1, \dots, \theta_k) = \max. \quad (4)$$

Таким образом, вообще говоря, оценки t_1, \dots, t_k получаются как решения k уравнений правдоподобия¹⁾

$$\frac{\partial L}{\partial \theta_1} = \frac{\partial L}{\partial \theta_2} = \dots = \frac{\partial L}{\partial \theta_k} = 0. \quad (5)$$

Решения уравнений правдоподобия, не являющиеся постоянными (т. е. существенно зависящие от выборочных значений x_1, \dots, x_n), называются *оценками максимального правдоподобия* и имеют при весьма общих условиях следующие свойства (в случае, когда имеется лишь один неизвестный параметр, т. е. $k=1$):

Теорема I. Если существует эффективная (и потому несмещенная и состоятельная) оценка $t_{эфф}$ для параметра θ , то уравнение правдоподобия имеет единственное решение, равное $t_{эфф}$.

Теорема II. Уравнение правдоподобия имеет решение $t_{пр}$, являющееся состоятельной оценкой для θ , т. е. такое, что $t_{пр} \xrightarrow[n \rightarrow \infty]{\text{по вер.}} \theta$. Это решение имеет асимптотически нормальное распределение с параметрами $\mu = \theta$ и $\sigma^2 = \sigma_0^2$, где σ_0^2 определено формулой (10.9.2).

Говорят, что оценка максимального правдоподобия $t_{пр}$, существование которой устанавливает теорема II, есть *асимптотически эффективная* оценка для θ . Подчеркнем, что при любом конечном значении n ни $\mathfrak{M}\{t_{пр}\}$, ни $\sigma^2\{t_{пр}\}$ не обязаны существовать. Более того, если $\mathfrak{M}\{t_{пр}\}$ существует, то это математическое ожидание не обязательно равняется θ , так что $t_{пр}$ необязательно является несмещенной оценкой. Иногда бывает удобно сделать такую оценку несмещенной или, может быть, умножить ее на какой-либо множитель g_n , такой, что $g_n \sim 1$ и $g_n \xrightarrow[n \rightarrow \infty]{} 1$ (такой множитель не меняет коэффициента изменчивости).

Упражнение 1. Покажите, что формулу (10.9.2) можно записать также в виде

$$\sigma_0^2 = \frac{-1}{\mathfrak{M}\left\{\frac{\partial^2 L}{\partial \theta^2}\right\}}, \quad (6)$$

где L определено формулой (4).

¹⁾ Если φ имеет при некоторых значениях параметров точки разрыва, то может оказаться, что одна из этих точек и дает максимум функции L (см. задачу 54).

Пример 1. Для биномиального распределения (4.3.4) функция правдоподобия имеет вид

$$P(x_1, \dots, x_n; \theta) = \\ = \binom{v}{x_1} \theta^{x_1} (1-\theta)^{v-x_1} \binom{v}{x_2} \theta^{x_2} (1-\theta)^{v-x_2} \dots \binom{v}{x_n} \theta^{x_n} (1-\theta)^{v-x_n}. \quad (7)$$

Будем считать здесь v известным, как это обычно имеет место, и тогда оценка максимального правдоподобия для θ согласно (5) должна быть определена из уравнения

$$\frac{\partial L}{\partial v} = \frac{\partial}{\partial \theta} \left[\ln \prod_{i=1}^n \binom{v}{x_i} + \right. \\ \left. + (x_1 + \dots + x_n) \ln \theta + (nv - (x_1 + \dots + x_n)) \ln (1-\theta) \right] = \\ = \frac{x_1 + \dots + x_n}{\theta} - \frac{nv - (x_1 + \dots + x_n)}{1-\theta} = 0, \quad (8)$$

откуда

$$\theta \approx t = \frac{x_1 + \dots + x_n}{nv} = \frac{\bar{x}}{v}.$$

Проверьте теоремы I и II (см. также результат, полученный в упражнении 4 § 10.9).

Пример 2. Для распределения Пуассона (4.3.5) функция правдоподобия имеет вид

$$P(x_1, \dots, x_n; \mu) = \frac{e^{-\mu} \mu^{x_1}}{x_1!} \cdot \frac{e^{-\mu} \mu^{x_2}}{x_2!} \dots \frac{e^{-\mu} \mu^{x_n}}{x_n!}. \quad (9)$$

Согласно (5) оценка максимального правдоподобия для μ должна быть определена из уравнения

$$\frac{\partial L}{\partial \mu} = \frac{\partial}{\partial \mu} \left[-n\mu + (x_1 + \dots + x_n) \ln \mu - \ln \prod_{i=1}^n x_i! \right] = \\ = -n + \frac{x_1 + \dots + x_n}{\mu} = 0, \quad (10)$$

откуда

$$\mu \approx m = \frac{x_1 + \dots + x_n}{n} = \bar{x}.$$

Проверьте теоремы I и II (см. также результат, полученный в упражнении 5 § 10.9).

Пример 3. Часто среди значений x_1, \dots, x_n в выражениях (8) или (10) возможные значения 0, 1, 2... встречаются по несколько раз. В таких случаях вычисления можно упростить, подсчитав числа n_i появлений результатов i . Тогда $\sum n_i = n$ и

$$\bar{x} = \frac{x_1 + \dots + x_n}{n} = \frac{\sum_i n_i i}{\sum_i n_i}. \quad (11)$$

Например, используя значения из примера § 10.4, по формуле (11) получим

$$\mu \approx m = \frac{10097}{2608} = 3,87.$$

Теоремы I и II можно обобщить на случай наличия нескольких параметров. Если рассмотреть сначала частный случай, в котором

$$\mathfrak{M} \left\{ \frac{\partial^2 L}{\partial \theta_i \partial \theta_k} \right\} = 0 \text{ при всех } i \neq k, \quad (12)$$

то обобщение формулы (10.9.2) имеет вид при $\theta_i \approx t_i$

$$\sigma^2 \{t_i\} \geq \sigma_{0i}^2 = \frac{1}{n \mathfrak{M} \left\{ \frac{\partial^2 \ln \varphi}{\partial \theta_i^2} \right\}} > 0, \quad i = 1, 2, \dots, k. \quad (13)$$

Упражнение 2. Покажите, что формула (13) может быть записана также в виде

$$\sigma_{0i}^2 = \frac{-1}{\mathfrak{M} \left\{ \frac{\partial^2 L}{\partial \theta_i^2} \right\}}, \quad (14)$$

где функция L определена формулой (4).

Множество оценок t_1, \dots, t_k , для которого при всех t_i в формуле (13) достигается знак равенства, называется множеством *совместно-эффективных оценок*. Теоремы I и II с заменой слова „эффективный“ на „совместно-эффективные“ остаются справедливыми для случая нескольких параметров, причем если условие (12) выполнено, то минимальные дисперсии оценок определяются формулой (13).

Если условие (12) не выполняется, то следует рассмотреть матрицу вторых моментов $M_0^{(T)}$ для множества совместно-эффективных оценок t_1, \dots, t_k (см. пример 2 § 6.4) и заменить в формуле (13) величины σ_{0i}^2 на диагональные элементы матрицы $M_0^{(T)}$, получаемой, в результате естественного обобщения формулы (14), в виде

$$M_0^{(T)} = \mathfrak{M} \{ (t_r - \theta_r) (t_s - \theta_s) \} = - \left\{ \mathfrak{M} \left\{ \frac{\partial^2 L}{\partial \theta_r \partial \theta_s} \right\} \right\}^{-1}. \quad (15)$$

Упражнение 3. Покажите, что при выполнении условия (12) формула (15) превращается в формулу (14).

Пример 4. Для распределения Кэптейна (10.3.2) функция правдоподобия имеет вид

$$\begin{aligned} & P(x_1, \dots, x_n; \mu, \sigma) = \\ & = \left(\frac{1}{V 2\pi \sigma} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} [(G(x_1) - \mu)^2 + \dots + (G(x_n) - \mu)^2] \right\} \times \\ & \quad \times G'(x_1) \dots G'(x_n). \end{aligned} \quad (16)$$

Согласно формуле (5) оценка максимального правдоподобия для μ определяется из уравнения

$$\begin{aligned} \frac{\partial L}{\partial \mu} &= \frac{\partial}{\partial \mu} \left[-n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (G(x_i) - \mu)^2 + \ln \prod_{i=1}^n G'(x_i) \right] = \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n (G(x_i) - \mu) = 0, \end{aligned} \quad (17)$$

откуда

$$\mu \approx m = \frac{G(x_1) + \dots + G(x_n)}{n} = \overline{G(x)}.$$

Оценка максимального правдоподобия для σ определяется из уравнения

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (G(x_i) - \mu)^2 = 0,$$

откуда

$$\sigma \approx s_1 = \left[\frac{1}{n} \sum_{i=1}^n (G(x_i) - m)^2 \right]^{1/2} = \left[\overline{(G(x) - m)^2} \right]^{1/2}. \quad (18)$$

Наконец, поскольку $\mathfrak{M}\{G(x)\} = \mu$ и $\mathfrak{M}\{(G(x) - \mu)^2\} = \sigma^2$, имеем

$$\begin{aligned} \mathfrak{M}\left\{\frac{\partial^2 L}{\partial \mu^2}\right\} &= \mathfrak{M}\left\{-\frac{n}{\sigma^2}\right\} = -\frac{n}{\sigma^2}, \\ \mathfrak{M}\left\{\frac{\partial^2 L}{\partial \sigma^2}\right\} &= \mathfrak{M}\left\{\frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_{i=1}^n (G(x_i) - \mu)^2\right\} = -\frac{2n}{\sigma^2}, \\ \mathfrak{M}\left\{\frac{\partial^2 L}{\partial \sigma \partial \mu}\right\} &= \mathfrak{M}\left\{-\frac{2}{\sigma^3} \sum_{i=1}^n (G(x_i) - \mu)\right\} = 0. \end{aligned} \quad (19)$$

Итак, согласно (15) имеем

$$\sigma_{0\mu}^2 = \frac{\sigma^2}{n}, \quad \sigma_{0\sigma}^2 = \frac{\sigma^2}{2n}. \quad (20)$$

Проверьте теоремы I и II (см. также результаты примера и упражнений 1 и 2 § 10.9).

Упражнение 4. Покажите, что если в распределении Кэптейна считать подлежащим оценке параметром кроме μ еще $\theta = \sigma^2$ (но не само σ), то оценкой максимального правдоподобия для σ^2 будет s_1^2 , где величина s_1 , определяемая формулой (18), есть оценка максимального правдоподобия для σ . Обобщая этот результат, покажите, что если $\theta \approx t_{\text{пр}}$, то $f(\theta) \approx f(t_{\text{пр}})$. Заметим, что если t есть состоятельная оценка для θ , то $f(t)$ есть также состоятельная оценка для $f(\theta)$, но что свойства несмещенности и эффективности при таком преобразовании могут и не сохраняться (см. уравнение 3 § 10.9).

Из предыдущих примеров мы видим, что если оцениваемый параметр есть среднее значение μ , то наилучшая оценка m для параметра μ есть среднее арифметическое $m = \bar{x}$. Однако это не всегда верно.

Пример 5. Наиболее яркую иллюстрацию последнего утверждения дает распределение Коши (4.4.9). Как было показано в примере 3 § 8.1, \bar{x} имеет то же самое распределение, что и x , так что $m = \bar{x}$ не является даже состоятельной оценкой для μ ; действительно, x не содержит какой-либо информации о значении μ сверх того, что может дать единичное измерение.

Упражнение 5. Покажите, что для распределения Лапласа (4.4.10) с $\mathcal{M}\{x\} = \mu$, $\sigma\{x\} = \sqrt{2}$ α оценками максимального правдоподобия для параметров μ и α являются $\mu \approx m = x_M$, $\alpha \approx a = \sqrt{|x - x_M|} = \frac{|x_1 - x_M| + \dots + |x_n - x_M|}{n}$, где x_M обозначает выборочную медиану (см. стр. 65), т. е. если $x_1 \leq x_2 \leq \dots \leq x_n$, то при нечетном n $x_M = x_k$, где $k = \frac{n+1}{2}$, а при четном n x_M может быть любым числом, заключенным между $x_{n/2}$ и $x_{n/2+1}$.

*** § 10.11.** Для некоторых распределений существуют так называемые *достаточные* оценки, т. е. оценки, содержащие при любом значении n всю информацию об оцениваемом параметре, которую вообще можно извлечь из выборки x_1, \dots, x_n .

Выражаясь математически, такая оценка t_D имеет следующее свойство: если в функции правдоподобия (10.10.3) перейти от переменных x_1, \dots, x_n к переменной $t_D = t_D(x_1, \dots, x_n)$ и $n-1$ другим переменным $u_1(x_1, \dots, x_n), \dots, u_{n-1}(x_1, \dots, x_n)$, то в случае только одного неизвестного параметра θ , этот параметр *не* входит в выражение для условного распределения вероятностей величины u_1, \dots, u_{n-1} при условии, что t_D имеет фиксированное значение t_D . Другими словами, функция правдоподобия может быть записана в виде

$$P(x_1, \dots, x_n; \theta) = g(t_D; \theta) h(u_1, \dots, u_{n-1} | t_D), \quad (1)$$

где g дает частное распределение величины t_D , которое зависит от θ , а h дает условное распределение величин u_1, \dots, u_{n-1} , которое не зависит от θ . Таким образом, если значение t_D известно, то сведения о значениях величин u_1, \dots, u_{n-1} не могут дать никакой дальнейшей информации о значении параметра θ , а это и означает, что вся информация о θ , содержащаяся в выборке, содержится целиком в достаточной оценке t_D .

Обобщение на случай нескольких параметров очевидно: множество оценок t_{D_1}, \dots, t_{D_l} для параметров $\theta_1, \dots, \theta_l$ ($l \leq k$) называется множеством *совместно-достаточных* оценок для параметров $\theta_1, \dots, \theta_l$, если можно произвести преобразование от переменных x_1, \dots, x_n к переменным $t_{D_1}(x_1, \dots, x_n), \dots, t_{D_l}(x_1, \dots, x_n)$,

$u_1(x_1, \dots, x_n), \dots, u_{n-1}(x_1, \dots, x_n)$ так, что функция правдоподобия примет вид

$$P(x_1, \dots, x_n; \theta_1, \dots, \theta_k) = g(t_{D_1}, \dots, t_{D_l}; \theta_1, \dots, \theta_l) \times \\ \times h(u_1, \dots, u_{n-l} | t_{D_1}, \dots, t_{D_l}; \theta_{l+1}, \dots, \theta_k), \quad (2)$$

где $\theta_1, \dots, \theta_l$ не входят во второй множитель, дающий условное распределение величин u_1, \dots, u_{n-l} .

Теорема III. Если существует множество совместно-достаточных оценок, то любое решение уравнений правдоподобия для $\theta_1, \dots, \theta_l$ является функцией от этих оценок t_{D_1}, \dots, t_{D_l} . Эта теорема непосредственно следует из формул (2) и (10.10.5).

Пример 1. Формулу (10.10.9) можно записать в виде

$$P(x_1, \dots, x_n; \mu) = \frac{e^{-n\mu} (n\mu)^{nm}}{(nm)!} \frac{(nm)!}{n^{nm} x_1! \dots x_n!}. \quad (3)$$

Здесь первый множитель дает вероятность того, что случайная величина $x_1 + \dots + x_n$ имеет значение nm (см. упражнение 5 § 4.11). Согласно закону умножения V второй множитель дает условную вероятность получения величин x_1, \dots, x_n при условии, что их сумма имеет значение nm . Поскольку параметр μ не входит во второй множитель, m есть достаточная оценка для μ .

Пример 2. Из (7.7.11) следует, что величины m и $s_1 \left(= \frac{q}{\sqrt{n}} \right)$, определяемые формулами (7.7.4) и (10.9.8), суть совместно-достаточные оценки для параметров μ и σ распределения Кэптейна (10.3.2).

Упражнение. Покажите, что в случае биномиального распределения величина $\frac{x}{y}$, определяемая формулой (10.10.8), есть достаточная оценка для θ .

Пример 3. Как пример оценки, не являющейся достаточной, укажем выборочную медиану [в случае распределения Лапласа (см. упражнение 5 § 10.10)].

§ 10.12. Метод максимума правдоподобия иногда приводит к уравнениям, решение которых затруднительно или требует очень большого числа вычислений. В таких случаях может быть полезен другой метод — *метод моментов*. Этот метод основан на непосредственном обобщении теоремы Хинчина, пример 4 § 8.1, гласящем, что если момент μ_i порядка i , определяемый формулой (5.2.9), существует в строгом смысле, т. е. если соответствующая сумма или интеграл абсолютно сходится, то *выборочный момент m_i порядка i сходится по вероятности к μ_i при $n \rightarrow \infty$* :

$$m_i = \overline{x^i} = \frac{x_1^i + \dots + x_n^i}{n} \xrightarrow[n \rightarrow \infty]{\text{по вер.}} \mu_i, \quad i = 1, 2, \dots \quad (1)$$

Другими словами, m_i есть состоятельная оценка для μ_i . Любой момент есть функция параметров $\theta_1, \dots, \theta_k$. Чтобы получить оценки для параметров $\theta_1, \dots, \theta_k$ по методу моментов, надо разрешить относительно $\theta_1, \dots, \theta_k$ уравнения

$$\mu_i \approx m_i, \quad i = 1, 2, \dots, k, \quad (2)$$

в результате чего получим

$$\theta_i \approx t_i(m_1, \dots, m_k) = t_i(x_1, \dots, x_n). \quad (3)$$

У п р а ж н е н и е 1. Покажите, что метод моментов дает состоятельные оценки (см. упражнение 3 § 10.9).

У п р а ж н е н и е 2. Покажите, что для биномиального распределения и распределений Пуассона и Кэптейна оценки, получаемые методом максимума правдоподобия и методом моментов, совпадают.

Из центральной предельной теоремы следует, что при весьма больших условиях выборочный момент m_i при больших значениях n распределен асимптотически нормально с параметрами $\mu = \mu_i$ и $\sigma = \frac{\sigma \{x^i\}}{\sqrt{n}}$.

Аналогично можно показать, что оценки, получаемые методом моментов, при больших значениях n распределены асимптотически нормально с параметрами $\mu = \theta_i$ и $\sigma = \frac{c_i}{\sqrt{n}}$, где c_i суть некоторые постоянные. Таким образом, эти оценки должны иметь асимптотическую эффективность, определяемую формулой $e\{t_i\} = \frac{\sigma_0^2}{c_i^2/n}$. Однако

Фишер показал, что эта асимптотическая эффективность часто заметно меньше единицы, так что в таких случаях оценки, получаемые методом моментов, весьма неэффективны, т. е. не дают максимальной информации о параметрах, содержащейся в выборке. Тем не менее если увеличение числа наблюдаемых данных менее затруднительно, чем производство вычислений, необходимое для получения оценок максимального правдоподобия, то использование неэффективного метода моментов может оказаться предпочтительным.

П р и м е р. При нейтронной бомбардировке ядер урана начинается процесс расщепления, при котором ядро урана распадается на две меньшие части. При изучении этого явления в камере Вильсона обнаруживаются траектории двух частиц, исходящие из одной и той же точки, но имеющие противоположные направления. Обе траектории состоят из некоторого количества ветвей, получающихся вследствие столкновения частиц с молекулами газа в камере.

Рассмотрим случайную величину x , равную числу ветвей в одной траектории. Ядро урана распадается на две частицы различного рода, одна из которых имеет несколько большую массу, чем другая. Поэтому рассуждения примеров 4 и 5 § 4.3 могут быть применены к

распределению числа ветвей в траекториях частиц обоих видов, но в этих распределениях средние значения μ' и μ'' должны быть различны. Более того, данная траектория может с одинаковой вероятностью относиться к той или другой частице, так что для x мы получаем так называемое двойное распределение Пуассона

$$\varphi_i = \frac{1}{2} \frac{e^{-\mu'} \mu'^i}{i!} + \frac{1}{2} \frac{e^{-\mu''} \mu''^i}{i!}, \quad i=0, 1, 2, \dots \quad (4)$$

Здесь подлежат оценке два параметра μ' и μ'' , так что следует вычислить первые два момента распределения (4). По формулам (5.1.6) и (5.3.11) получаем

$$\mu_0 = 1, \quad \mu_1 = \frac{1}{2} \mu' + \frac{1}{2} \mu'', \quad \mu_2 = \mu_1 + \frac{1}{2} \mu'^2 + \frac{1}{2} \mu''^2. \quad (5)$$

Решение этих двух уравнений имеет вид

$$\left. \begin{array}{l} \mu' \\ \mu'' \end{array} \right\} = \mu_1 \pm \sqrt{\mu_2 - \mu_1 - \mu_1^2}. \quad (6)$$

В табл. 5 приведено распределение величины x для траекторий $n=327$ частиц, получившихся при расщеплении урана, а именно: указано количество n_i траекторий, имеющих i ветвей¹⁾.

Таблица 5

i	n_i	$n_i i$	$n_i i^2$
0	28	0	0
1	47	47	47
2	81	162	324
3	67	201	603
4	53	212	848
5	24	120	600
6	13	78	468
7	8	56	392
8	3	24	192
9	2	18	162
10	1	10	100
	327	928	3736

По данным таблицы получаем эмпирические моменты

$$\mu_1 \approx m_1 = \frac{1}{327} \sum_{i=0}^{10} n_i i = 2,8379, \quad (7)$$

$$\mu_2 \approx m_2 = \frac{1}{327} \sum_{i=0}^{10} n_i i^2 = 11,425.$$

¹⁾ B. Øggild, Brostrøm and Lauritzin, 'Danske Vid. Selsk. Mat.-fys. Medd., Vol. XIII, № 4, 1940.

По формуле (6) получаем

$$\mu' \approx 3,5684, \quad \mu'' \approx 2,1075. \quad (8)$$

Таким образом, два средних значения существенно различны.

У п р а ж н е н и е 3. Вычислите по формулам (4) и (8) теоретическое распределение $\nu_i = n\varphi_i$ и покажите графически, что согласие этого распределения с эмпирическим распределением табл. 5 удовлетворительное. Вычислите также обычное распределение Пуассона (4.3.5), для которого по формуле (7) $\mu \approx m_1 = 2,8379$, и покажите, что это распределение плохо согласуется с эмпирическим распределением табл. 5.

*У п р а ж н е н и е 4. Распределение (4) есть лишь частный случай общего двойного распределения Пуассона

$$\varphi_i = \gamma_1 \frac{e^{-\mu'} \mu'^i}{i!} + \gamma_2 \frac{e^{-\mu''} \mu''^i}{i!}, \quad \gamma_1 + \gamma_2 = 1, \quad (9)$$

имеющего три независимых, подлежащих оценке параметра μ' , μ'' и γ_1 . Полагая для сокращения

$$\beta_1 = \mu_1, \quad \beta_2 = \mu_2 - \mu_1, \quad \beta_3 = \mu_3 - 3\mu_2 + 2\mu_1, \quad (10)$$

покажите, что μ'' есть корень уравнения *)

$$\begin{aligned} &(\beta_2 - \beta_1^2) x^5 - (\beta_3 - \beta_1^3) x^4 + 2(\beta_1\beta_3 - \beta_2^2) x^3 + \\ &+ 2\beta_3(\beta_2 - \beta_1^2) x^2 - (\beta_3^2 - \beta_1^3) x + \beta_3(\beta_1\beta_3 - \beta_2^2) = 0. \end{aligned} \quad (11)$$

Покажите, далее, что μ' , γ_1 и γ_2 определяются формулами

$$\mu' = \frac{\beta_3 - \beta_1 \mu''^2}{\beta_2 - \mu''^2}, \quad \gamma_1 = \frac{\mu'' - \beta_1}{\mu'' - \mu'}, \quad \gamma_2 = \frac{\beta_1 - \mu'}{\mu' - \mu''}. \quad (12)$$

*У п р а ж н е н и е 5. Обозначая через b_1, b_2, b_3 эмпирические значения $\beta_1, \beta_2, \beta_3$, получите по данным табл. 5 значения

$$m_3 = \frac{1}{327} \sum_{i=0}^{10} n_i i^3 = 56,398; \quad b_1 = 2,8379; \quad b_2 = 8,5872; \quad b_3 = 27,798 \quad (13)$$

и вычислите значения величин

$$\gamma_1 \approx 0,27413; \quad \gamma_2 \approx 0,72587; \quad \mu' \approx 4,0276; \quad \mu'' \approx 2,3886. \quad (14)$$

Наконец, вычислите по формулам (9) и (14) теоретическое распределение $\nu_i = n\varphi_i$ и покажите графически, что оно совпадает с эмпирическим распределением табл. 5 так же хорошо, как и частное двойное распределение Пуассона для которого $\gamma_1 = \gamma_2 = 1/2$. Таким образом, в этом случае мы не можем эмпирически отличить распределение общего вида от распределения частного вида. Однако из теоретических соображений последнее предпочтительнее.

*) На самом деле μ' и μ'' являются корнями квадратного уравнения

$$(\beta_2 - \beta_1^2) x^2 + (\beta_1\beta_2 - \beta_3) x + (\beta_1\beta_3 - \beta_2^2) = 0.$$

Этот квадратный трехчлен входит множителем в левую часть уравнения, приведенного в тексте. (Прим. ред.)

11. ПРИЛОЖЕНИЕ ТЕОРИИ ВЕРОЯТНОСТЕЙ К ТЕОРИИ ОШИБОК

§ 11.1. Под теорией ошибок — несколько неудачное, но общепотребительное название — понимают специальную часть статистики, посвященную численному определению физических величин. Однако, поскольку при этом всегда используется нормальное распределение, методы, излагаемые в этой главе, могут применяться и в других задачах, когда можно считать, что нормальное распределение или его обобщение, данное Кэптейном, дает удовлетворительное описание рассматриваемого статистического явления.

В теории ошибок рассматриваются следующие четыре вопроса:

I. Что следует понимать под „истинным“ значением физической величины?

II. Как практически оценить истинные значения по данным измерений?

III. Какую степень надежности можно приписать полученным оценкам?

IV. Как можно сравнить оценки, полученные из различных серий измерений?

§ 11.2. Прежде чем отвечать на вопрос I, следует, вообще говоря, заметить, что *физическая* величина, как величина, определяемая с помощью некоторого процесса измерения, в противоположность математической величине, никогда не определяется настолько точно, чтобы можно было для „истинного“ значения этой величины предпочесть какое-нибудь одно определенное число по сравнению со всеми остальными числами. Поскольку совершенство наших органов чувств и всех наших измерительных инструментов ограничено, всегда имеется некоторое ограничение малости доступных нашему наблюдению величин. В результате этого мы можем получить при измерении число лишь с конечным количеством знаков. Иначе говоря, *каждое эмпирически полученное число есть целое число, если его выразить в наименьших доступных измерению единицах ϵ* . Таким образом, „истинным“ значением величины можно назвать любое действительное число из некоторого интервала длины ϵ . Далее, эта неточность определения, как правило, оказывается малой по сравнению с неточностями, возникающими вследствие ошибок измерений, описанных в § 1.2. Тем не менее, поскольку неограниченное совершенствование

нашей измерительной техники принципиально представляется возможным, для целей математического описания естественно отвлечься от указанной неточности определения физической величины и идеализировать результаты измерений, избирая одно определенное число в качестве *единственного* истинного значения рассматриваемой величины. Именно такие истинные значения, а не значения, полученные непосредственно из наблюдений, входят в нашу модель реального мира (§ 1.1), например в законы классической физики. В квантовой теории от такой идеализации отказываются как от чрезмерно грубой для описания атомных явлений. Вместо такого причинного описания, в котором все величины имеют определенные значения, в квантовой теории используют статистическое описание с помощью функций распределения, отличных от несобственных (см. § 4.17). Сравнивая наши представления о понятиях „истинного значения“ и „вероятности“, мы видим, что понятие „вероятности“ ничем не отличается от понятия „истинного значения“ каких-либо других физических величин. Вероятности суть просто истинные значения некоторых специальных величин, а именно относительных частот.

§ 11.3. Мы должны теперь пренебречь неточностями, присущими самим определяемым величинам, и рассматривать лишь те неточности, которые проявляются при измерениях. Как было описано в § 1.2, несовершенство наших измерительных инструментов, а также возмущающие факторы всегда приводят к тому, что каждый результат измерений есть *случайная величина*, так как при повторных измерениях одной и той же величины, не изменяющейся в процессе измерения, всегда получаются различные результаты. Это обстоятельство можно выразить, говоря, что измерения подвержены *ошибкам*, в результате которых измеренные значения более или менее отклоняются от истинных значений. Так будет, конечно, лишь в случае, если измерительные инструменты не слишком грубы. Если, например, расстояние между двумя черточками на стальной штанге равно $2,0078$ м, и это расстояние измеряется с помощью линейки, имеющей деления лишь через каждый метр, то каждое измерение даст результат „ 2 м“. Поэтому всегда предполагается, что измерительный инструмент выбран так, чтобы минимальная поддающаяся измерению этим инструментом единица была мала по сравнению с измеряемой величиной.

Ошибки делятся на три группы: (I) грубые ошибки, (II) систематические ошибки, (III) статистические или случайные ошибки.

Грубые ошибки суть ошибки в отчетах показаний инструмента, в вычислениях, ошибки, созданные неправильным использованием инструмента или просто возникшие из-за недостатка осторожности у наблюдателя. Конечно, такие ошибки должны быть исключены. Результаты измерений, содержащие грубые ошибки, обычно хорошо заметны, так как они совершенно не похожи на другие результаты измерений. (В артиллерии, например, говорят о „шалых“ выстрелах.)

В последующем мы будем предполагать, что наблюдения, содержащие грубые ошибки, уже исключены (см., однако, § 11.17).

Систематические ошибки суть ошибки, созданные одной или несколькими определенными причинами, действующими по определенным законам и, как правило, в определенном направлении. Если измерение повторяется при постоянных условиях, то возникает одна и та же систематическая ошибка. Следовательно, в противоположность грубым ошибкам систематические ошибки не обнаруживаются в каком-либо несогласии между различными результатами, а лишь изменяют все эти результаты на одно и то же слагаемое. Однако если законы, управляющие систематическими ошибками, известны, то эти ошибки можно вычислить и рассматривать как *поправки* к измеренным значениям. Большинство систематических ошибок создаются самими измерительными инструментами. Если, например, длина, измеряемая линейкой, немного короче, чем расстояние между делениями на линейке, то каждое измерение длины будет давать завышенный результат. Однако такую ошибку нельзя обнаружить, если только измерение не будет повторено с другой линейкой. Сравнение результатов, полученных с помощью различных измерительных инструментов, является наиболее эффективным методом для выявления систематических ошибок. В последующем мы будем предполагать, что имеющиеся в результатах наблюдений известные систематические ошибки исправлены.

Случайные ошибки суть все ошибки, отличные от рассмотренных выше. Эти ошибки не обнаруживают каких-либо регулярностей, или же такие регулярности нам неизвестны. Иногда слово „ошибки“ употребляется только для систематических ошибок, а для случайных ошибок используется слово „неточности“. Большинство случайных ошибок зависит от интерполяции, неизбежной при отсчетах по шкале, от необходимости регулирования измерительных приборов, а также вследствие всех возмущающих факторов, описанных в § 1.2. Вообще говоря, характеристической чертой случайных ошибок, в противоположность систематическим ошибкам, является то, что положительные и отрицательные значения таких ошибок одинаково вероятны. Однако возможны ошибки, имеющие асимметричное распределение; такие ошибки называются *односторонними*. При подробном исследовании часто удается выяснить, что односторонние ошибки суть ошибки систематические. Приведем в качестве примера односторонней ошибки искривление оси оптического инструмента.

Однако различие между описанными группами ошибок не является абсолютным. При внимательном исследовании в некоторых случайных ошибках удается обнаружить регулярности. При этом ошибка, которая ранее считалась случайной, может оказаться систематической. Случайно может появиться также и очень большая ошибка, которую было бы неправильно относить к разряду грубых ошибок. Поэтому на практике не всегда бывает легко определить, следует ли забра-

ковать измерение, заметно отклоняющееся от других измерений (см § 11.17).

§ 11.4 С данной физической величиной x и с данным методом измерения свяжем некоторую функцию распределения $\Phi(x)$ (см. примечание 2 на стр. 147).

Подчеркнем, что $\Phi(x)$ *зависит также от метода измерения*. Действительно, физическую величину нельзя считать определенной, прежде чем не установлен метод измерения. Ответы на четыре вопроса § 11.1 нельзя получить независимо от функции распределения $\Phi(x)$. Однако опыт показывает, что большинство физических измерений — после исключения грубых и систематических ошибок — распределено приблизительно нормально, что может быть объяснено теоретически (см. пример I § 10.3). Поэтому теория ошибок основывается на нормальном распределении

$$d\Phi = \varphi(x) dx = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx, \quad (1)$$

хотя, строго говоря, каждая физическая величина есть величина дискретная. Однако, если статистические флуктуации велики по сравнению с доступной измерению единицей, функция распределения будет иметь большое количество очень малых скачков и поэтому может быть аппроксимирована непрерывным распределением (см. стр. 41).

Поскольку, во-первых, параметр μ равен математическому ожиданию величины x , дающему порядок величины значений x (см. IV § 9.3), и, во-вторых, нормальное распределение симметрично относительно точки $x = \mu$, естественно принимать μ по определению за „истинное“ значение физической величины x . Подчеркнем, что по этому определению истинное значение зависит также от используемого метода измерения. Поэтому возникает важная задача о том, согласуются ли друг с другом, в пределах статистических флуктуаций, два различных „истинных“ значения, полученных двумя различными измерительными методами (см. § 11.15). Их значимое несогласие показало бы, что измерялась не одна и та же величина или что допущены некоторые систематические ошибки. Тогда можно получить истинные значения, установив и учтя эти систематические ошибки.

§ 11.5. Рассмотрим теперь следующий вопрос: как по данным измерений оценить численное значение μ ? Физические измерения делятся на две группы — непосредственные и косвенные измерения. При непосредственном измерении, как, например, при измерении длины, массы или при измерении силы тока с помощью амперметра, численный результат получается прямо из наблюдаемых данных. При косвенных измерениях, как, например, при измерении удельного сопротивления $\rho = \frac{\pi R d^2}{4l}$ проволоки с полным сопротивлением R , длиной l и диаметром d , результат вычисляется по данным наблюдений, так

как подлежащая измерению величина есть функция некоторых других величин, измеряемых непосредственно. Сначала мы будем рассматривать исключительно непосредственные измерения, а рассмотрение косвенных измерений отложим до § 11.22. Чтобы определить истинное значение рассматриваемой величины x , вследствие статистических флуктуаций недостаточно осуществить лишь одно измерение величины x . Лишь в тех случаях, когда заранее, например по предыдущим измерениям, можно ожидать, что статистические флуктуации малы, достаточно иметь одно измерение x_1 . В соответствии с методом максимума правдоподобия в этом случае x_1 будет наилучшей оценкой для μ . Вообще же следует повторить измерение несколько раз, получив при этом результаты x_1, x_2, \dots, x_n . Как можно видеть из примера 4 § 10.10, наилучшей оценкой для истинного значения μ будет

$$\mu \approx m = \bar{x} = \frac{x_1 + \dots + x_n}{n} = \frac{[x]}{n} \quad (1)$$

(при этом надо положить $G(x) = x$).

Как показано в примере § 10.9, при всех значениях n величина m есть несмещенная эффективная оценка для μ . Символ $[x]$ в формуле (1) означает

$$[x] = \sum_{i=1}^n x_i. \quad (2)$$

Этот символ, введенный Гауссом, широко используется в теории ошибок. В теории ошибок также употребляется термин „истинные ошибки“¹⁾ для разностей

$$\epsilon_i = x_i - \mu, \quad i = 1, 2, \dots, n \quad (3)$$

и термин „кажущиеся ошибки“ для разностей

$$v_i = x_i - \bar{x}, \quad i = 1, 2, \dots, n. \quad (4)$$

Упражнение 1. Покажите, что

$$[v] = \sum_{i=1}^n v_i = 0. \quad (5)$$

Далее, покажите, что

$$\sum_{i=1}^n (x_i - x')^2 = [v'v'] = [vv] + n(\bar{x} - x')^2, \quad (6)$$

где $v' = x_i - x'$, а x' — произвольное число.

¹⁾ Ошибка есть величина, которую следует прибавить к „истинному значению“, чтобы получить измеренное значение; ошибка с обратным знаком называется поправкой:

истинное значение + ошибка = измеренное значение,
измеренное значение + поправка = истинное значение.

Поскольку в функции правдоподобия μ входит лишь в выражение $\sum_{i=1}^n (x_i - \mu)^2 = [\varepsilon\varepsilon]$ (см. (10.10.16)), то ясно, что в случае нормального распределения наилучшая оценка для μ получается из условия, что кажущиеся ошибки удовлетворяют соотношению

$$[vv] = \sum_{i=1}^n (x_i - \bar{x})^2 = \min [v'v']. \quad (7)$$

Этот известный принцип называется *методом наименьших квадратов*. Из формулы (6) видно, что $x' = \bar{x}$ действительно соответствует наименьшему возможному значению $[v'v']$.

У п р а ж н е н и е 2. Покажите, что для кажущихся ошибок (4)

$$\mathfrak{M} \{v_i\} = 0, \quad i = 1, 2, \dots, n. \quad (8)$$

Покажите, далее, что

$$v_i = \sum_{j=1}^n \left(\delta_{ij} - \frac{1}{n} \right) x_j, \quad \delta_{ij} = \begin{cases} 1 & i=j, \\ 0 & i \neq j, \end{cases} \quad (9)$$

так что

$$\sigma^2 \{v_i\} = \left(1 - \frac{1}{n}\right)^2 \sigma^2 + \frac{1}{n^2} \sigma^2 + \dots + \frac{1}{n^2} \sigma^2 = \frac{n-1}{n} \sigma^2; \quad i = 1, 2, \dots, n. \quad (10)$$

Наконец, покажите, что

$$\mathfrak{M} \{[vv]\} = (n-1) \sigma^2, \quad (11)$$

что согласуется с формулой (7.7.13), так как $q^2 = [vv]$.

Из упражнения 2 следует, что поскольку каждая кажущаяся ошибка v_i есть сумма независимых нормально распределенных величин, она имеет нормальное распределение с параметрами 0 и $\sqrt{\frac{n-1}{n}} \sigma$. Однако вследствие соотношения (5) кажущиеся ошибки v_i не независимы.

* У п р а ж н е н и е 3. Покажите, что

$$\rho \{v_i, v_j\} = \frac{-1}{n-1}, \quad i \neq j \quad (12)$$

и

$$\rho \{v_i, \bar{x} - \mu\} = 0. \quad (13)$$

§ 11.6. Как показано в примере 4 § 10.10, наилучшая оценка параметра σ определяется формулой

$$\sigma \approx s_1 = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{[vv]}{n}}. \quad (1)$$

Однако из упражнения 2 § 10.9 следует, что s_1 является лишь асимптотически несмещенной и эффективной оценкой для параметра σ , и, конечно, существует много других оценок для этого параметра, также являющихся асимптотически несмещенными и эффективными, как, например, все оценки вида $g_n s_1$, где $g_n \sim 1$ и $g_n \xrightarrow{n \rightarrow \infty} 1$, имеющие тот же самый коэффициент изменчивости, что и s_1 . В статистической практике принято использовать оценку s , определяемую формулой (7.7.16), т. е. полагать $g_n = \sqrt{\frac{n}{n-1}}$:

$$\sigma \approx s = \sqrt{\frac{n}{n-1}} s_1 = \sqrt{\frac{[sv]}{n-1}}, \quad (2)$$

причем $n-1 = f$ называется *числом степеней свободы*. Вопрос об использовании оценки s_1 или оценки s зависит от личного выбора. Однако, как будет показано в § 11.10, оценка s имеет некоторые преимущества по сравнению с оценкой s_1 , вследствие чего, как правило, используется оценка s .

§ 11.7. Две оценки \bar{x} и s для параметров μ и σ сами суть случайные величины, подверженные статистическим флуктуациям, т. е. приобретают новые значения, если n измерений, по которым они вычислены, повторить заново. Чтобы оценить точность этих оценок, мы определим дисперсии \bar{x} и s . Как показано в § 7.5, величина \bar{x} при всех значениях n имеет нормальное распределение со средним значением μ и стандартным отклонением $\frac{\sigma}{\sqrt{n}}$, т. е.

$$\sigma\{\bar{x}\} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}} = \sqrt{\frac{[sv]}{n(n-1)}}; \quad (1)$$

далее, мы видели, что $\sqrt{n-1}s = q$ имеет q -распределение (7.7.12), которое при больших значениях n приблизительно нормально. Согласно формулам (7.7.18) и (7.7.19) величина s распределена приблизительно нормально со средним значением σ и стандартным отклонением $\frac{\sigma}{\sqrt{2(n-1)}}$, так что

$$\sigma\{s\} \sim \frac{\sigma}{\sqrt{2(n-1)}} \approx \frac{s}{\sqrt{2(n-1)}} = \frac{\sqrt{[sv]}}{\sqrt{2(n-1)}}. \quad (2)$$

Пример 1. Заметим, что с точки зрения формулы (2), если не было бы каких-либо других причин предпочитать оценку s , а не оценку s_1 , эти оценки можно было бы считать одинаково хорошими, так как при не слишком малых значениях n разность этих оценок

мала по сравнению с их стандартным отклонением:

$$\begin{aligned} s - s_1 &= \sqrt{[vv]} \left(\frac{1}{\sqrt{n-1}} - \frac{1}{\sqrt{n}} \right) = \\ &= s \left[1 - \left(1 - \frac{1}{n} \right)^{1/2} \right] = s \left[\frac{1}{2n} + \dots \right], \end{aligned}$$

что действительно мало по сравнению с $\sigma\{s\} \approx s \left(\frac{1}{\sqrt{2n}} + \dots \right)$.

Другой аргумент, часто заставляющий предпочитать оценку s , а не оценку s_1 , заключается в том, что s^2 есть несмещенная оценка для σ^2 , а s_1^2 этим свойством не обладает (см. упражнение 1 § 10,9). Однако, во-первых, более естественно считать подлежащим оценке параметром σ , а не σ^2 ; но тогда ни s , ни s_1 не будут несмещенными оценками (см. упражнение 2 § 10,9). Во-вторых, и s и s_1 сходятся по вероятности к σ , а это означает, что при больших значениях n масса вероятности в распределениях обеих оценок, имеющих их одинаковые коэффициенты изменчивости, сконцентрирована около значения σ ; но тогда их разность будет мала по сравнению с их дисперсиями, как это и было показано выше. С другой стороны, $\sigma\{s\}$ настолько медленно стремится к нулю, что фактически наблюдаемые значения не обязаны быть очень близкими к σ , если только n не имеет чрезвычайно большого значения. Основная же причина для предпочтения оценки s , а не s_1 , описана в § 11.10.

Таким образом, окончательным результатом, извлеченным из измерений x_1, \dots, x_n , является:

Оценка истинного значения

$$\mu \approx \bar{x} = \frac{[x]}{n} \text{ со стандартным отклонением } \sigma\{\bar{x}\} \approx \frac{s}{\sqrt{n}}. \quad (3)$$

Оценка стандартного отклонения

$$\sigma \approx s = \sqrt{\frac{[vv]}{n-1}} \text{ со стандартным отклонением } \sigma\{s\} \approx \frac{s}{\sqrt{2(n-1)}}. \quad (4)$$

Часто стандартное отклонение оценки называется *стандартной ошибкой*, *средней квадратической ошибкой* или просто *средней ошибкой*. Само стандартное отклонение σ в старых работах носит наименование *стандартной* или *средней ошибки единичного измерения*, что, однако, является довольно неудачным наименованием.

Подчеркнем, что для оценки точности значений \bar{x} и s всегда должны быть даны их стандартные отклонения, или, во всяком случае, число n измерений, по которым эти значения получены.

Пример 2. Иногда результат дается в виде

$$\mu = \bar{x} \pm \frac{s}{\sqrt{n}} \text{ или } \mu = \bar{x} \pm 100 \frac{s}{x \sqrt{n}} \%. \quad (5)$$

Однако, во-первых, мы имеем $\mu \approx \bar{x}$, а не $\mu = \bar{x}$. Во-вторых, μ вполне может иметь значения вне указанных интервалов и все еще быть совместимым с наблюдаемыми значениями \bar{x} и s (см. § 11.9). Таким образом, использование записи (5) неудобно, и от него следует отказаться, предпочитая запись (3).

Как было указано в § 7.4, иногда, особенно в старых работах, вместо параметра σ используется другой параметр $\rho = 0,67449\sigma$, который дает 50%₀-ные доверительные пределы для среднего значения. Тогда результат записывается в виде

$$\mu = \bar{x} \pm 0,674 \frac{s}{\sqrt{n}} = \bar{x} \pm \frac{r}{\sqrt{n}}, \quad r = 0,674 s. \quad (6)$$

Поскольку результат измерения может с одинаковой вероятностью оказаться как внутри, так и вне пятидесятипроцентного доверительного интервала, утверждения, подобные (6), несомненно, приводят к переоценке точности значения \bar{x} ; поэтому такие утверждения использовать не следует.

Пример 3. Вместо оценки s , определяемой формулой (4), используются также и другие оценки для σ , как, например, оценка s_2 , определяемая формулой (10.9.15). Для грубой оценки, чего часто достаточно в инженерном деле, весьма удобно использовать наибольшее x_{\max} и наименьшее x_{\min} значения среди данных n измерений и полагать $\mu \approx m_1 = \frac{1}{2}(x_{\max} + x_{\min})$ и $\sigma \approx s_3 = \frac{1}{\alpha_n}(x_{\max} - x_{\min})$, причем α_n определяется так, чтобы $M\{s_3\} = \sigma$ ($w = x_{\max} - x_{\min}$ называется размахом). Можно показать, что m_1 и s_3 суть состоятельные, но весьма неэффективные оценки для параметров μ и σ . Функция α_n табулирована¹⁾, например: $\alpha_{10} = 3,03$; $\alpha_{30} = 4,09$; $\alpha_{100} = 5,02$ и $\alpha_{500} = 6,07$.

*У п р а ж и е н и е. Иногда, например в артиллерии, используется оценка

$$s_4 = \frac{\sqrt{\pi}}{2} \frac{d_1 + \dots + d_{n-1}}{n-1}, \quad d_i = |x_i - x_{i+1}|,$$

причем d_i называются *последовательными или разностями*. Покажите, что s_4 есть несмещенная оценка для σ . С помощью нескольких длинных выкладок можно показать, что s_4 есть состоятельная оценка с эффективностью $e\{s_4\} = 0,61 \frac{n-1}{n}$.

§ 11.8. Если параметры μ и σ известны, то по табл. II (см. § 7.4) при всех значениях P могут быть найдены *доверительные пределы*, т. е. пределы (предполагаемые, как правило, симметричными относи-

1) Описание и литературу см. в книге Г. Крамер, Математические методы статистики, стр. 403, 1948. Таблица чисел α_n при $n=1-1000$ дана А. Н. С. Tippett, *Biometrika*, 17, 364, 1925.

тельно значения μ), *внутри* которых x может оказаться с заданной вероятностью $1 - P$. Вследствие статистических флуктуаций величин \bar{x} и s их значения нельзя непосредственно подставлять вместо μ и σ в формулы для доверительных пределов, хотя это и делается часто. Однако имеется t -распределение (7.9.3), не содержащее самих параметров μ и σ и позволяющее оценить доверительные пределы для нового измерения x_{n+1} по значениям \bar{x} и s , вычисленным по выборке x_1, \dots, x_n :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (1)$$

Мы образуем для этого новую случайную величину

$$t = \frac{\bar{x} - x_{n+1}}{\sqrt{1 + \frac{1}{n}} s}, \quad (2)$$

которая, как показано в упражнении 2 § 7.9, имеет t -распределение (7, 9, 3) с $f = n - 1$. Если $t(P, f)$ обозначает решение уравнения $P(|t| \geq t) = P$, которое протабулировано в табл. III приложения, то

$$P \left[-t(P, f) \leq \frac{\bar{x} - x_{n+1}}{\sqrt{1 + \frac{1}{n}} s} \leq t(P, f) \right] = 1 - P. \quad (3)$$

Но эту формулу можно записать также, разрешив неравенства относительно x_{n+1} , в виде

$$P \left[\bar{x} - \sqrt{1 + \frac{1}{n}} s t(P, f) \leq x_{n+1} \leq \bar{x} + \sqrt{1 + \frac{1}{n}} s t(P, f) \right] = 1 - P. \quad (4)$$

Таким образом, при заданном значении P по данной выборке можно получить следующую оценку доверительных пределов для нового измерения x_{n+1} :

$$\bar{x} - \sqrt{1 + \frac{1}{n}} s t(P, f), \quad \bar{x} + \sqrt{1 + \frac{1}{n}} s t(P, f). \quad (5)$$

Поскольку при $n \rightarrow \infty$ t -распределение стремится к нормированному нормальному распределению, множитель $\sqrt{1 + \frac{1}{n}} t(P, f)$ при $n \rightarrow \infty$ стремится к соответствующему значению, получаемому по табл. II. На практике принято выбирать $P = 5\%$.

Пример. Для 10 измерений, приведенных ниже, в примере § 11.11, мы имеем $\bar{x} = 4,0759$ и $s = 0,0039$. Поскольку $f = 10 - 1 = 9$

и $t(5\%/0,9) = 2,262$, по табл. III, имеем $\sqrt{1 + \frac{1}{10}} t(5\%/0,9) = 2,371$, что заметно больше, чем $t(5\%/0, \infty) = 1,96$.

Таким образом, мы получаем доверительные пределы для нового измерения в виде $4,0759 \pm 0,0039 \cdot 2,371 = 4,0759 \pm 0,0092$.

§ 11.9. Формула (11.7.3) дает наилучшую оценку \bar{x} для истинного значения μ и стандартное отклонение этой оценки $\frac{s}{\sqrt{n}}$. При этом μ есть определенная, хотя и неизвестная постоянная, т. е. параметр, а не случайная величина. Таким образом, нельзя говорить о вероятности того, что μ лежит между некоторыми пределами, скажем, между пределами $\bar{x} \pm \frac{s}{\sqrt{n}}$. Однако можно искать значения μ , совместные с наблюдаемыми значениями. Для этой цели вновь можно использовать t -распределение (7.9.3), образовав случайную величину

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}, \quad (1)$$

которая, как показано в упражнении 2 § 7.9, имеет распределение (7.9.3) с числом степеней свободы $f = n - 1$. Если μ меняется, то меняется и значение t , и поэтому имеет смысл считать определенное значение μ совместным с наблюдаемой выборкой x_1, \dots, x_n , если соответствующее значение t лежит в некоторых пределах, в качестве которых принято брать пятипроцентные доверительные пределы. Таким образом, если $t(P, f)$ имеет тот же смысл, что и в § 11.8, μ можно считать совместным с выборкой, если выполняется условие

$$P \left[-t(P, f) \leq \frac{\bar{x} - \mu}{s/\sqrt{n}} \leq t(P, f) \right] = 1 - P. \quad (2)$$

Разрешая неравенства относительно μ , мы сможем записать это условие в виде

$$P \left[\bar{x} - \frac{s}{\sqrt{n}} t(P, f) \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}} t(P, f) \right] = 1 - P. \quad (3)$$

Такие пределы для μ называются *доверительными пределами*. Подчеркнем, что доверительные пределы в этом случае не следует путать с понятием доверительных пределов, рассматривавшимся в предыдущих параграфах. Последние суть вероятные пределы для случайной величины, первые же есть пределы совместности значений параметра с наблюдаемыми данными.

Пример. Для 10 измерений, приведенных ниже, в примере § 11.11, имеем $\bar{x} = 4,0759$ и $s/\sqrt{10} = 0,0012$. Поскольку $f = 10 - 1 = 9$ и по табл. III $t(5\%/0,9) = 2,262$, то мы получаем довери-

тельные пределы для истинного значения в виде $4,0759 \pm 0,0027 = (4,0732; 4,0786)$.

§ 11.10. Осуществление большого количества измерений, необходимых для получения надежной оценки параметра σ по формуле (11.7.4), к сожалению, не всегда возможно или неудобно на практике. Однако часто имеется большое количество коротких рядов измерений, относящихся к различным значениям μ , но полученных при одинаковых условиях и поэтому имеющих одинаковую точность, т. е. одно и то же значение σ . Поскольку распределение величины $q = \sqrt{[vv]}$ (см. (7.7.12)) зависит лишь от σ и от $f (= n - 1)$, но не от μ , может быть удобно рассматривать в качестве первичной величины эту величину q , а не непосредственно измеренные значения x_i , по которым вычисляется q . В первом из этих подходов n измерений x_1, \dots, x_n представляют собой n наблюдаемых значений случайной величины x , во втором подходе — лишь одно измерение случайной величины q . Таким образом, можно ожидать, что при этих двух подходах для одного и того же параметра σ получатся различные оценки.

Пусть имеется m рядов измерений величины x , состоящих соответственно из n_1, \dots, n_m измерений. Пусть для этих m рядов величина $q = \sqrt{[vv]}$ имеет значения $q_1 = \sqrt{[vv]_{n_1}}, \dots, q_m = \sqrt{[vv]_{n_m}}$. Разделив q_i на число степеней свободы $f_i = n_i - 1$, получим m оценок для σ $s_1 = \frac{q_1}{\sqrt{f_1}}, \dots, s_m = \frac{q_m}{\sqrt{f_m}}$, полученных по каждому ряду измерений в отдельности. В предположении, что m рядов измерений независимы, получим функцию правдоподобия (10.10.3) согласно формуле (7.7.12) в виде

$$P(q_1, \dots, q_m; \sigma, f_1, \dots, f_m) = \frac{\text{const}}{\sigma^m} \left(\frac{q_1}{\sigma}\right)^{f_1-1} \dots \left(\frac{q_m}{\sigma}\right)^{f_m-1} \exp \left[-\frac{1}{2\sigma^2} (q_1^2 + \dots + q_m^2) \right]. \quad (1)$$

Таким образом, оценка максимального правдоподобия для σ есть решение уравнения

$$\frac{\partial L}{\partial \sigma} = \frac{\partial}{\partial \sigma} \ln P = -\frac{f}{\sigma} + \frac{q^2}{\sigma^3} = 0, \quad (2)$$

$$f = f_1 + \dots + f_m = n - m, \quad q^2 = q_1^2 + \dots + q_m^2,$$

так что эта оценка имеет вид

$$\sigma \approx s = \frac{q}{\sqrt{f}} = \sqrt{\frac{[vv]_n}{n-m}} = \sqrt{\frac{(n_1-1)s_1^2 + \dots + (n_m-1)s_m^2}{(n_1-1) + \dots + (n_m-1)}}. \quad (3)$$

Согласно формуле (10.10.6) асимптотическая дисперсия величины s , распределенной асимптотически нормально со средним значением σ , дается формулой

$$\sigma_0^2 \{s\} = \frac{-1}{\mathfrak{M} \left\{ \frac{\partial^2 L}{\partial \sigma^2} \right\}} = \frac{-1}{\mathfrak{M} \left\{ \frac{f}{\sigma^2} - 3 \frac{q^2}{\sigma^4} \right\}} = \frac{\sigma^2}{2f} \quad (4)$$

(проверьте). Таким образом,

$$\sigma \{s\} \sim \frac{\sigma}{\sqrt{2f}} \approx \frac{s}{\sqrt{2(n-m)}}. \quad (5)$$

При $m=1$ эти формулы сводятся к формуле (11.7.4). Мы видим, что хотя для единичного ряда наблюдений ($m=1$) несущественно, делить ли на n или на f , в случае использования для оценки σ нескольких рядов наблюдений это существенно, так как суммируются числа степеней свободы, а не числа наблюдений в сериях. *Это является единственной причиной того, что использование оценки s может оказаться более удобным, чем использование оценки s_1 .*

Описанный метод оценки параметра σ практически очень полезен и должен был бы использоваться гораздо шире, чем это имеет место в современной литературе. Так, например, часто для исключения грубых ошибок измерение повторяют, но в таких случаях повторное измерение недостаточно для оценки σ . Если же имеется ряд повторных измерений с одной и той же точностью, т. е. с одним и тем же σ (но, возможно, с различными значениями μ), то такие наблюдения могут быть использованы для точной оценки σ .

§ 11.11. Численное вычисление величин \bar{x} и s может быть упрощено, если принять за начало отсчета некоторое значение x^0 . Тогда

$$\bar{x} = \frac{[x]}{n} = \frac{[x^0 + v^0]}{n} = x^0 + \frac{[v^0]}{n}, \quad (1)$$

где

$$v_i^0 = x_i - x^0. \quad (2)$$

Для проверки вычисления можно вычислить кажущиеся ошибки $v_i = x_i - \bar{x}$ и проверить, удовлетворяют ли они соотношению (11.5.5), т. е. $[v] = 0$. Далее, полагая в формуле (11.5.6) $x' = x^0$, имеем

$$[v^2] = [v^0 v^0] - n(\bar{x} - x^0)^2 = [v^0 v^0] - \frac{[v^0]^2}{n}. \quad (3)$$

Пример. При 10 отсчетах по микрометру получены результаты x_1, \dots, x_{10} , приведенные в нижеследующей таблице. Используя (1) и (3), располагаем необходимые вычисления в следующем порядке:

i	x_i	v_i^0	$v_i^{0^2}$
1	4,078	$3 \cdot 10^{-3}$	$9 \cdot 10^{-6}$
2	4,080	5	25
3	4,071	-4	16
4	4,076	1	1
5	4,081	6	36
6	4,077	2	4
7	4,075	0	0
8	4,073	-2	4
9	4,079	4	16
10	4,069	-6	36
		$9 \cdot 10^{-3}$	$147 \cdot 10^{-6}$

$$x^0 = 4,075$$

$$[v^0] = 9 \cdot 10^{-3}$$

$$[v^0 v^0] = 147 \cdot 10^{-6}$$

$$[vv] = (147 - 81/10) \cdot 10^{-6} = 138,9 \cdot 10^{-6}$$

$$\bar{x} = 4,075 + \frac{9}{10} \cdot 10^{-3} = 4,0759$$

$$s = \sqrt{\frac{138,9}{9}} \cdot 10^{-3} = 0,00393$$

$$\frac{s}{\sqrt{10}} = 0,00124$$

$$\frac{s}{\sqrt{18}} = 0,000926$$

Результат:

$$\mu \approx \bar{x} = 4,0759 \text{ со стандартной ошибкой } 0,0012,$$

$$\sigma \approx s = 0,0039 \text{ со стандартной ошибкой } 0,0009.$$

Из этого примера видно, что в большинстве вычислений участвуют лишь малые целые числа, так что большинство вычислений можно делать в уме.

§ 11.12. Если наблюдения уже группированы и n столь велико, что длины интервалов группировки Δt_i малы, то можно с хорошим приближением заменить все значения, попавшие в один интервал группировки, значением t_i средней точки этого интервала. Таким образом, получим (см. § 10.5)

$$\bar{x} \sim \frac{n_1 t_1 + \dots + n_m t_m}{n_1 + \dots + n_m} = \frac{[nt]}{[n]} \quad (1)$$

и

$$[vv] \sim n_1 (t_1 - \bar{x})^2 + \dots + n_m (t_m - \bar{x})^2 = [n(t - \bar{x})^2]. \quad (2)$$

Если принять за начало отсчета значение x^0 , равное одному из чисел t_i , скажем, t^0 , то вычисления вновь будут очень простыми, особенно если длины всех интервалов группировки одинаковы. Действительно, полагая $t_i - t^0 = r_i \Delta t$, имеем

$$\bar{x} = t^0 + \Delta t \frac{[nr]}{n}, \quad (3)$$

$$[vv] = (\Delta t)^2 \left([nr^2] - \frac{[nr]^2}{n} \right)$$

(проверьте).

Пример. При обработке результатов 96 выстрелов по данным примера § 10.5 расположим вычисления, выбрав $\Delta t = 10$, следующим образом:

Боковое отклонение

t_i	n_i	r_i	r_i^2	$n_i r_i$	$n_i r_i^2$
-30	0	-3	9	0	0
-20	2	-2	4	-4	8
-10	9	-1	1	-9	9
0	28	0	0	0	0
10	30	1	1	30	30
20	21	2	4	42	84
30	5	3	9	15	45
40	1	4	16	4	16
50	0	5	25	0	0
	96			78	192

Отклонение по высоте

t_i	n_i	r_i	r_i^2	$n_i r_i$	$n_i r_i^2$
-60	0	-6	36	0	0
-50	3	-5	25	-15	75
-40	5	-4	16	-20	80
-30	13	-3	9	-39	117
-20	18	-2	4	-36	72
-10	21	-1	1	-21	21
0	21	0	0	0	0
10	10	1	1	10	10
20	5	2	4	10	20
30	0	3	9	0	0
	96			-111	395

$t^0 = 0$	$t^0 = 0$
$[nr] = 78$	$[nr] = -111$
$[nr^2] = 192$	$[nr^2] = 395$
$[vv] = 100 \left(192 - \frac{78^2}{96} \right) =$ $= 12863$	$[vv] = 100 \left(395 - \frac{111^2}{96} \right) =$ $= 26666$
$\bar{x} = 0 + \frac{10 \cdot 78}{96} =$ $= 8,125 (8,278)$	$\bar{x} = 0 - \frac{10 \cdot 111}{96} =$ $= -11,562 (-11,969)$
$s = \sqrt{\frac{12863}{95}} =$ $= 11,636 (11,393)$	$s = \sqrt{\frac{26666}{95}} =$ $= 16,754 (16,992)$
$\frac{s}{\sqrt{96}} = 1,188$	$\frac{s}{\sqrt{96}} = 1,710$
$\frac{s}{\sqrt{190}} = 0,844$	$\frac{s}{\sqrt{190}} = 1,215$

Результат:

Боковое отклонение

$$\mu \approx \bar{x} = 8,1 \text{ см со стандартной ошибкой } 1,2 \text{ см,}$$

$$\sigma \approx s = 11,6 \text{ см со стандартной ошибкой } 0,8 \text{ см.}$$

Отклонение по высоте

$$\mu \approx \bar{x} = -11,6 \text{ см со стандартной ошибкой } 1,7 \text{ см,}$$

$$\sigma \approx s = 16,8 \text{ см со стандартной ошибкой } 1,2 \text{ см.}$$

Для сравнения мы вычислили также точные значения \bar{x} и s (цифры в скобках) по формулам (11.7.3) и (11.7.4). Можно видеть, что ошибки, создаваемые группировкой, малы по сравнению со статистическими флуктуациями, характеризующимися стандартными ошибками. При вычислении ожидаемых теоретических распределений в примере § 10.5 мы использовали точные значения, но приближенные значения дали бы те же самые графики.

*** § 11.13. Коэффициент корреляции.** Если имеется одновременное измерение двух необязательно независимых физических величин, то теория ошибок базируется на двумерном нормальном распределении

(см. § 7.6), в котором мы будем теперь писать x вместо t и y вместо u . Применяя метод максимума правдоподобия к выборке, состоящей из n пар измерений $(x_1, y_1), \dots, (x_n, y_n)$, мы получим следующие оценки для параметров μ_x, μ_y, σ_x и σ_y :

$$\mu_x \approx \bar{x} = \frac{[x]}{n}, \quad \mu_y \approx \bar{y} = \frac{[y]}{n} \quad (1)$$

$$\sigma_x \approx s_x = \sqrt{\frac{[v_x v_x]}{n-1}}, \quad \sigma_y \approx s_y = \sqrt{\frac{[v_y v_y]}{n-1}} \quad (2)$$

$$v_{x_i} \approx x_i - \bar{x}, \quad v_{y_i} = y_i - \bar{y}.$$

Здесь в оценках для стандартных отклонений n , как и прежде, заменено на $n - 1$. Для коэффициента корреляции находим

$$\rho \approx r = \frac{[(x - \bar{x})(y - \bar{y})]}{(n-1)s_x s_y} = \frac{[v_x v_y]}{\sqrt{[v_x v_x][v_y v_y]}} \quad (3)$$

У п р а ж е н и е. Проверьте формулы (1) — (3) и найдите соответствующие стандартные ошибки. Однако, поскольку оценка r для ρ распределена не нормально, даже если n и очень велико, следует быть осторожным при использовании стандартной ошибки для r .

Для численного вычисления суммы $[v_x v_y]$ имеются следующие формулы, соответствующие аналогичным формулам для $[v u]$:

$$[v_x v_y] = [v_x^0 v_y^0] - \frac{[v_x^0][v_y^0]}{n}, \quad (4)$$

$$[v_x v_y] = [xy] - \frac{[x][y]}{n}, \quad (5)$$

$$[v_x v_y] = \Delta t_x \Delta t_y \left([n_{xy} r_x r_y] - \frac{[n_x r_x][n_y r_y]}{n} \right) \quad (6)$$

(проверьте).

Здесь $v_{x_i}^0 = x_i - x^0$, $v_{y_i}^0 = y_i - y^0$, $t_{x_i} - t_x^0 = r_x \Delta t_x$, $t_{y_i} - t_y^0 = r_y \Delta t_y$, $n_{x_i} = \sum_j n_{x_i y_j}$, $n_{y_i} = \sum_j n_{x_j y_i}$, причем (6) относится к *двумерной группировке* (см. § 10.5), которая состоит в следующем. Мы делим соответствующий интервал на оси x на $m_x (< n)$ не обязательно равных частей Δt_{x_i} , и интервал на оси y на $m_y (< n)$ не обязательно равных частей Δt_{y_j} .

Таким образом, мы получаем в плоскости $xу$ $m_x m_y$ прямоугольников, средние точки которых обозначаем (t_{x_i}, t_{y_j}) . Для каждого прямоугольника подсчитываем число $n_{x_i y_j}$ измерений, результаты которых лежат в прямоугольнике

$$t_{x_i} - \frac{\Delta t_{x_i}}{2} < x \leq t_{x_i} + \frac{\Delta t_{x_i}}{2}, \quad t_{y_j} - \frac{\Delta t_{y_j}}{2} < y \leq t_{y_j} + \frac{\Delta t_{y_j}}{2}. \quad (7)$$

Пример. При обработке результатов 96 выстрелов по данным § 10.5 мы имеем следующие значения:

Боковое отклонение x

	r_x					r_y					Всего n_y
	-30	-20	-10	0	10	20	30	40	50		
Отклонение по высоте y	-60	0	0	0	0	0	0	0	0	0	0
	-50	0	0	0	1	0	2	0	0	0	3
	-40	0	0	1	1	1	2	0	0	0	5
	-30	0	1	1	3	5	2	1	0	0	13
	-20	0	1	3	7	3	2	2	0	0	18
	-10	0	0	2	6	10	3	0	0	0	21
	0	0	0	1	6	6	6	1	1	0	21
	10	0	0	0	3	3	3	1	0	0	10
	20	0	0	1	1	2	1	0	0	0	5
	30	0	0	0	0	0	0	0	0	0	0
Всего n_x	0	2	9	28	30	21	5	1	0	96	

Такая таблица называется *корреляционной таблицей*. Распределения сумм столбцов и сумм строк суть *эмпирические частные распределения* (дающие частоты одной величины независимо от значений другой величины — см. § 4.13). Мы видим, что эти распределения совпадают с двумя распределениями из примера § 11.12.

Чтобы вычислить сумму $[n_{xy} r_x r_y]$, мы сначала вычисляем при каждом *фиксированном* x_i сумму $\sum_j n_{x_i y_j} r_{y_j} = [n_{x_i y} r_y]$, затем умножаем результат на r_{x_i} и суммируем по i . Используя результаты примера § 11.12, расположим вычисления следующим образом:

r_x	$[n_{x_i y} r_y]$	$[n_{x_i y} r_y] r_{xi}$
-3	0	0
-2	-5	10
-1	-13	13
0	-33	0
1	-28	-28
2	-26	-52
3	-6	-18
4	0	0
5	0	0
	-111	-75

$$[n_{xy} r_x r_y] = -75, \quad [v_x v_y] = 100 \left(-75 + \frac{78 \cdot 111}{96} \right) = 1518,7,$$

$$r = \frac{1518,7}{\sqrt{12863 \cdot 26666}} = 0,0820.$$

Заметим, что сумма чисел во втором столбце дает способ проверки, так как

$$\sum_i \sum_j n_{x_i y_j} r_{y_j} = \sum_j \left(\sum_i n_{x_i y_j} \right) r_{y_j} = \sum_j n_{y_j} r_{y_j} = [n_y r_y] = -111$$

согласно результатам примера § 11.12. На практике полное вычисление производится, конечно, по одной схеме, составленной из схем примера § 11.12 и схемы этого примера.

§ 11.14. В некоторых видах *статистического анализа* бывает нужно решить, равно ли нулю истинное значение нормально распределенной величины. Чтобы проверить, совместима ли оценка \bar{x} истинного значения μ с гипотезой $\mu = 0$, мы образуем новую случайную величину

$$t = \frac{\bar{x}}{s/\sqrt{n}}, \quad (1)$$

которая, как показано в упражнении 2 § 7.9, имеет t -распределение (7.9.3) с числом степеней свободы $f = n - 1$, если $\mu = 0$. Тогда по табл. III можно проверить, лежит ли наблюдаемое значение t в некоторых разумных пределах. На практике обычно выбирается пятипроцентный предел, который означает, что значение \bar{x} следует считать *значимо* отклоняющимся от нуля, т. е. отклоняющимся более, чем этого следует ожидать благодаря статистическим флуктуациям, если

соответствующее значение t лежит вне интервала, для которого $P(|t| \geq t) = 5\%$.

Пример. Рассматривая результаты 96 выстрелов по данным примера § 10.5, можно задать вопрос, являются ли отклонения пробойн от точки прицеливания $(0,0)$ настолько большими, что это указывает на систематические ошибки в прицельном механизме пулемета, или же они таковы, каких следует ожидать благодаря статистическим флуктуациям. По значениям, приведенным в примере § 10.5, мы получаем для бокового отклонения $t_x = \frac{8,125}{1,188} = 6,84$ и для отклонения по высоте $t_y = -\frac{11,562}{1,710} = -6,76$. Поскольку $f = 96 - 1 = 95$, из табл. III следует, что эти значения t лежат далеко за пятипроцентным пределом 1,986. Более того, при столь большом f величина t имеет распределение, очень близкое к нормальному, так что может быть использована табл. II, причем обнаруживается, что полученные выше значения t лежат даже за пределом, соответствующим коэффициенту доверия $P = 10^{-9}$. Следовательно, отклонения высоко значимы, так что рассмотренная серия выстрелов указывает на необходимость внимательного изучения причин ошибки.

§ 11.15. В других видах *статистического анализа* условия экспериментов изменяются таким образом, чтобы выяснить, оказывает ли влияние на результат экспериментов некоторые факторы (см. § 9.5). Осуществляя соответственно n_1 и n_2 одинаково точных измерений до и после изменения условий эксперимента, мы получим для истинного значения две оценки \bar{x}_1 и \bar{x}_2 . Тогда возникает проблема проверки того, совместимы ли эти оценки с гипотезой, что истинное значение исследуемой величины не изменилось, т. е. является ли разность этих оценок больше, чем этого следовало ожидать благодаря статистическим флуктуациям. Для проверки этого образуем случайную величину

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad s = \sqrt{\frac{f_1 s_1^2 + f_2 s_2^2}{f_1 + f_2}}, \quad (1)$$

$$f_1 = n_1 - 1, \quad f_2 = n_2 - 1,$$

которая, как показано в упражнении 3 § 7.9, имеет t -распределение (7.9.3) с числом степеней свободы $f = f_1 + f_2$. (Заметим, что как показано в § 11.10, s есть наилучшая оценка для σ , которую можно получить, объединив $n_1 + n_2$ измерений.) С помощью табл. III можно проверить, лежит ли значение t в разумных пределах. На практике принято выбирать пятипроцентные гределы и считать разность $\bar{x}_1 - \bar{x}_2$ *значимой*, если соответствующее значение лежит вне этих пределов. В этом случае нельзя считать, что \bar{x}_1 и \bar{x}_2 суть оценки одного и того

же истинного значения, или, другими словами, *чем больше значение t , тем более значимый эффект производит рассматриваемый фактор.*

Пример. При измерении некоторых спектральных линий возникает вопрос, была ли наблюдаена новая линия. Каждое измерение повторяется четыре раза. В первом измерении были получены следующие значения:

$$50,339, \quad 50,340, \quad 50,337 \quad \text{и} \quad 50,339,$$

для этих измерений

$$\bar{x}_1 = 50,33875 \sim 50,339 \quad \text{и} \quad s_1 = 0,00126 \sim 0,001;$$

во втором измерении получены значения

$$50,339, \quad 50,330, \quad 50,333 \quad \text{и} \quad 50,333,$$

для которых

$$\bar{x}_2 = 50,33375 \sim 50,334 \quad \text{и} \quad s_2 = 0,00378 \sim 0,004.$$

Таким образом, мы получаем $s = \sqrt{\frac{3s_1^2 + 3s_2^2}{6}} = 0,00281$ и

$$t = \frac{0,00500}{0,00281} \sqrt{2} = 2,52.$$

Поскольку $f = 4 + 4 - 2 = 6$, из табл. III видно, что значение t лежит вне пятипроцентного предела 2,447, но не превосходит однопроцентного предела 3,707. Следовательно, \bar{x}_1 и \bar{x}_2 отличаются значительно, и мы можем считать эти значения относящимися к различным спектральным линиям, хотя надежность этого вывода не очень велика.

§ 11.16. Слишком большое значение t в предыдущем параграфе может быть объяснено также тем, что измерения в двух различных сериях *не* одинаково точны, т. е. что эти измерения относятся к двум различным значениям σ . Однако это может быть проверено отдельно. Во многих видах *статистического анализа* важно проверить, является ли разница между двумя оценками для σ , полученными при измерениях двумя различными методами, *значимой*, т. е. большей, чем этого следует ожидать благодаря статистическим флуктуациям; иначе говоря, надо проверить, *одинаково ли точны два метода измерения*, т. е. имеют ли распределения вероятностей их результатов один и тот же параметр σ . Пусть s_1 и s_2 суть две оценки для σ , полученные по n_1 и по n_2 измерениям соответственно. Тогда образуем новую случайную величину

$$w^2 = \left(\frac{s_1}{s_2} \right)^2, \quad s_1 \geq s_2, \quad (1)$$

которая (см. § 7.11) имеет распределение (7.11.4) с числами степеней свободы $f_1 = n_1 - 1$, $f_2 = n_2 - 1$. По табл. V можно проверить, лежит ли наблюдаемое значение w^2 в разумных пределах. На практике выбирается пятипроцентный предел, $P(w^2 \geq w^2) = 5^0/0$, и если w^2 пре-

вышает этот предел, то разница между s_1 и s_2 считается *значимой*, т. е. два метода измерения не могут считаться одинаково точными.

Пример 1. В измерении спектральных линий, рассмотренном в примере § 11.15, получаем $s_1 = 0,00126$ и $s_2 = 0,00378$, т. е.

$$\omega^2 = \left(\frac{0,00378}{0,00126} \right)^2 = 9,00. \text{ Поскольку } f_1 = f_2 = 4 - 1 = 3, \text{ по табл.}$$

V мы убеждаемся, что полученное значение ω^2 не превышает пятипроцентного предела 9,28. Таким образом, разница между s_1 и s_2 не значима и, следовательно, нет причин считать, что измерения обладают различной точностью.

Пример 2. Для результатов 96 выстрелов в примере § 10.5 получаем $s_1 = 16,992$ и $s_2 = 11,393$, откуда $\omega^2 = (16,992/11,393)^2 = 2,23$. Из табл. V видно, что при $f_1 = f_2 = 96 - 1 = 95$ пятипроцентный предел меньше, чем 1,70. Таким образом, разница между s_1 и s_2 значима, и дисперсии боковых отклонений и отклонений по высоте нельзя считать равными.

§ 11.17. В практических приложениях теории ошибок важно, что измерения, содержащие грубые ошибки, бракуются, так как *даже единственная грубая ошибка может совершенно исказить результат ряда измерений*. Однако, поскольку теория ошибок базируется на нормальном распределении, допускающем в принципе появление сколь угодно больших ошибок, хотя и с пренебрежимо малыми вероятностями (см. табл. II), возникает вопрос, как отличить *грубые* ошибки от больших *случайных* ошибок. *Единственным надежным методом является браковка подозрительных измерений в процессе самого измерения*, если какие-либо обстоятельства заставляют думать, что условия эксперимента перестают быть неизменными (вибрации, внезапные изменения температуры и т. д.). В таких случаях необходимо тщательное исследование, чтобы решить, следует ли забраковать результат измерения. Однако на практике такое исследование может быть затруднительно или даже невозможно, и поскольку грубые ошибки, как правило, бросаются в глаза благодаря резкому их отличию от других измерений, *предложение считать величину ошибки единственным решающим критерием для браковки грубых ошибок оказывается очень заманчивым*, особенно если измерения производятся малоопытными наблюдателями.

Было предложено много правил для браковки грубых ошибок; особенный интерес уделяли этому в артиллерии¹⁾. Если бы параметры μ и σ были известны, можно было бы использовать подходящий доверительный предел, например $0,1^0/0$ — пределы $\mu \pm 3,29\sigma$. Однако,

¹⁾ При этом говорят о правилах браковки высказывающихся значений. Более детальное изложение таких правил см., например, в книге Arley, *Danske Vid. Selsk. Mat.-fys. Medd.*, т. XVIII, № 3, 1940, гл. IV, или Cranz, *Lehrbuch der Ballistik*, т. I, стр. 420.

как правило, могут быть использованы лишь оценки \bar{x} и s , так что наиболее естественно рассматривать относительные отклонения

$$r_i = \frac{x_i - \bar{x}}{\sqrt{\frac{n-1}{n}} s}, \quad (1)$$

каждое из которых, как показано в § 7.10, имеет частное распределение в виде r -распределения (7.10.2) с $f = n - 2$ степенями свободы. При больших значениях n это распределение практически нормально, так что пределы для r приближаются к пределам, получаемым по нормальному распределению. С помощью табл. IV можно проверить, лежат ли относительные отклонения в разумных пределах. При этом можно считать подходящими 0,1% пределы, $P(|r| \geq r) = 0,1\%$, и, следовательно, браковать измерение, если соответствующее значение r оказывается вне этих пределов.

Пример. Для 10 измерений из примера § 11.11 находим следующие относительные отклонения:

0,563	0,295
1,100	—0,241
—1,315	—0,778
0,027	0,832
1,368	—1,851

Поскольку $f = 10 - 2 = 8$, по табл. IV можно убедиться, что все эти значения r лежат внутри 0,1% пределов 2,616 и даже внутри 5% пределов 1,895. Поэтому нет никаких уважительных причин подозревать, что какое-либо из имеющихся измерений содержит грубые ошибки.

Заметим, что любое правило браковки грубых ошибок следует использовать с некоторой осторожностью, особенно при малых значениях n .

Упражнение. Покажите, что для четырех измерений 21,790; 21,789; 21,789; 21,789 первое относительное отклонение равно максимальному значению r , именно $\sqrt{3}$. С этой точки зрения первое измерение, конечно, браковать еще не следует.

*** § 11.18.** Важным видом *статистического анализа* является исследование зависимости или *корреляции* между некоторыми явлениями (см. § 9.5). Если можно предполагать, что совместная функция распределения величин x и y нормальна, то нужно найти способ проверки гипотезы о том, что истинное значение коэффициента корреляции $\rho\{x, y\}$ равно нулю. Для проверки этого вычисляем наилучшую оценку для ρ — эмпирический коэффициент корреляции r_{xy} , определяемый формулой (11.13.3), — и образуем новую случайную величину

$$r = \sqrt{n-1} r_{xy}. \quad (1)$$

Можно показать, что r имеет при $\rho=0$ r -распределение (7.10.2) с числом степеней свободы $f=n-2$ ¹⁾. С помощью табл. IV можно проверить, лежит ли r в разумных пределах. Выбирая в качестве таковых пятипроцентные пределы, $P(|r| \geq r) = 5\%$, будем считать наблюдаемое значение r *значимо* отклоняющимся от нуля, если оно оказывается вне этих пределов; иначе говоря, в этом случае гипотезу $\rho=0$ следует считать несовместимой с имеющейся выборкой, так что величины x и y следует считать коррелированными.

Пример. По результатам 96 выстрелов из пулемета, приведенным в примере § 10.5, в примере § 11.13 была получена оценка $r_{xy} = 0,0820$ коэффициента корреляции между боковыми отклонениями попаданий и отклонениями по высоте. Таким образом $r = \sqrt{95} \cdot 0,0820 = 0,799$. Поскольку $f = 96 - 2 = 94$, по табл. IV видно, что полученное значение r лежит ниже пятипроцентного предела, равного 1,956. Более того, поскольку r -распределение при больших значениях f очень близко к нормальному, может быть использована табл. II, которая показывает, что полученное значение r лежит даже ниже сорокапроцентного предела, равного 0,8416. Таким образом, r не отклоняется значимо от нуля, и, следовательно, между боковыми отклонениями и отклонениями по высоте попаданий в цель из исследуемого пулемета корреляции не обнаруживается.

§ 11.19. Общие характерные черты методов, описанных в § 11.14—11.18, заключаются, во-первых, в том, что в этих методах не предполагается каких-либо сведений об истинных значениях параметров, известных на практике лишь очень редко; во-вторых, эти методы учитывают наличие статистических флуктуаций, что очень полезно, так как на практике часто приходится ограничиваться *малыми* выборками, состоящими лишь из небольшого количества наблюдений; в-третьих, эти методы дают возможность полного контроля надежности статистических выводов. Однако надо подчеркнуть, что различные рассмотренные критерии позволяют делать лишь *отрицательные* выводы. Так, например, если гипотеза заключается в том, что истинное значение равно нулю, а наблюдения дают значение t , имеющее лишь очень малую вероятность, то очевидно, что гипотеза не находится в хорошем согласии с данными наблюдений. Но если рассмотреть две гипотезы относительно истинного значения, заключающиеся в том, что это значение равно μ_1 или μ_2 , и если наблюдения дают для этих двух гипотез значения t , лежащие, скажем, внутри соответственно 90 и 70% пределов, то из этого нельзя заключить, что гипотеза μ_1 лучше, чем гипотеза μ_2 , а можно лишь сказать, что обе гипотезы совместимы с наблюдаемыми данными. Более того, следует вновь

¹⁾ См., например, К р а м е р, Математические методы статистики, § 29.7, 1948. Таблица, дающая непосредственно распределение для r_{xy} , приведена в книге Fisher and Yates, Statistical Tables, Table VI.

подчеркнуть, что выбор именно пятипроцентного предела для отбрасывания неправдоподобных гипотез совершенно произволен.

Единственное, что говорит в пользу выбора пятипроцентного значения, это то, что на практике это значение оказывается вполне подходящим: с одной стороны, оно достаточно велико для фактического отбрасывания „неверных“ гипотез, с другой стороны, оно достаточно мало, так что приводит к браковке лишь немногих „верных“ гипотез, большие отклонения от которых все же могут случайным образом возникнуть.

Если, однако, в нашем распоряжении нет таблиц, используемых в различных критериях, проверка гипотез может быть произведена более примитивным способом — в пренебрежении статистическими флуктуациями величины s . Так, если нужно выяснить, согласуются ли друг с другом два средних значения, то это можно сделать приближенно, вычислив дисперсию их разности и проверив значимость этой разности с помощью допустимых пределов для нормального распределения. Если разность во много раз превосходит свою дисперсию, то она, несомненно, значима, но если она того же порядка, что и дисперсия, и число наблюдений мало, как это имеет место в примере § 11.15, то надежные выводы могут быть получены лишь с помощью таблицы значений t , так как в таких случаях статистическими флуктуациями величины s пренебрегать нельзя.

§ 11.20. В практических приложениях теории ошибок большое значение имеет вопрос о том, сколь много следует сделать наблюдений. Конечно, это зависит от конкретных обстоятельств, в особенности от тех сведений о статистических флуктуациях, которые могут быть известны заранее. *При точных измерениях следует требовать, чтобы дисперсия среднего значения \bar{x} была мала по сравнению с \bar{x} и чтобы дисперсия стандартного отклонения s была мала по сравнению с s .* Вообще говоря, результаты, полученные по 10 измерениям, можно считать надежными, а более чем 20 измерений производят лишь очень редко, что диктуется соображениями экономии как времени, так и затрачиваемых средств.

Поскольку дисперсия среднего значения \bar{x} стремится к нулю при неограниченном возрастании числа наблюдений n , в принципе можно было бы думать, что неограниченное увеличение точности определения истинного значения возможно скорее при увеличении числа измерений, чем при использовании более точных измерительных инструментов. Однако такое мнение ошибочно, хотя оно и встречается довольно часто.

Пример. При измерении расстояния между двумя отметками на стальной балке, описанном в § 11.3, мы получаем для \bar{x} дисперсию, в точности равную нулю, но вряд ли кто-либо сделал при этом заключение, что истинная длина в точности равна 2 м. Если измерить

затем эту же самую длину с помощью линейки, разделенной лишь на сантиметры, то дисперсия величины \bar{x} при достаточно большом количестве измерений может иметь порядок 10^{-4} м. Однако вряд ли кто-либо мог бы ожидать, что полученная при таком измерении длина будет совпадать со значением 2,0078 м, полученным при измерении с помощью более точных измерительных инструментов.

Следует запомнить, во-первых, что \bar{x} и s суть лишь оценки параметров теоретического распределения, с помощью которых мы описываем имеющиеся наблюдения, и хотя дисперсии этих оценок стремятся к нулю при $n \rightarrow \infty$, само s не стремится к нулю, а сходится по вероятности к постоянной σ , отличной от нуля. Во-вторых, следует помнить, что мы совершенно произвольно *определили* истинное значение как среднее значение μ соответствующей функции распределения.

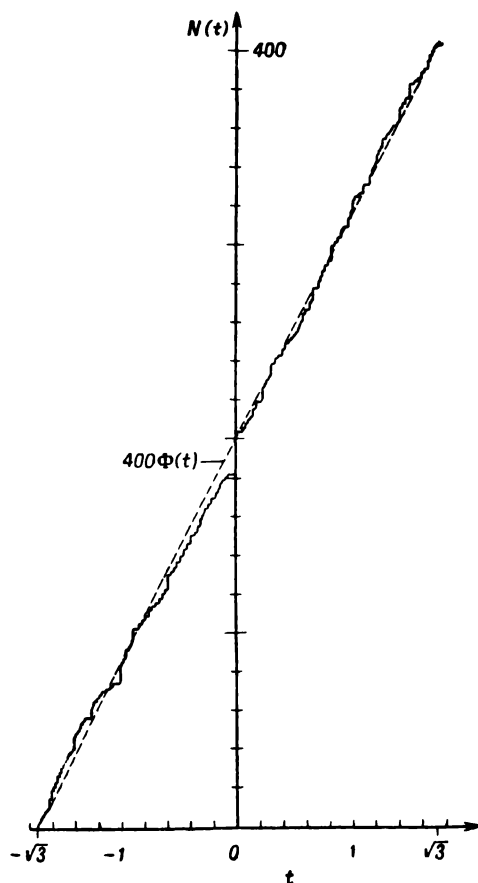
Любое значение, скажем, в пределах $\mu \pm \sigma$ может быть столь же законно названо „истинным“ значением. В-третьих, следует помнить, что при нашем определении истинное значение зависит также от метода измерения. Так, например, если иметь дело с тремя различными методами измерения, то мы будем иметь три различные случайные величины, каждую со своей собственной функцией распределения. Наконец, следует помнить, что использование среднего арифметического как наилучшей оценки для параметра μ , основано на той гипотезе, что теорию ошибок можно базировать на нормальном распределении. Но эта гипотеза может быть проверена лишь с некоторой степенью надежности (см. следующий параграф), поскольку любое *эмпирически* найденное распределение дискретно, и поэтому если колебания измеренных значений велики по сравнению с наименьшей доступной измерению единицей, то эмпирическое распределение лишь с некоторым приближением может быть описано посредством непрерывного распределения, в частности нормального.

По всем этим причинам, вообще говоря, бессмысленно сохранять в значениях \bar{x} и s излишнее количество значащих цифр сверх одной, добавленной к количеству значащих цифр в результатах отдельных измерений.

§ 11.21. В этом параграфе будет показано, как можно проверить основную гипотезу теории ошибок — гипотезу о том, что результаты непосредственных измерений могут быть удовлетворительно описаны с помощью нормального распределения.

Простейший прием заключается в построении гистограммы и ее сравнении с графиком нормальной плотности вероятности $\frac{1}{\sigma} \phi \left(\frac{x - \mu}{\sigma} \right)$, в которой для μ и σ берутся значения \bar{x} и s и для вычисления которой используется табл. I. Этот прием был проиллюстрирован в примере § 10.5.

Другой и более совершенный прием заключается в построении полигона сумм и его сравнении с графиком нормальной функции распределения $\Psi\left(\frac{x-\mu}{\sigma}\right)$, в которой для μ и σ берутся значения \bar{x} и s . Этот прием был продемонстрирован в примере § 10.6.



Фиг. 19.

Еще более совершенный прием заключается в использовании метода выпрямленной диаграммы, изложенного в § 10.7, при котором график нормальной функции распределения преобразуется в прямую линию. Этот прием был проиллюстрирован в примере 1 § 10.7.

Как правило, на практике мы располагаем лишь малыми выборками, состоящими из небольшого количества результатов измерений. В таких случаях указанные методы непригодны, так как они основаны на использовании большого количества наблюдений. Однако часто можно располагать большим количеством малых выборок, соответствующих различным значениям как μ , так и σ , но в которых измеренные величины и использованные методы измерения имеют одну и ту же природу, так что можно с уверенностью предполагать, что функции распределения однотипны, именно, согласно основной гипотезе, нормальны.

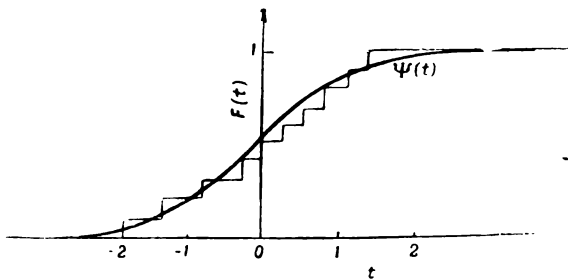
Тогда можно вычислить все относительные ошибки (11.17.1) и сравнить их распределение с теоретическим r -распределением, определяемым формулой (7.10.2) и протабулированным в табл. IV *).

*) Табл. IV в конце книги недостаточно подробна для этой цели. О более подробной таблице см. примечание на стр. 110. (Прим. ред.)

Пример 1. На фиг. 19 изображен полигон сумм $400 F(t) = N(t)$ для 400 относительных ошибок, вычисленных по 100 выборкам, состоящим каждая из четырех измерений спектральных линий¹⁾. Для сравнения на график нанесена соответствующая теоретическая кривая $400 \Phi(t)$, которая согласно формуле (7.10.2) при $f = 4 - 2 = 2$ изображается отрезком прямой, соединяющим точки $(-\sqrt{3}, 0)$ и $(\sqrt{3}, 400)$, и₂ означает, что при $f = 2$ относительные ошибки равномерно распределены по интервалу $-\sqrt{3}, +\sqrt{3}$. Из графика видно, что согласие очень хорошее, так как в этом случае колебания столь малы, что их нельзя считать большими по сравнению с единицей измерения — условии, обычно необходимое для аппроксимации эмпирического распределения с помощью непрерывной функции распределения.

Следует заметить, что, строго говоря, распределения относительных ошибок с номерами 1, 2, 3 и 4 надо рассматривать по отдельности, так как в противном случае относительные ошибки нельзя считать независимыми (см. работу Арлея¹⁾). Однако если количество выборок велико, то этой зависимостью можно пренебречь, как это доказано в цитированной работе Арлея¹⁾.

Пример 2. Наконец, можно также подсчитать относительные ошибки для *единичной* выборки и сравнить их полигон сумм с нормированной нормальной функцией распределения $\Psi(t)$. Однако, поскольку g -распределение лишь приблизительно нормально при больших



Ф и г. 20.

значениях n и поскольку относительные ошибки в пределах единичной выборки не независимы, такое сравнение может дать лишь грубый критерий для проверки того, имеют ли рассматриваемые величины нормальное распределение¹⁾. На фиг. 20 изображен полигон сумм для 10 результатов измерений, приведенных в примере § 11.11, и для сравнения изображена функция $\Psi(t)$. График показывает, что согласие неплохое.

¹⁾ Этот пример заимствован из работы Arley, *Danske Vid. Selsk. Mat.-fys. Medd.*, т. XVIII, № 3, 1940.

§ 11.22. До сих пор рассматривались лишь непосредственные измерения, но, как было указано в § 11.5, большинство физических измерений суть измерения косвенные, т. е. результаты получаются не непосредственно по данным измерений, а рассматриваемая величина z есть функция других величин x_1, x_2, \dots, x_v , измеряемых непосредственно. Считая x_1, \dots, x_v независимыми и допустив, что указанные в § 6.5 условия выполнены, мы можем считать, что среднее значение и дисперсия величины z с хорошим приближением определяются формулами (6.5.9) и (6.5.10). Пользуясь вместо параметров оценками

($\mu_1 \approx \bar{x}_1, \dots, \mu_v \approx \bar{x}_v, \sigma_1 \approx s_1, \dots, \sigma_v \approx s_v$), получим

$$\mathfrak{M}\{z\} \approx \bar{z} \sim f(\bar{x}_1, \dots, \bar{x}_v), \quad (1)$$

$$\sigma^2\{z\} \approx s^2 \sim \left(\frac{\partial f}{\partial x_1}\right)^2 s_1^2 + \dots + \left(\frac{\partial f}{\partial x_v}\right)^2 s_v^2, \quad (2)$$

причем в (2) частные производные берутся в точке $(\bar{x}_1, \dots, \bar{x}_v)$.

Во многих случаях, встречающихся особенно часто в физике и в технологии, z есть так называемая *логарифмируемая* функция, т. е.

$$z = c \cdot x_1^{k_1} \cdot x_2^{k_2} \cdot \dots \cdot x_v^{k_v}, \quad (3)$$

где k_1, \dots, k_v суть положительные или отрицательные постоянные, а c также есть постоянная. Тогда по формулам (6.5.9) и (6.5.10) получим

$$\sigma^2\{z\} \approx \mathfrak{M}^2\{z\} \left[k_1^2 \left(\frac{\sigma_1}{\mu_1}\right)^2 + \dots + k_v^2 \left(\frac{\sigma_v}{\mu_v}\right)^2 \right], \quad (4)$$

что может быть записано также в виде

$$\left(\frac{\sigma_z}{\mu_z}\right)^2 \approx k_1^2 \left(\frac{\sigma_1}{\mu_1}\right)^2 + \dots + k_v^2 \left(\frac{\sigma_v}{\mu_v}\right)^2. \quad (5)$$

Упражнение. Проверьте эти формулы. Покажите также, что если

$$k_1 \frac{\sigma_1}{\mu_1} \approx k_2 \frac{\sigma_2}{\mu_2} \approx \dots \approx k_v \frac{\sigma_v}{\mu_v},$$

то формула (5) может быть записана приближенно в виде

$$\frac{\sigma_z}{\mu_z} \sim \frac{1}{\sqrt{v}} \left(k_1 \frac{\sigma_1}{\mu_1} + \dots + k_v \frac{\sigma_v}{\mu_v} \right). \quad (6)$$

Эта формула широко распространена, особенно в инженерном деле; она всегда дает несколько меньшее значение, чем формула (5), за исключением случаев, когда отдельные слагаемые строго равны (проверьте). Поэтому всегда следует использовать формулу (5), применение которой почти так же просто, как и формулы (6).

Пример 1. Рассмотрим измерение удельного сопротивления $\rho = \pi R d^2 / 4l$ и предположим, что полное сопротивление R , диаметр d и длина l были измерены с коэффициентами изменчивости 2% , 3%

и 3⁰/₀ соответственно. По формуле (5) получаем

$$\left(\frac{\sigma_\rho}{\mu_\rho}\right)^2 \approx \left(\frac{\sigma_R}{\mu_R}\right)^2 + 4 \left(\frac{\sigma_d}{\mu_d}\right)^2 + \left(\frac{\sigma_l}{\mu_l}\right)^2 = 49,$$

т. е.

$$\frac{\sigma_\rho}{\mu_\rho} \approx 7^0/0.$$

По формуле (6) мы получили бы $\sigma_\rho/\mu_\rho \sim 11/\sqrt{3^0/0} = 6^0/0$.

Формула (5) часто используется при планировании измерений. Так, если коэффициент изменчивости величины z не должен превосходить заданного значения и если, кроме того, считать, что отдельные измерения должны производить одинаковый вклад в коэффициент изменчивости величины z , то можно найти коэффициенты изменчивости, с которыми следует производить непосредственные измерения величин x_1, \dots, x_v .

Пример 2. Если в примере 1 потребовать, чтобы коэффициент изменчивости величины ρ не превосходил 1⁰/₀, то получаем

$$\begin{aligned} \left(\frac{\sigma_R}{\mu_R}\right)^2 &< \frac{1}{3} \cdot 0,01^2, & \text{т. е. } \frac{\sigma_R}{\mu_R} &< 0,6^0/0, \\ 4 \left(\frac{\sigma_\alpha}{\mu_\alpha}\right)^2 &< \frac{1}{3} \cdot 0,01^2, & \text{т. е. } \frac{\sigma_\alpha}{\mu_\alpha} &< 0,3^0/0, \\ \left(\frac{\sigma_l}{\mu_l}\right)^2 &< \frac{1}{3} \cdot 0,01^2, & \text{т. е. } \frac{\sigma_l}{\mu_l} &< 0,6^0/0. \end{aligned}$$

Следует заметить, что на практике равномерное распределение дисперсии z по отдельным измерениям не всегда удобно использовать, поскольку при планировании измерений следует уделять внимание имеющимся сведениям о возможности уменьшить дисперсии различных непосредственных измерений.

12. ПРИЛОЖЕНИЕ ТЕОРИИ ВЕРОЯТНОСТЕЙ К ТЕОРИИ УРАВНИВАНИЯ

§ 12.1. В теории ошибок, изложенной в разделе 11, рассматривался простейший случай n независимых измерений одной и той же физической величины. В теории уравнивания рассматривается более общий случай n независимых нормально распределенных измерений *различных* величин, не независимых, но, как известно заранее, связанных некоторыми соотношениями. Так, при составлении земельных планов рассматриваются три точки A , B и B , отмеченные на участке, и если измеряются три угла треугольника ABB , то их сумма должна быть равна 180° . Но, как было рассказано в § 11.3, результат любого физического измерения есть случайная величина, так что указанная сумма, вообще говоря, будет несколько отличаться от 180° .

Рассмотрим общий случай, в котором осуществляются n независимых измерений, дающих в результате значения l_1, \dots, l_n n физических величин l_1, \dots, l_n , истинные значения (т. е. средние значения) которых $\lambda_1, \dots, \lambda_n$ удовлетворяют $r < n$ соотношениям¹⁾

$$F_j(\lambda_1, \dots, \lambda_n) = 0, \quad j = 1, 2, \dots, r, \quad (1)$$

называемым *фундаментальными уравнениями*. Если в эти уравнения вместо $\lambda_1, \dots, \lambda_n$ подставить наблюдаемые значения l_1, \dots, l_n , то уравнения, вообще говоря, не будут удовлетворяться. Целью теории уравнивания является получение по измеренным значениям наилучших оценок $\bar{l}_1, \dots, \bar{l}_n$ параметров $\lambda_1, \dots, \lambda_n$, которые подобно самим параметрам удовлетворяли бы фундаментальным уравнениям²⁾. Эта задача решается с помощью метода максимума правдоподобия (§ 10.10). Если стандартное отклонение величины l_i обозначить через σ_i , то

1) См. примечание 2 на стр. 147. В этом разделе вследствие большого количества имеющихся величин более удобно обозначать средние значения соответствующими греческими буквами, а их оценки — соответствующими латинскими буквами с черточкой сверху, поскольку эти оценки будут получены как простые обобщения средних арифметических значений, именно как средние взвешенные значения (см. § 6.4).

2) Решение этой задачи называется несколько неудачно „уравниванием“, что может дать повод к такому неправильному пониманию, будто следует немного „изменить“ измеренные значения, чтобы они удовлетворяли соотношению (1). Таким образом, термин „теория уравнивания“, сейчас всеупотребительный, следует признать неудачным.

Функция правдоподобия (10.10.3) примет вид

$$\begin{aligned}
 P(l_1, \dots, l_n; \lambda_1, \dots, \lambda_n; \sigma_1, \dots, \sigma_n) &= \\
 &= \left(\frac{1}{\sqrt{2\pi}} \right)^n \frac{1}{\sigma_1 \dots \sigma_n} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \frac{(l_i - \lambda_i)^2}{\sigma_i^2} \right\} = \\
 &= \left(\frac{1}{\sqrt{2\pi}} \right)^n \frac{1}{\sigma_1 \dots \sigma_n} \exp \left\{ -\frac{1}{2} \left[\frac{1}{\sigma^2} \varepsilon \varepsilon \right] \right\}, \quad (2)
 \end{aligned}$$

где

$$\varepsilon_i = l_i - \lambda_i, \quad i = 1, 2, \dots, n \quad (3)$$

— так называемые „истинные“ ошибки системы, а $[\]$ — символ Гаусса для суммы (11.5.2), употребляемый как в теории ошибок, так и в теории уравнивания. Так как на практике отношения чисел σ_i , как правило, известны, удобно ввести n положительных чисел p_1, \dots, p_n и коэффициент пропорциональности σ , определяемые формулой

$$p_1 \sigma_1^2 = p_2 \sigma_2^2 = \dots = p_n \sigma_n^2 = \sigma^2. \quad (4)$$

Числа p_i называются *весами* n измерений, а (произвольный) коэффициент пропорциональности σ — *стандартным отклонением* (или средней ошибкой измерения единичного веса).

Поскольку $\sigma^2 \left\{ \frac{[x]}{n} \right\} = \frac{\sigma^2 \{x\}}{n}$, вес среднего арифметического n равнозначных измерений в n раз больше, чем вес единичного измерения. Это обстоятельство и служит причиной наименования „веса“, поскольку среднее арифметическое содержит все n измерений и поэтому имеет такой же вес, как все эти n измерений вместе (см. пример 1 § 12.9). Далее, на практике веса часто появляются именно таким образом, поскольку каждая величина l_i измеряется некоторое количество n_i раз, и полученное среднее арифметическое подставляется в уравнения как непосредственно измеренное значение l_i с весом n_i .

Вводя веса в формулу (2), получим

$$P(l_1, \dots, l_n; \lambda_1, \dots, \lambda_n; \sigma) = \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n (p_1 \dots p_n)^{1/2} \exp \left[-\frac{1}{2\sigma^2} [p\varepsilon\varepsilon] \right]. \quad (5)$$

Согласно формуле (10.10.5) наилучшие оценки $\bar{l}_1, \dots, \bar{l}_n$ для параметров $\lambda_1, \dots, \lambda_n$ суть такие их значения, при которых величина (5) имеет максимум при фиксированных значениях l_1, \dots, l_n и которые в то же время удовлетворяют фундаментальным уравнениям (1). Поскольку $\lambda_1, \dots, \lambda_n$ входят лишь в сумму $[p\varepsilon\varepsilon]$, предыдущее утверждение эквивалентно тому, что $[p\varepsilon\varepsilon]$ должно быть минимально. Ошибки

$$v_i = \bar{l}_i - l_i, \quad i = 1, 2, \dots, n, \quad (6)$$

при которых $[p\varepsilon\varepsilon]$ достигает минимального значения, называются

кажущимися поправками системы ¹⁾, и мы получаем

$$[p\bar{v}\bar{v}] = p_1(\bar{l}_1 - l_1)^2 + \dots + p_n(\bar{l}_n - l_n)^2 = \min, \quad (7)$$

причем $[p\bar{v}\bar{v}]$ называется *взвешенной суммой квадратов ошибок*. Принцип (7) называется *методом наименьших квадратов* и является обобщением принципа (11.5.7). Подчеркнем, что в (7) $l_1, \dots, l_n, p_1, \dots, p_n$ суть известные величины, а $\bar{l}_1, \dots, \bar{l}_n$ — неизвестные величины.

§ 12.2. Различают два метода уравнивания: *уравнивание с помощью элементов* (уравнивание косвенных измерений) и *уравнивание с помощью коррелят* (уравнивание условных измерений).

При *уравнивании с помощью элементов* фундаментальные уравнения (12.1.1) задаются в *параметрической форме*, т. е. истинные значения $\lambda_1, \dots, \lambda_n$ задаются в виде функций некоторого числа $m < n$ независимых параметров ξ_1, \dots, ξ_m , однозначно определяющих систему и называющихся ее *элементами*:

$$\lambda_i = f_i(\xi_1, \dots, \xi_m), \quad i = 1, 2, \dots, n \quad (n > m). \quad (1)$$

Так, например, плоский треугольник определяется одной стороной и двумя углами или тремя сторонами, но не тремя углами. Вообще говоря, элементы можно выбирать многими различными способами, и в каждом отдельном случае следует решить, какое избрать множество элементов. При этом следует соблюдать лишь такие условия, что система должна быть однозначно определена выбранными элементами и что они должны быть независимыми, не связанными никакими ограничениями за исключением того, что их значения принадлежат некоторым интервалам (область их определения). Так, в случае плоского треугольника в качестве элементов нельзя выбрать три угла, так как, во-первых, треугольник не определяется однозначно своими тремя углами и, во-вторых, эти углы не независимы, а связаны условием того, что их сумма равна 180° .

Если система уже измерена, то проблема заключается в нахождении наилучших оценок $\bar{x}_1, \dots, \bar{x}_n$ для истинных значений ξ_1, \dots, ξ_m m элементов x_1, \dots, x_n , т. е. нахождении их значений, которые, с одной стороны, удовлетворяют r фундаментальным уравнениям (1) и, с другой стороны, обращают в минимум сумму $[p\bar{v}\bar{v}]$. Однако это не всегда эквивалентно непосредственному измерению самих элементов, что может быть даже и невозможным, и поэтому их приходится измерять не непосредственно, а посредством измерения других частей системы, являющихся функциями от этих элементов. Измерив столько величин, сколько элементов имеет система, мы получаем m уравнений

¹⁾ Заметим, что величины v_1, \dots, v_n , определяемые формулой (6), не являются ошибками, а суть поправки (см. примечание на стр. 166). В теории уравнивания более естественно рассматривать не ошибки, а поправки.

для определения m неизвестных элементов, которые, таким образом, остается определить математически. Однако, как было рассказано в § 11.5, одного измерения неизвестной величины недостаточно. Поэтому, чтобы уменьшить влияние статистических флуктуаций и оценить их величину, измеряют большее количество величин, чем число неизвестных; в последующем мы предполагаем, что $n > m$. Говорят, что при этом имеют дело с *переопределенной* системой, а дополнительные измерения называются *избыточными*.

При *уравнивании с помощью коррелят* фундаментальные уравнения (12.1.1) задаются в виде $r < n$ уравнений, не разрешенных относительно параметров λ :

$$f_j(\lambda_1, \dots, \lambda_n) = 0, \quad j = 1, 2, \dots, r \quad (n > r). \quad (2)$$

Имея здесь n неизвестных, связанных r соотношениями, в принципе можно произвольно выбрать $n - r = m$ независимых параметров λ , считать их элементами и выразить остальные $n - m = r$ параметров λ через выбранные элементы. Следовательно, опять можно решать проблему, применяя уравнивание с помощью элементов, но идея уравнивания с помощью коррелят как раз и заключается в том, чтобы избежать решения уравнений (2), оперируя непосредственно с неразрешенными уравнениями. Обратное, в принципе можно исключить элементы из (1) и получить, таким образом, фундаментальные уравнения в форме (2). *Таким образом, уравнивание с помощью элементов и уравнивание с помощью коррелят суть лишь два различных метода решения одной и той же проблемы, дающие, конечно, одни и те же результаты.* В практике встречаются и смешанные методы уравнивания, при которых некоторые из фундаментальных уравнений задаются в параметрической форме, а другие — в виде уравнений, не разрешенных относительно параметров λ . Здесь мы будем рассматривать лишь два указанных основных метода¹⁾.

Подставляя в фундаментальные уравнения (1) или (2) вместо параметров $\lambda_1, \dots, \lambda_n$ их оценки $\bar{l}_1, \dots, \bar{l}_n$, определяемые из формулы (12.1.6), и вместо параметров ξ_1, \dots, ξ_m их оценки $\bar{x}_1, \dots, \bar{x}_n$, мы получим при уравнивании с помощью элементов уравнения

$$l_i + v_i = f_i(\bar{x}_1, \dots, \bar{x}_n), \quad i = 1, 2, \dots, n, \quad (3)$$

т. е. n уравнений с $n + m$ неизвестными v_1, \dots, v_n и $\bar{x}_1, \dots, \bar{x}_n$.

При уравнивании с помощью коррелят получаем уравнения

$$f_j(l_1 + v_1, \dots, l_n + v_n) = 0, \quad j = 1, 2, \dots, r, \quad (4)$$

т. е. r уравнений с n неизвестными v_1, \dots, v_n . В обоих случаях число получаемых уравнений меньше числа неизвестных. Недостаю-

¹⁾ Изложение смешанных методов уравнивания читатель найдет в учебниках по теории уравнивания, указанных в списке литературы.

щие m уравнений получаются из условия $[p_{vv}] = \min$, как это вскоре будет показано.

Уравнения (3) или (4) называются *условными уравнениями*, и мы будем здесь предполагать, что эти уравнения *линейны* (нелинейный случай будет рассмотрен отдельно в § 12.13):

$$l_i + v_i = a_{i0} + a_{i1}\bar{x}_1 + a_{i2}\bar{x}_2 + \dots + a_{im}\bar{x}_m, \quad i = 1, 2, \dots, n \quad (5)$$

или

$$\{a_{j0} + a_{j1}(l_1 + v_1) + a_{j2}(l_2 + v_2) + \dots + a_{jn}(l_n + v_n) = 0, \\ j = 1, 2, \dots, r. \quad (6)$$

Но при этом не только неудобно выписывать все эти уравнения, но и ясное изложение метода уравнивания затруднительно из-за длинных выкладок. Поэтому мы предпочтем пользоваться более простой и удобной записью, располагая для этого идеальным средством в виде *матричной* символики, описанной вкратце в приложении 2.

Уравнивание с помощью элементов

§ 12.3. Введем следующие матрицы:

$$L_{n1} = \begin{Bmatrix} l_1 \\ l_2 \\ \cdot \\ \cdot \\ l_n \end{Bmatrix}, \quad V_{n1} = \begin{Bmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_n \end{Bmatrix}, \quad A_{nm} = \begin{Bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{Bmatrix}, \\ A_{n1}^0 = \begin{Bmatrix} a_{10} \\ a_{20} \\ \cdot \\ \cdot \\ a_{n0} \end{Bmatrix}, \quad \bar{X}_{m1} = \begin{Bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \cdot \\ \cdot \\ \bar{x}_m \end{Bmatrix} \quad (1)$$

Тогда условные уравнения (12.2.5) можно записать в виде

$$\bar{L}_{n1} = L + V = A^0 + A \cdot \bar{X}. \quad (2)$$

Обозначим для сокращения

$$N_{n1} = L - A^0 \quad (3)$$

и запишем условные уравнения в виде

$$V_{n1} = -N + A \cdot \bar{X}. \quad (4)$$

В (4) мы имеем n уравнений относительно $n + m$ неизвестных V и \bar{X} , а N и A суть заданные величины. Недостающие m уравнений получим из условия $[p\bar{v}v] = \min$, поскольку для этого необходимо, чтобы все m частных производных от $[p\bar{v}v]$ по $\bar{x}_1, \dots, \bar{x}_m$ равнялись нулю:

$$\frac{\partial}{\partial \bar{x}_j} [p\bar{v}v] = 2 \left(p_1 v_1 \frac{\partial v_1}{\partial \bar{x}_j} + \dots + p_n v_n \frac{\partial v_n}{\partial \bar{x}_j} \right) = 0, \quad j = 1, 2, \dots, m. \quad (5)$$

Из (4) видно, что

$$\left\{ \frac{\partial v_r}{\partial \bar{x}_s} \right\} = \{a_{rs}\} = A, \quad (6)$$

так что, введя матрицу весов, как диагональную матрицу

$$P_{nn} = \begin{Bmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & p_n \end{Bmatrix}, \quad (7)$$

мы сможем записать уравнения (5) в виде

$$A^* \cdot P \cdot V = 0. \quad (8)$$

У п р а ж н е н и е 1. Проверьте это. Покажите также, что если ввести m матриц

$$A_{n1}^1 = \begin{Bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{Bmatrix}, \quad A_{n1}^2 = \begin{Bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{Bmatrix}, \quad \dots, \quad A_{n1}^m = \begin{Bmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{Bmatrix}, \quad (9)$$

то m уравнений (8) могут быть записаны в виде

$$[pa_1v] = [pa_2v] = \dots = [pa_mv] = 0. \quad (10)$$

У п р а ж н е н и е 2. Покажите, что $[p\bar{v}v]$ можно записать также в виде

$$[p\bar{v}v] = V^* \cdot P \cdot V. \quad (11)$$

Подставляя V из (4) в (8), мы сможем исключить V и получить уравнения, содержащие только \bar{X} :

$$A^* \cdot P \cdot A \cdot \bar{X} = A^* \cdot P \cdot N. \quad (12)$$

Введем для сокращения матрицу

$$A_{mn}^* \cdot P_{nn} \cdot A_{nm} = B_{mm}, \quad (13)$$

которая симметрична, так как

$$B^* = (A^* \cdot P \cdot A)^* = A^* \cdot P^* \cdot A^{**} = A^* \cdot P \cdot A = B. \quad (14)$$

Таким образом, окончательно получаем

$$B \cdot \bar{X} = A^* \cdot P \cdot N. \quad (15)$$

Эти m уравнений относительно m неизвестных величин \bar{x} называются *нормальными уравнениями* системы.

Упражнение 3. Покажите, что с помощью обозначений (9) формулу (15) можно записать в виде

$$[p a_j a_1] \bar{x}_1 + [p a_j a_2] \bar{x}_2 + \dots + [p a_j a_m] \bar{x}_m = [p a_j n], \quad j = 1, 2, \dots, m, \quad (16)$$

что является обычной формой нормальных уравнений, введенной еще Гауссом. Иногда вместо $[puz]$ употребляют символ $[uz]$.

Из формулы (15) видно, что метод наименьших квадратов действительно дает единственное решение, получаемое разрешением уравнений (15) относительно \bar{X} , что можно осуществить умножением слева на B^{-1} :

$$\bar{X}_{m1} = C_{mn} \cdot N_{n1}, \quad (17)$$

$$C = B^{-1} \cdot A^* \cdot P. \quad (18)$$

Здесь мы предполагаем, что детерминант $|B|$ не равен нулю, так как при $|B| = 0$ из теории линейных уравнений следует, что между величинами x имеются некоторые соотношения, а это противоречило бы нашему предположению о том, что элементы суть независимые величины. Зная величины \bar{X} , определяемые формулой (17), можно, наконец, получить величины V по формуле (4) и величины \bar{L} по формуле (2):

$$V = (-E + A \cdot C) \cdot N, \quad (19)$$

$$\bar{L} = A^0 + A \cdot C \cdot N. \quad (20)$$

Упражнение 4. Покажите, что при $n = m$ эти формулы сводятся к следующим:

$$\bar{X} = A^{-1} \cdot N, \quad \bar{L} = L \quad \text{и} \quad V = 0. \quad (21)$$

§ 12.4. Остается проверить, что величины V , определяемые формулой (19), действительно обращают сумму $[p v v]$ в минимум. Пусть V' — какое-либо другое множество поправок, а X' — соответствующее множество элементов, т. е. согласно (12.3.4)

$$V' = -N + A \cdot X'. \quad (1)$$

Вычитая (12.3.4) из (1), получим

$$V' - V = A \cdot (X' - \bar{X}). \quad (2)$$

Используя (12.3.11), получим в качестве обобщения формулы (11.5.6),

поскольку $V' = (V' - V) + V$, следующее соотношение:

$$[pv'v'] = V'^* \cdot P \cdot V' = V^* \cdot P \cdot V + (V' - V)^* \cdot P \cdot (V' - V) + V^* P \cdot A \cdot (\bar{X}' - \bar{X}) + (\bar{X}' - \bar{X})^* \cdot A^* \cdot P \cdot V = [pvv] + [p(v' - v)^2], \quad (3)$$

причем третье и четвертое слагаемое в силу формулы (12.3.8) равны нулю. Но последнее слагаемое неотрицательно и обращается в нуль лишь при $V' = V$. Таким образом, $[pv'v']$ достигает минимума при $V' = V$.

У п р а ж н е н и е. Покажите, что

$$[p\varepsilon\varepsilon] = q^2 + (\bar{X} - \Xi)^* \cdot B \cdot (\bar{X} - \Xi), \quad (4)$$

где

$$q = V[\overline{pvv}], \quad \Xi_{m1} = \left\{ \begin{array}{c} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_m \end{array} \right\}. \quad (5)$$

§ 12.5. Чтобы найти наилучшую оценку s для параметра σ , надо решить уравнение максимального правдоподобия, в котором P определено формулой (12.1.5):

$$\frac{\partial}{\partial \sigma} \ln P = \frac{\partial}{\partial \sigma} \left(\text{const} - n \ln \left[\sigma - \frac{[p\varepsilon\varepsilon]}{2\sigma^2} \right] \right) = -\frac{n}{\sigma} + \frac{[p\varepsilon\varepsilon]}{\sigma^3} = 0,$$

откуда

$$\sigma \approx s_1 = \sqrt{\frac{[pvv]}{n}}. \quad (1)$$

По тем причинам, которые изложены в § 11.6—§11.7, здесь также удобно использовать другую оценку, s , имеющую такой же коэффициент изменчивости, как и s_1 , и поэтому также являющуюся асимптотически эффективной оценкой для σ :

$$\sigma \approx s = \sqrt{\frac{n}{n-m}} s_1 = \sqrt{\frac{[pvv]}{n-m}}, \quad (2)$$

где $n - m$ есть число степеней свободы, так как n величин v связаны m соотношениями (12.3.8).

*§ 12.6. Совместное распределение величин $\bar{x}_1, \dots, \bar{x}_m$ и $q = \sqrt{[pvv]}$ получается как обобщение формулы (7.7.11):

$$\begin{aligned} d\Phi &= \varphi'_m(\bar{x}_1, \dots, \bar{x}_m, q) d\bar{x}_1 \dots d\bar{x}_m dq = \\ &= \left\{ \frac{V|B|}{(V2\pi\sigma)^m} \exp \left[-\frac{1}{2\sigma^2} (\bar{X} - \Xi)^* \cdot B \cdot (\bar{X} - \Xi) \right] \cdot d\bar{x}_1 \dots d\bar{x}_m \right\} \times \\ &\times \left\{ \frac{1}{\left(\frac{f-2}{2}\right)! 2^{\frac{f-2}{2}}} \left(\frac{q}{\sigma}\right)^{f-1} \exp \left[-\frac{q^2}{2\sigma^2} \right] \frac{dq}{\sigma} \right\}, \quad f = n - m. \quad (1) \end{aligned}$$

У п р а ж н е н и е. Проверьте это. (Доказательство аналогично доказательству формулы (7.7.11). Введите в (12.1.5) вместо l_1, \dots, l_n новые n величин $\bar{x}_1, \dots, \bar{x}_m, q$ и u_i , полагая $l_i = \bar{l}_i - \frac{q}{V p_i} u_i, i = 1, 2, \dots, n$; используйте формулу (12.4.4).)

Формула (1) показывает, что $\bar{x}_1, \dots, \bar{x}_m$ независимы от q , что $\bar{x}_1, \dots, \bar{x}_m$ и q образуют множество совместно достаточных оценок для ξ_1, \dots, ξ_m и σ (см. § 10.11), что $\bar{x}_1, \dots, \bar{x}_m$ имеют нормальное распределение со средними значениями $\bar{\Xi}$ и матрицей вторых моментов

$$M(\bar{x}) = M\{x_r - \xi_r, (x_s - \xi_s)\} = \sigma^2 B^{-1}, \quad (2)$$

(см. § 7.6), т. е. что необходимым и достаточным условием взаимной независимости величин $\bar{x}_1, \dots, \bar{x}_m$ является диагональный вид матрицы B . Наконец, формула (1) показывает, что q имеет q -распределение (7.7.12) с числом степеней свободы $f = n - m$, что согласуется с тем обстоятельством, что n величин v связаны m соотношениями (12.3.8). Как показано в § 7.7, поскольку $s = \frac{q}{\sqrt{f}}$, то

$$M\{s^2\} = \sigma^2, \quad \sigma^2\{s^2\} = \frac{2\sigma^4}{n-m} \quad (3)$$

(см. задачу 69) и

$$M\{s\} \sim \sqrt{1 - \frac{1}{2(n-m)}} \sigma, \quad \sigma^2\{s\} \sim \frac{\sigma^2}{2(n-m)} \quad (4)$$

при $f = n - m \gg 1$

§ 12.7. Оценки $\bar{x}_1, \dots, \bar{x}_m$ и s суть, конечно, случайные величины, подверженные статистическим флуктуациям, так как новые n измерений l'_1, \dots, l'_n приведут к несколько иным оценкам $\bar{x}'_1, \dots, \bar{x}'_m, s'$. Поэтому следует также оценить величину этих статистических флуктуаций, для чего надо образовать оценки для дисперсий величин $\bar{x}_1, \dots, \bar{x}_m$ и s . Эти оценки можно получить, пользуясь матрицей вторых моментов величин $\bar{x}_1, \dots, \bar{x}_m$, приведенной в формуле (12.6.2). Таким образом,

$$\sigma\{\bar{x}_j\} = \sqrt{(B^{-1})_{jj}} \sigma \approx \sqrt{(B^{-1})_{jj}} \sqrt{\frac{[p_{vv}]}{n-m}}, \quad j = 1, 2, \dots, m. \quad (1)$$

Подчеркнем, что, вообще говоря, величины \bar{x} коррелированы, так как B^{-1} обычно не является диагональной матрицей. Однако если B^{-1} есть диагональная матрица, то $\rho\{\bar{x}_i, \bar{x}_j\} = 0$ при $i \neq j$, так что величины \bar{x} некоррелированы, и B также есть диагональная матрица. Как показывает формула (12.6.1), при этом величины \bar{x} даже независимы.

У п р а ж н е н и е 1. Покажите, что (12.6.2) следует также из (6.4.23), так как согласно (12.1.5) имеем

$$M^{(L)} = \sigma^2 P^{-1}, \quad (2)$$

а \bar{X} есть линейная функция от L , определяемая формулой (12.3.17).

У п р а ж н е н и е 2. Покажите, что (12.6.2) следует также из (10.10.15).

Пример. В теории уравнивания принято называть множество линейных функций y_1, \dots, y_k от результатов прямых измерений l_1, \dots, l_n *свободными функциями*, если величины y попарно некоррелированы, т. е. $\rho\{y_i, y_j\} = 0$ при $i \neq j$. По формулам (6.4.23) и (2) имеем

$$M_{kk}^{(Y)} = \sigma^2 F \cdot P^{-1} \cdot F^*, \quad (3)$$

откуда следует, что для того, чтобы функции y были свободными функциями, необходимо и достаточно, чтобы матрица $F \cdot P^{-1} \cdot F^*$ была диагональной. Однако, поскольку предполагается, что величины l имеют нормальное распределение, величины y также имеют нормальное распределение, так что формула (3) показывает, что в случае нормально распределенных величин три понятия — взаимно независимые, попарные некоррелированные и свободные функции — эквивалентны (см. пример 2 § 7.6).

§ 12.8. Если имеется множество функций от \bar{X} , как, например, величины \bar{L} , определенные формулой (12.3.2), и желательно вычислить дисперсии этих функций, пользуясь дисперсиями величин \bar{X} , определяемыми формулой (12.6.2), то следует помнить, что простой закон сложения дисперсий (6.4.4), вообще говоря, неприменим, и должен быть использован общий закон сложения дисперсий (6.4.3), поскольку величины \bar{X} обычно не некоррелированы. Однако, используя матричную символику, необходимое вычисление можно проделать автоматически. Так, если]

$$Z_{k1} = G_{k1}^0 + G_{km} \cdot \bar{X}_{m1} \quad (1)$$

есть множество новых k случайных величин, также нормально распределенных, то из (6.4.23) и (12.6.2) немедленно получаем

$$M_{kk}^{(Z)} = \sigma^2 G \cdot B^{-1} \cdot G^*. \quad (2)$$

У п р а ж н е н и е. Покажите, что будет получен тот же самый результат, если сначала с помощью формулы (12.3.17) выразить Z как функцию от L , а затем использовать формулы (6.4.23) и (12.7.2).

В частности, поскольку из формулы (12.3.2) следует, что $G(\bar{L}) = A$, по формуле (2) получаем

$$M_{nn}^{(L)} = \sigma^2 A \cdot B^{-1} \cdot A^*, \quad (3)$$

причем эта матрица, вообще говоря, не является диагональной. Таким образом, оценки \bar{L} обычно не некоррелированы (т. е. не независимы). По формулам (3) и (12.5.2) получаем

$$\sigma \{ \bar{l}_i \} = \sqrt{(A \cdot B^{-1} \cdot A^*)_{ii}} \sigma \approx \sqrt{(A \cdot B^{-1} \cdot A^*)_{ii}} \sqrt{\frac{[\rho\sigma\sigma]}{n-m}}. \quad (4)$$

Как доказывается в задаче 68, для контроля можно использовать соотношение

$$\sum_{i=1}^n p_i \sigma^2 \{ \bar{l}_i \} = m \sigma^2. \quad (5)$$

Пример. Легко впасть в ошибку, полагая $\sigma \{ \bar{l}_i \} = \sigma \{ l_i \} = \frac{\sigma}{\sqrt{p_i}}$.

Ошибочность этого можно усмотреть из того, что согласно (12.1.4) $\sum_{i=1}^n p_i \sigma^2 \{ l_i \} = n \sigma^2$, в отличие от формулы (5).

§ 12.9. Для стандартного отклонения величины s по формуле (12.6.4) имеем

$$\sigma \{ s \} \sim \frac{\sigma}{\sqrt{2(n-m)}} \approx \frac{s}{\sqrt{2(n-m)}}, \quad (1)$$

если $f = n - m$ не слишком мало; в противном случае следует использовать точную формулу (7.7.19).

У п р а ж н е н и е 1. Покажите, что согласно формуле (10.10.14) дисперсия оценки максимального правдоподобия s_1 , определяемой формулой (12.5.1), асимптотически равна

$$\sigma^2 \{ s_1 \} = \frac{\sigma^2}{2n}, \quad (2)$$

откуда также следует формула (1).

Окончательный результат n измерений l_1, \dots, l_n выглядит теперь следующим образом:

Оценка для истинного значения:

$$\bar{x}_j \approx \bar{x}_j \text{ со стандартным отклонением } \sigma \{ \bar{x}_j \} \approx \sqrt{(B^{-1})_{jj}} s, \quad j=1, 2, \dots, m. \quad (3)$$

Оценка для истинного значения:

$$\bar{l}_i \approx \bar{l}_i \text{ со стандартным отклонением } \sigma \{ \bar{l}_i \} \approx \sqrt{(A \cdot B^{-1} \cdot A^*)_{jj}} s, \quad i=1, 2, \dots, n \quad (4)$$

Оценка для стандартного отклонения:

$$s \approx s \text{ со стандартным отклонением } \sigma \{ s \} \approx \frac{s}{\sqrt{2(n-m)}}. \quad (5)$$

(Стандартные отклонения оценок часто называются *стандартами* или *средними ошибками*.)

Упражнение 2. Покажите, что

$$[p\sigma\sigma] = N^* \cdot P \cdot N - N^* \cdot P \cdot A \cdot \bar{X} = [pnn] - [pna_1] \bar{x}_1 - \dots - [pna_m] \bar{x}_m, \quad (6)$$

$$[p\sigma\sigma] = N^* \cdot P \cdot N - \bar{X}^* \cdot B \cdot \bar{X}, \quad (7)$$

$$[p\sigma\sigma] = [p\ell\ell] - [p\bar{\ell}\bar{\ell}] + 2A^0 \cdot P \cdot V. \quad (8)$$

Эти выражения могут быть использованы или для вычисления $[p\sigma\sigma]$ или для контроля за таким вычислением.

* Упражнение 3. Покажите, что

$$\frac{\bar{x}_j - \xi_j}{\sqrt{(B^{-1})_{jj}} s}, \quad j=1, 2, \dots, m, \quad \text{и} \quad \frac{\bar{\ell}_i - \lambda_i}{\sqrt{(A \cdot B^{-1} A^*)_{ii}} s}, \quad i=1, 2, \dots, m, \quad (9)$$

имеют распределение t с $f = n - m$ степенями свободы (см. § 7.9). Найдите выражения доверительных интервалов для параметров $\xi_1, \dots, \xi_m, \lambda_1, \dots, \lambda_n$ (см. § 11.9).

Пример 1. Пусть l_1, \dots, l_n суть n , не обязательно равнозначных, измерений некоторой величины x , так что число элементов m равно 1, и фундаментальные уравнения имеют вид

$$\lambda_1 = \lambda_2 = \dots = \lambda_n = \xi.$$

В этом случае имеем

$$A^0 = 0, \quad A_{nm} = A_{n1} = \begin{Bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{Bmatrix},$$

$$B_{nm} = B_{11} = \{11 \dots 1\} \begin{Bmatrix} p_1 0 & \dots & 0 \\ 0 & p_2 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & p_n \end{Bmatrix} \begin{Bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{Bmatrix} = [p],$$

$$B^{-1} = \frac{1}{[p]},$$

$$C_{mn} = C_{1n} = \frac{1}{[p]} \{11 \dots 1\} \begin{Bmatrix} p_1 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & p_n \end{Bmatrix} = \left\{ \frac{p_1}{[p]}, \frac{p_2}{[p]}, \dots, \frac{p_n}{[p]} \right\},$$

так что получаем

$$\xi \approx \bar{x} = \frac{[p\ell]}{[p]}, \quad (10)$$

что является *взвешенным средним арифметическим*. Далее,

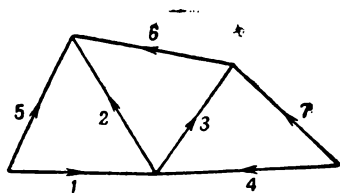
$$\sigma \approx s = \sqrt{\frac{[p\nu\nu]}{n-1}}, \quad (11)$$

$$\sigma\{\bar{x}\} \approx \sqrt{\frac{[p\nu\nu]}{[p](n-1)}}, \quad (12)$$

$$\sigma\{s\} \approx \frac{s}{\sqrt{2(n-1)}}. \quad (13)$$

Заметим, что формула (12) совпадает с результатом упражнения 2 § 6.4 и, более того, показывает, что *вес среднего арифметического также и в общем случае равен сумме весов отдельных измерений* (см. § 12.1). При $p_1 = p_2 = \dots = p_n = 1$ формулы (10) — (13) сводятся к соответствующим формулам (11.7.3) и (11.7.4).

Пример 2. На земельном участке были измерены разности высот пяти точек, показанных на фиг. 21 (стрелки указывают направления убывания высот). Измеренные значения (в метрах) равны



Фиг. 21.

$$L = \begin{Bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \\ l_6 \\ l_7 \end{Bmatrix} = \begin{Bmatrix} 20,21 \\ 40,07 \\ 34,17 \\ 35,84 \\ 60,40 \\ 5,87 \\ 69,99 \end{Bmatrix}.$$

Как показывает детальное исследование метода измерений, используемого в таких задачах, веса естественно выбирать обратно пропорциональными соответствующим расстояниям. Считая эти расстояния пропорциональными числам 1000, 1110, 910, 1250, 1000, 1110 и 1000, мы сможем принять в качестве матрицы весов

$$P_{77} = \begin{Bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1,1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{Bmatrix}.$$

Выбирая за элементы первую, вторую, третью и четвертую разности высот, мы получим фундаментальные уравнения в виде

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix}.$$

Далее находим

$$B = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 2,8 & -0,9 & 0 \\ 0 & -0,9 & 3 & 1 \\ 0 & 0 & 1 & 1,8 \end{pmatrix}, \quad |B| = 17,324,$$

$$B^{-1} = \frac{1}{17,324} \begin{pmatrix} 10,862 & -4,4 & -1,62 & 0,9 \\ -4,4 & 8,8 & 3,24 & -1,8 \\ -1,62 & 3,24 & 8,28 & -4,6 \\ 0,9 & -1,8 & -4,6 & 12,18 \end{pmatrix},$$

$$A^* \cdot P \cdot N = \begin{pmatrix} 80,610 \\ 101,746 \\ 102,294 \\ 98,662 \end{pmatrix},$$

и получаем следующие результаты:

$$\Xi \approx \bar{X} = \begin{pmatrix} 20,260 \\ 40,090 \\ 34,185 \\ 35,821 \end{pmatrix} \quad \begin{array}{c} \text{со стандартными} \\ \text{отклонениями} \end{array} \quad \begin{pmatrix} 0,039 \\ 0,035 \\ 0,034 \\ 0,041 \end{pmatrix}$$

$$\Lambda \approx \bar{L} = \begin{pmatrix} 20,260 \\ 40,090 \\ 34,185 \\ 35,821 \\ 60,350 \\ 5,905 \\ 70,006 \end{pmatrix} \quad \begin{array}{c} \text{со стандартными} \\ \text{отклонениями} \end{array} \quad \begin{pmatrix} 0,039 \\ 0,035 \\ 0,034 \\ 0,041 \\ 0,039 \\ 0,039 \\ 0,040 \end{pmatrix},$$

и $\sigma \approx s = 0,0493$ со стандартным отклонением 0,0201. (Проверьте численные вычисления.)

* Пример 3. В примере 2 мы имели $f = n - m = 7 - 4 = 3$ и $t(5\%/0,3) = 3,182$ согласно табл. III. Таким образом, пятипроцентные

доверительные пределы получаются умножением полученных стандартных отклонений на 3,182.

У п р а ж н е н и е 4. При практических расчетах в примере 2 лучше было бы выбрать в качестве элементов истинные поправки, а не истинные значения разностей высот с номерами 1—4, чтобы вычисления можно было производить лишь с малыми числами. Прделайте все соответствующие вычисления.

* **12.10.** Если желательно проверить, не содержит ли измерение грубой ошибки, то можно вычислить соответствующее относительное отклонение и с помощью r -распределения проверить, лежит ли оно в разумных пределах (см. § 7.10).

У п р а ж н е н и е 1. Покажите, что матрица вторых моментов для V определяется формулой

$$M_{nn}^{(V)} = \sigma^2 (A \cdot C - E) \cdot P^{-1} \cdot (A \cdot C - E)^* = \sigma^2 T, \quad (1)$$

где T есть сокращенное обозначение для сложной матрицы в формуле (1).

Относительные отклонения определяются формулой

$$r_i = \frac{\bar{l}_i - l_i}{\sqrt{t_{ii}} s} = \sqrt{\frac{n-m}{t_{ii}}} \frac{v_i}{\sqrt{[p\sigma\sigma]}}, \quad i = 1, 2, \dots, n, \quad (2)$$

и можно показать¹⁾, что частное распределение каждого r_i есть r -распределение (7.10.2) с $f = n - m - 1$, в согласии с тем обстоятельством, что величины r связаны $m + 1$ соотношениями.

У п р а ж н е н и е 2. Покажите, что формула (12.3.8) представляет m соотношений между величинами r . Далее, покажите, что

$$[ptr^2] = \sum_{i=1}^n p_i t_{ii} r_i^2 = n - m. \quad (3)$$

Более того, относительные отклонения можно здесь использовать также для проверки гипотезы о том, что наблюдаемые данные имеют нормальное распределение, так как часто имеется в распоряжении много малых выборок с различными истинными значениями, весами и стандартными отклонениями, но с одним и тем же числом степеней свободы f (см. § 11.21).

Уравнивание с помощью коррелят

§ 12.11. Если ввести матрицы

$$L_{n1} = \begin{Bmatrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{Bmatrix}, \quad V_{n1} = \begin{Bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{Bmatrix}, \quad A_{r1}^0 = \begin{Bmatrix} a_{10} \\ a_{20} \\ \vdots \\ a_{r0} \end{Bmatrix}, \quad A_{rn} = \begin{Bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \dots & a_{rn} \end{Bmatrix}, \quad (1)$$

¹⁾ См. Arley, *Danske Vid. Selsk. Mat.-fys. Medd.*, т. XVIII, № 3, 1940, гл. II.

то условные уравнения (12.2.6) можно записать в виде

$$A^0 + A \cdot (L + V) = 0. \quad (2)$$

Для сокращения удобно обозначить

$$A^0 + A \cdot L = N_{r1}. \quad (3)$$

Тогда уравнения (2) примут вид, аналогичный уравнениям (12.3.4),

$$N + A \cdot V = 0. \quad (4)$$

Здесь имеется r уравнений с $n > r$ неизвестными величинами v . Поскольку эти величины не независимы, а связаны r соотношениями (4), условие, необходимое для того, чтобы $[pvv]$ было минимально, в этом случае заключается не в том, что частные производные суммы $[pvv]$ по всем v равны нулю (что привело бы к условию $V = 0$). Необходимое условие заключается в том, что полный дифференциал суммы $[pvv]$ должен равняться нулю, т. е., если использовать формулу (12.3.11), в том, что

$$d[pvv] = [2pvdv] = 2V^* \cdot P \cdot dV = 0, \quad (5)$$

где dV обозначает матрицу

$$dV_{n1} = \begin{Bmatrix} dv_1 \\ dv_2 \\ \vdots \\ dv_n \end{Bmatrix}. \quad (6)$$

Дифференцируя (4), получим

$$A \cdot dV = 0. \quad (7)$$

Исключим теперь из уравнений (5) и (7) величины dV с помощью метода множителей Лагранжа, называемых здесь *коррелятами* k_1, k_2, \dots, k_r и образующих матрицу коррелят

$$K_{r1} = \begin{Bmatrix} k_1 \\ k_2 \\ \vdots \\ k_r \end{Bmatrix}. \quad (8)$$

При произвольной матрице K из формул (5) и (7) получаем

$$V^* \cdot P \cdot dV - K^* \cdot A \cdot dV = (V^* \cdot P - K^* \cdot A) \cdot dV = 0. \quad (9)$$

Метод, предложенный Лагранжем, заключается теперь в подборе таких значений доселе произвольных коррелят, чтобы

$$V^* \cdot P - K^* \cdot A = 0, \quad (10)$$

откуда

$$V = P^{-1} \cdot A^* \cdot K. \quad (11)$$

n уравнений (11) с $n - r$ неизвестными V и K называются *уравнениями коррелят*. Добавив r уравнений (4), мы имеем столько же уравнений, сколько неизвестных.

Пример. n уравнений (10) можно получить следующим образом. Поскольку между n величинами v имеется r соотношений, можно выбрать $n - r$ из этих величин и выразить через них остальные r величин. Вводя полученные выражения в сумму $[pvv]$ и приравнявая нулю частные производные от полученного выражения по $n - r$ выбранным свободным величинам v , мы получим $n - r$ уравнений, которые вместе с r уравнениями (4) составят систему n уравнений относительно неизвестных величин v . Этих вычислений в методе Лагранжа удаётся избежать. Пусть в качестве свободных величин v выбраны величины v_1, \dots, v_{n-r} . Тогда определим сначала r коррелят так, чтобы коэффициенты при последних r дифференциалах в (9) равнялись нулю; затем, поскольку дифференциалы dv_1, \dots, dv_{n-r} являются теперь независимыми, соотношение (9) будет означать, что эти значения коррелят должны быть таковы, что и при первых $n - r$ дифференциалах коэффициенты равны нулю. Но это обстоятельство и выражается n уравнениями (10).

Вводя формулы (11) в (4), мы исключим V и получим r уравнений

$$N + A \cdot P^{-1} \cdot A^* \cdot K = 0. \quad (12)$$

Аналогично (12.3.13), положим для сокращения

$$A_{rn} \cdot P_{nn}^{-1} \cdot A_{nr}^* = B_{rr}, \quad (13)$$

причем эта матрица будет *симметричной*, так как

$$B^* = (A \cdot P^{-1} \cdot A^*)^* = A^{**} \cdot P^{-1*} \cdot A^* = A \cdot P^{-1} \cdot A^* = B. \quad (14)$$

Таким образом, окончательно получаем

$$B \cdot K = -N. \quad (15)$$

Эти r уравнений относительно r неизвестных коррелят k , аналогичные уравнениям (12.3.15), называются *нормальными уравнениями системы*.

У п р а ж н е н и е 1. Введя r матриц, аналогичных (12.3.9),

$$A_{n1}^1 = \begin{Bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1n} \end{Bmatrix}, \quad A_{n1}^2 = \begin{Bmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2n} \end{Bmatrix}, \dots, \quad A_{n1}^r = \begin{Bmatrix} a_{r1} \\ a_{r2} \\ \vdots \\ a_{rn} \end{Bmatrix}, \quad (16)$$

покажите, что уравнения (15) могут быть записаны в виде

$$\left[\frac{a_j a_1}{p} \right] k_1 + \left[\frac{a_j a_2}{p} \right] k_2 + \dots + \left[\frac{a_j a_r}{p} \right] k_r + n_j = 0, \quad j = 1, 2, \dots, r, \quad (17)$$

что является обычной формой нормальных уравнений, введенной еще Гауссом.

Иногда вместо символа $\left[\frac{yz}{p}\right]$, введенного здесь, употребляют символ (yz) , и вместо символа n_j — символ $(a_j n)$, так что нормальные уравнения принимают формально такой же вид, что и при уравнивании с помощью элементов (12.3.16).

Из формулы (15) видно, что метод наименьших квадратов в комбинации с методом множителей Лагранжа приводит к единственному решению, получаемому из (15) умножением слева на B^{-1} 1).

$$K = -B^{-1} \cdot N. \quad (18)$$

Из (11) окончательно получаем V :

$$V = -P^{-1} \cdot A^* \cdot B^{-1} \cdot N, \quad (19)$$

так что

$$\bar{L} = L + V = L - P^{-1} \cdot A^* \cdot B^{-1} \cdot N. \quad (20)$$

Упражнение 2. Покажите, что формулу (20) можно записать также в виде

$$\bar{L} = -P^{-1} \cdot A^* \cdot B^{-1} \cdot A^0 + (E - P^{-1} \cdot A^* \cdot B^{-1} \cdot A) \cdot L. \quad (21)$$

§ 12.12. Как было указано в § 12.2, уравнивание с помощью коррелят приводит к тем же самым результатам, что и уравнивание с помощью элементов. Следовательно, можно не доказывать заново, что значения v , определяемые формулой (12.11.19), действительно обращают сумму $[pvv]$ в минимум. Поскольку число элементов равно $m = n - r$, согласно предыдущим формулам имеем также

$$\sigma \approx s = \sqrt{\frac{[pvv]}{r}} \quad (1)$$

с асимптотическим стандартным отклонением

$$\sigma \{s\} \approx \frac{s}{\sqrt{2r}}. \quad (2)$$

Упражнение 1. Покажите, что

$$[pvv] = -N^* \cdot K, \quad (3)$$

$$[pvv] = K^* \cdot B \cdot K, \quad (4)$$

$$[pvv] = N^* \cdot B^{-1} \cdot N, \quad (5)$$

причем эти выражения можно использовать или для вычисления $[pvv]$, или для контроля такого вычисления.

Упражнение 2. Покажите, что матрицы вторых моментов для K и \bar{L} определяются формулами

$$M_{rr}^{(K)} = \sigma^2 B^{-1}. \quad (6)$$

$$M_{nn}^{(\bar{L})} = \sigma^2 (P^{-1} - P^{-1} \cdot A^* \cdot B^{-1} \cdot A \cdot P^{-1}). \quad (7)$$

1) Здесь детерминант $|B| \neq 0$ по тем же причинам, что и при уравнивании с помощью элементов (см. стр. 198).

Из (7) получаем

$$\begin{aligned} \sigma \{\bar{L}_i\} &= \sqrt{(P^{-1} - P^{-1} \cdot A^* \cdot B^{-1} \cdot A \cdot P^{-1})_{ii}} \sigma \approx \\ &\approx \sqrt{(P^{-1} - P^{-1} \cdot A^* \cdot B^{-1} \cdot A \cdot P^{-1})_{ii}} \sqrt{\frac{[\rho v v]}{r}}. \end{aligned} \quad (8)$$

Для контроля можно использовать формулу (12.8.5), которая благодаря тому, что $m = n - r$, принимает теперь вид

$$\sum_{i=1}^n \rho_i \sigma^2 \{\bar{L}_i\} = (n - r) \sigma^2. \quad (9)$$

Пример. В задаче, рассмотренной в примере 2 § 12.9, имелось три соотношения между истинными значениями семи результатов измерений, именно:

$$\begin{aligned} \lambda_1 + \lambda_2 - \lambda_5 &= 0, \\ \lambda_2 - \lambda_3 - \lambda_6 &= 0, \\ \lambda_3 + \lambda_4 - \lambda_7 &= 0, \end{aligned}$$

т. е.

$$A^0 = 0 \text{ и } A_{37} = \begin{pmatrix} 1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 \end{pmatrix},$$

откуда

$$N_{31} = A \cdot L = \begin{pmatrix} -0,12 \\ 0,03 \\ 0,02 \end{pmatrix}$$

и

$$\begin{aligned} B_{33} = A \cdot P^{-1} \cdot A^* = A \cdot \begin{pmatrix} 1 & 0 & 0 \\ 1,1111 & 1,1111 & 0 \\ 0 & -0,90909 & 0,90909 \\ 0 & 0 & 1,25 \\ -1 & 0 & 0 \\ 0 & -1,1111 & 0 \\ 0 & 0 & -1 \end{pmatrix} = \\ = \begin{pmatrix} 3,1111 & 1,1111 & 0 \\ 1,1111 & 3,1313 & -0,90909 \\ 0 & -0,90909 & 3,1591 \end{pmatrix} \end{aligned}$$

при значениях L и P , приведенных в примере 2 § 12.9. Отсюда получаем

$$B^{-1} = \begin{pmatrix} 0,37302 & -0,14442 & -0,04156 \\ -0,14442 & 0,40439 & 0,11637 \\ -0,04156 & 0,11637 & 0,35003 \end{pmatrix}$$

и

$$K = -B^{-1} \cdot N = \begin{Bmatrix} 0,04993 \\ -0,03179 \\ -0,01548 \end{Bmatrix}, \text{ т. е. } V = P^{-1} \cdot A^* \cdot K = \begin{Bmatrix} 0,050 \\ 0,020 \\ 0,015 \\ -0,019 \\ -0,050 \\ 0,035 \\ 0,016 \end{Bmatrix}.$$

Отсюда окончательно получаем

$$\bar{L} = \begin{Bmatrix} 20,260 \\ 40,090 \\ 34,185 \\ 35,821 \\ 60,350 \\ 5,905 \\ 70,006 \end{Bmatrix},$$

что полностью согласуется с результатами, полученными в примере 2 § 12.9 (проверьте все численные вычисления).

§ 12.13. Заметим, что развита здесь теория уравнивания пригодна лишь при условии, что фундаментальные уравнения (12.2.1) или (12.2.2) действительно линейны, т. е. имеют вид (12.2.5) или (12.2.6). Однако на практике часто встречаются задачи, в которых исходные фундаментальные уравнения нелинейны. Тогда первой задачей является преобразование фундаментальных уравнений к линейному виду. В большинстве случаев условия, указанные в § 6.5, удовлетворяются, т. е. ошибки столь малы, что все наши функции можно с хорошим приближением заменять их соответствующими касательными плоскостями, или, иными словами, в разложениях этих функций в ряды Тейлора можно пренебречь всеми членами, кроме линейных.

При уравнивании с помощью элементов в качестве последних выберем истинные поправки, а не истинные значения (см. упражнение 4 § 12.9). Для этой цели определим сначала приближенные значения наилучших оценок $x_1^0, x_2^0, \dots, x_m^0$ для элементов. Затем получим

$$\bar{x}_j = x_j^0 + \Delta x_j, \quad j = 1, 2, \dots, m \quad (1)$$

и предположим, что все Δx настолько малы, что их вторыми и более высокими степенями можно пренебречь. Такое условие, конечно, будет выполнено при подходящем выборе значений x_1^0, \dots, x_m^0 . Разлагая функции в условных уравнениях в ряды Тейлора и сохраняя лишь

линейные члены этих рядов, мы получим линеаризованные уравнения

$$l_i + v_i = f_i(x_1^0 + \Delta x_1, \dots, x_m^0 + \Delta x_m) = f_i(x_1^0, \dots, x_m^0) + \\ + \left(\frac{\partial f_i}{\partial x_1}\right)_0 \Delta x_1 + \dots + \left(\frac{\partial f_i}{\partial x_m}\right)_0 \Delta x_m + \dots, \quad i = 1, 2, \dots, n, \quad (2)$$

в которых частные производные берутся при $\Delta x_1 = \dots = \Delta x_n = 0$. Вводя обозначения

$$f_i(x_1^0, \dots, x_m^0) = a_{i0}, \quad i = 1, 2, \dots, n \quad (3)$$

и

$$\left(\frac{\partial f_i}{\partial x_j}\right)_0 = a_{ij} \quad \begin{matrix} i = 1, 2, \dots, n, \\ j = 1, 2, \dots, m, \end{matrix} \quad (4)$$

мы получим уравнения (12.3.2), в которых вместо \bar{x}_j стоят Δx_j . Из теории следует, что величины Δx имеют нормальное распределение, но, как было указано в § 7.5, при наших условиях величины \bar{x} также имеют нормальное распределение.

При уравнивании с помощью коррелят непосредственно разложим все функции по степеням величин v и получим, пренебрегая высшими степенями величин v , следующие уравнения:

$$f_j(l_1, \dots, l_n) + \left(\frac{\partial f_j}{\partial l_1}\right)_0 v_1 + \dots + \left(\frac{\partial f_j}{\partial l_n}\right)_0 v_n + \dots = 0, \\ j = 1, 2, \dots, m. \quad (5)$$

Чтобы величины v были столь малы, что их высшими степенями можно смело пренебрегать, надо предположить, что измеренные величины очень близки к наилучшим оценкам их истинных значений, получаемым при последующих вычислениях. Для точных измерений это условие всегда выполняется, но при грубых измерениях вполне может случиться, что полученные величины v этому условию не удовлетворяют. В таких случаях следует повторить вычисления, разлагая функции в ряды не по измеренным значениям, а по другим более подходящим значениям.

Вводя в уравнениях (5) обозначения

$$f_j(l_1, \dots, l_n) = n_j, \quad j = 1, 2, \dots, m \quad (6)$$

и

$$\left(\frac{\partial f_j}{\partial l_i}\right)_0 = a_{ji}, \quad \begin{matrix} i = 1, 2, \dots, n, \\ j = 1, 2, \dots, m, \end{matrix} \quad (7)$$

мы получим уравнения (12.11.4).

Пример¹⁾. В треугольнике ABC были измерены стороны a и b и углы A и B , причем получены значения $l_1 = a = 52,3$ см, $l_2 = b = 33,4$ см, $l_3 = A = 62^\circ, 2$, $l_4 = B = 34^\circ, 7$. В этом случае между

¹⁾ Дальнейшие примеры нелинейных задач читатель найдет в учебниках по теории уравнивания, указанных в списке литературы.

четырьмя измерениями ($n=4$) имеется одно соотношение ($m=1$)

$$f(l_1 + v_1, \dots, l_4 + v_4) = (l_1 + v_1) \sin(l_4 + v_4) - (l_2 + v_2) \sin(l_3 + v_3) = 0.$$

(Подставляя численные измеренные значения, можно видеть, что $a \sin B - b \sin A = 0,22 \neq 0$, так что условие уравнение не удовлетворяется). Поскольку

$$\frac{\partial f}{\partial l_1} = \sin l_4, \quad \frac{\partial f}{\partial l_2} = -\sin l_3, \quad \frac{\partial f}{\partial l_3} = -l_2 \frac{\pi}{180} \cos l_3, \quad \frac{\partial f}{\partial l_4} = l_1 \frac{\pi}{180} \cos l_4,$$

уравнение (5) принимает вид

$$l_1 \sin l_4 - l_2 \sin l_3 + \sin l_4 \cdot v_1 - \sin l_3 \cdot v_2 - \frac{\pi}{180} l_2 \cos l_3 \cdot v_3 + \frac{\pi}{180} l_1 \cos l_4 \cdot v_4 = 0$$

или, после подстановки численных значений,

$$0,22 + 0,57v_1 - 0,88v_2 - 0,27v_3 + 0,75v_4 = 0,$$

$$\text{т. е. } A_{11}^0 = N = 0,22, \quad A_{14} = \{0,57 \quad -0,88 \quad -0,27 \quad 0,75\}.$$

Предполагая, что точность измерения углов вдвое больше, чем точность измерения сторон, получим

$$P_{44} = \begin{pmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Остальные вычисления дают следующие результаты:

$$P^{-1} \cdot A^* = \begin{pmatrix} 2,28 \\ -3,52 \\ -0,27 \\ 0,75 \end{pmatrix}, \quad B_{11} = A \cdot P^{-1} A^* = 5,03, \quad B^{-1} = \frac{1}{5,03},$$

$$K_{11} = -0,0437,$$

$$V = \begin{pmatrix} -0,10 \\ 0,15 \\ 0,01 \\ -0,03 \end{pmatrix}, \quad \text{т. е. } \begin{pmatrix} \bar{a} \\ \bar{b} \\ \bar{A} \\ \bar{B} \end{pmatrix} = \begin{pmatrix} 52,20 \text{ см} \\ 33,55 \text{ см} \\ 62^\circ,21 \\ 34^\circ,67 \end{pmatrix}.$$

Для контроля вычисляем $52,20 \sin 34^\circ,67 - 33,55 \sin 62^\circ,21 = 0,02 \sim 0$, что лежит в пределах точности вычислений. Наконец, вычисляя стандартные отклонения, получим в результате

$$\lambda \approx \bar{L} = \left\{ \begin{array}{l} 52,20 \text{ см} \\ 33,55 \text{ см} \\ 62^\circ,21 \\ 34^\circ,67 \end{array} \right\} \text{ со стандартными отклонениями } \left\{ \begin{array}{l} 0,16 \text{ см} \\ 0,12 \text{ см} \\ 0^\circ,10 \\ 0^\circ,09 \end{array} \right\}$$

$\sigma \approx s = 0,10$ со стандартным отклонением 0,07. (Проверьте все численные вычисления.)

§ 12.14. Завершим изложение теории уравнивания замечанием о том, когда удобнее использовать уравнивание с помощью элементов и когда — уравнивание с помощью коррелят. Поскольку основной труд в решении задачи уравнивания состоит в численных вычислениях для решения «нормальных уравнений»¹⁾, предпочтительнее использовать, как правило, тот вид уравнивания, который приводит к наименьшему числу уравнений. Если число соотношений между истинными значениями наблюдаемых величин равно r , то уравнивание с помощью элементов требует решения $m = n - r$ уравнений, где m — число элементов, а уравнивание с помощью коррелят требует решения r уравнений, причем r — число коррелят. За исключением специальных случаев, уравнивание с помощью элементов предпочтительнее, если $n - r < r$, т. е. если $r > \frac{n}{2}$, а уравнивание с помощью коррелят предпочтительнее, если $r < \frac{n}{2}$.

§ 12.15. Анализ регрессии. Наиболее важной проблемой, встречающейся во многих областях науки, является нахождение связи между двумя или большим количеством величин с помощью статистических методов. Эту проблему можно трактовать различными способами, в зависимости от природы задачи и целей анализа. Одним из основных способов является анализ регрессии, являющийся частным приложением теории уравнивания.

Рассмотрим здесь лишь случай двух величин x и y , из которых x рассматривается как независимая переменная, а y — как зависимая переменная. x может быть как статистической, так и не статистической величиной (последнее, как правило, означает лишь то, что дисперсия величины x в рассматриваемой проблеме пренебрежимо мала). Как уже было указано, последний случай есть лишь частный случай первого, в котором x имеет несобственное распределение (см. пример 1 § 4.3). Предположим, что условное распределение величины y при каждом значении x нормально (или что непосредственно заданная величина y' преобразована в новую величину y , так что это условие выполняется — см. § 10.3). Далее, предположим, что регрессия y по x (см. § 5.5) есть известная функция $f(x)$, содержащая некоторое количество параметров a_1, \dots, a_m ; причем предполагается также, что

¹⁾ Наиболее удобные для таких вычислений технические приемы читатель найдет в книгах, указанных в списке литературы.

эти параметры входят в $f(x)$ линейно:

$$\begin{aligned} \eta(x) = \mathfrak{M}\{y|x\} &= \int_{-\infty}^{\infty} y\varphi(y|x) dy = f(x; a_1, \dots, a_m) = \\ &= a_1 f_1(x) + \dots + a_m f_m(x). \end{aligned} \quad (1)$$

Наконец, предположим, что дисперсия $\sigma^2\{y|x\}$ величины y при фиксированном значении x либо есть постоянная, либо пропорциональна известной функции $g^2(x)$:

$$\sigma^2\{y|x\} = \int_{-\infty}^{\infty} [y - \eta(x)]^2 \varphi(y|x) dy = \sigma^2 g^2(x). \quad (2)$$

Функции $f_1(x), \dots, f_m(x)$ и $g(x)$ либо получаются из теоретических соображений, либо берутся в качестве гипотез, возникших при рассмотрении графика, на котором нанесены соответствующие значения величин x и y . Тогда возникает задача получения наилучших оценок a_1, \dots, a_m и s для параметров a_1, \dots, a_m и σ по наблюдаемым значениям y , предполагаемым стохастически независимыми. Эту задачу можно считать решенной, поскольку ее, очевидно, можно рассматривать как задачу уравнивания с помощью элементов, причем элементами являются параметры a_1, \dots, a_m . Пусть x_1, \dots, x_n суть заданные (не обязательно все различные) значения x , а y_1, \dots, y_n — соответствующие наблюдаемые значения y .

Тогда можно просто пользоваться формулами § 12.3—§ 12.9, изменив в них обозначения следующим образом:

$$\begin{aligned} \{l_1, \dots, l_n\} &\longleftrightarrow \{y_1, \dots, y_n\} \\ \{\lambda_1, \dots, \lambda_n\} &\longleftrightarrow \{\eta_1, \dots, \eta_n\} \\ \{\bar{l}_1, \dots, \bar{l}_n\} &\longleftrightarrow \{\bar{y}_1, \dots, \bar{y}_n\} \\ \{p_1, \dots, p_n\} &\longleftrightarrow \left\{ \frac{1}{g^2(x_1)}, \dots, \frac{1}{g^2(x_n)} \right\} \\ \{\xi_1, \dots, \xi_m\} &\longleftrightarrow \{a_1, \dots, a_m\} \\ \{\bar{x}_1, \dots, \bar{x}_m\} &\longleftrightarrow \{a_1, \dots, a_m\} \\ \{a_{10}, \dots, a_{n0}\} &\longleftrightarrow \{0, \dots, 0\} \\ \left\{ \begin{array}{c} a_{11}, \dots, a_{1m} \\ \cdot \\ \cdot \\ \cdot \\ a_{n1}, \dots, a_{nm} \end{array} \right\} &\longleftrightarrow \left\{ \begin{array}{c} f_1(x_1), \dots, f_m(x_1) \\ \cdot \\ \cdot \\ \cdot \\ f_1(x_n), \dots, f_m(x_n) \end{array} \right\}. \end{aligned} \quad (3)$$

Пример 1. В простейшем случае анализа регрессии регрессия $\eta(x)$ величины y по x есть линейная функция от x

$$\eta(x) = a + \beta x, \quad (4)$$

т. е. в формуле (1) $a_1 = a$, $a_2 = \beta$, $f_1(x) = 1$, $f_2(x) = x$. Если предположить для простоты, что $g(x) \equiv 1$, то наилучшие оценки

$\bar{y}_1, \dots, \bar{y}_n$ для величин η_1, \dots, η_n даются следующими условными уравнениями:

$$\begin{aligned} \bar{y}_1 &= a + bx_1, \\ &\dots \dots \dots \\ \bar{y}_n &= a + bx_n. \end{aligned} \quad (5)$$

Таким образом,

$$A^0 = 0, \quad A_{n2} = \begin{Bmatrix} 1 & x_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{Bmatrix}, \quad P = E, \quad (6)$$

и

$$\begin{aligned} B_{22} &= A^* \cdot P \cdot A = \begin{Bmatrix} n & [x] \\ [x] & [x^2] \end{Bmatrix}, \quad |B| = n[x^2] - [x]^2 = n[x^2] - n^2 \bar{x}^2 = \\ &= n[(x - \bar{x})^2], \quad \bar{x} = \frac{[x]}{n}. \\ B_{22}^{-1} &= \frac{1}{n[(x - \bar{x})^2]} \begin{Bmatrix} [x^2] & -[x] \\ -[x] & n \end{Bmatrix}. \end{aligned} \quad (7)$$

Следовательно,

$$\begin{aligned} \begin{Bmatrix} a \\ \beta \end{Bmatrix} \approx \begin{Bmatrix} a \\ b \end{Bmatrix} &= B^{-1} \cdot A^* \cdot Y = \begin{Bmatrix} \frac{[x^2][y] - [x][xy]}{n[(x - \bar{x})^2]} \\ \frac{n[xy] - [x][y]}{n[(x - \bar{x})^2]} \end{Bmatrix} = \begin{Bmatrix} \frac{[x^2]\bar{y} - [xy]\bar{x}}{[(x - \bar{x})^2]} \\ \frac{[xy] - n\bar{x}\bar{y}}{[(x - \bar{x})^2]} \end{Bmatrix}, \quad (8) \\ \bar{y} &= \frac{[y]}{n}. \end{aligned}$$

Далее,

$$\sigma \approx s = \sqrt{\frac{[uv]}{n-2}}, \quad v_i = \bar{y}_i - y_i, \quad i = 1, 2, \dots, n \quad (9)$$

и

$$\sigma\{a\} \approx \sqrt{\frac{[x^2]}{n[(x - \bar{x})^2]}} s, \quad \sigma\{b\} \approx \frac{s}{\sqrt{[(x - \bar{x})^2]}}, \quad \sigma\{s\} \approx \frac{s}{\sqrt{2n-4}}. \quad (10)$$

Упражнение 1. На практике может оказаться более удобной следующая запись формулы (4):

$$\eta(x) = a + \beta(x - \bar{x}), \quad \bar{x} = \frac{[x]}{n}. \quad (4a)$$

Причина для предпочтения формулы (4a) по сравнению с формулой (4) заключается частью в том, что это может упростить численные вычисления, и частью в том, что получающиеся при этом оценки a и b будут статистически независимыми. Докажите последнее утверждение, выведя формулы, соответствующие формулам (5) — (10).

Пример 2. Эксперименты показывают, что при истечении воды из сосуда через небольшое отверстие квадрат объема V вытекающей за секунду воды есть линейная функция высоты h воды в сосуде, отсчитываемой от уровня отверстия, т. е.

$$V^2 = a + \beta h,$$

причем буквы, как и всегда в физических уравнениях, обозначают истинные значения величин. Таким образом, V^2 соответствует нашему y и h — нашему x . Строго говоря, и V^2 и h подвержены экспериментальным ошибкам, но эти ошибки для h можно здесь считать пренебрежимо малыми. Далее, V^2 можно с уверенностью считать нормально распределенной величиной с постоянной дисперсией. Таким образом, все условия применимости анализа регрессии выполнены, и по экспериментальным значениям, приведенным в первых двух столбцах табл. 6, можно проделать все вычисления в соответствии с формулами примера 1. Однако, чтобы упростить численные вычисления, более удобно сначала принять предварительные значения α и β , скажем, α_0 и β_0 , определив их, например, по графику. Тогда запишем

$$V^2 = \alpha_0 + \alpha' + (\beta_0 + \beta')h \quad \text{или} \quad V^2 - (\alpha_0 + \beta_0 h) = \alpha' + \beta' h.$$

Теперь уже применим формулы примера 1 к $y = V^2 - (\alpha_0 + \beta_0 h)$ и $h = x$. При численных значениях

$$\alpha_0 = -0,4320, \quad \beta_0 = 0,5035$$

мы получаем значения, приведенные в 3—6 столбцах табл. 6. По значениям из табл. 6 и по формулам примера 1 находим

$$\begin{aligned} \alpha' &= -0,007727 \approx -0,0077 \\ b' &= 0,001826 \approx 0,0018, \end{aligned}$$

т. е.

$$\begin{aligned} \alpha &= \alpha_0 + \alpha' = -0,4397, \\ \beta &= \beta_0 + b' = 0,5053. \end{aligned}$$

По этим значениям получаем значения \bar{y}_i , v_i и v_i^2 , приведенные в столбцах 7—9, так что

$$\begin{aligned} \sigma \approx s &= \sqrt{\frac{287,25}{10-2}} \cdot 10^{-3} = 0,005992 \approx 0,0060, \\ \sigma \{a\} &\approx \sqrt{\frac{111,0}{10 \times 6,7}} s = 0,007712 \approx 0,0077, \\ \sigma \{b\} &\approx \frac{s}{\sqrt{6,7}} = 0,002315 \approx 0,0023, \\ \sigma \{s\} &\approx \frac{s}{\sqrt{20-4}} = 0,001498 \approx 0,0015. \end{aligned}$$

Упражнение 2. Пусть x обозначает скорость автомобиля (в милях в час), а y — соответствующую длину пути торможения (в футах). $\mathfrak{M}\{y|x\}$ зависит от x , от среднего времени реакции шофера α и от эффективности тормозов. В течение времени α автомобиль продвигается на расстояние αx . Далее, как получено из экспериментов, расстояние, проходимое с момента включения тормозов, пропорционально квадрату скорости. Поэтому следует ожидать, что регрессия y по x определяется формулой

$$\eta(x) = \alpha x + \beta x^2.$$

Таблица 6

t_1	V_i^2	$a_0 + 3\beta t_1$	$y_i = V_i^2 - (a_0 + 3\beta t_1)$	x_i^2	$x_i y_i$	\bar{y}_i	$v_0 = \bar{y}_i - y_i$	v_i^2
4,530 сж	1,852 см ⁶ /сек ²	1,8488	+3,2 · 10 ⁻³	20,2	+14,4 · 10 ⁻³	+0,51 · 10 ⁻³	-2,69 · 10 ⁻³	7,24 · 10 ⁻⁶
4,241	1,703	1,7034	-0,4	17,6	-1,8	-0,02	+0,38	0,14
3,952	1,553	1,5579	-4,9	16,0	-19,6	-0,54	+4,36	19,01
3,663	1,408	1,4123	-4,3	13,7	-15,9	-1,06	+3,24	10,50
3,374	1,274	1,2668	+7,2	11,5	+24,5	-1,59	-8,79	77,26
3,085	1,112	1,1213	-9,3	9,6	-28,8	-2,11	+7,19	51,70
2,796	0,972	0,9758	-3,8	7,8	-10,6	-2,64	+1,16	1,35
2,507	0,832	0,8308	+1,7	6,2	+4,2	-3,17	-4,87	23,72
2,218	0,688	0,6848	+3,2	4,8	+7,0	-3,70	-6,90	47,61
1,929	0,528	0,5392	-11,2	3,6	-21,3	-4,22	+6,98	48,72

$[x] = 32,295$
 $x = 3,2295$

$[y] = -0,0186$
 $y = -0,00186$

$[x^2] = 111,0$
 $[xy] = -0,0479$

$[vv] = 287,25 \cdot 10^{-6}$

Далее, основной вклад в дисперсию величины y производит дисперсия σ времени реакции шофера, равного в среднем α . Пренебрегая вкладом в дисперсию, производимым слагаемым βx^2 , по сравнению со вкладом слагаемого αx , мы должны ожидать, что

$$\sigma \{y | x\} = \sigma \alpha x.$$

Найдите наилучшие оценки для параметров α, β и σ по данным, приведенным в табл. 7.

Т а б л и ц а 7

x (в милях в час)	4	8	10	12	14	16	18	20	22	24
y (в футах)	2	16	18	20	36	40	42	56	66	70

Дальнейшее изложение анализа регрессии и его обобщения на случай нескольких независимых переменных (означающего лишь замену единственной переменной x в формулах (1) и (2) несколькими переменными x_1, \dots, x_k), читатель найдет в книгах, указанных в списке литературы.

ПРИЛОЖЕНИЕ 1

$n!$

Под $n!$ (читается „ n факториал“), где n — целое положительное число, понимается число ¹⁾

$$n! = n(n-1)(n-2) \dots 3 \cdot 2 \cdot 1.$$

С помощью интегрирования по частям находим, что для целых положительных n

$$n! = \int_0^{\infty} t^n e^{-t} dt \quad (e = 2,71828\dots);$$

полагая этот интеграл равным $f(n)$, находим

$$f(n) = - \int_0^{\infty} t^n d(e^{-t}) = - t^n e^{-t} \Big|_0^{\infty} + n \int_0^{\infty} t^{n-1} e^{-t} dt = n f(n-1),$$

так как $\lim_{t \rightarrow \infty} t^n e^{-t} = 0$. Повторяя этот процесс, получим

$$f(n) = n(n-1) \dots 3 \cdot 2 \cdot \int_0^{\infty} e^{-t} dt = n!$$

Можно показать, что этот интеграл существует для всех значений $n > -1$. Поэтому можно определить $n!$ для всех $n > -1$, положив его равным этому интегралу. Функция $\Gamma(n) = (n-1)!$ называется

гамма-функцией. В частности, $0! = \int_0^{\infty} e^{-t} dt = 1$.

Можно показать, что для больших значений n имеет место так называемая формула Стирлинга

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \exp\left[\frac{\theta}{12n}\right], \quad |\theta| < 1,$$

где θ есть некоторое число, зависящее от n . Если положить $\theta = 0$, то ошибка, даваемая полученным выражением, будет меньше $10^0/0$ уже для $n = 1$.

¹⁾ Более подробное изложение и доказательства см. Уиттекер и Ватсон, Курс современного анализа, т. II, ГТТИ, 1934.

С помощью подстановки $ax^2 = t$ интеграл

$$\int_0^{\infty} x^n e^{-ax^2} dx,$$

часто полезный в теории вероятностей, можно выразить через предыдущий интеграл и обратно. Получим

$$\begin{aligned} \int_0^{\infty} x^n e^{-ax^2} dx &= \frac{1}{\frac{n+1}{2}} \int_0^{\infty} t^{\frac{n}{2}} e^{-t} \frac{dt}{2\sqrt{t}} = \\ &= \frac{1}{2a^{\frac{n+1}{2}}} \int_0^{\infty} t^{\frac{n-1}{2}} e^{-t} dt = \frac{1}{2a^{\frac{n+1}{2}}} \left(\frac{n-1}{2}\right)!. \end{aligned}$$

Полагая $n=0$, $a=1$ и $x^2 = \frac{t^2}{2}$, мы получим, таким образом, в силу (7.1.2)

$$\frac{1}{2} \left(-\frac{1}{2}\right)! = \int_0^{\infty} e^{-x^2} dx = \frac{1}{2\sqrt{2}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \frac{\sqrt{\pi}}{2},$$

т. е. $\left(-\frac{1}{2}\right)! = \sqrt{\pi}$. Отсюда находим $\left(\frac{1}{2}\right)! = \frac{1}{2} \left(-\frac{1}{2}\right)! = \frac{\sqrt{\pi}}{2}$,

$$\left(\frac{3}{2}\right)! = \frac{3}{2} \left(\frac{1}{2}\right)! = \frac{3}{4} \sqrt{\pi}, \quad \left(\frac{5}{2}\right)! = \frac{5}{2} \left(\frac{3}{2}\right)! = \frac{15}{8} \sqrt{\pi}$$

и т. д.

ПРИЛОЖЕНИЕ 2

Теория матриц

Дадим краткую сводку основных понятий и теорем теории матриц¹⁾.

1. Под *матрицей* понимается прямоугольная таблица чисел, называемых *элементами*. Мы будем обозначать матрицы большими жирными буквами, а элементы маленькими:

$$A = A_{mn} = \{a_{rs}\} = \begin{Bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{Bmatrix}.$$

Мы говорим, что матрица A_{mn} имеет m строк и n столбцов. При $n=1$ матрица A называется *матрицей-столбцом*, при $m=1$ — *матрицей-строкой*. При $m=n=1$ матрица A состоит из одного числа. Элемент матрицы A , стоящий в r -ой строке и s -ом столбце, мы будем обозначать также через $(A)_{rs}$.

2. Будем говорить, что матрицы A и B тождественны $A=B$, если они имеют одно и то же число строк и одно и то же число столбцов и $a_{rs}=b_{rs}$ при всех r и s .

3. *Нулевая матрица* O есть матрица, все элементы которой равны нулю $a_{rs}=0$. Подчеркнем, что существует бесконечно много различных нулевых матриц — по одной на каждую пару значений m и n .

4. Поменяв ролями строки и столбцы в матрице A_{mn} , мы получим новую матрицу, имеющую n строк и m столбцов. Она называется *транспонированной* по отношению к матрице A и обозначается через A_{nm}^* .

5. Под матрицей αA , где α — произвольное число, подразумевается матрица, полученная умножением всех элементов матрицы A на α , т. е.

$$\alpha A = \alpha_i^r \{a_{rs}\} = \{\alpha a_{rs}\}.$$

6. Если матрицы A и B имеют одинаковое число строк и одинаковое число столбцов, то их *суммой* называется матрица

$$C_{mn} = A_{mn} + B_{mn}$$

1) Более детальное изложение, а также доказательства приведенных теорем читатель может найти в книгах: Фрезер, Дункан и Коллар, Теория матриц, Издательство иностранной литературы, Москва, 1950; Курант и Гильберт, Методы математической физики, т. I, гл. 1, ГИИТЛ, 1951.

с элементами

$$c_{rs} = a_{rs} + b_{rs}.$$

При этом определении сохраняются обычные свойства суммы и разности.

7. Если число столбцов матрицы A_{mn} равно числу строк матрицы B_{np} , то их *произведением*

$$C_{mp} = A_{mn} \cdot B_{np}$$

называется матрица с элементами

$$c_{rs} = a_{r1}b_{1s} + a_{r2}b_{2s} + \dots + a_{rn}b_{ns} = \sum_{i=1}^n a_{ri}b_{is}.$$

Таким образом, произведение двух матриц имеет столько же строк, сколько первый множитель, и столько столбцов, сколько второй множитель. При этом определении сохраняется ассоциативный закон

$$(A_{mn} \cdot B_{np}) \cdot C_{pq} = A_{mn} \cdot (B_{np} \cdot C_{pq})$$

и дистрибутивный закон.

$$\begin{aligned} A_{mn} \cdot (B_{np} + C_{np}) &= A_{mn} \cdot B_{np} + A_{mn} \cdot C_{np}, \\ (A_{mn} + B_{mn}) \cdot C_{np} &= A_{mn} \cdot C_{np} + B_{mn} \cdot C_{np}. \end{aligned}$$

Коммутативный закон $A \cdot B = B \cdot A$ места, однако, не имеет. Так, например, если число столбцов матрицы B отлично от числа строк матрицы A , то произведение $B \cdot A$ нельзя даже образовать, в то время как произведение $A \cdot B$ существует.

8. Для транспонирования произведения существует правило

$$(A_{mn} \cdot B_{np})^* = B_{pn}^* \cdot A_{nm}^*,$$

которое может быть непосредственно распространено на произведение любого конечного числа множителей.

9. *Квадратной* матрицей A_{nn} называется матрица, число строк которой равно числу ее столбцов.

10. Квадратная матрица называется *симметричной*, если $A^* = A$, т. е. $a_{rs} = a_{sr}$ для всех r и s . Для произвольной матрицы A оба произведения $A \cdot A^*$ и $A^* \cdot A$ существуют и являются симметричными, хотя и не обязательно тождественными матрицами.

11. *Диагональной* матрицей называется квадратная матрица, для которой $a_{rs} = 0$ при $r \neq s$. Элементы a_{rr} называются *диагональными элементами*.

12. *Единичной матрицей* называется диагональная матрица, для которой $a_{rr} = 1$ при всех r . Она обозначается через E . Для любой матрицы

$$A \cdot E = E \cdot A = A$$

при условии, что соответствующие произведения существуют. Подчеркнем, что существует бесконечно много единичных матриц, по одной на каждое значение n .

13. Под *детерминантом* $|A|$ квадратной матрицы A понимается детерминант $|a_{rs}|$. В частности, если A есть диагональная матрица, то $|A| = a_{11} a_{22} \dots a_{nn}$. Всегда имеет место равенство $|A^*| = |A|$.

14. Если $|A| \neq 0$, то существует одна и только одна матрица $A^{-1} = \{(A^{-1})_{rs}\}$, такая, что

$$A \cdot A^{-1} = A^{-1} \cdot A = E.$$

Эта матрица называется *обратной* по отношению к матрице A . Элементы матрицы A^{-1} определяются равенством

$$(A^{-1})_{sr} = \frac{K_{rs}}{|A|},$$

где K_{rs} — алгебраическое дополнение элемента a_{rs} в детерминанте $|A|$, т. е. умноженный на $(-1)^{r+s}$ минор, полученный из $|A|$ вычеркиванием его r -ой строки и s -го столбца. В частности, если A есть диагональная матрица, то матрица A^{-1} тоже является диагональной и

$$(A^{-1})_{rr} = \frac{1}{a_{rr}}.$$

15. Если $|A| \neq 0$, то матричное уравнение

$$A \cdot X = B$$

имеет одно и только одно решение, которое мы получим, если умножим обе части последнего равенства слева на A^{-1} :

$$A^{-1} \cdot A \cdot X = X = A^{-1} \cdot B.$$

Точно так же матричное уравнение

$$Y \cdot A = B$$

имеет при $|A| \neq 0$ одно и только одно решение, получаемое с помощью умножения справа на A^{-1} :

$$Y \cdot A \cdot A^{-1} = Y = B \cdot A^{-1}.$$

Отметим, что, вообще говоря, $X \neq Y$.

16. Если матрицы A^{-1} и B^{-1} существуют, то

$$(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}.$$

17. Однородная квадратичная форма $\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$ ($a_{ij} = a_{ji}$)

может быть записана в виде $X^* \cdot A \cdot X$, где $A = \{a_{rs}\}$, а X есть матрица-столбец с элементами x_1, \dots, x_n .

18. *Ортогональной* матрицей называется квадратная матрица, для которой

$$A \cdot A^* = A^* \cdot A = E.$$

Для ортогональной матрицы $|A| = \pm 1$ и $A^* = A^{-1}$.

19. Если ввести в выражении $X^* \cdot A \cdot X$ новые переменные, положив

$$X_{n1} = F_{nn} \cdot Y_{n1},$$

то получим

$$X^* \cdot A \cdot X = Y^* \cdot B \cdot Y, \quad B = F^* \cdot A \cdot F.$$

Для произвольной симметричной матрицы A существует ортогональная матрица F такая, что матрица B является диагональной. В этом случае мы имеем $|A| = |B| = b_{11}b_{22} \dots b_{nn}$.

Таблица I
Нормальное распределение (§ 7.1)

t	$\psi(t)$	$\Psi(t)$	$2\Psi(t)-1$
0,0	0,39894	0,50000	0,00000
0,1	39695	53983	07966
0,2	39104	57926	15852
0,3	38139	61791	23582
0,4	36827	65542	31084
0,5	0,35207	0,69146	0,38292
0,6	33322	72575	45150
0,7	31225	75804	51608
0,8	28969	78814	57628
0,9	26609	81594	63188
1,0	0,24197	0,84134	0,68268
1,1	21785	86433	72866
1,2	19419	88493	76986
1,3	17137	90320	80640
1,4	14973	91924	83848
1,5	0,12952	0,93319	0,86638
1,6	11092	94520	89040
1,7	09405	95543	91086
1,8	07895	96407	92814
1,9	06562	97128	94256
2,0	0,05399	0,97725	0,95450
2,1	04398	98214	96428
2,2	03547	98610	97220
2,3	02833	98928	97856
2,4	02239	99180	98360
2,5	0,01753	0,99379	0,98758
2,6	01358	99534	99068
2,7	01042	99653	99306
2,8	00792	99744	99488
2,9	00595	99813	99626
3,0	0,00443	0,99865	0,99730
3,1	00327	99903	99806
3,2	00238	99931	99862
3,3	00172	99952	99904
3,4	00123	99966	99932
3,5	0,00087	0,99977	0,99954
3,6	00061	99984	99968
3,7	00042	99989	99978
3,8	00029	99993	99986
3,9	00020	99995	99990
4,0	0,00013	0,99997	0,99994

Таблица II
Нормальное распределение
(§ 7.4)

p	a
1,0	0
0,9	0,12566
0,8	0,25335
0,7	0,38532
0,6	0,52440
0,5	0,67449
0,4	0,84162
0,3	1,03643
0,2	1,28155
0,1	1,64485
0,05	1,95996
0,01	2,57583
0,001	3,29053
10 ⁻⁴	3,89059
10 ⁻⁵	4,41717
10 ⁻⁶	4,89164
10 ⁻⁷	5,32672
10 ⁻⁸	5,73073
10 ⁻⁹	6,10941

Таблица III
t-распределение (§ 7. 9)

P	0,1	0,05	0,01	0,001
f				
1	6,314	12,706	63,657	636,619
2	2,920	4,303	9,925	31,598
3	2,353	3,182	5,841	12,941
4	2,132	2,776	4,604	8,610
5	2,015	2,571	4,032	6,859
6	1,943	2,447	3,707	5,959
7	1,895	2,365	3,499	5,405
8	1,860	2,306	3,355	5,041
9	1,833	2,262	3,250	4,781
10	1,812	2,228	3,169	4,587
11	1,796	2,201	3,106	4,437
12	1,782	2,179	3,055	4,318
13	1,771	2,160	3,012	4,221
14	1,761	2,145	2,977	4,140
15	1,753	2,131	2,947	4,073
16	1,746	2,120	2,921	4,015
17	1,740	2,110	2,898	3,965
18	1,734	2,101	2,878	3,922
19	1,729	2,093	2,861	3,883
20	1,725	2,086	2,845	3,850
21	1,721	2,080	2,831	3,819
22	1,717	2,074	2,819	3,792
23	1,714	2,069	2,807	3,767
24	1,711	2,064	2,797	3,745
25	1,708	2,060	2,787	3,725
26	1,706	2,056	2,779	3,707
27	1,703	2,052	2,771	3,690
28	1,701	2,048	2,763	3,674
29	1,699	2,045	2,756	3,659
30	1,697	2,042	2,750	3,646
35	1,689	2,030	2,724	3,591
40	1,684	2,021	2,704	3,551
45	1,679	2,014	2,689	3,522
50	1,676	2,008	2,677	3,497
60	1,671	2,000	2,660	3,460
70	1,667	1,995	2,648	3,436
80	1,664	1,990	2,639	3,416
90	1,662	1,987	2,632	3,401
100	1,660	1,984	2,626	3,391
120	1,658	1,980	2,617	3,373
∞	1,645	1,960	2,576	3,291

Таблица IV
r-распределение (§ 7. 10)

P	0,1	0,05	0,01	0,001
f				
1	1,397	1,409	1,414	1,414
2	1,559	1,645	1,715	1,730
3	1,611	1,757	1,918	1,982
4	1,631	1,814	2,051	2,178
5	1,640	1,848	2,142	2,329
6	1,644	1,870	2,208	2,447
7	1,647	1,885	2,256	2,540
8	1,648	1,895	2,294	2,616
9	1,649	1,903	2,324	2,678
10	1,649	1,910	2,348	2,730
11	1,649	1,916	2,368	2,774
12	1,649	1,920	2,385	2,812
13	1,649	1,923	2,399	2,845
14	1,649	1,926	2,412	2,874
15	1,649	1,928	2,423	2,899
16	1,649	1,931	2,432	2,921
17	1,649	1,933	2,440	2,941
18	1,649	1,935	2,447	2,959
19	1,649	1,936	2,454	2,975
20	1,649	1,937	2,460	2,990
21	1,649	1,938	2,465	3,003
22	1,648	1,940	2,470	3,015
23	1,648	1,941	2,475	3,026
24	1,648	1,941	2,479	3,037
25	1,648	1,942	2,483	3,047
26	1,648	1,943	2,487	3,056
27	1,648	1,943	2,490	3,064
28	1,648	1,944	2,492	3,071
29	1,648	1,945	2,495	3,078
30	1,648	1,945	2,498	3,085
35	1,648	1,948	2,509	3,113
40	1,648	1,949	2,518	3,134
45	1,647	1,950	2,524	3,152
50	1,647	1,951	2,529	3,166
60	1,646	1,953	2,537	3,186
70	1,646	1,954	2,542	3,201
80	1,646	1,955	2,547	3,211
90	1,646	1,956	2,550	3,220
100	1,646	1,956	2,553	3,227
120	1,646	1,957	2,556	3,237
∞	1,645	1,960	2,576	3,291

Таблица V
 ω^2 -распределение (§ 7.11): $P=0,05$

f_1	1	2	3	4	5	6	8	12	24	∞
f_2										
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,35	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

ЗАДАЧИ

Задачи к главам 1—3

1. Из шести колод карт, каждая из которых содержит 52 карты, выбирается наудачу по одной карте из каждой колоды. Найти вероятность того, что четыре карты окажутся красной масти, а две — черной.

2. Из колоды карт в 52 карты выбираются наудачу шесть карт. Найти вероятность того, что четыре карты окажутся красной масти, а две — черной.

3. Найти вероятность ровно одного выпадения шестерки в четырех бросках игральной кости.

4. Выбраны наудачу четыре целых положительных числа. Найти вероятность того, что они имеют общий множитель. (Использовать формулу

$$\prod_{i=1}^{\infty} \left(1 - \frac{1}{p_i^4}\right) = \frac{90}{\pi^4},$$

где p_i является i -ым простым числом.)

5. Олень убегает от охотника с постоянной скоростью v . Вероятность попасть в оленя с расстояния $d > v$ предполагается равной $\frac{v^2}{d^2}$. Охотник стреляет в оленя один раз с расстояния $d = 2v$ и в случае промаха стреляет еще раз с расстояния $d = 3v$ и т. д. Всего, однако, он может сделать не больше n выстрелов. Какова вероятность того, что охотник попадает в оленя?

6. Вывести общие формулы, аналогичные закону сложения IV, для

$$P(A+B+C) \text{ и } P(A+B+C+D).$$

7. Найти вероятность, того, что *по крайней мере* один игрок в бридж получит при сдаче 13 карт одной масти. (При игре в бридж колода карт в 52 карты сдается поровну между четверьмя игроками. Использовать результат задачи 6.)

8. У одного из игроков в бридж оказалось два туза. Какова вероятность того, что у его партнера окажется по крайней мере один из оставшихся двух тузов?

9. Вероятность наступления некоторого события равна $\frac{1}{10}$. Найти вероятность того, что это событие наступит *по крайней мере* 10 раз в 100 случаях.

10. Ожидается прибытие двух судов с бананами. Статистика показывает, что в 1% случаев груз бананов портится в дороге. Найти вероятность того, что а) оба судна придут с неиспорченным грузом, б) только одно судно придет с неиспорченным грузом и в) оба судна придут с испорченным грузом.

11. Какова вероятность торпедировать судно, если всего может быть выпущено три торпеды, а вероятность поражения одной торпедой равна 0,25?

12. Человек стоит в начале координат числовой оси. Он бросает монету и после каждого броска делает один шаг вправо при выпадении решетки и один шаг влево при выпадении герба. Обозначим через x абсциссу, соответствующую положению человека после 10 бросков. Каковы возможные значения величины x и каковы вероятности этих значений?

13. Торпедный катер находится на расстоянии d от цели. Вероятность поражения цели с этого расстояния одной торпедой равна 0,05. Катер приближается к цели на половину этого расстояния и выпускает три торпеды. Предполагая, что вероятность поражения обратно пропорциональна квадрату расстояния до цели, найти а) вероятность p_a трех попаданий, б) вероятность p_b двух попаданий, в) вероятность p_v по крайней мере двух попаданий, г) вероятность p_r по крайней мере одного попадания, д) вероятность p_d не попасть в цель ни одного раза.

14. Вероятность поражения фиксированной цели одним выстрелом с расстояния d равна 0,5. На каком расстоянии от цели должна быть расположена батарея из четырех орудий, чтобы вероятность ровно двух попаданий в залпе была равна $\frac{3}{32}$? (Вероятность попадания предполагается обратно пропорциональной квадрату расстояния.)

15. К автобусной остановке каждые четыре минуты подходит автобус линии A и каждые шесть минут автобус линии B . Предполагая времена ожидания автобусов этих двух линий независимыми, найти вероятность того, что а) первый подошедший автобус окажется автобусом линии A , б) автобус какой-нибудь линии подойдет в течение двух минут.

16. Две батареи A и B с одной и той же вероятностью попадания p делают каждая n выстрелов по общей цели. Если среди $2n$ выстрелов было ровно два попадания, найти вероятность того, что а) оба попадания были сделаны батареей A , б) одно из попаданий было сделано батареей A , а другое — батареей B и в) оба попадания были сделаны батареей B .

17. Вероятность того, что человек, проживший p лет, умрет до того, как ему исполнится $p+1$ год, равна P_p . Найти вероятность того, что человек, проживший 50 лет, умрет в течение следующих 5 лет.

18. A и B играют в „орел и решку“. A бросает первым и выигрывает при выпадении решетки. В случае выпадения герба монету бросает B и выигрывает в случае выпадения решетки. Если же снова выпадает герб, то монету бросает опять A и т. д. Если после $2n$ бросков не выигрывает ни A , ни B , то выигрывает третье лицо B . Найти вероятности P_A, P_B, P_B выигрыша A, B и B .

19. В карточной игре используются только карты 1, 2, ..., 10 одной определенной масти. В игре участвует пять человек, и она заключается в том, что карты раздаются игрокам в определенном порядке, так что каждый получает по две карты. Участник выигрывает или проигрывает в зависимости от того, равна сумма очков на его двух картах 11 или нет. Какова вероятность того, что а) p заранее фиксированных игроков выигрывают, а остальные $5-p$ проигрывают, б) k -ый игрок первым набирает 11 очков, в) ни один из пяти игроков не набирает 11 очков?

20. Пусть вероятность того, что погода в данный день будет такой же как и в предыдущий день (дождливой или нет), равна p и пусть вероятность того, что первый день года — дождливый, равна P . Найти вероятность P_n того, что n -ый день — дождливый, и найти предел вероятности P_n при $n \rightarrow \infty$.

Задачи к главам 4—6

21. Три батареи с вероятностями попадания, равными соответственно 0,1, 0,2 и 0,3, делают каждая по одному выстрелу. Найти вероятности всех возможных значений числа попаданий и соответствующую функцию распределения.

22. Пусть случайная величина x есть число бросаний монеты, продолжающихся до первого выпадения решетки. Найти функцию распределения, среднее значение и дисперсию.

23. Пусть случайная величина x имеет плотность вероятности

$$\varphi(t) = \begin{cases} 0 & \text{при } -\infty < t \leq -\frac{1}{n} \\ n + n^2t & \text{при } -\frac{1}{n} \leq t \leq 0 \\ n - n^2t & \text{при } 0 \leq t \leq \frac{1}{n} \\ 0 & \text{при } \frac{1}{n} \leq t < \infty. \end{cases}$$

Найти $\Phi(t)$, μ и σ . Начертить графики $\varphi(t)$ и $\Phi(t)$. Показать, что $\Phi(t) \rightarrow \varepsilon(t)$ при $n \rightarrow \infty$.

24. На числовой прямой выбираются наудачу два числа, лежащие между 0 и 1. Найти вероятность того, что их произведение x меньше, чем t , $0 < t < 1$. Найти среднее значение μ и стандартное отклонение σ случайной величины x .

25. На окружности наудачу выбираются точки A , B и C . Найти вероятность того, что треугольник ABC будет остроугольным.

26. Три числа a , b и c , заключенные между 0 и 1, выбираются наудачу. Найти вероятность того, что уравнение

$$ax^2 + 2bx + c = 0$$

имеет действительные корни.

27. Случайная величина x имеет биномиальное распределение, причем $\mu = 2$ и $\sigma = \frac{3}{2}$. Найти вероятности

$$P(3 < x < 5), P(3 \leq x < 5), P(3 < x \leq 5) \text{ и } P(3 \leq x \leq 5).$$

28. Пять лиц, которым поставлены в соответствие пять различных чисел, рассаживаются наудачу в пяти креслах, занумерованных теми же числами. Пусть x есть случайная величина, равная числу лиц из этой пятерки, которые попали в кресло со „своим“ номером. Найти функцию распределения случайной величины x , а также ее среднее значение μ и стандартное отклонение σ .

29. Урна содержит четыре красных и три белых шара. Из нее наудачу вынимаются четыре шара. Пусть x есть случайная величина, равная числу красных шаров среди четырех вынутых. Найти среднее значение и дисперсию случайной величины x . Решить ту же задачу при условии, что шары вынимаются из урны один за другим, причем, прежде чем вынуть следующий шар, мы опускаем обратно в урну вынутый ранее.

30. Шесть шаров занумерованы цифрами от 1 до 6. Из них наудачу выбирается три шара. Пусть случайная величина x является наибольшим из номеров выбранных шаров. Начертить график функции распределения случайной величины x и найти ее среднее значение и дисперсию.

31. Случайная величина x задана своей плотностью вероятности $\varphi(t)$. Найти плотности вероятностей случайных величин $y = \frac{1}{x}$ и $z = \cos x$. Предположить, в частности, что случайная величина x имеет нормальное распределение.

32. Случайная величина x имеет плотность вероятности

$$\varphi(t) = \frac{a}{e^t + e^{-t}}.$$

Определить значение a , после чего найти вероятность того, что наибольшее из двух наблюдений будет меньше единицы.

33. Пусть случайная величина x с произвольным распределением имеет конечное среднее значение μ и конечное стандартное отклонение σ . Найти случайную величину x' с теми же μ и σ , что и x , имеющую прямоугольное распределение.

34. Случайная величина x , принимающая только положительные значения, имеет логарифмически-нормальное распределение

$$\frac{d\Phi}{dt} = \varphi(t) = \frac{1}{\sqrt{2\pi} \beta t} \exp \left[-\frac{(\ln t - a)^2}{2\beta^2} \right].$$

Найти ее среднее значение и дисперсию.

35. Пусть случайная величина x , принимающая только положительные значения, имеет плотность вероятности

$$\frac{d\Phi}{dt} = \varphi(t) = e^{-t},$$

а x_1, \dots, x_n суть n независимых наблюдений величины x . Найти распределение случайной величины

$$z_n = x_1 + x_2 + \dots + x_n,$$

и ее среднее значение и дисперсию.

36. Пусть случайная величина x имеет заданную функцию распределения $\Phi(t)$, для которой $\mu = 0$ и σ конечно. Показать, что

$$\Phi(t) \leq \frac{\sigma^2}{\sigma^2 + t^2} \quad \text{при } t < 0,$$

$$\Phi(t) \geq \frac{t^2}{\sigma^2 + t^2} \quad \text{при } t > 0.$$

Показать на примерах, что эти неравенства не могут быть улучшены.

37. Пусть дискретные случайные величины x и y принимают соответственно значения x_1, \dots, x_k и y_1, \dots, y_l с вероятностями p_1, \dots, p_k и q_1, \dots, q_l . Пусть вероятность одновременного выполнения равенств $x = x_i$ и $y = y_j$ равна φ_{ij} . Определим τ^2 , положив

$$\tau^2 = \frac{1}{l-1} \sum_{i=1}^k \sum_{j=1}^l \frac{(\varphi_{ij} - p_i q_j)^2}{p_i q_j}.$$

Показать, что 1) $0 \leq \tau^2 \leq 1$, 2) необходимым и достаточным условием для $\tau^2 = 0$ является независимость величин x и y , 3) необходимым и достаточным условием для $\tau^2 = 1$ является полная зависимость y от x , т. е. то, что любому определенному значению x с вероятностью, равной единице, соответствует некоторое определенное значение y .

38. Пусть случайная величина x есть число красных карт у одного игрока в бридж, y — число красных карт у его партнера. Найти коэффициент корреляции $\rho \{x, y\}$.

39. Пусть $z = (x, y)$ есть произвольная двумерная случайная величина, такая, что моменты $\mu_x, \mu_y, \mu_{xx}, \mu_{yy}, \mu_{xy}$ первого и второго порядка для обоих распределений конечны и $\sigma_x > 0, \sigma_y > 0, |\rho| < 1$. Показать, что равномерное распределение массы по площади, ограниченной так называемым эллипсом рассеяния

$$\frac{1}{1 - \rho^2} \left(\frac{(x - \mu_x)^2}{\sigma_x^2} - \frac{2\rho(x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right) = 4,$$

имеет те же моменты первого и второго порядка, что и z . Показать, что прямые средней квадратической регрессии являются диаметрами этого эллипса, сопряженными с координатными осями.

40. Пусть случайные величины x_1 и x_2 имеют распределение Коши соответственно с параметрами μ_1, α_1 и μ_2, α_2 . Показать, что величина $x = x_1 + x_2$ имеет распределение Коши с параметрами $\mu = \mu_1 + \mu_2, \alpha = \alpha_1 + \alpha_2$.

41. Пусть при размножении бактерий делением вероятность бактерии разделиться за интервал времени dt равна λdt . Найти вероятности $P_i(t); i = 1, 2, 3, \dots$ наличия i бактерий через время t , если в момент $t=0$ была одна бактерия. Найти $\mathcal{M}\{x\}$ и $\sigma\{x\}$.

42. Решить задачу 12 с помощью производящих функций.

43. Урна содержит ν шаров, занумерованных числами от 1 до ν . Один шар выбирается наудачу, замечается его номер, и шар возвращается в урну. Этот эксперимент повторяется λ раз. Найти с помощью производящих функций вероятность того, что сумма λ замеченных номеров имеет заданное значение.

44. Найти производящую функцию для биномиального распределения. Применить результат к доказательству того, что сумма двух биномиально распределенных величин имеет биномиальное распределение.

45. Найти производящую функцию для распределения Пуассона. Применить результат для доказательства того, что сумма двух величин с пуассоновскими распределениями имеет распределение Пуассона.

Задачи к главам 7—8

46. Пусть случайная величина x имеет нормальное распределение со средним значением $\mu = 0$. Найти распределение величины $y = \epsilon x$.

47. Будем предполагать, что при стрельбе из орудия боковое отклонение x и отклонение по дальности y независимы и нормально распределены с одной и той же дисперсией σ^2 . Бесконечно большая горизонтальная цель с началом координат в центре рассеяния разделена четырьмя прямыми с уравнениями $x = 0,6745 \sigma, x = -0,6745 \sigma, y = 0,6745 \sigma, y = -0,6745 \sigma$ на части. Найти вероятность поражения каждой части.

48. M_1 бобов одного сорта S_1 смешиваются с N_2 бобами другого сорта S_2 . Числа M_1 и M_2 предполагаются достаточно большими, а длина бобов обоих сортов — распределенной нормально соответственно с параметрами μ_1, σ_1 и μ_2, σ_2 . Найти распределение длины бобов в смеси и его среднее значение и дисперсию.

49. Рост людей призывного возраста предполагается нормально распределенным со средним значением $\mu = 170$ см и стандартным отклонением $\sigma = 5$ см. Медицинская комиссия признает негодными для несения воинской службы всех людей, рост которых меньше 155 см. Найти распределение роста людей, признанных годными, его среднее значение и стандартное отклонение. Будут ли последние больше или меньше, чем соответственно 170 и 5?

50. Случайная величина x нормально распределена с параметрами $\mu = 0$ и σ . Сделаем два независимых наблюдения величины x и обозначим через z наибольшее из этих наблюдений. Показать, что

$$\mathcal{M}\{z\} = \frac{\sigma}{\sqrt{\pi}}.$$

51. Пусть x_1, x_2, \dots, x_n суть ν независимых наблюдений нормально распределенной случайной величины со средним значением $\mu = 0$ и стандартным отклонением $\sigma = 1$. Обозначим через z наибольшее из этих значений. Найти функцию распределения и плотность вероятности случайной величины z .

52. Пусть случайные величины x, y, z независимы и нормально распределены с одними и теми же параметрами $\mu = 0$ и σ . Найти плотность вероятности величины $x^2 + y^2 + z^2$. ■

53. Пусть случайные величины x и y независимы и нормально распределены. Найти коэффициент корреляции случайных величин $x + y$ и $x - y$.

Задачи к главам 10—11

54. Показать, что для распределения

$$\frac{d\Phi}{dt} = \varphi(t) = \frac{1}{2(1 - e^{-\theta})} e^{-|t|}, \quad -\theta \leq t \leq \theta$$

оценка максимального правдоподобия для θ равна наибольшему из чисел $|x_1|, \dots, |x_n|$.

55. Пусть некоторое событие имеет вероятность θ и встречается в x из n испытаний. Показать, что $\frac{x}{n}$ является наилучшей оценкой для θ .

56. 15 измерений начальной скорости снаряда дали следующие результаты (в метрах в секунду):

444,9	441,0	442,1
444,2	439,5	443,8
439,8	444,1	440,2
442,6	440,9	444,9
441,8	441,3	443,7

Вычислить наилучшие оценки истинного значения, дисперсии и дисперсий этих оценок.

57. 15 измерений широты в Кейптауне дали следующие результаты:

— 33°56'3",48	— 33°56'3",50	— 33°56'3",50
3",32	3",09	2",98
3",07	3",28	3",27
3",20	3",30	3",25
3",11	3",30	3",27

Вычислить наилучшие оценки истинного значения, дисперсии и дисперсий этих оценок.

58. При стрельбе из винтовки измерение боковых отклонений и отклонений по высоте от точки $(0, 0)$ дало следующие результаты:

Номер	Боковое отклонение	Отклонение по высоте
1	20	— 6,5
2	— 4	6,5
3	24	0
4	16	— 5
5	— 10,5	1,5
6	4,5	— 30
7	8	— 26
8	12,5	— 15
9	10	0,5
10	6,5	— 3
11	6	— 13
12	1,5	— 1

Найти центр рассеяния, две дисперсии и дисперсии их оценок.

59. Проверить, может ли какое-либо из измерений, приведенных в задаче 58, содержать грубую ошибку.

60. Проверить, значимо ли отклоняется средний центр рассеяния в задаче 58 от $(0, 0)$.

61. Проверить, существует ли в задаче 58 какая-нибудь корреляция между боковым отклонением и отклонением по высоте.

62. В двух различных сериях выстрелов были обнаружены следующие боковые отклонения от точки $(0, 0)$ (в сантиметрах):

Первая серия	Вторая серия
1 — 3,5	1 16
2 — 9	2 12
3 15,5	3 16
4 — 4	4 14,5
5 6,5	5 — 1
6 — 2,5	6 20
	7 21,5
	8 — 4
	9 9

Найти две оценки дисперсии, основанные на каждой серии в отдельности. Затем найти наилучшую оценку, основанную на всех 15 измерениях (см. § 11.10).

63. Проверить, значима ли в задаче 62 разница между оценками дисперсии по разным сериям.

64. Проверить, значима ли в задаче 62 разница между двумя средними арифметическими значениями.

65. Плотность d воздуха при 0°C и давлении 760 мм следующим образом определяется по весу m (в граммах) воздуха в сосуде объема V см³ при температуре $t^\circ\text{C}$ и давлении p мм:

$$d = \frac{m}{V} \frac{760}{p} \left(1 + \frac{t}{273} \right).$$

Найти среднее значение, дисперсию и коэффициент изменчивости величины d , если

$$m = 2,4875 \text{ г} \quad V = 980,3 \text{ см}^3, \quad p = 741,5 \text{ мм}, \quad t = 21,4^\circ\text{C}, \\ m = 0,0023 \text{ г}, \quad \sigma_V = 1,3 \text{ см}^3, \quad \sigma_p = 0,9 \text{ мм}, \quad \sigma_t = 0,3^\circ\text{C}.$$

Задачи к главе 12

66. Показать, что

$$[p:\varepsilon] = [pvv] + E^*PHE, \quad H = AB^{-1}A^*P,$$

где A и B определены в § 12.3, а $E_{n1} = \Lambda - L$ — матрица, образованная истинными поправками (использовать (12.4.4)).

67. След квадратной матрицы A_{nn} определяется как сумма ее диагональных элементов:

$$A = \sum_{i=1}^n a_{ii}.$$

Показать, что след $(AB) = \text{след}(BA)$ для любых двух матриц A и B , для которых оба произведения существуют.

68. Доказать, что

$$\sum_{i=1}^n p_i \sigma^2 \{\bar{l}_i\} = \sigma^2 \text{ след } (PAB^{-1}A^*),$$

где $\sigma^2 \{\bar{l}_i\}$ определяется соотношением (12.8.3). Затем доказать (12.8.5), используя результат задачи 67.

69. Доказать, что

$$\mathfrak{M} \{E^*PHE\} = \sigma^2 \text{ след } H$$

(см. задачу 66). Затем показать, что

$$\mathfrak{M} \{[p\sigma\sigma]\} = (n - m) \sigma^2.$$

70. Для того чтобы найти расстояние x между точками A и B , измеряется база $\mu = AC$ и углы $\alpha = \angle BAC$ и $\beta = \angle BCA$. Пусть m , a и b являются соответственно средними арифметическими значениями результатов измерения величин μ , α и β . Предполагая $\sigma\{m\}$, $\sigma\{a\}$ и $\sigma\{b\}$ известным, определить $\sigma\{x\}$.

71. Для определения сторон прямоугольного треугольника измеряются два катета a и b и гипотенуза c . Предполагая все три измерения равноточными, найти наилучшие оценки для всех трех сторон.

72. При измерении трех сторон и высоты равностороннего треугольника были получены соответственно следующие результаты:

$$\{l_1, l_2, l_3, l_4\} = \{5,01 \quad 5,02 \quad 5,00 \quad 4,30\}.$$

Найти наилучшую оценку стороны, применяя уравнивание с помощью элементов.

73. Решить задачу 72, применяя уравнивание с помощью коррелят.

74. При пробной стрельбе из орудия при возрастающей скорости ветра, перпендикулярного направлению стрельбы, были замечены следующие значения бокового отклонения в 10 выстрелах, производившихся через равные промежутки времени:

(1)	57,2	(6)	59,8
(2)	58,0	(7)	60,4
(3)	58,1	(8)	60,0
(4)	59,1	(9)	60,0
(5)	59,3	(10)	62,2

Предполагая боковое отклонение линейной функцией времени $f(t) = \alpha + \beta t$ найти наилучшие оценки параметров α и β . (Определить также параметры α и β графически).

75. Для четырех точек A, B, C, D , лежащих на одной прямой, измеряются шесть расстояний AB, BC, CD, AC, AD, ED , причем получаются соответственно следующие результаты:

$$\{l_1, \dots, l_6\} = \{3,17 \quad 1,12 \quad 2,25 \quad 4,31 \quad 6,51 \quad 3,36\}.$$

Найти наилучшие оценки этих расстояний, применяя уравнивание с помощью элементов.

76. Решить задачу 75, применяя уравнивание с помощью коррелят.

77. Число телеграмм, отправленных в Германии (в миллионах) за 1925—1934 гг., дано в следующей таблице:

1925	50	1930	34
1926	47	1931	27
1927	48	1932	23
1928	43	1933	22
1929	40	1934	21

Предполагая это число линейной функцией времени

$$f(t) = \alpha + \beta(t - 1925),$$

найти наилучшие оценки параметров α и β . (Определить также параметры α и β графически.)

78. Измерение четырех углов четырехугольника дало следующие результаты:

$$\{l_1, \dots, l_4\} = \{50^\circ 12' 37'' \quad 112^\circ 17' 19'' \quad 120^\circ 47' 26'' \quad 76^\circ 46' 18''\},$$

где l_1, l_2, l_3, l_4 являются соответственно средними арифметическими 3, 4, 2 и 2 равноточных измерений. Найти наилучшие оценки этих углов, применяя уравнивание с помощью элементов.

79. Решить задачу 78, применяя уравнивание с помощью коррелят.

80. Измерение ординат трех точек A, B, C , имеющих соответственно абсциссы 0, 1 и 2, дало следующие результаты:

$$\{1,95 \quad 2,29 \quad 2,14\}.$$

Предполагая, что точки A, B, C лежат на одной окружности

$$(x - \alpha)^2 + y^2 = \rho^2$$

с центром на оси x , найти наилучшие оценки параметров α и ρ .

81. Измерение расстояний от точки $P = \{\xi_1, \xi_2\}$ до точек $(0, 0)$, $(5, 0)$, $(0, 4)$, $(3, 6)$ дало следующие результаты:

$$\{l_1, l_2, l_3, l_4\} = \{3,60 \quad 4,26 \quad 2,20 \quad 3,17\}.$$

Найти наилучшие оценки величин ξ_1 и ξ_2 , применяя уравнивание с помощью элементов.

82. Измерение сторон a и b и углов A и C в треугольнике ABC дало следующие результаты:

$$\{a, b, A, C\} = \{35,1 \text{ см} \quad 61,9 \text{ см} \quad 34^\circ,1 \quad 42^\circ,6\}.$$

Предполагая, что углы измерялись с точностью вдвое большей, чем стороны, найти наилучшие оценки величин a, b, A, C , применяя уравнивание с помощью элементов.

83. Решить задачу 82, применяя уравнивание с помощью коррелят.

84. При измерении расстояний с помощью нивелира с горизонтальными черточками расстояние y является линейной функцией

$$y = \alpha + \beta x$$

части x шеста, наблюдаемой между горизонтальными черточками нивелира. Измерение значений x , соответствующих заданным значениям y , равным 40 м, 60 м, 80 м, 100 м, дало следующие результаты:

$$\{0,335 \text{ м} \quad 0,502 \text{ м} \quad 0,671 \text{ м} \quad 0,841 \text{ м}\}.$$

Найти наилучшие оценки параметров α и β .

85. Измерение скоростей y (в км/сек) и расстояний x (в миллионах парсек) для десяти внегалактических туманностей дало следующие результаты:

x	y	x	y
1,20	630	9,12	4820
1,82	890	10,97	5230
3,31	2350	14,45	7500
7,24	3810	22,91	11800
8,92	4630	36,31	19600

Предполагая y линейной функцией от x

$$y = \alpha + \beta x,$$

найти наилучшие оценки параметров α и β . (Определить также параметры α и β графически.)

86. Давление p газа и его объем v связаны равенством вида

$$pv^\gamma = \text{const.}$$

Для следующих данных

p (кг/см ²)	0,5	1,0	1,5	2,0	2,5	3,0
v (л)	1,62	1,00	0,75	0,62	0,52	0,46

найти наилучшую оценку величины γ , подбирая подходящую прямую для логарифмов величин p и v и считая p независимым переменным.

87. Среднее число y детей в семье при заданной продолжительности брака x (в годах) было в 1920 г. в Норвегии следующим:

x	0—1	5—6	10—11	15—16	20—21	25—26	30—31
y	0,48	2,09	3,26	4,33	5,14	5,63	5,77

Предполагая y многочленом третьей степени от x :

$$y = a_0 + a_1x + a_2x^2 + a_3x^3,$$

найти наилучшие оценки параметров a_0, a_1, a_2, a_3 .

88. Длины λ волн спектральных линий H_α, H_β, \dots водорода подчиняются закономерности, найденной Бальмером:

$$\frac{1}{\lambda} = R \left(\frac{1}{2^2} - \frac{1}{m^2} \right), \quad \text{т. е. } \lambda = \frac{4}{R} \frac{m^2}{m^2 - 4},$$

где $m = 3, 4, \dots$, а R называется константой Ридберга для водорода. По следующим экспериментальным значениям величин λ (в ангстремах, 1 ангстрем = 10^{-8} см)

m	λ
3	6562,79
4	4861,33
5	4340,47
6	4101,74

найти наилучшую оценку постоянной R (в см⁻¹).

ПРИМЕЧАНИЕ РЕДАКТОРА

При сравнении в § 10.6—10.7 и в примере 1 § 11.21 полигона сумм с теоретическим распределением авторы книги ограничиваются только качественной оценкой их близости, не приводя никаких методов оценки допустимых при этом отклонений. В этом дополнении будет кратко изложен разработанный советскими математиками метод, позволяющий оценить эти отклонения.

Примем за меру отклонения полигона сумм $F(t)$ от теоретической интегральной кривой распределения $\Phi(t)$

$$D_n = \max_{-\infty < t < \infty} |\Phi(t) - F(t)|$$

(n — число наблюдений). Как показал А. Н. Колмогоров, в случае непрерывного закона распределения $\Phi(t)$ закон распределения

$$P(D_n \sqrt{Vn} \leq \lambda) = K_n(\lambda)$$

не зависит от $\Phi(t)$ и при $n \rightarrow \infty$ сходится к предельному закону

$$K_n(\lambda) \rightarrow K(\lambda) = \sum_{m=-\infty}^{\infty} (-1)^m e^{-2m^2 \lambda^2}.$$

Используя это предельное соотношение, мы можем написать при больших n следующее приближенное равенство

$$P \left\{ \Phi(t) - \frac{\lambda}{\sqrt{n}} \leq F(t) \leq \Phi(t) + \frac{\lambda}{\sqrt{n}} \right\} \approx K(\lambda).$$

Ниже приведена таблица функции $K(\lambda)$, заимствованная из работы Н. В. Смирнова¹⁾.

Для оценки расхождения между полигоном сумм и функцией распределения $\Phi(t)$ поступают следующим образом. Получив с помощью графика $D_n = \max |\Phi(t) - F(t)|$, вычисляют $\lambda = D_n \sqrt{Vn}$. Затем по таблице находят $P = K(\lambda)$. Используя критерий А (гл. 9), считают, что гипотеза, заключающаяся в том, что случайная величина x подчиняется закону распределения $\Phi(t)$, находится в согласии с опытом, если, например, $P \leq 0,95$, и не находится в согласии с опытом, если

¹⁾ Бюлл. МГУ (А) 2, вып. 2 (1939).

$P > 0,95$. Например, на фиг. 19 (стр. 188) мы имеем $n = 400$, $D_n = 0,05$, откуда $\lambda = D_n \sqrt{n} = 1,00$ и $P = 0,73$, следовательно, отклонения надо признать случайными¹⁾.

Надо заметить, что вышеприведенный метод разработан для сравнения полигона сумм с фиксированным законом распределения $\Phi(t)$. В случае же, когда $\Phi(t)$ зависит от некоторых параметров, значения которых подбираются по результатам выборки (как это имеет место в § 10.6—10.7), этот метод будет давать преувеличенную надежность.

Заметим, наконец, что закон распределения Колмогорова $K(\lambda)$ может служить также для того, чтобы определить, получены ли две независимые выборки объема n_1 и n_2 из одной и той же генеральной совокупности с непрерывным законом распределения. Пусть $F_1(t)$ и $F_2(t)$ два полигона сумм, построенные по независимым выборкам объема n_1 и n_2 . Как показал Н. В. Смирнов, если обе выборки получены из одной и той же генеральной совокупности с непрерывным законом распределения $\Phi(\lambda)$, то закон распределения

$$P \left\{ D_{n_1, n_2} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} < \lambda \right\} = K_{n_1, n_2}(\lambda),$$

где

$$D_{n_1, n_2} = \max_{-\infty < t < \infty} |F_1(t) - F_2(t)|,$$

не зависит от $\Phi(t)$ и при $n_1 \rightarrow \infty$ и $n_2 \rightarrow \infty$ имеет пределом закон $K(\lambda)$.

¹⁾ Этот критерий применим в том случае, когда отдельные наблюдения независимы. Хотя в этом примере отдельные наблюдения зависимы, но, как говорится в примере 1 на стр. 189, эта зависимость настолько слаба при большом числе наблюдений, что ею можно пренебречь.

Таблица функции $K(\lambda)$

λ	$K(\lambda)$	λ	$K(\lambda)$	λ	$K(\lambda)$	λ	$K(\lambda)$
0,32	0,0000	0,71	0,3055	1,10	0,8223	1,49	0,9764
0,33	0,0001	0,72	0,3223	1,11	0,8300	1,50	0,9778
0,34	0,0002	0,73	0,3391	1,12	0,8374	1,51	0,9791
0,35	0,0003	0,74	0,3560	1,13	0,8445	1,52	0,9803
0,36	0,0005	0,75	0,3728	1,14	0,8514	1,53	0,9815
0,37	0,0008	0,76	0,3896	1,15	0,8580	1,54	0,9826
0,38	0,0013	0,77	0,4064	1,16	0,8644	1,55	0,9836
0,39	0,0019	0,78	0,4230	1,17	0,8706	1,56	0,9846
0,40	0,0028	0,79	0,4395	1,18	0,8765	1,57	0,9855
0,41	0,0040	0,80	0,4559	1,19	0,8823	1,58	0,9864
0,42	0,0055	0,81	0,4720	1,20	0,8878	1,59	0,9873
0,43	0,0074	0,82	0,4880	1,21	0,8930	1,60	0,9880
0,44	0,0097	0,83	0,5038	1,22	0,8981	1,61	0,9888
0,45	0,0126	0,84	0,5194	1,23	0,9030	1,62	0,9895
0,46	0,0160	0,85	0,5347	1,24	0,9076	1,63	0,9902
0,47	0,0200	0,86	0,5497	1,25	0,9121	1,64	0,9908
0,48	0,0247	0,87	0,5645	1,26	0,9164	1,65	0,9914
0,49	0,0300	0,88	0,5791	1,27	0,9206	1,66	0,9919
0,50	0,0361	0,89	0,5933	1,28	0,9245	1,67	0,9924
0,51	0,0428	0,90	0,6073	1,29	0,9283	1,68	0,9929
0,52	0,0503	0,91	0,6209	1,30	0,9319	1,69	0,9934
0,53	0,0585	0,92	0,6343	1,31	0,9354	1,70	0,9938
0,54	0,0675	0,93	0,6473	1,32	0,9387	1,71	0,9942
0,55	0,0772	0,94	0,6601	1,33	0,9418	1,72	0,9946
0,56	0,0876	0,95	0,6725	1,34	0,9449	1,73	0,9950
0,57	0,0987	0,96	0,6846	1,35	0,9478	1,74	0,9953
0,58	0,1104	0,97	0,6964	1,36	0,9505	1,75	0,9956
0,59	0,1228	0,98	0,7079	1,37	0,9531	1,76	0,9959
0,60	0,1357	0,99	0,7191	1,38	0,9557	1,77	0,9962
0,61	0,1492	1,00	0,7300	1,39	0,9580	1,78	0,9965
0,62	0,1632	1,01	0,7406	1,40	0,9603	1,79	0,9967
0,63	0,1778	1,02	0,7508	1,41	0,9625	1,80	0,9969
0,64	0,1927	1,03	0,7608	1,42	0,9646	1,81	0,9971
0,65	0,2080	1,04	0,7704	1,43	0,9665	1,82	0,9973
0,66	0,2236	1,05	0,7798	1,44	0,9684	1,83	0,9975
0,67	0,2396	1,06	0,7889	1,45	0,9702	1,84	0,9977
0,68	0,2558	1,07	0,7976	1,46	0,9718	1,85	0,9979
0,69	0,2722	1,08	0,8061	1,47	0,9734	1,86	0,9980
0,70	0,2888	1,09	0,8143	1,48	0,9750	1,87	0,9981

Продолжение табл. функции $K(\lambda)$

λ	$K(\lambda)$	λ	$K(\lambda)$	λ	$K(\lambda)$	λ	$K(\lambda)$
1,88	0,9983	1,94	0,9989	2,00	0,9993	2,15	0,9998
1,89	0,9984	1,95	0,9990	2,02	0,9994	2,20	0,9999
1,90	0,9985	1,96	0,9991	2,04	0,9995	2,25	0,9999
1,91	0,9986	1,97	0,9991	2,06	0,9996	2,30	0,9999
1,92	0,9987	1,98	0,9992	2,08	0,9997	2,31	1,0000
1,93	0,9988	1,99	0,9993	2,10	0,9997		

СПИСОК ЛИТЕРАТУРЫ

Главы 1—9 (Общая теория)

- А. Н. Колмогоров, Основные понятия теории вероятностей, ОНТИ, 1936.
- Г. Крамер, Случайные величины и распределения вероятностей, Москва, 1947.
- Р. Мизес, Вероятность и статистика, 1930.
- Т. Фрай, Теория вероятностей для инженеров, ГТТИ, 1934.
- E. Borel et al., *Traité du calcul des probabilités et de ses applications*, т. 1—4, Paris, 1924—1939.
- E. Czuber, *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung, Statistik und Lebensversicherung*, Leipzig, 1908—1910, т. I—II.
- E. Czuber, *Die Entwicklung der Wahrscheinlichkeitstheorie und ihrer Anwendungen*, *Jahresbericht der deutschen Mathver.*, т. 7, № 2, стр. 1—279, 1899.
- M. Fréchet, Exposé et discussion de quelques recherches récentes sur les fondements du calcul des probabilités. *Théorie des probabilités*, вып. II, стр. 23, *Actualités scientifiques et industrielles*; № 735, Paris, 1938.
- M. Fréchet, The Diverse Definitions of Probability, *Journal of Unified Science*, т. 8, стр. 7, 1939.
- R. von Mises, *Wahrscheinlichkeitsrechnung und ihre Anwendung in der statistik, Fehlertheorie und in der theoretischen Physik*, Leipzig — Wien, 1931.

Стохастические процессы и предельные теоремы

- В. Феллер, К теории стохастических процессов, *Успехи матем. наук*, вып. 5, 1938.
- А. Я. Хинчин, Асимптотические законы теории вероятностей, ОНТИ, 1936.
- С. Чандрасекар, Стохастические проблемы в физике и астрономии, Москва, 1947.
- N. Arley, *On the Theory of Stochastic Processes and Their Application to the Theory of Cosmic Radiation*, New York, 1948.
- M. S. Bartlett, *Stochastic Processes*, Chapel Hill N. C., 1947.
- P. Lévy, *Théorie de l'addition des variables aléatoires*, Paris, 1937.
- O. Lundberg, *On Random Processes and Their Application to Sickness and Accident Statistics*, Uppsala, 1940.
- J. E. Moyal, *Stochastic Processes and Statistical Physics*, *J. Roy. Stat. Soc.*, т. B 11, 1949.

Глава 9 (Отношение теории вероятностей к опыту и ее практическое значение)

- E. Borel, Valeur pratique et philosophie des probabilités, *Traité du calcul des probabilités*, т. 4, вып. 3, Paris, 1939.
- C. Cranz, Lehrbuch der Ballistik, т. I, гл. 10, стр. 385—457, Berlin, 1925.
- R. Fürth, Einführung in die theoretische Physik, Vienna, 1936.
- T. J. Hayes, Elements of Ordnance, New York, 1938.
- J. Haag, Applications du calcul des probabilités au tir, *Traité du calcul des probabilités*, т. 4, вып. 1, Paris, 1926.

Глава 10 (Приложение теории вероятностей к статистике)

- Г. Крамер, Математические методы статистики, Москва, 1948.
- O. L. Davies, editor, Statistical Methods in Research and Production, London, 1947.
- R. A. Fisher, Statistical Methods for Research Workers, 7-е изд., London, 1938.
- R. A. Fisher, On the Mathematical Foundations of Theoretical Statistics, *Phil. Trans. of the Roy. Soc. of London*, т. A 222, стр. 309, 1921.
- R. A. Fisher, Theory of Statistical Estimation, *Proc. of the Cambridge Phil. Soc.*, т. 22, стр. 700, 1925.
- R. A. Fisher and F. Yates, Statistical Tables, London, 1938.
- J. C. Kapteyn and M. J. van Uven, Skew Frequency-Curves in Biology and Statistics, Groningen, 1916.
- M. G. Kendall, The Advanced Theory of Statistics, т. I и II, London, 1943—1946.
- G. U. Yule and M. G. Kendall, An Introduction to the Theory of Statistics, 12-е изд., London, 1940.

Главы 11—12 (Приложение теории вероятностей к теории ошибок и теории уравнивания)

- Уиттекер и Робинсон, Математическая обработка результатов наблюдений, Москва, 1935.
- N. Arley, On the Distribution of Relative Errors from a Normal Population of Errors, *Danske Vid. Selsk. Mat.-fys. Medd.*, т. XVIII, № 3, 1940.
- E. Czuber, Theorie der Beobachtungsfehler, Leipzig, 1891.
- F. Helmert, Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate, Leipzig, 1907.
- C. Runge and H. König, Vorlesungen über numerisches Rechnen, Berlin, 1924.

§ 12.15 (Анализ регрессии)

- M. Ezekiel, Methods of Correlation Analysis, 2-е изд., New York, 1941.
- R. A. Fisher, Statistical Methods for Research Workers, 7-е изд., гл. V, стр. 134—176, London, 1938.
- M. G. Kendall, The Advanced Theory of Statistics, т. II, гл. 22, London, 1946.
- G. U. Yule and M. G. Kendall, An Introduction to the Theory of Statistics, 12-е изд., гл. 17, London, 1940.

УКАЗАТЕЛЬ

- Амплитуды вероятности 63
Анализ регрессии 214
— статистический 126
- Величина нормированная 75
— ограниченная 66
Веса 88, 193
Вероятность 17
— абсолютная 26
— апостериорная 17
— априорная 17
— условная 26
Взвешенная сумма квадратов ошибок 194
Выборка 127
- Гистограмма 135
— накопленная 140
Группировка 134, 178
- Дециль 74
Дисперсия 74
Дифференциал вероятности 43
- Задача Бернулли 32
Закон больших чисел 15, 115
— Максвелла-Больцмана 48, 70, 97
Значимость 126, 180, 181, 182, 185
- Измерения избыточные 195
— косвенные 165
— непосредственные 165
Интеграл ошибок 93
Интервал группировки 134
Истинное значение 165, 187
- Квантиль 74
Квартиль 74
Корреляты 207
Корреляционная таблица 179
Корреляция 184
Коэффициент изменчивости 75
— корреляции 81, 177, 184
- Коэффициент регрессии 80
Кривая ошибок 92
— регрессии 79
- Матрица весов 197
— вторых моментов 88
— корреляты 207
— корреляционная 89
Медиана 65
Мера точности 92
Метод максимума правдоподобия 152
— моментов 152, 158
— наименьших квадратов 167, 194
— выпрямленной диаграммы 142
Мода 65
Моменты 71
— абсолютные 71
— смешанные, второго порядка 81
— факториальные 71
— центральные 71, 81
- Неравенство Чебышева 113
Нормирующий множитель 44
- Отклонение, вероятное 97
— среднее 97
— — квадратическое 74
— стандартное 74
Оценки 130
— асимптотически эффективные 153
— достаточные 157
— максимального правдоподобия 153
— несмещенные 148
— параметров 130
— совместно-достаточные 157
— совместно-эффективные 155
— состоятельные 148
— эффективные 149
- Ошибка 163
— грубая 163
— истинная 166

- Ошибка кажущаяся 166
 — систематическая 164
 — средняя 169
 — квадратическая 169
 — стандартная 169
 — случайная 164
- Парадокс Бертраана 22
 Параметры 27
 Переопределенная система 195
 Плотность вероятности 42
 Плотность потока вероятности 58]
 Полигон сумм 140
 — частот 135
 Полная аддитивность 30
 Полуширота 74
 Поправка 166
 — кажущаяся 193
 Последовательные разности 170
 Пределы доверительные 96, 170
 Прямые средней квадратической регрессии 80
- Распределение биномиальное 39
 — выборки 133, 135
 — Гельмерта 106
 — каноническое Гиббса 60
 — Коши 44
 — Кэптейна 131
 — Лапласа 44
 — логарифмически-нормальное 132
 — несобственное 39
 — нормальное 44
 — отрицательно-биномиальное 41
 — Паскаля 40
 — Пирсона 132
 — Поля 41
 — прямоугольное 43
 — Пуассона 39
 — равномерное 43
 — Стьюдента 109
 — условное 56
 — частное 51, 54
 — эмпирическое 133
 — q 106
 — r 110
 — t 109
 — w^2 112
- Регрессия 79
- Семинварианты 79
 Случайные явления 12
 События достоверные 33
 — невозможные 24
 — независимые 29
- События несовместимые 29
 — противоположные или допол-
 нительные 31
- Спектр 38, 42
 Среднее арифметическое 88
 — взвешенное 88
 — значение 65
- Статистика 148
 Статистическое описание 11
 Степень свободы 106, 168
 Стохастические процессы 57
 Сходимость по вероятности 113
- Теорема Байеса 33
 — Бернулли 115, 124, 133, 136, 140
 — Муавра 118
 — Хинчина 116, 158
 — центральная предельная 119,
 131, 159
 — эргодическая 62
- Теория ошибок 161
- Уравнения коррелят 208
 — нормальные 198, 208
 — правдоподобия 153
 — условные 196
 — фундаментальные 192
- Уравнивание с помощью коррелят 206
 — — — элементов 196
- Формула Лапласа 117, 146
 — Пуассона 122
 — Стирлинга 107, 220
- Функция Дирака 45
 — правдоподобия 152
 — производящая 71
 — — моментов 72
 — распределения, дискретная 38
 — — совместная 49
 — свободная 201
 — характеристическая 72
- Характеристики расположения 65
 — рассеяния 74
- Частота абсолютная 15
 — относительная 15
- Широта 74
 — семи-интерквартильная 74
 — семи-интерквартильная 74
- Элементы 194
 Эффективность 149

ОГЛАВЛЕНИЕ

Предисловие к русскому переводу	5
Из предисловия авторов к первому датскому изданию	9
1. Понятие вероятности	11
2. Основания теории вероятностей	23
3. Элементарные теоремы	29
4. Случайные величины и функции распределения	35
5. Среднее значение и дисперсия	65
6. Среднее значение и дисперсия суммы, произведения и других функций случайных величин	83
7. Нормальное распределение	92
8. Пределные теоремы	113
9. Отношение теории вероятностей к опыту и ее практическое значение	123
10. Приложение теории вероятностей к статистике	129
11. Приложение теории вероятностей к теории ошибок	162
12. Приложение теории вероятностей к теории уравнивания	192
Приложение 1. $n!$	220
Приложение 2. Теория матриц	222
Таблица I. Нормальное распределение (§ 7.1)	226
Таблица II. Нормальное распределение (§ 7.4)	226
Таблица III. t -распределение (§ 7.9)	227
Таблица IV. r -распределение (§ 7.10)	227
Таблица V. ω^2 -распределение (§ 7.11)	228
Задачи	229
Примечание редактора	239
Список литературы	243
Указатель	245

Редактор Б. А. Севастьянов
Технический редактор Н. Печникова
Корректор Р. Новик

Сдано в производство 23/1 1951 г.
Подписано к печати 24/III 1951 г.
A02498. Бумага $60 \times 92\frac{1}{16} = 7,8$ бум. л.
15,3 печ. л. Уч.-издат. л. 16,4.
Изд. № 1,993 Цена 15 руб. Зак. № 2279

Первая Образцовая типография имени
А. А. Жданова Главполиграфиздата
при Совете Министров СССР. Москва,
Валовая, 28.

Отпечатано в типографии
Тимирязевской с.-х. академии.

