

Теория вероятностей

■
В.Н.ТУТУБАЛИН

В. Н. ТУТУБАЛИН



Теория вероятностей

*КРАТКИЙ КУРС
И НАУЧНО-МЕТОДИЧЕСКИЕ ЗАМЕЧАНИЯ*

*ИЗДАТЕЛЬСТВО
МОСКОВСКОГО
УНИВЕРСИТЕТА
1972*

Книга состоит из двух частей. Первая часть представляет собой краткий курс теории вероятностей с элементами математической статистики, предназначенный для студентов естественных специальностей университетов и пединститутов. Интенсивное использование аппарата математического анализа и линейной алгебры позволяет при небольшом объеме охватить значительный материал, включая доказательство центральной предельной теоремы и теорию метода наименьших квадратов. Вторая часть книги написана по материалам работы со слушателями факультета повышения квалификации преподавателей вузов. Она содержит научные и методические замечания, которые полезно иметь в виду преподавателю теории вероятностей при изложении материала, охваченного первой частью книги.

Учебное пособие предназначено студентам указанных специальностей, а также будет полезно преподавателям и всем интересующимся теорией вероятностей и ее приложениями.

Печатается по постановлению
Редакционно-издательского совета
Московского университета

Предисловие

Первая часть данной книги возникла на основе курса лекций, которые автор читал для студентов-механиков механико-математического факультета МГУ. Этот курс рассчитан на один семестр, следовательно, должен быть «кратким». С другой стороны, в курс необходимо включить кроме чисто математической части обсуждение вопросов применимости теории вероятностей и элементы математической статистики. Создается затруднение, выход из которого облегчается тем, что большая часть понятий и теорем элементарной теории вероятностей является, с современной точки зрения, конкретизацией более общих математических понятий. Используя высокую математическую культуру, которую полагается иметь студентам мехмата, удастся охватить довольно большой материал, включая интеграл Лебега, доказательство центральной предельной теоремы и общую теорию метода наименьших квадратов.

Собственно математическая часть курса построена по принципу выведения наибольшего числа следствий из наименьшего числа общих теорем. Поэтому возможно, что первая часть книги будет удобным пособием для подготовки к экзаменам. Для ее понимания требуется знание хорошего курса математического анализа и элементов линейной алгебры. Таким образом, книгой могут пользоваться студенты естественнонаучных специальностей с усиленной математической подготовкой.

Вторая часть книги создана на основе опыта преподавания на факультете повышения квалификации преподавателей вузов. Она имеет форму замечаний к соответствующим параграфам первой части. Ее цель — ответить (насколько возможно) на вопрос о том, что должен иметь в виду преподаватель теории вероятностей, который собирается изложить предмет в объеме первой части книги. Замечания к § 1—7 имеют в основном методический характер (включая указания на некоторые «подводные камни», т. е. логические пробелы, которые иногда встречаются). Замечания к § 8—12 по большей части излагают связанный с первой частью научный материал.

Это объясняется тем, что элементарные понятия теории вероятностей изложены в § 1—7 первой части книги по существу исчерпывающим образом, чего совсем нельзя сказать о более серьезном материале § 8—12.

Итак, первая часть книги предназначена для студентов, вторая часть — для преподавателей. Автор надеется, однако, что книга представит определенный интерес также для знакомых с основами современной математики инженеров и естествоиспытателей. Будучи математиком по профессии, автор попытался встать на точку зрения естествоиспытателя при изложении основ теории вероятностей. Понятия и теоремы, приведенные в книге, оцениваются с этой точки зрения. Критически рассматривается вопрос об области применимости изложенного материала. Не следует думать, что по данному вопросу автор имеет какое-то свое особое мнение. В книге излагаются взгляды, принятые в настоящее время большинством специалистов, но только в соответствии с требованиями момента слегка подчеркивается критическая сторона.

Первая часть книги не предполагает знакомства с литературой по теории вероятностей. Для понимания второй части книги формально достаточно владения первой частью. Однако читатель должен иметь в виду, что данная книга отнюдь не заменяет более подробных учебников Гнеденко [8] и Феллера [22]. Тот, кто интересуется приложениями теории вероятностей, обычно не читает учебник подряд, а разыскивает в нем пример, похожий на то, что его интересует. Следовательно, подробный учебник по теории вероятностей должен содержать более или менее универсальный набор примеров. В данной книге примеры есть, их даже много. Однако их набор далеко не универсален.

Определения, леммы, теоремы и рисунки нумеруются двумя числами, из которых первое указывает номер параграфа. Во второй части номер снабжается штрихом, например теорема 8.1' находится в замечаниях к § 8.

§ 1

ДИСКРЕТНОЕ ПРОСТРАНСТВО ЭЛЕМЕНТАРНЫХ СОБЫТИЙ

1.1. Введение. Теория вероятностей изучает математические модели случайных экспериментов, т. е. таких экспериментов, исход которых не вполне однозначно определяется условиями опыта. Любимым примером случайного эксперимента является бросание монеты: монета может упасть либо гербом кверху, либо цифрой. В дальнейшем мы увидим, что практически можно воспроизвести вероятностную структуру любого случайного эксперимента, бросая монету достаточно много раз.

Если бы теория вероятностей позволяла сказать нечто содержательное об исходе любого эксперимента, про который известно только лишь, что его исход неоднозначен, то она была бы, конечно, наукой наук. На самом деле ее роль гораздо скромнее. Теория вероятностей имеет дело не с любыми случайными экспериментами, а лишь с экспериментами, обладающими свойством статистической устойчивости, или устойчивости частот.

Это свойство описывается следующим образом. Обозначим через A один из возможных исходов случайного эксперимента. Повторим этот эксперимент n раз и обозначим через μ_A число наступлений исхода A в этих n экспериментах. Тогда отношение μ_A/n называется частотой события¹ A , а свойство устойчивости частот заключается в том, что при большом n частота события A лишь слегка колеблется (при изменении n) около некоторого числа. Далее, если сделать несколько серий экспериментов, то частоты $\mu_A^{(i)}/n_i$ (где через n_i обозначено число экспериментов в i -ой серии, а через $\mu_A^{(i)}$ число наступлений события A в этой серии) будут близки между собой, лишь бы только числа n_i были все достаточно велики. Например, К. Пирсон бросал монету 24 000

¹ Слова «событие» и «исход эксперимента» означают одно и то же.

раз и при этом герб выпал 12 012 раз. Отношение $\frac{12\,012}{24\,000}$, очевидно, очень близко к $\frac{1}{2}$.

В книге Ф. Мостеллера, Р. Рурке и Дж. Томаса [21] сообщается следующий результат десяти серий по 1000 бросаний монеты: число выпадений герба равнялось соответственно 502, 518, 497, 529, 504, 476, 507, 528, 504, 529 (см. [21], стр. 91).

Частоты выпадений герба группируются, следовательно, около $\frac{1}{2}$. Число, около которого колеблется частота события A , называется *вероятностью события A* и обозначается через $P(A)$. Например, нет сомнений в том, что вероятность выпадения герба при бросании монеты равна $\frac{1}{2}$.

Сделанное описание свойства устойчивости частот и определение вероятности вряд ли кому-либо покажутся удовлетворительными. Нельзя избежать вопроса о том, насколько сильно при данном конечном n частота μ_A/n события A может отличаться от вероятности $P(A)$? Вообще, каким образом следует производить проверку устойчивости частот, решая вопрос о применимости методов теории вероятностей к данному конкретному явлению? Следует сказать, что в настоящее время нет достаточно общего научного ответа на эти вопросы. На первый вопрос — о допустимом различии μ_A/n и $P(A)$ — мы дадим впоследствии ответ, но при дополнительном условии независимости результатов каждого из n экспериментов (о независимости будет говориться дальше). Для проверки независимости есть частичные критерии, которые, однако, действуют также лишь при соблюдении некоторых дополнительных условий. Еще более сложен вопрос об устойчивости частот в разных сериях экспериментов. Здесь трудность состоит прежде всего в том, как выделять эти серии. Например, мы могли бы договориться сделать подряд большое число экспериментов, относя к первой серии эксперименты с простыми номерами, ко второй серии — эксперименты с четными номерами, к третьей серии — эксперименты с номерами, делящимися на 3, и т. д. Очевидно, что способов выделять серии слишком много для того, чтобы их все можно было перепробовать. Кроме того, если пробовать все способы, то мы непременно должны рассмотреть и такой, когда в одну серию попадут те опыты, в которых событие A наступило, а в другую — те опыты, в которых оно не наступило. В первой серии частота события A будет равна 1, а во второй — 0, так что эти частоты будут резко различны. Указанная трудность является существенной, и лишь в самые последние годы намечены научные способы ее преодоления. Однако эти способы не доведены до практических рекомендаций.

Из сказанного ясно, что вопрос о применимости вероятностных методов в каждом отдельном случае решается на интуитивном уровне (интуиция, конечно, основана на личном и общенаучном опыте). Научная добросовестность требует от каждого исследователя применения доступных методов проверки статисти-

ческой устойчивости¹, но наличие ее редко можно вполне гарантировать.

Все мыслимые эксперименты можно разделить на три группы. К первой группе относятся хорошие эксперименты, в которых обеспечивается полная устойчивость исхода опыта. Ко второй группе относятся эксперименты похуже, где полной устойчивости нет, но есть статистическая устойчивость. К третьей группе относятся совсем плохие эксперименты, когда нет и статистической устойчивости. В первой группе все ясно без теории вероятностей. В третьей группе она бесполезна. Вторая группа составляет настоящую сферу применения теории вероятностей, но мы вряд ли когда-нибудь можем быть вполне уверены, что интересующий нас эксперимент относится ко второй, а не к третьей группе.

Прогресс физики привел, однако, к тому, что теперь более чем когда-либо прежде верят в важность теории вероятностей: по существующим представлениям эксперименты на квантово-механическом уровне, выясняющие наиболее фундаментальные законы природы, относятся именно ко второй группе.

1.2. Вероятностное пространство. В современной математической теории вероятностей предпочитают оставлять проблему статистической устойчивости¹, в стороне и рассматривают математическую модель, в которой отражены все возможные исходы эксперимента и считаются известными связанные с этим экспериментом вероятности. Наиболее простой вид эта модель имеет в том случае, когда множество² возможных исходов эксперимента конечно или счетно³. Этот случай называется дискретным; его мы и рассмотрим.

Определение 1.1. *Пространством элементарных событий* называется любое конечное или счетное множество.

В дальнейшем пространство элементарных событий обозначается буквой Ω , его элементы — буквами $\omega_1, \omega_2, \dots$; тот факт, что элемент ω_i входит в множество Ω , записывается в виде $\omega_i \in \Omega$, а тот факт, что множество Ω состоит из элементов $\omega_1, \omega_2, \dots$ и только из них, записывается в виде $\Omega = \{\omega_1, \omega_2, \dots\}$ или $\Omega = \{\omega_i, i=1, 2, \dots\}$.

Определение 1.2. Каждому $\omega_i \in \Omega$ отвечает число $P(\omega_i)$, называемое вероятностью элементарного исхода ω_i .

Аксиомы 1.1. $0 \leq P(\omega_i) \leq 1$,

$$1.2. \sum_{\omega_i \in \Omega} P(\omega_i) = 1.$$

Определение 1.3. Любое подмножество⁴ множества элементарных событий называется *событием*.

¹ Вместо «статистическая устойчивость» говорят также «статистическая однородность».

² Слова «множество» и «совокупность» означают одно и то же.

³ Счетным называется множество, элементы которого можно поставить во взаимно однозначное соответствие натуральным числам.

⁴ Слова «подмножество» и «часть» означают одно и то же.

В дальнейшем утверждение « A есть подмножество Ω » записывается в виде $A \subseteq \Omega$. В том случае, когда $A \subseteq \Omega$, но $A \neq \Omega$, пишут $A \subset \Omega$.

Определение 1. 4. Вероятностью $P(A)$ любого события A называется сумма вероятностей элементарных событий, входящих в событие A , иначе говоря:

$$P(A) = \sum_{\omega_i \in A} P(\omega_i).$$

Замечание. Теория вероятностей возникла еще в XVII в., и язык, на котором выражалось содержание этой науки, первоначально сильно отличался от принятого сейчас языка, основные термины которого введены в только что изложенных определениях и аксиомах. Современный язык есть, очевидно, язык теории множеств. Однако большинство задач, которые читатель найдет в современных учебниках и задачниках по теории вероятностей, сформулированы на старом языке. Это не случайно: только по задачам изучающий теорию вероятностей знакомится с теми типичными ситуациями, когда ее можно практически применять, а теоретико-множественный язык исключил бы всякое описание практических ситуаций.

Следовательно, при изучении теории вероятностей надо прежде всего научиться переводить условия задач с традиционного языка на современный. При этом возможны различные результаты. Ситуация здесь совершенно такая же, как та, с которой сталкивается школьник при переводе задач на язык систем уравнений. Пусть, например, нужно решить задачу: «У меня в левом кармане вдвое больше денег, чем в правом, а всего 1 руб. 80 коп. Сколько денег в каждом кармане?» Тогда одна система уравнений будет:

$$\begin{cases} x = 2y, \\ x + y = 180, \end{cases}$$

где x копеек в левом кармане, y копеек в правом. Другая система состоит из одного уравнения

$$x + 2x = 180,$$

где x копеек в правом кармане, $2x$ — в левом.

С математической точки зрения это две различные системы уравнений. Тем не менее каждая из них правильно решает одну и ту же задачу, хотя этого и нельзя доказать в рамках математического доказательства.

Точно так же при решении одной задачи по теории вероятностей возможно введение разных пространств элементарных событий и разных вероятностей P , причем обычно не следует искать им обоснование на математическом уровне строгости.

1.3. Правила перевода. Переводу учатся на примерах, но правила также играют некоторую роль.

Правило 1. Пространство элементарных событий Ω есть совокупность всех мыслимых исходов опыта; при этом считается, что исходы регистрируются возможно более подробно.

Правило 2. В том случае, когда из каких-либо соображений симметрии ясно, что все элементарные исходы равновероятны, т. е. $P(\omega_i)$ не зависит от ω_i , множество элементарных событий Ω будет конечным (в силу аксиом 1.1 и 1.2), а $P(\omega_i) = \frac{1}{N}$, где N — число элементов Ω . Если событие A содержит M элементарных событий, то $P(A) = \frac{M}{N}$.

З а м е ч а н и е. Если правило 2 применимо, то говорят, что имеется задача «на классическую вероятность». Элементарные события¹, входящие в событие A , называются «благоприятными», и формула $P(A) = \frac{M}{N}$ на традиционном языке выражается так: «вероятность события равна отношению числа благоприятных исходов к числу всех возможных».

Рассмотрим примеры.

Пример 1.1. Бросают две игральные кости. Чему равна вероятность того, что сумма очков, выпавших на обеих костях, не превзойдет 5?

Решение. Пусть n_1 очков выпало на первой кости, n_2 — на второй. Пространство элементарных событий есть множество пар (n_1, n_2) :

$$\Omega = \{(n_1, n_2): n_1, n_2 = 1, 2, 3, 4, 5, 6\}.$$

Интересующее нас событие A имеет вид

$$A = \{(n_1, n_2): n_1, n_2 = 1, 2, 3, 4; n_1 + n_2 \leq 5\}.$$

Множество Ω содержит 36 элементов, множество A — 10 элементов. В силу наших представлений о правильных игровых костях применимо правило 2. Значит $P(A) = \frac{10}{36} = \frac{5}{18}$.

Пример 1.2. Среди 25 экзаменационных билетов имеется 5 «счастливых» и 20 «несчастливых». Студенты подходят за билетами один за другим по очереди. У кого больше вероятность вытащить счастливый билет: у того, кто подошел за билетами первым, или у того, кто подошел вторым?

Решение. Пусть «счастливые» билеты имеют номера 1, 2, 3, 4, 5. Обозначим через i_1 номер билета, взятого первым студентом, через i_2 — номер билета, взятого вторым студентом. Тогда

$$\Omega = \{(i_1, i_2): i_1, i_2 = 1, \dots, 25, i_1 \neq i_2\}.$$

¹ Слова «элементарное событие» и «элементарный исход» означают одно и то же.

Кроме того, все элементарные события равновероятны. Событие A : «первый студент взял «счастливый» билет» имеет вид

$$A = \{(i_1, i_2) : i_1 = 1, 2, 3, 4, 5; i_2 = 1, \dots, 25, i_1 \neq i_2\}.$$

Событие B : «второй студент взял «счастливый» билет» имеет вид

$$B = \{(i_1, i_2) : i_1 = 1, \dots, 25; i_2 = 1, 2, 3, 4, 5, i_1 \neq i_2\}.$$

Каждое из событий A и B содержит по 120 элементов, Ω содержит 600 элементов. Следовательно, $P(A) = P(B) = \frac{1}{5}$.

Несколько позже мы рассмотрим решение этой задачи с помощью другого Ω и других правил перевода.

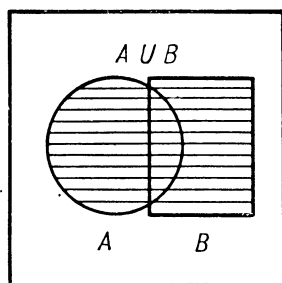


Рис. 1.1

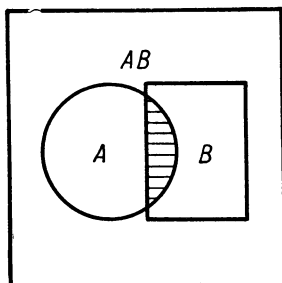


Рис. 1.2

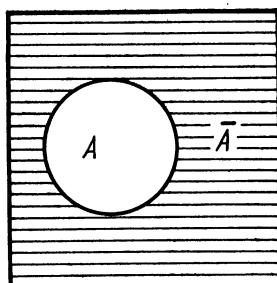


Рис. 1.3

1.4. Операции над событиями. Определение 1.5. Суммой $A \cup B$ событий A и B называется их теоретико-множественное объединение, т. е. событие, состоящее из таких элементарных исходов, которые входят в событие A или в событие B , или в оба вместе.

Определение 1.6. Произведением AB событий A и B называется их теоретико-множественное пересечение, т. е. событие, состоящее из таких элементарных исходов, которые входят одновременно в оба события A и B (AB обозначается также через $A \cap B$).

Определение 1.7. Противоположным событием \bar{A} для события A называется теоретико-множественное дополнение A до Ω , т. е. событие, состоящее из таких элементарных исходов, которые не входят в событие A .

Уже Аристотель по существу пользовался для изучения теоретико-множественных операций моделью, в которой множества изображаются в виде фигур на плоскости. При некотором навыке использование этой модели никогда не приводит к ошибкам (т. е. к результатам, неверным для произвольных множеств). На рис. 1.1, 1.2 и 1.3 изображены $A \cup B$, AB и \bar{A} .

В том случае, когда $AB = \emptyset$ (где \emptyset обозначает пустое множество¹ или множество, не содержащее элементов), мы будем писать $A+B$ вместо $A \cup B$ (в отличие от других наших обозначений, это не является вполне общепринятым). Из определения 1.1, очевидно, вытекает

Теорема 1.1. $P(A+B) = P(A) + P(B)$, если $AB = \emptyset$.

В случае общего (недискретного) пространства элементарных событий эта теорема принимается в качестве аксиомы.

В общем случае (не обязательно $AB = \emptyset$) имеем:

Теорема 1.2. $P(A \cup B) = P(A) + P(B) - P(AB)$.

Действительно,

$$P(A \cup B) = \sum_{\omega_i \in A \cup B} P(\omega_i),$$

но в сумме $P(A) + P(B)$ вероятности $P(\omega_i)$ для $\omega_i \in AB$ учитываются дважды. Поэтому, если из $P(A) + P(B)$ вычесть $P(AB)$, то получится как раз $P(A \cup B)$.

Дадим теперь образец выражения на традиционном языке. Говорят, что «событие A наступило», если опыт закончился одним из элементарных исходов, входящих в событие A . Следовательно, $A \cup B$ наступает, если наступает хотя бы одно из событий A или B ; AB наступает, если наступают оба события A и B ; \bar{A} наступает, если A не наступает.

Читателю рекомендуется доказать самостоятельно следующие свойства операций над событиями:

1. $(A \cup B)^c = A^c \cap B^c$.
2. $AB \cup C = (A \cup C)(B \cup C)$.
3. $\overline{A \cup B} = \bar{A} \bar{B}$.
4. $\overline{AB} = \bar{A} \cup \bar{B}$.
5. $A(B_1 + \dots + B_n) = AB_1 + \dots + AB_n$.

§ 2

УСЛОВНАЯ ВЕРОЯТНОСТЬ. НЕЗАВИСИМОСТЬ. ОСНОВНЫЕ ФОРМУЛЫ

2.1. Условная вероятность. Легко заметить, что хотя в аксиомах и определениях предыдущего параграфа частотная интерпретация вероятности явно не участвует, но тем не менее все эти аксиомы и определения мотивированы теми свойствами, которые должна иметь вероятность, определяемая по частоте.

¹ Пустое множество называется еще невозможным событием; очевидно, надо считать, что $P(\emptyset) = 0$.

Например, если событие A состоит из некоторых элементарных событий ω_i , то частота наступления события A в серии экспериментов равна, очевидно, сумме частот элементарных событий $\omega_i \in A$. Отсюда возникает определение 1.4:

$$P(A) = \sum_{\omega_i \in A} P(\omega_i).$$

Если $A = \Omega$, то частота этого события, очевидно, равна 1 (событие Ω называется достоверным). Поэтому вводится аксиома 1.2:

$$\sum_{\omega_i \in \Omega} P(\omega_i) = 1.$$

Применим теперь соображения, связанные с частотами, для того чтобы выяснить, каким должно быть математическое определение условной вероятности. Пусть в результате эксперимента могут наступить два события A и B . Условной частотой события A при условии, что событие B наступило, называется частота события A , вычисленная не по совокупности всех экспериментов, а лишь по совокупности тех экспериментов, в которых наступило событие B . Иными словами, если n — число всех экспериментов, μ_B — число наступлений события B , а μ_{AB} — число наступления события AB (или число таких экспериментов, в которых наступило и событие B , и событие A), то условная частота есть

$$\frac{\mu_{AB}}{\mu_B} = \frac{\frac{\mu_{AB}}{n}}{\frac{\mu_B}{n}}.$$

Левая часть этого выражения интерпретируется как приближенное значение условной вероятности $P(A/B)$ наступления события A при условии, что B наступило, отношение μ_{AB}/n приближенно равно $P(AB)$, а отношение $\frac{\mu_B}{n} \approx P(B)$. Изложенные соображения мотивируют введение следующего математического определения.

Определение 2.1. $P(A/B) = \frac{P(AB)}{P(B)}$.

Разумеется, это определение имеет смысл только в случае $P(B) \neq 0$.

Умножением на $P(B)$ обеих частей определения условной вероятности получается

Теорема 2.1. $P(AB) = P(B)P(A/B)$, иными словами, вероятность совместного наступления двух событий равна произведению вероятности одного из них на условную вероятность другого при условии, что первое наступило.

Почему столь очевидное утверждение мы считаем нужным выделить в виде теоремы? Дело в том, что роль этого математически тривиального утверждения вовсе не математическая. На самом деле оно играет роль одного из правил перевода на язык пространства элементарных событий. Оно применяется в тех случаях, когда нужно определить вероятность на множестве элементарных событий, причем из содержания задачи ясны значения условных вероятностей.

Пример 2.1. Рассмотрим в качестве примера другое решение задачи 1.2 предыдущего параграфа. В предыдущем параграфе мы считали, что наблюдению доступны номера билетов, взятых первым и вторым студентами. Тогда мы вправе взять в качестве множества элементарных событий множество пар $\{(i_1, i_2): i_1 \neq i_2, i_1, i_2 = 1, 2, \dots, 25\}$ номеров билетов. Можно, однако, предположить, что мы лишены возможности наблюдать номера билетов, а можем лишь определить, какие билеты попались — счастливые или нет (например, допустим, что мы наблюдаем только выражение лица студента, взявшего билет).

Если обозначить через $1_c, 1_n, 2_c, 2_n$ события, состоящие соответственно в том, что

1-й студент взял счастливый билет;

1-й » » несчастливый » ;

2-й » » счастливый » ;

2-й » » несчастливый » ;

то пространство элементарных событий состоит из четырех событий

$$1_c 2_c, \quad 1_c 2_n, \quad 1_n 2_c, \quad 1_n 2_n.$$

Остается ввести вероятности на множестве этих элементарных событий. Напомним еще раз, что с логической точки зрения утверждение, приписывающее вероятности элементарным событиям, является не теоремой (которую можно доказать), а определением (которого нельзя доказать, но разумность которого должна быть мотивирована). Воспользуемся теоремой 2.1. Если бы вероятности элементарных событий были уже введены, то должно было бы выполняться соотношение

$$P(1_c 2_c) = P(1_c) P(2_c/1_c). \quad (2.1)$$

С другой стороны, ясно, что $P(1_c) = \frac{1}{5}$, поскольку первый студент выбирает один из 25 билетов, из которых 5 счастливых. Далее, ясно, что $P(2_c/1_c) = \frac{4}{24}$, поскольку второму приходится выбирать из 24 билетов, среди которых (при условии, что первый вытащил счастливый билет) только 4 счастливых. Определим теперь $P(1_c 2_c)$ так, чтобы равенство (2.1) выполнялось:

$$P(1_c 2_c) = P(1_c) P(2_c/1_c) = \frac{1}{5} \cdot \frac{4}{24}$$

и аналогично

$$P(1_c 2_n) = P(1_c) P(2_n/1_c) = \frac{1}{5} \cdot \frac{20}{24},$$

$$P(1_n 2_c) = P(1_n) P(2_c/1_n) = \frac{20}{25} \cdot \frac{5}{24},$$

$$P(1_n 2_n) = P(1_n) P(2_n/1_n) = \frac{20}{25} \cdot \frac{19}{24}.$$

После того как вероятности элементарных событий определены, применяем определение 1.4 и получаем

$$P(2_c) = P(1_c 2_c) + P(1_n 2_c) = \frac{120}{600} = \frac{1}{5},$$

т. е. получается тот же самый ответ, что и в § 1.

З а м е ч а н и е. Способ, примененный в § 1, позволяет аналогично подсчитать вероятность того, что 3-й (4-й, 5-й и т. д.) студент получит счастливый билет. Все эти вероятности равны $\frac{1}{5}$. Только что изложенный способ приводит к более сложным вычислениям. Таким образом, иногда лучше ввести пространство из большего числа элементарных событий, но так, чтобы эти события были равновероятны.

2.2. Независимость. Естественно сказать, что событие A не зависит от события B , если условная вероятность события A при условии B равна безусловной вероятности события A

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

С помощью определения 2.1 получаем, что в этом случае $P(AB) = P(A) \cdot P(B)$. Следовательно, если A не зависит от B , то и B не зависит от A , поскольку последнее равенство симметрично относительно A и B . Поэтому дается следующее

Определение 2.2. События A и B называются *независимыми*, если выполнено равенство

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

Смысл этого определения заключается в том, что если произошло одно из независимых событий, то это никак не влияет на вероятность другого события. Но в таком случае, если первое событие не произошло, то это также не должно влиять на вероятность второго. Действительно, имеет место

Теорема 2.2. Если события A и B независимы, то события \bar{A} и B также независимы.

Доказательство. $P(\bar{A}\bar{B}) = P(B) - P(AB) = P(B) - P(A) \times P(B) = (1 - P(A)) P(B) = P(\bar{A}) P(B)$.

Следствие. Если A и B независимы, то \bar{A} и \bar{B} также независимы.

Вообще, введенное определением 2.2 понятие независимости обладает всеми свойствами, которых требует интуиция. Для того чтобы определение независимости n событий A_1, A_2, \dots, A_n было столь же хорошим, его надо вводить следующим образом.

Определение 2.3. События A_1, A_2, \dots, A_n называются *независимыми в совокупности*, если для любых k из них ($k \leq n$) выполняется соотношение

$$P(A_{i_1}, A_{i_2}, \dots, A_{i_k}) = \prod_{j=1}^k P(A_{i_j}).$$

Если это соотношение выполняется только при $k=2$, то события называются *попарно независимыми*. Читателю предоставляется привести пример, показывающий, что из попарной независимости не следует независимость в совокупности.

2.3. Прямое произведение вероятностных пространств. Понятие независимости событий, лежащих в одном и том же вероятностном пространстве, может показаться несколько искусственным, так как оно формально никак не связано с различными случайными экспериментами, проводимыми независимо друг от друга. Достаточно лишь слегка изменить вероятности отдельных элементарных событий, и равенство $P(AB) = P(A)P(B)$ перестанет выполняться. Но на самом деле вероятностное пространство, связанное с независимыми событиями, строится обычно специальным образом. Пусть имеются два вероятностных пространства: $\Omega^{(1)}$ и $\Omega^{(2)}$, причем их элементам $\omega_i^{(1)}$ и $\omega_j^{(2)}$ отвечают вероятности $P^{(1)}(\omega_i^{(1)})$ и $P^{(2)}(\omega_j^{(2)})$. Можно представить себе $\Omega^{(1)}$ как вероятностное пространство, связанное с одним случайным экспериментом, а $\Omega^{(2)}$ — с другим.

Определение 2.4. *Прямым произведением* вероятностных пространств $\Omega^{(1)}$ и $\Omega^{(2)}$ называется пространство $\Omega = \Omega^{(1)} \times \Omega^{(2)}$, состоящее из всевозможных пар вида $\omega = (\omega_i^{(1)}, \omega_j^{(2)})$, причем вероятность каждого элементарного события $\omega \in \Omega$ определяется по формуле

$$P(\omega) = P(\omega_i^{(1)}, \omega_j^{(2)}) = P^{(1)}(\omega_i^{(1)}) P^{(2)}(\omega_j^{(2)}).$$

События, связанные с исходом первого случайного эксперимента, можно описывать не только как подмножества $\Omega^{(1)}$, но и как подмножества $\Omega = \Omega^{(1)} \times \Omega^{(2)}$, поступая следующим образом. Пусть $A^{(1)} \subseteq \Omega^{(1)}$. Рассмотрим подмножество $\tilde{A}^{(1)}$ множества Ω следующего вида:

$$\tilde{A}^{(1)} = \{(\omega_i^{(1)}, \omega_j^{(2)}) : \omega_i^{(1)} \in A^{(1)}, \omega_j^{(2)} \in \Omega^{(2)}\}.$$

Аналогично, если $A^{(2)} \subseteq \Omega^{(2)}$, положим

$$\tilde{A}^{(2)} = \{(\omega_i^{(1)}, \omega_j^{(2)}) : \omega_i^{(1)} \in \Omega^{(1)}, \omega_j^{(2)} \in A^{(2)}\}.$$

Множества $\tilde{A}^{(1)}$ и $\tilde{A}^{(2)}$ называются цилиндрическими. Очевидно, что $P(\tilde{A}^{(1)}) = P^{(1)}(A^{(1)})$ и $P(\tilde{A}^{(2)}) = P^{(2)}(A^{(2)})$.

Теорема 2.3. При любых $A^{(1)} \subseteq \Omega^{(1)}$ и $A^{(2)} \subseteq \Omega^{(2)}$ события $\tilde{A}^{(1)}$ и $\tilde{A}^{(2)}$ независимы.

Доказательство. Поскольку

$$\tilde{A}^{(1)} \tilde{A}^{(2)} = \{(\omega_i^{(1)}, \omega_j^{(2)}) : \omega_i^{(1)} \in A^{(1)}, \omega_j^{(2)} \in A^{(2)}\},$$

имеем

$$\begin{aligned} P\{\tilde{A}^{(1)} \tilde{A}^{(2)}\} &= \sum_{\omega_i^{(1)} \in A^{(1)}} \sum_{\omega_j^{(2)} \in A^{(2)}} P\{(\omega_i^{(1)}, \omega_j^{(2)})\} = \\ &= \sum_{\omega_i^{(1)} \in A^{(1)}} \sum_{\omega_j^{(2)} \in A^{(2)}} \bar{P}^{(1)}(\omega_i^{(1)}) P^{(2)}(\omega_j^{(2)}) = \\ &= \left\{ \sum_{\omega_i^{(1)} \in A^{(1)}} P^{(1)}(\omega_i^{(1)}) \right\} \cdot \left\{ \sum_{\omega_j^{(2)} \in A^{(2)}} P^{(2)}(\omega_j^{(2)}) \right\} = \\ &= P^{(1)}(A^{(1)}) P^{(2)}(A^{(2)}) = P(\tilde{A}^{(1)}) P(\tilde{A}^{(2)}), \end{aligned}$$

что и требовалось доказать.

Таким образом, при описании событий, связанных с каждым из двух случайных экспериментов, в терминах одного пространства элементарных событий (прямого произведения $\Omega = \Omega^{(1)} \times \Omega^{(2)}$) получаются независимые события в смысле определения 2.2. Независимость этих событий сохранится, если как угодно изменять вероятности $P^{(1)}$ и $P^{(2)}$.

Изложенная конструкция прямого произведения имеет большое значение в теории вероятностей.

2.4. Формула полной вероятности. Пусть пространство Ω представляется в виде суммы

$$\Omega = H_1 + H_2 + \dots + H_n$$

попарно непересекающихся ($H_i H_j = \emptyset$ при $i \neq j$) событий H_1, H_2, \dots, H_n , A — любое событие (на классическом языке говорят, что H_1, H_2, \dots, H_n есть полная группа событий).

$$\text{Теорема 2.4. } P(A) = \sum_{i=1}^n P(H_i) P(A/H_i)$$

(эта формула называется формулой полной вероятности).

Доказательство. Легко видеть, что

$$A = AH_1 + AH_2 + \dots + AH_n,$$

причем $(AH_i)(AH_j) = \emptyset$ при $i \neq j$. Следовательно,

$$P(A) = \sum_{i=1}^n P(AH_i) = \sum_{i=1}^n P(H_i) P(A/H_i)$$

(см. теорему 2.1).

С точки зрения математика вновь выведенная формула является тривиальной, однако она может применяться для задания разумным образом вероятностей на множестве элементарных событий.

Пример 2.2. Три машины производят болты, причем первая машина производит 20% всей продукции, вторая машина — 30% и третья машина — 50%. Доля брака в продукции первой машины 5%, в продукции второй машины 2%, в продукции третьей — 1%. Чему равна вероятность того, что наудачу взятый болт окажется дефектным?

Решение. Эксперимент может состоять в том, что проверяется наудачу взятый болт. Элементарных исходов два: D — болт дефектный и \bar{D} — болт годный. Обозначим через H_i ($i=1, 2, 3$) событие, состоящее в том, что болт сделан i -той машиной. Тогда имеем

$$P(D) = \sum_{i=1}^3 P(H_i) P(D/H_i) = 0,2 \cdot 0,05 + 0,3 \cdot 0,02 + 0,5 \cdot 0,01 = 0,021.$$

Приведенное решение может казаться не вполне убедительным, хотя против него трудно привести какие-либо возражения. Для того чтобы убедиться в правильности ответа, приведем другое решение.

Рассмотрим в качестве Ω совокупность всех изготовленных тремя машинами болтов. По условию Ω представляется в виде $\Omega = H_1 + H_2 + H_3$, где через H_i обозначено множество болтов, изготовленных i -той машиной. Обозначим через D множество дефектных болтов. Снова «наудачу взятый болт» мы истолкуем так, что все элементы Ω равновероятны. Обозначая через $N(C)$ число элементов любого множества C , имеем

$$P(D) = \frac{N(D)}{N(\Omega)}.$$

В силу условия задачи

$$N(H_1) = 0,2 N(\Omega), \quad N(H_2) = 0,3 N(\Omega),$$

$$N(H_3) = 0,5 N(\Omega), \quad N(DH_1) = 0,05 N(H_1),$$

$$N(DH_2) = 0,02 N(H_2), \quad N(DH_3) = 0,01 N(H_3).$$

Следовательно,

$$\begin{aligned} P(D) &= \frac{N(DH_1) + N(DH_2) + N(DH_3)}{N(\Omega)} = \\ &= 0,05 \frac{N(H_1)}{N(\Omega)} + 0,02 \frac{N(H_2)}{N(\Omega)} + 0,01 \frac{N(H_3)}{N(\Omega)}, \end{aligned}$$

т. е. мы получаем тот же самый ответ.

Мы видим, что в некоторых задачах формула полной вероятности позволяет получать ответ, минуя построение пространства элементарных событий.

2.5. Формула Байеса. В условиях теоремы 2.4 имеет место формула

$$P(H_i/A) = \frac{P(AH_i)}{P(A)} = \frac{P(A/H_i) P(H_i)}{\sum_{j=1}^n P(A/H_j) P(H_j)}.$$

Эта формула называется формулой Байеса.

Формулу Байеса иногда пытаются интерпретировать следующим образом. Пусть при начале некоторого научного исследования у нас имеются n гипотез H_1, H_2, \dots, H_n о природе изучаемого объекта, причем мы приписываем им вероятности $P(H_1), \dots, P(H_n)$ (эти вероятности называются априорными). Затем мы делаем эксперимент, в результате которого может наступить или не наступить событие A . Если событие A наступает, то мы переоцениваем нашу веру в справедливость каждой гипотезы, заменяя вероятности $P(H_i)$ на вероятности $P(H_i/A)$ (эти вероятности называются апостериорными). Так мы продолжаем, пока для некоторого $i=i_0$ апостериорная вероятность гипотезы H_{i_0} делается (почти) равной единице. Тогда гипотеза H_{i_0} на самом деле верна.

Надо сказать, конечно, что подобная «формализация» процесса научного исследования в большинстве случаев бессмысленна, потому что неоткуда взять априорные вероятности гипотез. Имеется, однако, случай, когда такая схема может быть полезной. Речь идет о медицинской диагностике. Пусть в некоторую клинику обращаются больные, у которых может быть одна из болезней H_1, H_2, \dots, H_n .

Обозначим через A комплекс симптомов для данного больного. В таком случае априорные вероятности $P(H_i)$ и условные вероятности $P(A/H_i)$ могут быть экспериментально найдены на основании собранной за прошлые годы статистики. При этом $P(H_i)$ равна примерно частоте болезни H_i среди больных данной клиники, а $P(A/H_i)$ — частота наблюдения комплекса симптомов A у больных с болезнью H_i в данной клинике. Поскольку речь идет о статистике за прошлые годы, то можно считать имеющие-

ся статистические данные почти достоверными (например, брать лишь данные, подтвержденные патологоанатомическим исследованием). Формально применение формулы Байеса не вызывает здесь сомнения, и трудности начинаются только при попытке практической реализации этого плана. Существенная трудность состоит в следующем: посмотрим, насколько часто наблюдались в прошлом больные с данным комплексом симптомов A . Предположим, что в клинике имеется материал о 10 000 больных по каждой болезни H_i . Сколько же существует возможных комплексов симптомов? Не будет преувеличением сказать, что при записи двоичным кодом комплекса симптомов требуется не менее 50 двоичных знаков (в тех случаях, когда ответ двузначен — есть данный симптом или нет — требуется один двоичный знак, результаты же измерений, анализов должны разбиваться на группы, а тут может потребоваться два-три двоичных знака для каждого результата). Тогда возможных наборов симптомов будет 2^{50} , т. е. около 10^{15} , и в среднем на каждый набор приходится $10^4 \cdot 10^{-15} = 10^{-11}$, т. е. практически нуль наблюдений.

При отсутствии наблюдений нельзя определить условные вероятности $P(A/H_i)$. Таким образом, прямое применение формулы Байеса невозможно, а задача машинной диагностики оказывается трудной научной задачей, требующей далеко не тривиальных усилий. Такова судьба не одной формулы Байеса: при ближайшем рассмотрении применение любого метода теории вероятностей требует нетривиальных усилий, если, конечно, стремиться получать заслуживающие доверия результаты.

§ 3

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ ОСНОВНЫЕ ХАРАКТЕРИСТИКИ

3.1. Случайные величины. Мы рассматриваем дискретное (т. е. конечное или счетное) пространство элементарных событий Ω , элементам которого ω отвечают вероятности $P(\omega)$.

Определение 3.1. *Случайной величиной* ξ называется функция $\xi(\omega)$, определенная на множестве Ω и принимающая вещественные или комплексные значения.

Для определенности мы будем рассматривать случайные величины с вещественными значениями. Среди возможных значений $\xi(\omega)$, отвечающих различным $\omega \in \Omega$, не обязательно все различны. Обозначим различные возможные значения случайной величины ξ через $a_1, a_2, \dots, a_n, \dots$ (при этом $a_1, a_2, \dots, a_n, \dots$ — вещественные числа, не обязательно расположенные в порядке возрастания или убывания, так как счетное множество вещественных чисел не всегда можно записать в определенном порядке по величине; например, этого нельзя сделать с множеством всех рациональных чисел).

Введем обозначение¹:

$$\{\xi = a_i\} = \{\omega : \omega \in \Omega, \quad \xi(\omega) = a_i\}.$$

Очевидно, $\{\xi = a_i\}$ есть подмножество множества Ω , следовательно, событие. Обозначим через p_i вероятность этого события:

$$p_i = P\{\xi = a_i\} = \sum_{\omega: \xi(\omega) = a_i} P(\omega)$$

(запись под знаком Σ означает, что суммирование берется по ω таким, что $\xi(\omega) = a_i$, при этом порядок суммирования безразличен, поскольку слагаемые неотрицательны).

Таблица вида

$$\begin{pmatrix} a_1 a_2 \dots a_n \dots \\ p_1 p_2 \dots p_n \dots \end{pmatrix},$$

где в верхней строчке стоят возможные значения случайной величины ξ , а в нижней строчке под каждым значением стоит вероятность $p_i = P\{\xi = a_i\}$ того, что случайная величина ξ принимает это значение, называется *распределением случайной величины* ξ .

В приложениях теории вероятностей, как правило, имеют дело не с самими случайными величинами, а с их распределениями. Это связано с тем, что обычно в результате случайного эксперимента регистрируется значение случайной величины $\xi(\omega)$, но не регистрируется, каким элементарным исходом ω закончился сам опыт. Регистрируется, следовательно, значение функции, но не значение аргумента. То обстоятельство, что при разных ω случайная величина $\xi(\omega)$ может принимать одно и то же значение, неожиданно оказывается необычайно существенным: множество $\{a_1, a_2, \dots\}$ возможных значений случайной величины может быть гораздо проще, чем все множество Ω . Поэтому может получиться так, что мы не в состоянии узнать из опыта вероятности $P(\omega)$, но можем определить по частотам вероятности $p_i = P\{\xi = a_i\}$.

Поясним сказанное примером. Пусть бросается игральная кость и наблюдаемая нами случайная величина ξ есть число выпавших очков. Введем пространство Ω на основании следующих соображений. Хорошо известно, что движение твердого тела вполне определится, если в некоторый момент задать шесть параметров, определяющих его положение в пространстве, вместе со скоростями изменения этих параметров. Будем понимать под элементарным событием ω набор этих двенадцати чисел (измеренных в тот момент, когда мы выпускаем кость из рук), записанных с таким числом десятичных знаков, которого достаточно для определения, какой гранью кверху в конце концов остановится кость. Тогда, зная ω , мы знаем и ξ , т. е. ξ есть функция от ω .

¹ В обозначениях теории множеств двоеточие заменяет слова «такие, что». Обозначение $\{\omega : \omega \in \Omega, \xi(\omega) = a_i\}$ подробно читается так: «множество элементов ω таких, что $\omega \in \Omega$ и $\xi(\omega) = a_i$ ».

Множество Ω возможных значений ω конечно, так как в опыте бросания кости ее положение и скорость может колебаться лишь в конечных пределах, а мы хотим регистрировать эти параметры лишь с конечным числом десятичных знаков. Но совершенно очевидно, что $\xi(\omega)$ наблюдать легко, в то время как при регистрации ω возникают непреодолимые трудности. Кроме того, возможных значений ω очень много, а возможных значений $\xi(\omega)$ всего шесть. Отсюда видно, насколько сильным может быть упрощение при переходе от случайной величины к ее распределению.

Значения $\{a_1, a_2, \dots, a_n, \dots\}$ случайной величины могут быть любыми. Вероятности p_i , очевидно, неотрицательны, а кроме того,

$$\begin{aligned} \sum p_i &= \sum_{a_i} \mathbf{P} \{ \xi = a_i \} = \sum_{a_i} \left(\sum_{\omega: \xi(\omega)=a_i} \mathbf{P}(\omega) \right) = \\ &= \sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1. \end{aligned}$$

Обратно, любой таблице

$$\begin{pmatrix} a_1, a_2, \dots, a_n, \dots \\ p_1, p_2, \dots, p_n, \dots \end{pmatrix},$$

где числа $a_1, a_2, \dots, a_n, \dots$ различны между собой, $p_i \geq 0$ и $\sum p_i = 1$, отвечает случайная величина, имеющая эту таблицу своим распределением. Для доказательства достаточно положить

$$\Omega = (a_1, a_2, \dots, a_n, \dots), \quad \mathbf{P}(a_i) = p_i, \quad \xi(a_i) = a_i.$$

В приложениях редукция к распределению случайной величины часто бывает недостаточной. Действительно, возможных значений $a_1, a_2, \dots, a_n, \dots$ может быть все же еще слишком много, чтобы можно было определить из опыта их вероятности p_i .

В нашей математической модели множество $\{a_1, a_2, \dots\}$ может быть даже бесконечным (счетным). Но если даже оно конечно, но содержит много элементов, определение всех вероятностей p_i может быть невозможно. Желательно поэтому охарактеризовать распределение несколькими параметрами, которые затем можно было бы определять экспериментально. Эта задача чрезвычайно важна и будет изучаться на протяжении всего курса, а сейчас мы займемся формальным введением важнейших параметров.

3.2. Математическое ожидание. Пусть дана случайная величина $\xi(\omega)$.

Определение 3.2. Если ряд $\sum_{\omega \in \Omega} \xi(\omega) \mathbf{P}(\omega)$ сходится абсолютно, то его сумма

$$M\xi = \sum_{\omega \in \Omega} \xi(\omega) P(\omega)$$

называется *математическим ожиданием* случайной величины ξ .

Замечание. Если указанный ряд сходится неабсолютно (т. е. условно), то говорят, что случайная величина не имеет математического ожидания. Действительно, в этом случае значение суммы зависит от порядка слагаемых, в то время как среди элементарных событий ω обычно нельзя установить естественного отношения порядка. Если указанный ряд сходится к $+\infty$ или $-\infty$ (независимо от порядка слагаемых), то иногда полагают соответственно $M\xi = +\infty$ или $M\xi = -\infty$. Мы, однако, и в этом случае будем говорить, что случайная величина не имеет математического ожидания.

Пусть $a_1, a_2, \dots, a_n, \dots$ — возможные значения случайной величины, $p_1, p_2, \dots, p_n, \dots$ — вероятности этих значений.

Теорема 3.1. Ряды $\sum_{\omega \in \Omega} \xi(\omega) P(\omega)$ и $\sum a_i p_i$ одновременно сходятся или не сходятся абсолютно. В случае абсолютной сходимости

$$M\xi = \sum a_i p_i.$$

Таким образом, математическое ожидание выражается через распределение случайной величины: нужно значения случайной величины умножить на их вероятности и сложить все полученные произведения.

Доказательство. Пусть ряд $\sum_{\omega \in \Omega} \xi(\omega) P(\omega)$ абсолютно сходится. В таком случае его члены можно как угодно переставлять и группировать. Имеем поэтому:

$$\begin{aligned} \sum_{\omega \in \Omega} \xi(\omega) P(\omega) &= \sum_{a_i} \left(\sum_{\omega: \xi(\omega)=a_i} \xi(\omega) P(\omega) \right) = \\ &= \sum_{a_i} (a_i \sum_{\omega: \xi(\omega)=a_i} P(\omega)) = \sum_{a_i} a_i P\{\xi = a_i\} = \sum a_i p_i. \end{aligned}$$

Аналогично

$$\sum_{\omega \in \Omega} |\xi(\omega)| P(\omega) = \sum_{a_i} |a_i| p_i,$$

откуда видно, что ряды $\sum_{\omega \in \Omega} \xi(\omega) P(\omega)$ и $\sum a_i p_i$ одновременно сходятся или не сходятся абсолютно. Теорема доказана.

На равенстве $M\xi = \sum a_i p_i$ основаны различные интерпретации понятия математического ожидания. Потребность интерпретации математических понятий в других терминах связана с тем интуи-

тивно осознаваемым фактом, что все физико-математические науки едины, так что было бы хорошо, если бы в них было поменьше различных понятий. Действительно, если мы в точки прямой линии с абсциссами $a_1, a_2, \dots, a_n, \dots$ положим массы p_1, p_2, \dots, p_n , то, учитывая, что $\sum p_i = 1$, найдем, что $M\xi = \sum a_i p_i$ есть абсцисса центра тяжести этой системы материальных точек.

Интересно установить также связь с понятиями функционального анализа. На математическое ожидание можно смотреть как на функционал, т. е. как на операцию, ставящую в соответствие функции $\xi = \xi(\omega)$ число $M\xi$. Этот функционал оказывается линейным, иными словами имеет место

Теорема 3.2. Пусть ξ и η — две случайные величины такие, что $M\xi$ и $M\eta$ существуют, а a и b — любые числа. Тогда у линейной комбинации $a\xi + b\eta$ также существует математическое ожидание, причем

$$M(a\xi + b\eta) = aM\xi + bM\eta.$$

Доказательство. Имеем

$$\begin{aligned} aM\xi + bM\eta &= a \sum_{\omega \in \Omega} \xi(\omega) P(\omega) + b \sum_{\omega \in \Omega} \eta(\omega) P(\omega) = \\ &= \sum_{\omega \in \Omega} (a\xi(\omega) + b\eta(\omega)) P(\omega) = \sum_{\omega \in \Omega} (a\xi + b\eta)(\omega) P(\omega) = M(a\xi + b\eta). \end{aligned}$$

Поясним, что мы воспользовались тем, что сумма абсолютно сходящихся рядов есть абсолютно сходящийся ряд и через $(a\xi + b\eta)(\omega)$ обозначили значение случайной величины $a\xi + b\eta$ на $\omega \in \Omega$.

Теорема 3.3. Пусть $f(x)$ — любая функция вещественного переменного, ξ — случайная величина с распределением

$$\begin{pmatrix} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix}.$$

В таком случае

$$Mf(\xi) = \sum_{a_i} f(a_i) p_i$$

(в предположении, что последний ряд сходится абсолютно).

Доказательство. Имеем

$$\begin{aligned} \sum_{\omega \in \Omega} f(\xi(\omega)) P(\omega) &= \sum_{a_i} \left(\sum_{\omega: \xi(\omega) = a_i} f(\xi(\omega)) P(\omega) \right) = \\ &= \sum_{a_i} \left(f(a_i) \sum_{\omega: \xi(\omega) = a_i} P(\omega) \right) = \sum_{a_i} f(a_i) p_i \end{aligned}$$

(проведенные операции законны, так как их применение к ряду $\sum_{\omega \in \Omega} |f(\xi_i(\omega))| \mathbf{P}(\omega)$ дает сходящийся ряд $\sum_{a_i} |f(a_i)| p_i$).

Пример 3.1. Азартными играми называются игры, в которых выигрыш игроков зависит от исхода случайного эксперимента. Если при k -том повторении игры выигрыш первого игрока есть случайная величина ξ_k , а всего сыграно n игр, то общий выигрыш первого игрока есть, очевидно, $\sum_{k=1}^n \xi_k$. Посмотрим, чему примерно равняется его доход в среднем на одну игру, т. е. вычислим $\frac{1}{n} \sum_{k=1}^n \xi_k$. Естественно считать, что случайные величины ξ_k все имеют одинаковые распределения. Пусть их значения суть a_1, a_2, \dots , а $\mathbf{P}\{\xi_k = a_i\} = p_i$. Следовательно, в сумме $\sum_{k=1}^n \xi_k$ отдельные слагаемые принимают значения a_1, a_2, \dots . Обозначим через $N(a_i)$ число этих слагаемых, равных a_i . Тогда имеем

$$\frac{1}{n} \sum_{k=1}^n \xi_k = \sum_{a_i} a_i \frac{N(a_i)}{n}.$$

При большом n естественно ожидать, что $\frac{N(a_i)}{n} \approx p_i$. Поэтому

$$\frac{1}{n} \sum_{k=1}^n \xi_k \approx \sum_{a_i} a_i p_i = M\xi_k.$$

Следовательно, если $M\xi_k > 0$, то игра для первого игрока выгодна, а если $M\xi_k < 0$, то невыгодна. Применим эти соображения для исследования следующей известной игры.

Играют двое. Первый игрок зажимает в кулаке одну из двух монет: 10 или 20 коп. Второй пытается угадать, какая монета спрятана, и если он угадывает правильно, то он ее получает, а если неправильно — платит первому игроку 15 коп. Спрашивается, как наиболее выгодно поступать первому игроку.

Возможные исходы одной игры представлены следующей таблицей:

II \ I	10	20
	10 20	-10 +15 -20

Эта таблица, называемая матрицей игры, означает следующее. Если, например, первый игрок выбирает «стратегию» 20 коп., а второй игрок выбирает «стратегию» 10 коп., то выигрыш первого игрока + 15 коп. может быть прочитан на пересечении столбца, отвечающего выбранной стратегии первого игрока, и строки, отвечающей стратегии второго.

Первый игрок может менять свою стратегию от игры к игре согласно какому-нибудь правилу. Однако, по принятой в теории игр концепции, мы ждем от противника большой проницательности и не надеемся его запутать: он непременно разгадает наше правило и выиграет. Единственная наша надежда — воспользоваться тем, чего никак нельзя предвидеть, а именно производимым скрытно от противника случайным экспериментом. Будем считать поэтому, что первый игрок выбирает стратегию 10 коп. с вероятностью p , а 20 коп. — с вероятностью $1-p$, и постараемся подобрать p оптимальным образом. Для этого будем искать

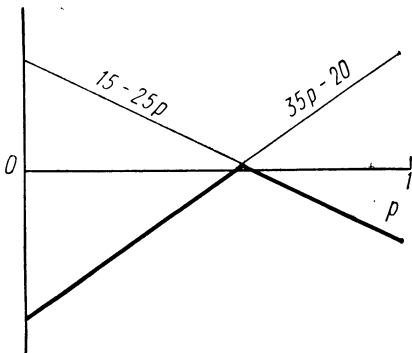


Рис. 3.1

математическое ожидание выигрыша. Если второй игрок выбирает стратегию 10 коп., то с вероятностью p значение выигрыша первого игрока есть (-10) , а с вероятностью $1-p$ есть $+15$. Следовательно, математическое ожидание равно

$$(-10)p + 15(1-p) = 15 - 25p.$$

Если второй игрок выбирает стратегию 20 коп., то аналогично получаем для математического ожидания выражение

$$15p - 20(1-p) = 35p - 20.$$

Поэтому первому игроку гарантируется средний выигрыш на одну игру, равный

$$\varphi(p) = \min(15 - 25p, 35p - 20). \quad (3.1)$$

При этом p пробегает значения от 0 до 1. График функции $\varphi(p)$ изображен на рис. 3.1. Очевидно, $\max_p \varphi(p)$ достигается в точке, где

$$15 - 25p = 35p - 20,$$

откуда

$$p = \frac{7}{12}; \quad \varphi\left(\frac{7}{12}\right) = \frac{5}{12}.$$

Таким образом, пряча 10 коп. с вероятностью $7/12$, а 20 коп. с вероятностью $5/12$, первый игрок может гарантировать себе средний выигрыш за одну игру, равный $5/12$ коп.

Если мы будем рассуждать теперь с точки зрения второго игрока, то таким же образом придем к выбору вероятности π того, что второй игрок выберет стратегию 10 коп. Если первый игрок прячет 10 коп., то средний выигрыш второго есть

$$10\pi - 15(1 - \pi) = 25\pi - 15.$$

Если же первый игрок прячет 20 коп., получим

$$-15\pi + 20(1 - \pi) = 20 - 35\pi.$$

Число π надо выбрать так, чтобы достигался

$$\max_{\pi} \{ \min (25\pi - 15, 20 - 35\pi) \}. \quad (3.2)$$

При $\pi = \frac{7}{12}$ выражение (3.2) равно $\left(-\frac{5}{12}\right)$.

Таким образом, что бы ни делал первый игрок, второй игрок может не проигрывать в среднем за одну игру более $\frac{5}{12}$ коп. По принятой в теории игр модели оба противника должны удовлетвориться тем, что они могут заведомо себе обеспечить, и не отступать от стратегий, соответственно $p = \frac{7}{12}$ и $\pi = \frac{7}{12}$.

Существование таких стратегий является частным случаем одной из основных теорем теории игр.

Итак, с точки зрения первого игрока, описанная игра является выгодной: можно выиграть в среднем $5/12$ коп. за одну игру. Однако при практической реализации этого способа обогащения неизбежно возникает вопрос, как велико может быть отклонение фактического выигрыша от его математического ожидания. В частности, сколько раз нужно повторить игру, чтобы с высокой вероятностью фактический выигрыш был не менее одного рубля? На эти вопросы можно ответить, но для этого нужно значительно продвинуть теоретический материал курса.

3.3. Дисперсия. Нам необходимо ввести какую-то характеристику отклонения случайной величины от ее среднего значения.

Определение 3.3. *Дисперсией* $D\xi$ случайной величины ξ называется математическое ожидание случайной величины $(\xi - M\xi)^2$

$$D\xi = M\{(\xi - M\xi)^2\}.$$

З а м е ч а н и е. В правой части последнего выражения фигурные скобки всегда опускаются; вообще $M\xi^2$ означает $M(\xi)^2$, а квадрат математического ожидания $M\xi$ записывается в виде $(M\xi)^2$.

По теореме 3.3. получаем следующее выражение для дисперсии:

$$D\xi = M(\xi - M\xi)^2 = \sum_{a_i} (a_i - M\xi)^2 p_i.$$

Раскрывая скобки, имеем

$$\begin{aligned} D\xi &= M(\xi - M\xi)^2 = M(\xi^2 - 2\xi M\xi + (M\xi)^2) = M\xi^2 - M(2\xi M\xi) + \\ &+ M(M\xi)^2 = M\xi^2 - 2M\xi M\xi + (M\xi)^2 = M\xi^2 - (M\xi)^2 \end{aligned}$$

(здесь мы воспользовались теоремой 3.2 и интерпретировали постоянную величину $M\xi$ как случайную величину, равную $M\xi$ при всех $\omega \in \Omega$).

Таким образом, получаем полезную формулу

$$D\xi = M\xi^2 - (M\xi)^2 = \Sigma a_i^2 p_i - (\Sigma a_i p_i)^2.$$

Почему в качестве меры отклонения от математического ожидания выбрано $M(\xi - M\xi)^2$?

С тем же правом можно было бы выбрать, например, $M|\xi - M\xi|$ или $M|\xi - M\xi|^3$, или $M(\xi - M\xi)^4$.

Некоторый ответ на этот вопрос дает механическая аналогия: если $M\xi$ имеет смысл центра тяжести, то

$$D\xi = \Sigma (a_i - M\xi)^2 p_i$$

имеет смысл момента инерции относительно центра тяжести. Величины типа

$$M|\xi - M\xi| = \Sigma |a_i - M\xi| p_i$$

вообще никогда не встречаются в механике (абсолютная величина $|x|$ не является аналитической функцией x).

Но полный ответ на этот вопрос будет получен при дальнейшем изучении теории вероятностей: дело в том, что именно дисперсия, а не какая-либо другая из возможных мер отклонения от среднего, входит в формулировку важнейшей в теории вероятностей центральной предельной теоремы.

3.4. Независимые случайные величины. Пусть в результате опыта могут наблюдаться две случайные величины ξ и η . Слова «могут наблюдаться», очевидно, можно понимать в том смысле, что для любых двух числовых множеств A и B мы можем сказать, произошло или не произошло каждое из двух событий $\{\xi \in A\}$ и $\{\eta \in B\}$. «Независимость» случайных величин интуитивно понимается так, что, зная результат наблюдения над одной случайной величиной, мы ничего не можем сказать дополнительно о другой случайной величине. Этим мотивируется

Определение 3.4. Две случайных величины ξ и η называются *независимыми*, если для любых двух числовых множеств A и B события $\{\xi \in A\}$ и $\{\eta \in B\}$ независимы.

Теорема 3.4. Пусть $a_1, a_2, \dots, a_n, \dots$ — возможные значения случайной величины ξ и $P\{\xi=a_i\}=p_i$, $b_1, b_2, \dots, b_m, \dots$ — возможные значения случайной величины η и $P\{\eta=b_j\}=q_j$. Для независимости случайных величин ξ и η необходимо и достаточно, чтобы события $\{\xi=a_i\}$ и $\{\eta=b_j\}$ были независимы при любых a_i и b_j , т. е. чтобы

$$P\{(\omega: \xi(\omega)=a_i, \eta(\omega)=b_j)\} = P_i\{\xi=a_i, \eta=b_j\} = \\ = P\{\xi=a_i\}P\{\eta=b_j\}=p_iq_j.$$

Доказательство. Для любых A и B имеем

$$P\{\xi \in A, \eta \in B\} = \sum_{a_i \in A, b_j \in B} P\{\xi=a_i, \eta=b_j\} = \\ = \sum_{a_i \in A, b_j \in B} P\{\xi=a_i\}P\{\eta=b_j\} = \sum_{a_i \in A} P\{\xi=a_i\} \sum_{b_j \in B} P\{\eta=b_j\} = \\ = P\{\xi \in A\}P\{\eta \in B\},$$

что и требовалось доказать.

Теорема 3.5. Любые функции $f(\xi)$ и $g(\eta)$ от независимых случайных величин ξ и η являются независимыми случайными величинами.

Доказательство. Обозначая через f^{-1} и g^{-1} операции взятия полного прообраза, имеем

$$P\{f(\xi) \in A, g(\eta) \in B\} = P\{\xi \in f^{-1}(A), \eta \in g^{-1}(B)\} = \\ = P\{\xi \in f^{-1}(A)\}P\{\eta \in g^{-1}(B)\} = P\{f(\xi) \in A\}P\{g(\eta) \in B\}.$$

Определение 3.5. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ называются независимыми в совокупности, если для любых числовых множеств A_1, A_2, \dots, A_n события $\{\xi_1 \in A_1\}, \{\xi_2 \in A_2\}, \dots, \{\xi_n \in A_n\}$ независимы в совокупности.

Теорема 3.6. Пусть $a_1^{(k)}, a_2^{(k)}, \dots, a_n^{(k)}, \dots$ — значения случайной величины ξ_k . Для независимости случайных величин $\xi_1, \xi_2, \dots, \xi_n$ достаточно, чтобы

$$P\{\xi_1=a_1^{(1)}, \xi_2=a_2^{(2)}, \dots, \xi_n=a_n^{(n)}\} = \prod_{j=1}^n P\{\xi_j=a_j^{(j)}\}. \quad (3.3)$$

Доказательство. Требуется доказать, что при любом $k \leq n$ и любых множествах A_{i_1}, \dots, A_{i_k}

$$P\{\xi_{i_1} \in A_{i_1}, \dots, \xi_{i_k} \in A_{i_k}\} = \prod_{s=1}^k P\{\xi_{i_s} \in A_{i_s}\}.$$

Заметим, что, полагая $A_i = (-\infty, \infty)$ для $i \in \overline{(i_1, \dots, i_k)}$ и $A_i = A_{i_s}$, если $i = i_s$, имеем

$$\{\xi_{i_1} \in A_{i_1}, \dots, \xi_{i_k} \in A_{i_k}\} = \{\xi_1 \in A_1, \xi_2 \in A_2, \dots, \xi_n \in A_n\}.$$

Таким образом, событие, относящееся к случайным величинам $\xi_{i_1}, \dots, \xi_{i_k}$, можно заменить событием, относящимся ко всем случайным величинам ξ_1, \dots, ξ_n . Дальнейшие выкладки такие же, как и при доказательстве теоремы 3.4.

З а м е ч а н и е. Таким образом, независимость в совокупности n случайных величин может быть сформулирована в виде равенства (3.3), относящегося сразу к n случайным величинам. В то же время независимость в совокупности n событий определяется иначе (определение 2.3). Это связано с тем, что в настоящей книге принято традиционное определение независимости событий (см. в методических замечаниях о том, почему это сделано и как можно было бы изменить традиционное определение).

Теорема 3.7. *Если случайные величины ξ и η независимы и существуют $M\xi$ и $M\eta$, то существует математическое ожидание произведения $\xi\eta$ и*

$$M\xi\eta = M\xi M\eta.$$

Доказательство. Имеем

$$\begin{aligned} M\xi\eta &= \sum_{\omega \in \Omega} \xi(\omega) \eta(\omega) P(\omega) = \sum_{a_i, b_j} \left\{ \sum_{\omega: \substack{\xi(\omega)=a_i \\ \eta(\omega)=b_j}} \xi(\omega) \eta(\omega) P(\omega) \right\} = \\ &= \sum_{a_i, b_j} \left\{ a_i b_j \sum_{\omega: \substack{\xi(\omega)=a_i \\ \eta(\omega)=b_j}} P(\omega) \right\} = \sum_{a_i, b_j} a_i b_j P\{\xi = a_i, \eta = b_j\} = \\ &= \left(\sum_{a_i} a_i P\{\xi = a_i\} \right) \left(\sum_{b_j} b_j P\{\eta = b_j\} \right) = M\xi M\eta \end{aligned}$$

(проведенные выкладки законны, так как их применение к ряду $\sum_{\omega \in \Omega} |\xi(\omega)| \cdot |\eta(\omega)| P(\omega)$ приводит его к виду $M|\xi| M|\eta|$, следовательно, в силу существования $M|\xi|$ и $M|\eta|$, все наши выкладки делаются с абсолютно сходящимися рядами).

3.5. Совместное распределение. Если имеется две случайных величины ξ и η , причем ξ принимает значения $a_1, a_2, \dots, a_n, \dots$, а η принимает значения $b_1, b_2, \dots, b_m, \dots$, то рассматривают события

$$\{\omega : \xi(\omega) = a_i, \eta(\omega) = b_j\} = \{\xi = a_i, \eta = b_j\}$$

и вероятности этих событий. Набор векторов (a_i, b_i) вместе с вероятностями $P\{\xi = a_i, \eta = b_j\}$ образует так называемое совместное распределение случайных величин ξ и η . Если $f(x, y)$ — лю-

бая функция двух переменных, то совершенно так же, как теореме 3.3, можно доказать следующую теорему.

Теорема 3.8.

$$Mf(\xi, \eta) = \sum_{a_i, b_j} f(a_i, b_j) P\{\xi = a_i, \eta = b_j\}.$$

В частности, полагая $f(x, y) = 1$, если $x = a_{i_0}$, и $f(x, y) = 0$, если $x \neq a_{i_0}$, имеем

$$\begin{aligned} Mf(\xi, \eta) &= P\{f(\xi, \eta) = 1\} = P\{\xi = a_{i_0}\} = \\ &= \sum_{a_i, b_j} f(a_i, b_j) P\{\xi = a_i, \eta = b_j\} = \sum_{b_j} P\{\xi = a_{i_0}, \eta = b_j\}, \end{aligned}$$

т. е. распределение случайной величины ξ получается из совместного распределения путем суммирования по значениям второй случайной величины. Впрочем, этот факт легко установить и непосредственно.

Если случайная величина ξ имеет n различных значений, а случайная величина η имеет m различных значений, то для задания совместного распределения ξ и η надо задать nm вероятностей $P\{\xi = a_i, \eta = b_j\}$.

В случае независимых случайных величин

$$P\{\xi = a_i, \eta = b_j\} = P\{\xi = a_i\} P\{\eta = b_j\},$$

так что достаточно задать $m+n$ вероятностей $P\{\xi = a_i\}$ и $P\{\eta = b_j\}$. Для нескольких случайных величин соответствующая разница становится еще более резкой. Этим объясняется то, что в приложениях предпочитают вероятностные модели с независимыми случайными величинами.

3.6. Свойства дисперсии. Теорема 3.9. $D(c\xi) = c^2 D\xi$, где c — константа.

Доказательство очевидно.

Рассмотрим вопрос о дисперсии суммы $\xi + \eta$. Имеем

$$\begin{aligned} D(\xi + \eta) &= M\{\xi + \eta - M(\xi + \eta)\}^2 = M\{(\xi - M\xi) + (\eta - M\eta)\}^2 = \\ &= M(\xi - M\xi)^2 + M(\eta - M\eta)^2 + 2M\{(\xi - M\xi)(\eta - M\eta)\}. \end{aligned}$$

Выражение $M\{(\xi - M\xi)(\eta - M\eta)\} = \text{cov}(\xi, \eta)$ называется *ковариацией* величин ξ и η . Для независимых и случайных величин

$$\text{cov}(\xi, \eta) = M(\xi - M\xi)M(\eta - M\eta) = 0$$

(теоремы 3.5 и 3.7). Мы получили следующее утверждение

Теорема 3.10. Для любых случайных величин ξ и η

$$D(\xi + \eta) = D\xi + D\eta + 2\text{cov}(\xi, \eta).$$

Для независимых ξ и η

$$D(\xi + \eta) = D\xi + D\eta.$$

§ 4

НЕРАВЕНСТВО ЧЕБЫШЕВА. ЗАКОН БОЛЬШИХ ЧИСЕЛ. ИСПЫТАНИЯ БЕРНУЛЛИ. ТЕОРЕМА ПУАССОНА

Содержательные теоремы теории вероятностей могут быть получены, если рассматривать не одно событие или случайную величину, а много. Действительно, пока имеется одно событие, все, что можно о нем сказать,—это, что оно или произойдет, или нет, а об одной случайной величине — что она примет одно из своих возможных значений. Однако о большом числе случайных событий (случайных величин) можно сделать некоторые практически достоверные выводы. Для этого, очевидно необходим аппарат, позволяющий с тем или иным приближением получать выводы о вероятностях разных событий, связанных с большим числом случайных величин. Мы вскоре увидим, что только что доказанная простая теорема

$$D(\xi + \eta) = D\xi + D\eta$$

для независимых ξ и η является важным звеном этого аппарата. Вторым его звеном является знаменитое неравенство Чебышева.

4.1. Неравенство Чебышева. Дисперсия есть мера отклонения случайной величины от ее математического ожидания. Если дисперсия $D\xi$ мала, то большие отклонения ξ от $M\xi$ маловероятны. Точно этот факт выражается следующей теоремой.

Теорема 4.1. *Имеет место следующее неравенство, называемое неравенством Чебышева¹:*

$$P\{|\xi - M\xi| \geq \varepsilon\} \leq \frac{D\xi}{\varepsilon^2}$$

для любого $\varepsilon > 0$.

Доказательство. Имеем

$$D\xi = \sum_{a_i} (a_i - M\xi)^2 P\{\xi = a_i\},$$

¹ Обозначение $P\{|\xi - M\xi| \geq \varepsilon\}$ есть сокращенное обозначение для

$$P\{\omega : |\xi(\omega) - M\xi| \geq \varepsilon\} = P\{\omega : \xi(\omega) = a_i, |a_i - M\xi| \geq \varepsilon\} =$$

$$= \sum_{a_i : |a_i - M\xi| \geq \varepsilon} P\{\xi = a_i\}.$$

где a_i — возможные значения случайной величины ξ . Если суммирование по всем a_i заменить суммированием по всем a_i таким, что $|a_i - M\xi| \geq \varepsilon$, то сумма не увеличится.

Таким образом,

$$\begin{aligned} D\xi &\geq \sum_{a_i: |a_i - M\xi| \geq \varepsilon} (a_i - M\xi)^2 P\{\xi = a_i\} \geq \\ &\geq \varepsilon^2 \sum_{a_i: |a_i - M\xi| \geq \varepsilon} P\{\xi = a_i\} = \varepsilon^2 P\{|\xi - M\xi| \geq \varepsilon\}, \end{aligned}$$

что, очевидно, эквивалентно доказываемому утверждению.

Определение 4.1. Говорят, что последовательность ξ_1, ξ_2, \dots случайных величин *сходится по вероятности к нулю*: $\xi_n \rightarrow 0$ (вер), если для любого $\varepsilon > 0$, при $n \rightarrow \infty$

$$P\{|\xi_n| > \varepsilon\} \rightarrow 0.$$

Определение 4.2. Говорят, что последовательность случайных величин ξ_1, ξ_2, \dots *сходится по вероятности к (случайной или неслучайной) величине a* : $\xi_n \rightarrow a$ (вер), если последовательность разностей $\xi_1 - a_1, \xi_2 - a_2, \dots$ сходится к нулю по вероятности.

Теорема 4.2. Если для последовательности случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$ имеем $D\xi_n \rightarrow 0$, то последовательность $\xi_1 - M\xi_1, \xi_2 - M\xi_2, \dots$ сходится к нулю по вероятности.

Доказательство. В силу неравенства Чебышева, при $n \rightarrow \infty$

$$P\{|\xi_n - M\xi_n| > \varepsilon\} \leq \frac{D\xi_n}{\varepsilon^2} \rightarrow 0$$

для любого $\varepsilon > 0$.

4.2. Закон больших чисел. Случайные величины возникают в приложениях как результаты измерений, причем либо сами измерения подвержены случайным ошибкам, либо объекты измерения случайным образом выбираются из некоторой совокупности. Давно было замечено, что в то время как результаты отдельных измерений $\xi_1, \xi_2, \dots, \xi_n$ могут колебаться сильно, их средние арифметические $\frac{1}{n}(\xi_1 + \xi_2 + \dots + \xi_n)$ обнаруживают гораздо большую устойчивость (а, конечно, естествоиспытатель стремится получать устойчивые, т. е. мало меняющиеся в разных сериях опытов, характеристики). В частном случае этот экспериментальный факт уже отмечался ранее. Действительно, рассмотрим некоторое событие A , которое может произойти или не произойти в результате опыта, и положим $\xi_i = 1$, если в i -том опыте событие A произошло, и $\xi_i = 0$, если в i -том опыте событие A не произошло. Тогда $\xi_1 + \dots + \xi_n$ есть, очевидно, число наступлений события A в n опытах, а $\frac{1}{n}(\xi_1 + \dots + \xi_n)$ есть частота наступления события A . На яв-

лении устойчивости частот основаны все применения теории вероятностей.

Если явление устойчивости средних имеет место в действительности, то в математической модели, с помощью которой мы изучаем случайные явления, должна существовать отражающая этот факт теорема. В условии этой теоремы нужно ввести некоторые ограничения на случайные величины ξ_1, \dots, ξ_n . Эти ограничения делятся на две группы. Одна группа предположений — одинаковость распределений всех случайных величин $\xi_1, \xi_2, \dots, \xi_n$. Достаточно требовать это свойство в сильно ослабленной форме: мы предположим, что одинаковы математические ожидания

$$M\xi_1 = M\xi_2 = \dots = M\xi_n = a$$

и что дисперсии случайных величин ограничены одним и тем же числом

$$D\xi_i < c, \quad i = 1, 2, \dots, n, \dots$$

Вторая группа предположений — предположения о независимости величин $\xi_1, \xi_2, \dots, \xi_n, \dots$. Нам достаточно будет предположить, что эти величины попарно независимы (т. е. любые две ξ_i и ξ_j при $i \neq j$ независимы). Тогда, очевидно,

$$D(\xi_1 + \xi_2 + \dots + \xi_n) = D\xi_1 + D\xi_2 + \dots + D\xi_n.$$

Сформулируем знаменитый закон больших чисел в форме Чебышева.

Теорема 4.3. Пусть случайные величины $\xi_1, \xi_2, \dots, \xi_n, \dots$ попарно независимы и $D\xi_i < c$. В таком случае для любого положительного ε имеем при $n \rightarrow \infty$

$$P \left\{ \left| \frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} - \frac{M\xi_1 + M\xi_2 + \dots + M\xi_n}{n} \right| > \varepsilon \right\} \rightarrow 0.$$

Доказательство. Поскольку

$$M \left(\frac{\xi_1 + \dots + \xi_n}{n} \right) = \frac{M\xi_1 + \dots + M\xi_n}{n},$$

достаточно, в силу теоремы 4.2, установить, что дисперсия величины $\frac{1}{n} (\xi_1 + \dots + \xi_n)$ стремится к нулю. Имеем при $n \rightarrow \infty$

$$D \left(\frac{\xi_1 + \dots + \xi_n}{n} \right) = \frac{1}{n^2} \sum_{i=1}^n D\xi_i \leq \frac{nc}{n^2} = \frac{c}{n} \rightarrow 0,$$

что и требовалось доказать.

Следствие. Если $M\xi_1 = \dots = M\xi_n = a$, то для любого $\varepsilon > 0$ при $n \rightarrow \infty$

$$P \left\{ \left| \frac{\xi_1 + \dots + \xi_n}{n} - a \right| > \varepsilon \right\} \rightarrow 0.$$

З а м е ч а н и е. Это следствие и является выражением устойчивости средних.

Прежде считалось иногда, что из этой математической теоремы вытекает экспериментальный факт устойчивости средних.

Логически это, конечно, неверно: ни из какой теоремы математики, механики или физики логически не вытекает тот или иной результат эксперимента, поскольку никогда на практике нельзя гарантировать выполнение условия теоремы. Тем не менее, все существование физико-математических наук основано на том, что следствия из хорошо подобранной математической модели обычно находят свое подтверждение в практической деятельности. Посмотрим, каким же образом интерпретировать в виде частоты ту вероятность, о которой закон больших чисел утверждает, что она стремится к нулю. Для начала нам даны случайные величины $\xi_1, \xi_2, \dots, \xi_n$. Это означает, что серию из n экспериментов можно много раз повторять, причем эти случайные величины будут при каждом повторении серии из n опытов получать свои значения. Предполагается, что можно будет говорить о вероятности того, что

$$\{\xi_1 \in A_1, \xi_2 \in A_2, \dots, \xi_n \in A_n\}, \quad (4.1)$$

где A_1, \dots, A_n — произвольные числовые множества, т. е. частоты наступления этого события будут устойчивы. Более того, предполагается, что можно будет говорить о вероятности

$$P\{(\xi_1, \dots, \xi_n) \in C^{(n)}\} \quad (4.2)$$

того, что вектор (ξ_1, \dots, ξ_n) попадет в любое подмножество $C^{(n)}$ n -мерного евклидова пространства (предполагается в том смысле, что эта вероятность должна допускать частотную интерпретацию). Положим

$$C^{(n)} = \left\{ x = (x_1, \dots, x_n) : \left| \frac{x_1 + \dots + x_n}{n} - a \right| > \varepsilon \right\}.$$

Тогда, в условиях теоремы 4.3,

$$P\{(\xi_1, \dots, \xi_n) \in C^{(n)}\} = P\left\{ \left| \frac{\xi_1 + \dots + \xi_n}{n} - a \right| > \varepsilon \right\}$$

при большом n будет сколь угодно малой, т. е. событие $(\xi_1, \dots, \xi_n) \in C^{(n)}$ будет очень маловероятным. Следовательно, подавляющее большинство серий из n опытов приведет к тому, что вектор, составленный из их результатов, не попадет в $C^{(n)}$.

Отметим, что чрезвычайно трудно экспериментально проверить предположение о существовании вероятностей (4.1) и (4.2) — ведь это надо сделать для всевозможных подмножеств A_1, \dots, A_n и $C^{(n)}$. Таким образом, возможность рассматривать результаты n экспериментов ξ_1, \dots, ξ_n как набор случайных величин определяется обычно интуитивно, исходя из имеющихся у исследователя опыта и общих представлений.

4.3. Испытания Бернулли. В теории вероятностей имеет большое значение простая схема случайных экспериментов, называемая схемой Бернулли. На традиционном языке ее определение выглядит следующим образом: «испытаниями Бернулли называются независимые испытания с двумя исходами и с вероятностью успеха, не меняющейся от испытания к испытанию».

Для читателя, привыкшего к теоретико-множественному языку, это «определение» выглядит скорее как заклинание и требует перевода на привычный язык. Нужно, следовательно, ввести пространство элементарных событий.

В случае одного испытания (т. е. опыта) с двумя исходами (т. е. возможными результатами) пространство элементарных событий состоит из двух элементов. Один из этих элементов назовем «успехом» и будем обозначать единицей, другой назовем «неудачей» и будем обозначать нулем. Пусть вероятность успеха есть p : $P(1)=p$, тогда вероятность неудачи есть $q=1-p$: $P(0)=q=1-p$.

Интерес представляет, конечно, не один опыт, а несколько. Расшифруем выражение «независимые испытания». Согласно п. 2.3, соответствующее пространство элементарных событий является прямым произведением n (в случае n испытаний) экземпляров пространства

$$\Omega = (0, 1), \quad P(0) = p, \quad P(1) = q = 1 - p.$$

Следовательно, в случае n испытаний пространство элементарных событий Ω состоит из последовательностей

$$\omega = \underbrace{110 \dots 01}_n$$

нулей и единиц длины n , причем для получения вероятности $P(\omega)$ нужно каждой единице поставить в соответствие p , каждому нулю — q и полученные числа перемножить. Иными словами, если через $\mu(\omega)$ обозначить число единиц в последовательности ω , то получим

$$P(\omega) = p^{\mu(\omega)} q^{n-\mu(\omega)}.$$

Поэтому на современном языке следует дать
Определение 4.3. *Испытания Бернулли* — это

$$\Omega = (0, 1) \times (0, 1) \times \dots \times (0, 1) \quad (n \text{ раз}),$$

причем для любого $\omega \in \Omega$

$$P(\omega) = p^{\mu(\omega)} q^{n-\mu(\omega)}, \quad 0 \leq p \leq 1, \quad q = 1 - p.$$

Сравним достоинства и недостатки двух определений испытаний Бернулли. Старое определение не вполне ясно, но с его помощью можно узнавать, в каких конкретных ситуациях речь идет об испытаниях Бернулли. Испытаниями Бернулли будут, на-

пример, бросания монеты (герб — успех), стрельба в цель несколькими одинаково метких стрелков (попадание — успех), наблюдения за погодой, проводимые в данный день (скажем, 7 мая) каждого года (дождь — успех, нет дождя — неудача). Определению испытаний Бернулли не будут удовлетворять бросания поразному искривленных монет (от бросания к бросанию меняется вероятность успеха), стрельба в цель при наличии корректировки (нет независимости результатов отдельных выстрелов и постоянства вероятности успеха), наблюдения за погодой в последовательные дни одного года (нет независимости). В то же время определение 4.3 ясно и удобно для математических выводов, но, если строго им ограничиться, то оно совершенно не дает пути применения схемы Бернулли.

Выведем (разумеется, из определения 4.3) формулу для вероятности $P\{\mu = m\}$, обозначаемой еще $P_n(m)$, т. е. для вероятности того, что при испытаниях Бернулли будет ровно m успехов.

Теорема 4.4

$$P_n(m) = P\{\mu = m\} = C_n^m p^m q^{n-m}$$

Доказательство. Запишем цепочку равенств

$$P\{\mu = m\} = P\{\omega : \mu(\omega) = m\} = \sum_{\omega: \mu(\omega)=m} P(\omega) = p^m q^{n-m} N(m),$$

где через $N(m)$ обозначено число элементарных событий таких, что $\mu(\omega) = m$. Очевидно, $N(m) = C_n^m$ (число сочетаний по m элементов из n), так как элементарное событие, содержащее m единиц, вполне определяется указанием мест, на которых стоят единицы, в последовательности нулей и единиц длины n .

Введем случайные величины μ_k :

$$\mu_k = \begin{cases} 1, & \text{если в } k\text{-том испытании произошел успех,} \\ 0, & \text{» » » » произошла неудача.} \end{cases}$$

Иными словами, $\mu_k(\omega)$ есть символ, стоящий на k -том месте в последовательности ω . Очевидно, что

$$P\{\mu_1 = a_1, \mu_2 = a_2, \dots, \mu_n = a_n\} = p^{a_1} q^{1-a_1} p^{a_2} q^{1-a_2} \dots p^{a_n} q^{1-a_n},$$

где a_i принимают значения 0 и 1. Следовательно,

$$P\{\mu_1 = a_1, \mu_2 = a_2, \dots, \mu_n = a_n\} = \prod_{i=1}^n P\{\mu_i = a_i\},$$

т. е. случайные величины μ_1, \dots, μ_n независимы в совокупности. Очевидно, что $M_{\mu_k} = p$, $D_{\mu_k} = pq$.

Поскольку

$$\mu = \mu_1 + \mu_2 + \dots + \mu_n, \quad (4.3)$$

имеем

$$M\mu = n p, \quad D\mu = n p q.$$

З а м е ч а н и е. Распределение случайной величины μ (теорема 4.4) называется биномиальным. Оно определяется двумя числами (параметрами): n — число испытаний и p — вероятность успеха при одном испытании. В приложениях n обычно известно. Для определения p обычно считают приближенно, что $p \approx \frac{\mu}{n}$.

Очевидно, $\frac{\mu}{n}$ есть частота успеха. В силу равенства (4.3) и закона больших чисел, имеем $\frac{\mu}{n} \rightarrow p$ (вер.). Однако, как велико может быть различие между $\frac{\mu}{n}$ и p при данном конечном n , мы пока не знаем: это будет выяснено с помощью центральной предельной теоремы.

4.4. Теорема Пуассона. На практике обычно представляет интерес вероятность вида

$$P\{a \leq \mu \leq b\} = \sum_{a \leq m \leq b} P\{\mu = m\} = \sum_{a \leq m \leq b} C_n^m p^m q^{n-m}. \quad (4.4)$$

При больших n , a и b вычисление суммы (4.4) является очень сложным. Естественно желание составить таблицы для ее вычисления. Очевидно,

$$P\{a \leq \mu \leq b\} = P\{0 \leq \mu \leq b\} - P\{0 \leq \mu \leq a-1\}.$$

Поэтому достаточно составить таблицы для вычисления $P\{0 \leq \mu \leq b\}$. Эта вероятность зависит от n , p и b , а следовательно, таблицы должны иметь три входа. Поэтому таблицы имеют слишком большой объем и ими неудобно пользоваться.

Достаточно подробные таблицы на русском языке до сих пор не изданы. Из широко доступных изданий можно указать цитированную в § 1 книгу [21], где имеются краткие таблицы биномиального распределения, пригодные, однако, лишь для учебных целей. Хорошие приближенные выражения указаны в книге Л. Н. Большева и Н. В. Смирнова [4]. Но эта книга рассчитана на специалистов. Практически большую ценность имеют приближенные выражения для биномиальных вероятностей, полученные при различных предположениях о параметрах n и p . Таких приближений два. Соответствующие утверждения называются теоремой Муавра—Лапласа и теоремой Пуассона. Сейчас мы рассмотрим гораздо более простую (хотя полученную позже) теорему Пуассона.

Теорема Пуассона касается случая, когда число испытаний n велико, а вероятность успеха p мала, причем произведение np имеет порядок нескольких единиц. Грубое и не всегда верное правило для применения теоремы Пуассона состоит в том, что n должно быть порядка не менее нескольких десятков, а лучше сотен, а произведение np должно заключаться между 0 и 10. При больших np рекомендуется применять теорему Муавра—Лапласа. Однако при желании получить некоторую гарантированную точность следует обратиться к указанной книге Л. Н. Большева и Н. В. Смирнова, так как этой грубой рекомендации особенно доверять нельзя.

Выразить на математическом языке то утверждение, что некоторое число велико, а некоторое другое мало, можно только, используя понятие предельного перехода. Следовательно, величины n и p надо сделать переменными, стремящимися соответственно к ∞ и 0. Но в последовательности испытаний Бернулли вероятность успеха должна быть постоянной. Поэтому теорему Пуассона нельзя сформулировать с помощью одной последовательности испытаний Бернулли. Приходится рассмотреть последовательность серий испытаний Бернулли: в первой серии ($n=1$) имеется всего одно испытание с вероятностью успеха p_1 , число успехов в ней обозначим μ_1 (очевидно, либо $\mu_1=1$, либо $\mu_1=0$); далее, во второй серии ($n=2$) имеется 2 испытания, каждое с вероятностью успеха p_2 , и число успехов μ_2 , и т. д. В n -ной серии имеется n испытаний с вероятностью успеха p_n и μ_n успехов.

Теорема 4.5 (теорема Пуассона). Пусть при $n \rightarrow \infty$ $p_n \rightarrow 0$ таким образом, что $np_n \rightarrow \lambda$, где λ фиксированное неотрицательное число. Тогда для любого фиксированного $k=0, 1, 2, \dots$ при $n \rightarrow \infty$

$$P\{\mu_n = k\} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}.$$

Доказательство. Имеем $p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right)$,

$$\begin{aligned} P\{\mu_n = k\} &= C_n^k p_n^k (1 - p_n)^{n-k} = \\ &= \frac{n(n-1) \dots (n-k+1)}{k!} \left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right]^k \left[1 - \frac{\lambda}{n} - o\left(\frac{1}{n}\right) \right]^{n-k}. \end{aligned}$$

При $n \rightarrow \infty$ имеем

$$\begin{aligned} n(n-1) \dots (n-k+1) \left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right]^k &= \\ = \lambda^k \frac{n(n-1) \dots (n-k+1)}{n^k} (1 + o(1))^k &\rightarrow \lambda^k, \\ \left[1 - \frac{\lambda}{n} - o\left(\frac{1}{n}\right) \right]^{n-k} &\rightarrow e^{-\lambda}. \end{aligned}$$

Отсюда вытекает утверждение теоремы.

З а м е ч а н и е. Мы видим, что доказательство теоремы Пуассона почти тривиально. Выше было отмечено, что приближение биномиального закона, даваемое теоремой Пуассона, не является особенно хорошим. В чем же тогда значение этой теоремы? Дело в том, что, с точки зрения применений, математические теоремы бывают хорошими и плохими в следующем смысле: хорошие теоремы продолжают действовать, если даже нарушать их условия, а плохие сразу перестают быть верными при нарушении условий. Теорема Пуассона является в этом смысле хорошей и даже превосходной. Формально она относится к испытаниям Бернулли, но можно довольно сильно нарушать условия схемы Бернулли (т. е. допускать переменную вероятность успеха и даже не слишком сильную зависимость результатов отдельных испытаний), не нарушая окончательного вывода теоремы Пуассона.

Определение 4.4. Говорят, что случайная величина ξ , принимающая значения 0, 1, 2, ..., имеет *распределение Пуассона* с параметром λ , если

$$P\{\xi = k\} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Таким образом, при довольно слабых ограничениях число успехов в последовательности испытаний (не обязательно испытаний Бернулли) будет мало отличаться от распределения Пуассона. Это утверждение не удастся адекватно выразить на математическом языке, т. е. в виде теоремы, хотя бы потому, что неадекватно выражать приблизительное равенство, которым можно пользоваться на практике в терминах предельного перехода. Мы видим здесь один из примеров принципиального несовершенства нынешнего математического языка.

Более точные аппроксимации биномиального закона, упоминавшиеся выше, имеют то преимущество, что они более точны, и тот недостаток, что они более сложны. При отклонениях от схемы Бернулли они сразу теряют смысл (так обычно бывает с более сложными в математическом отношении фактами). В то же время простое распределение Пуассона обладает сравнительно универсальной применимостью. Последнее утверждение мы понимаем в том смысле, что если экспериментальные данные показывают, что закон Пуассона неприменим, в то время как, сообразно со здравым смыслом, он должен был бы действовать, то естественнее подвергнуть сомнению статистическую устойчивость наших данных, чем искать какой-то другой закон распределения. Следовательно, почти тривиальное, с математической точки зрения, доказательство теоремы 4.5 нужно рассматривать как эвристический прием, приводящий нас к одному из универсальных (в рамках применимости теории вероятностей) законов природы.

Пример 4.1. Пусть известно, что на выпечку 1000 сладких булочек с изюмом полагается 10 000 изюмин. Найти распределе-

ние числа изюмин в какой-то случайным образом выбранной булочке.

Решение. Рассмотрим следующую схему испытаний. Всего будет $n=10\,000$ испытаний (по числу изюмин). Испытание с номером k будет состоять в том, что мы определяем, попала ли изюмина с номером k в нашу случайно выбранную булочку (заметим, что на покупку булочки в магазине вполне можно смотреть как на случайный выбор). Тогда, поскольку всего булочек 1000, вероятность того, что k -тая изюмина попала именно в нашу булочку, есть $p=1/1000$. Испытания не вполне независимы, так как если в нашу булочку попали изюмины с номерами от 1 до нескольких тысяч, то она будет состоять целиком из изюма, так что остальные изюмины просто не смогут там поместиться. Однако ясно, что при хорошем перемешивании теста такой случай почти невозможен. Вообще, идеализация, в которой изюмины считаются точками, достаточно хороша для нашей задачи. Применяем поэтому распределение Пуассона с параметром

$$\lambda = np = 10\,000 \cdot \frac{1}{1000} = 10$$

(заметим, что λ есть среднее число изюмин, приходящихся на одну булочку). Следовательно,

$$P\{\mu = k\} = \frac{10^k}{k!} e^{-10}.$$

В частности $P\{\mu=0\}$, т. е. вероятность того, что нам достанется булочка вовсе без изюма, есть e^{-10} , что примерно есть $0,5 \cdot 10^{-4}$.

З а м е ч а н и е. Если случайная величина ξ подчиняется закону Пуассона, то

$$P\{\xi \leq m\} = \sum_{k=0}^m \frac{\lambda^k}{k!} e^{-\lambda}. \quad (4.5)$$

Выражение (4.5) зависит от двух параметров m и λ . Поэтому таблица для выражения (4.5) будет таблицей с двумя входами, что несравненно удобнее, чем таблица с тремя входами для биномиального закона. Хорошие таблицы для закона Пуассона имеются в уже цитированной книге Л. Н. Большева и И. В. Смирнова, а также в книге Я. Янко «Математико-статистические таблицы». Москва, Госстатиздат, 1961 (перевод с чешского языка).

Ясно, что пример 4.1 имеет весьма общий характер: можно говорить вместо изюмин в булочках, например, о числе бактерий в капле воды, взятой из хорошо перемешанного ведра, о числе больных данной (неинфекционной) болезнью в данном городском районе и т. д. Если мы предположим, что атомы радиоактивного вещества распадаются независимо друг от друга, причем в тече-

ние данного интервала времени распад отдельного атома происходит с вероятностью p , а всего к началу интервала времени было n атомов, то для числа распадов получим, естественно, распределение Пуассона с параметром $\lambda=np$.

В каких же случаях нельзя ожидать закона Пуассона? Одна из типичных ситуаций, когда он нарушается,— это сильная зависимость результатов отдельных испытаний. Например, если при возникновении аварии одного элемента оборудования почти наверняка выходят из строя связанные с ним элементы, то общее число вышедших из строя элементов не будет подчиняться закону Пуассона. Очевидно, то же самое справедливо для случая числа больных инфекционной болезнью, а также для случая числа покупателей, пришедших в магазин за дефицитным товаром, и т. д.

§ 5

СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ

При проведении научных исследований часто приходится рассуждать по схеме силлогизма. Напомним в удобной для нас форме это понятие. Пусть исследователь имеет о природе некоторого явления несколько гипотез. Одна из них — обозначим ее H_0 — может быть по какой-то причине для него особенно дорога или важна. Будем называть гипотезу H_0 основной. Кроме того, исследователь может допустить еще ряд гипотез H_1, H_2, \dots . Эти гипотезы мы будем называть альтернативными. Для проверки того, какая гипотеза — основная H_0 или одна из альтернативных H_1, H_2, \dots — на самом деле справедлива, исследователь прибегает к помощи опыта. Пусть в результате опыта может наступить или не наступить определенное событие S . Известно, однако, что если на самом деле верна гипотеза H_0 , то событие S наступить никак не может: если S имеет место, то неверна H_0 . Допустим, что в результате опыта S все же наступило. Тогда очевидно, что H_0 неверна. Такая схема рассуждений:

если S имеет место, то H_0 неверна

представляет собой один из видов силлогизма. Очевидно, что понятие силлогизма выделяет в формализованном виде некоторую — хотя и очень небольшую — часть того, что на самом деле происходит при научном исследовании.

Теория статистической проверки гипотез изучает примерно эту же часть научного мышления, но в условиях, когда имеется случайность (предполагается, конечно, статистически устойчивая случайность, для описания которой разумно пользоваться понятиями и методами теории вероятностей).

В частности, пусть известно не то, что при верной гипотезе H_0 событие S наступить вообще не может, а что его вероятность

$P(S/H_0)$ имеет смысл и мала (несмотря на внешнее сходство записи, $P(S/H_0)$ не имеет ничего общего с условной вероятностью, так как гипотезе H_0 мы не собираемся приписывать никакой вероятности — при нашем подходе она либо верна, либо неверна).

Ясно, что если первая посылка силлогизма верна «в большинстве случаев», то его заключение (при верной второй посылке) верно также «в большинстве случаев». Следовательно, если при наступлении S мы будем отвергать H_0 , то мы лишь изредка будем ошибаться. Поскольку аккуратный исследователь много раз проверяет свои выводы, то в конечном счете истина возобладает. Однако для того чтобы понять, почему нельзя вообще не ошибаться, надо несколько уточнить рассматриваемую ситуацию.

Итак, у нас имеется гипотеза H_0 о природе некоторого явления, которую мы по каким-то причинам выделяем и называем основной, противопоставляя ее множеству $\{H_\lambda\}$ альтернативных гипотез (λ может принимать значения 1, 2, ... или вещественные значения, а в принципе пробегать любое множество).

Далее, имеется опыт, результат которого x есть элемент некоторого множества X , называемого выборочным пространством. Например, если опыт состоит в пересчете каких-то предметов, то x — неотрицательное целое число, а $X = \{0, 1, 2, \dots\}$. Если же опыт состоит в проведении какого-то измерения, то часто естественно считать, что его результат x может быть любым вещественным числом, а X — множество всех вещественных чисел. В случае нескольких измерений x — вектор, а X — многомерное пространство.

Связь между гипотезами H_λ и результатом опыта x состоит в следующем. Предполагается, что в X выделен достаточно широкий класс подмножеств $A \subseteq X$ таких, что при любой верной гипотезе H_λ определены вероятности

$$P\{x \in A/H_\lambda\},$$

т. е. вероятности того, что результат опыта x попадет в A , если на самом деле верна гипотеза H_λ (предполагается, следовательно, соответствующая статистическая устойчивость).

Формально процесс проверки гипотезы H_0 состоит в том, что выбирается некоторое множество S (называемое критическим для гипотезы H_0) и делается опыт. Если результат опыта $x \in S$, то гипотеза H_0 отвергается. Посмотрим, каким условиям должно удовлетворять S .

Хорошо было бы, если бы $P\{x \in S/H_0\} = 0$ (тогда бы мы никогда не отвергали верную гипотезу), а $P\{x \in S/H_\lambda\} = 1$ при $\lambda \neq 0$ (тогда мы всегда бы отвергали H_0 , если на самом деле верна любая из гипотез H_λ , $\lambda \neq 0$). Однако в практически интересных случаях, для того чтобы $P\{x \in S/H_0\} = 0$, множество S должно быть пустым. Но тогда и $P\{x \in S/H_\lambda\} = 0$ для любого λ и вся процедура бесполезна.

Поэтому исследователю приходится, скрепя сердце, допускать ненулевые значения $P\{x \in S/H_0\}$. Единственное, что он может сделать,— выбрать заранее «уровень значимости», т. е. некоторое число $\alpha > 0$, и потребовать, чтобы

$$P\{x \in S/H_0\} \leq \alpha. \quad (5.1)$$

Если мы дорожим гипотезой H_0 и не хотим ее отвергнуть понапрасну, то α должно быть малым. Каким конкретно—довольно безразлично, поэтому можно уговориться выбирать одно из значений 0,05; 0,01 или 0,001, как обычно и делается. Наличие этих трех всеми признанных значений сокращает объем необходимых статистических таблиц. А поскольку никто не настаивает на том, что только эти значения α допустимы, нет и искушения бороться за равноправие других значений α .

Итак, сначала назначается α , затем выбирается S , удовлетворяющее (5.1), и, наконец, делается опыт. Очевидно, что $P\{x \in S/H_0\}$ (обозначаемая также через $P\{S/H_0\}$) есть вероятность напрасно отвергнуть H_0 (когда она верна). Такая ошибка называется ошибкой первого рода. Из (5.1) следует, что вероятность ошибки первого рода не превосходит уровня значимости α . Если $x \in S$, где S удовлетворяет (5.1), то говорят: «гипотеза H_0 отвергается на уровне значимости α ».

Если $x \in S$, то, казалось бы, следует сказать «гипотеза H_0 принимается». Но каждый статистик знает, что если гипотеза не отвергается одним способом, то, возможно, она будет отвергнута другим, и можно только сказать, что «гипотеза H_0 на уровне значимости α не отвергается». Но, конечно, основной интерес представляют те случаи, когда гипотеза отвергается или не отвергается одинаково на всех разумных уровнях значимости (указанных выше). Такие случаи бывают достаточно часто.

Кроме ошибки первого рода возможна еще ошибка второго рода, которая состоит в том, что гипотеза H_0 не отвергается, когда на самом деле она не верна, а верна одна из гипотез H_λ . Вероятность этой ошибки $\beta(\lambda)$ есть, очевидно,

$$\beta(\lambda) = P\{x \notin S/H_\lambda\} = 1 - P\{x \in S/H_\lambda\} = P\{\bar{S}/H_\lambda\}.$$

Функция $1 - \beta(\lambda) = P\{x \in S/H_\lambda\}$, равная вероятности отвергнуть гипотезу H_0 , если на самом деле верна гипотеза H_λ , называется *функцией мощности статистического критерия S* .

Как только что объяснялось, мы бы желали, чтобы $P\{S/H_0\} = 0$, а $P\{S/H_\lambda\} = 1$ при $\lambda \neq 0$. Иными словами, было бы хорошо, если бы функция мощности имела вид, показанный сплошной линией на рис. 5.1. На самом деле приходится удовлетворяться функцией мощности, имеющей вид вроде показанного на рис. 5.1 пунктиром.

Каким следует выбирать S ? Естественно, надо среди S , удовлетворяющих условию (5.1), выбрать такое \bar{S} , при котором

$$P\{\tilde{S}/H_\lambda\} = \max_S P\{S/H_\lambda\},$$

где \max берется по всем S , удовлетворяющим (5.1). Но $\tilde{S}=S_\lambda$, при котором достигается максимум, будет, вообще говоря, зависеть от λ . Если S_λ не зависит от λ , то $\tilde{S}=S_\lambda$ есть, очевидно, наилучший критерий. Существование наилучшего критерия — вещь

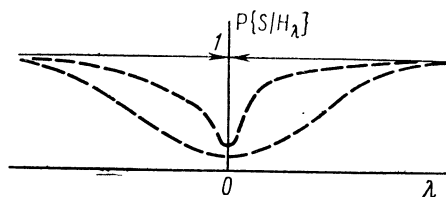


Рис. 5.1

довольно редкая. Если наилучший критерий есть, то он, во-первых, описан во всех учебниках математической статистики, а, во-вторых, к нему обычно нетрудно прийти, опираясь на соображения здравого смысла. Именно так обстоит дело в приводимом ниже примере.

Пример 5.1. Пусть нам известно, что при выпечке сладких булочек по государственному стандарту полагается на 1000 булочек 10 000 изюмин. Мы, однако, подозреваем, что изюм мог (по крайней мере, частично) разойтись по непредусмотренным законом каналам, и желаем это проверить. С этой целью мы покупаем одну булочку и пересчитываем в ней изюм. Если изюмин слишком мало, мы укрепляемся в своих подозрениях.

Попробуем формализовать эту процедуру с помощью только что введенных понятий.

Начнем с гипотез H_0 и H_λ . Выберем параметр λ следующим образом: λ принимает значения на отрезке $[0, 1]$ и обозначает долю украденного изюма. Гипотеза H_0 отвечает $\lambda=0$ и означает, что ничего не украдено.

Опыт состоит в том, что мы пересчитываем изюмины в купленной булочке. Выборочное пространство X (множество всех возможных исходов опыта) состоит из чисел $x=0, 1, \dots, 10\,000$, но нам удобнее считать его состоящим из всех чисел $0, 1, 2, \dots$ (считая, что значения $x > 10\,000$ встречаются с нулевой вероятностью).

Вероятности $P\{x = k/H_\lambda\}$ можно вычислить, как мы видели в предыдущем параграфе, применяя распределение Пуассона с параметром, равным среднему числу изюмин, приходящихся на одну булочку, т. е. при верной гипотезе H_λ , пуассоновский параметр есть

$$\frac{10\,000(1-\lambda)}{1\,000} = 10(1-\lambda).$$

Таким образом,

$$P\{x = k/H_\lambda\} = \frac{[10(1-\lambda)]^k}{k!} e^{-10(1-\lambda)}.$$

Перейдем к вопросу о выборе α . Вспомним, что α ограничивает сверху вероятность ошибочно отвергнуть нулевую гипотезу, т. е., в нашем случае, ошибочно обвинить невинного человека. Поэтому приемлемо лишь значение $\alpha=0$, но в этом случае вся функция мощности будет равна 0, т. е. мы не сможем обвинить и виноватого. Эти соображения показывают, что статистические методы вряд ли пригодны для решения вопроса о возбуждении обвинения (тем более, для решения вопроса о виновности). Посмотрим, что же все-таки юни могут дать. С этой целью испробуем два значения α : $\alpha=0,01$ и $\alpha=0,001$.

Главный вопрос — как выбирать критическое множество S . Ясно, что хищение изюма проявится в том, что изюма в булочке будет слишком мало. Иными словами, критическое множество S должно иметь вид

$$S = \{x : x \leq k\},$$

где k следует выбрать из условия

$$P\{S/H_0\} = P\{x \leq k/H_0\} \leq \alpha.$$

Задача о выборе k по заданному α очень легко решается с помощью таблиц распределения Пуассона. При $\alpha=0,01$ имеем $S = \{x : x \leq 3\}$, а при $\alpha=0,001$ имеем $S = \{x : x \leq 1\}$. Функция мощности дается следующей таблицей:

Параметр λ (доля украденного изюма)	Функция мощности $P\{x \in S/H_\lambda\}$ (вероятность возбуждения обвинения)		
	$\alpha = 0,01$ $S = \{x : x \leq 3\}$	$\alpha = 0,001$ $S = \{x : x \leq 1\}$	$\alpha = 0,15$ $S = \{x : x \leq 6\}$
0,0	0,010	0,00050	0,13
0,1	0,021	0,0013	
0,2	0,042	0,0031	0,76
0,3	0,082	0,0073	
0,4	0,15	0,017	
0,5	0,27	0,041	
0,6	0,43	0,092	
0,7	0,65	0,20	
0,8	0,86	0,41	1
0,9	0,98	0,74	
1,0	1	1	

Из таблицы видно, в частности, что если вероятность ложного обвинения ограничить сверху числом 0,001, то она на самом деле будет равна 0,00050. При этом того, кто украл половину изюма ($\lambda=0,5$), мы обвиним с вероятностью 0,041. Правда, за счет покупки и исследования нескольких булочек можно было бы приблизить функцию мощности к 1 при $\lambda \neq 0$, не увеличивая ее свыше 0,001 при $\lambda=0$. Однако ясно, что одними статистическими методами здесь не обойтись.

Нельзя ли все же извлечь из статистики некоторую пользу? Договоримся решать вопрос об обвинении не при помощи статистики, а при помощи прямого наблюдения. Но в таком случае, если допустить, что 80% всех работников честны и лишь 20% нечестны, то 80% рабочего времени «наблюдателя» будет потеряно впустую. Будем теперь проверять нашу гипотезу H_0 на совершенно ином уровне значимости $\alpha=0,15$, договорившись, что отбрасывание гипотезы не означает возбуждения обвинения, а лишь установление в соответствующем месте наблюдения. Как видно из таблицы, вероятность ошибки первого рода, т. е. напрасной посылки «наблюдателя» есть 0,13 и, таким образом, лишь $0,13 \cdot 0,80 = 10,4\%$ рабочего времени «наблюдателя» будет потеряно впустую. С другой стороны, если $\lambda=0,5$, то вероятность посылки наблюдателя и тем самым обнаружения хищения (при его совершении в следующий раз) равна 0,76, что вполне удовлетворительно.

Рассмотренное положение вообще характерно для применения статистических методов: не решая до конца научной или технической задачи, они позволяют ценой сравнительно небольших расходов наметить объект или план углубленного научного исследования. Таким образом, мы убедились в справедливости афоризма «статистике часто принадлежит первое слово, но никогда последнее».

§ 6

АКСИОМАТИКА КОЛМОГорова. ИНТЕГРАЛ ЛЕБЕГА

Хотя наблюдаемые в эксперименте величины могут быть измерены лишь с ограниченной точностью, все основные понятия науки связаны с идеей числового континуума. Таким образом, обычно считается, что возможными значениями одной измеряемой величины могут быть все действительные числа (принадлежащие некоторому интервалу); если измеряется несколько величин, то получающийся вектор может принимать любое значение из некоторой области евклидова пространства. По-видимому, континуум гораздо проще изучать, чем дискретные числовые множества. Соответственно, непрерывные модели проще и дают лучшее приближение к реальности.

Теория вероятностей не может держаться в стороне от этой общенаучной традиции. Следовательно, нужно допустить случайные величины, значениями которых могут быть все точки некоторого интервала или всей числовой прямой. Но если случайную величину $\xi = \xi(\omega)$ понимать как функцию от $\omega \in \Omega$, то Ω в таком случае не может быть счетным множеством (разумно рассматривать лишь однозначные функции, когда $\omega \in \Omega$ отвечает лишь одно число $\xi(\omega)$). Поэтому нельзя приписать каждой точке $\omega \in \Omega$ вероятность $P(\omega)$, отличную от нуля. Действительно, допустив, что все $P(\omega) > 0$, рассмотрим для каждого натурального n мно-
жество

$$\Omega_n = \left\{ \omega : \frac{1}{n+1} < P(\omega) \leq \frac{1}{n} \right\}.$$

Если каждое из этих множеств конечно, то $\Omega = \Omega_1 + \Omega_2 + \dots + \Omega_n + \dots$ не более чем счетно (хорошо известно, что объединение счетного числа не более чем счетных множеств — не более чем счетно). Если одно из множеств Ω_n бесконечно, то, очевидно, $\sum P(\omega) = \infty$, а следовательно, и $\sum P(\omega) = \infty$, $\omega \in \Omega_n$, в то время как было бы естественно требовать, чтобы $P(\Omega) = 1$. Поэтому, приписывая всем точкам из Ω (кроме, может быть, счетного числа) нулевые вероятности $P(\omega)$, мы должны каким-то образом научиться складывать из $\omega \in \Omega$ события ненулевой вероятности.

Это вполне возможно. Например, все точки отрезка $[0, 10]$ имеют нулевую длину, но из них прекрасно «складывается» отрезок длины 10. Однако определение длины отрезка идет не через определение длин различных точек и их континуального суммирования, а через сравнение длин различных отрезков с длиной отрезка единичной длины. По-видимому, континуальное суммирование нельзя определить корректно. В теории меры и теории вероятностей допускается лишь счетное суммирование.

6.1. Аксиоматика Колмогорова. В общей аксиоматике теории вероятностей сохраняется понятие множества элементарных событий Ω (которое не обязано быть счетным) и понятие события A как подмножества $\Omega : A \subseteq \Omega$. Однако не требуется, чтобы любое подмножество Ω было событием. Требуется лишь, чтобы теоретико-множественные операции, производимые над событиями в счетном числе, приводили опять к событиям. Множество всех рассматриваемых событий обычно обозначается буквой \mathfrak{B} . Мы потребуем, чтобы \mathfrak{B} удовлетворяло некоторой системе аксиом.

Аксиомы. 6.1. $\Omega \in \mathfrak{B}$, $\emptyset \in \mathfrak{B}$.

6.2. Если $A \in \mathfrak{B}$, то $\bar{A} \in \mathfrak{B}$.

6.3. Если $A_1, A_2, \dots, A_n, \dots \in \mathfrak{B}$, то

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \dots \cup A_n \cup \dots \in \mathfrak{B},$$

$$\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \dots \cap A_n \cap \dots = A_1 A_2 \dots A_n \dots \in \mathfrak{B}.$$

Множество \mathfrak{B} подмножеств Ω , удовлетворяющее аксиомам 6.1, 6.2 и 6.3, называется σ -алгеброй (сигма-алгеброй).

Кратко аксиомы 6.1, 6.2 и 6.3 можно выразить так: множество событий является σ -алгеброй подмножеств множества элементарных событий Ω .

Из понятий, связанных с σ -алгебрами множеств, нам понадобится только понятие наименьшей σ -алгебры, содержащей данные множества. Пусть даны множества B_α , являющиеся подмножествами Ω (где индекс α пробегает некое множество), которые, вообще говоря, σ -алгебры не образуют. Рассмотрим все σ -алгебры \mathfrak{B}_β , содержащие множества B_α (это значит, что все множества B_α являются элементами каждой σ -алгебры \mathfrak{B}_β). Такие \mathfrak{B}_β существуют: одним из примеров является σ -алгебра всех подмножеств Ω . Рассмотрим теперь

$$\mathfrak{M} = \bigcap_{\beta} \mathfrak{B}_\beta.$$

Очевидно, что

- 1) \mathfrak{M} — σ -алгебра.
- 2) все $B_\alpha \in \mathfrak{M}$,
- 3) если какая-нибудь σ -алгебра содержит все B_α , то \mathfrak{M} содержится в этой σ -алгебре (действительно, эта σ -алгебра совпадает с одной из \mathfrak{B}_β).

Свойства 1), 2) и 3) дают основание называть \mathfrak{M} наименьшей σ -алгеброй, содержащей данные множества B_α .

Пример 6.1. Рассмотрим эксперимент, состоящий в том, что на отрезок $[0, 1]$ случайным образом бросается точка. Опишем пространство элементарных событий, связанных с этим экспериментом. Ясно, что возможными его исходами являются точки отрезка $[0, 1]$. Поэтому $\Omega = [0, 1]$. Что же нужно считать σ -алгеброй событий?

Правило перевода. Событиями считаются те множества элементарных событий, которые естественно считать наблюдаемыми при эксперименте.

Ясно, что в примере с бросанием точки на отрезок модель выглядела бы крайне неуклюже, если бы нельзя было считать наблюдаемыми интервалы вида $0 \leq a < \omega < b \leq 1$, где ω — результат опыта. Итак, интервалы $\{\omega : a < \omega < b\}$ должны входить в σ -алгебру событий \mathfrak{B} . Отсюда вытекает, что в качестве \mathfrak{B} должна быть взята σ -алгебра, совпадающая с наименьшей σ -алгеброй, содержащей все интервалы, или еще более широкая σ -алгебра.

Определение 6.1. Наименьшая σ -алгебра, содержащая все интервалы, называется σ -алгеброй борелевских подмножеств отрезка $[0, 1]$.

Посмотрим, насколько богата σ -алгебра борелевских подмножеств:

1) Отдельные точки являются борелевскими подмножествами. Действительно, точку $c \in (0, 1)$ можно подставить в виде пересечения интервалов $\bigcap_{n=n_0}^{\infty} \left(c - \frac{1}{n}, c + \frac{1}{n}\right)$, где $n=1, 2, \dots$.

Поскольку интервалы $\left(c - \frac{1}{n}, c + \frac{1}{n}\right)$ при $n \geq n_0$ являются (борелевскими) подмножествами отрезка $[0, 1]$, из аксиомы 6.3 вытекает наше утверждение.

2) Множество рациональных точек является борелевским. Действительно, оно является счетным объединением отдельных точек.

3) Множество иррациональных точек является борелевским. Действительно, оно является дополнением к множеству рациональных точек.

4) Любое открытое множество является борелевским. Действительно, всякое открытое множество есть сумма счетного числа интервалов.

5) Любое замкнутое множество является борелевским (как дополнение к открытому).

6) Если $f(\omega)$ непрерывная функция на отрезке $[0, 1]$, то любое множество вида $\{\omega: f(\omega) \leq c\}$, где c — число, является борелевским (как замкнутое).

Довольно трудно привести пример множества, не являющегося борелевским. Ясно, что все множества, имеющие хоть какой-то практический интерес, являются борелевскими. Поэтому, если взять в качестве \mathfrak{B} борелевскую σ -алгебру, то все возможные прикладные потребности будут обеспечены с большим запасом. Действительно, практически неотличимо даже иррациональное число от рационального.

Введем теперь последние аксиомы аксиоматики Колмогорова. Они состоят в том, что каждому событию A (т. е. каждому $A \in \mathfrak{B}$) ставится в соответствие число $P(A)$, называемое вероятностью события A , удовлетворяющее условиям:

Аксиома 6.4. $P(\Omega) = 1$, $P(\emptyset) = 0$.

Аксиома 6.5. Если множества $A_1, A_2, \dots, A_n, \dots$ не пересекаются: $A_i A_j = \emptyset$, то¹

$$P\{A_1 + A_2 + \dots + A_n + \dots\} = \sum_{i=1}^{\infty} P(A_i). \quad (6.1)$$

Равенство (6.1) называется свойством *счетной аддитивности* (или *полной аддитивности*) вероятностей (или вероятностной меры) P .

¹ Знаком $+$ в данной книге обозначается сумма непересекающихся множеств.

Итак, общая модель Колмогорова состоит из трех элементов $(\Omega, \mathfrak{B}, P)$, где Ω — любое множество, \mathfrak{B} — некоторая σ -алгебра его подмножеств, P — счетно-аддитивная вероятностная мера, определенная на \mathfrak{B} .

Продолжим рассмотрение примера 6.1. Нам осталось ввести вероятность на σ -алгебре \mathfrak{B} борелевских подмножеств отрезка $[0, 1]$. Для этого мы должны обратиться к толкованию слов «на отрезок случайным образом бросается точка». Их обычно толкуют так, что вероятность попадания точки на интервал $(a, b) \subseteq [0, 1]$ не зависит от положения этого интервала, а лишь от его длины $|a - b|$. Посмотрим, что получится, если в качестве вероятности $P\{\omega: a < \omega \leq b\}$ взять длину $|a - b|$. Во-первых, $P(\Omega) = 1$, $P(\emptyset) = 0$. Во-вторых, если (a, b) и (c, d) не пересекаются, то $P\{(a, b) + (c, d)\}$, т. е. длина суммы двух интервалов, очевидно, равна сумме $P\{(a, b)\} + P\{(c, d)\}$.

Трудности, которые нас здесь встречают, состоят в том, что, во-первых, мы пока не знаем, что такое длина произвольного борелевского множества, а, во-вторых, не умеем обосновать равенство (6.1). Эти трудности связаны со сложностью борелевских множеств. Но зачем нам понадобились борелевские множества? Они возникли из простых и естественных интервалов после применения к ним в счетном числе теоретико-множественных операций. Зачем, однако, требовать, чтобы к событиям можно было применять теоретико-множественные операции в счетном, а не только в конечном числе? (Ведь счетное (бесконечное) число событий нельзя реально наблюдать.)

Ответ на последний вопрос состоит в том, что при допущении счетного числа теоретико-множественных операций получается особенно прозрачная и красивая математическая теория. Но в таком случае дело математиков сводить концы с концами, т. е. показывать в разных конкретных случаях, что предложенная модель со счетным числом операций в самом деле непротиворечива.

В применении к борелевским множествам отрезка вопрос вполне решается с помощью введения понятия меры Лебега: именно, показано, что на некотором широком классе множеств, содержащем борелевские множества, можно определить вероятностную меру P , во всех практически интересных случаях совпадающую с суммарной длиной составляющих множество интервалов, причем так, чтобы выполнять равенство (6.1).

Читатель может при желании ознакомиться с полным изложением теории меры в книге А. Н. Колмогорова и С. В. Фомина [13]. Однако естествоиспытателю достаточно знать, что математики умеют каким-то образом справляться с построением непротиворечивых моделей, удовлетворяющих аксиомам 6.1—6.5, причем в результате их усилий σ -алгебра \mathfrak{B} (в практически важных случаях) с избытком достаточна для применений. Интересно

отметить, что эта σ -алгебра появилась именно для преодоления математических трудностей в общей теории интегрирования.

На первый взгляд кажется, что можно было потребовать, чтобы любому подмножеству A множества Ω отвечала вероятность $P(A)$, удовлетворяющая (6.1). Однако в наиболее интересных случаях этого нельзя добиться: если бы мы постулировали (6.1) для любых $A \in \Omega$, то мы бы построили теорию, не имеющую практически важных приложений. Поэтому приходится ограничивать класс множеств A σ -алгеброй \mathfrak{B} , которая, впрочем, в тех случаях, которые мы будем рассматривать, заведомо включает все имеющие какой-то практический смысл множества.

Определение 6.2. Говорят, что имеется задача на *геометрическую вероятность*, если множество элементарных событий Ω есть область в евклидовом пространстве, имеющая конечный объем, и вероятность любой ее «подобласти» A задается формулой

$$P(A) = \frac{V(A)}{V(\Omega)},$$

где V обозначает объем (в двумерном случае — площадь). Конечно, математик заметил бы, что речь идет не о любом подмножестве A области Ω , а только о борелевском¹, и не об объеме, а о лебеговской мере объема.

Пример 6.2. Рассмотрим теперь следующую «задачу о встрече». Два человека A и B договорились встретиться в условленном месте, причем каждый из них приходит туда независимо от другого в случайный момент между 12 и 13 часами, ждет 20 минут ($= \frac{1}{3}$ часа) и, если второй за это время не появился, уходит. Найти вероятность того, что встреча произойдет.

В рамках общей теории следует сначала перевести задачу на язык пространства элементарных событий. Пусть x — момент прихода A на условленное место; $12 \leq x \leq 13$, y — момент прихода B , $12 \leq y \leq 13$. В качестве Ω можно, очевидно, взять квадрат на плоскости:

$$\Omega = \{(x, y): 12 \leq x \leq 13, 12 \leq y \leq 13\}.$$

В качестве \mathfrak{B} стандартным образом берем борелевские множества квадрата (практически это все фигуры на плоскости, так как каждая фигура есть сумма счетного числа прямоугольников, а прямоугольники — по определению, борелевские множества). Для введения $P(C)$ сначала введем ее на прямоугольниках. Если

¹ В многомерном случае борелевскими подмножествами называются элементы наименьшей σ -алгебры, содержащей все параллелепипеды:

$$\{x = (x_1, \dots, x_n): a_1 < x_1 < b_1, \dots, a_n < x_n < b_n\},$$

где $a_i < b_i$, а в остальном a_i и b_i — любые числа.

$$C = \{(x, y) : a < x < b, c < y < d\},$$

то

$$P(C) = P\{a < x < b, c < y < d\} = P\{a < x < b\} \cdot P\{c < y < d\},$$

поскольку в условии задачи упомянуто о независимости моментов прихода x и y . Но, как и в примере 6.1, слова «случайный момент» следует понимать следующим образом:

$$P\{a < x < b\} = |a - b|, \quad P\{c < y < d\} = |c - d|.$$

Следовательно, $P(C)$ есть площадь прямоугольника. Любая фигура складывается из (счетного числа) прямоугольников. Поэтому естественно считать, что для любого C имеем $P(C) = V(C)$, где V обозначает площадь (площадь всего Ω есть 1). Вероятность встречи есть вероятность следующего события C :

$$C = \{(x, y) : |x - y| \leq \frac{1}{3}\}.$$

Легко вычислить, что

$$P(C) = V(C) = \frac{5}{9}.$$

6.2. Случайные величины. События, т. е. элементы \mathfrak{B} , называются еще «измеримыми множествами» (это название связано с тем, что только элементам \mathfrak{B} приписывается вероятность, т. е. они входят в область определения вероятностной меры P). Измеримость есть (часто довольно далекое) отражение возможности наблюдать событие A в эксперименте: идея состоит в том, что те события $A \subseteq \Omega$ имеют вероятность $P(A)$, для которых при каждом повторении эксперимента мы можем сказать, наступило событие A или не наступило, а следовательно, определить вероятность по частоте. Однако требования математической модели (аксиомы 6.1—6.3) заставляют нас считать измеримыми не только те события, которые в самом деле можно наблюдать, но и любые элементы наименьшей содержащей их σ -алгебры (см. пример 6.1). Поэтому практическая интерпретация полученных с помощью модели результатов требует осторожности.

Например, в примере 6.1 вероятность того, что наудачу брошенная точка окажется рациональной, есть нуль, так как множество рациональных точек счетно, а вероятность попадания в любую точку равна нулю. Следовательно, наудачу брошенная точка с вероятностью 1 иррациональна. Однако вряд ли этот математический факт может иметь прикладную интерпретацию.

Мы сохраняем определение случайной величины ξ как функции $\xi(\omega)$. Однако естественно потребовать, чтобы событие $\{\omega : a < \xi(\omega) < b\}$, где a и b — числа, было наблюдаемым, т. е. элементом \mathfrak{B} .

Определение 6.3. *Случайной величиной* называется измеримая функция, определенная на Ω , т. е. функция $\xi(\omega)$, обладающая свойством¹

$$\{\omega : a < \xi(\omega) < b\} = \xi^{-1}\{(a, b)\} \in \mathfrak{B}$$

для любых чисел $a < b$.

Теорема 6.1. *Если ξ — случайная величина, то полный прообраз $\xi^{-1}(B)$ любого борелевского подмножества прямой B есть элемент \mathfrak{B} .*

Доказательство. Заметим, что операция взятия полного прообраза перестановочна с теоретико-множественными операциями:

$$\xi^{-1}(\bar{A}) = \overline{\xi^{-1}(A)}, \quad (6.2)$$

$$\xi^{-1}\left(\bigcup_{\alpha} A_{\alpha}\right) = \bigcup_{\alpha} \xi^{-1}(A_{\alpha}), \quad (6.3)$$

$$\xi^{-1}\left(\bigcap_{\alpha} A_{\alpha}\right) = \bigcap_{\alpha} \xi^{-1}(A_{\alpha}), \quad (6.4)$$

где α пробегает любое (не обязательно счетное) множество.

Обозначим через \mathfrak{M} множество таких подмножеств $B \subseteq (-\infty, \infty)$, для которых $\xi^{-1}(B) \in \mathfrak{B}$. В силу определения 6.3, \mathfrak{M} содержит все интервалы.

В силу равенств (6.2) — (6.4) и аксиом 6.1—6.3, \mathfrak{M} есть σ -алгебра.

Поскольку σ -алгебра борелевских подмножеств прямой есть наименьшая σ -алгебра, содержащая все интервалы, то \mathfrak{M} содержит все борелевские множества, что и требовалось доказать.

Замечание 1. В дальнейшем множества вида: $\{\omega : a < \xi(\omega) < b\}$, $\{\omega : \xi(\omega) < b\}$ и т. д. сокращенно обозначаются в виде $\{a < \xi < b\}$, $\{\xi < b\}$ и т. д.

Замечание 2. Поскольку

$$\{\omega : a < \xi(\omega) < b\} = \{a < \xi < b\} = \{\xi < b\} \cap \{\xi \leq a\}$$

и

$$\{\xi \leq a\} = \bigcap_{n \geq 1} \left\{ \xi < a + \frac{1}{n} \right\},$$

то для выполнения формулировки определения 6.3 достаточно потребовать, чтобы для любого b

$$\{\omega : \xi(\omega) < b\} = \{\xi < b\} \in \mathfrak{B}.$$

6.3. Интеграл Лебега. Сейчас мы укажем простой и общий способ построения понятия интеграла от измеримой функции, определенной на пространстве с мерой.

¹ Через ξ^{-1} обозначается полный прообраз отображения ξ , отображающего Ω в множество вещественных чисел $(-\infty, \infty)$.

Определение 6.4. Функция $\xi(\omega)$, определенная на пространстве Ω , называется *простой*, если пространство Ω можно представить в виде счетной суммы

$$\Omega = A_1 + A_2 + \dots + A_n + \dots, \quad A_i A_j = \emptyset$$

попарно непересекающихся множеств $A_i \in \mathfrak{B}$ причем на каждом A_i функция $\xi(\omega)$ принимает постоянное значение:

$$\xi(\omega) = c_i, \text{ если } \omega \in A_i.$$

Определение 6.5. Индикатором множества A называется функция

$$I_A(\omega) = \begin{cases} 1, & \text{если } \omega \in A, \\ 0, & \text{если } \omega \notin A. \end{cases}$$

Очевидно, простая функция $\xi(\omega)$ представима в виде

$$\xi(\omega) = \sum_{i=1}^{\infty} c_i I_{A_i}(\omega).$$

Определение 6.6. Интегралом $\int_{\Omega} \xi(\omega) P(d\omega)$ от простой функции $\xi(\omega)$ по мере P называется выражение

$$\int_{\Omega} \xi(\omega) P(d\omega) = \sum_{i=1}^{\infty} c_i P(A_i),$$

если ряд в правой части сходится абсолютно.

Замечание 1. Мы видим, что интеграл Лебега $\int_{\Omega} \xi(\omega) P(d\omega)$ является для простых функций (случайных величин) $\xi(\omega)$ аналогом математического ожидания.

Теорема 6.2. Если простая функция $\xi(\omega)$ допускает кроме представления (6.5) другое представление

$$\xi(\omega) = \sum_{j=1}^{\infty} d_j I_{B_j}(\omega),$$

где $B_j B_k = \emptyset$ при $j \neq k$, $\Omega = B_1 + B_2 + \dots + B_n + \dots$, то

$$\sum_{i=1}^{\infty} c_i P(A_i) = \sum_{j=1}^{\infty} d_j P(B_j),$$

т. е. значение интеграла Лебега зависит только от самой функции $\xi(\omega)$, но не от выбора ее представления в виде (6.5).

Доказательство. Положим $D_{ij} = A_i B_j$ и заметим, что на множестве D_{ij} значение f_{ij} функции $\xi(\omega)$ равно сразу c_i и d_j . Поэтому имеем (в силу 6.1)

$$\begin{aligned}\sum_{i,j=1}^{\infty} f_{ij} \mathbf{P}(D_{ij}) &= \sum_{i=1}^{\infty} \left(\sum_{j=1}^{\infty} f_{ij} \mathbf{P}(D_{ij}) \right) = \\ &= \sum_{i=1}^{\infty} \left(c_i \sum_{j=1}^{\infty} \mathbf{P}(A_i B_j) \right) = \sum_{i=1}^{\infty} c_i \mathbf{P}(A_i),\end{aligned}$$

так как $A_i = A_i B_1 + A_i B_2 + \dots + A_i B_j + \dots$ и члены абсолютно сходящегося ряда можно суммировать в любом порядке. Аналогично доказывается, что

$$\sum_{j=1}^{\infty} d_j \mathbf{P}(B_j) = \sum_{i,j=1}^{\infty} f_{ij} \mathbf{P}(D_{ij}).$$

Теорема доказана.

Теорема 6.3. Если $\xi(\omega)$ и $\eta(\omega)$ — простые функции, причем существуют интегралы

$$\int_{\Omega} \xi(\omega) \mathbf{P}(d\omega) \quad \text{и} \quad \int_{\Omega} \eta(\omega) \mathbf{P}(d\omega),$$

то $\xi(\omega) + \eta(\omega)$ — простая функция и

$$\int_{\Omega} (\xi(\omega) + \eta(\omega)) \mathbf{P}(d\omega) = \int_{\Omega} \xi(\omega) \mathbf{P}(d\omega) + \int_{\Omega} \eta(\omega) \mathbf{P}(d\omega).$$

Доказательство. Теорема очевидна, если

$$\xi(\omega) = \sum_{i=1}^{\infty} c_i I_{A_i}(\omega), \quad \eta(\omega) = \sum_{j=1}^{\infty} d_j I_{B_j}(\omega),$$

т. е. ξ и η являются линейными комбинациями индикаторов одних и тех же множеств.

Если $\xi(\omega)$ и $\eta(\omega)$ представлены в виде линейной комбинации индикаторов множеств $A_1, A_2, \dots, A_n, \dots$ и $B_1, B_2, \dots, B_m, \dots$, то представим обе функции в виде линейной комбинации индикаторов множеств $D_{ij} = A_i B_j$ как это только что делалось при доказательстве теоремы 6.2.

Теорема 6.4. Если $\xi(\omega)$ — простая функция, то

$$\left| \int_{\Omega} \xi(\omega) \mathbf{P}(d\omega) \right| \leq \sup_{\omega \in \Omega} |\xi(\omega)|.$$

Действительно,

$$\left| \sum_{i=1}^{\infty} c_i \mathbf{P}(A_i) \right| \leq \sup_i |c_i| \cdot \sum_{i=1}^{\infty} \mathbf{P}(A_i) = \sup_i |c_i| = \sup_{\omega \in \Omega} |\xi(\omega)|.$$

Следствие. Если последовательность простых случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$ равномерно сходится к случайной величине ξ , то последовательность интегралов

$$\int_{\Omega} \xi_1(\omega) \mathbf{P}(d\omega), \int_{\Omega} \xi_2(\omega) \mathbf{P}(d\omega), \dots, \int_{\Omega} \xi_n(\omega) \mathbf{P}(d\omega) \dots$$

фундаментальна в смысле Коши.

Действительно,

$$\begin{aligned} & \left| \int_{\Omega} \xi_n(\omega) \mathbf{P}(d\omega) - \int_{\Omega} \xi_m(\omega) \mathbf{P}(d\omega) \right| = \\ & = \left| \int_{\Omega} (\xi_n(\omega) - \xi_m(\omega)) \mathbf{P}(d\omega) \right| \leq \sup_{\omega \in \Omega} |\xi_n(\omega) - \xi_m(\omega)| \rightarrow 0 \end{aligned}$$

при $n, m \rightarrow \infty$.

Лемма 6.1. Для любой измеримой функции $\xi(\omega)$ существует последовательность $\xi_n(\omega)$ простых функций, сходящаяся к $\xi(\omega)$ равномерно.

Доказательство. Положим для каждого натурального n и целого k

$$\begin{aligned} A_k^{(n)} &= \left\{ \omega : \frac{k}{n} \leq \xi(\omega) < \frac{k+1}{n} \right\} = \\ &= \left\{ \frac{k}{n} \leq \xi < \frac{k+1}{n} \right\} = \xi^{-1} \left\{ \left[\frac{k}{n}, \frac{k+1}{n} \right) \right\} \in \mathfrak{B}. \end{aligned}$$

Положим

$$\xi_n(\omega) = \sum_{k=-\infty}^{\infty} \frac{k}{n} I_{A_k^{(n)}}(\omega).$$

Тогда

$$|\xi_n(\omega) - \xi(\omega)| \leq \frac{1}{n}.$$

Определение 6.7. Пусть $\xi_n(\omega)$ — любая последовательность простых функций, сходящаяся к случайной величине $\xi(\omega)$ равномерно. Положим

$$\int_{\Omega} \xi(\omega) \mathbf{P}(d\omega) = \lim_{n \rightarrow \infty} \int_{\Omega} \xi_n(\omega) \mathbf{P}(d\omega).$$

В силу следствия из теоремы 6.4 и леммы 6.1, определение корректно. Очевидно, что все интегралы

$$\int_{\Omega} \xi_n(\omega) \mathbf{P}(d\omega),$$

начиная с достаточно большого n , одновременно существуют или не существуют. В последнем случае говорят, что $\int_{\Omega} \xi(\omega) \mathbf{P}(d\omega)$ не существует.

Теорема 6.5. Пусть $\xi(\omega)$ и $\eta(\omega)$ — случайные величины и интегралы $\int_{\Omega} \xi(\omega) \mathbf{P}(d\omega)$ и $\int_{\Omega} \eta(\omega) \mathbf{P}(d\omega)$ существуют. Тогда существует

$$\int_{\Omega} (\xi(\omega) + \eta(\omega)) \mathbf{P}(d\omega) = \int_{\Omega} \xi(\omega) \mathbf{P}(d\omega) + \int_{\Omega} \eta(\omega) \mathbf{P}(d\omega).$$

Доказательство непосредственно вытекает из определения 6.7 и теоремы 6.3.

З а м е ч а н и е. Используя последовательность $\xi_n(\omega)$, определенную в лемме 6.1, получим

$$\int_{\Omega} \xi_n(\omega) \mathbf{P}(d\omega) = \lim_{n \rightarrow \infty} \sum_{k=-\infty}^{\infty} \frac{k}{n} \mathbf{P} \left\{ \frac{k}{n} \leq \xi < \frac{k+1}{n} \right\}.$$

Сравним интеграл Лебега и интеграл Римана для функций, определенных на отрезке $[0, 1]$. Если функция $f(x)$ непрерывна на отрезке $[0, 1]$, то, разбивая отрезок на достаточно мелкие части $A_i^{(n)} = [x_i^{(n)}, x_{i+1}^{(n)})$, получим, что последовательность

$$f_n(x) = \sum_i f(x_i^{(n)}) I_{A_i^{(n)}}(x)$$

равномерно сходится к функции $f(x)$.

Понимая под $\mathbf{P}(A_i^{(n)})$ длину интервала $A_i^{(n)}$, имеем интеграл Лебега:

$$\begin{aligned} (L) \int_0^1 f(x) \mathbf{P}(dx) &= \lim_{n \rightarrow \infty} (L) \int_0^1 f_n(x) \mathbf{P}(dx) = \\ &= \lim_{n \rightarrow \infty} \sum_i f(x_i^{(n)}) \mathbf{P}(A_i^{(n)}). \end{aligned}$$

Однако последняя сумма есть риманова сумма для риманова интеграла $(R) \int_0^1 f(x) dx$. Следовательно, интеграл Лебега равен интегралу Римана (если подынтегральная функция непрерывна и под мерой \mathbf{P} понимается длина).

Во всех случаях, когда интеграл Лебега можно практически вычислить, он сводится либо к суммированию ряда (как в определении 6.6), либо к вычислению интеграла Римана.

В теории вероятностей интеграл Лебега служит для выражения в единой форме всех форм понятия математического ожидания, которые исторически сложились в этой науке.

Определение 6.8. Математическим ожиданием $M\xi$ случайной величины ξ называется

$$M\xi = \int_{\Omega} \xi(\omega) P(d\omega)$$

(если этот интеграл не существует, т. е. расходится интеграл $\int_{\Omega} |\xi(\omega)| P(d\omega)$, то говорят, что случайная величина не имеет математического ожидания).

Очевидно, что для любого числа c

$$M(c\xi) = cM\xi$$

(доказательство предоставляется читателю).

Следовательно, из теоремы 6.5 вытекает следующая

Теорема 6.6. Пусть случайные величины ξ и η имеют математические ожидания $M\xi$ и $M\eta$, а a и b — любые числа. Тогда

$$M(a\xi + b\eta) = aM\xi + bM\eta.$$

Эта теорема является аналогом теоремы 3.2, установленной для дискретных случайных величин. В следующем § 7 мы установим аналоги для всего содержания § 3.

§ 7

РАСПРЕДЕЛЕНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН

7.1. Определения и основные теоремы. При рассмотрении дискретных случайных величин отмечалось, что, как правило, сама случайная величина $\xi(\omega)$ как функция от ω не является наблюдаемой в эксперименте. Поэтому практическое значение имеет лишь распределение случайной величины ξ . Мы переходим к введению этого понятия в общем случае.

Замечание. Наше изложение не будет зависеть от того, является ли ξ вещественной случайной величиной, или комплексной, или, наконец, вектором произвольной конечной размерности.

Дадим следующее

Определение 7.1. Отображение $\xi: \Omega \rightarrow R^n$ множества элементарных событий Ω в евклидово n -мерное пространство R^n называется *векторной случайной величиной*, если

$$\xi^{-1}(B) \in \mathfrak{B}$$

для любого борелевского $B \subseteq R^n$.

(Напомним, что борелевскими подмножествами n -мерного евклидова пространства называются элементы наименьшей σ -алгебры, содержащей все параллелепипеды.)

Точка n -мерного евклидова пространства R^n будет обозначаться через x независимо от размерности. Иногда мы будем явно выписывать координаты, например, $\xi(\omega) = (\xi_1(\omega), \dots, \xi_n(\omega))$ или $x = (x_1, \dots, x_n)$.

Определение 7.2. Мера μ_ξ , определенная на борелевских подмножествах R^n равенством

$$\mu_\xi(B) = P\{\omega: \xi(\omega) \in B\} = P\{\xi \in B\} = P\{\xi^{-1}(B)\},$$

называется *распределением* случайной величины ξ .

Определение 7.3. Функция $f(x)$ точки $x \in R^n$, отображающая R^n в множество вещественных чисел $(-\infty, \infty)$, называется *измеримой по Борелю*, если для любого борелевского $C \subseteq (-\infty, \infty)$ множество $\{x: f(x) \in C\} = f^{-1}(C)$ является борелевским в R^n .

Практически все функции являются измеримыми по Борелю. Докажем, например, что любая непрерывная функция $f(x)$ измерима по Борелю. Рассмотрим класс K множеств $C \subseteq (-\infty, \infty)$ таких, что $f^{-1}(C)$ есть борелевское подмножество R^n . Класс K содержит все открытые множества: если C открыто, то $f^{-1}(C)$ тоже открыто в силу непрерывности f , но все открытые подмножества R^n являются борелевскими¹. Ясно также (см. аналогичное рассуждение в доказательстве теоремы 6.1), что класс K замкнут относительно теоретико-множественных операций, производимых в счетном числе. Следовательно, класс K содержит все борелевские множества.

Теорема 7.1. Пусть $\xi = \xi(\omega)$ (векторная) случайная величина, f — измеримая по Борелю функция. Тогда суперпозиция двух функций

$$f(\xi) = f(\xi(\omega))$$

является случайной величиной.

Доказательство. Пусть C — борелевское подмножество $(-\infty, \infty)$. Имеем

$$\{\omega: f(\xi(\omega)) \in C\} = \{\omega: \xi(\omega) \in f^{-1}(C)\}.$$

Но $f^{-1}(C)$ — борелевское подмножество R^n . Следовательно

$$\{\omega: \xi(\omega) \in f^{-1}(C)\} = \xi^{-1}(f^{-1}(C)) \in \mathfrak{B},$$

что и требовалось доказать.

¹ Для доказательства заметим, что а) объединение счетного числа параллелепипедов есть борелевское множество и что б) всякое открытое множество B можно представить в виде объединения возрастающей последовательности множеств $A_1 \subseteq A_2 \subseteq \dots \subseteq A_m \subseteq \dots$, где каждое A_m является суммой не более чем счетного числа параллелепипедов (разбивая пространство R^n на параллелепипеды диаметром $< 1/m$ и относя к A_m все такие параллелепипеды, содержащиеся в B).

Определение 7.4. Интеграл

$$\int_{R^n} f(x) \mu_{\xi}(dx)$$

понимается как введенный в предыдущем параграфе интеграл Лебега относительно меры μ_{ξ} в R^n .

Докажем теперь теорему.

Теорема 7.2. В условиях теоремы 7.1

$$M f(\xi(\omega)) = \int_{R^n} f(x) \mu_{\xi}(dx),$$

причем левая и правая части или одновременно существуют или не существуют.

Доказательство. По определению математического ожидания, имеем

$$\begin{aligned} M f(\xi(\omega)) &= \int_{\Omega} f(\xi(\omega)) P(d\omega) = \\ &= \lim_{n \rightarrow \infty} \sum_{k=-\infty}^{\infty} \frac{k}{n} P \left\{ \omega : \frac{k}{n} \leq f(\xi(\omega)) < \frac{k+1}{n} \right\} = \\ &= \lim_{n \rightarrow \infty} \sum_{k=-\infty}^{\infty} \frac{k}{n} P \left\{ \omega : \xi(\omega) \in f^{-1} \left(\left[\frac{k}{n}, \frac{k+1}{n} \right) \right) \right\}. \end{aligned}$$

С другой стороны,

$$\begin{aligned} \int_{R^n} f(x) \mu_{\xi}(dx) &= \lim_{n \rightarrow \infty} \sum_{k=-\infty}^{\infty} \frac{k}{n} \mu_{\xi} \left\{ x : \frac{k}{n} \leq f(x) < \frac{k+1}{n} \right\} = \\ &= \lim_{n \rightarrow \infty} \sum_{k=-\infty}^{\infty} \frac{k}{n} \mu_{\xi} \left\{ f^{-1} \left(\left[\frac{k}{n}, \frac{k+1}{n} \right) \right) \right\}. \end{aligned}$$

В силу определения меры μ_{ξ} имеем

$$P \{ \omega : \xi(\omega) \in f^{-1}(C) \} = \mu_{\xi} \{ f^{-1}(C) \},$$

если множество $f^{-1}(C)$ борелевское. В силу измеримости по Борелю функции f множество $f^{-1} \left(\left[\frac{k}{n}, \frac{k+1}{n} \right) \right)$, т. е. полный прообраз интервала $\left[\frac{k}{n}, \frac{k+1}{n} \right)$, является борелевским. Поэтому интегралы $\int_{\Omega} f(\xi(\omega)) P(d\omega)$ и $\int_{R^n} f(x) \mu_{\xi}(dx)$ являются пределами равных между

собой сумм, следовательно, существуют одновременно, а если существуют, то совпадают. Теорема доказана.

З а м е ч а н и е. Таким образом, $Mf(\xi)$ выражается через распределение μ_ξ . По определению 7.2, $\mu_\xi(B) = P\{\xi \in B\}$. Таким образом, с известными оговорками (подобными тем, которые были сделаны в § 6.2 относительно наблюдаемости событий) $Mf(\xi)$ выражается через наблюдаемые величины. Например, полагая в случае размерности n , равной единице, $f(x) = x$, получим: если ξ одномерная случайная величина, то

$$M\xi = \int_{-\infty}^{\infty} x \mu_\xi(dx).$$

Аналогично,

$$D\xi = M(\xi - M\xi)^2 = \int_{-\infty}^{\infty} (x - M\xi)^2 \mu_\xi(dx).$$

Таким образом, для вычисления $Df(\xi)$ и $Mf(\xi)$ нам нужно научиться вычислять интегралы вида

$$\int_{R^n} f(x) \mu_\xi(dx).$$

7.2. Частные виды распределений. О мере μ_ξ мы знаем пока только, что это есть некоторая мера, определенная на борелевских множествах. Практически представляют интерес лишь распределения μ_ξ некоторых частных видов. Мы сейчас их опишем.

1) **Дискретный случай.** Если в пространстве R^n имеется не более чем счетное множество точек $x_1, x_2, \dots, x_m, \dots$, на котором сосредоточена мера μ_ξ (это означает, что $\mu_\xi\{(x_1, x_2, \dots, x_m, \dots)\} = 0$, где черта обозначает дополнение), то распределение μ_ξ называется дискретным. В этом случае μ_ξ полностью характеризуется мерами $\mu_\xi(x_i)$, приписанными отдельным точкам x_i . Действительно, для любого борелевского множества C , обозначая через Q множество $\{x_1, \dots, x_m, \dots\}$, имеем

$$\mu_\xi(C) = \mu_\xi(CQ) + \mu_\xi(C\bar{Q}).$$

Но так как $\mu_\xi(\bar{Q}) = 0$, то и $\mu_\xi(C\bar{Q}) = 0$, а так как Q счетно, то

$$\mu_\xi(CQ) = \sum_{x \in CQ} \mu_\xi(x) = \sum_{x_i \in C} \mu_\xi(x_i).$$

Случайная величина ξ устроена следующим образом: $P\{\xi \in \bar{Q}\} = \mu_\xi(\bar{Q}) = 0$, т. е. ξ принимает лишь значения из $Q = \{x_1, x_2, \dots, x_m, \dots\}$. При этом $\mu_\xi(x_i) = P\{\xi = x_i\}$ являются вероятностями отдельных значений.

Найдем значение интеграла $\int_{R^n} f(x) \mu_{\xi}(dx)$. Пусть $I_Q = 1$ при $x \in Q$ и $I_Q(x) = 0$ при $x \notin Q$. Имеем

$$f(x) = f(x) I_Q(x) + f(x) (1 - I_Q(x)).$$

Функция $g(x) = f(x) (1 - I_Q(x))$ равна нулю на множестве Q , на котором сосредоточена мера μ_{ξ} . Поэтому

$$\int_{R^n} g(x) \mu_{\xi}(dx) = 0.$$

Действительно,

$$\int_{R^n} g(x) \mu_{\xi}(dx) = \lim_{n \rightarrow \infty} \sum_{k=-\infty}^{\infty} \frac{k}{n} \mu_{\xi} \left\{ x: \frac{k}{n} \leq g(x) < \frac{k+1}{n} \right\},$$

но
$$\mu_{\xi} \left\{ x: \frac{k}{n} \leq g(x) < \frac{k+1}{n} \right\} = 0 \quad \text{при } k \neq 0,$$

так как

$$\left\{ x: \frac{k}{n} \leq g(x) < \frac{k+1}{n} \right\} \subseteq \{x: g(x) \neq 0\} \subseteq \bar{Q}.$$

Функция $f(x) I_Q(x)$ отлична от нуля лишь на счетном множестве Q , следовательно, является простой функцией. Поэтому

$$\begin{aligned} \int_{R^n} f(x) I_Q(x) \mu_{\xi}(dx) &= \sum_{c_i} c_i \mu_{\xi} \{x: f(x) I_Q(x) = c_i\} = \\ &= \sum_{c_i} c_i \left\{ \sum_{\substack{x_j \in Q \\ f(x_j) = c_i}} \mu_{\xi}(x_j) \right\} = \sum_{x_j \in Q} f(x_j) \mu_{\xi}(x_j) \end{aligned}$$

(последнее равенство получается путем перестановки членов ряда; это возможно в силу существования $Mf(\xi)$).

Таким образом, в дискретном случае

$$Mf(\xi) = \int_{R^n} f(x) \mu_{\xi}(dx) = \sum_{x_j} f(x_j) \mu_{\xi}(x_j),$$

где x_j пробегает множество Q возможных значений случайной (векторной) величины ξ , в частности, если ξ одномерна, то

$$M\xi = \sum_{x_j} x_j \mu_{\xi}(x_j) = \sum_{x_j} x_j P\{\xi = x_j\}.$$

Это вполне согласуется с тем, что мы имели прежде в дискретном случае (заметим, что сейчас Ω может быть любым,

но величина ξ имеет не более чем счетное множество различных значений).

2) Случай существования плотности распределения (абсолютно непрерывный случай). Будем обозначать через dx элемент объема в пространстве R^n .

Определение 7.5. Говорят, что случайная величина имеет *плотность распределения* $p_{\xi}(x)$, если распределение μ_{ξ} выражается в виде

$$\mu_{\xi}(B) = P\{\xi \in B\} = \int_B p_{\xi}(x) dx, \quad (7.1)$$

где $p_{\xi}(x)$ — измеримая по Борелю функция и равенство имеет место для любого борелевского множества $B \subseteq R^n$.

Интеграл в правой части равенства (7.1) понимается, вообще говоря, как интеграл Лебега по мере, совпадающей с объемом в R^n , но поскольку имеющие практическое значение плотности $p_{\xi}(x)$ бывают обычно непрерывны (во всяком случае, интегрируемы по Риману), то читатель практически ничего не потеряет, если будет считать этот интеграл интегралом Римана (в предыдущем параграфе упоминалось, что интегралы Лебега и Римана в этом случае совпадают).

Можно показать, что равенство (7.1) достаточно проверить лишь для параллелепипедов B , так как из справедливости его для параллелепипедов вытекает его справедливость для любых борелевских множеств B (по той причине, что любое борелевское множество складывается из параллелепипедов), но мы не будем давать строгого доказательства этого утверждения.

Отметим следующие свойства плотности распределения $p_{\xi}(x)$.

1. Можно считать, что $p_{\xi}(x) \geq 0$, так как если бы на некотором множестве C было бы $p_{\xi}(x) < 0$, то C было бы борелевским и тогда, с одной стороны,

$$\int_C p_{\xi}(x) dx = P\{\xi \in C\} \geq 0,$$

а с другой,

$$\int_C p_{\xi}(x) dx < 0,$$

если только объем множества C не равен нулю.

Таким образом, только на множестве C нулевого объема может быть $p_{\xi}(x) < 0$. Если мы заменим ее нулем на этом множестве, то равенство (7.1) не нарушится, так как

$$\int_B p_{\xi}(x) dx = \int_{BC} p_{\xi}(x) dx + \int_{B \setminus C} p_{\xi}(x) dx,$$

но $\int_{BC} p_{\xi}(x) dx = 0$, а на множестве $B\bar{C}$ значения $p_{\xi}(x)$ не меняются. Поэтому с самого начала можно ограничиться неотрицательными плотностями распределения.

$$2. \int_{R^n} p_{\xi}(x) dx = P\{\xi \in R^n\} = 1.$$

Любая измеримая по Борелю функция $p(x)$, обладающая свойствами 1 и 2, может быть плотностью распределения некоторой случайной величины. Действительно, положим $\Omega = R^n$, \mathfrak{B} есть σ -алгебра борелевских подмножеств R^n , тогда для $B \in \mathfrak{B}$

$$P(B) = \int_B p(x) dx.$$

Тогда совокупность $(\Omega, \mathfrak{B}, P)$ будет удовлетворять аксиомам 6.1—6.4, а в качестве случайной величины ξ мы возьмем $\xi(\omega) = \omega$. В результате будем иметь

$$\mu_{\xi}(B) = P\{\xi \in B\} = P\{\omega: \xi(\omega) \in B\} = P\{\omega \in B\} = \int_B p(x) dx.$$

Найдем выражение для интеграла $\int_{R^n} f(x) \mu_{\xi}(dx)$ в случае существования плотности $p(x)$. Пусть сначала функция $f(x)$ простая, т. е. принимает лишь счетное множество различных значений $c_1, c_2, \dots, c_n, \dots$. Тогда случайная величина $f(\xi(\omega))$ дискретна. По только что доказанному свойству дискретных случайных величин имеем

$$\begin{aligned} Mf(\xi) &= \int_{R^n} f(x) \mu_{\xi}(dx) = \sum_{c_i} c_i \cdot P\{f(\xi) = c_i\} = \\ &= \sum_{c_i} c_i P\{\xi \in f^{-1}(c_i)\} = \sum_{c_i} c_i \int_{f^{-1}(c_i)} p_{\xi}(x) dx = \\ &= \sum_{c_i} \int_{f^{-1}(c_i)} c_i p_{\xi}(x) dx = \sum_{c_i} \int_{f^{-1}(c_i)} f(x) p_{\xi}(x) dx = \int_{R^n} f(x) p_{\xi}(x) dx. \end{aligned}$$

Последнее равенство получается в силу того, что для простых функций $f(x)$ справедливо соотношение

$$\int_{A_1 + A_2 + \dots + A_n + \dots} f(x) p_{\xi}(x) dx = \sum_{i=1}^{\infty} \int_{A_i} f(x) p_{\xi}(x) dx,$$

если только борелевские множества A_1, A_2, \dots попарно не пересекаются. Итак,

$$M f(\xi) = \int_{R^n} f(x) \mu_{\xi}(dx) = \int_{R^n} f(x) p_{\xi}(x) dx$$

для простых функций $f(x)$. Предельный переход убеждает нас в том, что верна

Теорема 7.3. В случае существования плотности распределения $p_{\xi}(x)$ имеем для любой измеримой по Борелю функции $f(x)$

$$M f(\xi) = \int_{R^n} f(x) p_{\xi}(x) dx.$$

Замечание 1. В случае $n=1$, т. е. для одномерной случайной величины ξ , имеем, полагая $f(x) = x$,

$$M \xi = \int_{-\infty}^{\infty} x p_{\xi}(x) dx.$$

Замечание 2. Иногда распределения μ_{ξ} случайных величин пытаются делить на дискретные и «непрерывные». Поскольку в литературе не выработалось однозначного понимания термина «непрерывное распределение», мы им не будем пользоваться. Распределения, имеющие плотность, мы будем называть *абсолютно непрерывными* в соответствии с существующим вполне четким термином.

3) **Смешанный случай.** Практические потребности обычно бывают удовлетворены, если рассматривать распределения μ следующего вида:

$$\mu_{\xi} = p\mu_1 + q\mu_2, \quad p + q = 1,$$

где μ_1 дискретно, а распределение μ_2 абсолютно непрерывно.

Если распределение μ_1 сосредоточено в точках $\{x_1, x_2, \dots, x_m, \dots\}$, а плотность распределения μ_2 есть $p(x)$, то из определения интеграла Лебега получаем

$$\begin{aligned} M f(\xi) &= \int_{R^n} f(x) \mu_{\xi}(dx) = p \int_{R^n} f(x) \mu_1(dx) + q \int_{R^n} f(x) \mu_2(dx) = \\ &= p \sum_{x_i} f(x_i) \mu_1(x_i) + q \int_{R^n} f(x) p(x) dx. \end{aligned}$$

4) **Случай распределения, сосредоточенного на поверхности.** Может оказаться, что значения случайного вектора $\xi = \xi(\omega)$ являются точками некоторой поверхности S , лежащей в евклидовом пространстве R^n , причем на этой поверхности ξ имеет плотность распределения. Это означает, что для любого борелевского $A \subseteq S$ имеем

$$P\{\xi \in A\} = \int_A p_\xi(s) ds,$$

где ds — элемент объема поверхности S . Поскольку при изучении абсолютно непрерывного случая мы нигде не использовали того, что R^n — евклидово пространство, а не какая-то поверхность, имеем, очевидно,

$$Mf(\xi) = \int_{R^n} f(x) \mu_\xi(dx) = \int_S f(s) p_\xi(s) ds.$$

7.3. Функция и плотность распределения в одномерном случае.

Согласно основным определениям, распределение μ_ξ случайной величины ξ есть мера, определенная на борелевских множествах B . Следовательно, под знак μ_ξ можно в качестве аргумента подставлять борелевские множества B . Множество значений аргумента оказывается при этом крайне сложным. Было бы хорошо, если бы меру μ_ξ можно было задавать функцией, у которой аргумент устроен попроще (с точки зрения приложений такое определение является просто необходимым). Один способ, приводящий к некоторым упрощениям, — рассматривать либо абсолютно непрерывные распределения, задаваемые плотностью, т. е. функцией точки, либо дискретные распределения, задаваемые последовательностью (функцией целочисленного аргумента). Но в определение меры μ_ξ через плотность $p_\xi(x)$ входит интегрирование, что не всегда удобно. В случае, когда ξ имеет размерность 1, т. е. является случайной точкой прямой $(-\infty, \infty)$, существует более удобная форма задания распределения μ_ξ .

Определение 7.6. *Функцией распределения $F_\xi(x)$ случайной величины ξ называется функция, задаваемая выражением*

$$F_\xi(x) = \mu_\xi\{(-\infty, x)\} = P\{\xi < x\}, \\ -\infty < x < \infty.$$

Удобство функции распределения объясняется тем, что в одномерном случае, как правило, интересуются вероятностью $P\{\xi \in B\}$ отнюдь не для любых борелевских множеств B , а обычно для интервалов вида $B = \{x: a \leq x < b\}$, или $\{x: a < x \leq b\}$, или $\{x: a < x < b\}$ и т. п., а в крайнем случае для множеств B , являющихся суммами интервалов указанного типа. Очевидно, что

$$P\{a \leq \xi < b\} = P\{\xi < b\} - P\{\xi < a\} = F_\xi(b) - F_\xi(a).$$

Для выражения вероятности $P\{a < \xi < b\}$ через функцию распределения докажем лемму.

Лемма 7.1. *Пусть события $A_1, A_2, \dots, A_n, \dots$ образуют возрастающую последовательность, т. е.*

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots \subseteq A_n \subseteq \dots$$

и $A = \bigcup_{i=1}^{\infty} A_i$. В таком случае $P(A) = \lim_{n \rightarrow \infty} P(A_n)$.

Пусть события $B_1, B_2, \dots, B_n, \dots$ образуют убывающую последовательность, т. е.

$$B_1 \supseteq B_2 \supseteq \dots \supseteq B_n \supseteq \dots$$

и $B = \bigcap_{i=1}^{\infty} B_i$. В таком случае $P(B) = \lim_{n \rightarrow \infty} P(B_n)$.

Доказательство. Имеем

$$A = A_1 + A_2 \bar{A}_1 + A_3 \bar{A}_2 + \dots + A_n \bar{A}_{n-1} + \dots,$$

$$A_n = A_1 + A_2 \bar{A}_1 + \dots + A_n \bar{A}_{n-1}, \quad (A_n \bar{A}_{n-1}) (A_m \bar{A}_{m-1}) = \emptyset.$$

В силу счетной аддитивности меры P имеем

$$\begin{aligned} P(A) &= P(A_1) + P(A_1 \bar{A}_2) + \dots + P(A_n \bar{A}_{n-1}) + \dots = \\ &= \lim_{n \rightarrow \infty} \{P(A_1) + P(A_1 \bar{A}_2) + \dots + P(A_n \bar{A}_{n-1})\} = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

Далее имеем

$$B_1 = B + B_1 \bar{B}_2 + B_2 \bar{B}_3 + \dots + B_n \bar{B}_{n+1} + \dots$$

Следовательно,

$$P(B_1) = P(B) + P(B_1 \bar{B}_2) + \dots + P(B_n \bar{B}_{n+1}) + \dots,$$

откуда, в частности, следует, что при $n \rightarrow \infty$

$$\sum_{k=n}^{\infty} P(B_k \bar{B}_{k+1}) \rightarrow 0.$$

Поскольку

$$B_n = B + B_n \bar{B}_{n+1} + B_{n+1} \bar{B}_{n+2} + \dots,$$

получаем, что

$$P(B_n) - P(B) = \sum_{k=n}^{\infty} P(B_k \bar{B}_{k+1}) \rightarrow 0 \quad (n \rightarrow \infty).$$

Лемма доказана.

Поскольку

$$\{a < \xi < b\} = \bigcup_{n=1}^{\infty} \left\{ a + \frac{1}{n} \leq \xi < b \right\},$$

получаем следствие:

$$\begin{aligned} P\{a < \xi < b\} &= \lim_{n \rightarrow \infty} \left(F_{\xi}(b) - F_{\xi}\left(a + \frac{1}{n}\right) \right) = \\ &= F_{\xi}(b) - F_{\xi}(a + 0). \end{aligned}$$

Аналогично,

$$\begin{aligned} P\{\xi = a\} &= \lim_{n \rightarrow \infty} P\left\{a \leq \xi < a + \frac{1}{n}\right\} = \\ &= \lim_{n \rightarrow \infty} \left(F_{\xi}\left(a + \frac{1}{n}\right) - F_{\xi}(a) \right) = F_{\xi}(a + 0) - F_{\xi}(a). \end{aligned}$$

Отметим следующие свойства функции распределения:

1. Если $x_2 > x_1$, то

$$F_{\xi}(x_2) = P\{\xi < x_2\} = P\{\xi < x_1\} + P\{x_1 \leq \xi < x_2\} \geq F_{\xi}(x_1).$$

Таким образом, функция распределения монотонна.

$$2. \lim_{x \rightarrow -\infty} F_{\xi}(x) = \lim_{n \rightarrow \infty} P\{\xi < -n\} = P\left\{\bigcap_n (\xi < -n)\right\} = P\{\emptyset\} = 0$$

(поясним, что в силу отмеченной в свойстве 1 монотонности функции $F_{\xi}(x)$ можно заменить $\lim_{x \rightarrow -\infty}$, где x пробегает все вещественные значения, на $\lim_{n \rightarrow \infty}$, где n пробегает лишь целые значения, и затем воспользоваться леммой 7.1). Аналогично показывается, что

$$\lim_{x \rightarrow +\infty} F_{\xi}(x) = 1.$$

$$\begin{aligned} 3. \lim_{x \rightarrow a-0} F_{\xi}(x) &= \lim_{n \rightarrow \infty} F_{\xi}\left(a - \frac{1}{n}\right) = \lim_{n \rightarrow \infty} P\left\{\xi < a - \frac{1}{n}\right\} = \\ &= P\left\{\bigcup_{n=1}^{\infty} \left(\xi < a - \frac{1}{n}\right)\right\} = P\{\xi < a\} = F_{\xi}(a). \end{aligned}$$

Таким образом, функция распределения непрерывна слева.

Верно и обратное: для всякой функции $F(x)$, имеющей свойства 1—3, найдется случайная величина ξ такая, что $F(x) = F_{\xi}(x)$. Однако этого утверждения мы доказывать не будем.

Отметим, что если существует плотность распределения $p_{\xi}(x)$, то

$$F_{\xi}(x) = P\{\xi < x\} = \int_{-\infty}^x p_{\xi}(x) dx. \quad (7.2)$$

Следовательно, во всех точках непрерывности $p_{\xi}(x)$ имеем

$$p_{\xi}(x) = \frac{dF_{\xi}(x)}{dx}. \quad (7.3)$$

Однако в точках разрыва $p_{\xi}(x)$ равенство (7.3) может нарушаться. Поэтому эквивалентным определением плотности через функцию распределения может быть равенство (7.2), но не равенство (7.3).

Замечание. В процессе развития обозначений теории вероятностей получилось так, что интеграл, обозначаемый нами через $\int_{-\infty}^{\infty} f(x) \mu_{\xi}(dx)$, обычно обозначается через $\int_{-\infty}^{\infty} f(x) dF_{\xi}(x)$.

Иногда он истолковывается несколько иначе (как интеграл Римана — Стильтьеса или интеграл Лебега — Стильтьеса, причем в первом толковании в настоящее время нет необходимости, а второе толкование совпадает с нашим с точностью до названия). Читателю следует просто иметь в виду тождество

$$\int_{-\infty}^{\infty} f(x) dF_{\xi}(x) \equiv \int_{-\infty}^{\infty} f(x) \mu_{\xi}(dx).$$

Пример 7.1. *Показательное распределение.* Пусть имеется некий прибор, который включают в нулевой момент времени, а в случайный момент времени ξ он выходит из строя. Постараемся вывести общий вид функции распределения $P\{\xi < x\} = F_{\xi}(x)$. Нам удобней будет иметь дело с функцией $Q_{\xi}(x) = 1 - F_{\xi}(x) = P\{\xi \geq x\}$. Согласно простейшей модели прибор выйдет из строя, как только уровень нагрузки превысит некоторое допустимое значение N . Правдоподобно будет предположить, что вероятность того, что это случится в промежутке времени $[a, b]$ при условии, что этого не случилось до момента a , зависит лишь от длины интервала $[a, b]$ и при малых $b-a$ есть $\lambda(b-a) + o(b-a)$, где $\lambda > 0$ — число. Поэтому при малом Δx

$$\begin{aligned} P\{\xi \geq x + \Delta x\} &= P\{\xi \geq x\} P\{\xi \geq x + \Delta x | \xi \geq x\} = \\ &= P\{\xi \geq x\} \cdot \{1 - \lambda \Delta x - o(\Delta x)\}. \end{aligned}$$

Иначе говоря, для $Q_{\xi}(x) = P\{\xi \geq x\}$ получаем

$$Q_{\xi}(x + \Delta x) - Q_{\xi}(x) = -(\lambda \Delta x + o(\Delta x)) Q_{\xi}(x).$$

При обычных предположениях дифференцируемости функции $Q_{\xi}(x)$ имеем

$$\frac{dQ_{\xi}(x)}{dx} = -\lambda Q_{\xi}(x), \quad Q_{\xi}(x) = Ce^{-\lambda x}$$

и при естественном предположении $Q_{\xi}(0) = 1$ имеем

$$Q_{\xi}(x) = e^{-\lambda x}, \quad F_{\xi}(x) = 1 - e^{-\lambda x}, \quad x \geq 0.$$

Такое распределение называется показательным. Очевидно, что функция $p_{\xi}(x) = \lambda e^{-\lambda x}$ при $x \geq 0$ и $p_{\xi}(x) = 0$ при $x < 0$ является

плотностью показательного закона распределения, так как выполняется равенство (7.2). Легко вычислить, что

$$M\xi = \int_{-\infty}^{\infty} x p_{\xi}(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Считается, что показательный закон распределения справедлив во многих случаях, например для длительности работы электронных ламп. Заметим, что в основе приведенных для вывода показательного закона эвристических соображений лежит предположение о том, что длительность работы прибора определяется колебаниями нагрузки, но не старением самого прибора. Это предположение трудно проверяемо, а иногда и неверно. Кроме того, сделано некоторое предположение о распределении вероятностей различных нагрузок, которое также легко может быть неверным, хотя бы потому, что нагрузка, возможно, не обладает статистической устойчивостью, и потому бессмысленно говорить о ней в вероятностных терминах.

По всем этим причинам наличие показательного распределения для времени безаварийной работы является всегда гипотезой, которая должна быть подтверждена опытными данными. Подчеркнем, что таково же положение со всеми другими видами распределений, которые встречаются в теории вероятностей.

Однако гипотеза показательного закона очень привлекательна. Отметим две причины этого. Во-первых, показательный закон обладает гармоничным свойством самовоспроизводимости в следующем смысле. Допустим, что интересующий нас прибор состоит из n звеньев A_1, A_2, \dots, A_n , причем отказ любого звена приводит к отказу прибора, а моменты отказов ξ_i звеньев A_i независимы и распределены по показательному закону: $P\{\xi_i < x\} = 1 - e^{-\lambda_i x}$ с параметром λ_i . В таком случае момент отказа всего прибора есть, очевидно, $\xi = \min(\xi_1, \dots, \xi_n)$. Имеем

$$\begin{aligned} P\{\xi < x\} &= 1 - P\{\xi \geq x\} = 1 - P\{\min(\xi_1, \dots, \xi_n) \geq x\} = \\ &= 1 - P\{\xi_1 \geq x, \xi_2 \geq x, \dots, \xi_n \geq x\}, \end{aligned}$$

а в силу независимости случайных величин ξ_1, \dots, ξ_n (в общем случае сохраняется определение 3.5 независимых случайных величин) получаем

$$P\{\xi < x\} = 1 - \prod_{i=1}^n P\{\xi_i \geq x\} = 1 - e^{-(\lambda_1 + \dots + \lambda_n)x},$$

т. е. вновь получается показательный закон с параметром $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$.

Например, если в вычислительной машине имеется $n = 1\,000\,000$ элементов, для каждого из которых математическое

ожидание времени отказа $\frac{1}{\lambda_i} = 100\,000$ час (в году 8760 час), т. е. $\lambda_i = 10^{-5}$, то для всей машины $\lambda = 10$. Следовательно, вероятность $P\{\xi \geq 1\}$ того, что машина будет безотказно работать 1 час есть

$$P\{\xi \geq 1\} = e^{-\lambda} = e^{-10} \approx 0,5 \cdot 10^{-4},$$

если, конечно, предположить, что отказ одного элемента нарушает работу всей машины. Большие машины, следовательно, должны строиться с учетом ненадежности элементов.

Вторая причина привлекательности показательного закона — его связь с законом Пуассона. Естественно предположить, что число аварий $\mu_{[S, T]}$, которые происходят в большой системе за отрезок времени от S до T , имеет распределение Пуассона. Если $S < T < U$, то

$$\mu_{[S, U]} = \mu_{[S, T]} + \mu_{[T, U]},$$

а следовательно, для функции двух переменных $m(S, T) = M\mu_{[S, T]}$, т. е. для параметра распределения Пуассона, имеем

$$m(S, U) = m(S, T) + m(T, U).$$

Дальнейшее зависит от того, будем ли мы учитывать старение или нет. Если мы не учитываем старение, то $m(S, T)$ зависит лишь от разности $T - S$: $m(S, T) = f(T - S)$.

Но тогда, полагая $T - S = x$, $U - T = y$, имеем

$$f(x + y) = f(x) + f(y),$$

откуда вытекает, что функция f линейна:

$$f(x) = \lambda x,$$

где по смыслу задачи $\lambda \geq 0$. Вероятность $P\{\xi \geq T\}$ безаварийной работы на отрезке времени $[0, T]$ есть вероятность $P\{\mu_{[0, T]} = 0\} = e^{-m(0, T)}$, откуда

$$P\{\xi < T\} = 1 - P\{\xi \geq T\} = 1 - e^{-m(0, T)} = 1 - e^{-f(T)} = 1 - e^{-\lambda T}.$$

Если же мы желаем учитывать старение, то все определяется функцией $g(T) = m(0, T)$, так как, очевидно, $m(S, T) = g(T) - g(S)$. Функция $g(T)$ может быть любой (неубывающей, по смыслу задачи) функцией. В этом случае получаем для момента ξ первого отказа

$$P\{\xi < T\} = 1 - P\{\mu_{[0, T]} = 0\} = 1 - e^{-g(T)},$$

что является некоторым обобщением показательного закона.

В целом, показательный закон (возможно, в форме, учитывающей старение) является, вместе с тесно связанным с ним законом Пуассона, одним из универсальных вероятностных распределений, при нарушении которых оказывается под подозрением статистическая однородность.

7.4. Плотность распределения в многомерном случае. В многомерном случае функция распределения случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ иногда вводится равенством

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = P\{\xi_1 < x_1, \dots, \xi_n < x_n\}.$$

Можно показать, что задание функции распределения однозначно определяет меру μ_ξ . Однако в многомерном случае понятие функции распределения почти бесполезно. Дело в том, что в многомерном пространстве возможна не одна разумная система координат, как на прямой, а многие. Во-первых, это декартовы системы координат, получаемые ортогональным преобразованием базиса, и аффинные системы, получаемые линейным невырожденным преобразованием. Во-вторых, это различные криволинейные координаты типа сферических, цилиндрических и т. п.

Общий принцип состоит поэтому в использовании таких величин, которые преобразовывались бы сравнительно просто при преобразовании системы координат. Между тем, уже при ортогональном преобразовании координат функция распределения преобразуется чрезвычайно сложно. Для того чтобы записать соответствующую формулу, фактически приходится делать переход к мере μ_ξ . Далее, в криволинейных системах (например, в полярных координатах на плоскости) понятие функции распределения вовсе бессмысленно. Поэтому в настоящей книге предлагается рассматривать либо меру μ (которая определяется независимо от системы координат равенством $\mu(B) = P\{\xi \in B\}$ и потому инвариантна относительно преобразований координат), либо, в случае существования плотности, — плотность $p_\xi(x)$ меры μ_ξ . Как преобразуется плотность, мы сейчас увидим.

Пусть имеется n гладких функций $y_i = f_i(x_1, \dots, x_n)$ от n вещественных переменных x_1, \dots, x_n , задающих отображение $f: R^n \rightarrow R^n$. Якобианом Df отображения f называется функция, значение которой в точке x , обозначаемое $Df(x)$, задается формулой

$$Df(x) = \det \left\| \frac{\partial f_i}{\partial x_j} \right\|, \quad i, j = 1, \dots, n$$

(т. е. $Df(x)$ есть определитель матрицы Якоби).

Известно, что если $Df(x) \neq 0$, то в окрестности точки $f(x)$ существует гладкое обратное отображение $f^{-1}(y)$. При этом

$$Df^{-1}(f(x)) Df(x) = 1.$$

Если f есть взаимно однозначное отображение $R^n \rightarrow R^n$, то имеет место также формула замены переменных в определенном интеграле:

$$\int_B \varphi(x) dx = \int_{f^{-1}(B)} \varphi(f(x)) |Df(x)| dx,$$

где B — любая область в R^n .

Пусть теперь $\xi = (\xi_1, \dots, \xi_n)$ — случайный вектор, распределение которого имеет плотность p_ξ , f — взаимно однозначное отображение R^n на R^n , $\eta = f(\xi)$.

Теорема 7.4. Распределение вектора η имеет плотность p_η , выражаемую формулой

$$p_\eta(x) = p_\xi(f^{-1}(x)) |Df^{-1}(x)|.$$

Доказательство. Имеем

$$\begin{aligned} \int_B p_\xi(f^{-1}(x)) |Df^{-1}(x)| dx &= \int_{f^{-1}(B)} p_\xi\{f^{-1}(f(x))\} |Df^{-1}(x)| \cdot |Df(x)| dx = \\ &= \int_{f^{-1}(B)} p_\xi(x) dx = P\{\xi \in f^{-1}(B)\} = P\{f(\xi) \in B\} = P\{\eta \in B\}, \end{aligned}$$

что и требовалось доказать.

Как получить из многомерной плотности $p_\xi(x) = p_{\xi_1 \dots \xi_n}(x_1, \dots, x_n)$ одномерную плотность распределения $p_{\xi_i}(x_i)$ для одной из компонент ξ_i вектора $\xi = (\xi_1, \dots, \xi_n)$?

Теорема 7.5.

$$p_{\xi_i}(x_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_{\xi_1 \dots \xi_n}(x_1, \dots, x_n) dx_2 \dots dx_n.$$

Доказательство. Для любого одномерного борелевского множества B имеем

$$\begin{aligned} P\{\xi_i \in B\} &= P\{\xi_i \in B, -\infty < \xi_2 < \infty, \dots, -\infty < \xi_n < \infty\} = \\ &= \int_{B_i} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_{\xi_1 \dots \xi_n}(x_1, \dots, x_n) dx_1 \dots dx_n = \\ &= \int_{B_i} \left\{ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_{\xi_1 \dots \xi_n}(x_1, \dots, x_n) dx_2 \dots dx_n \right\} dx_1, \end{aligned}$$

что и требовалось доказать.

Пример 7.2. Распределение суммы случайных величин. Пусть двумерный вектор $\xi = (\xi_1, \xi_2)$ имеет плотность распределения $p_\xi(x) = p_{\xi_1 \xi_2}(x_1, x_2)$. Рассмотрим преобразование $f: (x_1, x_2) \rightarrow (x_1 + x_2, x_2)$. Тогда

$$(\xi_1, \xi_2) \rightarrow (\eta_1, \eta_2) = (\xi_1 + \xi_2, \xi_2).$$

Так как $Df = 1$, имеем

$$p_\eta(x) = p_{\eta_1 \eta_2}(x_1, x_2) = p_\xi(f^{-1}(x)) = p_{\xi_1 \xi_2}(x_1 - x_2, x_2).$$

Следовательно,

$$p_{\xi_1 + \xi_2}(x_1) = p_{\eta_1}(x_1) = \int_{-\infty}^{\infty} p_{\xi_1 \xi_2}(x_1 - x_2, x_2) dx_2.$$

7.5. Перенос на общий случай основных понятий. Понятие независимости случайных величин специально вводилось таким образом, чтобы сделать тривиальным переход от дискретного случая к общему. Определение 3.4 и теорема 3.5 сохраняются без изменений, только числовые множества A и B надо считать борелевскими, а функции f и g — измеримыми по Борелю. Аналогом теоремы 3.6 является следующая

Теорема 7.6. Пусть плотность распределения $p_{\xi}(x)$ случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ представляется в виде

$$p_{\xi}(x) = p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{j=1}^n p_{\xi_j}(x_j),$$

где $p_{\xi_j}(x_j)$ — соответствующие одномерные плотности. Тогда случайные величины ξ_1, \dots, ξ_n независимы в совокупности.

Доказательство. Проверим определение 3.5. Имеем для любых борелевских A_{i_1}, \dots, A_{i_k}

$$\mathbf{P}\{\xi_{i_1} \in A_{i_1}, \dots, \xi_{i_k} \in A_{i_k}\} = \mathbf{P}\{\xi_{i_s} \in A_{i_s}, s = 1, \dots, k, \xi_i \in (-\infty, \infty),$$

$$\begin{aligned} i \neq i_s\} &= \int_{A_{i_1}} \dots \int_{A_{i_k}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p_{\xi_{i_1}}(x_{i_1}) \dots p_{\xi_{i_k}}(x_{i_k}) \times \\ &\times dx_{i_1} \dots dx_{i_k} \prod_{i \neq i_s} p_{\xi_i}(x_i) dx_i = \prod_{s=1}^k \int_{A_{i_s}} p_{\xi_{i_s}}(x_{i_s}) dx_{i_s} = \\ &= \prod_{s=1}^k \mathbf{P}\{\xi_{i_s} \in A_{i_s}\}, \end{aligned}$$

что и требовалось доказать (поясним, что мы воспользовались представлением многократного интеграла в виде произведения однократных и равенством $\int_{-\infty}^{\infty} p_{\xi_i}(x_i) dx_i = 1$).

Пример 7.3. Формула композиции. Для независимых случайных величин

$$p_{\xi_1 \xi_2}(x_1, x_2) = p_{\xi_1}(x_1) p_{\xi_2}(\tilde{x}_2),$$

поэтому плотность их суммы есть

$$p_{\xi_1 + \xi_2}(x_1) = \int_{-\infty}^{\infty} p_{\xi_1}(x_1 - x_2) p_{\xi_2}(x_2) dx_2.$$

Эта формула, называемая формулой композиции, или свертки, имеет известную аналогию с формулой полной вероятности: каждому значению x_2 величины ξ_2 ставится в соответствие значение $x_1 - x_2$ величины ξ_1 , перемножаются значения плотностей вероятностей и все интегрируется. Разумеется, справедлива и формула

$$p_{\xi_1 + \xi_2}(x_1) = \int_{-\infty}^{\infty} p_{\xi_1}(x_2) p_{\xi_2}(x_1 - x_2) dx_2. \quad (7.4)$$

В дискретном случае большую роль играла формула $M(\xi\eta) = M\xi \cdot M\eta$ для независимых ξ и η . Обобщим эту формулу на случай произвольных независимых ξ и η . Полагая $\xi_n = \frac{k}{n}$, если $\frac{k}{n} \leq \xi < \frac{k+1}{n}$, $\eta_n = \frac{k}{n}$, если $\frac{k}{n} \leq \eta < \frac{k+1}{n}$, получим, что ξ_n и η_n независимы как функции от независимых случайных величин, следовательно, $M\xi_n \eta_n = M\xi_n M\eta_n$. Очевидно, что $M\xi_n \rightarrow M\xi$, $M\eta_n \rightarrow M\eta$. Покажем, что $M\xi_n \eta_n \rightarrow M\xi\eta$. Имеем

$$\begin{aligned} M\xi\eta - M\xi_n \eta_n &= M(\xi\eta - \xi_n \eta_n) = M(\xi\eta - \xi_n \eta + \xi_n \eta - \xi_n \eta_n) = \\ &= M(\xi - \xi_n)\eta + M\xi_n(\eta - \eta_n). \end{aligned}$$

Имеем $|\xi - \xi_n| \leq \frac{1}{n}$, откуда $|\xi_n| \leq |\xi| + \frac{1}{n}$, $|\eta - \eta_n| \leq \frac{1}{n}$. Поэтому

$$|M\xi\eta - M\xi_n \eta_n| \leq \frac{1}{n} M|\eta| + \frac{1}{n} \left(M|\xi| + \frac{1}{n} \right) \rightarrow 0$$

при $n \rightarrow \infty$, что и требовалось доказать.

Определение и свойства дисперсии автоматически переносятся на общий случай. Для доказательства неравенства Чебышева напишем ряд неравенств:

$$\begin{aligned} D\xi &= M(\xi - M\xi)^2 = \int_{-\infty}^{\infty} (x - M\xi)^2 \mu_{\xi}(dx) \geq \\ &\geq \int_{|x - M\xi| \geq \varepsilon} (x - M\xi)^2 \mu_{\xi}(dx) \geq \varepsilon^2 \int_{|x - M\xi| \geq \varepsilon} \mu_{\xi}(dx) = \\ &= \varepsilon^2 \int_{-\infty}^{\infty} I_{\varepsilon}(x) \mu_{\xi}(dx), \end{aligned}$$

где

$$I_{\varepsilon}(x) = \begin{cases} 1, & \text{если } |x - M\xi| \geq \varepsilon, \\ 0, & \text{если } |x - M\xi| < \varepsilon. \end{cases}$$

Поэтому

$$\int_{-\infty}^{\infty} I_{\varepsilon}(x) \mu_{\xi}(dx) = M I_{\varepsilon}(\xi) = P\{I_{\varepsilon}(\xi) = 1\} = P\{|\xi - M\xi| \geq \varepsilon\},$$

откуда и вытекает неравенство Чебышева.

Теоремы о законе больших чисел переносятся на общий случай автоматически.

§ 8

ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

8.1. «Чудо» Лапласа. Мы познакомились в предыдущих параграфах с основными понятиями теории вероятностей: пространством элементарных событий, случайной величиной, распределением случайной величины, математическим ожиданием. Эти понятия в своих взаимосвязях образуют известную гармонию, и основные их связи мы выражали рядом теорем.

Однако эти теоремы при анализе их содержания обнаруживают полную тривиальность, поскольку фактически являются переформулировкой известных теорем математического анализа о свойствах функций (отображений), меры и интеграла (не столь важно, что в курсах математического анализа эти факты выражаются иногда не на наилучшем языке).

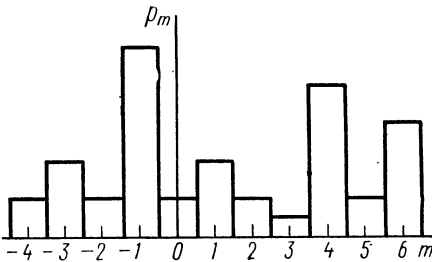


Рис. 8.1

Между тем, за любой научной признается право на существование только в том случае, если в ней имеются результаты, которые без нее предвидеть невозможно. При первом знакомстве такие результаты воспринимаются почти как «чудо». Лучшим «чудом» в современной теории вероятностей продолжает оставаться открытая Лапласом центральная предельная теорема. Состоит оно в следующем.

Рассмотрим сумму $S_n = \xi_1 + \xi_2 + \dots + \xi_n$ независимых случайных величин $\xi_1, \xi_2, \dots, \xi_n$. Для начала предположим, что эти случайные величины все имеют одинаковое распределение, причем принимают только целочисленные значения $0, \pm 1, \pm 2, \dots$ (последнее ограничение целочисленности не является существенным с точки зрения естествоиспытателя, поскольку результаты любых измерений выражаются дробью с конечным числом десятичных знаков, а тогда при подходящем выборе единицы измерения их можно

считать целыми числами). Распределение $p_m = P\{\xi_k = m\}$ каждой из величин ξ_1, \dots, ξ_n можно изобразить следующим образом. Нарисуем для каждого m прямоугольник, середина основания которого есть точка m , длина основания равна 1, а площадь есть p_m . Получится совершенно произвольный ряд прямоугольников (с высотами p_m), подчиненный лишь тому условию, что сумма всех их

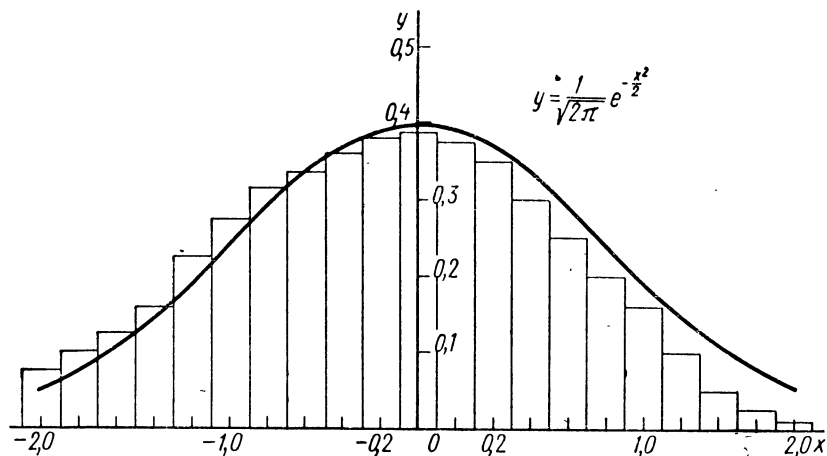


Рис. 8.2

площадей равна 1 (рис. 8.1). Попробуем изобразить таким же образом вероятности отдельных значений суммы $S_n = \xi_1 + \xi_2 + \dots + \xi_n$ при довольно большом n . При этом мы обнаружим, прежде всего, что даже если случайные величины $\{\xi_k\}$ принимали всего два значения 0 и 1, то значениями суммы S_n могут быть числа от 0 до n , следовательно, эти значения при большом n просто не поместятся в прежнем масштабе на рис. 8.1, и мы будем принуждены изменить масштаб. Иными словами, нужно сделать линейное преобразование оси абсцисс, т. е. вместо значений случайной величины $S_n = \xi_1 + \dots + \xi_n$ откладывать значения величины $S'_n = \frac{1}{b_n} (S_n - a_n)$, где a_n и b_n — некоторые числа, зависящие от n .

Лаплас открыл, что получится нечто замечательное, если положить

$$a_n = M(\xi_1 + \dots + \xi_n) = nx, \quad a = M\xi_k$$

$$b_n = \sqrt{D(\xi_1 + \dots + \xi_n)} = \sigma\sqrt{n}, \quad \sigma^2 = D\xi_k, \quad \sigma = \sqrt{D\xi_k}.$$

Случайную величину

$$s'_n = \frac{\xi_1 + \dots + \xi_n - M(\xi_1 + \dots + \xi_n)}{\sqrt{D(\xi_1 + \dots + \xi_n)}} = \frac{S_n - MS_n}{\sqrt{DS_n}} = \frac{S_n - na}{\sigma\sqrt{n}}$$

мы будем называть нормированной суммой. Очевидно, что $M s_n^* = 0$, $D s_n^* = 1$, а значениями величины s_n^* являются числа $x_n(m) = \frac{m - na}{\sigma \sqrt{n}}$, причем для любого целого m

$$P\{S_n = m\} = P\left\{s_n^* = \frac{m - na}{\sigma \sqrt{n}}\right\} = P\{s_n^* = x_n(m)\}.$$

Отложим (рис. 8.2) по оси абсцисс значения $x_n(m)$ и изобразим вероятности

$$P\{S_n = m\} = P\{s_n^* = x_n(m)\}$$

этих значений прямоугольниками, середины оснований которых лежат в точках $x_n(m)$, длины оснований равны расстоянию

$$x_n(m+1) - x_n(m) = \frac{1}{\sigma \sqrt{n}}$$

между соседними точками, а площади равны $P\{s_n^* = x_n(m)\}$. Высоты этих прямоугольников равны, следовательно,

$$\sigma \sqrt{n} P\{s_n^* = x_n(m)\}.$$

При этом произойдет нечто удивительное. Верхние основания этих прямоугольников почти точно лягут на некоторую раз навсегда вычисленную кривую, задаваемую уравнением

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Иными словами, при $n \rightarrow \infty$ имеем

$$\sigma \sqrt{n} P\{s_n^* = x_n(m)\} \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2(m)}{2}}. \quad (8.1)$$

Но в таком случае, очевидно,

$$\begin{aligned} P\{A \leq s_n^* \leq B\} &= \sum_{A \leq x_n(m) \leq B} P\{s_n^* = x_n(m)\} \approx \\ &\approx \sum_{A \leq x_n(m) \leq B} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2(m)}{2}} \cdot \frac{1}{\sigma \sqrt{n}} = \sum_{A \leq x_n(m) \leq B} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2(m)}{2}} \times \\ &\times (x_n(m+1) - x_n(m)) \rightarrow \int_A^B \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy. \end{aligned}$$

Введем функцию

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy,$$

называемую функцией Лапласа¹. Тогда

$$P\{A \leq s_n^* \leq B\} \approx \Phi(B) - \Phi(A). \quad (8.2)$$

В этом и состоит «чудо» Лапласа. Собственно, «чудом» является соотношение (8.1), а (8.2) является тривиальным следствием из него. Однако формулировка полученного результата в виде (8.1) неудобна тем, что такое «чудо» имеет исключения.

Действительно, пусть величины ξ_1, \dots, ξ_n принимают с ненулевой вероятностью только четные значения. Тогда при m нечетном

$$P\{s_n^* = x_n(m)\} = P\{S_n = m\} = 0$$

и равенство (8.1) неверно.

Можно было бы доказать, что указанное исключение по существу является единственным. Удвоение единицы измерения приведет наши случайные величины к величинам, принимающим также нечетные значения. Тогда (8.1) будет верно, а следовательно, верно и (8.2). Иными словами, получается так, что прямоугольники, отвечающие нечетным m , пропадают, но зато соседние прямоугольники, отвечающие четным m , удваиваются и (8.2) сохраняется.

Но мы не будем уточнять доказательство для этого случая, так как, на первый взгляд, все равно не ясно, как сформулировать (8.1) для общих (недискретных) случайных величин. Правда, с точки зрения естествоиспытателя, между дискретным и общим случаем не должно быть особых различий: результаты эксперимента, с одной стороны, всегда дискретны (измеряются с конечной точностью), а с другой стороны, по принятому во всей науке подходу,

¹ Читателю следует иметь в виду, что иногда авторы таблиц понимают под функцией Лапласа $\Phi(x)$ несколько иные выражения:

$$\Phi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy, \quad \Phi(x) = \int_{-x}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

и даже

$$\Phi(x) = \int_{-x\sqrt{2}}^{x\sqrt{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

Прежде чем пользоваться таблицей, надо посмотреть на ее заголовок, где всегда написано, что именно понимается под функцией $\Phi(x)$.

их можно считать и непрерывными. Поэтому должен существовать аналог (8.1), по крайней мере для случая существования плотностей распределения. Так и есть на самом деле, но этот случай вызывает некоторое усложнение доказательства. Наоборот, соотношение (8.2) имеет смысл и может быть доказано для самого общего случая.

Итак, в отличие от (8.1), соотношение (8.2), во-первых, не имеет исключений, а во-вторых, справедливо для недискретных величин. Поэтому мы направим наши основные усилия на точную формулировку и доказательство (8.2) в виде математической теоремы.

8.2. Нормальное распределение. Хорошо известно, что неотрицательная функция

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

обладает свойством $\int_{-\infty}^{\infty} \varphi(x) dx = 1$. Поэтому $\varphi(x)$ можно считать плотностью распределения некоторой случайной величины.

Определение 8.1. Говорят, что (одномерная) случайная величина ξ имеет *нормальное распределение с параметрами* $(0, 1)$, если

$$p_{\xi}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Коротко пишут иногда $\xi \sim N(0, 1)$. Легко видеть, что $M\xi = 0$, $D\xi = 1$.

Определение 8.2. Говорят, что случайная величина $\eta = \sigma\xi + a$, где $\sigma > 0$, a — любое число, $\xi \sim N(0, 1)$, имеет *нормальное распределение с параметрами* a, σ .

Коротко пишут иногда $\eta \sim N(a, \sigma)$. Очевидно, что $M\eta = a$, $D\eta = \sigma^2$.

Из теоремы 7.4 вытекает, что

$$p_{\eta}(x) = \frac{1}{\sigma} p_{\xi}\left(\frac{x-a}{\sigma}\right) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

8.3. Основной аппарат — характеристические функции. Пока что мы имеем для вычисления распределения суммы случайных величин лишь один способ — применение формулы композиции. Ясно, что при большом числе слагаемых этот способ, связанный с многократным интегрированием, совершенно безнадежен. Мы сейчас изложим другой способ, по существу также открытый Лапласом.

Определение 8.3. *Характеристической функцией* случайной величины ξ называется функция $f_{\xi}(t)$ вещественного переменного t , определяемая равенством

$$f_{\xi}(t) = M e^{it\xi} = \int_{-\infty}^{\infty} e^{itx} \mu_{\xi}(dx).$$

З а м е ч а н и е. Интеграл Лебега от функции с комплексными значениями определяется как сумма интеграла от вещественной части и интеграла от мнимой части, умноженного на $i = \sqrt{-1}$.

Очевидно, что $|f_{\xi}(t)| \leq 1$.

Теорема 8.1. Если случайные величины ξ_1, \dots, ξ_n независимы в совокупности, то

$$f_{\xi_1 + \dots + \xi_n}(t) = f_{\xi_1}(t) f_{\xi_2}(t) \dots f_{\xi_n}(t).$$

Доказательство. Имеем

$$\begin{aligned} f_{\xi_1 + \dots + \xi_n}(t) &= M \{e^{it(\xi_1 + \xi_2 + \dots + \xi_n)}\} = \\ &= M \left\{ \prod_{k=1}^n e^{it\xi_k} \right\} = \prod_{k=1}^n (M e^{it\xi_k}) = \prod_{k=1}^n f_{\xi_k}(t), \end{aligned}$$

поскольку величины $e^{it\xi_1}, \dots, e^{it\xi_n}$ независимы как функции от независимых случайных величин ξ_1, \dots, ξ_n (читателю рекомендуется обосновать утверждение теоремы о математическом ожидании произведения независимых случайных величин для величин с комплексными значениями).

Важную роль играют так называемые «моменты» случайной величины.

Определение 8.4. Моментом k -того порядка случайной величины ξ называется $M\xi^k$; центральным моментом (k -того порядка) называется $M(\xi - M\xi)^k$; абсолютным моментом называется $M|\xi|^k$; центральным абсолютным моментом называется $M|\xi - M\xi|^k$.

Теорема 8.2. Если $M\xi^k$ существует, то

$$M\xi^k = i^{-k} \frac{d^k}{dt^k} f_{\xi}(0).$$

Доказательство.

$$\frac{d^k}{dt^k} f_{\xi}(0) = \frac{d^k}{dt^k} \left(\int_{-\infty}^{\infty} e^{itx} \mu_{\xi}(dx) \right) \Big|_{t=0} = \int_{-\infty}^{\infty} (ix)^k e^{itx} \Big|_{t=0} \mu_{\xi}(dx) = i^k M\xi^k$$

(перемена порядка дифференцирования и интегрирования законна потому, что в силу существования $M\xi^k$ интеграл

$$\int_{-\infty}^{\infty} |x|^k \mu_{\xi}^*(dx) < \infty,$$

а следовательно, интеграл

$$\int_{-\infty}^{\infty} (ix)^k e^{itx} \mu_{\xi}^*(dx),$$

получившийся после дифференцирования, сходится равномерно (и абсолютно).

Теорема 8.3. Если $\eta = \sigma \xi + a$, где a и σ — числа, то

$$f_{\eta}(t) = e^{ita} f_{\xi}(t\sigma).$$

Доказательство

$$f_{\eta}(t) = M e^{it\eta} = e^{ita} M e^{it\sigma\xi} = e^{ita} f_{\xi}(t\sigma).$$

Теорема 8.4. Характеристическая функция $f(t)$ нормального распределения с параметрами $(0, 1)$ есть $e^{-\frac{t^2}{2}}$.

Доказательство

$$f(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} \cos tx \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Имеем

$$\begin{aligned} f'(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x (-\sin tx) e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sin tx \cdot e^{-\frac{x^2}{2}} d\left(-\frac{x^2}{2}\right) = \\ &= \frac{1}{\sqrt{2\pi}} \left(\sin tx \cdot e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} t \cos tx \cdot e^{-\frac{x^2}{2}} dx \right) = -tf(t), \end{aligned}$$

откуда $\frac{f'(t)}{f(t)} = -t$, $f(t) = Ce^{-\frac{t^2}{2}}$, и, подставляя $t = 0$, получаем $C = 1$, что и требовалось доказать.

П л а н д а л ь н е й ш е г о и з л о ж е н и я. Мы собираемся исследовать поведение суммы $S_n = \xi_1 + \dots + \xi_n$ и для

$$s_n^* = \frac{S_n - MS_n}{DS_n}$$

установить соотношение

$$P\{A \leq s_n^* \leq B\} \rightarrow \Phi(B) - \Phi(A),$$

$$(n \rightarrow \infty),$$

$$(8.3)$$

предельный переход в котором выражает на математическом языке знак приближенного равенства в (8.2). Вычислить распределение s_n^* сложно, но характеристическую функцию $f_{s_n^*}(t)$ легко вычислить, применяя теоремы 8.1 и 8.3. Забегая вперед, заметим, что при $n \rightarrow \infty$

$$f_{s_n^*}(t) \rightarrow e^{-\frac{t^2}{2}}, \quad (8.4)$$

т. е., в силу теоремы 8.4, $f_{s_n^*}(t)$ сходится при $n \rightarrow \infty$ к характеристической функции нормального закона $N(0, 1)$. Из соотношения (8.4) мы можем получить доказательство (8.3), для чего, впрочем, достаточно, чтобы для любого x

$$F_{s_n^*}(x) = P\{s_n^* < x\} \rightarrow \Phi(x), \quad (8.5)$$

т. е. из сходимости характеристических функций (8.4) можно вывести сходимость функций распределения (8.5). Для обоснования этого нам придется доказать соответствующую теорему.

8.4. Теорема о непрерывности. Напомним, что *преобразованием Фурье* суммируемой функции $\varphi(x)$, т. е. такой функции, что $\int_{-\infty}^{\infty} |\varphi(x)| dx < \infty$, называется функция

$$\tilde{\varphi}(t) = \int_{-\infty}^{\infty} e^{itx} \varphi(x) dx.$$

Если функция $\varphi(x)$ обращается в нуль вне некоторого конечного отрезка и имеет непрерывную производную порядка k , то с помощью интегрирования по частям получаем

$$\tilde{\varphi}^{(k)}(t) = (it)^k \tilde{\varphi}(t),$$

откуда вытекает, ввиду ограниченности $\tilde{\varphi}^{(k)}(t)$, что при $k \geq 2$ функция $\tilde{\varphi}(t)$ суммируема. Поэтому

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

где интеграл понимается в обычном смысле. Будем называть функции, обращающиеся в нуль вне конечного отрезка и имеющие две непрерывные производные, *финитными*. Имеем следующую лемму:

Лемма 8.1. *Преобразование Фурье $\tilde{\varphi}(t)$ финитной функции есть суммируемая функция и*

$$\varphi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \tilde{\varphi}(t) dt.$$

Лемма 8.2. Пусть $\varphi(x)$ — финитная функция, ξ — случайная величина, $f_{\xi}(t) = M e^{it\xi}$. Тогда

$$M\varphi(\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f_{\xi}(-t) \tilde{\varphi}(t) dt.$$

Доказательство. Мы можем написать следующие преобразования:

$$\begin{aligned} M\varphi(\xi) &= \int_{-\infty}^{\infty} \varphi(x) \mu_{\xi}(dx) = \int_{-\infty}^{\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \tilde{\varphi}(t) dt \right) \mu_{\xi}(dx) = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{\varphi}(t) \left(\int_{-\infty}^{\infty} e^{-itx} \mu_{\xi}(dx) \right) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} f_{\xi}(-t) \tilde{\varphi}(t) dt, \end{aligned}$$

так как для непрерывных функций интеграл Лебега совпадает с интегралом Римана, а для интеграла Римана известна теорема о перемене порядка интегрирования.

Лемма 8.3. Пусть последовательность характеристических функций $f_{\xi_1}(t), \dots, f_{\xi_n}(t), \dots$, случайных величин $\xi_1, \dots, \xi_n, \dots$ сходится к характеристической функции $f_{\xi}(t)$ случайной величины ξ , причем сходимость $f_{\xi_n}(t) \rightarrow f_{\xi}(t)$ равномерна в каждом конечном интервале $|t| < T$ (T фиксировано). Тогда для любой финитной функции $\varphi(x)$

$$M\varphi(\xi_n) \rightarrow M\varphi(\xi).$$

Доказательство. Для любого T справедливо равенство

$$\begin{aligned} M\varphi(\xi_n) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} f_{\xi_n}(-t) \tilde{\varphi}(t) dt = \frac{1}{2\pi} \int_{-T}^T f_{\xi_n}(-t) \tilde{\varphi}(t) dt + \\ &+ \frac{1}{2\pi} \int_{|t|>T} f_{\xi_n}(-t) \tilde{\varphi}(t) dt. \end{aligned}$$

Поэтому

$$\begin{aligned} |M\varphi(\xi_n) - M\varphi(\xi)| &\leq \left| \frac{1}{2\pi} \int_{-T}^T (f_{\xi_n}(-t) - f_{\xi}(-t)) \tilde{\varphi}(t) dt \right| + \\ &+ \frac{1}{2\pi} \int_{|t|>T} |f_{\xi_n}(-t)| \cdot |\tilde{\varphi}(t)| dt + \frac{1}{2\pi} \int_{|t|>T} |f_{\xi}(-t)| \cdot |\tilde{\varphi}(t)| dt. \end{aligned}$$

Первый интеграл стремится к нулю при $n \rightarrow \infty$ и любом T , второй и третий могут быть сделаны сколь угодно малыми при всех n выбором достаточно большого T .

Теорема 8.5. Пусть в условиях леммы 8.3. функция распределения $F_{\xi}(x)$ непрерывна. В таком случае при $n \rightarrow \infty$

$$F_{\xi_n}(x) = P\{\xi_n < x\} \rightarrow F_{\xi}(x),$$

причем сходимость равномерна по x при $-\infty < x < \infty$.

Доказательство. Для любого конечного интервала $[A, B]$ и любого $\varepsilon > 0$ существует финитная функция $\varphi_{\varepsilon}(x)$, равная 0 вне $[A, B]$ и 1 на $[A + \varepsilon, B - \varepsilon]$, заключенная между 0 и 1 в остальных точках отрезка $[A, B]$. Очевидно, что

$$M\varphi_{\varepsilon}(\xi_n) \leq P\{A \leq \xi_n < B\},$$

$$M\varphi_{\varepsilon}(\xi_n) \geq P\{A + \varepsilon \leq \xi_n < B - \varepsilon\}.$$

Следовательно,

$$\liminf_{n \rightarrow \infty} P\{A \leq \xi_n < B\} \geq F_{\xi}(B - \varepsilon) - F_{\xi}(A + \varepsilon).$$

Аналогично доказывается, что

$$\limsup_{n \rightarrow \infty} P\{A \leq \xi_n < B\} \leq F_{\xi}(B + \varepsilon) - F_{\xi}(A - \varepsilon).$$

Поскольку $F_{\xi}(x)$ непрерывна, то при $\varepsilon \rightarrow 0$ имеем

$$\lim_{n \rightarrow \infty} P\{A \leq \xi_n < B\} = \lim_{n \rightarrow \infty} P\{A \leq \xi_n < B\} = F_{\xi}(B) - F_{\xi}(A).$$

Следовательно,

$$\lim_{n \rightarrow \infty} (F_{\xi_n}(B) - F_{\xi_n}(A)) = F_{\xi}(B) - F_{\xi}(A).$$

Покажем теперь, что для любого x

$$F_{\xi_n}(x) \rightarrow F_{\xi}(x). \quad (8.6)$$

Пусть $\varepsilon > 0$; выберем A_0 и B_0 так, чтобы

$$F_{\xi}(B_0) - F_{\xi}(A_0) \geq 1 - \varepsilon.$$

В таком случае для больших n

$$F_{\xi_n}(B_0) - F_{\xi_n}(A_0) \geq 1 - 2\varepsilon.$$

Следовательно, $F_{\xi_n}(A_0) \leq 2\varepsilon$. Имеем

$$\begin{aligned} F_{\xi}(x) - F_{\xi_n}(x) &= [F_{\xi}(x) - F_{\xi}(A_0)] - \\ &- [F_{\xi_n}(x) - F_{\xi_n}(A_0)] + F_{\xi}(A_0) - F_{\xi_n}(A_0). \end{aligned}$$

Поскольку по доказанному выше

$$[F_{\xi}(x) - F_{\xi}(A_0)] - [F_{\xi_n}(x) - F_{\xi_n}(A_0)] \rightarrow 0 \quad (n \rightarrow \infty),$$

получаем

$$\lim_{n \rightarrow \infty} |F_{\xi}(x) - F_{\xi_n}(x)| \leq F_{\xi}(A_0) + F_{\xi_n}(A_0) \leq 3\varepsilon;$$

переходя к пределу при $\varepsilon \rightarrow 0$, получаем

$$F_{\xi_n}(x) \rightarrow F_{\xi}(x).$$

Докажем, наконец, что сходимость в (8.6) равномерна. Выберем точки $x_1 < x_2 < \dots < x_N$ такие, что для данного $\varepsilon > 0$

$$F_{\xi}(x_{i+1}) - F_{\xi}(x_i) < \varepsilon, \quad i = 1, 2, \dots, N-1;$$

$$F_{\xi}(x_1) - F_{\xi}(-\infty) = F_{\xi}(x_1) < \varepsilon;$$

$$F_{\xi}(\infty) - F_{\xi}(x_N) = 1 - F_{\xi}(x_N) < \varepsilon.$$

Поскольку при каждом фиксированном x верно (8.6), найдется номер n_0 такой, что при $n > n_0$

$$|F_{\xi}(x_i) - F_{\xi_n}(x_i)| < \varepsilon, \quad i = 1, \dots, N.$$

Пусть $x \in (-\infty, \infty)$ и найдется i такое, что $x_i \leq x < x_{i+1}$. Тогда

$$\begin{aligned} |F_{\xi}(x) - F_{\xi_n}(x)| &\leq |F_{\xi}(x_i) - F_{\xi_n}(x_i)| + |F_{\xi}(x) - F_{\xi}(x_i)| + \\ &+ |F_{\xi_n}(x) - F_{\xi_n}(x_i)| \leq 2\varepsilon + |F_{\xi_n}(x) - F_{\xi_n}(x_i)|. \end{aligned}$$

Однако

$$\begin{aligned} |F_{\xi_n}(x) - F_{\xi_n}(x_i)| &\leq |F_{\xi_n}(x_{i+1}) - F_{\xi_n}(x_i)| \leq |F_{\xi}(x_{i+1}) - \\ &- F_{\xi}(x_i)| + |F_{\xi_n}(x_{i+1}) - F_{\xi}(x_{i+1})| + |F_{\xi_n}(x_i) - F_{\xi}(x_i)| \leq 3\varepsilon, \end{aligned}$$

откуда

$$|F_{\xi}(x) - F_{\xi_n}(x)| \leq 5\varepsilon,$$

что и требовалось доказать. В случаях $x < x_1$ и $x > x_N$ соответствующая оценка очевидна. Теорема доказана.

Замечание 1. Не представляет труда доказать обратное предложение: если при $n \rightarrow \infty$

$$F_{\xi_n}(x) \rightarrow F_{\xi}(x),$$

то последовательность характеристических функций

$$f_{\xi_n}(t) \rightarrow f_{\xi}(t)$$

равномерно в каждом конечном интервале $|t| < T$. Это делается путем аппроксимации функции e^{itx} в выражении

$$f_{\xi_n}(t) = \int_{-\infty}^{\infty} e^{itx} \mu_{\xi_n}(dx) \quad (8.7)$$

функциями, постоянными на интервалах (как при построении римановой интегральной суммы). Правда, тут возникает небольшая трудность, так как пределы интеграла (8.7) бесконечны. Читателю рекомендуется преодолеть ее самостоятельно; мы же не будем приводить полного доказательства, поскольку высказанное утверждение в дальнейшем не используется (см. Гнеденко [8]).

Замечание 2. Из теоремы 8.5 следует, конечно, что если у двух законов распределения совпадают характеристические функции, то совпадают и сами законы распределения. Применим этот факт для доказательства следующей теоремы.

Теорема 8.6. *Сумма двух независимых случайных величин $\xi_1 \sim N(a_1, \sigma_1^2)$ и $\xi_2 \sim N(a_2, \sigma_2^2)$ имеет нормальное распределение $N(a_1 + a_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.*

Доказательство. В силу теорем 8.3, 8.4 и определения 8.1 имеем

$$f_{\xi_1}(t) = e^{ita_1} e^{-\frac{t^2 \sigma_1^2}{2}}, \quad f_{\xi_2}(t) = e^{ita_2} e^{-\frac{t^2 \sigma_2^2}{2}}.$$

Следовательно,

$$f_{\xi_1 + \xi_2}(t) = e^{it(a_1 + a_2)} e^{-\frac{t^2}{2}(\sigma_1^2 + \sigma_2^2)},$$

что отвечает $N(a_1 + a_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

Замечание. Поскольку $a_i = M\xi_i$, $\sigma_i = \sqrt{D\xi_i}$, $i = 1, 2$, то из общих свойств математических ожиданий и дисперсий следует, что

$$M(\xi_1 + \xi_2) = a_1 + a_2, \quad D(\xi_1 + \xi_2) = \sigma_1^2 + \sigma_2^2.$$

Применение характеристических функций нужно лишь, чтобы установить нормальность распределения $\xi_1 + \xi_2$.

8.5. Формулировка и доказательство центральной предельной теоремы. Мы сформулируем эту теорему в *условиях Ляпунова*. Эти условия подобраны таким образом, чтобы обеспечить сходимость соответствующей последовательности характеристических функций к функции $e^{-t^2/2}$. Они являются, следовательно, достаточными условиями справедливости теоремы. Однако эти условия чрезвычайно близки и к необходимым.

Сформулируем условия теоремы Ляпунова. Пусть дана последовательность независимых случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$, для каждой из которых существует математическое ожидание $M\xi_k = a_k$, дисперсия $D\xi_k = \sigma_k^2$ и третий центральный абсолютный момент $M|\xi_k - a_k|^3$. Положим

$$B_n^2 = D(\xi_1 + \dots + \xi_n) = \sum_{k=1}^n D\xi_k = \sum_{k=1}^n \sigma_k^2.$$

Пусть выполнено следующее условие Ляпунова:

$$\frac{1}{B_n^3} \sum_{i=1}^n M |\xi_k - a_k|^3 \rightarrow 0 \quad (n \rightarrow \infty). \quad (8.8)$$

Тогда справедлива следующая теорема, называемая теоремой Ляпунова (естественнее было бы называть ее центральной предельной теоремой в форме Ляпунова).

Теорема 8.7. При выполнении условия (8.8)

$$P(s_n^* < x) \rightarrow \Phi(x)$$

при $n \rightarrow \infty$ равномерно по x , где через s_n^* обозначена нормированная сумма случайных величин ξ_1, \dots, ξ_n :

$$s_n^* = \frac{1}{B_n} (\xi_1 + \dots + \xi_n - (a_1 + \dots + a_n)).$$

Доказательство. Рассмотрим случайные величины

$$\eta_{nk} = \frac{1}{B_n} (\xi_k - a_k), \quad k = 1, \dots, n.$$

Очевидно, что $s_n^* = \sum_{k=1}^n \eta_{nk}$. При этом

$$M\eta_{nk} = 0, \quad D\eta_{nk} = \frac{1}{B_n^2} D\xi_k, \quad M|\eta_{nk}|^3 = \frac{1}{B_n^3} M|\xi_k - a_k|^3.$$

Поэтому (см. теорему 8.2 и ее доказательство) характеристическая функция $f_{nk}(t)$ случайной величины η_{nk} обладает свойствами

$$\begin{aligned} \frac{d}{dt} f_{nk}(t) |_{t=0} &= 0, \\ \frac{d^2}{dt^2} f_{nk}(t) |_{t=0} &= -\frac{1}{B_n^2} D\xi_k, \\ \left| \frac{d^3}{dt^3} f_{nk}(t) \right| &\leq M|\eta_{nk}|^3 = \frac{1}{B_n^3} M|\xi_k - a_k|^3. \end{aligned}$$

Применяя формулу Тейлора с остаточным членом для вещественной и мнимой части функции $f_{nk}(t)$, получаем следующее утверждение. Равномерно в любом интервале $|t| < T$

$$f_{nk}(t) = 1 - \frac{D\xi_k}{B_n^2} \frac{t^2}{2} + R_{nk}(t),$$

где

$$|R_{nk}(t)| \leq C \frac{M |\xi_k - a_k|^3}{B_n^3} = CM |\eta_{nk}|^3.$$

Из условия Ляпунова (8.8) следует, что $M |\eta_{nk}|^3 \rightarrow 0$ при $n \rightarrow \infty$ равномерно по k , но тогда и $\frac{D\xi_k}{B_n^2} = D\eta_{nk} \rightarrow 0$ при $n \rightarrow \infty$ равномерно по k . Действительно, для любого $\varepsilon > 0$

$$\begin{aligned} D\eta_{nk} &= M\eta_{nk}^2 \leq \varepsilon^2 P\{|\eta_{nk}| < \varepsilon\} + \int_{|x| \geq \varepsilon} x^2 \mu_{\eta_{nk}}(dx) \leq \\ &\leq \varepsilon^2 + \frac{1}{\varepsilon} \int_{|x| \geq \varepsilon} |x|^3 \mu_{\eta_{nk}}(dx) \leq \varepsilon^2 + \frac{1}{\varepsilon} M |\eta_{nk}|^3 \rightarrow \varepsilon^2 \end{aligned}$$

при $n \rightarrow \infty$ равномерно по k .

Положим $z_{nk} = -\frac{D\xi_k}{B_n^2} \frac{t^2}{2} + R_{nk}(t)$. При $n \rightarrow \infty$ имеем $|z_{nk}| \rightarrow 0$ равномерно по k ,

$$f_{nk}(t) = 1 + z_{nk},$$

$$\ln f_{nk}(t) = z_{nk} - \frac{z_{nk}^2}{2} + \frac{z_{nk}^3}{3} - \dots,$$

$$|\ln f_{nk}(t) - z_{nk}| \leq \frac{|z_{nk}|^2}{2} \sum_{s=0}^{\infty} |z_{nk}|^s = |z_{nk}|^2 \frac{1}{2(1 - |z_{nk}|)},$$

т. е. при достаточно большом n

$$|\ln f_{nk}(t) - z_{nk}| \leq |z_{nk}|^2. \quad (8.9)$$

Вычислим теперь характеристическую функцию $f_{s_n}^*(t) = \prod_{k=1}^n f_{nk}(t)$.

Очевидно, одно из значений логарифма комплексной величины $f_{s_n}^*(t)$ имеет вид

$$\ln f_{s_n}^*(t) = \sum_{k=1}^n \ln f_{nk}(t) = \sum_{k=1}^n z_{nk} + \sum_{k=1}^n (\ln f_{nk}(t) - z_{nk}).$$

Имеем при $n \rightarrow \infty$

$$\sum_{k=1}^n z_{nk} = - \sum_{k=1}^n \frac{D\xi_k}{B_n^2} \frac{t^2}{2} + \sum_{k=1}^n R_{nk}(t) = -\frac{t^2}{2} + \sum_{k=1}^n R_{nk}(t).$$

Однако (именно здесь действует условие Ляпунова)

$$\left| \sum_{k=1}^n R_{nk}(t) \right| \leq \sum_{k=1}^n |R_{nk}(t)| \leq C \sum_{k=1}^n \frac{M |\xi_k - a_k|^3}{B_n^3} \rightarrow 0.$$

Далее, в силу (8.9),

$$\begin{aligned} \left| \sum_{k=1}^n (\ln f_{nk}(t) - z_{nk}) \right| &\leq \sum_{k=1}^n |z_{nk}|^2 \leq \max_{k=1, \dots, n} |z_{nk}| \cdot \sum_{k=1}^n |z_{nk}| \leq \\ &\leq \max_{k=1, \dots, n} |z_{nk}| \cdot \left(\sum_{k=1}^n \frac{D\xi_k}{B_n^2} \cdot \frac{t^2}{2} + \sum_{k=1}^n |R_{nk}(t)| \right) \rightarrow 0 \end{aligned}$$

при $n \rightarrow \infty$ в силу того, что $\max_{k=1, \dots, n} |z_{nk}| \rightarrow 0$.

Таким образом, $\ln f_{s_n}^*(t) \rightarrow -\frac{t^2}{2}$, следовательно, $f_{s_n}^*(t) \rightarrow e^{-t^2/2}$.

В силу теоремы 8.5, из этого вытекает доказываемая теорема.

З а м е ч а н и е. Условие Ляпунова (8.8) не только обеспечивает соотношение

$$\sum_{k=1}^n |R_{nk}(t)| \rightarrow 0,$$

но из него вытекает еще, что

$$D\eta_{nk} = \frac{D\xi_k}{B_n^2} = \frac{D\xi_k}{\sum_{k=1}^n D\xi_k} \rightarrow 0,$$

т. е. что дисперсия каждой случайной величины ξ_k составляет лишь малую долю в общей дисперсии суммы $\xi_1 + \dots + \xi_n$. Если бы это было не так, например величина ξ_1 колебалась бы существенно сильнее, чем остальные величины ξ_2, \dots, ξ_n , то закон распределения суммы $\xi_1 + \dots + \xi_n$ определялся бы в основном величиной ξ_1 . В этом случае не было бы основания ожидать нормального распределения суммы. Если считать случайные величины примерно равноправными, то это естественно в математической форме выразить так: существуют константы c_1, c_2 и d_1, d_2 такие, что для всех k

$$0 < c_1 \leq D\xi_k \leq c_2, \quad 0 < d_1 \leq M |\xi_k - a_k|^3 \leq d_2.$$

В этом случае

$$B_n^2 = \sum_{k=1}^n D\xi_k \geq c_1 n,$$

а

$$\sum_{k=1}^n M |\xi_k - a_k|^3 \leq d_2 n.$$

Поэтому

$$\frac{\sum_{k=1}^n M |\xi_k - a_k|^3}{B_n^3} \leq \frac{d_2 n}{(c_1 n)^{3/2}} \sim \frac{1}{\sqrt{n}} \rightarrow 0,$$

т. е. условие Ляпунова выполнено. В частности, оно выполнено для одинаково распределенных случайных величин, имеющих третий момент.

Частным случаем теоремы Ляпунова является следующая теорема Муавра — Лапласа.

Теорема 8.8. Пусть μ — число успехов в n испытаниях Бернулли с вероятностью успеха в отдельном испытании p . При $n \rightarrow \infty$ равномерно по x

$$P \left\{ \frac{\mu - np}{\sqrt{npq}} < x \right\} \rightarrow \Phi(x).$$

Доказательство. Вспомним, что

$$\mu = \mu_1 + \mu_2 + \dots + \mu_n,$$

где случайные величины μ_k принимают значения 1, если в k -том испытании был успех, и значение 0 в противном случае. Величины μ_1, \dots, μ_n независимы, одинаково распределены, $M\mu_k = p$, $D\mu_k = pq$, $q = 1 - p$. Применяя теорему Ляпунова, получаем доказываемое утверждение.

З а м е ч а н и е. Отметим один случай, когда нельзя применять центральную предельную теорему. Часто интересуются такими значениями x , при которых величина вероятности

$$P \left\{ \frac{\mu - np}{\sqrt{npq}} < x \right\} = P(x)$$

будет очень близкой к нулю или к единице. Например, если x таково, что $P(x) \approx 0,001$, то нам желательно, чтобы ошибка от приближенной замены $P(x)$ на $\Phi(x)$ составляла несколько процентов от 0,001. Этого, как правило, не будет. Можно утверждать, что разность

$$P(x) - \Phi(x) \rightarrow 0,$$

но неверно, что отношение

$$\frac{P(x)}{\Phi(x)} \rightarrow 1$$

равномерно по всем x . Практически отсюда следует, что пользоваться $\Phi(x)$ в качестве приближенного значения для случая ма-

лых $P(x)$ (или близких к 1 $P(x)$) ни в коем случае не следует. Например, пусть мы проектируем некоторую систему, рассчитывая ее на данное значение x , причем если окажется, что наступит событие

$$\left\{ \frac{\mu - np}{\sqrt{npq}} \geq x \right\} = \{ \mu \geq np + x \sqrt{npq} \},$$

то это ведет к аварии. Тогда мы, естественно, желаем, чтобы x удовлетворяло условию

$$P \left\{ \frac{\mu - np}{\sqrt{npq}} \geq x \right\} = 1 - P(x) < \alpha,$$

где α — малое число. Если α мало (имеет порядок 0,001), то замена $1 - P(x)$ на $1 - \Phi(x)$ приведет к ошибочной оценке $1 - P(x)$, причем ошибка может быть порядка сотен процентов (т. е. в несколько раз).

В теории вероятностей имеются теоремы, называемые теоремами о больших отклонениях, которые теоретически должны применяться вместо центральной предельной теоремы в рассматриваемом случае (см. книгу В. Феллера [22], т. 2). Однако степень их практической полезности не ясна. Прежде всего, это связано с тем, что для вычисления вероятности редкого и опасного события рискованно пользоваться упрощенной вероятностной схемой типа испытаний Бернулли. (Это не значит, что во всех случаях нужно стараться искать улучшенную более сложную вероятностную схему: более сложная схема легко может дать еще худший результат, чем более простая. Лучше, может быть, вовсе отказаться от вероятностных оценок.)

Далее, асимптотические формулы, даваемые теоремами о больших отклонениях, начинают действовать при слишком большом числе испытаний. В общем оценка малых (или близких к 1) вероятностей требует особой аккуратности (как чисто математической, так и естественнонаучной) и далеко не всегда является научно возможной при современном уровне развития науки.

8.6. Что можно сказать о применимости нормального закона?

При обсуждении теоремы Пуассона мы выяснили разницу между «хорошей» и «плохой» математической теоремой (с точки зрения естествоиспытателя): «хорошая» теорема не боится некоторого нарушения ее условий. Центральная предельная теорема в этом смысле является «превосходной». Ее основное условие — предположение о независимости слагаемых $\xi_1, \xi_2, \dots, \xi_n$. Это предположение можно довольно сильно ослабить. Математические работы с такой целью были начаты С. Н. Бернштейном и сейчас достигли большого развития, причем некоторые направления в этой области не потеряли до сих пор научного интереса.

По общему мнению, нормально распределены ошибки измерений. Каждое измерение зависит от многих, частично зависимых, частично независимых факторов, которые дают вклад в суммар-

ную ошибку. Не обязательно, конечно, суммарная ошибка является суммой $\xi_1 + \dots + \xi_n$ различных факторов. Правильнее ее считать некоторой функцией $f(\xi_1, \dots, \xi_n)$ от них. Однако величины $\xi_1, \xi_2, \dots, \xi_n$ естественно считать мало отклоняющимися от каких-то неслучайных значений a_1, \dots, a_n :

$$\xi_k = a_k + \eta_k; \quad k = 1, \dots, n,$$

где η_k малы. Тогда

$$f(\xi_1, \dots, \xi_n) \approx f(a_1, \dots, a_n) + \sum_{k=1}^n \frac{\partial f(a_1, \dots, a_n)}{\partial a_k} \eta_k.$$

Сумма, стоящая в правой части, естественно наводит на мысль о нормальности распределения.

Нужно сказать, что в развитых экспериментальных науках (астрономия, геодезия) применяются специальные приемы типа усреднения результатов измерений, которые наряду с повышением качества измерений обеспечивают нормальность ошибки.

В общем нормальный закон является бесспорно самым универсальным среди всех законов распределения. Не нужно только требовать от нормального закона больше, чем он может дать. Например, в конце предыдущего пункта отмечено, что нельзя ожидать хорошего согласия с нормальным законом на «хвостах» распределения, т. е. в области слишком больших или слишком малых значений случайной величины (для конкретности это объяснялось для испытаний Бернулли, но так бывает в любых вероятностных схемах и обычно на практике).

Грубейшей ошибкой является применение нормального закона в тех условиях, когда наблюдения имеют некоторую неопределенность, но эта неопределенность не есть случайность в смысле прикладной теории вероятностей, т. е. не обладает статистической устойчивостью. Например, содержание полезного компонента в руде может постепенно закономерно возрастать от периферии к центру месторождения. Крайне нелепо было бы обрабатывать данные отдельных его определений (подверженных как случайным ошибкам, так и закономерному возрастанию) таким образом, как если бы это были реализации одной и той же случайной величины, подчиненной нормальному закону распределения.

§ 9

ПРИМЕНЕНИЯ ЦЕНТРАЛЬНОЙ ПРЕДЕЛЬНОЙ ТЕОРЕМЫ

Обыкновенно центральная предельная теорема используется по следующей схеме. Мы интересуемся вероятностью того, что сумма $S_n = \xi_1 + \dots + \xi_n$ примет значение, лежащее в некотором интервале $[A, B]$: $P\{A \leq S_n \leq B\}$. Вычисляем MS_n и DS_n , нормированную сумму s_n^* и пользуемся очевидным тождеством

$$P\{A \leq S_n \leq B\} = P\left\{\frac{A - MS_n}{\sqrt{DS_n}} \leq s_n^* \leq \frac{B - MS_n}{\sqrt{DS_n}}\right\}.$$

Правую часть последнего равенства приближенно заменяем на

$$\Phi\left(\frac{B - MS_n}{\sqrt{DS_n}}\right) - \Phi\left(\frac{A - MS_n}{\sqrt{DS_n}}\right),$$

где Φ — функция Лапласа. Оценка ошибки при этом производится в чрезвычайно редких случаях (в основном, в чисто математических работах).

9.1. Правило 3σ . С помощью таблиц нормального закона легко устанавливаем, что $\Phi(3) - \Phi(-3) = 0,997$. Иными словами, если

$$\frac{A - MS_n}{\sqrt{DS_n}} = -3, \quad \frac{B - MS_n}{\sqrt{DS_n}} = +3$$

или, что то же самое,

$$A = MS_n - 3\sqrt{DS_n}, \quad B = MS_n + 3\sqrt{DS_n},$$

то

$$P\{A \leq S_n \leq B\} \approx 0,997.$$

Иначе говоря,

$$P\{|S_n - MS_n| \leq 3\sqrt{DS_n}\} \approx 0,997, \quad (9.1)$$

т. е. практически достоверно, что S_n отклоняется от MS_n не более чем на $3\sqrt{DS_n}$. Обычно обозначают $\sqrt{DS_n} = \sigma$, поэтому изложенное правило называется «правилом 3σ ».

Для отклонения на $2\sigma = 2\sqrt{DS_n}$ имеем

$$P\{|S_n - MS_n| \leq 2\sqrt{DS_n}\} \approx 0,965.$$

Конечно, как было отмечено в конце предыдущего параграфа, оценка очень близких к 1 вероятностей с помощью центральной предельной теоремы обычно дает неправильный результат. Поэтому, строго говоря, «правило 3σ » действует лишь в том случае, когда в точности выполняется нормальный закон распределения.

Вместо значения 0,997 вероятность (9.1) легко может равняться 0,990, а то и 0,980 (могут быть, в зависимости от конкретных случайных величин ξ_1, \dots, ξ_n , и большие отклонения). Это надо иметь в виду при использовании изложенного правила.

Для одинаково распределенных случайных величин с $M\xi_i = a \neq 0$ величина $MS_n = na$ имеет порядок n , а величина $\sqrt{DS_n} = \sqrt{nD\xi_i}$ имеет порядок \sqrt{n} . Таким образом, при больших n преобладающую роль играет неслучайное число $MS_n = na$, так как отклонение S_n от na имеет порядок \sqrt{n} .

9.2. Применение к среднему арифметическому. Пусть ξ_1, \dots, ξ_n — независимые случайные величины, причем $M\xi_i = a$, $D\xi_i = \sigma^2$ не зави-

сят от номера i , $\varepsilon > 0$. В таком случае закон больших чисел утверждает, что при $n \rightarrow \infty$

$$P \left\{ \left| \frac{\xi_1 + \dots + \xi_n}{n} - a \right| > \varepsilon \right\} \rightarrow 0.$$

Строго говоря, для применения центральной предельной теоремы в том виде, в каком мы ее знаем, величины ξ_1, \dots, ξ_n должны быть независимы в совокупности, в то время как для закона больших чисел нужна лишь попарная независимость. Однако, предполагая, что центральная предельная теорема имеет место, получаем

$$\begin{aligned} & P \left\{ \left| \frac{\xi_1 + \dots + \xi_n}{n} - a \right| > \varepsilon \right\} = \\ &= P \left\{ \left| \frac{\xi_1 + \dots + \xi_n - na}{\sigma \sqrt{n}} \right| > \frac{\varepsilon \sqrt{n}}{\sigma} \right\} = \\ &= P \left\{ s_n^* < -\frac{\varepsilon \sqrt{n}}{\sigma} \right\} + P \left\{ s_n^* > \frac{\varepsilon \sqrt{n}}{\sigma} \right\} \approx 2\Phi \left(-\frac{\varepsilon \sqrt{n}}{\sigma} \right). \end{aligned}$$

Например, при $\frac{\varepsilon \sqrt{n}}{\sigma} = 3$, т. е. $\varepsilon = \frac{3\sigma}{\sqrt{n}}$, эта вероятность равна 0,997 (в той мере, в какой можно пользоваться правилом 3σ). Заметим, что ε убывает, как корень $\sqrt{\frac{1}{n}}$. Если интерпретировать ξ_1, \dots, ξ_n как наблюдения, то мы видим, что ошибка от замены a на $\frac{1}{n}(\xi_1 + \dots + \xi_n)$ обратно пропорциональна квадратному корню из числа наблюдений. Любопытно, что этот вид зависимости от числа наблюдений почти всегда имеет место в рамках чисто статистического подхода. Иногда бывает лучше сделать более точный прибор, чем надеяться увеличить точность за счет увеличения числа наблюдений. Статистикой надо пользоваться тогда, когда исчерпаны технические возможности. Любопытно, что в этом случае и ошибки наблюдений лучше укладываются в теоретико-вероятностные модели.

9.3. Вероятность и частота. Мы можем, наконец, ответить в рамках модели испытаний Бернулли), насколько сильно может отличаться частота от вероятности при данном числе наблюдений n . Имеем

$$\begin{aligned} P \left\{ \left| \frac{\mu}{n} - p \right| > \varepsilon \right\} &= P \left\{ \left| \frac{\mu - np}{\sqrt{npq}} \right| > \frac{\varepsilon \sqrt{n}}{\sqrt{pq}} \right\} \approx \\ &\approx 2\Phi \left(-\frac{\varepsilon \sqrt{n}}{\sqrt{pq}} \right). \end{aligned}$$

Поскольку при $p + q = 1$, очевидно, $pq \leq \frac{1}{4}$, эта вероятность, очевидно, не превосходит $2\Phi(-2\varepsilon\sqrt{n})$. Иными словами,

$$\begin{aligned} P\left\{\frac{\mu}{n} - \varepsilon \leq p \leq \frac{\mu}{n} + \varepsilon\right\} &\approx \\ &\approx 1 - 2\Phi\left(-\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) \leq 1 - 2\Phi(-2\varepsilon\sqrt{n}). \end{aligned}$$

Таким образом, зная число успехов μ в n испытаниях Бернулли, мы можем построить интервал

$$\left(\frac{\mu}{n} - \varepsilon, \frac{\mu}{n} + \varepsilon\right),$$

который будет накрывать неизвестное значение вероятности p успеха в одном испытании с любой заданной вероятностью $1 - \alpha$ (α — малое, но положительное число). Для этого нужно выбрать $\varepsilon = \varepsilon_\alpha$ из соотношения: $2\Phi(-2\varepsilon_\alpha\sqrt{n}) \leq \alpha$. Тогда

$$P\left\{\frac{\mu}{n} - \varepsilon_\alpha \leq p \leq \frac{\mu}{n} + \varepsilon_\alpha\right\} \geq 1 - \alpha.$$

Интервал $\left(\frac{\mu}{n} - \varepsilon_\alpha, \frac{\mu}{n} + \varepsilon_\alpha\right)$ называется *доверительным интервалом* для p с коэффициентом доверия (коэффициентом надежности) $1 - \alpha$. Итак, доверительный интервал есть интервал со случайными (зависящими от μ) концами, который накрывает неслучайное, но нам неизвестное, значение p с вероятностью $\geq 1 - \alpha$.

Каким нужно выбирать α ? Обычно берут одно из значений 0,05; 0,01; 0,001. Смысл ε_α — указать порядок возможной ошибки от замены p на $\frac{\mu}{n}$. При переходе от $\alpha = 0,05$ к $\alpha = 0,001$ доверительный интервал удлиняется примерно в полтора раза (но при $\alpha = 0,001$ оценка $P\left\{\frac{\mu}{n} - \varepsilon_\alpha \leq p \leq \frac{\mu}{n} + \varepsilon_\alpha\right\}$ с помощью теоремы Муавра — Лапласа становится крайне ненадежной). Поэтому довольно безразлично, какое значение α выбрать.

Можно представить себе, что мы, не хотим увеличить размер доверительного интервала, заменив $\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}$ на $2\varepsilon\sqrt{n}$. В таком случае, поскольку речь все равно идет лишь о порядке длины, можно (при больших n) вставить вместо p значение $\frac{\mu}{n}$. Иными словами, можно ε_α выбрать из соотношения

$$2\Phi\left(-\frac{\varepsilon_{\alpha}\sqrt{n}}{\sqrt{\frac{\mu}{n}\left(1-\frac{\mu}{n}\right)}}\right)\leq\alpha.$$

9.4. Сравнение вероятностей. Принципиальный вопрос проверки статистической однородности обычно решается следующим образом. Пусть мы провели n опытов, в каждом из которых могло наступить или не наступить событие A . Разделим имеющийся у нас материал на две группы (группы, конечно, должны выбираться из естественнонаучных соображений, в зависимости от наших представлений, каким именно образом может нарушаться статистическая однородность).

Пусть в первой группе будет n_1 испытаний и μ_1 успехов, во второй группе n_2 испытаний и μ_2 успехов. Конечно, частоты $\frac{\mu_1}{n_1}$ и $\frac{\mu_2}{n_2}$, как правило, будут различны. Спрашивается, можно ли считать, что это различие вызвано чисто случайными причинами или оно настолько велико, что мы должны отвергнуть гипотезу $p_1=p_2$ равенства вероятностей успеха. Построим критерий для проверки гипотезы $p_1=p_2$.

Очевидно, критерий должен иметь вид

$$\left|\frac{\mu_1}{n_1}-\frac{\mu_2}{n_2}\right|>x_{\alpha},$$

где x_{α} надо выбрать таким образом, чтобы вероятность

$$P\left\{\left|\frac{\mu_1}{n_1}-\frac{\mu_2}{n_2}\right|>x_{\alpha}\mid p_1=p_2\right\}\leq\alpha,$$

причем α — уровень значимости.

При верной гипотезе $p_1=p_2=p$ имеем

$$M\left(\frac{\mu_1}{n_1}\right)=p, \quad M\left(\frac{\mu_2}{n_2}\right)=p, \quad D\left(\frac{\mu_1}{n_1}\right)=\frac{pq}{n_1}, \quad D\left(\frac{\mu_2}{n_2}\right)=\frac{pq}{n_2},$$

где p обозначает общее значение p_1 и p_2 (нам p неизвестно), $q=1-p$.

Считая, что величины μ_1 и μ_2 независимы, мы получаем, что разность $\frac{\mu_1}{n_1}-\frac{\mu_2}{n_2}$ двух приблизительно нормальных независимых величин имеет примерно нормальное распределение с параметрами

$$M\left(\frac{\mu_1}{n_1}-\frac{\mu_2}{n_2}\right)=0, \quad \sqrt{D\left(\frac{\mu_1}{n_1}-\frac{\mu_2}{n_2}\right)}=\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

Поэтому x_{α} следует выбрать из соотношения

$$2\Phi\left(-\frac{x_\alpha}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}\right)=\alpha. \quad (9.2)$$

Беда здесь заключается в том, что нам неизвестно p . Поэтому разумно подставить вместо p его оценку:

$$p = \frac{\mu_1 + \mu_2}{n_1 + n_2}.$$

Обозначим через x'_α найденное таким образом значение x_α . Если $n_1 + n_2$ велико, то $\frac{\mu_1 + \mu_2}{n_1 + n_2}$ близко к p и x'_α близко к x_α . Поэтому при замене x_α на x'_α мы будем проверять гипотезу примерно на уровне значимости α .

9.5. Окончание исследования игры в 10 и 20 коп. В примере 3.1 § 3 мы нашли, что первому игроку выгодно прятать 10 коп. с вероятностью 7/12 и 20 коп. с вероятностью 5/12. При этом второму игроку безразлично, как поступать, в смысле математического ожидания выигрыша, и мы предположим, что он с вероятностью 1/2 называет 10 коп. или 20 коп. Спрашивается, сколько раз надо повторить игру, чтобы выигрыш первого игрока был не менее 1 рубля с вероятностью, скажем, 0,975.

Случайная величина ξ_i , равная выигрышу первого игрока при i -том повторении игры, имеет следующее распределение:

$$\begin{pmatrix} -10 & +15 & -20 \\ 7/24 & 1/2 & 5/24 \end{pmatrix}.$$

Мы помним, что $M\xi_i = 5/12$; вычислим

$$D\xi_i = M\xi_i^2 - (M\xi_i)^2 \approx 225.$$

Число повторений игры n надо выбрать из соотношения

$$P\{S_n = \xi_1 + \dots + \xi_n \geq 100\} = P\left\{s_n^* \geq \frac{100 - \frac{5}{12}n}{\sqrt{225n}}\right\} = 0,975.$$

Иными словами,

$$\left(100 - \frac{5}{12}n\right) \frac{1}{\sqrt{225n}} \leq -1,96.$$

Отсюда подбором получаем $n \approx 5150$.

Следовательно, если повторять игру 10 раз в минуту, то придется играть без отдыха 8,5 часа, чтобы выиграть рубль. В случае, если первый игрок не сумеет точно выдержать вероятности 7/12 и 5/12, а сообразит на 1/2 и 1/2, то он вообще ничего не выиграет. Ясно, что легче и проще заработать рубль каким-либо другим способом.

§ 10

ВЫБОРКА. ОЦЕНКА ПАРАМЕТРОВ

10.1. Выборка. Математические модели теории вероятностей исходят из пространства элементарных событий с заданной заранее вероятностной мерой. Ожидается, что вероятности тех или иных событий, равные значениям этой меры, можно будет интерпретировать как частоты. При практическом применении теории вероятностей естественно возникают два вопроса: 1) можно ли говорить о вероятностях рассматриваемых событий, т. е. обладают ли их частоты свойством статистической устойчивости; 2) как найти эти вероятности из экспериментальных данных.

Оба эти вопроса мы уже рассматривали применительно к испытаниям Бернулли. Сейчас мы рассмотрим несколько более общую ситуацию, в которой результат отдельного опыта не обязательно является одним из двух событий «успех» или «неудача», а может быть значением измерения, т. е. вещественным числом. Конечно, следует рассмотреть не один эксперимент, а несколько.

Итак, пусть исследователь произвел n опытов, каждый из которых состоял в некотором измерении, и получил результаты x_1, x_2, \dots, x_n . Спрашивается, надо ли их обрабатывать статистически и каким образом. Сам способ статистической обработки, равно как и ценность ее результатов, полностью зависит от положенной в основу вероятностной модели, которая должна объяснить вероятностную структуру наблюдений x_1, x_2, \dots, x_n .

Простейшая и наиболее надежная модель относится к тому случаю, когда заранее известно, что во всех n опытах измерялась одна и та же величина (например, среднее расстояние от Земли до Солнца), так что различие в наблюдениях x_1, \dots, x_n объясняется лишь ошибками наблюдений. Далее, предполагается, что во всех n опытах условия эксперимента оставались одинаковыми, а сами эксперименты были независимыми друг от друга. Заметим, что при измерении расстояния от Земли до Солнца наблюдатели, жившие в XVIII и XX вв., пользовались совершенно различными техническими средствами, так что условия эксперимента нельзя считать одинаковыми. (Следует рассматривать лишь измерения, проводимые с помощью примерно одинаковых средств.)

Сформулированные на естественнонаучном языке предпосылки о неизменности измеряемой величины и условий опыта и о независимости друг от друга результатов эксперимента на языке теории вероятностей превращаются в следующую основную предпосылку: *результаты n экспериментов x_1, \dots, x_n являются независимыми в совокупности случайными величинами с одной и той же функцией распределения $F(x)$* . Кратко это предположение выражают так: *числа x_1, \dots, x_n образуют выборку с теоретическим законом распределения $F(x)$* , или *числа x_1, \dots, x_n образуют выборку из распределения $F(x)$* .

Строго говоря, результаты наблюдений следовало бы обозначить $\xi_1, \xi_2, \dots, \xi_n$, если придерживаться введенных ранее обозначений. Однако в математической статистике сложилась досадная традиция, в силу которой x_1, x_2, \dots, x_n обозначают иногда *случайные величины*, а иногда *неслучайные числа* (например, переменные интегрирования) и притом в одной и той же формуле. Читатель должен научиться отличать эти два случая. Поскольку настоящая книга должна облегчить обращение к имеющимся учебникам по математической статистике, пришлось и здесь принять эту традицию.

Например, если каждая из случайных величин x_1, \dots, x_n имеет плотность распределения $p(x)$, A — подмножество n -мерного евклидова пространства, то

$$P\{(x_1, \dots, x_n) \in A\} = \int \dots \int_A p(x_1) \dots p(x_n) dx_1 \dots dx_n, \quad (10.1)$$

причем в левой части этого выражения x_1, \dots, x_n обозначают случайные величины, а в правой части — переменные интегрирования.

Если мы говорим, что x_1, x_2, \dots, x_n образуют выборку, то тем самым мы принимаем некоторую вероятностную модель для результатов наблюдения. В частности, мы объявляем, что имеет смысл рассматривать вероятности, стоящие в левой части (10.1). Иными словами, мы объявляем, что в принципе совокупность из n наблюдений может быть повторена сколько угодно раз, причем n -мерный вектор, получающийся при каждом повторении n наблюдений, будет попадать в подмножество A с частотой, примерно равной $P\{(x_1, \dots, x_n) \in A\}$.

Грубой ошибкой, следовательно, является применение понятия выборки и связанных с этим понятием методов статистической обработки данных в тех случаях, когда говорить о повторениях совокупности из n наблюдений в условиях статистической однородности не имеет смысла. Несколько позже мы будем говорить о возможном изменении теоретико-вероятностной модели в том случае, когда условия опыта систематически меняются (метод наименьших квадратов).

Сейчас же мы предположим, что x_1, \dots, x_n образуют выборку, и поставим вопрос о нахождении закона распределения $F(x)$. Практически обычно бывает, что известны лишь результаты опытов x_1, \dots, x_n и по этим результатам нужно найти $F(x)$. Основным принципом здесь является принцип подбора $F(x)$ из некоторого параметрического семейства законов распределения. Предполагается, что $F(x)$ не может быть произвольной функцией распределения, а определяется небольшим числом (не более трех-четырех) параметров. Например, нормальный закон распределения $N(a, \sigma)$ определяется двумя параметрами — математическим ожиданием a (в статистике математическое ожидание называют *средним*) и дисперсией σ^2 . Закон Пуассона определяется единственным парамет-

ром — средним λ . Чаще всего в качестве параметров берутся моменты

$$a_k = \int_{-\infty}^{\infty} x^k dF(x), \quad k = 1, 2, \dots$$

Величины a_k называются *теоретическими моментами*. Очевидно, что среднее a есть момент a_1 , а дисперсия $\sigma^2 = a_2 - a_1^2$.

Общий подход состоит в том, что семейство законов распределения, из которого нужно подобрать $F(x)$, определяют заранее из каких-либо соображений (например, решают, что теоретический закон является нормальным), а параметры подбирают исходя из конкретных значений x_1, \dots, x_n . В статистике вводятся (в отличие от только что введенных «теоретических» или «истинных») понятия эмпирической (выборочной) функции распределения, эмпирических (выборочных) среднего, дисперсии, моментов и т. д. Удобным мнемоническим правилом для запоминания всех этих определений является следующее: образуем фиктивную «случайную величину», принимающую каждое значение x_1, \dots, x_n с вероятностью $\frac{1}{n}$; тогда «выборочные» характеристики являются соответствующими характеристиками этой случайной величины. Например, *эмпирической функцией распределения* $F_n(x)$ называется функция распределения этой фиктивной случайной величины

$$F_n(x) = \frac{\text{число } x_i \text{ таких, что } x_i < x}{n},$$

где x — любое вещественное число.

Эмпирическим (выборочным) средним называется

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i;$$

эмпирической дисперсией называется величина

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - (\bar{x})^2;$$

эмпирическим k -тым моментом α_k называется величина

$$\alpha_k = \frac{1}{n} \sum_{i=1}^n x_i^k.$$

Важность эмпирических характеристик заключается в том, что они близки (при большом n) к соответствующим теоретическим величинам. Конечно, поскольку выборочные характеристики явля-

ются случайными величинами, а теоретические характеристики — числа, то близость может пониматься лишь в смысле сходимости по вероятности.

Лемма 10.1. Пусть последовательность случайных величин $\{\xi_n\}$ сходится по вероятности к числу a , последовательность $\{\eta_n\}$ сходится по вероятности к числу b , а функция $f(x, y)$ двух переменных x и y непрерывна в точке $x=a, y=b$. В таком случае

$$f(\xi_n, \eta_n) \rightarrow f(a, b)$$

по вероятности.

Доказательство. Для любого $\varepsilon > 0$ выберем $\delta > 0$ такое, что при $|x-a| < \delta, |y-b| < \delta$

$$|f(x, y) - f(a, b)| < \varepsilon.$$

Имеем следующее включение для событий:

$$\{|f(\xi_n, \eta_n) - f(a, b)| > \varepsilon\} \subseteq \{|\xi_n - a| \geq \delta\} \cup \{|\eta_n - b| \geq \delta\}.$$

Следовательно,

$$\begin{aligned} P\{|f(\xi_n, \eta_n) - f(a, b)| > \varepsilon\} &\leq \\ &\leq P\{|\xi_n - a| \geq \delta\} + P\{|\eta_n - b| \geq \delta\}. \end{aligned}$$

Но в силу условия леммы каждое слагаемое правой части последнего выражения стремится к нулю при $n \rightarrow \infty$. Лемма доказана.

Теорема 10.1. При $n \rightarrow \infty$ имеют место следующие соотношения (где сходимость понимается в смысле сходимости по вероятности):

$$1) F_n(x) \rightarrow F(x) \text{ при любом } x, -\infty < x < \infty.$$

$$2) \alpha_k = \frac{1}{n} \sum_{i=1}^n x_i^k \rightarrow \alpha_k = \int_{-\infty}^{\infty} x^k dF(x),$$

если $\alpha_{2k} < \infty$, в частности, если $\alpha_2 < \infty$,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \rightarrow a = \int_{-\infty}^{\infty} x dF(x);$$

$$3) S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \rightarrow \sigma^2 = \int_{-\infty}^{\infty} (x - a)^2 dF(x).$$

Доказательство.

1) Фиксируем $x, -\infty < x < \infty$ и введем функцию

$$I_x(y) = \begin{cases} 1, & \text{если } y < x, \\ 0, & \text{если } y \geq x. \end{cases}$$

Очевидно, что

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_x(x_i) \rightarrow M I_x(x_i),$$

так как $I_x(x_i)$ являются независимыми случайными величинами (как функции от независимых случайных величин x_1, \dots, x_n) и имеют конечную дисперсию. Далее,

$$M I_x(x_i) = 1 \cdot P\{I_x(x_i) = 1\} = P\{x_i < x\} = F(x),$$

так что утверждение 1) доказано.

2) Если $a_{2h} < \infty$, то существует конечная дисперсия $D(x_i^k)$, так что 2) вытекает из закона больших чисел.

3) Применим лемму 10.1 к выражению

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2,$$

полагая

$$\xi_n = \frac{1}{n} \sum_{i=1}^n x_{i1}^2 \rightarrow a_2 = M x_{i1}^2 = \int_{-\infty}^{\infty} x^2 dF(x),$$

$$\eta_n = \bar{x} \rightarrow M x_i = a = \int_{-\infty}^{\infty} x dF(x)$$

и считая, что $f(\xi_n, \eta_n) = \xi_n^2 - \eta_n^2$. Утверждение доказано.

10.2. Основные понятия теории оценок. Оценкой вообще называется функция $\gamma = \gamma(x_1, \dots, x_n)$, зависящая только от наблюдавшихся в опыте значений x_1, \dots, x_n . Но нас интересуют не любые оценки, а только близкие в каком-то смысле к оцениваемому параметру c .

Определение 10.1. Оценка $\gamma = \gamma(x_1, \dots, x_n)$ параметра c называется *состоятельной* оценкой параметра c , если при $n \rightarrow \infty$

$$\gamma(x_1, \dots, x_n) \rightarrow c$$

в смысле сходимости по вероятности.

Теорему 10.1 можно сформулировать следующим образом: *выборочные характеристики являются состоятельными оценками соответствующих теоретических характеристик.*

Требование состоятельности предъявляется ко всем практическим используемым оценкам. Часто считается полезным также следующее свойство несмещенности.

Определение 10.2. Оценка $\gamma = \gamma(x_1, \dots, x_n)$ называется *несмещенной*, если при любом n

$$\mathbf{M} \gamma(x_1, \dots, x_n) = c.$$

Таким образом, в случае несмещенной оценки отклонение $\gamma - c$ носит несистематический характер. Впрочем, обычно вполне достаточно, если при $n \rightarrow \infty$

$$\mathbf{M} \gamma(x_1, \dots, x_n) \rightarrow c. \quad (10.2)$$

Если выполнено (10.2) и при $n \rightarrow \infty$ дисперсия

$$\mathbf{D} \gamma(x_1, \dots, x_n) \rightarrow 0, \quad (10.3)$$

то оценка γ является состоятельной оценкой параметра c (это вытекает из неравенства Чебышева). Большинство практически применимых оценок обладают свойствами (10.2) и (10.3).

Заметим, что

$$\mathbf{M} \bar{x} = \frac{1}{n} \sum_{i=1}^n \mathbf{M} x_i = a = \int_{-\infty}^{\infty} x dF(x),$$

т. е. \bar{x} является несмещенной оценкой для a . Величина

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

не является несмещенной оценкой для σ^2 . В связи с этим вместо S^2 иногда применяют

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Теорема 10.2. Величина s^2 является несмещенной оценкой для σ^2 .

Доказательство. Положим $x_i = y_i + a$, где $a = \mathbf{M} x_i$. Тогда, очевидно,

$$\mathbf{M} y_i = 0, \quad \mathbf{D} y_i = \mathbf{D} x_i = \sigma^2,$$

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

Заметим, что в силу независимости при $i \neq j$

$$\mathbf{M} (y_i y_j) = \mathbf{M} y_i \mathbf{M} y_j = 0.$$

Имеем

$$\mathbf{M} \sum_{i=1}^n (y_i - \bar{y})^2 = \mathbf{M} \left\{ \sum_{i=1}^n y_i^2 - n (\bar{y})^2 \right\}.$$

Вычислим

$$M(\bar{y})^2 = \frac{1}{n^2} M \left\{ \sum_{i=1}^n y_i^2 + 2 \sum_{i < j} y_i y_j \right\} = \frac{M y_i^2}{n} = \frac{\sigma^2}{n}.$$

Следовательно,

$$Ms^2 = \frac{1}{n-1} M \left\{ \sum_{i=1}^n (y_i - \bar{y})^2 \right\} = \frac{nM y_i^2 - nM(\bar{y})^2}{n-1} = \sigma^2.$$

Теорема доказана.

З а м е ч а н и е. Очевидно, что при $n \rightarrow \infty$ величина $s^2 = \frac{n}{n-1} S^2$ стремится к тому же пределу σ^2 , что и S^2 , т. е. s^2 является несмещенной состоятельной оценкой для σ^2 . Однако $s = \sqrt{s^2}$ не будет несмещенной оценкой для σ , так что использование s вместо S связано не столько с несмещенностью s^2 , сколько с тем, что в случае нормального теоретического закона распределения именно s входит во все важнейшие формулы, как мы это вскоре увидим.

Важнейшим понятием теории оценок является понятие доверительного интервала, с которым мы уже встречались при оценке вероятности успеха в испытаниях Бернулли.

О п р е д е л е н и е 10.3. Пусть $\gamma_1(x_1, \dots, x_n)$, $\gamma_2(x_1, \dots, x_n)$ — такие функции от выборочных значений x_1, \dots, x_n , что $\gamma_1(x_1, \dots, x_n) \leq \gamma_2(x_1, \dots, x_n)$ при любых возможных x_1, \dots, x_n и для параметра c имеем

$$P \{ \gamma_1(x_1, \dots, x_n) \leq c \leq \gamma_2(x_1, \dots, x_n) \} \geq 1 - \alpha,$$

где α — некоторое (небольшое) число. В таком случае интервал $[\gamma_1, \gamma_2]$ называют *доверительным интервалом* для параметра c с коэффициентом доверия (надежности) $1 - \alpha$.

Практическая схема применения введенных понятий обычно такова. С самого начала имеется выборка x_1, \dots, x_n . Из тех или иных теоретических или экспериментальных соображений определяют, в каком параметрическом семействе нужно искать теоретический закон распределения $F(x)$. Затем с помощью выборочных характеристик приближенно определяют параметры $F(x)$, после подстановки которых получаем некоторый новый закон распределения $\tilde{F}(x)$. Однако, если мы хотим делать какие-то выводы из найденного закона, нам важно знать, с какой точностью он найден, т. е. насколько сильно может отличаться $\tilde{F}(x)$ от $F(x)$. Например, если мы хотим знать число x_β такое, что $F(x_\beta) = \beta$ (где β — некоторое небольшое число), то мы можем в качестве приближения к x_β взять число \tilde{x}_β такое, что $\tilde{F}(\tilde{x}_\beta) = \beta$. Однако нам важно знать, как велика может быть разница между x_β и \tilde{x}_β . С этой целью нуж-

но вычислить доверительные интервалы, показывающие примерный порядок ошибки определения каждого из параметров, и установить, насколько сильно может колебаться \tilde{x}_β при колебании параметров в пределах своих доверительных интервалов. Точный доверительный интервал для x_β установить таким образом не всегда возможно, поскольку не всегда известно совместное распределение оценок параметров. Тем не менее, порядок возможной ошибки от замены x_β на \tilde{x}_β нужно всегда иметь в виду.

10.3. Доверительные интервалы для моментов. Оценкой момента $a_k = \int_{-\infty}^{\infty} x^k dF(x)$ является эмпирический момент $\alpha_k = \frac{1}{n} \sum_{i=1}^n x_i^k$.

Применяя центральную предельную теорему, имеем

$$\begin{aligned} P\{|\alpha_k - a_k| \geq \varepsilon\} &= P\left\{\left|\frac{\sum_{i=1}^n (x_i^k - M x_i^k)}{n}\right| \geq \varepsilon\right\} = \\ &= P\left\{\left|\frac{\sum_{i=1}^n (x_i^k - M x_i^k)}{\sqrt{nD(x_i^k)}}\right| \geq \frac{\varepsilon \sqrt{n}}{\sqrt{D(x_i^k)}}\right\} \approx 2\Phi\left(-\frac{\varepsilon \sqrt{n}}{\sqrt{D(x_i^k)}}\right), \end{aligned}$$

что позволяет сразу построить доверительный интервал для a_k , если только известна дисперсия

$$D(x_i^k) = M x_i^{2k} - (M x_i^k)^2 = a_{2k} - (a_k)^2.$$

Для нахождения $D(x_i^k)$ обычно нельзя предложить ничего лучшего, чем заменить a_{2k} и a_k их оценками α_{2k} и α_k . Учитывая невысокие требования к точности построения доверительного интервала, можно обычно ограничиться этим приемом при больших n (порядка нескольких десятков). Таким образом, при оценке точности определения a_k по выборке приходится привлекать a_{2k} . При очень большом a_{2k} замена a_{2k} на α_{2k} для применения центральной предельной теоремы может привести к ошибке даже при нескольких десятках наблюдений. Мы вскоре увидим, что при выборках из нормального закона существует точный способ построения доверительных интервалов. Пока же отметим специально случай $k=1$:

$$P\{|\bar{x} - a| \geq \varepsilon\} \approx 2\Phi\left(-\frac{\varepsilon \sqrt{n}}{\sigma}\right),$$

где вместо σ часто бывает возможным подставить любую из оценок

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad S = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

10.4. Метод Монте-Карло. Мы отмечали в § 1, что самым надежным случайным экспериментом является бросание монеты. Положим $\xi=1$, если в i -том бросании выпал герб и $\xi_i=0$, если в i -том бросании выпала цифра. Пусть N — некоторое большое число (порядка 40—50). Рассмотрим случайную величину

$$\zeta = \sum_{i=1}^N \frac{\xi_i}{2^i},$$

т. е. двоичную дробь, знаками которой являются числа $\xi_1, \xi_2, \dots, \xi_N$. Очевидно, ζ имеет 2^N различных равновероятных значений, лежащих на отрезке $[0, 1]$. Практически распределение ζ неотлично от равномерного распределения на отрезке $[0, 1]$, т. е. распределения, плотность которого равна 1 на отрезке $[0, 1]$ и нулю вне этого отрезка. Пусть теперь $F(x)$ — любая непрерывная функция распределения. Положим

$$\eta = F^{-1}(\zeta).$$

В таком случае для любого x имеем

$$F_\eta(x) = P\{\eta < x\} = P\{F^{-1}(\zeta) < x\} = P\{\zeta < F(x)\} = F(x),$$

поскольку при $0 \leq y \leq 1$

$$P\{\zeta < y\} = y.$$

Таким образом, с помощью бросаний монеты ξ_1, \dots, ξ_N и аналитических операций можно образовать случайную величину η , имеющую наперед заданное распределение $F(x)$. Повторяя бросание монеты Nn раз, можно образовать выборку η_1, \dots, η_n из закона распределения $F(x)$. Беря функции вида $f_1(\eta_1, \dots, \eta_n), \dots, f_k(\eta_1, \dots, \eta_n)$, можно образовать более сложные многомерные распределения. Нет сомнения в том, что таких распределений достаточно для любых практических целей.

Описанный здесь процесс называется методом моделирования, или методом Монте-Карло. Сложность, однако, состоит в том, что при решении реальных задач необходимо бросать монету так быстро, чтобы это было сравнимо со скоростью работы ЭВМ. К сожалению, такого способа не существует, и современные датчики случайных чисел выдают в сущности результаты арифметических операций над вполне определенными неслучайными числами. В силу каких-то мало понятных свойств чисел получаемые таким образом псевдослучайные числа могут применяться для практических целей так же, как и настоящие случайные числа. Например, если разлагать число π в десятичную дробь, то с помощью статистических критериев невозможно заметить в этой последовательности цифр никаких отклонений от случайной последовательности цифр $0, \dots, 9$. Было также давно замечено, что если взять последние цифры многозначных логарифмов натуральных чисел, то получается неплохая

модель равномерного распределения на отрезке $[0, 1]$. Все эти факты не имеют пока полного научного объяснения, и где проходит граница применимости псевдослучайных чисел, не ясно. Мы рассмотрим одно применение теории статистических оценок, предполагая, что при моделировании мы имеем дело с настоящими случайными числами, хотя надо со всей серьезностью предупредить, что при использовании псевдослучайных чисел возможны существенные отклонения от излагаемых результатов.

Пусть требуется вычислить интеграл

$$I = \int_0^1 \int_0^1 \dots \int_0^1 f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

от некоторой функции $f(x) = f(x_1, \dots, x_n)$, заданной на n -мерном единичном кубе K . Предположим, что $\zeta_1, \zeta_2, \dots, \zeta_N$ — случайные величины, имеющие каждая равномерное распределение на K . Очевидно, что

$$f(\zeta_1), f(\zeta_2), \dots, f(\zeta_N) \quad (10.4)$$

образуют выборку. Имеем

$$M f(\zeta_i) = \int_0^1 \dots \int_0^1 f(x_1, \dots, x_n) dx_1 \dots dx_n = I.$$

Поэтому искомое значение интеграла I есть теоретическое среднее выборки (10.4). Имеем, согласно п. 10.3,

$$P\left\{\left|\frac{1}{N} \sum_{i=1}^N f(\zeta_i) - I\right| > \varepsilon\right\} \approx 2\Phi\left(-\frac{\varepsilon\sqrt{N}}{\sigma}\right),$$

где $\sigma^2 = Df(\zeta_i)$.

Следовательно, полагая, например,

$$2\Phi\left(-\frac{\varepsilon\sqrt{N}}{\sigma}\right) = 0,997,$$

откуда $\frac{\varepsilon\sqrt{N}}{\sigma} = 3$, или $\varepsilon = \frac{3\sigma}{\sqrt{N}}$, мы видим, что точность ε деления I по выборке имеет порядок $\frac{1}{\sqrt{N}}$. Заметим, что для этой оценки важна не гладкость функции $f(x_1, \dots, x_n)$, а лишь дисперсия

$$\sigma^2 = \int_0^1 \dots \int_0^1 (f(x_1, \dots, x_n) - I)^2 dx_1 \dots dx_n.$$

Таким образом, за счет N моделирований n -мерных случайных векторов ζ_1, \dots, ζ_N и N вычислений значений $f(\zeta_i)$, $i=1, \dots, N$ дости-

гается точность $\frac{3\sigma}{\sqrt{N}}$. Сравним этот результат с классическим способом вычисления интегралов с помощью аппроксимации суммами.

В случае $n=1$ метод Симпсона по N точкам дает (для гладких функций) точность порядка $\frac{1}{N^4}$. Однако если в случае $n>1$ перемennых применять метод Симпсона по каждому переменному хотя бы в 10 точках, то потребуется 10^n вычислений значений функции $f(x_1, \dots, x_n)$. При таком способе вычисление интеграла I при возрастании n сначала сделается невозможным из-за накопления ошибок, а затем из-за недостаточного быстродействия машины. Метод же Монте-Карло при возрастании размерности требует nN моделирований и N вычислений значений функции $f(\xi_i)$ вместо 10^n вычислений значений функции. Ясно, что при больших размерностях n метод Монте-Карло является единственно возможным.

При его применении желательно, конечно, иметь оценку сверху дисперсии σ^2 .

10.5. Практические приемы работы с выборками. Одним из полезнейших изобретений для статистической обработки выборок является вероятностная бумага. При изменении шкалы отсчета $x \rightarrow \sigma x + a$ закон распределения $F(x)$ преобразуется в закон распределения $F\left(\frac{x-a}{\sigma}\right)$. Поэтому имеющее физический смысл семейство законов распределения должно как минимум содержать параметры a и σ .

Уговоримся рисовать график функции $F(x)$ в таком масштабе по оси ординат, чтобы этот график был прямой линией. Тогда все законы $F\left(\frac{x-a}{\sigma}\right)$, отличающиеся от $F(x)$ линейной заменой аргумента, также будут изображаться прямыми линиями. Изготовленная таким способом бумага называется вероятностной. Способ оцифровки оси ординат для ее изготовления ясен из рис. 10.1. Имея таблицу функции $F(x)$, вероятностную бумагу легко приготовить на любом листе миллиметровки, а на худой конец — на листе из тетради в клетку. В случае нормального закона $F(x)$ бумага называется нормальной.

На практике на нормальной бумаге изображается эмпирическая функция распределения $F_n(x)$. Если выборка имеет теоретическое нормальное распределение, то ступенчатая функция $F_n(x)$

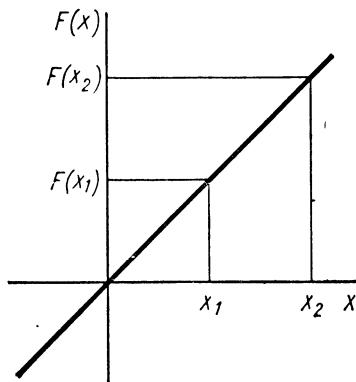


Рис. 10.1

должна быть близка к прямой. Используя прозрачную линейку, нетрудно провести соответствующую прямую. При некотором навыке бывает достаточно ограничиться этим способом проверки нормальности распределения. Явно заметное отклонение от прямой может позволить обнаружить не только тот случай, когда теоретическое распределение не будет нормальным, но и случай, когда наблюдения вовсе не образуют выборки (т. е. нарушается соответствующая статистическая однородность). По-видимому, последовательность приемов должна быть такова: проверка нормальности с помощью нормальной бумаги; затем, при плохом согласии с нормальным законом — проверка статистической однородности и лишь затем (в случае наличия однородности) подбор других семейств распределений.

В случае согласия с нормальным законом можно считать, что проведенная на глаз прямая близка к теоретическому нормальному закону $\Phi\left(\frac{x-a}{\sigma}\right)$. С помощью графика этой прямой легко без вычислений найти a и σ . Действительно, значение x_0 такое, что $\Phi\left(\frac{x_0-a}{\sigma}\right) = 0,5$, соответствует $\frac{x_0-a}{\sigma} = 0$, т. е. $x_0 = a$, так как $\Phi(0) = 0,5$. Ввиду того что $\Phi(1) = 0,84$, $\Phi(-1) = 0,16$, легко находимые графически решения x_{-1} и x_{+1} уравнений

$$\Phi\left(\frac{x_{-1}-x_0}{\sigma}\right) = 0,16; \quad \Phi\left(\frac{x_{+1}-x_0}{\sigma}\right) = 0,84$$

обладают свойствами:

$$x_{-1} - x_0 = -\sigma, \quad x_{+1} - x_0 = +\sigma,$$

откуда $\sigma = \frac{1}{2}(x_{+1} - x_{-1})$. Этим способом оценок обычно ограничиваются на практике. Он может служить также для контроля вычислений \bar{x} и s .

Следует заметить, что при оценке возможной разницы между x_0 и a может встретиться следующая ошибка. Оценивают на глаз, с какой точностью может быть проведена прямая линия, приближающая $F_n(x)$, и каковы, следовательно, возможные колебания точки x_0 (точки пересечения этой прямой с линией $y=0,5$). Этот способ обманчив, так как уже в самой эмпирической функции распределения $F_n(x)$ содержится некоторая случайность. Поэтому следует пользоваться методом доверительных интервалов.

В данной книге лишь намечены практические приемы работы. Исчерпывающее изложение практической стороны дела вместе с примерами использования нормальной бумаги можно найти в книге А. Хальда [24]. При большом числе наблюдений вместо эмпирической функции распределения можно построить так называемую гистограмму, которая по виду напоминает плотность распределения. Короче всего определить гистограмму так: на обыч-

ной (невероятностной) бумаге строится эмпирическая функция распределения, затем она сглаживается ломаной линией, а гистограмма является графиком производной последней ломаной линии. По поводу работы с гистограммами читатель также отсылается к книге А. Хальда [24].

Не так просто, как кажется на первый взгляд, найти в доступной литературе данные, на которых можно было бы попрактиковаться в статистической обработке. От плохих наблюдений нельзя ожидать нормального закона. Даже при хороших наблюдениях экспериментаторы часто выписывают данные таким образом, что фактически лишь одна цифра является значащей, поскольку следующие цифры сильно колеблются и в них нет веры. Обрабатывать грубо округленные данные с помощью нормальной бумаги бессмысленно. Нужно, чтобы наблюдения записывались хотя бы с двумя значащими цифрами.

Примером данных, вполне оправдывающих гипотезу нормальности, являются данные Милликена по измерению заряда электрона (задачник Л. Д. Мешалкина, задача 456). На этих данных можно практически освоить приемы, рассматриваемые в этом и следующих параграфах. Например, построив эмпирическую функцию распределения на нормальной бумаге, читатель увидит, что в средней части эмпирическая функция распределения прекрасно укладывается на прямую, а в «хвостах», т. е. в областях значений, близких к 0 или 1, укладывается хуже. Последнее не всегда свидетельствует о плохом согласии с нормальным законом и является обычным.

§ 11

ОБЩАЯ ЛИНЕЙНАЯ МОДЕЛЬ, СВЯЗАННАЯ С НОРМАЛЬНЫМ РАСПРЕДЕЛЕНИЕМ ОШИБОК НАБЛЮДЕНИЙ

В этом параграфе мы изложим основные приемы обработки наблюдений в том случае, когда ошибки наблюдений считаются нормально распределенными случайными величинами с нулевым математическим ожиданием. Мы рассмотрим как случай выборки (т. е. случай наблюдений одной и той же величины), так и случай наблюдений разных, но связанных между собой величин. Начнем с математического введения.

11.1. Основные распределения, связанные с нормальным законом. Свойства проекций нормального вектора.

Определение 11.1. *Распределением Пирсона* (или распределением χ^2) с n степенями свободы называется распределение случайной величины

$$\chi_n^2 = \xi_1^2 + \xi_2^2 + \dots + \xi_n^2,$$

где ξ_1, \dots, ξ_n — независимые случайные величины, имеющие каждая нормальное распределение $N(0, 1)$.

Определение 11.2. *Распределением Стьюдента* (или *t-распределением*) с n степенями свободы называется распределение случайной величины

$$t_n = \frac{\xi}{\sqrt{\frac{1}{n} \chi_n^2}},$$

где ξ — нормальная случайная величина с параметрами $(0, 1)$, χ_n^2 — случайная величина, имеющая распределение χ^2 с n степенями свободы и не зависящая от ξ .

Определение 11.3. *Распределением Фишера* (или *F-распределением*) с (m, n) степенями свободы называется распределение случайной величины

$$F_{m,n} = \frac{\frac{1}{m} \chi_m^2}{\frac{1}{n} \chi_n^2},$$

где в числителе и знаменателе стоят независимые случайные величины, имеющие распределения χ^2 с m и n степенями свободы (соответственно).

При обработке наблюдений постоянно приходится иметь дело с таблицами только что введенных распределений. Наоборот, явные выражения для их плотностей почти не используются на практике и поэтому мы их не приводим (читатель может найти их в любом учебнике математической статистики). Остановимся коротко лишь на приближенном выражении этих распределений при большом числе степеней свободы.

В силу центральной предельной теоремы, при больших n распределение χ_n^2 является приблизительно нормальным. Нетрудно подсчитать, что $M \chi_n^2 = n$, $D \chi_n^2 = 2n$, так что самую простую аппроксимацию распределения χ_n^2 получаем, если считать, что распределение величины

$$\frac{\chi_n^2 - n}{\sqrt{2n}}$$

приближенно равно $N(0, 1)$. Согласие с нормальным распределением улучшится, если заменить $\sqrt{2n}$ на $\sqrt{2n-1}$, но мы не будем вдаваться здесь в подробности точного определения функции распределения χ_n^2 (см., например, таблицы Л. И. Большева и Н. В. Смирнова [4]).

В силу закона больших чисел $\frac{1}{n} \chi_n^2 \rightarrow 1$ по вероятности, так что естественно ожидать, что при больших n распределение величины t_n практически совпадает с распределением числителя ξ , т. е.

с распределением $N(0, 1)$. Обычно считают, что при $n > 30$ практически можно пользоваться вместо распределения t_n распределением $N(0, 1)$, но это, разумеется, неверно для вероятностей, очень близких к 0 или 1. Правда, обычно нельзя всерьез утверждать, что в области таких вероятностей совершенно точно действует и распределение Стьюдента.

Гораздо сложнее обстоит дело с табулированием распределения Фишера (по той простой причине, что вероятность $P\{F_{m,n} < x\}$ зависит от трех параметров m , n и x , а следовательно, нужна таблица с тремя входами). В доступной литературе имеются лишь таблицы процентных точек F -распределения, т. е. таблицы точек $x(m, n, \alpha)$, таких, что

$$P\{F_{m,n} < x(m, n, \alpha)\} = \alpha$$

для нескольких значений α . Обычно этого бывает достаточно.

Применения указанных распределений основаны на следующей лемме и следствиях из нее.

Лемма 11.1. Пусть $\xi = (\xi_1, \dots, \xi_n)$ — случайный вектор, компоненты которого ξ_1, \dots, ξ_n независимы (в совокупности) и имеют каждая распределение $N(0, \sigma)$. Если U — ортогональное преобразование евклидова пространства, то распределение вектора

$$\eta = U\xi$$

совпадает с распределением вектора ξ .

Доказательство. Имеем для плотности распределения

$$\begin{aligned} p_\xi(x) &= p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \\ &= \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{x_i^2}{2\sigma^2}} = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n e^{-\frac{(x, x)}{2\sigma^2}}. \end{aligned}$$

Далее

$$\begin{aligned} p_\eta(x) &= p_{U\xi}(x) = p_\xi(U^{-1}x) = \\ &= \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} (U^{-1}x, U^{-1}x) \right\} = \\ &= \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} (x, x) \right\} = p_\xi(x) \end{aligned}$$

в силу ортогональности U (якобиан преобразования равен единице и $(U^{-1}x, U^{-1}x) = (x, x)$). Лемма доказана.

Выведем из этой леммы ряд (весьма важных) следствий. Эти следствия будут относиться к вектору $\xi + b$, отличающемуся от ξ добавлением неслучайного вектора b . Буквой L с различными индексами мы будем обозначать линейные подпространства n -мерно-го евклидова пространства, а знаком \perp — ортогональность.

Следствие 1. Если $L_1 \perp L_2$, то проекции

$$\text{proj}_{L_1}(\xi + b) \text{ и } \text{proj}_{L_2}(\xi + b)$$

являются независимыми случайными векторами.

Доказательство. Поскольку

$$\text{proj}_{L_i}(\xi + b) = \text{proj}_{L_i}\xi + \text{proj}_{L_i}b, \quad i = 1, 2,$$

и b — неслучайный вектор, то достаточно доказать, что независимы $\text{proj}_{L_1}\xi$ и $\text{proj}_{L_2}\xi$. Выберем в пространстве ортонормированный базис $f_1, f_2, \dots, f_{k_1}, f_{k_1+1}, \dots, f_{k_1+k_2}, f_{k_1+k_2+1}, \dots, f_n$, в котором векторы f_1, \dots, f_{k_1} являются базисом в L_1 (размерность $\dim L_1 = k_1$), а векторы $f_{k_1+1}, \dots, f_{k_1+k_2}$ являются базисом в L_2 ($\dim L_2 = k_2$). В таком случае

$$\text{proj}_{L_1}\xi = \sum_{s=1}^{k_1} (\xi, f_s) f_s,$$

$$\text{proj}_{L_2}\xi = \sum_{s=k_1+1}^{k_1+k_2} (\xi, f_s) f_s.$$

Но, в силу леммы 11.1, величины

$$(\xi, f_1), (\xi, f_2), \dots, (\xi, f_n)$$

имеют такое же распределение, как и величины $\xi_1, \xi_2, \dots, \xi_n$, в частности независимы. Поэтому $\text{proj}_{L_1}\xi$ и $\text{proj}_{L_2}\xi$ независимы как функции от независимых случайных величин.

Следствие 2. Пусть $b \perp L$. Тогда $\|\text{proj}_L(\xi + b)\|^2 = \|\text{proj}_L\xi\|^2$ имеет распределение $\sigma^2 \chi^2_{\dim L}$, где $\sigma^2 = D\xi_i$, $\chi^2_{\dim L}$ — случайная величина, имеющая распределение χ^2 с числом степеней свободы, равным размерности $\dim L$.

Доказательство. Положим $L = L_1$, чтобы использовать обозначения доказательства предыдущего следствия. Тогда

$$\|\text{proj}_{L_1}\xi\|^2 = \sum_{s=1}^{k_1} (\xi, f_s)^2,$$

что имеет такое же распределение, как

$$\sum_{s=1}^{k_1} \xi_s^2 = \sigma^2 \sum_{s=1}^{k_1} \left(\frac{\xi_s}{\sigma} \right)^2 = \sigma^2 \chi_{k_1}^2,$$

что и требовалось доказать.

Следствие 3. Если $b \perp L_1$, $b \perp L_2$, причем $L_1 \perp L_2$, то отношение

$$\frac{\frac{1}{\dim L_1} \|\text{proj}_{L_1}(\xi + b)\|^2}{\frac{1}{\dim L_2} \|\text{proj}_{L_2}(\xi + b)\|^2} \quad (11.1)$$

имеет распределение Фишера $F_{\dim L_1, \dim L_2}$.

Доказательство. В силу следствия 1 числитель и знаменатель (11.1) независимы, а в силу следствия 2 имеют, соответственно, распределения $\frac{1}{\dim L_1} \sigma^2 \chi^2_{\dim L_1}$ и $\frac{1}{\dim L_2} \sigma^2 \chi^2_{\dim L_2}$. При делении числителя на знаменатель общий множитель σ^2 сокращается и получается утверждение следствия.

11.2. Общая линейная модель и метод наименьших квадратов.

Пусть мы имеем совокупность наблюдений $\eta_1, \eta_2, \dots, \eta_n$, вообще говоря, различных величин a_1, a_2, \dots, a_n . Каждое наблюдение $\eta_i = a_i + \delta_i$ складывается из значения a_i наблюдаемой величины и ошибки наблюдения δ_i . В математической статистике не остается ничего другого, как предположить, что наблюдения η_i не имеют систематической ошибки, т. е. $M\delta_i = 0$. Кроме того, обычно предполагается, что все дисперсии $D\delta_i = \sigma^2$ одинаковы (но нам неизвестны) и распределения δ_i нормальны: $\delta_i \sim N(0, \sigma)$. Несколько более общий случай, когда дисперсии $D\delta_i$ неодинаковы, но известны их отношения, т. е.

$$D\delta_i = \sigma^2 w_i,$$

легко сводится к предыдущему. Действительно, положим

$$\xi_i = \frac{\eta_i}{\sqrt{w_i}} = \frac{a_i}{\sqrt{w_i}} + \frac{\delta_i}{\sqrt{w_i}}.$$

Тогда дисперсии $D\xi_i$ одинаковы, а роль чисел a_1, \dots, a_n играют числа $\frac{a_i}{\sqrt{w_i}}$. Заметим, что величины

$$\frac{1}{w_1}, \frac{1}{w_2}, \dots, \frac{1}{w_n},$$

обратные дисперсиям наблюдений, называются весами наблюдений. Веса предполагаются заранее известными.

Как может случиться, что дисперсии $\sigma^2 w_i$ отдельных наблюдений неизвестны, но веса w_i известны? Один из возможных вариантов состоит в следующем. Исследователь хочет измерить величины a_1, \dots, a_n , но результаты единичных измерений этих величин столь плохи, что он даже не записывает отдельных наблюдений, а берет лишь среднее

$$\eta_i = \frac{z_{i1} + z_{i2} + \dots + z_{im_i}}{m_i}$$

из m_i наблюдений величины a_i . В силу условий эксперимента числа m_1, \dots, m_n вполне могут быть различными. Если при этом желают применить статистическую обработку, т. е. допустить соответствующую статистическую однородность, то естественно считать, что дисперсии $D\eta_i$ обратно пропорциональны числам m_i . Таким образом, числа наблюдений m_1, m_2, \dots, m_n оказываются весами наблюдений.

В дальнейшем мы будем считать веса одинаковыми.

Итак, наша модель имеет вид

$$\eta_i = a_i + \delta_i, \quad \delta_i \sim N(0, \sigma),$$

где σ неизвестно и все $\delta_1, \dots, \delta_n$ считаются независимыми случайными величинами.

Если числа a_1, a_2, \dots, a_n совершенно произвольны, то наблюдения $\eta_1, \eta_2, \dots, \eta_n$ бессмысленно обрабатывать совместно. На другом языке бессмысленность совместной обработки можно выразить так: ничто не мешает предположить, что $\sigma=0$, т. е. ошибок наблюдений нет, так что $\eta_i=a_i$. Наблюдения имеет смысл обрабатывать совместно только при наличии какой-то дополнительной информации о наблюдаемых величинах a_1, a_2, \dots, a_n . Чрезвычайно удобным и общим способом математически строгого («формального») выражения этой информации является следующий: предполагается, что вектор $a=(a_1, \dots, a_n)$ лежит в некотором известном линейном подмногообразии M (напомним, что линейным подмногообразием M называется множество вида $M=L+c$, где L — линейное подпространство и $c=(c_1, \dots, c_n)$ — фиксированный вектор). Переходя от величин η_1, \dots, η_n и a_1, \dots, a_n к величинам $\eta_1-c_1, \dots, \eta_n-c_n$ и a_1-c_1, \dots, a_n-c_n , мы всегда можем без потери общности считать, что $c=0$, т. е. $M=L$ есть линейное подпространство. Итак, наши предположения имеют следующий вид:

$$\eta = a + \delta, \quad a \in L, \quad (11.2)$$

где $\eta=(\eta_1, \dots, \eta_n)$ — вектор наблюдений, $a=(a_1, \dots, a_n)$ — вектор, который мы хотим определить из наблюдений, $\delta=(\delta_1, \dots, \delta_n)$ — вектор ошибок. При этом $\delta_1, \dots, \delta_n$ независимы и

$$\delta_i \sim N(0, \sigma). \quad (11.3)$$

Для определения вектора a Гаусс предложил следующий метод максимального правдоподобия. Рассмотрим плотность распределения

$$\begin{aligned} p_\eta(x) &= p_{\eta_1, \dots, \eta_n}(x_1, \dots, x_n) = \\ &= \prod_{i=1}^n \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(x_i - a_i)^2}{2\sigma^2} \right\} = \\ &= \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n \exp \left\{ -\frac{(x - a, x - a)}{2\sigma^2} \right\}. \end{aligned}$$

Если в выражение $p_{\eta}^{-}(x)$ вместо переменной интегрирования x подставить наблюдаемый вектор η , то функция от a, η

$$p_{\eta}(a, \eta) = \left(\frac{1}{\sigma \sqrt{2\pi}} \right)^n \exp \left\{ - \frac{(\eta - a, \eta - a)}{2\sigma^2} \right\}$$

называется *функцией правдоподобия*. В качестве оценки $\hat{a} = \hat{a}(\eta)$ неизвестного значения a следует взять ту точку, в которой (при заданном η) достигается максимум функции правдоподобия $p_{\eta}(a, \eta)$:

$$p_{\eta}(\hat{a}(\eta), \eta) = \max_{a \in L} p_{\eta}(a, \eta),$$

причем максимум берется по всем $a \in L$ (поскольку заранее известно, что $a \in L$).

Очевидно, что $\hat{a}(\eta)$ обладает тем свойством, что

$$(\eta - \hat{a}(\eta), \eta - \hat{a}(\eta)) = \|\eta - \hat{a}(\eta)\|^2 = \min_{a \in L} \|\eta - a\|^2.$$

Другими словами,

$$\hat{a}(\eta) = \text{proj}_L \eta \quad (11.4)$$

является оценкой максимального правдоподобия для неизвестного вектора a .

Поскольку это решение получено путем минимизации $\|\eta - a\|^2$, то в рассматриваемом случае говорят также о методе наименьших квадратов.

Кроме оценки для a , желательно бывает получить оценку для дисперсии σ^2 отдельного наблюдения. С этой целью заметим, что случайная величина

$$\Delta^2 = \|\eta - \text{proj}_L \eta\|,$$

называемая «кажущейся ошибкой» (происхождение этого названия будет ясно несколько позже), совпадает с

$$\|\text{proj}_{L'} \eta\|^2,$$

где L' есть ортогональное дополнение к L . Следовательно, в силу того, что $a \perp L'$, Δ^2 имеет распределение

$$\sigma^2 \chi_{\dim L'}^2 = \sigma^2 \chi_{n - \dim L}^2.$$

Рассмотрим величину

$$s^2 = \frac{1}{n - \dim L} \Delta^2.$$

Очевидно, что

$$Ms^2 = \frac{1}{n - \dim L} \sigma^2 M\chi_{n - \dim L}^2 = \sigma^2.$$

Следовательно, s^2 является несмещенной оценкой для σ^2 . Отклонение s^2 от σ^2 определяется отклонением случайной величины

$$\frac{1}{n - \dim L} \chi_{n - \dim L}^2$$

от единицы. Следовательно, при большом числе $n - \dim L$ степеней свободы это отклонение будет малым. Например, если число наблюдений $n \rightarrow \infty$, а размерность $\dim L$ подпространства L (в котором лежит вектор математических ожиданий наблюдений) остается ограниченной, то s^2 является состоятельной оценкой для σ^2 .

З а м е ч а н и е. Оценка $\hat{a}(\eta)$ для a связана с проекцией вектора η на L ; оценка s^2 для σ^2 связана с проекцией вектора η на $L' \perp L$. Следовательно, $\hat{a}(\eta)$ и s^2 являются независимыми случайными величинами (следствие 1 леммы 11.1). При этом

$$\hat{a}(\eta) = \text{proj}_L \eta = \text{proj}_L (a + \delta) = a + \text{proj}_L \delta,$$

так что квадрат нормы

$$\|a - \hat{a}(\eta)\|^2 = \|\text{proj}_L \delta\|^2$$

имеет распределение $\sigma^2 \chi_{\dim L}^2$. Распределение отношения

$$\frac{\frac{1}{\dim L} \|a - \hat{a}(\eta)\|^2}{s^2}$$

есть распределение Фишера $F_{\dim L, \dim L'}$.

Мы рассмотрели общую задачу об оценке параметров в случае линейной модели. Значение этой задачи можно представить себе только после рассмотрения важнейших примеров применения этой общей задачи.

11.3. Нормальная выборка (теория ошибок). Важнейшим частным случаем общей теории является случай нормальной выборки, когда измеряется все время одна и та же величина, т. е. $a_1 = a_2 = \dots = a_n = a$. Очевидно, что в этом случае

$$L = \{(a_1, a_2, \dots, a_n) = ae\sqrt{n}, \quad -\infty < a < \infty\},$$

где через e обозначен нормированный вектор

$$e = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right).$$

Оценкой для вектора $ae\sqrt{n}$ будет

$$\text{proj}_L \eta = (\eta, e) e = \frac{\sum_{i=1}^n \eta_i}{\sqrt{n}} \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right) = (\bar{\eta}, \dots, \bar{\eta}),$$

где $\bar{\eta} = \frac{1}{n} \sum_{i=1}^n \eta_i$. Следовательно, оценкой для вектора $(a_1, \dots, a_n) = (a, \dots, a)$ является вектор $(\bar{\eta}, \dots, \bar{\eta})$, иначе говоря, оценкой для числа a является $\bar{\eta}$.

Заметим, что вместо обозначения $\eta = (\eta_1, \dots, \eta_n)$ обычно употребляется обозначение $x = (x_1, \dots, x_n)$ для результатов наблюдений. Тогда оценкой для a будет \bar{x} .

Оценкой для σ^2 будет (в силу того, что $\dim L = 1$)

$$s^2 = \frac{1}{n-1} \|(x_1, \dots, x_n) - (\bar{x}, \dots, \bar{x})\|^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Заметим, что случайная величина $\frac{1}{\sigma} (\bar{x} - a) \sqrt{n}$ имеет, очевидно, распределение $N(0, 1)$. Случайная величина $\frac{1}{\sigma^2} s^2$ имеет распределение $\frac{1}{n-1} \chi_{n-1}^2$. Поэтому величина

$$t_{n-1} = \frac{\frac{1}{\sigma} (\bar{x} - a) \sqrt{n}}{\sqrt{\frac{1}{\sigma^2} s^2}} = \frac{(\bar{x} - a) \sqrt{n}}{s}$$

имеет распределение Стьюдента с $(n-1)$ степенью свободы. Это важнейшее свойство нормальной выборки.

Построим, используя этот факт, доверительный интервал для среднего a . С помощью таблиц распределения Стьюдента легко найти по заданному α число t_α такое, что

$$P\{|t_{n-1}| \leq t_\alpha\} = 1 - \alpha.$$

Подставим вместо t_{n-1} величину $\frac{(\bar{x} - a) \sqrt{n}}{s}$ и решим относительно a получившееся неравенство. Находим

$$P\left\{\bar{x} - \frac{s}{\sqrt{n}} t_\alpha \leq a \leq \bar{x} + \frac{s}{\sqrt{n}} t_\alpha\right\} = 1 - \alpha. \quad (11.5)$$

Если бы мы пользовались нормальным распределением для \bar{x} , считая, что $\sigma = s$, то вместо нахождения t_α по таблицам распределения Стьюдента мы должны были бы найти его по таблицам нормального закона. Значения t_α оказались бы при этом меньше. Следовательно, применение распределения Стьюдента дает более «осторожный» (т. е. более широкий) доверительный интервал.

Далее, $\frac{(n-1)s^2}{\sigma^2}$ имеет распределение χ_{n-1}^2 . Найдем числа $x_1(\alpha)$ и $x_2(\alpha)$ такие, что

$$P\{x_1(\alpha) \leq \chi_{n-1}^2 \leq x_2(\alpha)\} = 1 - \alpha$$

(этим соотношением $x_1(\alpha)$ и $x_2(\alpha)$ определены неоднозначно; обычно их выбирают таким образом, чтобы

$$P\{\chi_{n-1}^2 \leq x_1(\alpha)\} = \frac{\alpha}{2}, \quad P\{\chi_{n-1}^2 \geq x_2(\alpha)\} = \frac{\alpha}{2}.$$

Подставляя $\frac{(n-1)s^2}{\sigma^2}$ вместо χ_{n-1}^2 и решая относительно σ^2 получившееся неравенство, имеем

$$P\left\{\frac{(n-1)s^2}{x_2(\alpha)} \leq \sigma^2 \leq \frac{(n-1)s^2}{x_1(\alpha)}\right\} = 1 - \alpha. \quad (11.6)$$

Мы получили, таким образом, доверительный интервал для дисперсии σ^2 .

Так называемая «теория ошибок» сводится в сущности к соотношениям (11.5) и (11.6).

11.4. Проверка некоторых гипотез. Гипотеза $a=a_0$.

Пусть имеется выборка x_1, x_2, \dots, x_n из нормального закона $N(a, \sigma)$ и мы желаем проверить гипотезу $a=a_0$, где a_0 — некоторое известное число. Гипотеза $H_0: a=a_0$ может иметь различные альтернативы. Наиболее часто встречаются две альтернативы:

- а) $a > a_0$, т. е. если неверно, что $a=a_0$, то $a > a_0$ (переменой знака эта альтернатива переводится в альтернативу $a < a_0$);
- б) $a \neq a_0$, т. е. если неверно, что $a=a_0$, то возможны оба случая: $a < a_0$ и $a > a_0$.

Первая альтернатива а) называется односторонней; вторая б) — двусторонней.

Поскольку оценкой для a является \bar{x} , то из общих соображений следует, что при альтернативе а) гипотезу $H_0: a=a_0$ следует отклонять, если $\bar{x} - a_0 > c_1$ (и не следует отклонять, если $\bar{x} < a_0$), а при альтернативе б) гипотезу H_0 следует отклонять, если $|\bar{x} - a_0| > c_2$, где числа c_1 и c_2 должны быть выбраны каким-то разумным способом. «Разумность» здесь означает, конечно, соблюдение заданного уровня значимости α . Рассмотрим сначала вопрос о выборе c_1 . Если гипотеза $H_0: a=a_0$ верна, то величина

$$t_{n-1} = \frac{(\bar{x} - a_0)\sqrt{n}}{s}$$

имеет распределение Стьюдента. Найдем с помощью таблиц число d_α такое, что

$$P\{t_{n-1} \geq d_\alpha\} = \alpha.$$

Тогда при верной гипотезе H_0

$$P \left\{ \bar{x} - a_0 \geq \frac{sd_\alpha}{\sqrt{n}} \right\} = \alpha.$$

Иными словами, $c_1 = \frac{sd_\alpha}{\sqrt{n}}$. Итак, выбор числа c_1 учитывает характеристику s для точности отдельного наблюдения и число наблюдений n .

В случае альтернативы б) надо, очевидно, вместо d_α взять число d'_α , определяемое из соотношения

$$P \{ |t_{n-1}| \geq d'_\alpha \} = \alpha.$$

Из определения распределения Стьюдента вытекает, что при любом z

$$P \{ |t_{n-1}| \geq z \} = 2P \{ t_{n-1} \geq z \}.$$

Следовательно, $d_{\alpha/2} = d'_\alpha$. Величина d_α называется односторонней доверительной границей, величина d'_α — двусторонней.

Гипотеза $\sigma_1 = \sigma_2$. Пусть имеются две независимые выборки x_1, x_2, \dots, x_n и y_1, y_2, \dots, y_m из распределений $N(a_1, \sigma_1)$ и $N(a_2, \sigma_2)$. Рассмотрим проверку гипотезы $\sigma_1 = \sigma_2$. Удобно записать эту гипотезу в виде $\sigma_1/\sigma_2 = 1$ и применить для ее проверки распределение Фишера. Рассмотрим отношение

$$\frac{s_x^2}{s_y^2} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}{\frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2}.$$

Числитель имеет распределение $\frac{\sigma_1^2}{n-1} \chi_{n-1}^2$, знаменатель — распределение $\frac{\sigma_2^2}{m-1} \chi_{m-1}^2$. Если гипотеза $\sigma_1 = \sigma_2$ верна, то отношение

$\frac{s_x^2}{s_y^2}$ имеет распределение Фишера $F_{n-1, m-1}$. Для распределения

Фишера имеются таблицы односторонних и двусторонних доверительных границ, которые позволяют, точно так же, как в предыдущем случае, определить критическую область.

Гипотеза $a_1 = a_2$.

Пусть, как и в предыдущем случае, имеются две выборки, и мы собираемся проверить гипотезу $a_1 = a_2$ (при альтернативе, для определенности, $a_1 \neq a_2$). Ясно, что эту гипотезу следует отклонить, если $|\bar{x} - \bar{y}| > c$, разумным образом выбирая c . Если $\sigma_1 \neq \sigma_2$, то рас-

смаатриваемая задача называется проблемой Беренса — Фишера, и для ее решения до сих пор не существует теоретически и практически безупречного подхода. Это, впрочем, не должно особенно смущать заинтересованного в практическом применении естествоиспытателя, который (при достаточном количестве наблюдений) вполне может считать, что дисперсия разности $\bar{x} - \bar{y}$ примерно равна $\frac{s_x^2}{n} + \frac{s_y^2}{m}$, и применить распределение $N(0, 1)$ для отношения

$$(\bar{x} - \bar{y}) \left(\frac{s_x^2}{n} + \frac{s_y^2}{m} \right)^{-\frac{1}{2}}.$$

При умеренном числе наблюдений можно применить более точные способы (см., например, таблицы Я. Янко [29]).

Мы же рассмотрим случай, когда известно, что $\sigma_1 = \sigma_2 = \sigma$ (обычно сначала проверяется гипотеза $\sigma_1 = \sigma_2$ и, если она не отвергается, то считают возможным применить излагаемый метод). Если это так, то при верной гипотезе $a_1 = a_2$ разность $\bar{x} - \bar{y}$ имеет нормальное распределение $N\left(0, \sigma \sqrt{\frac{1}{m} + \frac{1}{n}}\right)$. За оценку для σ можно взять следующую величину:

$$s^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}.$$

Действительно, $(n-1)s_x^2$ имеет распределение $\sigma^2 \chi_{n-1}^2$, $(m-1)s_y^2$ имеет распределение $\sigma^2 \chi_{m-1}^2$; в силу независимости двух выборок $(n+m-2)s^2$ имеет распределение $\sigma^2 \chi_{n+m-2}^2$ и $Ms^2 = \sigma^2$. Кроме того, $\bar{x} - \bar{y}$ и s^2 независимы (как функции от четырех независимых в совокупности величин \bar{x} , \bar{y} , s_x^2 , s_y^2). Поэтому статистика (статистикой называется любая функция от результатов наблюдений)

$$t_{n+m-2} = \frac{\bar{x} - \bar{y}}{s \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

имеет распределение Стьюдента с $(n+m-2)$ степенями свободы (при верной гипотезе $a_1 = a_2$). Как воспользоваться найденным распределением для проверки гипотезы, объяснено на примере гипотезы a).

Мы видим, что процедура проверки гипотез сводится к отысканию статистики, распределение которой не зависит от параметров закона распределения наблюдений, так как для такого не зависящего ни от чего распределения можно заранее составить таблицы. Сложность проблемы Беренса — Фишера состоит в том, что не удастся освободиться от зависимости от двух параметров σ_1 и σ_2 практически приемлемым способом.

З а м е ч а н и е. При практической проверке гипотез обычно недостаточно знания одного уровня значимости. Требуется знать (хотя бы приблизительно) функцию мощности (см. § 5, стр. 43). Для всех рассмотренных примеров проверки гипотез существуют таблицы и графики для вычислений их мощностей (см., например, таблицы Я. Янко [29]). Мы, однако, не будем рассматривать соответствующих «нецентральных» распределений Стьюдента, Пирсона и Фишера, на которых основано вычисление функций мощности. Приближенное представление о функции мощности можно получить, считая дисперсию σ^2 отдельного наблюдения точно совпадающей с s^2 . При известной σ^2 для вычисления функции мощности достаточно использовать уже рассмотренные распределения. Читатель, интересующийся точными выражениями для функций мощности, может найти их, например, в книге Г. Шеффе [27].

§ 12

ДАЛЬНЕЙШИЕ ПРИМЕНЕНИЯ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

Общая линейная модель, рассмотренная в предыдущем параграфе, позволяет охватить с единой точки зрения классические приемы обработки наблюдений. Мы кратко рассмотрим их сейчас.

12.1. Уравнивание измерений. Пусть наши измерения x_i имеют вид

$$x_i = a_i + \delta_i,$$

где $\delta_1, \dots, \delta_n$ — имеющие распределение $N(0, \sigma)$ независимые ошибки наблюдений, а относительно истинных значений измеряемых величин a_1, a_2, \dots, a_n известно, что они связаны системой уравнений

$$F_k(a_1, a_2, \dots, a_n) = 0, \quad k = 1, \dots, m, \quad (12.1)$$

где, разумеется, рассматривается лишь случай $m < n$ (в противном случае мы могли бы найти величины a_1, \dots, a_n прямо из системы уравнений без всяких измерений). Обычно считается, что в точке (a_1, \dots, a_n) матрица Якоби

$$\left\| \frac{\partial F_k}{\partial a_j} \right\|, \quad k = 1, \dots, m; \quad j = 1, \dots, n$$

системы (12.1) имеет наивысший возможный ранг m . Ошибки измерений $\delta_i = x_i - a_i$ считают настолько малыми, что их квадратами можно пренебречь.

Ставится вопрос о том, как, используя информацию о неизвестных величинах a_1, \dots, a_n , уточнить их полученные из измерений значения x_1, \dots, x_n , т. е. найти такие поправки $\Delta x_1, \dots, \Delta x_n$, чтобы так называемые «уравненные» измерения

$$x'_i = x_i + \Delta x_i,$$

во-первых, удовлетворяли уравнениям

$$F_k(x'_1, \dots, x'_n) = 0, \quad k = 1, \dots, m,$$

а во-вторых, были бы в каком-то разумном смысле ближе к истинным значениям a_1, \dots, a_n , чем «неуровненные», т. е. полученные прямо из наблюдений величины x_1, \dots, x_n .

Высокоточные измерения проводятся обычно в несколько этапов, так что хорошее приближение a_1^0, \dots, a_n^0 к величинам a_1, \dots, a_n можно считать уже известным. Поэтому с практической точки зрения система (12.1) эквивалентна системе

$$F_k(a_1, \dots, a_n) \approx F_k(a_1^0, \dots, a_n^0) + \sum_{i=1}^n \frac{\partial F_k(a_1^0, \dots, a_n^0)}{\partial a_i} (a_i - a_i^0) = 0, \quad (12.2)$$

которая получается линеаризацией системы (12.1). Однако система (12.2) относительно неизвестных a_i является линейной. Следовательно, вектор (a_1, \dots, a_n) должен лежать в некотором линейном подмногообразии M , т. е. мы имеем частный случай общей линейной модели.

Если точка (a_1^0, \dots, a_n^0) лежит на поверхности

$$F_k(a_1^0, \dots, a_n^0) = 0, \quad k = 1, \dots, m, \quad (12.3)$$

то линейное подмногообразие M , выделяемое уравнениями (12.2), является касательным к этой поверхности. В качестве приближенного значения для вектора $a = (a_1, \dots, a_n)$ мы должны, согласно общей теории метода наименьших квадратов, взять вектор

$$x' = (x'_1, x'_2, \dots, x'_n) = \text{proj}_M x = \text{proj}_M (x_1, \dots, x_n),$$

где $x = (x_1, \dots, x_n)$.

Если замена поверхности (12.1) ее касательной плоскостью (12.2) практически не оказывает влияния, то вектор $\text{proj}_M x$ можно принять за вектор уравненных измерений. Квадрат расстояния

$$\sum_{i=1}^n (x'_i - a_i)^2 = \|\text{proj}_M x - a\|^2 = \|\text{proj}_M (a + \delta) - a\|^2 \approx \|\text{proj}_M \delta\|^2$$

имеет распределение $\sigma^2 \chi^2_{\dim M}$ (приближенное равенство стоит потому, что лишь приближенно верно, что $a \in M$). В то же время квадрат расстояния

$$\sum_{i=1}^n (x_i - a_i)^2 = \|\delta\|^2$$

имеет распределение $\sigma^2 \chi_n^2$. При малой размерности $\dim M = n - m$ разница между $\chi_{\dim M}^2$ и χ_n^2 может быть очень чувствительной.

Если (a_1^0, \dots, a_n^0) не лежит на поверхности

$$F_k(a_1^0, \dots, a_n^0) = 0, \quad k = 1, \dots, m, \quad (12.3)$$

то нетрудно проверить, что проекция на многообразие M , задаваемое уравнениями (12.2), лишь на величину порядка $(a_i - a_i^0)^2$ отличается от проекции на касательную плоскость в точке (a_1, \dots, a_n) . Поэтому, полагая, что квадратами неточностей можно пренебречь, получаем, что и в этом случае при применении метода наименьших квадратов практически получится тот же результат. По-видимому, то же самое получится, если не считать, что точка (a_1^0, \dots, a_n^0) известна заранее, а допустить, что она получается из тех же самых наблюдений x_1, \dots, x_n , например, следующим образом: полагаем $a_1^0 = x_1, \dots, a_{n-m}^0 = x_{n-m}$ и затем m оставшихся значений $a_{n-m+1}^0, \dots, a_n^0$ находим из уравнений (12.3).

Мы ограничиваемся здесь этими краткими замечаниями об уравнивании измерений. В областях науки с давно культивируемой статистической обработкой (например, в геодезии) уравнивание измерений превратилось в особую науку, изложение которой выходит за рамки данной книги. Изложение математической теории уравнивания измерений можно найти в книге Ю. В. Линника [15]. Здесь мы ограничимся лишь приведением простейшего примера.

Пример 12.1. Два груза a_1 и a_2 сначала взвесили порознь, причем получились результаты x_1 и x_2 , а затем положили на чашку весов вместе и получили результат взвешивания x_3 . Дать оценки для весов a_1 и a_2 .

Решение. Примем модель $x_i = a_i + \delta_i$, где $i = 1, 2, 3$. Очевидно, что вектор $a = (a_1, a_2, a_3)$ лежит в подпространстве $L = \{a = (a_1, a_2, a_3) : a_3 = a_1 + a_2\}$. Ортогональный вектор к этому подпространству есть $e = \left\{ \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}} \right\}$, поэтому

$$\begin{aligned} \text{proj}_L x &= x - (x, e) e = x - \frac{x_1 + x_2 - x_3}{\sqrt{3}} \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}} \right) = \\ &= \left\{ x_1 - \frac{x_1 + x_2 - x_3}{3}, x_2 - \frac{x_1 + x_2 - x_3}{3}, x_3 + \frac{x_1 + x_2 - x_3}{3} \right\}. \end{aligned}$$

Следовательно,

$$a_1 \approx \frac{2x_1 + (x_3 - x_2)}{3}, \quad a_2 \approx \frac{2x_2 + (x_3 - x_1)}{3}, \quad a_3 \approx \frac{x_1 + x_2 + 2x_3}{3}.$$

Заметим, что в предшествовавшие общие рассуждения a_1, a_2, \dots, a_n входили на равных правах, в то время как сейчас нам интересны веса a_1 и a_2 и совершенно не интересна сумма весов

$a_3 = a_1 + a_2$. Красивая оценка выигрыша в точности, связанная с заменой $\sigma^2 \chi_n^2$ на $\sigma^2 \chi_{\dim M}^2 = \sigma^2 \chi_{n-m}^2$, в рассматриваемом случае лишена смысла. Используя модель $x_i = a_i + \delta_i$, имеем

$$\hat{a}_1 = \frac{2x_1 + (x_3 - x_2)}{3} = a_1 + \frac{1}{3} (2\delta_1 + \delta_3 - \delta_2),$$

$$\hat{a}_2 = \frac{2x_2 + (x_3 - x_1)}{3} = a_2 + \frac{1}{3} (-\delta_1 + 2\delta_2 + \delta_3).$$

Дисперсии оценок \hat{a}_1 и \hat{a}_2 равны $\frac{2}{3} \sigma^2$, где $\sigma^2 = D\delta_i$. При отсутствии третьего взвешивания x_3 дисперсия оценок x_1 и x_2 для a_1 и a_2 равнялась бы, очевидно, σ^2 . Наличие третьего взвешивания x_3 позволяет дать некую оценку для σ^2 , исходя из того, что

$$s^2 = \|(x, e) e\|^2 = (x, e)^2 = \frac{(x_1 + x_2 - x_3)^2}{3}$$

имеет распределение $\sigma^2 \chi_1^2$. С помощью таблиц распределения χ_1^2 находим

$$P \left\{ 0,0158 \leq \frac{(x, e)^2}{\sigma^2} \leq 2,706 \right\} = 0,8.$$

Следовательно,

$$P \{ 0,37 (x, e)^2 \leq \sigma^2 \leq 63,3 (x, e)^2 \} = 0,8.$$

Очевидно, при малом числе степеней свободы (в данном случае при одной степени свободы) оценка для σ^2 получается чрезвычайно грубой.

12.2. Определение параметров из эксперимента. Близкой к только что рассмотренной задаче является задача определения параметров. Пусть условия эксперимента характеризуются переменной величиной t в том смысле, что результат наблюдения $x(t)$ имеет вид

$$x(t) = f(t, c_1, c_2, \dots, c_k) + \delta(t),$$

где вид функции $f(t, c_1, \dots, c_k)$ нам известен, а неизвестны лишь значения параметров c_1, \dots, c_k . Например, зависимость длины $l(t)$ стержня от температуры t имеет вид

$$l(t) = c_0 + c_1 t,$$

где параметры c_0 и c_1 обычно приходится определять из опыта. Для определения параметров c_1, \dots, c_k мы по своему произволу даем переменной t значения t_1, t_2, \dots, t_n . При этом t_1, \dots, t_n не обязаны все быть различными, однако требуется, чтобы различных значений среди t_1, \dots, t_n было (гораздо) больше, чем требуется для однозначного решения системы

$$x(t_i) = f(t_i, c_1, \dots, c_k), \quad i = 1, \dots, n$$

относительно c_1, \dots, c_k . Подчеркнем, что значения t_1, \dots, t_n считаются неслучайными и известными точно.

Принимается модель, согласно которой результаты наблюдений имеют вид

$$x_i = f(t_i, c_1, \dots, c_k) + \delta_i$$

с обычными предположениями об ошибках δ_i (независимость и нормальность $N(0, \sigma)$).

Предположим, что известно некоторое приближение c_1^0, \dots, c_k^0 для параметров c_1, \dots, c_k . Пренебрегая членами порядка выше первого, имеем

$$\begin{aligned} f(t_i, c_1, \dots, c_k) &= f(t_i, c_1^0, \dots, c_k^0) + \\ &+ \sum_{j=1}^k \frac{\partial f(t_i, c_1^0, \dots, c_k^0)}{\partial c_j} (c_j - c_j^0). \end{aligned}$$

Полагая $y_i = x_i - f(t_i, c_1^0, \dots, c_k^0)$, имеем

$$y_i = \sum_{j=1}^k \frac{\partial f(t_i, c_1^0, \dots, c_k^0)}{\partial c_j} (c_j - c_j^0) + \delta_i.$$

Поскольку $M\delta_i = 0$, то вектор (My_1, \dots, My_k) лежит в гиперплоскости, натянутой на k векторов:

$$\left\{ \frac{\partial f(t_1, c_1^0, \dots, c_k^0)}{\partial c_j}, \frac{\partial f(t_2, c_1^0, \dots, c_k^0)}{\partial c_j}, \dots, \frac{\partial f(t_n, c_1^0, \dots, c_k^0)}{\partial c_j} \right\},$$

$$j = 1, 2, \dots, k.$$

Таким образом, мы вновь имеем линейную модель.

Ничто не мешает считать функцию $f(t, c_1, \dots, c_k)$ векторной и интерпретировать ее как положение планеты в момент времени t , считая параметры c_1, \dots, c_k параметрами траектории.

Мы, естественно, не можем вдаваться здесь в подробности, которые, как и в случае уравнивания измерений, вновь составляют целую науку.

12.3. Сглаживание измерений. Задача определения параметров физического закона, кратко описанная в предыдущем пункте, требует знания этого закона с точностью до параметров. Часто, однако, приходится иметь дело с таким случаем, когда несовершенство наших знаний столь глубоко, что об истинном характере наблюдаемой закономерности мы не имеем представления. Тем не менее истинные закономерности обычно имеют «регулярный» характер, в то время как случайные ошибки, их искажающие, — «нерегулярный». В настоящее время нет возможности на научном уровне строгости сказать, что такое «регулярный» или «нерегулярный»

характер. В зависимости от формализации этих понятий мы будем приходить к тому или иному способу статистической обработки.

Самая «регулярная» из функций — несомненно, константа. Если мы предположим, что во всех наблюдениях измеряется одно и то же, а разброс данных объясняется чистой случайностью, то в качестве модели мы получим выборку, и статистическая обработка сведется к определению закона распределения. Конечно, при этом делается трудно проверяемое предположение о статистической однородности и независимости отдельных наблюдений.

Бывает, однако, и так, что статистическая однородность наблюдений явно отсутствует, как в рассмотренном в предыдущем пункте примере, когда на результат эксперимента влияла переменная t , выбираемая нами по нашему усмотрению. В этом случае в модель включается предположение о статистической однородности не самих наблюдений, а лишь ошибок наблюдений δ_i .

Если же закон явления $f(t, c_1, \dots, c_k)$ неизвестен, расплывчатое предположение о «регулярности» обычно интерпретируют как предположение о том, что неизвестная функция $f(t, c_1, \dots, c_k)$ является попросту многочленом. На чем основано и к чему приводит данное предположение?

Известна математическая теорема, согласно которой любая непрерывная функция на любом отрезке с любой степенью точности может быть аппроксимирована многочленом достаточно высокой степени. На первый взгляд, троекратное повторение слова «любой» способно убедить каждого в том, что ничего лучшего и не требуется. Однако возможность этого троекратного повторения достигается в сущности путем замены в словосочетании «достаточно высокой степени» слова «достаточно» на слово «чрезвычайно».

А между тем график многочлена чрезвычайно высокой степени, пожалуй, никто не назовет «регулярной» функцией. Далее, достигнув хорошего приближения многочленом, скажем синусоиды, на одном отрезке, мы увидим с огорчением, что согласие резко ухудшится, если рассматривать эти же функции на более широком отрезке. Следовательно, экстраполяция экспериментальной зависимости, полученной сглаживанием с помощью многочлена, рискованна. Наконец, коэффициенты многочлена высокой степени плохо определяются по данным, содержащим погрешности, так что ожидаемого освобождения от случайных ошибок при резком повышении степени многочлена может и не получиться.

Все, что было сказано об отрицательных свойствах сглаживания многочленами, разумеется, относится и к другим классам сглаживающих функций.

Итак, для успешного сглаживания многочленом нужно, чтобы настоящая зависимость не слишком отличалась от многочлена, чтобы сглаживание производилось на не слишком большом отрезке и многочленом не слишком высокой степени.

Сглаживание обычно имеет две цели: 1) поточнее определить экспериментальные данные за счет освобождения от случайных

ошибок, допущенных в каждом отдельном опыте; 2) редуцировать большое количество экспериментальных данных к немногим коэффициентам многочлена. Надо сказать, что эти задачи часто могут быть с тем же успехом (или неуспехом) решены без использования методов математической статистики, просто путем проведения на глаз гладкой кривой. Однако при этом вряд ли можно судить о точности проведения этой кривой, т. е. о соответствующих доверительных интервалах для истинных значений наблюдаемой функции в разных точках. Математическую статистику ее сторонники характеризуют словами: «здравый смысл плюс точность». Точнее было бы сказать: «плюс оценка возможной точности».

Перейдем к математической стороне дела. Положим

$$P_m(t) = c_0 + c_1 t + \dots + c_m t^m,$$

где c_0, c_1, \dots, c_m — неизвестные нам коэффициенты многочлена (степень m многочлена пока считается известной). Примем следующую модель для результатов наблюдений:

$$x_i = P_m(t_i) + \delta_i, \quad i = 1, \dots, n,$$

где δ_i — независимые случайные величины с распределением $N(0, \sigma)$. Требуется по результатам наблюдений $x = (x_1, \dots, x_n)$ определить коэффициенты c_0, c_1, \dots, c_m и σ .

Для решения задачи введем векторы

$$T^0 = (1, \dots, 1), \quad T^1 = (t_1, \dots, t_n), \dots,$$

$$T^m = (t_1^m, \dots, t_n^m), \quad \delta = (\delta_1, \dots, \delta_n).$$

В таком случае имеем

$$x = \sum_{l=0}^m c_l T^l + \delta.$$

Обозначая через L линейную оболочку

$$L = L(T^0, T^1, \dots, T^m)$$

векторов T^0, T^1, \dots, T^m , мы видим, что вновь имеем дело с общей линейной моделью.

Заметим, что мы не требуем, чтобы все t_1, t_2, \dots, t_n были различны: нам нужно лишь, чтобы векторы T^0, T^1, \dots, T^m были линейно независимы (в частности, $m+1 \leq n$).

Ясно, что если степень многочлена $m \geq n-1$, где n — число наблюдений, то многочлен можно провести точно через все наблюдаемые точки. При этом никакого отделения «регулярной» компоненты от «нерегулярных» ошибок не получится.

По общему правилу, оценкой для вектора $\sum_{l=0}^m c_l T^l$ будет проекция $\text{proj}_L x$. Следовательно, оценками для c_l будут числа γ_l такие, что

$$\sum_{l=0}^m \gamma_l T^l = \text{proj}_L x,$$

иными словами

$$\begin{aligned} \|x - \sum_{l=0}^m \gamma_l T^l\|^2 &= \min_{c_0, \dots, c_m} \|x - \sum_{l=0}^m c_l T^l\|^2 = \\ &= \min_{c_0, \dots, c_m} \sum_{i=1}^n (x_i - c_0 - c_1 t_i - \dots - c_m t_i^m)^2. \end{aligned}$$

Следовательно, γ_i являются решением следующей системы так называемых «нормальных» уравнений:

$$\frac{\partial}{\partial \gamma_k} \sum_{i=1}^n (x_i - \gamma_0 - \gamma_1 t_i - \dots - \gamma_m t_i^m)^2 = 0, \quad k = 0, \dots, m$$

или

$$\sum_{i=1}^n (x_i - \gamma_0 - \gamma_1 t_i - \dots - \gamma_m t_i^m) t_i^k = 0, \quad k = 0, \dots, m.$$

Заметим, что величина

$$\Delta^2 = \|x - \text{proj}_L x\|^2 = \sum_{i=1}^n (x_i - \gamma_0 - \gamma_1 t_i - \dots - \gamma_m t_i^m)^2,$$

называется «кажущейся ошибкой», имеет распределение $\sigma^2 \chi_{n-m-1}^2$ (правильно, но слишком длинно называть Δ^2 «суммой квадратов кажущихся ошибок»: поскольку значения $\gamma_0, \gamma_1, \dots, \gamma_m$ не есть истинные значения c_0, c_1, \dots, c_m , то и разность $x_i - \gamma_0 - \dots - \gamma_m t_i^m$ не есть истинная ошибка наблюдения δ_i , а есть «кажущаяся ошибка»).

Для построения доверительных интервалов для неизвестных коэффициентов многочлена предпочтительнее записать его не в виде $c_0 + c_1 t + \dots + c_m t^m$, т. е. в виде линейной комбинации многочленов $1, t, \dots, t^m$, а сделать переход к другому базису. Положим

$$W^0 = T^0,$$

$$W^1 = T^1 - \frac{(T^1, W^0)}{(W^0, W^0)} W^0,$$

$$\dots \dots \dots$$

$$W^m = T^m - \frac{(T^m, W^0)}{(W^0, W^0)} W^0 - \dots - \frac{(T^m, W^{m-1})}{(W^{m-1}, W^{m-1})} W^{m-1}.$$

Для случая равноотстоящих точек t_1, t_2, \dots, t_n , т. е. когда $t_1 - t_2 = t_2 - t_3 = \dots = t_{n-1} - t_n$, такая ортогонализация может быть сделана раз навсегда для каждого значения n . Соответствующие полиномы называются ортогональными полиномами Чебышева (см., например, [4]). Но вообще ортогонализация удлинняет вычисления, а также имеет то неприятное свойство, что при добавлении наблюдений должна переделываться заново.

В новом базисе, очевидно,

$$a = \sum_{l=0}^m c_l T^l = \sum_{l=0}^m b_l W^l,$$

где b_l связаны с c_l линейным преобразованием. Оценками для величин b_l будут величины β_l такие, что

$$\sum_{l=0}^m \beta_l W^l = \text{proj}_L x.$$

В силу ортогональности W^0, \dots, W^m имеем

$$\beta_l = \frac{(x, W^l)}{(W^l, W^l)}, \quad l = 0, \dots, m.$$

Так как $x = a + \delta$, то

$$\beta_l = \frac{(a + \delta, W^l)}{(W^l, W^l)} = \frac{(a, W^l)}{(W^l, W^l)} + \frac{(\delta, W^l)}{(W^l, W^l)} = b_l + \frac{(\delta, W^l)}{(W^l, W^l)}.$$

Заметим, что β_l связана с проекцией $x = a + \delta$ на L , а кажущаяся ошибка Δ^2 — с проекцией x на ортогональное дополнение к L . Следовательно, β_l и Δ^2 — независимые случайные величины. Нетрудно подсчитать, что

$$D(\beta_l - b_l) = \frac{\sigma^2}{(W^l, W^l)},$$

где $\sigma^2 = D\delta_i$ — дисперсия ошибки отдельного наблюдения. Следовательно, отношение

$$\frac{(\beta_l - b_l) \sqrt{(W^l, W^l)}}{\sqrt{\frac{1}{n-m-1} \Delta^2}} = t_{n-m-1}$$

имеет распределение Стьюдента с $(n-m-1)$ степенями свободы. Это обстоятельство позволяет легко написать доверительный интервал для b_l . Поскольку оценки β_l , $l=0, \dots, m$, независимы в совокупности, отношение

$$\frac{\frac{1}{m} \sum_{l=0}^m (\beta_l - b_l)^2 (W^l, W^l)}{\frac{1}{n-m-1} \Delta^2} = F_{m, n-m-1}$$

имеет распределение Фишера с указанным числом степеней свободы. Поэтому по заданному α легко указать такое F_α , что

$$P \left\{ \frac{1}{m} \sum_{l=0}^m (\beta_l - b_l)^2 (W^l, W^l) < \frac{F_\alpha}{n-m-1} \Delta^2 \right\} = 1 - \alpha.$$

Таким образом, доверительная область для вектора (b_0, \dots, b_m) является эллипсоидом.

Теперь рассмотрим вопрос о выборе степени m сглаживающего многочлена. Здесь нет строгих правил. Интуитивно ясно, что чем меньше m , при котором отклонения наблюдений от многочлена можно считать чисто случайными, тем лучше. Надо начать с небольшой степени ($m=1, 2$, вряд ли более 3) и увеличивать m лишь при наличии достаточных экспериментальных оснований.

Приведем статистический критерий для решения этого вопроса. Пусть мы начали приближение многочленом $P_{m_1}(t)$ степени m_1 , а затем решили попробовать многочлен $P_{m_2}(t)$ степени $m_2 > m_1$. Если наблюдаемый эффект действительно задается гладкой функцией, то даже при не слишком большом увеличении степени многочлена точность приближения обычно резко растет. Поэтому имеются известные основания считать, что многочленом $P_{m_2}(t)$ истинная зависимость может быть описана точно. Иначе это можно выразить следующим образом.

Положим $L_1 = L(T^0, T^1, \dots, T^{m_1})$, $L_2 = L(T^0, T^1, \dots, T^{m_2})$, $m_2 > m_1$. Если $x = (x_1, \dots, x_n)$ — наши наблюдения, $a = (Mx_1, \dots, Mx_n)$ заведомо входит в L_2 . Возможно, однако, что на самом деле $a \in L_1$, т. е. достаточно ограничиться многочленом степени m_1 .

Проверим эту гипотезу. Конечно, полагая

$$\Delta_1^2 = \|x - \text{proj}_{L_1} x\|^2, \quad \Delta_2^2 = \|x - \text{proj}_{L_2} x\|^2,$$

мы всегда будем иметь $\Delta_2^2 < \Delta_1^2$, поскольку $L_2 \supset L_1$. Но, спрашивается, достаточно ли сильно уменьшение кажущейся ошибки при переходе от m_1 к m_2 , чтобы поверить в необходимость такого перехода, т. е. в то, что $a \in L_1$? Для ответа на этот вопрос нужно составить некоторую статистику с заранее известным распределением. Заметим, что, обозначая через $L_2 \setminus L_1$ ортогональное дополнение в L_2 к подпространству L_1 , получаем

$$\Delta_1^2 - \Delta_2^2 = \|\text{proj}_{L_2 \setminus L_1} x\|^2,$$

$$\dim L_2 \setminus L_1 = m_2 - m_1.$$

Если верна гипотеза $a \in L_1$, то

$$\text{proj}_{L_2 \setminus L_1} x = \text{proj}_{L_2 \setminus L_1} (a + \delta) = \text{proj}_{L_2 \setminus L_1} \delta.$$

Поэтому $\Delta_1^2 - \Delta_2^2$ имеет распределение $\sigma^2 \chi_{m_2 - m_1}^2$. В то же время, поскольку $a \in L_2$, Δ_2^2 имеет распределение $\sigma^2 \chi_{n - m_2 - 1}^2$. Следовательно, при верной гипотезе $a \in L_1$ отношение

$$\frac{\frac{1}{m_2 - m_1} (\Delta_1^2 - \Delta_2^2)}{\frac{1}{n - m_2 - 1} \Delta_2^2} = F_{m_2 - m_1, n - m_2 - 1}$$

имеет распределение Фишера с указанным числом степеней свободы. Поэтому гипотеза $a \in L_1$ может быть проверена стандартным путем, и при необходимости отклонить ее следует увеличить степень многочлена, правда, возможно не с m_1 до m_2 , а, например, с m_1 до $m_1 + 1$. На этот счет нет точных рекомендаций, равно как и рекомендаций о значениях m_1 и m_2 . Разумеется, желательно, чтобы данные примерно одинакового происхождения сглаживались многочленами одинаковой степени. Только проверка результатов на нескольких массивах данных может дать уверенность в устойчивости результата.

12.4. Линейная регрессия. Многочлен $\gamma_0 + \gamma_1 t + \dots + \gamma_m t^m$, сглаживающий экспериментальные данные, проведенный методом наименьших квадратов, называется *линией параболической регрессии* порядка m . Если $m=1$, то прямая линия $\gamma_0 + \gamma_1 t$ называется *линией регрессии* (подразумевается первого порядка). Если переменная t , характеризующая условия опыта, меняется в нешироких пределах, то любую зависимость можно приближенно считать линейной. Поэтому проведение линии регрессии (обычно — на глаз, иногда — с помощью вычислений) является одним из любимых приемов первичного осмысливания результатов эксперимента. С другой стороны, этот прием находит некоторое применение при оценке согласия между теорией и экспериментом. Мы рассмотрим вскоре пример такого рода, но сначала выпишем явно соответствующие формулы.

Имеем следующую модель результатов эксперимента:

$$x_i = c_0 + c_1 t_i + \delta_i, \quad i = 1, \dots, n,$$

где δ_i — ошибки наблюдений. Вводя векторы

$$x = (x_1, \dots, x_n), \quad T^0 = (1, \dots, 1),$$

$$T^1 = (t_1, \dots, t_n), \quad \delta = (\delta_1, \dots, \delta_n),$$

запишем эту модель в виде

$$x = c_0 T^0 + c_1 T^1 + \delta.$$

В соответствии с процедурой, изложенной в предыдущем пункте, ортогонализируем систему векторов T^0 и T^1 :

$$W^0 = T^0,$$

$$W^1 = T^1 - \frac{(T^1, W^0)}{(W^0, W^0)} T^0 = T^1 - \left(\frac{1}{n} \sum_{i=1}^n t_i \right) T^0 = T^1 - \bar{T},$$

где через \bar{T} обозначен вектор, все компоненты которого равны $\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i$. В обозначениях W^0, W^1 имеем

$$x = c_0 W^0 + c_1 (W^1 + \bar{t} W^0) = b_0 W^0 + b_1 W^1 + \delta,$$

где $b_0 = c_0 + c_1 \bar{t}$, $b_1 = c_1$. Если иметь в виду эти соотношения, то, конечно, безразлично, рассуждать ли в терминах c_0 и c_1 или в терминах b_0 и b_1 . Сохраним обозначение $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ (хотя x_1, x_2, \dots, x_n теперь не образуют выборки, мы используем старое обозначение для их среднего арифметического). Имеем следующие оценки β_0 и β_1 для b_0 и b_1 :

$$\beta_0 = \frac{(x, W^0)}{(W^0, W^0)} = \bar{x},$$

$$\beta_1 = \frac{(x, W^1)}{(W^1, W^1)} = \frac{(x - \bar{x} W^0, W^1)}{(W^1, W^1)} = \frac{\sum_{i=1}^n (x_i - \bar{x})(t_i - \bar{t})}{\sum_{i=1}^n (t_i - \bar{t})^2}.$$

Мы имеем точное равенство

$$x = \beta_0 W^0 + \beta_1 W^1 + (x - \beta_0 W^0 - \beta_1 W^1).$$

Если β_0 близко к b_0 , β_1 близко к b_1 , то член $x - \beta_0 W^0 - \beta_1 W^1$ близок к δ . Допуская, что ошибки малы, имеем

$$x \approx \beta_0 W^0 + \beta_1 W^1,$$

или, подставляя $\beta_0 = \bar{x}$, получаем

$$x_i - \bar{x} \approx \beta_1 (t_i - \bar{t}).$$

Так обычно записывается линия регрессии.

Из предыдущего пункта известно, что кажущаяся ошибка

$$\Delta^2 = \sum_{i=1}^n (x_i - \bar{x} - \beta_1 (t_i - \bar{t}))^2$$

имеет распределение $\alpha^2 \chi^2_{n-2}$. Кроме того, оценки β_0 и β_1 не зависят от Δ^2 и друг от друга, причем разности $\beta_0 - b_0$ и $\beta_1 - b_1$ имеют нормальные распределения

$$N\left(0, \frac{\sigma}{\sqrt{(W^0, W^0)}}\right), \quad N\left(0, \frac{\sigma}{\sqrt{(W^1, W^1)}}\right).$$

Следовательно, отношения

$$\frac{(\beta_0 - b_0) \sqrt{(W^0, W^0)}}{\Delta} \sqrt{n-2}, \quad \frac{(\beta_1 - b_1) \sqrt{(W^1, W^1)}}{\Delta} \sqrt{n-2}$$

имеют каждое распределение Стюдента t_{n-2} . Поэтому распределение Стюдента может быть применено при проверке любой из гипотез $b_0 = \tilde{b}_0$ или $b_1 = \tilde{b}_1$, где \tilde{b}_0 и \tilde{b}_1 — заданные числа. Для одновременной проверки этих двух гипотез естественно применить статистику

$$\frac{\frac{1}{2} [(\beta_0 - \tilde{b}_0)^2 (W^0, W^0) + (\beta_1 - \tilde{b}_1)^2 (W^1, W^1)]}{\frac{1}{n-2} \Delta^2},$$

которая (при верной гипотезе $b_0 = \tilde{b}_0$, $b_1 = \tilde{b}_1$) имеет распределение Фишера $F_{2, n-2}$.

После этих замечаний рассмотрим задачу экспериментальной проверки теории.

12.5. Проверка согласия между теорией и наблюдениями.

Пусть теоретически предсказаны значения t_1, t_2, \dots, t_n для измеряемой величины, а фактические наблюдения дали значения x_1, x_2, \dots, x_n . Спрашивается, можно ли считать, что, с точностью до ошибок измерений, наблюдения подтверждают теорию? Иначе говоря, можно ли расхождения $x_1 - t_1, x_2 - t_2, \dots, x_n - t_n$ отнести целиком за счет ошибок наблюдений?

Если бы математическая статистика давала объективный прием для решения этого вопроса, пригодный во всех случаях, то она, очевидно, была бы наукой наук. Но легко видеть, что это не так. Действительно, если положить $x_i - t_i = \delta_i$, где δ_i — независимые случайные величины с распределением $N(0, \sigma)$, то в принципе для $\delta_1, \delta_2, \dots, \delta_n$ возможны любые значения. Статистические приемы основаны на том, что некоторые области этих значений (критические области) объявляются маловероятными. Однако вероятность любого конкретного набора значений $\delta_1 = y_1, \dots, \delta_n = y_n$ равна нулю, так что любой набор значений можно было бы объявить невозможным. Но только что любой набор значений был объявлен возможным.

Выход из формального противоречия состоит в том, что критические области выбираются не произвольно, а исходя из представления о возможных альтернативах для гипотезы, состоящей в том, что $x_1 - t_1, \dots, x_n - t_n$ являются ошибками измерений. Иначе

говоря, предлагаются и проверяются различные вероятностные модели для результатов наблюдений.

Поскольку получаемые выводы целиком зависят от исследуемых моделей, не может быть и речи о полной объективности статистических методов. Они занимают некоторое среднее место между (недостижимой) полной объективностью и чисто субъективной оценкой на глаз. Заметим, что при правильном применении статистических методов и добросовестности субъективной оценки обычно не возникает противоречия между этими двумя подходами. Отчасти роль статистических методов состоит в том, чтобы помочь приобрести необходимые навыки для субъективной оценки.

Итак, мы должны указать в рамках статистической модели для результатов наблюдений проверяемую гипотезу и класс возможных альтернатив. Излагаемый здесь подход, основанный на понятии линии регрессии, не является единственно возможным. Например, указываемые методы проверки не всегда выявляют наличие зависимости (скажем, периодических колебаний) между разностями $x_1 - t_1, x_2 - t_2, \dots, x_n - t_n$. Однако мы должны на чем-то остановиться для определенности.

Мы будем рассматривать основную (нулевую) гипотезу вида

$$x_i = t_i + \delta_i, \quad i = 1, \dots, n,$$

где $\delta_1, \dots, \delta_n$ — независимые случайные величины, имеющие распределение $N(0, \sigma)$. Возможными альтернативами мы будем считать альтернативы двух видов:

1) $x_i = c_0 + c_1 t_i + \delta_i$, где $c_0 \neq 0$ и (или) $c_1 \neq 1$.

2) $x_i = P_m(t_i) + \delta_i$, где $P_m(t)$ — многочлены от t степени $m > 1$.

Посмотрим, как практически могут быть интерпретированы эти альтернативы.

Среди альтернатив вида 1) есть одна весьма специальная, именно гипотеза $c_1 = 0$. Эта гипотеза означает, что x_i вообще не зависят от t_i , т. е. предсказываемый теорией эффект вовсе не наблюдается. Говорить о серьезной проверке теории можно, очевидно, только в том случае, когда есть достаточные основания для того, чтобы отвергнуть гипотезу $c_1 = 0$.

Гипотеза вида 1) при $c_0 \neq 0, c_1 = 1$ говорит, очевидно, о том, что у наблюдений x_i есть систематическая ошибка c_0 .

Гипотеза вида 1) при $c_1 \neq 1$ говорит, что предсказываемый теорией эффект существует, но (при значимом отличии c_1 от 1) количественное согласие отсутствует. При малом диапазоне изменения t_1, \dots, t_n практически любая зависимость от t_1, \dots, t_n является линейной, так что при c_1 , значимо отличном от 1, получается, очевидно, лишь довольно слабое подтверждение (если не опровержение) теории.

Альтернатива вида 2) говорит о том, что эффект имеет нелинейный характер, в отличие от предсказываемого теорией эффекта, равного просто t_i . Если альтернативу 2) приходится принять в ущерб линейной альтернативе 1), то это создает сильное подозрение в том, что на самом деле наблюдаемый эффект зависит не только от предсказанного эффекта t_i , но еще и от каких-то других обстоятельств, возможно, не контролируемых в эксперименте. Может, конечно, быть и нелинейная зависимость только от t_i .

Изложенные только что способы истолкования альтернативных гипотез 1) и 2) не следует, конечно, принимать догматически; это скорее лишь примеры возможных выводов, которые можно извлечь из статистического исследования. В зависимости от конкретных условий делаемые выводы могут стать совершенно иными (читатель, разумеется, заметил, что тот факт, что $c_1 \neq 1$, может быть истолкован иногда как подтверждение, а иногда как опровержение теории — в зависимости от того, чего мы хотим: только лишь заметить эффект или точно выразить его количественно).

Способы проверки гипотезы $x_i = t_i + \delta_i$ при альтернативе 1) изложены в предыдущем пункте. Надо альтернативу 1) записать несколько иначе:

$$x_i - b_0 = b_1 (t_i - \bar{t}) + \delta_i,$$

где

$$b_0 = c_0 + b_1 \bar{t}, \quad \bar{t} = \frac{1}{n} \sum_{i=1}^n t_i, \quad b_1 = c_1.$$

Тогда проверяемая гипотеза $c_0 = 0, c_1 = 1$ переписывается в виде $b_0 = \bar{t}, b_1 = 1$. Для проверки надо воспользоваться оценками β_0 и β_1 и их свойствами, указанными в предыдущем пункте.

Способ проверки линейной гипотезы 1) при нелинейной альтернативе 2) указан в пункте 12.3. Надо иметь в виду, что при проверке соответствия между теорией и экспериментом возможны весьма разнообразные сочетания основных и альтернативных гипотез, например, основная гипотеза: $x_i = t_i + \delta_i$, альтернативная: $x_i = t_i + ct_i^2 + \delta_i$, где $c \neq 0$. Эта альтернатива (и многие другие) может быть рассмотрена в рамках общей линейной модели. Бывают ситуации, не укладывающиеся в эти рамки. Но, по-видимому, теоретическое рассмотрение всех возможных альтернатив оказалось бы практически бесполезным.

Рассмотрим пример проверки соответствия между теорией и экспериментом. Речь идет о проверке общей теории относительно по отклонению луча света в поле тяготения. Данные для статистической обработки взяты из сборника статей Альберта Эйнштейна [28]. Схема опыта, проведенного под руководством Эддингтона, состоит в следующем (рис. 12.1). Пусть звезда лежит примерно в плоскости земной орбиты. Тогда в тот момент, когда Земля

находится в положении I, звезда видна в некотором направлении I. Через полгода луч света от звезды, идущий к Земле, находящейся в положении II, должен испытать отклонение в поле тяготения Солнца, так что звезда будет видна в направлении II. Иначе говоря, звезда испытает как бы перемещение, которое можно вычислить теоретически. Конечно, наблюдать звезду из положения II можно лишь в момент полного солнечного затмения.

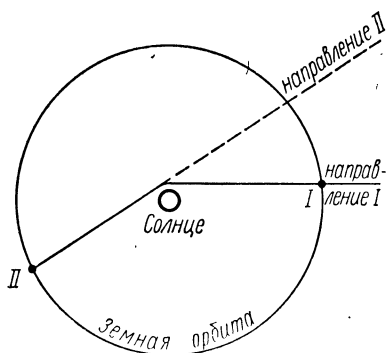


Рис. 12.1

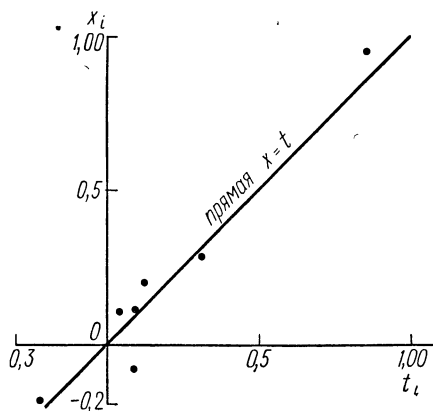


Рис. 12.2

Для наблюдения было выбрано 7 звезд. Их видимые перемещения (векторы на небесной сфере, которые из-за малости можно считать векторами на плоскости) разлагались по двум осям координат. Получились следующие результаты (в угловых секундах):

Первая координата

t_i (вычисленная) $-0,22 + 0,31 + 0,10 + 0,12 + 0,04 + 0,09 + 0,85$

x_i (наблюдённая) $-0,19 + 0,29 + 0,11 + 0,20 + 0,10 - 0,08 + 0,95$

Вторая координата

t_i (вычисленная) $+0,02 - 0,43 + 0,74 + 0,87 + 0,40 + 0,32 - 0,09$

x_i (наблюдённая) $+0,16 - 0,46 + 0,83 + 1,00 + 0,57 + 0,35 - 0,27$

Заметим, что часть результатов измерений была забракована Эддингтоном из-за технической погрешности и в книге Альберта Эйнштейна [28] не приводится.

Мы рассмотрим обработку данных по первой координате, рекомендуя читателю сделать то же самое для второй.

Первое, что следует сделать — это нарисовать в прямоугольной системе координат точки (t_i, x_i) (см. рис. 12.2). Глядя на этот график, всякий, кто имеет некоторый опыт глазомерной оценки, скажет, что согласие с теорией превосходное. Посмотрим, что же дают статистические расчеты.

Основные вычисления сводятся к вычислению следующих величин:

$$1) \bar{t} = \frac{1}{n} \sum_{i=1}^n t_i = 0,184,$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 0,197;$$

$$2) \|x - \bar{x}\|^2 = \sum_{i=1}^n (x_i - \bar{x})^2 = 0,819,$$

$$\|T^1 - \bar{T}\|^2 = \sum_{i=1}^n (t_i - \bar{t})^2 = 0,664;$$

$$3) (x - \bar{x}, T^1 - \bar{T}) = \sum_{i=1}^n (x_i - \bar{x})(t_i - \bar{t}) = 0,718,$$

где число наблюдений $n=7$.

Через эти величины легко выражаются все другие, нужные нам. Например, если обозначить через L линейную оболочку T^0 и T^1 , то «кажущаяся ошибка» Δ^2 есть

$$\begin{aligned} \Delta^2 &= \|x - \text{proj}_L x\|^2 = \|x - \bar{x} - \\ &- \text{proj}_{T^1 - \bar{T}}(x - \bar{x})\|^2 = \|x - \bar{x}\|^2 - \\ &- \frac{(x - \bar{x}, T^1 - \bar{T})^2}{\|T^1 - \bar{T}\|^2} = 0,819 - 0,776 = 0,043. \end{aligned}$$

Оценка для σ^2 (дисперсии отдельного наблюдения) есть $\frac{\Delta^2}{n-2}\Delta^2$, оценка для σ есть

$$s = \sqrt{\frac{1}{n-2} \Delta^2} = \sqrt{\frac{1}{5} \cdot 0,043} = 0,090.$$

Таким образом, ошибка отдельного наблюдения имеет порядок 0,1 угловой секунды. Поэтому сотые доли угловой секунды заведомо ненадежны. Тем не менее, наблюдения приводятся с двумя значащими цифрами, что только и делает возможной их статистическую обработку, основанную на нормальном распределении. Вообще все в этих опытах, включая отмеченную сейчас сравнительно мелкую подробность, свидетельствует о накопленной веками развития науки культуре точных измерений.

Оценки β_0 и β_1 имеют значения

$$\begin{aligned} [\beta_0] &= \bar{x} = 0,197, \\ \beta_1 &= \frac{(x - \bar{x}, T^1 - \bar{T})}{\|T^1 - \bar{T}\|^2} = 1,081. \end{aligned}$$

Приступим теперь к проверке гипотез.

Начнем с обидной для Эддингтона и Эйнштейна гипотезы $b_1=0$ (т. е. гипотезы об отсутствии связи между теорией и наблюдениями). Статистика

$$\frac{(\beta_1 - b_1) \|T^1 - \bar{T}\|}{\Delta} \sqrt{n-2} = t_{n-2}$$

имеет распределение Стьюдента с $n-2=5$ степенями свободы. При $b_1=0$ получаем

$$t_5 = 9,75.$$

Вероятность $P\{|t_5| \geq 9,75\}$ обычно не приводится в таблицах распределения Стьюдента. Однако с помощью таблиц Большева и Смирнова [4] можно установить, что эта вероятность равна примерно $2 \cdot 10^{-4}$. Таким образом, обидная гипотеза $b_1=0$, несомненно, должна быть отвергнута.

Теперь проверим согласие с гипотезой $x_i = t_i + \delta_i$, т. е. гипотезы $b_0 = \bar{t}$ и $b_1 = 1$.

Для проверки гипотезы $b_0 = \bar{t}$ значение статистики равно

$$\frac{(\beta_0 - \bar{t}) \sqrt{(W^0, W^0)}}{\Delta} \sqrt{n-2} = 0,38$$

(заметим, что $(W^0, W^0) = n=7$).

Для проверки гипотезы $b_1 = 1$ значение статистики равно

$$\frac{(\beta_1 - 1) \|T^1 - \bar{T}\|}{\Delta} \sqrt{n-2} = 0,73.$$

Оба эти значения заведомо незначимы для распределения Стьюдента с пятью степенями свободы (но и не настолько малы, чтобы можно было заподозрить подгонку наблюдений).

Для совместной гипотезы $b_0 = \bar{t}$, $b_1 = 1$ имеем следующую статистику:

$$\frac{\frac{1}{2} [(\beta_0 - \bar{t}) (W^0, W^0) + (\beta_1 - 1) \|T^1 - \bar{T}\|^2]}{\frac{1}{n-2} \Delta^2} = 0,35,$$

что также совершенно не значимо для распределения Фишера $F_{2,5}$.

Найдем еще доверительный интервал для σ . С помощью таблиц распределения χ^2_5 находим

$$P\left\{0,831 < \frac{\Delta^2}{\sigma^2} < 12,8\right\} = 0,95$$

или

$$P\{0,23 > \sigma > 0,056\} = 0,95.$$

Таким образом, мы вновь видим, что при малом числе наблюдений ($n=7$) дисперсия отдельного наблюдения оценивается весьма грубо. Заметим, что мы здесь применили чисто формально прием оценки дисперсии с помощью кажущейся ошибки после проведения линии регрессии. Дело в том, что, приняв модель

$$x_i = t_i + \delta_i, \quad i = 1, \dots, n,$$

мы имеем $\delta_i = x_i - t_i$, т. е. разности

$$x_1 - t_1, \quad x_2 - t_2, \dots, \quad x_n - t_n$$

образуют выборку из распределения $N(0, \sigma)$.

Поэтому величина

$$\sum_{i=1}^n (x_i - t_i)^2 = 0,050$$

имеет распределение $\sigma^2 \chi_n^2$. С помощью таблиц распределения χ_7^2 получаем

$$P\{0,17 > \sigma > 0,055\} = 0,95,$$

что несколько (но немногим) лучше.

Заметим, что все вычисления мы вели с расчетом на получение двух значащих цифр в окончательных результатах (т. е. в тех числах, с помощью которых мы входили в таблицы для определения значимости или незначимости тех или иных статистик). Не обязательно, чтобы обе значащие цифры были верными. Поэтому промежуточные вычисления в статистике обычно можно вести также с двумя значащими цифрами. Впрочем, если применять современные средства механизации вычислений, вопрос о числе значащих цифр почти полностью теряет интерес.

Введение

В первой части этой книги изложен элементарный курс теории вероятностей так, как его себе представляет автор. Цель второй части методических замечаний — объяснить, почему в соответствующем месте первой части выбран именно тот способ изложения, который там есть, чем этот способ отличается от других возможных или принятых в литературе, а также каких ошибок, по мнению автора, следует избегать. Традиция в преподавании, даже если речь идет о преподавании математики, закрепляет иногда весьма странные вещи. Рассмотрим, например, изложение вопроса об условном экстремуме в курсе математического анализа. Речь идет о следующей хорошо известной задаче: найти точки экстремума функции $F(x)$ на поверхности, выделяемой системой уравнений $G_1(x)=0, \dots, G_h(x)=0$, где $x=(x_1, \dots, x_n)$. Для безусловного экстремума (когда экстремум ищется во всем пространстве) необходимым является условие $\text{grad} F(x)=0$. В случае одного условия: поверхности $G(x)=0$ в точке экстремума $\text{grad} F(x)$ должен быть ортогонален к поверхности, так как в противном случае в касательной плоскости существует направление, вдоль которого производная от функции F отлична от нуля (тогда и производная вдоль кривой на поверхности, касающейся этого направления, отлична от нуля). Поскольку $\text{grad} G(x)$ есть нормаль к поверхности, в точке экстремума x_0 имеем при некотором λ

$$\text{grad} F(x_0) = \lambda \text{grad} G(x_0),$$

иначе говоря, при некотором λ

$$\text{grad} (F - \lambda G)(x_0) = 0.$$

В случае нескольких условий $G_1(x)=0, \dots, G_h(x)=0$ градиент должен лежать в нормальной плоскости к поверхности, т. е. в плос-

кости, натянутой на векторы $\text{grad} G_1(x_0), \dots, \text{grad} G_h(x_0)$. Имеем, следовательно, при некоторых $\lambda_1, \dots, \lambda_h$

$$\text{grad}(F - \lambda_1 G_1 - \dots - \lambda_h G_h) = 0,$$

в чем и заключается метод неопределенных множителей Лагранжа.

Любопытно, однако, посмотреть, что написано об этом в курсах математического анализа. По существу, разумеется, делается то же самое, но без использования понятий касательной плоскости, нормали и градиента, которые, впрочем, все равно имеются во всех этих курсах. Так, у Г. М. Фихтенгольца¹ изложение занимает четыре страницы и в заключении объявляет, что метод Лагранжа «следует рассматривать лишь как указание, облегчающее запоминание». Беда в том, что у Г. М. Фихтенгольца условный экстремум в первом томе, а градиент — в третьем. Но в крайнем случае, если нельзя перенести в первый том градиент, то можно было бы подождать с условным экстремумом до третьего тома. В курсе А. Я. Хинчина² условный экстремум излагается на шести страницах.

Вряд ли эти замечания повлияют на изложение вопроса об условном экстремуме в курсах математического анализа. Трудно поколебать сложившуюся традицию. Но что касается теории вероятностей, то здесь традиции преподавания только лишь складываются. Возможно, что предлагаемые здесь методические замечания в какой то мере повлияют на эти традиции (автор, разумеется, не уверен, что и эта книга не содержит какого-нибудь абсурда).

Вторая цель этих замечаний — обсудить возможности приложения теории вероятностей. Здесь есть две опасности.

Во-первых, применениями теории вероятностей может заниматься естествоиспытатель, недостаточно квалифицированный в математическом отношении. Имя возникающим в таком случае ошибкам — легион, и примеры приводить просто неинтересно.

Во-вторых, вполне квалифицированный математик может быть, к сожалению, лишен здравого смысла естествоиспытателя и предлагать применять теорию вероятностей во всех случаях жизни, в том числе и в тех, когда она неприменима. Например, Дж. Кемени, Дж. Снелл и Дж. Томпсон³ объявляют предметом теории вероятностей, в частности, решение вопроса о том, будет ли сегодня дождь. На этом вопросе следует остановиться подробнее.

Прежде всего, если речь идет о том, будет ли дождь в определенный день, скажем 7 мая, какого-то достаточно далекого от настоящего момента года, то это есть типичный вопрос теории вероятностей. Лучшее, что здесь можно сделать, — это посмотреть

¹ См. Г. М. Фихтенголец. Курс дифференциального и интегрального исчисления, т. I, М., 1966, стр. 467—471.

² См. А. Я. Хинчин. Краткий курс математического анализа. М., 1957, стр. 453—458.

³ См. Дж. Кемени, Дж. Снелл, Дж. Томпсон. Введение в конечную математику. М., ИЛ, 1963, стр. 162.

по материалам прошлых метеорологических наблюдений, как часто бывает дождь 7 мая, и полученную частоту считать примерно равной вероятности дождя. Однако, по мере приближения условленного срока, такой ответ будет все менее нас удовлетворять. Например, когда 7 мая указанного года наступит и пройдет, мы будем знать ответ на наш вопрос совершенно точно.

Рассмотрим теперь случай, когда 5 мая мы спрашиваем о том, будет ли послезавтра дождь. Чтобы сделать этот вопрос предметом теории вероятностей, мы должны указать соответствующую статистически однородную совокупность. Можно было бы выбрать по материалам прошедших лет не все годы, а лишь те годы, в которые 5 мая была такая же синоптическая обстановка, как и в данном году, и вычислить частоту дождя 7 мая по этим годам. Но что такое «такая же синоптическая обстановка»? Если под этим понимать, скажем, полное совпадение синоптических карт, то мы наверняка ни одного подходящего года в прошлом не найдем. Если же не требовать полного совпадения, то вопрос перестает решаться объективно. Указанная трудность является существенной: несмотря на развитие теории вероятностей, прогнозы погоды продолжают оставаться плохими.

Что касается пропаганды теоретико-вероятностных и статистических методов, то когда-то, в давно прошедшие времена, надо было отстаивать само их право на существование. Однако сейчас их польза никем не оспаривается, и основное внимание следует обратить на ограничения, при которых эти методы дают в самом деле надежные результаты. По современным представлениям область применения теоретико-вероятностных методов ограничена явлениями, которым присуща статистическая устойчивость. Однако проверка статистической устойчивости трудна и всегда неполна; к тому же она часто дает отрицательный вывод. В результате в целых областях знания, например в геологии, нормой стал такой подход, при котором статистическая устойчивость вовсе не проверяется, что неизбежно приводит к серьезным ошибкам. К тому же пропаганда кибернетики, предпринятая нашими ведущими учеными, дала (в некоторых случаях) несколько неожиданный результат: теперь считается, что только машина (а не человек) способна получать объективные научные результаты.

В таких обстоятельствах долг преподавателя теории вероятностей вновь и вновь пропагандировать ту старую истину, которую еще Петр I пытался (безуспешно) внушить русским купцам: что торговать надо честно, без обмана, так как в конечном счете это для самих же себя выгодно.

Методически (но, конечно, не научно) здесь было бы полезно ввести терминологическое разграничение. Если исход эксперимента не вполне однозначно определяется его условиями, то можно было бы говорить о наличии «неопределенности», например завтра может быть или не быть дождь. Если эта неопределенность обладает свойством статистической устойчивости, то тогда можно сказать,

что имеется «случайность», т. е. случайное событие или случайный эксперимент. В научном смысле, разумеется, невозможно сколь-нибудь продвинуться только за счет введения нового термина: основной вопрос только переформулируется — вместо того, чтобы спросить, есть ли статистическая устойчивость, мы должны будем спросить, есть ли «случайность» или только «неопределенность».

Однако такое терминологическое разграничение позволяет использовать удачные в литературном отношении обороты, например: «Ваш фактический материал содержит неопределенность, очевидно потому, что плохо делались эксперименты. Однако в нем нет случайности в том смысле, как это слово понимается в теории вероятностей».

З а м е ч а н и я к § 1

К п. 1.1. Не так просто привести пример случайного эксперимента с известной вероятностью того или иного исхода. Пожалуй, только бросание монеты никогда не бралось под сомнение. При бросании кости вряд ли можно сомневаться в статистической устойчивости, однако в некоторых экспериментах с несомненностью обнаруживалось, что вероятности выпадения отдельных граней не равны $1/6$.

То же относится к рулетке. На некоторых аппаратах отдельные номера выпадают чаще (или реже), чем им полагалось бы при идеальной рулетке. На этом основан способ наживы за счет владельцев рулетки, который, по крайней мере однажды, был успешно применен. Кстати, для обнаружения отклонений рулетки от идеальной весьма полезны и необходимы методы теории вероятностей. Неверно, таким образом, что теория вероятностей учит, что играющий в азартные игры неизбежно разоряется — это справедливо только в идеальной модели азартной игры. К сожалению (для игроков), вероятностные методы одинаково доступны также и для хозяев рулетки. Последние спасаются тем, что часто меняют аппараты.

Переходя к более серьезным примерам, рассмотрим вероятность рождения ребенка мужского пола.

В рамках генетики было бы понятно, если бы эта вероятность равнялась $1/2$. Но на самом деле она несколько больше, причем обнаруживаются отклонения от статистической однородности. Например, достоверно известно, что после крупных войн частота рождения мальчиков увеличивается. Таким образом, эта вероятность точно не известна, а оценивается по частоте.

Блестящим примером из биологии являются опыты Менделя по расщеплению признаков, в которых вероятности появления разных признаков у гороха практически не отличались от вычисленных теоретически (см. задачник Л. Д. Мешалкина [8], стр. 58). Следует заметить, что теория вероятностей приносит, пожалуй,

максимальную научную пользу не тогда, когда обнаруживается полное согласие с ее схемами, а тогда, когда обнаруживаются отклонения. Так, колебания частоты рождения мальчиков по крайней мере ставят вопрос о механизмах регуляции численности мужских и женских особей биологического вида.

Частотная интерпретация понятия вероятности и понятие статистической устойчивости принадлежат Р. Мизесу. Р. Мизес оказал большое влияние на развитие теории вероятностей, в частности весьма наглядно продемонстрировал несовершенство ее старого языка. Например, в классической теории вероятностей имеется определение: «два события называются несовместимыми, если они не могут произойти оба вместе» и теорема: «вероятность суммы двух несовместных событий равна сумме вероятностей». Р. Мизес придумал следующий парадокс. Пусть некий теннисист может поехать на турнир либо в Москву, либо в Лондон, причем турниры там происходят одновременно. Вероятность того, что он займет первое место в Москве равна 0,9 (если, конечно, он туда поедет), а в Лондоне — 0,6. Чему равна вероятность того, что он займет где-либо первое место? Решение: согласно классической теории, события «выигрыш турнира в Москве» и «выигрыш турнира в Лондоне» несовместны, поэтому искомая вероятность есть $0,9 + 0,6 = 1,5$.

Несмотря на очевидную нелепость этого рассуждения, в старой теории вероятностей не было ничего, что бы его запрещало.

Таким образом, в настоящее время совершенно ясно, что изложение теории вероятностей на классическом языке является анахронизмом.

Какие же имеются альтернативы?

Первая альтернатива — язык теории множеств и теории меры в той форме, которая ему придана А. Н. Колмогоровым в знаменитой «аксиоматике Колмогорова». Этот язык пользуется почти универсальным признанием как в нашей стране, так и за рубежом. В частности, на нем написана данная книга. В рамках этого языка только что приведенный парадокс с теннисистом решается просто: вероятности 0,9 и 0,6 относятся к разным пространствам элементарных событий, так что не имеет смысла теорема сложения вероятностей. Столь же просто разрешаются другие известные парадоксы, например, парадокс Бертрана (см. учебник Б. В. Гнеденко [8]). Благодаря этому языку необычайно усилилось развитие теории вероятностей, о чем, однако, автор не берется говорить подробнее. Но является ли этот язык вполне безупречным? Это не так по ряду причин.

Во-первых, хорошо известно, что понятие множества, в таком виде, в котором оно используется здесь, ведет к парадоксам. Преодоление парадоксов языка математики всегда совершалось лишь за счет существенного научного развития (можно напомнить, например, преодоление парадоксов теории рядов за счет изобретения теории пределов и понятия сходимости). Именно поэтому математик относится с недоверием к попыткам «формализации» системы

понятий какой-либо естественной науки, так как при наличии существенного научного прогресса «формализация» придет сама собой, а без такого процесса останется чисто схоластическим упражнением. Для теории множеств еще нет удовлетворительного во всех отношениях способа преодоления парадоксов.

Имеется, правда, эмпирическое правило для избежания парадоксов, которое заключается в том, что лучше не говорить с слишком больших или слишком расплывчато заданных множествах. Например, запрещается говорить о «множестве всех множеств» или о «множестве всех множеств, для задания которых требуется не более 100 слов русского языка». Этому правилу заведомо подчиняются все множества, рассматриваемые в теории вероятностей в рамках колмогоровской аксиоматики, так что парадоксы теории множеств не особенно беспокоят нас в теории вероятностей.

Во-вторых, в результате развития теории случайных процессов обнаружилось, что в рамках колмогоровской аксиоматики приходится уделять слишком много внимания различным математическим затруднениям. По крайней мере в преподавании курса случайных процессов получается так, что на преодоление этих затруднений уходит столько времени и сил, что это наносит ущерб изучению сопоставимых с действительностью моделей.

В-третьих, в рамках этой аксиоматики ничего не говорится о том, как узнать, приложима ли вероятностная модель к данному конкретному явлению.

Вторая альтернатива классическому языку теории вероятностей — язык «теории коллективов» Р. Мизеса. Этот язык имеет гораздо меньшее распространение, чем язык аксиоматики Колмогорова. По существу он был, вероятно, отчасти вытеснен последним (подчеркнем, что, на наш взгляд, здесь следует говорить именно о взаимодействии языков, так как конкретное содержание и практические приложения теории вероятностей, по Мизесу и по Колмогорову, в сущности совпадают). Не излагается этот язык и в настоящей книге (см., например, [19]). В таком случае было бы недобросовестно давать здесь его критику, в чем, впрочем, и нет никакой необходимости, так как вряд ли кто-нибудь хочет сейчас преподавать теорию вероятностей по Мизесу. Однако любопытно обрисовать изменение общего тона критики теории Р. Мизеса за истекшие 40 лет. Первоначальный вариант теории Мизеса казался неприемлемым для математика¹ по двум причинам: он был логически противоречив (это противоречие отмечено в § 1 настоящей книги при обсуждении вопроса о проверке устойчивости частот в различных сериях опытов) и не содержал математической аксио-

¹ Выражение этих ранних взглядов на теорию Р. Мизеса см. в статьях А. Я. Хинчина «Учение Мизеса о вероятностях и принципы физической статистики». «Успехи физических наук», 1929, т. IX, вып. 2; «Частотная теория Р. Мизеса и современные идеи теории вероятностей». «Вопросы философии», 1961, № 1, стр. 91—102; № 2, стр. 77—89 (последняя статья написана около 1944—1946 гг.).

матики. Изучение реакции Р. Мизеса на математическую критику представляет известные трудности, так как содержащая наиболее полное изложение его теории книга [20] написана—пусть аккуратно и добросовестно — но все же не самим Р. Мизесом, а Хильдой Гейрингер. Достоверно известно, однако, что Р. Мизес не считал отсутствие аксиоматики недостатком. Любопытно, что в связи с выяснившимися недостатками формальных аксиоматических теорий (например, теоремой Геделя о неполноте) многие математики постепенно перестали считать наличие аксиоматики большим достоинством. Впрочем, последователи Р. Мизеса создали аксиоматику его теории, и ее можно читать в виде приложения к упомянутой выше книге [20].

Далее, в работах А. Н. Колмогорова и его учеников намечен и способ преодоления логического противоречия теории Р. Мизеса (естественно, на совсем другом научном уровне, в частности, с использованием теории алгоритмов).

С другой стороны, за истекшие 40 лет теория вероятностей развивалась на основе аксиоматики Колмогорова. В частности, возникла теория случайных процессов. Естественно, перед теорией Мизеса встал вопрос о включении нового материала. По-видимому, это возможно, и Х. Гейрингер даже бралась это сделать, но издатель цитированной книги [20] отказался напечатать соответствующий материал из-за слишком большого объема книги.

В целом, теперешнее отношение специалистов к языку теории Р. Мизеса можно сравнить с отношением к мертвому языку, на котором почему-то никто говорить не хочет, но — при соответствующих поправках и переделках — вполне можно было бы сказать все то, что говорят на живом языке.

В то же время несомненно, что идеи Р. Мизеса органически входят в современную теорию вероятностей, а дальнейшая разработка их оказывает влияние на фундаментальные представления этой науки.

В частности, условия практической применимости теории вероятностей сейчас трактуются по Р. Мизесу. Они изложены в § 1. Методически надо иметь в виду, что этот материал не может быть усвоен студентами, впервые знакомящимися с теорией вероятностей. К понятию статистической устойчивости необходимо возвращаться на протяжении всего курса — при интерпретации случайной величины как результата измерения, подверженного случайным ошибкам: при применении теоремы Муавра—Лапласа для проверки гипотезы о равенстве вероятностей успеха в двух сериях испытаний Бернулли; при изложении метода наименьших квадратов и т. д.

К. п. 1.2 и 1.3. Целесообразно начинать изложение теории вероятностей с дискретного случая, т. е. со случая конечного или счетного Ω . При этом каждому подмножеству A множества Ω может быть приписана вероятность в соответствии с определением 1.4. Свойство $P(A+B) = P(A) + P(B)$ для непересекающихся A и

B оказывается теоремой. Это является особенностью принятого способа изложения. Принятый в п. 1.2 набор определений и аксиом представляется наиболее экономным.

Подчеркивание того обстоятельства, что при решении задач по теории вероятностей прежде всего нужно сделать перевод на язык элементарных событий, чрезвычайно облегчает студентам понимание того, чего от них хотят. На упражнениях по теории вероятностей нужно сделать несколько задач на такой перевод (рекомендуется задачник Л. Д. Мешалкина [18]). Однако преподаватель теории вероятностей должен помнить о том, что язык элементарных событий изучают в основном для того, чтобы можно было всегда избежать парадоксов, а впоследствии обычно мыслят сразу интересующими нас событиями.

Замечания к § 2

К п. 2.1. При введении определений и аксиом в § 1 не было подчеркнуто, что они являются математической формулировкой свойств понятия вероятности, которых естественно ждать, если иметь в виду частотную интерпретацию этого понятия. Это важное обстоятельство отмечается лишь в начале § 2, что сделано по чисто психологическим соображениям: первый параграф и так перегружен материалами естественнонаучного содержания. В то же время важно соблюдать равновесие между чисто математическим и естественнонаучным материалом, так как при пренебрежении первым изложение воспринимается студентами как поверхностное, а пренебрежение вторым наносит ущерб правильной оценке роли теории вероятностей в науке.

Само определение условной вероятности, данное в § 2, формально говоря, не соответствует трактовке условной вероятности в аксиоматике Колмогорова. В применении к дискретному случаю, если строго держаться этой аксиоматики, нужно было бы рассмотреть вероятности $P(A/B)$ и $P(A/\bar{B})$, задаваемые определением 2.1, а под условной вероятностью понимать *случайную величину*, определенную на Ω и равную $P(A/B)$ для $\omega \in B$ и $P(A/\bar{B})$ для $\omega \notin B$. Однако различие между одной вероятностью $P(A/B)$ и двумя вероятностями: $P(A/B)$ и $P(A/\bar{B})$ в контексте § 2 воспринималось бы как чисто схоластическое (по крайней мере, автор не нашел способа его убедительно мотивировать). Реальным это различие можно сделать только рассматривая условные вероятности при условии, что задано значение некоторой случайной величины ξ . Соответствующее определение должно было бы выглядеть так: условной вероятностью $P(A/\xi)$ называется случайная величина, равная $P\{A | (\xi = a_i)\}$ для тех и только тех $\omega \in \Omega$, для которых $\xi(\omega) = a_i$ (здесь a_1, a_2, \dots — все возможные значения случайной величины ξ). В силу формулы полной вероятности (см. п. 2.4, § 2) имеем

$$P(A) = \sum_{a_i} P\{A | (\xi = a_i)\} P\{\xi = a_i\},$$

что можно записать в виде

$$P(A) = MP(A|\xi), \quad (2.1')$$

где M есть знак математического ожидания (примененного к случайной величине $P(A|\xi)$).

При таком способе изложения следовало бы подчеркнуть, что сами значения $\{a_i\}$ случайной величины ξ в определение условной вероятности $P(A|\xi)$ никак не входят: важны лишь множества $(\xi = a_i)$. Отсюда можно было бы мотивировать введение условной вероятности $P(A/\mathfrak{B})$, где \mathfrak{B} — разбиение пространства Ω , т. е. набор множеств $\{B_1, B_2, \dots, B_n, \dots\}$ таких, что $B_i B_j = \emptyset$ при $i \neq j$ и $B_1 + B_2 + \dots + B_n + \dots = \Omega$.

Далее следовало бы ввести понятие условного математического ожидания $M(\eta/\mathfrak{B})$, а в частном случае $M(\eta/\xi)$, где η — любая дискретная случайная величина, и вывести формулу

$$M\eta = M\{M(\eta|\mathfrak{B})\}, \quad (2.2')$$

являющуюся обобщением формулы (2.1').

Почему указанный путь не реализован в настоящей книге? Дело в том, что формулы (2.1') и (2.2') по-настоящему важны не в дискретном случае, а в случае произвольного пространства эле-

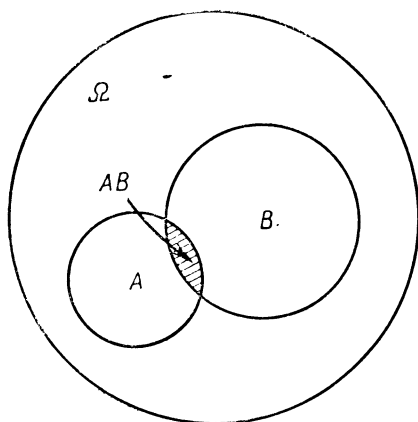


Рис. 2.1'

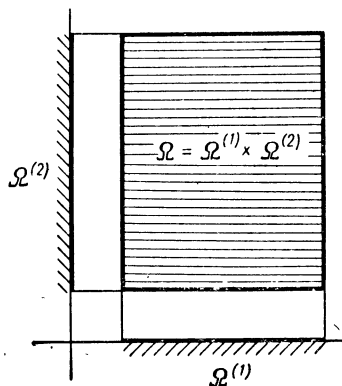


Рис. 2.2'

ментарных событий. Однако тогда изложение колмогоровской теории условных математических ожиданий предполагает владение теорией меры (в частности, знание теоремы Радона-Никодима), а кроме того, требует очень много времени (если, конечно, препода-

ватель добывается неформального владения этой теорией, т. е. слияния интуитивных представлений об условной вероятности с формальными определениями). Поэтому автор считает, что от общей теории условных вероятностей в элементарном курсе следует отказаться. Но если принять эту точку зрения, то, по мнению автора, формулы (2.1') и (2.2') должны показаться недостаточно

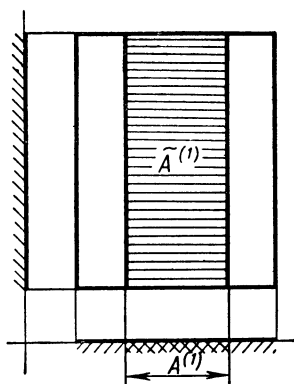


Рис. 2.3'

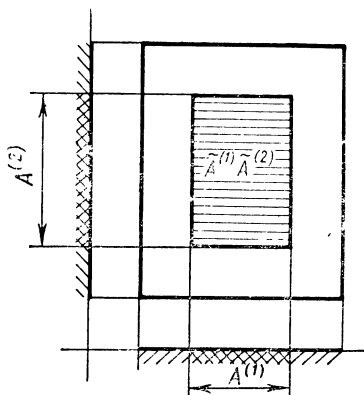


Рис. 2.4'

содержательными, хотя их, пожалуй, уже никто не назовет чистой схоластикой.

Два других соображения против расширенной, по сравнению с принятой в настоящей книге, трактовки понятия условной вероятности состоят в следующем.

Во-первых, наиболее реальные, с точки зрения естествоиспытателя, вероятностные модели связаны с независимыми событиями и случайными величинами, т. е. с тем случаем, когда условные вероятности равны безусловным. Это объясняется тем, что гораздо проще найти из опыта единственную вероятность $P(A)$, чем случайную величину $P(A/\xi)$, а кроме того, говоря о вероятности $P(A/\xi)$, мы лишним раз рискуем статистической устойчивостью: так как для того, чтобы эта условная вероятность имела смысл, нужно, чтобы частота наступления события A при условии, что $\xi = a_i$, была статистически устойчива при любом a_i . Таким образом, хотя многие после первого знакомства с понятием условной вероятности представляют себе мир в терминах этого понятия, так поступать чаще всего будет неправильно. Как известно, Декарт указывал, что можно придумать много разных способов, которыми господь бог — при желании — мог бы устроить мир, но никогда нельзя чисто умозрительно выбрать из них тот, которым он воспользовался на самом деле (чтобы это сделать, по Декарту, требуется эксперимент). Представление об условной ве-

роятности относится к числу таких мыслимых представлений, но только экспериментальная проверка устойчивости частот покажет, так ли устроено изучаемое явление на самом деле. Впрочем, пример Декарта наводит на грустные размышления: Декарт прилежно делал эксперименты, но это не мешало ему утверждать, что в сердце кровь обращается в пар и благодаря этому осуществляет кровообращение.

В частности, когда в элементарном курсе теории вероятностей из-за недостатка времени приходится выбирать между основными статистическими методами и теорией цепей Маркова, основанной на понятии условной вероятности, предпочтение следует отдать статистике. Бесспорно, что теория цепей Маркова имеет некоторые серьезные естественнонаучные применения, но бесспорно и то, что эта теория иногда применяется совершенно неверно. В качестве примера можно указать некоторые работы по «математической геологии», однако разбор их неуместен в настоящей книге.

Во-вторых, общая теория условных вероятностей связана, к сожалению, с рядом затруднений типа вопросов измеримости, что лишает ее того удобства, которое естествоиспытатель, разумеется, хочет видеть в методах исследования, которые дает ему математика.

Однако, по мнению автора, способ изложения, основанный на формулах (2.1') и (2.2'), обязателен в том случае, когда в дальнейшем предполагается изложение общей теории условных вероятностей или рассмотрение понятий энтропии и информации.

По поводу задач на понятие условной вероятности заметим, что до настоящего времени в советских и зарубежных задачниках попадаются совершенно бессмысленные задачи, вроде, например, следующей. «Охотник сидит в засаде и ждет медведя. Медведь может выскочить из-за первого куста с вероятностью 0,1, из-за второго куста — с вероятностью 0,2 и из-за третьего куста — с вероятностью 0,3. В первом случае охотник убивает его с вероятностью 0,5, во втором случае — с вероятностью 0,4, в третьем случае — с вероятностью 0,3. Найти вероятность того, что охотник убьет медведя».

В формулировке этой задачи весьма сомнительно предположение статистической устойчивости частот появления медведя из-за каждого куста, равно как и частот удачного выстрела охотника (без чего данные в задаче вероятности не имеют смысла). Но что можно гарантировать во всяком случае — это то, что эксперимент, обнаруживший статистическую устойчивость и давший приведенные значения вероятностей, никогда на самом деле не проводился. В целом задача наводит на мысль о всеобъемлющем характере понятия условной вероятности, применимость которого не требует экспериментальной проверки (ведь каждому ясно, что эксперимента не было). А поскольку изучающий теорию вероятностей знакомится с практическими примерами применения прежде всего из задач, то использование в преподавании задач, подобных приведенной выше, неизбежно создает неоправданные иллюзии. Многие задачи с

военным контекстом на понятие условной вероятности ничем не лучше «задачи про медведя».

К п. 2.2 и 2.3. Понятие независимости является фундаментальным для теории вероятностей. Сравнительно недавно, после того как колмогоровская аксиоматика дала возможность рассматривать теорию вероятностей как часть теории меры, одно время даже считалось, что отличительной чертой этой части теории меры является использование понятия независимости. В настоящее время независимость не господствует в теории вероятностей как в математической науке. В частности, в теории случайных процессов рассматриваются в основном зависимые события и величины. Однако почти в любой вероятностной модели, рассчитанной на применения, где-то, быть может довольно глубоко, спрятана независимость.

Как правило, независимые события наблюдаются в независимо друг от друга проводимых экспериментах. Соответствующая конструкция вероятностного пространства дана в п. 2.3. Обычно преподаватель рисует пространство Ω в виде подмножества (например, круга) на плоскости. На такой картинке независимые события выглядят вычурно: между вероятностями фигур A , B и AB должно выполняться соотношение (рис. 2.1')

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

Более естественная картинка получится, если использовать прямое произведение: на отрезке оси абсцисс изобразить пространство элементарных исходов $\Omega^{(1)}$ первого эксперимента, а на отрезке оси ординат — пространство $\Omega^{(2)}$, связанное со вторым экспериментом. Тогда $\Omega = \Omega^{(1)} \times \Omega^{(2)}$ есть показанный на рис. 2.2' прямоугольник. Если $A^{(1)}$ — событие, связанное с исходом первого эксперимента, то исход второго эксперимента может быть любым. Поэтому можно изображать $A^{(1)}$ также в виде прямоугольника $\tilde{A}^{(1)} = A^{(1)} \times \Omega^{(2)}$ (рис. 2.3'). Аналогично изображается событие $A^{(2)}$ в виде события $\tilde{A}^{(2)}$. Но тогда произведение $\tilde{A}^{(1)}\tilde{A}^{(2)}$ изображается в виде прямоугольника, являющегося пересечением $\tilde{A}^{(1)}$ и $\tilde{A}^{(2)}$. Равенство

$$P(\tilde{A}^{(1)}\tilde{A}^{(2)}) = P(\tilde{A}^{(1)})P(\tilde{A}^{(2)})$$

будет выполняться, если вероятность естественным образом интерпретировать как площадь (рис. 2.4').

Необходимо добиться, чтобы изучающий теорию вероятностей понимал полную естественность перехода от $A^{(1)}$ и $A^{(2)}$ к $\tilde{A}^{(1)}$ и $\tilde{A}^{(2)}$ в пространстве $\Omega = \Omega^{(1)} \times \Omega^{(2)}$.

Понятие независимости при решении задач по теории вероятностей используется для определения вероятностей на множестве элементарных событий, подобно другим «правилам перевода», упомянутым в § 1 и 2. Формально здесь редко делаются ошибки. Приведение примеров независимых событий не представляет труда: два или несколько случайных экспериментов, проводимых раз-

ными лицами в разных местах. Известную аккуратность надо соблюдать лишь при выборе самих экспериментов (чтобы они в самом деле были случайными в уже разъяснявшемся смысле).

Хуже обстоит дело с экспериментами, проводимыми одним и тем же лицом последовательно, например, с результатами последовательных измерений. Вообще говоря, здесь нет оснований ожидать независимости результатов различных экспериментов. Надо думать, что если эксперименты проводятся недостаточно тщательно, то независимости и не будет. В этом случае (плохих наблюдений) никакая статистическая обработка их не улучшит.

В том же случае, когда возможные причины колебаний результатов экспериментов тщательно устраняются, т. е. остаются лишь такие небольшие колебания, которые нельзя устранить без принципиального усовершенствования техники эксперимента, видимо, можно ожидать независимости результатов отдельных наблюдений. Таким образом, лишь при хороших экспериментах целесообразна статистическая обработка, которая может еще улучшить результаты. Бывают, однако, хорошие измерения, в которых проводимые через небольшие интервалы времени отсчеты дают зависимые результаты (например, в радиолокации). В таких случаях статистическая обработка существенно затрудняется.

К п. 2.4 и 2.5. Задачи на формулу полной вероятности, по традиции, обычно решаются без перехода к пространству элементарных событий. Это непоследовательно. Чтобы сохранить логическую состоятельность, можно применить следующий искусственный прием. Пусть в задаче требуется определить вероятность некоторого события D . Образует пространство элементарных событий из двух символов: D и \bar{D} . Введем $P(D)$ таким образом, чтобы оно не противоречило формуле полной вероятности, и положим

$$P(\bar{D}) = 1 - P(D).$$

Ответом на вопрос задачи является введенное по определению значение вероятности $P(D)$. Спасена и аксиоматическая модель. Этот способ перевода на язык пространства элементарных событий условий задач на формулу полной вероятности формально возможен. Он приведен в качестве первого способа решения задачи в примере 2.2. Более естественным является второй путь решения этой задачи. Однако при этом выпадает использование самой формулы полной вероятности, вместо которого производится подсчет числа элементов разных множеств.

В общем-то формула полной вероятности не имеет права на существование в курсе теории вероятностей в качестве самостоятельной теоремы. Это связано с тем, что она не только тривиальна с математической точки зрения, но и не дает ничего нового по сравнению с понятием условной вероятности для целей перевода условий задач на язык элементарных событий. Лишь традиционная привязанность к ней заставляет уделить ей место в лекцион-

ном курсе (всегда есть опасность, что кто-нибудь из преподавателей спросит ее на экзамене). В противном случае достаточно было бы разобрать ее на упражнениях.

Попытки описания научного прогресса с помощью формулы Байеса действительно делались. В применении к произвольным научным гипотезам это бессмысленно, хотя один крайний случай получается из формулы Байеса. Этот случай состоит в следующем. Пусть $P(A/H_1) > 0$, а $P(A/H_i) = 0$, $i = 2, 3, \dots$, т. е. событие A может произойти только в том случае, если верна гипотеза H_1 . Тогда, при любых апприорных вероятностях $P(H_1)$, \dots , $P(H_n)$, лишь бы только $P(H_1)$ была положительна (т. е. мы верили сколько-то в гипотезу H_1), имеем

$$P(H_1/A) = 1, P(H_i/A) = 0, i = 2, \dots, n.$$

Следовательно, если A произошло, то верна гипотеза H_1 — вывод весьма разумный, но очевидный и без использования формулы Байеса.

Во всяком случае, гипотезы H_1, H_2, \dots, H_n должны быть такими, чтобы условные вероятности $P(A/H_i)$ имели смысл, т. е. наблюдалась бы статистическая устойчивость. Это сильно ограничивает область применения формулы Байеса. Типичной задачей на эту формулу является следующая.

Пусть вынимается шар из урны, содержащей черные и белые шары, причем гипотеза H_i состоит в том, что отношение числа белых шаров в урне к числу всех шаров равно $\frac{i}{n}$, $i = 1, 2, \dots, n$.

Предположим, что с одинаковой вероятностью нам могли дать урну любого состава, т. е. что $P(H_i) = \frac{1}{n}$ для любого i . Спрашивается, чему равна вероятность H_i при условии, что вынутый шар оказался белым.

На тривиальном расчете мы не останавливаемся. Заметим, что если операцию вынимания шара повторять много раз (каждый раз возвращая вынутый шар в урну и хорошо перемешивая шары), то мы сможем почти точно выяснить долю белых шаров в урне. На языке апостериорных вероятностей это утверждение выглядит так: при вычислении апостериорной вероятности гипотезы H_i при условии, что задан результат k выниманий шара, получится, что при $k \rightarrow \infty$ для некоторого i эта апостериорная вероятность стремится к 1, а для остальных к нулю. (Мы не останавливаемся на точной формулировке и доказательстве последнего утверждения.) Этот вывод не зависит от апприорных вероятностей $P(H_i)$, лишь бы все они были положительны. На получении выводов, мало зависящих от апприорных вероятностей, основан байесовский подход в математической статистике. Этот подход имеет сравнительно немного сторонников, но в последнее время его популярность несколько возросла.

Применение формулы Байеса в задачах диагностики вызвано тем обстоятельством, что медицинская информация, касающаяся больного, в настоящее время слишком обширна.

Указанная в § 2 трудность, стоящая на пути формального применения теоремы Байеса, заставляет искать какие-то дополнительные допущения, при которых все-таки можно определить вероятности $P(A/H_i)$ по ограниченному статистическому материалу. Поскольку приходится делать дополнительные предположения, на результат всей процедуры не следует смотреть как на математически доказанный факт: это лишь некоторый вспомогательный способ суммировать для врача имеющуюся информацию. Автор книги не берется судить о достигнутых здесь результатах. Вообще оценка результатов математико-статистического исследования — очень трудное дело.

Формула Байеса может применяться также для решения задачи машинного узнавания букв. Роль комплекса симптомов в этом случае играет совокупность некоторых отличительных признаков буквы, которые легко могут быть выражены в виде, пригодном для введения в машину. Можно говорить о распознавании других объектов. В таком случае речь обычно идет о «распознавании образов». Каким образом человек распознает образы, т. е. узнает знакомые буквы, предметы, лица, в настоящее время совершенно неизвестно. Существующие математические теории строятся на основе использования понятия группы преобразований, но на самом деле преобразования, не меняющие восприятия образа, групповыми свойствами явно не обладают. О формуле Байеса можно в данной задаче сказать только одно — ее использования очевидным образом недостаточно для успешного распознавания образов.

Замечания к § 3

Курс математического анализа полон неожиданностей. Пусть, например,

$$\sum_{i=1}^{\infty} a_i = a_1 + a_2 + a_3 + \dots$$

абсолютно сходящийся ряд. Верно ли, что

$$\sum_{i=1}^{\infty} a_i = \sum_{k=1}^{\infty} a_{2k} + \sum_{k=1}^{\infty} a_{2k-1}, \quad (3.1')$$

т. е. можно ли сначала просуммировать члены ряда с четными номерами и к полученной сумме прибавить сумму членов с нечетными номерами? Пожалуй, в этом трудно сомневаться. В обоснование этого ссылаются обычно на теорему о перестановке членов ряда из курса математического анализа. Однако требуемое утверждение не следует из этой теоремы, так как в результате пере-

становки должен получиться снова ряд, а не сумма двух рядов вида (3.1'). Однако доказательство этой теоремы легко приспособить и для доказательства (3.1').

Пусть $\Omega = \{\omega\}$ — счетное множество элементарных событий, $P(\omega)$ — вероятность ω . Трудно удержаться от написания равенства

$$\sum_{\omega \in \Omega} P(\omega) = 1.$$

Но что означает символ $\sum_{\omega \in \Omega} P(\omega)$? На этот вопрос можно было бы

ответить так: пусть имеется взаимно однозначное отображение $i \rightarrow \omega_i$ натурального ряда на Ω . Будем понимать под суммой $\sum_{\omega \in \Omega} F(\omega)$, где $F(\omega)$ — некоторая функция от ω , сумму ряда

$\sum_{i=1}^{\infty} F(\omega_i)$. В силу упоминавшейся теоремы анализа, такое опре-

деление дает результат, не зависящий от выбора отображения $i \rightarrow \omega_i$. Пусть теперь $A_1, A_2, \dots, A_n, \dots$ — непересекающиеся подмножества Ω , причем $\Omega = A_1 + A_2 + \dots + A_n + \dots$. Тогда хотелось бы, чтобы (в случае абсолютной сходимости)

$$\sum_{\omega \in \Omega} F(\omega) = \sum_{\omega \in A_1} F(\omega) + \sum_{\omega \in A_2} F(\omega) + \dots + \sum_{\omega \in A_n} F(\omega) + \dots \quad (3.2')$$

Но это не вытекает из указанной теоремы анализа, хотя ее доказательство в сущности проходит. Рассмотрим еще пример, когда

$$A_1 + A_2 + \dots + A_n + \dots = \Omega \setminus \omega_0,$$

где ω_0 — некоторый элемент Ω . Хотелось бы, чтобы

$$\sum_{\omega \in \Omega} F(\omega) = \sum_{i=1}^{\infty} \left\{ \sum_{\omega \in A_i} F(\omega) \right\} + F(\omega_0). \quad (3.3')$$

Рассмотренные примеры доказывают необходимость иметь такое определение суммы $\sum_{\omega \in \Omega} F(\omega)$, которое охватывало бы как част-

ные случаи равенства (3.2') и (3.3'). По-видимому, нет ничего более простого, чем использование понятия трансфинитных чисел (счетной мощности). Если A — любое счетное вполне упорядоченное множество и $\alpha \rightarrow \omega_\alpha$ есть взаимно однозначное отображение A на Ω , то $\sum_{\alpha \in A} F(\omega_\alpha)$ определяется как предел по Шатуновскому

(см., например, [2 3]), и нужно доказывать, что в случае абсолютной сходимости

$$\sum_{\alpha \in A} F(\omega_\alpha) = \sum_{\beta \in B} F(\omega_\beta),$$

где B — другое вполне упорядоченное множество.

Доказательство все равно остается тем же самым. Так мы видим, что требуемое утверждение содержится в курсе анализа по существу, но не по форме. Поэтому преподаватель теории вероятностей не слишком погрешит против совести, если не будет обращаться на это внимания студентов, как это сделано в § 3 настоящей книги. Впрочем, для целей теории вероятностей достаточно, если угодно, обосновать равенство (3.2').

К п. 3.1. Все случайные величины, рассматриваемые в одном контексте, определены на одном множестве элементарных событий Ω . В противном случае нельзя избежать недоразумений. В частности, будет непонятно, что такое сумма случайных величин. Если случайная величина ξ_1 интерпретируется как результат одного опыта с пространством элементарных исходов $\Omega^{(1)}$, а ξ_2 — как результат другого опыта, связанного с пространством $\Omega^{(2)}$, то следует ввести пространство $\Omega^{(1)} \times \Omega^{(2)}$, на котором и рассматривать величины ξ_1 и ξ_2 .

К п. 3.2. Определение математического ожидания отличается от обычно принятого в дискретном случае. Обычно в качестве определения математического ожидания $M\xi$ случайной величины ξ , принимающей значения $a_1, a_2, \dots, a_n, \dots$ с вероятностями $p_1, p_2, \dots, p_n, \dots$ берется равенство

$$M\xi = \sum_{a_i} a_i p_i.$$

Кстати, обозначать пределы суммирования в последнем равенстве в виде $\sum_{i=1}^{\infty} a_i p_i$ неудобно, так как множество значений $\{a_i\}$ может быть конечным. При необходимости обозначить явно пределы суммирования мы предпочтем запись $\sum_{a_i} a_i p_i$.

Чем вызвано определение $M\xi$ в виде

$$M\xi = \sum_{\omega \in \Omega} \xi(\omega) P(\omega)?$$

При таком определении гораздо проще доказывать свойства математического ожидания. Рассмотрим, например, теорему

$$M(\xi + \eta) = M\xi + M\eta.$$

Обычно она доказывается следующим образом. Пусть

$$\begin{pmatrix} a_1, a_2, \dots, a_n, \dots \\ p_1, p_2, \dots, p_n, \dots \end{pmatrix}, \quad \begin{pmatrix} b_1, b_2, \dots, b_m, \dots \\ q_1, q_2, \dots, q_m, \dots \end{pmatrix}$$

распределения величин ξ и η . Обозначим через p_{ij} вероятность $P\{\xi=a_i, \eta=b_j\}$. Если значениями величин ξ и η являются числа a_i и b_j , то значениями суммы $\xi+\eta$ являются числа a_i+b_j , причем $P\{\xi+\eta=a_i+b_j\}=p_{ij}$. Тогда имеем

$$\begin{aligned} M(\xi + \eta) &= \sum_{a_i, b_j} (a_i + b_j) p_{ij} = \sum_{a_i, b_j} a_i p_{ij} + \sum_{a_i, b_j} b_j p_{ij} = \\ &= \sum_{a_i} a_i p_i + \sum_{b_j} b_j q_j = M\xi + M\eta. \end{aligned}$$

Доказательство это вообще неверно. Дело в том, что числа $\{a_1, a_2, \dots\}$ (так же, как числа $\{b_1, b_2, \dots\}$) между собой различны, иначе не имеет смысла вероятность $P\{\xi=a_i\}$, $P\{\eta=b_j\}$. Однако между числами a_i+b_j могут быть равные. Следовательно, либо не имеет смысла говорить о вероятности $P\{\xi+\eta=a_i+b_j\}$, либо во всяком случае неверно, что эта вероятность есть p_{ij} . Кроме того, из-за уже отмечавшегося несовершенства традиционного языка теории рядов в анализе не вполне понятно, на какие теоремы анализа следует ссылаться для обоснования выкладок с рядами. Впрочем, последнее обстоятельство сравнительно маловажно. Но первое затруднение существенно. На него мы попадаем вновь, желая доказать формулу $Mf(\xi) = \sum f(a_i) p_i$ (теорема 3.3). Неверно ведь, что $P\{f(\xi) = f(a_i)\} = p_i$. На самом деле

$$P\{f(\xi) = f(a_i)\} = \sum_{a_j: f(a_j)=f(a_i)} p_j.$$

Теорему 3.3 нужно выводить из последнего равенства следующим образом:

$$Mf(\xi) = \sum_{f(a_i)} f(a_i) \left\{ \sum_{a_j: f(a_j)=f(a_i)} p_j \right\} = \sum_{a_i} f(a_i) p_i,$$

но понимание правильности этого вывода требует некоторых усилий от студента (и, конечно, снова нельзя избежать формальных затруднений с рядами).

Та же история повторяется при выводе теоремы $M\xi\eta = M\xi M\eta$ для независимых случайных величин. Здесь вновь обычно утверждается, что

$$P\{\xi\eta = a_i b_j\} = P\{\xi = a_i, \eta = b_j\}.$$

Желание избавиться от серии этих затруднений привело к изменению определения математического ожидания по сравнению с обычно принятым (определение 3.2).

Отметим другую возможность. Сначала надо получить формулу

$$Mf(\xi) = \sum_{a_i} f(a_i) p_i \quad (3.4')$$

для того случая, когда величина ξ принимает векторные значения, т. е. является набором случайных величин. (Только что упомянутый вывод теоремы 3.3, разумеется, одинаково применим к векторным величинам.) Итак, пусть получена формула (3.4') для вектора $\xi = (\xi_1, \xi_2)$. Положим сначала

$$f_1(x) = x_1.$$

Тогда $f_1(\xi_1, \xi_2) = \xi_1$ и

$$Mf_1(\xi_1, \xi_2) = M\xi_1 = \sum_{a_i} f_1(a_i) p_i,$$

где под a_i понимаются возможные значения вектора $\xi = (\xi_1, \xi_2)$, $p_i = P\{\xi = (\xi_1, \xi_2 = a_i)\}$.

Аналогично, полагая $f_2(x) = x_2$, имеем

$$Mf_2(\xi_1, \xi_2) = M\xi_2 = \sum_{a_i} f_2(a_i) p_i.$$

Наконец, полагая

$$f_3(x) = x_1 + x_2 = f_1(x) + f_2(x),$$

получаем

$$\begin{aligned} Mf_3(\xi) &= M(\xi_1 + \xi_2) = \sum_{a_i} f_3(a_i) p_i = \sum_{a_i} (f_1(a_i) + f_2(a_i)) p_i = \\ &= \sum_{a_i} f_1(a_i) p_i + \sum_{a_i} f_2(a_i) p_i = M\xi_1 + M\xi_2. \end{aligned}$$

Пусть теперь значениями случайной величины ξ являются числа $c_1, c_2, \dots, c_n, \dots$, значениями случайной величины η — числа $d_1, d_2, \dots, d_m, \dots$ и величины ξ и η независимы. Рассмотрим векторную случайную величину $\zeta = (\xi, \eta)$, значениями которой являются векторы (c_i, d_j) , причем

$$P\{\zeta = (c_i, d_j)\} = P\{\xi = c_i, \eta = d_j\} = P\{\xi = c_i\} P\{\eta = d_j\}.$$

Полагая $f_4(x) = x_1 x_2$, имеем

$$\begin{aligned} Mf_4(\zeta) &= M\xi\eta = \sum_{c_i, d_j} f_4(c_i, d_j) P\{\zeta = (c_i, d_j)\} = \\ &= \sum_{c_i, d_j} c_i d_j P\{\xi = c_i\} P\{\eta = d_j\} = \\ &= \left(\sum_{c_i} c_i P\{\xi = c_i\} \right) \left(\sum_{d_j} d_j P\{\eta = d_j\} \right) = M\xi M\eta. \end{aligned}$$

Путь, принятый в § 3, представляется несколько менее формальным, чем только что изложенный. Там все основано на раз-

биениях множества Ω на различные части. Эти части указываются явно, и происходит сначала суммирование в пределах одной части, а затем — суммирование по всем получившимся частям.

Только что изложенный способ фактически также основан на разбиении на части некоторых сумм. Однако эти части являются множествами, на которых функции f_1, f_2, f_3, f_4 принимают постоянные значения, так что выделение частей осуществляется заданием функций f_i , что может быть несколько труднее при первом знакомстве с теорией вероятностей. Зато все свойства математического ожидания выступают в качестве следствий единственной формулы (3.4').

Применения понятия математического ожидания весьма многообразны. Мы остановимся коротко на теории игр и статистических решений, которая в последние годы стала довольно популярной. Идейное содержание теории игр по существу понятно из примера 3.1. Там описано такое поведение игроков, при котором каждый из них выбирает с некоторыми вероятностями доступные ему способы поведения — «стратегии», стараясь максимизировать математическое ожидание своего выигрыша. Важнейшей теоремой является теорема о существовании цены игры (в примере 3.1 цена игры есть $\frac{5}{12}$ коп.), т. е. такого числа c , что первый игрок, выбирая свое поведение, может сделать свой средний выигрыш равным c , что бы ни делал второй игрок, и в то же время второй игрок, выбирая свое поведение, может добиться, чтобы выигрыш первого игрока был не более c , что бы первый игрок ни делал. В теории игр предполагается, что оба игрока будут поступать именно таким образом, т. е. будут стараться получить то, что им гарантируется, и не стремиться спровоцировать противника на отказ от оптимально (в указанном только что смысле) выбранных вероятностей различных стратегий.

Надо сказать, что при этом всякое удовольствие от игры должно быть потеряно, так что психологически естественно именно спровоцировать противника (либо уж вовсе не играть). Однако в этом гораздо более сложном случае не известно содержательных математических результатов. Кроме того, после изложения центральной предельной теоремы пример 3.1 будет развит дальше, причем будет показано, что необходимо учитывать и дисперсию выигрыша. Получение более или менее гарантированного выигрыша в один рубль требует такого числа повторений рассмотренной в примере 3.1 игры, что лучше этот рубль заработать каким-либо другим путем. Между тем, в теории игр обычно не учитывается дисперсия.

Мы не говорим уже подробно о том, что наиболее интересные игры вообще не могут быть решены (т. е. найдены цена игры и оптимальное поведение игроков) даже с применением машинных вычислений.

Таким образом, модель поведения игроков, принятая в теории игр, не имеет всеобъемлющего характера, хотя и может, разумеется, быть правильной в отдельных частных случаях. Популярность

теории игр, очевидно, связана с тем обстоятельством, что многие судят о сфере приложений математической теории по ее названию, а не по содержанию. Название же в математике есть вещь довольно произвольная, а иногда даже обязанная случайной ассоциации (достаточно напомнить алгебраические термины: «кольца», «поля», «идеалы»).

С идеологией теории игр связана модернизация байесовского подхода математической статистике, называемая теорией статистических решений. Коротко суть подхода, здесь принятого, можно объяснить следующим образом. Рассмотрим задачу с урнами из замечаний к предыдущему параграфу (к пунктам 2.4 и 2.5). Гипотеза H_i состояла в том, что доля белых шаров в урне равна $\frac{i}{n}$. Оценка апостериорных вероятностей гипотез при условии, что вынутый из урны шар оказался белым, затруднялась тем обстоятельством, что нам неоткуда было взять априорные вероятности $P(H_i)$.

Применение понятий теории игр становится возможным, если считать, что нам заданы штрафы $S(H_i/H_j)$, которые придется уплатить статистику, если он объявит верной гипотезу H_i , в то время как на самом деле верна гипотеза H_j (конечно, естественно считать, что $S(H_i/H_j) = 0$). Кроме того, подобно тому, как это делается в теории игр, нужно разрешить статистику принимать не обязательно какую-нибудь одну из гипотез H_i , а допустить, что гипотезы H_1, \dots, H_n могут приниматься с некоторыми вероятностями p_1, \dots, p_n . Эти вероятности зависят, разумеется, от исхода опыта: если бы был вынут черный шар, то эти вероятности оказались бы другими: q_1, q_2, \dots, q_n . Вероятность π_{ij} того, что статистику придется уплатить штраф $S(H_i/H_j)$, дается равенством

$$\pi_{ij} = P(H_j) \{P(B/H_j) p_i + P(Ч/H_j) q_i\},$$

где $P(H_j)$ — априорная вероятность H_j , $P(B/H_j) = \frac{j}{n}$ — вероятность того, что из урны состава H_j будет вынут белый шар, $P(Ч/H_j) = \frac{n-j}{n}$ — аналогичная вероятность для черного шара. Риск, т. е. математическое ожидание штрафа, дается формулой

$$r = \sum_{i,j=1}^n S(H_i/H_j) \pi_{ij} = r \{P(H_i), p_i, q_i\}, \quad i, j = 1, \dots, n.$$

Очевидно, риск зависит от наборов вероятностей

$$\{P(H_1), \dots, P(H_n)\}, \{p_1, \dots, p_n\}, \{q_1, \dots, q_n\}.$$

Набор априорных вероятностей нам неизвестен. Основная идея теории статистических решений состоит в том, чтобы подбирать $\{p_1, \dots, p_n\}$ и $\{q_1, \dots, q_n\}$ таким образом, чтобы минимизировать

максимум риска, взятый по всем возможным априорным распределениям:

$$\max_{P(H_1), \dots, P(H_n)} r\{P(H_i), p_i, q_j\}$$

(эта функция зависит только от $p_i, q_j, i, j=1, \dots, n$, и предлагается найти минимум этой функции). Решение этой задачи, конечно, зависит от вида штрафов $S(H_i, H_j)$.

По поводу предлагаемого подхода можно сделать два замечания. Во-первых, здесь минимизируется риск при наихудшей стратегии. Поэтому этот подход (и связанная с ним оценка риска) может во многих случаях оказаться слишком осторожным. Но главное препятствие для применения такого подхода обычно состоит в том, что неоткуда взять штрафы $S(H_i/H_j)$ подобно тому, как неоткуда было взять априорные вероятности $P(H_i)$ ¹.

В общем теория статистических решений не может иметь в настоящее время особенно широких применений. Поэтому вряд ли следует включать ее в элементарный курс теории вероятностей, предназначенный для естествоиспытателей. Читателю, желающему подробней ознакомиться с теорией игр и статистических решений, можно рекомендовать, например, книгу Дж. Мак-Кинси [17].

К п. 3.4. Определение независимости случайных величин выбрано в форме определения 3.4 для того, чтобы оно ничем не отличалось от определения независимости в общем (недискретном) случае. Из этого определения немедленно вытекает теорема 3.5, верная и в общем случае. Полезно пояснить наглядный смысл этой теоремы: независимость случайных величин означает, что, зная одну из них, мы ничего дополнительно не можем сказать о другой. А если мы знаем не саму случайную величину ξ , а некоторую функцию от нее $f(\xi)$, то это меньше, чем знать ξ , так как при разных значениях ξ значения $f(\xi)$ могут совпадать. Следовательно, зная $f(\xi)$, мы ничего не можем сказать об η , но тогда ничего нельзя сказать и об $g(\eta)$, так как все сказанное о $g(\eta)$ говорит нечто и о η .

¹ Мы не говорим, конечно, о том случае, когда теорию решений пытаются применять в условиях, когда и вероятности типа $P(B/H_i)$ и $P(C/H_i)$ неоткуда взять. Рассмотрим следующий пример, правда, придуманный автором книги, но стилизованный под математическую геологию, потому что в этой области иногда можно встретить нечто подобное. Пусть нужно решить вопрос о том, является ли некое месторождение промышленным (гипотеза Π) или непромышленным (гипотеза H). С этой целью делается некоторое исследование, которое может иметь несколько результатов A_1, \dots, A_n . На основании этих результатов предлагается выбрать одну из гипотез Π и H с помощью теории статистических решений. В этой задаче сравнительно реально задание штрафов $S(\Pi/H)$ и $S(H/\Pi)$. Однако маловероятно, чтобы можно было говорить о вероятностях $P(A_i/\Pi)$ и $P(A_i/H)$, так как вряд ли здесь можно ожидать статистической устойчивости.

Во всяком случае, при оценке подобного подхода надо прежде всего обращать внимание на то, проверена статистическая устойчивость или нет.

На теорему 3.4 следует смотреть как на формулировку сравнительно легко проверяемого условия независимости случайных величин. Для недискретных величин, имеющих плотности распределения, имеет место аналогичная теорема.

За принятую в курсе непоследовательность при изложении основных вероятностей (см. замечания к п. 2.1) приходится расплачиваться при изложении независимости случайных величин (см. замечание к теореме 3.6). Можно было бы определить независимость n событий A_1, A_2, \dots, A_n следующим образом:

$$P(C_1 C_2 \dots C_n) = P(C_1) P(C_2) \dots P(C_n),$$

где каждое C_i может принимать два значения: A_i и \bar{A}_i . Это определение эквивалентно независимости случайных величин

$$I_{A_i}(\omega) = \begin{cases} 1, & \omega \in A_i \\ 0, & \omega \in \bar{A}_i, \end{cases}$$

называемых индикаторами событий A_i .

Замечания к § 4

К п. 4.1. Понятие сходимости по вероятности является основным понятием сходимости случайных величин, имеющим практическую интерпретацию. Обычно сопоставляются три понятия: сходимость по вероятности, сходимость с вероятностью 1 и сходимость законов распределения. О сходимости законов распределения речь впереди; что касается сходимости с вероятностью 1, то она определяется следующим образом: говорят, что последовательность случайных величин ξ_n сходится к случайной (или неслучайной) величине ξ с вероятностью 1, если

$$P\{\omega: \xi_n(\omega) \rightarrow \xi(\omega), n \rightarrow \infty\} = 1.$$

На понятии сходимости с вероятностью 1 основан ряд изящных теорем, принадлежащих главным образом А. Н. Колмогорову и А. Я. Хинчину и известных под названием теорем типа усиленного закона больших чисел и закона повторного логарифма. Теоремы типа усиленного закона больших чисел утверждают, что разность

$$\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} - M\left(\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n}\right)$$

сходится к нулю с вероятностью 1 (а не по вероятности, как утверждает теорема Чебышева).

Поскольку, как мы неоднократно замечали, случайные величины как функции от элементарного события обычно не наблюдаются, теоремы типа усиленного закона больших чисел вряд ли могут, с прикладной точки зрения, дать больше, чем обычный закон больших чисел.

Закон повторного логарифма формулируется довольно сложно. О нем можно прочесть в первом томе книги Феллера [22] (простейший случай), а также в книге Лозва [16].

К п. 4.2. Говоря о вероятности, относящейся к большому числу случайных величин, например, о

$$P \left\{ \left| \frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} - \frac{M\xi_1 + M\xi_2 + \dots + M\xi_n}{n} \right| > \varepsilon \right\}, \quad (4.1')$$

мы под символом P понимаем вероятность, определенную на том пространстве элементарных событий, на котором определены случайные величины ξ_1, \dots, ξ_n . Однако совершенно нетривиален вопрос о том, существует ли такое пространство. Пусть, например, мы хотим, чтобы все случайные величины ξ_1, \dots, ξ_n имели одинаковое распределение, т. е. принимали значения $a_1, a_2, \dots, a_n, \dots$ с вероятностями $p_1, p_2, \dots, p_n, \dots$ и были в совокупности независимы. Как построить соответствующее пространство Ω ?

В п. 3.1 мы отметили, что для одной случайной величины достаточно в качестве Ω взять множество $\{a_1, a_2, \dots, a_n, \dots\}$ и положить $P(a_i) = p_i$ и $\xi(a_i) = a_i$. Довольно ясно, в соответствии с пунктом 2.3, что для построения пары независимых случайных величин надо взять множество пар (a_i, a_j) , положить $P\{(a_i, a_j)\} = P(a_i)P(a_j) = p_i p_j$ (прямое произведение вероятностных пространств) и считать, что

$$\xi_1\{(a_i, a_j)\} = a_i, \quad \xi_2\{(a_i, a_j)\} = a_j$$

(доказательство независимости ξ_1 и ξ_2 предоставляется читателю). В таком случае для построения n случайных величин нужно взять n -кратное произведение пространств: элементарными событиями будут наборы $(a_{i_1}, a_{i_2}, \dots, a_{i_n})$, причем $P\{a_{i_1}, a_{i_2}, \dots, a_{i_n}\} = p_{i_1} p_{i_2} \dots p_{i_n}$, и $\xi_k\{(a_{i_1}, a_{i_2}, \dots, a_{i_n})\} = a_{i_k}$.

Но тогда при каждом n получается свое пространство элементарных событий и своя вероятность. Нельзя, следовательно, говорить о вероятности (4.1'), а надо говорить о вероятности

$$P_n \left\{ \left| \frac{\xi_1 + \dots + \xi_n}{n} - \frac{M\xi_1 + \dots + M\xi_n}{n} \right| > \varepsilon \right\}. \quad (4.2')$$

Но как быть с вероятностью, относящейся к случайным величинам ξ_1, \dots, ξ_k , где $k < n$? Их можно рассматривать на пространстве элементарных событий для k величин, $(k+1)$ величины, $(k+2)$ величин, \dots , n величин. Будут получаться вероятности P_k, P_{k+1}, \dots, P_n . Может быть, вместо (4.2') следует рассматривать

$$P_m \left\{ \left| \frac{\xi_1 + \dots + \xi_n}{n} - \frac{M\xi_1 + \dots + M\xi_n}{n} \right| > \varepsilon \right\} \quad (4.3')$$

для всех $m \geq n$? Почему-то ничего такого нет ни в одном учебнике теории вероятностей.

Естественно поэтому постараться все величины $\xi_1, \xi_2, \dots, \xi_n, \dots$ (в бесконечном числе) определить на одном пространстве. В качестве пространства можно было бы взять множество бесконечных последовательностей $(a_{i_1}, a_{i_2}, \dots, a_{i_n}, \dots)$. Но здесь нас ждет очень серьезное затруднение: это множество не будет счетным, так что на нем нельзя вести вероятности таким простым образом, как указано в § 1. Ясно, что если положить

$$P_i\{(a_{i_1} a_{i_2} \dots a_{i_n})\} = p_{i_1} p_{i_2} \dots p_{i_n} \dots,$$

то, как правило, в правой части получим нуль.

Выход из этого затруднения указывается одной из знаменитых теорем А. Н. Колмогорова (см. А. Н. Колмогоров [11]). Оказывается, что на множестве последовательностей все-таки можно задать вероятностную меру. Однако эта теорема лежит за пределами данной книги. Логически она не необходима: мы предполагаем, что все наши случайные величины заданы на одном пространстве элементарных событий, и нас мучает не логика, а только совесть: не рассматриваем ли мы нечто такое, чего вообще не может быть? Для успокоения совести достаточно знать о существовании теоремы Колмогорова, не умея ее точно формулировать и доказывать. Такое же положение повторяется при введении аксиоматики, основанной на теории меры: достаточно знать о существовании теории Лебега, из которой следует, что введенный аксиоматикой класс объектов в самом деле не пуст.

Чрезвычайно важно искоренить заблуждение, встречающееся иногда у недостаточно знакомых с теорией вероятностей инженеров и естествоиспытателей, что результат любого эксперимента можно рассматривать как случайную величину. В особо тяжелых случаях к этому присоединяется вера в нормальный закон распределения, а если уж сами случайные величины, не нормальны, то верят, например, что их логарифмы нормальны. Здесь нужно отмечать, что нормальности нет потому, что нет случайности, а случайность означает, что можно говорить о вероятности $P\{\xi \in A\}$ того, что случайная величина ξ примет значение, являющееся элементом числового множества A (A — практически любое множество чисел). Последнее означает, что событие $\{\xi \in A\}$ должно обладать статистической устойчивостью. При необходимости сослаться на авторитет или указать литературу лучше всего использовать гл. XIII книги А. Хальда [24]. Вообще, в странах с развитыми применениями математической статистики важность выделения статистически устойчивых экспериментов давно и хорошо понята.

Замечания к § 5

Математическая статистика — наука, устроенная довольно своеобразно. Если бы общие дефиниции различных наук не ка-

зались в настоящее время потерявшими всякий смысл, то пришлось бы сказать, что математическая статистика занимается определением по опытным данным тех самых вероятностей отдельных элементарных событий, которые в § 1 предполагались заданными.

Чаще, впрочем, приходится заниматься определением распределений случайных величин (мы несколько сдали первоначальную позицию и не требуем уже определения вероятностей всех элементарных событий, что является более сложным делом). Вдумываясь несколько глубже в реальные практические возможности математической статистики, мы будем принуждены сдать и эту позицию и сказать, что небезынтересно бывает хотя бы определить, какими не могут быть распределения случайных величин, наблюдаемых в опыте. Последняя задача формализуется как задача проверки статистических гипотез, рассмотренная в § 5. В конечном счете мы приходим к определению содержания науки исходя из действительно решаемых ею конкретных задач.

В математической статистике можно выделить, пожалуй, три основных понятия. В § 5 рассматриваются два из них — понятие статистической значимости (уровня значимости) и понятие функции мощности статистического критерия, или, что тоже самое, понятия ошибок первого и второго рода. Третье основное понятие статистики — понятие доверительного интервала — будет рассмотрено позже. По существу все эти понятия были известны еще Лапласу, но эволюция научного языка привела к тому, что теперь они выражаются совершенно иначе.

Язык, принятый в § 5, не является вполне общепринятым. Например, знаменитый английский статистик Р. Фишер называет рассуждение о том, что гипотеза либо верна с вероятностью 1, либо неверна (верна с вероятностью 0), жалким софизмом. Однако этот язык принят в настоящее время подавляющим большинством специалистов.

В процессе изучения статистических понятий ошибки первого и второго рода желательно рассматривать одновременно. Однако понятие ошибки второго рода требует указания, кроме основной гипотезы H_0 , множества альтернативных гипотез $\{H_\lambda, \lambda \neq 0\}$. Иногда альтернативные гипотезы трудно специфицировать. Например, пусть гипотеза H_0 состоит в том, что некоторые испытания независимы. Альтернативная гипотеза не может быть выражена в виде: «испытания зависимы», потому что указание альтернативной гипотезы H_λ должно позволить находить вероятности $P\{x \in A/H_\lambda\}$ того, что результат опыта x попадет в какое-то множество A при верной гипотезе H_λ . Следовательно, надо гораздо точнее задать род зависимости испытаний, т. е. задать не просто вероятности исходов отдельных испытаний, а совместную вероятность совокупности их исходов. Но это обычно трудно бывает сделать. Например, предположение о марковском характере зависимости не так уж часто бывает естественным.

В тех случаях, когда понятие альтернативной гипотезы расплывчато, имеет смысл лишь вероятность ошибки первого рода. В таких случаях важно лишь понятие уровня значимости.

Первое знакомство со статистической проверкой гипотез обычно затруднительно для учащегося. К этому вопросу необходимо несколько раз возвращаться на упражнениях. Вообще, упражнения на испытаниях Бернулли и теорема Пуассона должны быть в основном связаны с проверкой гипотез. Придумывание таких задач не составляет труда. Например, «В партии изделий не менее половины изделий должно быть первого сорта. При случайной проверке 10 изделий лишь одно из них оказалось первого сорта. Какой можно сделать вывод? Тот же вопрос, если 3 изделия оказались первого сорта».

На таких задачах можно показать, каким образом теория вероятностей приводит к практически достоверным выводам.

Задачи с ошибками первого и второго рода целесообразно основывать не на биномиальном распределении, а на более простом распределении Пуассона. Например, «Качество партии изделий считается удовлетворительным, если дефектные изделия составляют не более 2%. Сколько нужно испытать изделий из партии, для того чтобы партию, содержащую 3% дефектных изделий, отвергать с вероятностью не менее 0,99, а партию, содержащую 1% дефектных изделий, принимать с вероятностью не менее 0,95?»

Решение. Если проверять n изделий, то для партии с долей брака $p_1=0,01$ число дефектных изделий μ_1 подчиняется закону Пуассона с параметром $\lambda_1=np_1$, а для партии с долей брака $p_2=0,03$ число дефектных изделий μ_2 подчиняется закону Пуассона с параметром $\lambda_2=np_2$.

Каким образом должен решаться вопрос о принятии партии? Очевидно, единственно разумный способ состоит в следующем: выбирается некоторое число k такое, что если число дефектных изделий μ среди проверенных n изделий k или больше k , то партия отвергается, т. е. критическая область имеет вид $\{\mu \geq k\}$. Согласно условию задачи, n должно быть таким, чтобы существовало такое k , что

$$P\{\mu_1 < k\} \geq 0,95, \quad P\{\mu_2 \geq k\} \geq 0,99.$$

Иначе говоря,

$$\sum_{x=0}^{k-1} \frac{\lambda_1^x}{x!} e^{-\lambda_1} \geq 0,95; \quad \sum_{x=0}^{k-1} \frac{\lambda_2^x}{x!} e^{-\lambda_2} \leq 0,01,$$

где $\lambda_1=np_1$, $\lambda_2=np_2$. С помощью таблиц распределения Пуассона такое n легко находится подбором: сначала пробуем $n=500$, т. е. $\lambda_1=5$, $\lambda_2=15$. Для выполнения первого неравенства нужно взять

$k=10$, тогда $\sum_{x=0}^{k-1} \frac{\lambda_2^x}{x!} e^{-\lambda_2} = 0,070 > 0,01$. Пробуем $n=800$. Тогда

$\lambda_1=8$, $\lambda_2=24$. Для выполнения первого неравенства нужно взять $k=14$. Тогда вторая сумма равна 0,010, так что задача решена: нужно взять $n=800$. Следует обратить внимание на огромную величину n : маленькие вероятности $p_1=0,01$ и $p_2=0,03$ различать трудно. Приведенное решение задачи несколько условно из-за использования приближенной формулы Пуассона на правах точного выражения. Но практически довольно условно постоянство вероятностей p_1 и p_2 , так что применение более сложной аппроксимации биномиального распределения практически не оправдано, тем более, что практически интересно лишь приближенное знание числа n .

Чрезвычайно полезно также нарисовать график вероятности отвергнуть вторую партию с $p_2=0,03$ в зависимости от числа проверяемых изделий n и при условии, что

$$\sum_{x=0}^{k-1} \frac{\lambda_1^x}{x!} e^{-\lambda_1} \geq 0,95$$

(т. е. при заданной вероятности отвергнуть удовлетворительную партию). Этот график называется *оперативной характеристикой* статистического критерия $\{\mu \geq k\}$.

Заметим, что затронутому здесь вопросу о контроле качества продукции посвящена обширная литература. Наибольшим успехом в этой области является *последовательный анализ* Вальда [5].

Но все статистические методы контроля могут быть полезны лишь при наличии статистически устойчивого («статистически подконтрольного», как говорит в своей книге Хальд [24]) производства.

Замечания к § 6

В кратком курсе теории вероятностей, предназначенном для естествоиспытателей, не следует уделять много места теории меры и теоремам типа усиленного закона больших чисел, поскольку эти теоремы не имеют существенной естественнонаучной интерпретации (любопытно отметить, что в аксиоматике теории Мизеса, созданной его последователями, имеются специальные аксиомы, по существу запрещающие рассматривать такие теоремы).

С другой стороны, имеются веские основания для введения интеграла Лебега. Речь идет прежде всего о введении понятия математического ожидания. Нельзя обойтись без формулы

$$Mf(\xi) = \int_{-\infty}^{\infty} f(x) \mu_{\xi}(dx) \quad (6.1')$$

или хотя бы без ее варианта для случая, когда μ_{ξ} имеет плотность

$$Mf(\xi) = \int_{-\infty}^{\infty} f(x) p_{\xi}(x) dx.$$

Эту формулу нельзя, к сожалению, ввести просто в качестве определения математического ожидания. Дело в том, что случайная величина $\eta = f(\xi)$ имеет собственную плотность распределения $p_{\eta}(x)$, следовательно,

$$M\eta = Mf(\xi) = \int_{-\infty}^{\infty} xp_{\eta}(x) dx.$$

Следовательно, нужно уметь доказать, что если $\eta = f(\xi)$, то

$$\int_{-\infty}^{\infty} xp_{\eta}(x) dx = \int_{-\infty}^{\infty} f(x) p_{\xi}(x) dx.$$

Если функция f взаимно однозначно отображает R^1 в R^1 , то это просто. Но в случае неоднозначной функции это сложно. Анализ возможных способов доказательства этого равенства показывает, что они используют, по существу, способ построения интеграла Лебега. Методически недопустимо, чтобы существующие в математике достаточно общие и ясные понятия вводились в неявном виде. Кроме того, интеграл Лебега одинаково пригоден для одномерной и многомерной случайной величины ξ (см. А. Н. Колмогоров [11]).

Интеграл Лебега по способу своего построения, бесспорно, проще, чем интеграл Римана. Разберем, на чем основана эта простота. Римановы суммы состояются из значений интегрируемой функции, умноженных на длины отрезков. Лебегова сумма имеет вид

$$\sum_{k=-\infty}^{\infty} \frac{k}{n} P \left\{ \omega : \frac{k}{n} \leq \xi(\omega) < \frac{k+1}{n} \right\}.$$

Таким образом, считается возможным «измерять», т. е. приписывать значения меры P множествам вида $\left\{ \omega : \frac{k}{n} \leq \xi(\omega) < \frac{k+1}{n} \right\}$. Если даже $\Omega = \{\omega\}$ есть отрезок $[0,1]$, функция $\xi(\omega)$ гладкая, то все равно множества

$$\left\{ \omega : \frac{k}{n} \leq \xi(\omega) < \frac{k+1}{n} \right\}$$

могут быть сложными. Например, для функции $\omega^2 \sin \frac{1}{\omega}$ такие множества состоят из счетного объединения полуинтервалов. Счетные множества таят неожиданности. Например, не так просто до-

казать, что если интервал $[0,1)$ представлен в виде объединения счетного числа интервалов A_i :

$$[0, 1) = \bigcup_{i=1}^{\infty} A_i, \quad A_i A_j = \emptyset,$$

то длина $l\{(0,1)\}$ равна сумме длин $\sum_{i=1}^{\infty} l(A_i)$. Если интервалы A_i уложены на интервале $[0,1)$ по порядку, например

$$A_i = \left[1 - \frac{1}{i}, 1 - \frac{1}{i+1}\right), \quad i = 1, 2, \dots,$$

то утверждение очевидно. Однако необязательно концам интервалов A_i накапливаться именно к точке 1: они могут накапливаться к точке $1/2$:

$$A_i = \left[\frac{1}{2} - \frac{1}{i}, \frac{1}{2} - \frac{1}{i+1}\right), \quad i = 2, 3, \dots,$$

а к системе интервалов $A_i, i=2, 3, \dots$ можно прибавить еще интервал $\left[\frac{1}{2}, 1\right)$, чтобы покрыть весь отрезок $[0,1)$. Далее, можно считать, что точек накопления концов интервалов не одна, а счетное множество, и сами такие точки имеют точки накопления, притом тоже счетное множество и т. д. Неудивительно, что наше утверждение нуждается в нетривиальном доказательстве (в книге А. Н. Колмогорова и С. В. Фомина [13] аналогичный вопрос разобран в § 1 гл. V). Поэтому вопрос об «измерении» множеств типа $\left\{\omega : \frac{k}{n} \leq \xi(\omega) < \frac{k+1}{n}\right\}$ нетривиален.

Тем не менее, после построения лебеговской теории меры этот вопрос решен положительно. В дальнейшем было выяснено, что положительное решение вопроса распространяется не только на случай прямой, плоскости и евклидова пространства, но и на гораздо более широкий класс пространств. Таким образом, выяснилось, что модель, в которой вероятностная мера задана на некоторой σ -алгебре, имеет массу применений. В настоящей книге несколько «паразитически» используется это установленное прогрессом науки обстоятельство: просто постулируется аксиоматика Колмогорова (аксиомы 6.1—6.5) без приведения конкретных примеров. С чисто логической точки зрения такая позиция безупречна; для успокоения совести достаточно просто знать (без строгого доказательства), что постулировано не что-то искусственное, а вещь, имеющая большие применения. При приведении примеров (примеры 6.1 и 6.2) приходится признать, однако, отсутствие полных доказательств. Таким образом, общую методическую установку данной книги можно сформулировать как «интеграл Лебега без теории меры». Такая установка является заведомо приемлемой,

если курс теории меры имеется в учебном плане, хотя бы после курса теории вероятностей. В случае, если ни в каком курсе, кроме курса теории вероятностей, теории меры в том или ином виде нет, вопрос о допустимости такого подхода может быть спорным. При наличии времени нужно включать в курс теории вероятностей теорию меры и интеграла Лебега (используя, например, [13]).

Однако если такой возможности нет, то, по мнению автора, будет потеряно не слишком много: вряд ли естествоиспытателю приходится измерять сложные множества с помощью лебеговского продолжения меры. С другой стороны, построение лебеговского интеграла с использованием готовой теории меры настолько просто, обще и удобно, что должно быть известно каждому знакомому с основами современной математики естествоиспытателю.

К п. 6.1. Допущение счетных операций над событиями чрезвычайно удобно с чисто математической точки зрения, но явно выводит в самых простейших случаях (пример 6.1) за рамки наглядно представимых подмножеств множества Ω . Однако не существует какого-либо более простого средства обеспечить необходимые математические удобства. Например, если последовательность случайных величин (т. е. измеримых функций) $\xi_n(\omega)$ сходится при каждом ω к функции $\xi(\omega)$, то $\xi(\omega)$ измерима. Доказывается это так:

$$\{\omega : \xi(\omega) \leq c\} = \bigcap_{m=1}^{\infty} \left[\bigcup_{k=1}^{\infty} \left\{ \bigcap_{n=k}^{\infty} \xi_n(\omega) < c + \frac{1}{m} \right\} \right],$$

т. е. с полным использованием счетных теоретико-множественных сумм и пересечений.

Понятие наименьшей σ -алгебры, содержащей данные множества, довольно трудно определить, если исходить из того, что нужно рассматривать те множества, которые получаются из данных сначала счетным суммированием или пересечением, затем счетным суммированием или пересечением тех множеств, которые получились на первом этапе, и т. д. Здесь требуется применение трансфинитных чисел (см. П. С. Александров [1]). Данное в § 6 определение наименьшей σ -алгебры напоминает незаконные способы оперировать с теоретико-множественными терминами, но пока что признается достаточно законным.

При рассмотрении примера 6.1 употребляются понятия замкнутого и открытого множества и их простейшие свойства. Этот материал включается в курс математического анализа. Если на него почему-либо нежелательно ссылаться, то основной факт — что практически все функции на единичном отрезке измеримы по Борелю, можно проиллюстрировать, нарисовав график функции и показав, что $\{\omega : f(\omega) \leq c\}$ есть сумма конечного или счетного числа интервалов.

К п. 6.2. Для случая интеграла Лебега дается минимальный набор определений и теорем, необходимый для целей данной кни-

ги (практически, для введения определения математического ожидания). Более подробно см. [13].

В тех случаях, когда интеграл Римана существует, он равен интегралу Лебега. Можно было бы подумать, что современный курс математического анализа может вообще не содержать интеграла Римана, а только интеграл Лебега. Это, однако, заблуждение. Дело в том, что интеграл Лебега неестествен для ориентированных многообразий. Простейший пример, на котором можно почувствовать разницу, заключается в следующем. Если $x = \varphi(t)$, причем $\varphi(t_0) = a$, $\varphi(T) = b$, то

$$\int_a^b f(x) dx = \int_{t_0}^T f(\varphi(t)) \varphi'(t) dt. \quad (6.4')$$

для интеграла Римана и

$$\int_a^b f(x) dx = \int_{t_0}^T f(\varphi(t)) |\varphi'(t)| dt \quad (6.5')$$

для интеграла Лебега. Однако формула (6.4') верна всегда, а формула (6.5') верна лишь для монотонных функций $\varphi(t)$.

Интеграл $\int_a^b f(x) dx$ по отрезку $[a, b]$ существует таким образом, в двух видах — ориентированном и неориентированном. Ориентированный интеграл дает начало обобщениям на многообразия, в частности, знаменитой формуле Ньютона—Лейбница—Гаусса — Грина — Остроградского — Стокса — Пуанкаре. Неориентированный интеграл дает начало интегралу Лебега и общему понятию математического ожидания.

Замечания к § 7

К п. 7.1. Цель этого параграфа — перенести на общий случай свойства случайных величин и математических ожиданий, известные ранее для дискретного случая. (Формула

$$M(a\xi + b\eta) = aM\xi + bM\eta$$

уже установлена в конце § 6.) Иногда такое обобщение делается в два этапа: сначала все переносится на одномерные случайные величины и затем — на многомерные. Использование интеграла Лебега позволяет рассматривать сразу многомерный случай.

Возникает вопрос, следует ли при изложении теории вероятностей сначала рассматривать дискретный случай и лишь затем — общий. По мнению автора, с точки зрения чисто математической не требуется подготовительного курса дискретной теории вероятностей. При мало-мальски приличном изложении, наличии

учебных пособий и продуманном ведении упражнений студенты разберут любой курс. Однако для обсуждения принципиальных вопросов применимости теории вероятностей следует начинать с простой дискретной модели. Общая модель, с точки зрения естествоиспытателя, не является до конца естественной (см. начало п. 2 § 6), и внимание слишком связывается математическими трудностями.

Все множества, рассматриваемые в § 7, борелевские, все функции на R^n — измеримые по Борелю. Практически это не накладывает никаких ограничений. Можно было бы рассматривать вместо R^n в качестве множества значений случайных величин любое множество R , в котором выделена σ -алгебра подмножеств \mathfrak{M} . Тогда определение случайной величины со значениями из R выглядело бы так: $\xi = \xi(\omega)$ есть отображение $\Omega \rightarrow R$ такое, что для любого $M \in \mathfrak{M}$ прообраз $\xi^{-1}(M) \in \mathfrak{B}$. При этом на \mathfrak{M} индуцируется мера μ_ξ по формуле:

$$\mu_\xi(M) = P\{\omega: \xi(\omega) \in M\} = P\{\xi \in M\}.$$

При рассмотрении случайных точек на сфере, случайных матриц и т. д. имеются в виду определения именно по такой схеме. При этом автоматически сохраняются теоремы 7.1 и 7.2.

К п. 7.2 и 7.3. Что реально означает интеграл Лебега

$$\int_{R^n} f(x) \mu_\xi(dx),$$

т. е. каким образом его следует вычислять? Можно было бы опасаться, что построение книги по принципу «интеграл Лебега без теории меры» приведет в этом месте к неприятным последствиям. Однако это не так: никто не вычисляет интегралов Лебега, определяя сначала меру с помощью лебеговского продолжения. Частные случаи, в которых такое вычисление возможно, изложены в п. 7.2.

Случай существования плотности распределения противопоставляется дискретному случаю. В большинстве случаев именно так и бывает: если случайная величина недискретна, то она имеет плотность распределения. Однако не мешает описать строго возможные здесь случаи.

Определяемая по формуле $\mu_\xi(B) = P\{\xi \in B\}$ мера μ_ξ в евклидовом пространстве R^n сопоставляется с мерой Лебега l в этом пространстве (мера Лебега на несложных множествах совпадает с объемом, а на множествах, устроенных сложно, определяется с помощью лебеговского продолжения меры, но нам нет нужды входить в детали этого построения). Действительно, плотность $p_\xi(x)$ есть (измеримая по Борелю) функция, удовлетворяющая соотношению

$$P\{\xi \in B\} = \mu_\xi(B) = \int_B p_\xi(x) dx,$$

где под dx понимается элемент лебеговской меры (придерживаясь единых обозначений, следовало бы писать $\int_B \rho_{\xi}(x) l(dx)$). Из теории меры известен следующий факт. Пусть на некоторой σ -алгебре подмножеств определены две меры μ и ν . В таком случае меру μ можно представить в виде

$$\mu = \mu_1 + \mu_2,$$

где

1) мера μ_1 абсолютно непрерывна относительно меры ν , т. е. существует (измеримая) функция $\rho_1(x)$ такая, что¹:

$$\mu_1(B) = \int_B \rho_1(x) \nu(dx);$$

2) мера μ_2 сингулярна относительно меры ν , т. е. существует множество C_{μ_2} такое, что

$$\mu_2(\bar{C}_{\mu_2}) = 0, \quad \nu(C_{\mu_2}) = 0$$

(иными словами, мера μ_2 сосредоточена на множестве C_{μ_2} таком, что $\nu(C_{\mu_2}) = 0$).

Следовательно, вероятностная мера μ_{ξ} в евклидовом пространстве R^n представляется в виде

$$\mu_{\xi} = \mu_1 + \mu_2,$$

где мера μ_1 имеет плотность относительно меры Лебега l , а μ_2 сингулярна относительно l . В том случае, когда $\mu_{\xi} = \mu_1$, плотность меры μ_1 совпадает с плотностью меры μ_{ξ} , и этот случай предлагается называть абсолютно непрерывным.

Интересен вопрос, какими могут быть меры μ_2 , т. е. сингулярные меры. Достаточно рассмотреть случай, когда $\mu_{\xi} = \mu_2$, т. е. абсолютно непрерывная компонента μ_1 распределения μ_{ξ} отсутствует. В этом случае мера μ_{ξ} сосредоточена на множестве C таком, что $l(C) = 0$. Для случая пространства R^n размерности $n > 1$ такие множества возникают совершенно естественно: во-первых, это множества, состоящие из счетного числа точек, а во-вторых, все поверхности (кривые) размерности меньшей n . В случае $n = 1$ счетные множества, конечно, также имеют лебегову меру нуль, но приведение примеров других множеств нулевой меры оказывается сложнее. Пожалуй, самым лучшим примером здесь оказывается канторовское множество. Оно строится следующим образом. Берется отрезок $[0, 1]$, делится на три части:

¹ Функция $\rho_1(x)$ называется плотностью меры μ_1 относительно меры ν и иногда вводится обозначение

$$\rho_1(x) = \frac{d\mu_1}{d\nu}(x).$$

$$\left[0, \frac{1}{3}\right], \left(\frac{1}{3}, \frac{2}{3}\right), \left[\frac{2}{3}, 1\right]$$

и средний интервал выбрасывается. Каждый из двух оставшихся интервалов вновь делится на три части и средний (открытый) интервал выбрасывается. То, что остается после этой процедуры, повторенной счетное число раз, называется канторовским совершенным множеством C . Сумма длин выброшенных интервалов есть

$$\frac{1}{3} + 2 \cdot \frac{1}{9} + 4 \cdot \frac{1}{27} + \dots = \frac{1}{2} \sum_{n=1}^{\infty} \left(\frac{2}{3}\right)^n = 1.$$

Поэтому $l(C) = 0$. Канторовское множество может быть еще описано как множество точек $x \in [0, 1]$ таких, что в представлении

$$x = \sum_{n=1}^{\infty} \frac{x_n}{3^n}, \quad x_n = 0, 1, 2;$$

в виде троичной дроби «цифры» x_n не равны 1, а могут принимать лишь значения 0 и 2. Выброшенные при построении C интервалы называются смежными интервалами; каждый смежный интервал может быть охарактеризован как множество точек $x \in [0, 1]$ таких, что «цифры» x_n троичного разложения x выглядят так: несколько первых «цифр» x_1, \dots, x_k принимают значения 0 или 2, а «цифра» $x_{k+1} = 1$.

Один из удивительных математических фактов заключается в том, что испытания Бернулли (бросания монеты) можно интерпретировать как случайное бросание точки в канторовское множество C . Единственное, что нужно для этого позволить, — это рассматривать счетные последовательности испытаний Бернулли. Положим $\xi_n = 0$, если при n -ном бросании монеты выпала цифра, и $\xi_n = 2$, если при n -ном бросании монеты выпал герб. Составим

$$\xi = \sum_{n=1}^{\infty} \frac{\xi_n}{3^n}$$

и тогда ξ — случайная точка множества C . Рассмотрим функцию распределения

$$F_{\xi}(x) = P\{\xi < x\}.$$

Ясно, что если x меняется в пределах одного смежного интервала, то $P\{\xi < x\}$ остается постоянной. Следовательно, график монотонной функции $F_{\xi}(x)$ образует нечто вроде лестницы: горизонтальные площадки «ступеней» расположены над смежными интервалами. На каждом смежном интервале существует $\frac{dF_{\xi}(x)}{dx} = 0$. Сле-

довательно, на множестве лебеговской меры 1 имеем $\frac{dF_{\xi}(x)}{dx} = 0$.

В точках множества C эта производная не существует.

Таким образом, определять плотность $p_{\xi}(x)$ распределения случайной величины ξ равенством $p_{\xi}(x) = \frac{dF_{\xi}(x)}{dx}$ нельзя. Действительно, если потребовать выполнения этого равенства (следовательно, существования $\frac{dF_{\xi}(x)}{dx}$) при всех x , то не будут охвачены важные случаи, вроде показательного распределения (у которого не существует $\frac{dF_{\xi}(0)}{dx}$). Если же допускать нарушение этого равенства в отдельных точках x , то нужно прибавить, что эти исключительные точки образуют довольно редкое множество. Но какое именно? Обычно говорят, что лебеговская мера этого множества должна быть равна нулю. Однако в только что рассмотренном примере $\frac{dF_{\xi}(x)}{dx} = 0$ на множестве меры 1. Между тем явно нельзя считать, что плотность $p_{\xi}(x) = 0$. Можно показать, что следующее определение будет корректным: если $p_{\xi}(x) = \frac{dF_{\xi}(x)}{dx}$ существует во всех точках, кроме множества лебеговской меры 0, и

$$\int_{-\infty}^{\infty} p_{\xi}(x) dx = 1,$$

то $p_{\xi}(x)$ называется плотностью распределения. Однако методически проще дать интегральное определение плотности (определение 7.5), удобное в многомерном случае.

Иногда (например, при рассмотрении показательного распределения) требуется по функции $F_{\xi}(x)$ найти $p_{\xi}(x)$. Можно показать, что равенство

$$P\{\xi \in B\} = \int_B p_{\xi}(x) dx$$

не нужно проверять для всех борелевских множеств B . Достаточно проверить это равенство для всех интервалов. Если $B = [a, b)$, то

$$P\{\xi \in B\} = P\{\xi < b\} - P\{\xi < a\} = F_{\xi}(b) - F_{\xi}(a),$$

так что достаточно установить, что

$$F_{\xi}(b) - F_{\xi}(a) = \int_a^b p_{\xi}(x) dx.$$

Это эквивалентно (как легко видеть) равенству

$$F_{\xi}(x) = \int_{-\infty}^x p_{\xi}(y) dy.$$

Следовательно, это равенство можно считать определением плотности. Однако возможность перехода от класса борелевских множеств к классу интервалов требует обоснования (в § 7 этот вопрос намеренно смазан). Это обоснование делается следующим образом. Рассмотрим две функции множества B : $P\{\xi \in B\}$ и $\int_B p_{\xi}(x) dx$. Обе эти функции являются счетно аддитивными

функциями B (см., например, [13]). Нам дано, что эти две функции совпадают для интервалов. По теореме о единственности продолжения меры они совпадают и для всех борелевских множеств B .

Таким образом, приходится сослаться на две теоремы: о счетной аддитивности интеграла и о единственности продолжения меры.

Мы видим, что, приняв принцип «интеграл Лебега без теории меры», невозможно сделать изложение абсолютно строгим. Подчеркнем, однако, что «принцип» этот не является обязательным: дело тут лишь в экономии времени. Что приобретает будущий естествоиспытатель, если будет знать полное доказательство того, что из соотношения

$$F_{\xi}(x) = \int_{-\infty}^x p_{\xi}(y) dy$$

вытекает, что для любого борелевского B

$$P\{\xi \in B\} = \int_B p_{\xi}(y) dy?$$

Для интервалов и конечных сумм интервалов это непосредственно очевидно. Более сложные борелевские множества вряд ли имеют физический смысл. Фактически, например, для показательного распределения речь идет о том, получить ли эвристически формулу

$$F_{\xi}(x) = 1 - e^{-\lambda x}$$

и затем строго доказать, что $p_{\xi}(x) = \lambda e^{-\lambda x}$, или продолжить эвристические рассуждения на один шаг дальше и «нестрого» получить ту же формулу для $p_{\xi}(x)$. Однако при получении $F_{\xi}(x)$ столько «нестрогостей», что эта последняя не имеет никакого значения для справедливости результата.

В конце п. 7.3 (пример 7.1) читатель впервые сталкивается с эвристическим выводом закона распределения в теории вероят-

ностей на примере показательного закона. Иногда встречается заблуждение, согласно которому считается, что такие рассуждения, основанные на сравнительно произвольных вероятностных моделях, тем не менее являются строгими доказательствами того, что должен выполняться определенный закон распределения. Разъяснить ошибочность такого взгляда легко: собственно, он встречается лишь у тех, кто не понимает до конца использованных рассуждений.

Труднее сделать убедительной ту идею, что хорошие вероятностные закономерности продолжают оставаться справедливыми, даже если нет полного согласия с действительностью у той простой модели, на основании которой они выведены. Отмеченная гармоничность показательного закона: $\min(\xi_1, \dots, \xi_n)$ имеет показательное распределение, если каждая из независимых величин ξ_1, \dots, ξ_n имеет показательное распределение; показательный закон и закон Пуассона тесно связаны — не остается без отклика в природе. Беда в том, что так называемые «статистические» данные чаще всего будут отвергать гипотезу о показательном законе. Однако можно думать, что это иногда случается потому, что данные не являются «статистическими», т. е. у них нет статистической однородности.

К п. 7.4. Попытка строить теорию многомерных случайных величин на основании многомерной функции распределения обречена на неудачу [11]. В п. 7.4 описана одна из причин этого, связанная с невозможностью изменения системы координат. К этому можно прибавить следующее. В одномерном случае

$$M f(\xi) = \int_{-\infty}^{\infty} f(x) dF_{\xi}(x),$$

где интеграл можно понимать в смысле Римана—Стилтьеса. В многомерном случае пришлось бы рассматривать интегралы вида

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) d^n F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n).$$

Однако у n -кратного дифференциала

$$d^n F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$$

нет даже установившегося обозначения. Кроме того, по-видимому, ни у кого еще не хватило терпения выписать подряд все формулы, какие нужны для обоснования теории таких интегралов. Особенно «приятно» вычислять такие интегралы, если распределение случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ сосредоточено на некоторой поверхности (ср. случай 4) п. 7.2).

Пропадает и естественная характеристика класса функций, могущих быть функциями распределения: накладывается условие,

которое наглядно состоит в том, что для любого параллелепипеда C вероятность $P\{\xi \in C\}$, определенная по функции $F_{\xi_1 \dots \xi_n}(x_1, \dots, x_n)$ с помощью составления «смешанной разности» от нее, была бы неотрицательна (см. В. Феллер [22], т. 2, стр. 167).

Основным способом задания меры в многомерном случае является (кроме дискретного случая) задание плотности распределения.

При выводе основной теоремы 7.4 для преобразования плотности распределения требуется использовать понятие якобиана преобразования. Нельзя сказать, чтобы это понятие не было четким для современных курсов математического анализа. Однако эта четкость совершенно скрывается традиционно сложившимися обозначениями. Например, в обозначении $\frac{D(y_1, \dots, y_n)}{D(x_1, \dots, x_n)}$ совершенно не видно, в какой точке берется значение якобиана. В самом деле, уже для функций одной переменной записывается формула

$$\frac{dy}{dx} = \left[\frac{dx}{dy} \right]^{-1}.$$

Но что она означает?

Во-первых, никто не оспаривает, что $\frac{dy}{dx}$ и $\frac{dx}{dy}$ — это функции. Следовательно, написанную формулу следует понимать как равенство двух функций. Две функции одного и того же переменного называются равными, если равны их значения при одном и том же значении аргумента. Но функции $\frac{dy}{dx}$ и $\frac{dx}{dy}$ есть функции разных аргументов — соответственно, x и y . А что значит, что функции двух разных аргументов равны — этого, пожалуй, никто не знает. Можно было бы считать, что они равны, если равны их значения при подстановке одного и того же числа вместо аргумента. В таком случае $\sin x = \sin y$, но

$$\frac{dy}{dx} \neq \left[\frac{dx}{dy} \right]^{-1}.$$

Конечно, здесь речь идет о чисто языковой небрежности: небольшим размышлением можно установить, что хотят сказать следующее: если $f'(x) \neq 0$, то некоторая окрестность точки x взаимно однозначно отображается на окрестность точки $f(x)$, причем

$$(f^{-1})'(f(x)) = [f'(x)]^{-1}.$$

Более естественно это выглядит, если оператор производной обозначить не f' , а Df :

$$Df^{-1}(f(x)) = [Df(x)]^{-1}.$$

Во избежание путаницы предлагается то же самое обозначение использовать для якобиана.

Теорема 7.4 является общей и удобной. Например, из нее немедленно вытекает результат примера 7.2. Необходимо подчеркнуть, что в этой теореме f — взаимно однозначное отображение R^n в R^n . В случае не более чем счетной кратности отображения (т. е. $f^{-1}(x)$ счетно) имеем:

$$\rho_\eta(x) = \sum_{y \in f^{-1}(x)} \rho_\xi(y) |Df^{-1}(y)|.$$

Для непрерывных плотностей распределения теорему 7.4 можно вывести следующим образом. Будем обозначать через $U(x)$ окрестность точки x и через $\lim_{U(x) \downarrow x}$ предельный переход, при котором диаметр $U(x)$ стремится к нулю. Пусть $\eta = f(\xi)$. Очевидно, что для непрерывной ρ_η имеем

$$\rho_\eta(x) = \lim_{U(x) \downarrow x} \frac{P\{\eta \in U(x)\}}{V(U(x))}.$$

где V обозначает объем. Имеем далее

$$\begin{aligned} \lim_{U(x) \downarrow x} \frac{P\{\eta \in U(x)\}}{V(U(x))} &= \lim_{U(x) \downarrow x} \frac{P\{f(\xi) \in U(x)\}}{V(U(x))} = \\ &= \lim_{U(x) \downarrow x} \frac{P\{\xi \in f^{-1}(U(x))\}}{V\{f^{-1}(U(x))\}} \cdot \frac{V\{f^{-1}(U(x))\}}{V(U(x))} = \\ &= \lim_{U(f^{-1}(x)) \downarrow f^{-1}(x)} \frac{P\{\xi \in U(f^{-1}(x))\}}{V\{U(f^{-1}(x))\}} \times \\ &\times \lim_{U(x) \downarrow x} \frac{V\{f^{-1}(U(x))\}}{V(U(x))} = \rho_\xi(f^{-1}(x)) |Df^{-1}(x)|. \end{aligned}$$

В п. 7.4 принято иное доказательство, основанное на формуле замены переменных, справедливой для любых измеримых функций (в курсах анализа она доказывается для непрерывных функций, но ее распространение на измеримые функции делается самым стандартным предельным переходом).

При определении (определения 3.4 и 3.5) независимых случайных величин в дискретном случае участвовали произвольные множества A, B (определение 3.4) и A_1, \dots, A_n (определение 3.5). В общем случае слово «произвольные» надо заменить словом «борелевские», но для естествоиспытателя это безразлично. Теорема 3.5 вместе с доказательством остается в силе, только «любые» функции надо заменить на «измеримые по Борелю».

Достаточное условие независимости случайных величин, выраженное через плотности распределения, доказывается на основе теоремы о произвольности порядка интегрирования. Для интегралов Римана эта теорема должна быть известна читателю; в справедливость ее для интегралов Лебега (теорема Фубини) читатель легко поверит.

В § 1—7 речь шла об основных понятиях теории вероятностей и связях между ними, которые часто выражались в виде тривиальных теорем¹. Настоящее содержание теории вероятностей впервые серьезно затрагивается в § 8, посвященном центральной предельной теореме. Поскольку в части, касающейся основных определений, в настоящей книге изложен (хотя и кратко) по существу весь необходимый научный материал, то замечания к § 1—7 имели форму методических комментариев к отдельным пунктам этих параграфов. Ничего подобного нельзя сказать о материале § 8, так как он охватывает лишь небольшую часть научных результатов, связанных с центральной предельной теоремой. В интересах преподавателя теории вероятностей — прежде всего шире познакомиться с этими результатами. Поэтому замечания к § 8 своей основной целью имеют расширение научного содержания.

8.1'. Многомерное нормальное распределение. В распространенной учебной литературе недостаточно полно описан многомерный аналог нормального распределения, введенного в п. 8.2. Мы дадим вариант изложения, в котором за счет применения линейной алгебры достигается большая сжатость.

Определение 8.1'. Пусть $\xi = (\xi_1, \dots, \xi_n)$ — случайная величина с векторными значениями. Вектором математических ожиданий называется вектор

$$[m_\xi = (M \xi_1, \dots, M \xi_n),$$

матрицей ковариаций C_ξ называется квадратная симметричная матрица размером $n \times n$, элементами которой являются ковариации

$$[\text{cov}(\xi_i, \xi_j) = M(\xi_i - M \xi_i)(\xi_j - M \xi_j), \quad i, j = 1, \dots, n.$$

Лемма 8.1'. Пусть $x = (x_1, \dots, x_n)$ есть любой неслучайный вектор. Тогда

$$D(x, \xi) = (C_\xi x, x),$$

где D обозначает дисперсию, а (x, ξ) — скалярное произведение x на ξ .

Доказательство.

$$D(x, \xi) = D \left\{ \sum_{i=1}^n x_i \xi_i \right\} =$$

¹ Исключением является, в частности, закон больших чисел, доказательство которого хотя и просто, но совершенно нетривиально (в том смысле, что ничего подобного нет в математическом анализе, так что закон больших чисел является самостоятельной теоремой, а не перефразировкой известных утверждений).

$$= \sum_{i,j=1}^n x_i x_j \operatorname{cov}(\xi_i, \xi_j) = (C_{\xi} x, x),$$

где, разумеется, $\operatorname{cov}(\xi_i, \xi_i) = D\xi_i$.

Лемма 8.2'. Для того чтобы распределение вектора ξ было сосредоточено в некоторой гиперплоскости, необходимо и достаточно, чтобы матрица C_{ξ} была вырожденной.

Доказательство. Всякая гиперплоскость имеет вид

$$L = \{x = (x_1, \dots, x_n) : (a, x) = q\},$$

где $a = (a_1, \dots, a_n)$ — нормальный к этой гиперплоскости вектор и q — действительное число. Для того чтобы распределение вектора ξ было сосредоточено в L , необходимо и достаточно, чтобы при некотором q

$$P\{(a, \xi) = q\} = 1,$$

иначе говоря, чтобы $D(a, \xi) = 0$, т. е. $(C_{\xi} a, a) = 0$, что и означает вырожденность симметрической неотрицательно определенной (в силу леммы 8.1') матрицы C_{ξ} .

Лемма 8.3'. Пусть $\eta = A\xi + a$, где a — неслучайный вектор, A — неслучайная матрица. В таком случае

$$C_{\eta} = AC_{\xi}A',$$

где A' — транспонированная матрица A .

Доказательство. Имеем

$$\begin{aligned} D(x, \eta) &= D(x, A\xi + a) = D(x, A\xi) = \\ &= D(A'x, \xi) = (C_{\xi} A'x, A'x) = (AC_{\xi} A'x, x). \end{aligned}$$

С другой стороны, в силу леммы 8.1',

$$D(x, \eta) = (C_{\eta} x, x).$$

Лемма доказана.

Изложенные результаты имеют место для любого случайного вектора ξ (предполагается, конечно, существование дисперсий его компонент). Переходим теперь к определению и изучению нормально распределенных случайных векторов.

Определение 8.2'. Стандартным нормальным распределением называется распределение вектора $\xi = (\xi_1, \dots, \xi_n)$, компоненты которого независимы и имеют каждая одномерное нормальное распределение $N(0,1)$.

Замечание. Плотность $p_{\xi}(x)$ распределения вектора ξ имеет вид

$$p_{\xi}(x) = \prod_{i=1}^n p_{\xi_i}(x_i) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \right) = \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{(x, x)}{2}}.$$

Очевидно, что $C_{\xi} = E$, где E — единичная матрица.

Определение 8.3'. Нормальным распределением называется распределение вектора

$$\eta = A\xi + a,$$

где A — матрица, a — вектор, ξ — случайный вектор, имеющий стандартное нормальное распределение.

Очевидно, что $m_\eta = a$, $C_\eta = AC_\xi A' = AA'$.

Если матрица A вырождена, то вектор η распределен на некоторой гиперплоскости; следовательно, его распределение не имеет плотности в пространстве R^n . Если матрица A невырождена, то невырождена и матрица ковариаций

$$C_\eta = AA'$$

и справедлива следующая теорема.

Теорема 8.1'. Плотность $p_\eta(x)$ многомерного нормального распределения задается формулой

$$p_\eta(x) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det C_\eta}} \exp \left\{ -\frac{1}{2} (C_\eta^{-1} (x-a), (x-a)) \right\}.$$

Доказательство. Имеем (в силу теоремы 7.4)

$$\begin{aligned} p_\eta(x) &= \frac{1}{|\det A|} p_\xi(A^{-1}(x-a)) = \\ &= \frac{1}{(\sqrt{2\pi})^n |\det A|} \exp \left\{ -\frac{1}{2} (A^{-1}(x-a), A^{-1}(x-a)) \right\} = \\ &= \frac{1}{(\sqrt{2\pi})^n |\det A|} \exp \left\{ -\frac{1}{2} ((AA')^{-1}(x-a), x-a) \right\} \end{aligned}$$

(мы используем выражение $(A')^{-1} = (A^{-1})'$, которое легко получить, транспонируя равенство $AA^{-1} = E$). Учитывая, что $AA' = C_\eta$, а, следовательно, $|\det A| = \sqrt{\det C_\eta}$, получаем требуемую формулу.

Замечание. Распределение вектора η полностью определяется вектором m_η и матрицей C_η . Это распределение обозначается через $N(a, C_\eta)$ и называется многомерным нормальным распределением с параметрами a и C_η .

Пример. В двумерном случае матрица ковариаций вектора η имеет вид

$$C_\eta = \begin{pmatrix} D\eta_1 \operatorname{cov}(\eta_1, \eta_2) \\ \operatorname{cov}(\eta_1, \eta_2) & D\eta_2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

где обозначено

$$\sigma_1^2 = D\eta_1, \quad \sigma_2^2 = D\eta_2, \quad r = \frac{\operatorname{cov}(\eta_1, \eta_2)}{\sigma_1\sigma_2}$$

(величина r называется коэффициентом корреляции между η_1 и η_2).
Имеем

$$\det C_\eta = (1 - r^2) \sigma_1^2 \sigma_2^2, \quad C_\eta^{-1} = \frac{1}{1 - r^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{r}{\sigma_1 \sigma_2} \\ -\frac{r}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}.$$

Поэтому в двумерном случае плотность нормального закона имеет вид

$$p_\eta(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times \\ \times \exp \left\{ -\frac{1}{2(1-r^2)} \left[\frac{(x_1 - a_1)^2}{\sigma_1^2} - \frac{2r(x_1 - a_1)(x_2 - a_2)}{\sigma_1\sigma_2} + \frac{(x_2 - a_2)^2}{\sigma_2^2} \right] \right\}.$$

Лемма 8.4'. Если вектор $\eta = (\eta_1, \dots, \eta_m, \eta_{m+1}, \dots, \eta_n)$ имеет многомерное нормальное распределение, причем $\text{cov}(\eta_i, \eta_j) = 0$ при $i=1, \dots, m; j=m+1, \dots, n$, то случайные векторы (η_1, \dots, η_m) и $(\eta_{m+1}, \dots, \eta_n)$ независимы между собой.

Доказательство. Правый верхний и левый нижний углы матрицы C_η заполнены нулями. То же самое верно для обратной матрицы C_η^{-1} . Поэтому из теоремы 8.1' следует, что плотность распределения p_η вектора η распадается в произведение плотностей распределений векторов (η_1, \dots, η_m) и $(\eta_{m+1}, \dots, \eta_n)$. Лемма доказана.

З а м е ч а н и е. Таким образом, в нормальном случае если две случайные величины η_1 и η_2 некоррелированы, т. е. $\text{cov}(\eta_1, \eta_2) = 0$, то они и независимы. Этот вывод, однако, верен только в том случае, когда известно, что двумерное распределение (η_1, η_2) нормально. Нормальности каждой из величин η_1 и η_2 для указанного вывода недостаточно.

Лемма 8.5'. Если вектор (η_1, η_2) имеет двумерное нормальное распределение, причем $M\eta_1 = M\eta_2 = 0$, то случайные величины η_2 и $\delta = \eta_1 - \frac{\text{cov}(\eta_1, \eta_2)}{D\eta_2} \eta_2$ независимы между собой.

Доказательство. Имеем

$$\text{cov}(\delta, \eta_2) = \text{cov}(\eta_1, \eta_2) - \frac{\text{cov}(\eta_1, \eta_2)}{D\eta_2} D\eta_2 = 0.$$

Следствие. В случае двумерного нормального распределения имеет место равенство

$$\eta_1 - M\eta_1 = \frac{\text{cov}(\eta_1, \eta_2)}{D\eta_2} (\eta_2 - M\eta_2) + \delta,$$

где величина δ не зависит от η_2 .

Таким образом, если мы знаем величину $\eta_2 = x_2$, то мы можем утверждать, что при этом условии распределение η_1 есть распределение случайной величины $\delta + M\eta_1 + \frac{\text{cov}(\eta_1, \eta_2)}{D\eta_2}(x_2 - M\eta_2)$. Очевидно, что $M\delta = 0$,

$$D\delta = D\eta_1 - \left(\frac{\text{cov}(\eta_1, \eta_2)}{D\eta_2} \right)^2 D\eta_2 = D\eta_1(1 - r^2),$$

где r обозначает коэффициент корреляции между η_1 и η_2 . Иными словами, при известном значении $\eta_2 = x_2$ интуитивно понимаемое нами «условное распределение» величины η_1 является нормальным с математическим ожиданием $M\eta_1 + \frac{\text{cov}(\eta_1, \eta_2)}{D\eta_2}(x_2 - M\eta_2)$ и дисперсией $D\eta_1(1 - r^2)$.

Если отправляться от общепринятого определения условного распределения через отношение плотностей (см. Б. В. Гнеденко [8]), то получится, конечно, тот же самый результат, так как нетрудно проверить, что условное распределение δ при известном η_2 будет, в силу независимости, совпадать с безусловным также и при строгом способе введения условных распределений (иначе сам способ их введения никуда бы не годился).

Следовательно, при известном η_2 возможная неопределенность значения η_1 определяется неопределенностью величины δ . Дисперсия величины δ равна $D\eta_1(1 - r^2)$. Поэтому при r , близком к 1, мы по величине η_2 можем почти точно восстановить η_1 . Именно,

$$\eta_1 - M\eta_1 \approx \frac{\text{cov}(\eta_1, \eta_2)}{D\eta_2}(\eta_2 - M\eta_2).$$

Последнее уравнение называется уравнением регрессии. Многомерный аналог изложенного способа нахождения η_1 по η_2 получил название «регрессионного анализа». Необходимо отметить, однако, что предположение о нормальности многомерных распределений, лежащее в основе регрессионного анализа, редко проверяется, а чаще принимается без проверки. Соответственно, уменьшается доверие к полученным выводам. Мы вернемся к обсуждению этого вопроса при изложении метода наименьших квадратов.

8.2'. Недостатки понятия слабой сходимости. Понятие слабой сходимости, эквивалентное понятию сходимости функций распределения, позволяет сформулировать центральную предельную теорему в наиболее общем виде:

$$P\{A \leq s_n^* \leq B\} \rightarrow \Phi(B) - \Phi(A),$$

где s_n^* — нормированная сумма, A и B — любые числа, Φ — функция Лапласа. Такой способ формулировки этой теоремы в настоящее время является общепринятым; он принят и в основном тексте настоящей книги. Тем не менее, сходство распределения случайной величины s_n^* с нормальным не является полным.

Например, если слагаемые ξ_1, \dots, ξ_n дискретны, т. е. каждое из них принимает лишь счетное число значений, то нормированная сумма s_n^* также дискретна — её распределение сосредоточено на счетном множестве A_n . Между тем нормальное распределение имеет плотность, следовательно, вероятность попадания нормальной случайной величины η в любое счетное множество равно нулю. Итак,

$$P\{s_n^* \in A_n\} = 1, \quad P\{\eta \in A_n\} = 0.$$

Только для интервалов $[A, B]$ имеем

$$\begin{aligned} P\{s_n^* \in [A, B]\} &= P\{A \leq s_n^* \leq B\} \rightarrow \\ &\rightarrow P\{\eta \in [A, B]\} = \Phi(B) - \Phi(A), \end{aligned}$$

где $n \rightarrow \infty$. Следовательно, понятие сходимости, основанное на сходимости функций распределения, требует известной осторожности при его практическом использовании.

Однако по-настоящему яркий пример возможной неудачи при слишком формальном использовании такого понятия сходимости можно привести не в случае одномерных, а в случае двумерных законов распределения. Дело в том, что для одномерных случайных величин обычно приходится интересоваться вероятностью их попадания в интервалы, которая вполне характеризуется функцией распределения. В случае двумерного распределения функция распределения определяет вероятности попадания в прямоугольники со сторонами, параллельными осям координат. Но это далеко не все фигуры на плоскости, которые могут представлять практический интерес. Например, вероятность попадания в круг не выражается через функцию распределения. Таким образом, в двумерном случае само понятие функции распределения не имеет особого смысла. Это и приведет почти к парадоксу в нижеследующем примере. Этот пример показывает, что два распределения на плоскости могут иметь сколь угодно близкие функции распределения и тем не менее быть совершенно не похожими.

Пример 8.1'. Рассмотрим последовательность

$$\zeta_1 = \rho_1 e^{i\varphi_1}, \quad \zeta_2 = \rho_2 e^{i\varphi_2}, \dots, \quad \zeta_n = \rho_n e^{i\varphi_n}, \dots$$

случайных комплексных чисел, где модули $\rho_1, \rho_2, \dots, \rho_n, \dots$ и аргументы $\varphi_1, \varphi_2, \dots, \varphi_n, \dots$ представляют собой две последовательности одинаково распределенных случайных величин (модули и аргументы независимы между собой и друг от друга). Предположим, что аргументы распределены равномерно на отрезке $[0, 2\pi]$. Поставим вопрос об аналоге центральной предельной теоремы для произведения

$$\zeta_1 \zeta_2 \dots \zeta_n = \rho_1 \rho_2 \dots \rho_n \exp\{i(\varphi_1 + \dots + \varphi_n)\}.$$

Ясно, что для логарифма модуля

$$\ln |\zeta_1 \zeta_2 \dots \zeta_n| = \ln \rho_1 + \ln \rho_2 + \dots + \ln \rho_n$$

справедлива центральная предельная теорема, в то время как

$$\exp \{i(\varphi_1 + \dots + \varphi_n)\}$$

имеет равномерное распределение на единичной окружности $|z|=1$, причем модуль и аргумент произведения $\zeta_1 \zeta_2 \dots \zeta_n$ независимы как случайные величины. Поэтому распределение $\zeta_1 \zeta_2 \dots \zeta_n$ не меняется при вращениях вокруг начала координат (любые значения $\exp \{i(\varphi_1 + \dots + \varphi_n)\}$ равновероятны). Если положить $a = M \ln \rho_i$, $\sigma^2 = D \ln \rho_i$ (предполагая, следовательно, существование $M \ln \rho_i$ и $D \ln \rho_i$) и считать, что η_n — нормальная случайная величина с параметрами $(na, \sigma \sqrt{n})$, κ — случайная величина, равномерно распределенная на окружности, то распределение $\zeta_1 \zeta_2 \dots \zeta_n$ может быть аппроксимировано распределением случайной величины

$$e^{\eta_n} \exp(i\kappa).$$

Это и есть аналог центральной предельной теоремы. Теперь, зная правильный ответ, попробуем воспользоваться понятием слабой сходимости. Попытаемся «нормировать», т. е. преобразовать каким-либо образом распределение $\zeta_1 \zeta_2 \dots \zeta_n$ так, чтобы после преобразования двумерные функции распределения сходились к некоторой двумерной функции распределения. Положим

$$X_n = \rho_1 \rho_2 \dots \rho_n \cos(\varphi_1 + \varphi_2 + \dots + \varphi_n);$$

$$Y_n = \rho_1 \rho_2 \dots \rho_n \sin(\varphi_1 + \varphi_2 + \dots + \varphi_n),$$

т. е. $\zeta_1 \zeta_2 \dots \zeta_n = X_n + iY_n$.

Тогда

$$\ln |X_n| = \sum_{i=1}^n \ln \rho_i + \ln |\cos(\varphi_1 + \dots + \varphi_n)|;$$

$$\ln |Y_n| = \sum_{i=1}^n \ln \rho_i + \ln |\sin(\varphi_1 + \dots + \varphi_n)|.$$

Мы видим, что, ища преобразование нормировки, можно нормировать $\ln |X_n|$ и $\ln |Y_n|$, а затем потенцировать получающиеся выражения. Иными словами, кажется возможным нормировать пару (X_n, Y_n) путем перехода к паре (x_n^*, y_n^*) где

$$\begin{aligned} |x_n^*| &= \exp \left\{ \frac{\ln |X_n| - n M \ln \rho_i}{\sqrt{n D \rho_i}} \right\} = \\ &= \exp \left\{ \frac{\sum_{i=1}^n \ln \rho_i - na}{\sigma \sqrt{n}} + \frac{\ln |\cos(\varphi_1 + \dots + \varphi_n)|}{\sigma \sqrt{n}} \right\}, \end{aligned}$$

$$|y_n^*| = \exp \left\{ \frac{\ln |Y_n| - n M \ln \rho_l}{\sqrt{n D \rho_l}} \right\} = \\ = \exp \left\{ \frac{\sum_{l=1}^n \ln \rho_l - na}{\sigma \sqrt{n}} + \frac{\ln |\sin(\varphi_1 + \dots + \varphi_n)|}{\sigma \sqrt{n}} \right\},$$

а знаки x_n^* , y_n^* совпадают соответственно со знаками X_n и Y_n . Посмотрим, что дает такая нормировка.

Поскольку

$$\frac{\ln |\cos(\varphi_1 + \dots + \varphi_n)|}{\sigma \sqrt{n}} \rightarrow 0, \quad \frac{\ln |\sin(\varphi_1 + \dots + \varphi_n)|}{\sigma \sqrt{n}} \rightarrow 0$$

при $n \rightarrow \infty$ в смысле сходимости по вероятности, получаем

$$|P\{|x_n^*| < x, |y_n^*| < y\} - \\ - P\left\{\exp\left[\frac{\sum_{l=1}^n \ln \rho_l - na}{\sigma \sqrt{n}}\right] < \min(x, y)\right\}| \rightarrow 0.$$

Величина $\frac{1}{\sigma \sqrt{n}} \left(\sum_{l=1}^n \ln \rho_l - na \right)$ имеет стремящееся к нормальному $N(0,1)$ распределение. Нетрудно проверить, что случайные величины

$$\{|X_n|, |Y_n|\}, \{\operatorname{sgn} X_n, \operatorname{sgn} Y_n\}$$

независимы. Поэтому независимы и величины

$$\{|x_n^*|, |y_n^*|\}, \{\operatorname{sgn} x_n^*, \operatorname{sgn} y_n^*\},$$

причем, очевидно, $P\{\operatorname{sgn} x_n^* = 1\} = P\{\operatorname{sgn} x_n^* = -1\} = \frac{1}{2}$ и аналогично

для $\operatorname{sgn} y_n^*$. Обозначая через α и β случайные величины, принимающие значения ± 1 с вероятностью $1/2$ независимо друг от друга, а через η случайную величину, независимую от α и β и имеющую распределение $N(0,1)$, получаем: функция распределения пары случайных величин (x_n^*, y_n^*) стремится к функции распределения пары $(\alpha e^\eta, \beta e^\eta)$.

При попытке использовать практически этот результат мы должны будем заменить совместное распределение пары (x_n^*, y_n^*) его предельным значением, т. е. распределением $(\alpha e^\eta, \beta e^\eta)$, и получить приближенное распределение пары X_n, Y_n , выражая X_n и Y_n через x_n^* и y_n^* :

$$X_n^* = \operatorname{sgn} x_n^* \exp\{\sigma \sqrt{n} [\ln |x_n^*| + na]\},$$

$$Y_n = \operatorname{sgn} y_n^* \exp \{ \sigma \sqrt{n} \ln |y_n^*| + na \}$$

и заменяя в полученных выражениях x_n^* на αe^n и y_n^* на βe^n .

Ясно, что при этом функция распределения пары (X_n, Y_n) будет мало отличаться от аппроксимирующей функции распределения (верхняя грань разности между этими функциями распределения будет равна верхней грани разности между функциями распределения пар (x_n^*, y_n^*) и $(\alpha e^n, \beta e^n)$, т. е. будет стремиться к нулю). Однако $|\alpha e^n| = |\beta e^n|$, т. е. полученная аппроксимация распределения пары (X_n, Y_n) будет распределением такой двумерной величины (ξ_1, ξ_2) , что $|\xi_1| = |\xi_2|$, т. е. распределением, сосредоточенным на биссектрисах координатных углов. Однако мы выяснили раньше, что распределение вектора (X_n, Y_n) не меняется при вращениях, т. е. симметрично. Мы видим, что два совершенно непохожих распределения могут иметь близкие функции распределения, хотя аппроксимация одного распределения другим практически бессмысленна.

Необходимы, следовательно, такие теоремы, которые бы устанавливали нормальную аппроксимацию распределения суммы случайных величин не только в смысле разности между функциями распределения, но в каком-то более сильном смысле. Особую роль такие теоремы должны играть в случае многомерных распределений, когда к тому же рассматривается не просто суммирование случайных векторов, а другие операции (как в только что рассмотренном примере — умножение комплексных чисел). Исторически, однако, развитие теории предельных теорем шло по пути обобщения результатов, полученных для обычного суммирования одномерных случайных величин, так что прежде всего надо обратить внимание на соответствующие результаты для этого случая.

Такие результаты называются локальными предельными теоремами теории вероятностей. С самого начала здесь приходится рассматривать отдельно два случая: первый случай — суммы так называемых решетчатых случайных величин (случайные величины называются решетчатыми, если при подходящем линейном преобразовании шкалы отсчета их значения становятся целочисленными); второй случай — суммы величин, имеющих плотность распределения. Для сингулярных случайных величин (т. е. недискретных, но не имеющих и плотности распределения) локальных предельных теорем не известно. С чем мы тут имеем дело — с несовершенством ли математической теории или с непонятными пока свойствами пространства, которое допускает дискретную или непрерывную модель, но почему-то не допускает сингулярной — сказать в настоящее время невозможно. Ясно, конечно, что сингулярные распределения не могут иметь практического значения.

Методически в курс теории вероятностей могут включаться локальные предельные теоремы как вместе с интегральными (т. е. основанными на функциях распределения), так и вместо них.

В последнем случае интегральные предельные теоремы интерпретируются как следствия локальных (например, интегральная теорема Муавра—Лапласа выводится из локальной).

8.3'. Замечание о многомерных предельных теоремах. Понятие преобразования Фурье (характеристической функции) многомерного случайного вектора $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ формулируется естественно: это функция вещественного вектора $t = (t_1, t_2, \dots, t_n)$, задаваемая соотношением

$$f_{\xi}(t) = M e^{i(t, \xi)} = \int_{R^n} e^{i(t, x)} \mu_{\xi}(dx),$$

где (t, x) обозначает скалярное произведение векторов t и x .

Без особых изменений переносятся на многомерный случай свойства характеристических функций, в частности теорема непрерывности (из сходимости характеристических функций следует сходимость законов распределения). Сходимость законов распределения $\mu_n \rightarrow \mu$ понимается, конечно, в слабом смысле: для любой непрерывной ограниченной функции $\varphi(x)$

$$\int \varphi(x) \mu_n(dx) \rightarrow \int \varphi(x) \mu(dx).$$

Понятия вектора математических ожиданий m_{ξ} и матрицы ковариаций C_{ξ} для случайного вектора ξ были определены в п. 8.1'. Характеристическая функция $f_{\xi - m_{\xi}}$ вектора $\xi - m_{\xi}$ с нулевым математическим ожиданием, очевидно, при малых t имеет вид

$$f_{\xi - m_{\xi}}(t) = 1 - \frac{1}{2} (C_{\xi} t, t) + o(|t|^2),$$

где $|t|^2 = t_1^2 + \dots + t_n^2$.

Характеристическая функция нормального распределения с параметрами (a, C) есть

$$\exp \left\{ i(t, a) - \frac{1}{2} (Ct, t) \right\}.$$

Действительно, для стандартного (определение 8.2') нормального вектора ξ характеристическая функция есть $\exp \left\{ -\frac{1}{2} (t, t) \right\}$.

Тогда для вектора $\eta = A\xi + a$ имеем

$$\begin{aligned} M \exp(i(t, \eta)) &= \exp(i(t, a)) \exp\{i(t, A\xi)\} = \\ &= \exp(i(t, a)) \cdot \exp\{i(A't, \xi)\} = \\ &= \exp\{i(t, a)\} \exp\left\{-\frac{(A't, A't)}{2}\right\} = \\ &= \exp\{i(t, a)\} \exp\left\{-\frac{1}{2} (AA't, t)\right\} = \end{aligned}$$

$$= \exp \{i(t, a)\} \exp \left\{ -\frac{1}{2} (Ct, t) \right\},$$

где $C=AA'$ — матрица ковариаций.

Для одинаково распределенных случайных векторов $\xi_1, \xi_2, \dots, \xi_n, \dots$ автоматически получается центральная предельная теорема для нормированной суммы

$$s_n^* = \frac{1}{\sqrt{n}} (\xi_1 + \dots + \xi_n - M\xi_1 - \dots - M\xi_n).$$

Заметим, что нормировка производится путем деления на \sqrt{n} без участия матрицы ковариаций C_ξ , которая может быть вырожденной. Действительно,

$$\begin{aligned} f_{s_n^*}(t) &= \left[f_{\xi_k - M\xi_k} \left(\frac{t}{\sqrt{n}} \right) \right]^n = \\ &= \left[1 - \frac{1}{2n} (Ct, t) + o\left(\frac{1}{n}\right) \right]^n \rightarrow \\ &\rightarrow \exp \left\{ -\frac{1}{2} (Ct, t) \right\}, \quad C = C_{\xi_k}. \end{aligned}$$

Отсюда вытекает сходимость s_n^* к нормальному распределению с параметрами $(0, C)$.

Для неодинаково распределенных случайных векторов $\xi_1, \xi_2, \dots, \xi_n, \dots$ вопрос о центральной предельной теореме существенно усложняется. Дело в том, что матрица ковариаций C_n нормированной суммы s_n^* может не стремиться ни к какому пределу. Поэтому требуется включить в операцию нормировки линейное преобразование с матрицей $C_n^{-\frac{1}{2}}$.

По поводу определения $C_n^{-\frac{1}{2}}$ заметим, что любая функция $f(C)$ от симметрической матрицы C определяется следующим образом: если C привести ортогональным преобразованием к диагональному виду

$$C = U \Lambda U^{-1},$$

где

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ . & . & . & . \\ 0 & 0 & & \lambda_n \end{pmatrix},$$

то
где

$$f(C) = U f(\Lambda) U^{-1}$$

$$f(\Lambda) = \begin{pmatrix} f(\lambda_1) & 0 & \dots & 0 \\ 0 & f(\lambda_2) & \dots & 0 \\ 0 & 0 & \dots & f(\lambda_n) \end{pmatrix}.$$

Однако матрица C_n может быть вырожденной или стремиться к вырожденной при $n \rightarrow \infty$. Это может создать существенные затруднения. В учебной литературе вопрос о многомерной центральной предельной теореме рассматривается лишь в книге С. Н. Бернштейна «Теория вероятностей» [3]. Приоритет в получении таких теорем также принадлежит С. Н. Бернштейну.

Заметим, что несмотря на то, что эти теоремы выражаются на языке слабой сходимости, здесь не возникает ошибочной аппроксимации, вроде рассмотренной в п. 8.2'. Очевидно, наши представления о сущности явления сходимости к нормальному закону глубже, чем их выражение на языке математических теорем.

Замечания к § 9

Большой набор задач на применение центральной предельной теоремы можно найти в задачнике Л. Д. Мешалкина [18]. Однако многие задачи из этого задачника требуют более четкой формулировки. Таковы, например, задачи 225 и 231, которые следует переформулировать в терминах статистической проверки гипотез. Покажем на примере задачи 231, как это сделать.

Формулировка задачи № 231. Многие ботаники делали опыты по скрещиванию желтого (гибридного) гороха. По известной гипотезе Менделя вероятность появления зеленого гороха в таких опытах равна $1/4$. При 34153 опытах скрещивания в 8506 случаях был получен зеленый горох. Подтверждают ли эти данные гипотезу Менделя? Точнее говоря, при статистической проверке этой гипотезы должны ли мы ее отбросить на уровне значимости $\alpha = 0,05$? А на уровне значимости $\alpha = 0,10$?

Решение. Если $h = \frac{8506}{34153} = 0,2490$ — частота успеха, то при верной гипотезе Менделя $p = \frac{1}{4}$ разность

$$(h - p) \frac{1}{\sqrt{p(1-p)}}$$

имеет приблизительно нормальное распределение с математическим ожиданием $a = 0$ и $\sigma = \sqrt{\frac{1}{34153}} \approx 0,0054$.

Критическая область для проверки гипотезы $p = \frac{1}{4}$ на уровне значимости α есть

$$\left\{ h : \frac{|h - p|}{\sqrt{p(1-p)}} > \sigma x_\alpha \right\},$$

где x_α находится из соотношения

$$\Phi(-x_\alpha) + 1 - \Phi(x_\alpha) = \alpha.$$

При найденном значении $h = 0,2490$

$$\frac{|h - p|}{\sigma \sqrt{p(1-p)}} = \frac{0,0010 \cdot 4}{0,0054 \cdot \sqrt{3}} = 0,43.$$

При $\alpha = 0,05$ имеем $x_\alpha = 1,96$, следовательно, на 5% уровне значимости отвергать гипотезу Менделя $p = \frac{1}{4}$ нет оснований. При $\alpha = 0,10$, $x_\alpha = 1,64$, т. е. также не отвергается гипотеза. В сущности мы сделали переход к величине

$$\frac{h - p}{\sigma \sqrt{p(1-p)}},$$

имеющей примерно распределение $N(0, 1)$, и получили из опытных данных для этой величины значение 0,43. Согласие с распределением $N(0, 1)$ следует признать очень хорошим, особенно учитывая огромный объем экспериментального материала.

Дальнейшие примеры очень хорошего согласия экспериментальных данных с гипотезой Менделя читатель может найти, в частности, в работе А. Н. Колмогорова [12].

К п. 9.1. Странно, что правило «трех сигма» иногда рекомендуется для распределений, ничего общего не имеющих с нормальным. Корень этой ошибки — в слишком большой вере в универсальность нормального закона. Над этой верой смеялся еще А. Пуанкаре, когда он писал приблизительно так: «Все верят в нормальный закон: математики — потому, что думают, что физики наблюдают его на опыте; физики же — потому, что думают, что математики способны доказать теоретически, что нормальный закон должен выполняться». Однако правило «трех сигма» проникло в элементарные учебники, например, для экономистов, без упоминания нормальности распределения (и вообще без упоминания статистической устойчивости). С этим заблуждением преподаватель должен бороться.

К п. 9.2. Утвердился методически неправильный прием, когда студентов сначала учат оценивать вероятность

$$P \left\{ \left| \frac{\xi_1 + \dots + \xi_n}{n} - a \right| > \varepsilon \right\} \quad (9.1')$$

с помощью неравенства Чебышева, а потом учат делать то же самое с помощью центральной предельной теоремы. В результате студенты считают эти приемы равноправными. Между тем неравенство Чебышева дает для этой вероятности крайне грубую оценку сверху и притом тем грубее, чем меньше ε . Нормальное приближение дает примерно правильную по порядку оценку. Поэтому следует рассматривать лишь прием, основанный на центральной предельной теореме.

Вероятность (9.1') обычно интерпретируется следующим образом. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые наблюдения некоторой

величины a , которую мы хотим узнать. Если мы предположим, что наши наблюдения обладают свойством

$$M \xi_i = a, \quad (9.2')$$

т. е. отсутствует систематическая ошибка, то вероятность (9.1') есть вероятность отклонения среднего арифметического наших наблюдений от истинного значения a . При $n \rightarrow \infty$ получаем несколько парадоксальный вывод, что величину a можно узнать сколь угодно точно, даже если дисперсия $\sigma^2 = D \xi_i$ отдельного наблюдения велика. Восторг статистика от такого вывода несколько умеряется тем обстоятельством, что точность ε имеет вид

$$\varepsilon \sim (2 \div 3) \frac{\sigma}{\sqrt{n}}.$$

т. е. точность возрастает пропорционально \sqrt{n} . Но на самом деле важнее другое: в рамках чисто статистического подхода ничто не может гарантировать равенство (9.2'), которое является определением отсутствия систематической ошибки. У прибора с большой величиной σ наверняка будет и систематическая ошибка.

Замечания к § 10

Последние три параграфа книги посвящены математической статистике. Надо ясно представлять себе, что квалифицированное применение методов математической статистики требует привлечения специалистов в этой области. Данная же книга рассчитана на математически образованного естествоиспытателя, который не собирается стать специалистом по математической статистике. Однако ему необходимо четко представлять себе, в каких случаях эта наука может принести существенную пользу, и уметь воспользоваться помощью специалиста. Лучший способ для создания такого представления за ограниченное время — подробно разъяснить некоторые частные приемы математической статистики, доведя их до возможности приложений, но оставив в стороне другие приемы за недостатком времени. Сама структура математической статистики позволяет такое исчерпывающее изучение одних приемов за счет полного игнорирования других, так как в этой науке сравнительно небольшую роль играют общие математические теоремы, а отдельные приемы часто логически независимы друг от друга.

В данной книге выбраны для изучения приемы обработки результатов наблюдений. В § 10 излагаются основные приемы работы с выборками.

К п. 10.1. По-видимому, требуется неоднократно подчеркивать, что нельзя на любые результаты измерений x_1, \dots, x_n смотреть как на выборку. Методическим целям хорошо служит следующий со-

физм. Мы говорим, что выборку образуют результаты нескольких независимых измерений, проводимых в одинаковых условиях. Однако если мы контролируем все условия опыта, то у нас всегда будет получаться одно и то же число (не будет никакой неопределенности), а если мы контролируем не все условия опыта, то откуда мы можем знать, что они остаются одинаковыми?

Не существует исчерпывающего способа проверки того, что данные измерения x_1, \dots, x_n можно считать выборкой. Уверенность в правильности соответствующих методов работы приобретается в процессе накопления опыта. Одним из «чудес» теории вероятностей является возможность подбора кривых распределения путем определения их параметров по выборкам. Остановимся подробнее на том, что при этом происходит.

Пусть дана выборка x_1, x_2, \dots, x_n . В § 10 рекомендуется ее графическое представление с помощью эмпирической функции распределения $F_n(x)$. Не менее употребительным приемом является построение гистограммы. Он заключается в следующем. Отрезок оси абсцисс, на котором лежат наблюдения x_1, \dots, x_n , разбивается на несколько одинаковых интервалов A_1, A_2, \dots, A_m . Пусть μ_k — число наблюдений x_1, \dots, x_n , попавших в интервал A_k . Рассмотрим кусочно постоянную функцию, принимающую на интервале A_k значение $\frac{\mu_k}{n}$. График этой функции называется *гистограммой*.

Если $p(x)$ — плотность распределения, то

$$\frac{\mu_k}{n} \approx \int_{A_k} p(x) dx.$$

Предполагая, что на интервале A_k плотность $p(x)$ меняется незначительно, получаем

$$\int_{A_k} p(x) dx \approx p(x_k) l(A_k),$$

где $l(A_k)$ — длина A_k , x_k — любая точка интервала A_k . На этом основании считают иногда, что при большом числе наблюдений гистограмма (построенная в подходящем масштабе по оси ординат) стремится к плотности вероятности $p(x)$. Это неверно: для того чтобы это было так, нужно, чтобы выполнялось довольно сложное соотношение между числом наблюдений n и числом интервалов m , употребляемых для построения гистограммы. Тем не менее, гистограмма по внешнему виду напоминает плотность вероятности.

Заметим, что одним из достоинств эмпирической функции распределения $F_n(x)$ является то, что утверждение $F_n(x) \rightarrow F(x)$ верно без всяких ограничений. Кроме того, построение $F_n(x)$ целесообразно при небольших значениях n , когда строить гистограмму бессмысленно.

При применении обоих способов (как гистограммы, так и эмпирической функции распределения) выборка x_1, x_2, \dots, x_n наглядно характеризуется некоторым графиком. Что же происходит потом? О графике временно забывают; производят некоторые вычисления, обычно вычисления эмпирических моментов:

$$\alpha_k = \frac{1}{n} \sum_{i=1}^n x_i^k,$$

(при $k=1,2$, а иногда и при $k=3,4$). Полученные значения α_k подставляют в некие формулы, например в формулу плотности нормального закона:

$$p(x; a, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

вместо a подставляют α_1 , вместо σ — выражение $\sqrt{\alpha_2 - \alpha_1^2}$. Затем полученную плотность (функцию распределения) строят на том же графике, где была построена гистограмма (эмпирическая функция распределения). В результате гладкая кривая прекрасно приближает эмпирическую кривую. Если согласие оказывается не особенно хорошим, его можно улучшить, привлекая момент более высокого порядка. Например, вместо нормального закона $\Phi(x)$ (для функции распределения) берут закон, выражаемый следующей формулой:

$$\Phi\left(\frac{x-\bar{x}}{s}\right) - \frac{\beta_3}{3!} \Phi^{(3)}\left(\frac{x-\bar{x}}{s}\right) + \\ + \frac{\beta_4 - 3}{4!} \Phi^{(4)}\left(\frac{x-\bar{x}}{s}\right) + \frac{10}{6!} \beta_3^2 \Phi^{(6)}\left(\frac{x-\bar{x}}{s}\right),$$

где

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2,$$

$$\beta_3 = \frac{1}{s^3 n} \sum_{i=1}^n (x_i - \bar{x})^3, \quad \beta_4 = \frac{1}{s^4 n} \sum_{i=1}^n (x_i - \bar{x})^4,$$

а $\Phi^{(k)}(x)$ означает $\frac{d^k}{dx^k} \Phi(x)$. В результате приближение часто становится превосходным (подробнее см. [14]). Очевидно, конечно, что \bar{x} , s^2 , β_3 и β_4 выражаются через α_1 , α_2 , α_3 , α_4 . Таким образом, кажется возможным по крайней мере всю информацию, содержащуюся в выборке, свести к нескольким параметрам, например α_k .

Однако указанное улучшение нормального закона не является единственным средством сглаживания эмпирической функции распределения. Другую систему кривых, определяемых по четырем параметрам, предложил К. Пирсон (кривые Пирсона). Эти кривые часто столь же хорошо приближают эмпирические данные, как и только что указанные (см. [10]).

Имеются и другие типы кривых распределения, которые можно использовать с тем же правом.

Пока речь идет о приближении эмпирического распределения в области не слишком больших и не слишком малых вероятностей, более или менее безразлично, каким семейством кривых пользоваться. Однако часто практически наибольший интерес представляют именно «хвосты» функции распределения. Оценка «хвостов» с помощью разных семейств распределения приводит к резко различным результатам. Следует признать, что не существует имеющих научные основания общих способов для выбора того или иного семейства распределений. По-видимому, все они дают в области «хвостов» одинаково не заслуживающие доверия результаты.

Остановимся подробнее на свойствах эмпирической функции распределения $F_n(x)$. Пусть теоретическая функция распределения $F(x)$ монотонно строго возрастает. Рассмотрим статистику (т. е. функцию от результатов наблюдений) вида

$$\kappa_n = \sqrt{n} \sup_x |F_n(x) - F(x)|.$$

Значение этой статистики не изменится, если от выборки x_1, \dots, x_n перейти к выборке y_1, \dots, y_n , где $y_i = G(x_i)$, причем G — любая монотонно возрастающая функция, если, конечно, теоретическим законом распределения y_i считать

$$P\{y_i < x\} = P\{G(x_i) < x\} = P\{x_i < G^{-1}(x)\} = F(G^{-1}(x)).$$

Полагая $G(x) = F(x)$, мы получим, что величины y_i имеют равномерное распределение на отрезке $[0, 1]$. Следовательно, распределение вероятностей для статистики κ_n не зависит от вида функции $F(x)$.

Асимптотическое выражение

$$\lim_{n \rightarrow \infty} P\{\kappa_n < x\} = K(x)^1$$

найдено А. Н. Колмогоровым. Таблицы функции Колмогорова имеются во всех сборниках статистических таблиц.

Если мы собираемся проверить гипотезу о том, что $F(x)$ есть заданная функция, например функция Лапласа $\Phi(x)$, то мы можем сделать это с помощью статистики

$$\kappa_n = \sup_x |F_n(x) - \Phi(x)|,$$

¹ $K(x)$ — функция Колмогорова.

используя ее асимптотическое распределение $K(x)$ (имеются, впрочем, и таблицы распределений κ_n для конечных n).

Часто неправильно используется аналогичный прием, когда желательно проверить не то, что теоретический закон распределения является заранее заданной функцией, а то, что он принадлежит заданному параметрическому семейству. Например, составляют статистику

$$\kappa'_n = \sup_x \left| F_n(x) - \Phi\left(\frac{x - \bar{x}}{s}\right) \right|$$

и применяют для нее распределение $K(x)$. Это неверно, так как за счет подбора параметров \bar{x} и s по выборке x_1, \dots, x_n мы сильно приближаем закон распределения $\Phi\left(\frac{x - \bar{x}}{s}\right)$ к данной эмпирической функции распределения $F_n(x)$. Между тем, распределение Колмогорова относится к случаю, когда никакого подбора параметров не производится. Выше отмечалось, что за счет подбора четырех параметров можно обычно очень точно подогнать любой закон распределения к любым эмпирическим данным.

Область применения критерия Колмогорова ограничена случаем, когда теоретический закон распределения известен точно.

Замечания к § 11

В этом параграфе в п. 11.1 излагаются основные математические факты, на которых строится статистическая обработка наблюдений при нормально распределенных ошибках. Однако сфера применения этих фактов на самом деле шире и охватывает также асимптотически нормальные распределения. Соответственно шире область применений для распределений Пирсона, Стьюдента и Фишера, особенно для пирсоновского распределения χ^2 . Эти применения мы отчасти изложим в замечаниях к § 11.

11.1'. Применение распределения χ^2 к полиномиальному распределению. Пусть имеется n независимых испытаний, каждое из которых имеет m исходов A_1, A_2, \dots, A_m . Допустим, что в каждом отдельном испытании вероятности этих исходов равны p_1, p_2, \dots, p_m , где $p_i > 0$, $i=1, \dots, m$ и $p_1 + p_2 + \dots + p_m = 1$. Пусть x_1, x_2, \dots, x_m — соответственно, число наступлений событий A_1, A_2, \dots, A_m в n испытаниях. Тогда распределение вектора $x = (x_1, \dots, x_m)$, где, очевидно, $x_i = 0, 1, \dots, n$, причем $x_1 + x_2 + \dots + x_m = n$ называется полиномиальным распределением. Введем случайные векторы

$$\mu_k = (\xi_{k1}, \xi_{k2}, \dots, \xi_{km})$$

следующим образом. Если в k -том испытании произошло событие A_i , то $\xi_{ki} = 1$, а остальные $\xi_{kj} = 0$ при $j \neq i$. Тогда, очевидно,

$$x = \mu_1 + \mu_2 + \dots + \mu_n.$$

Найдем матрицу ковариаций C_x случайного вектора x . Поскольку μ_1, \dots, μ_n — независимые одинаково распределенные случайные векторы,

$$C_x = nC_\mu,$$

где C_μ — матрица ковариаций любого из векторов μ_h . Имеем

$$D\xi_{ki} = M\xi_{ki}^2 - (M\xi_{ki})^2 = p_i - p_i^2 = p_i q_i,$$

где $q_i = 1 - p_i$. Далее, при $i \neq j$

$$\text{cov}(\xi_{ki}, \xi_{kj}) = M(\xi_{ki}\xi_{kj}) - (M\xi_{ki})(M\xi_{kj}) = -p_i p_j,$$

поскольку всегда $\xi_{ki}\xi_{kj} = 0$, а значит, $M(\xi_{ki}\xi_{kj}) = 0$.

Таким образом,

$$C_\mu = \begin{pmatrix} p_1 q_1 & -p_1 p_2 & \dots & -p_1 p_n \\ -p_2 p_1 & p_2 q_2 & \dots & -p_2 p_n \\ \dots & \dots & \dots & \dots \\ -p_n p_1 & -p_n p_2 & \dots & p_n q_n \end{pmatrix}.$$

В силу многомерной центральной предельной теоремы, при большом числе испытаний n вектор

$$\frac{x - Mx}{\sqrt{n}} = \left(\frac{x_1 - np_1}{\sqrt{n}}, \frac{x_2 - np_2}{\sqrt{n}}, \dots, \frac{x_m - np_m}{\sqrt{n}} \right)$$

будет иметь близкое к нормальному распределение с параметрами $(0, C_\mu)$. Однако матрица C_μ вырождена, что отвечает тому факту, что распределение вектора x сосредоточено в гиперплоскости $x_1 + x_2 + \dots + x_m = n$. Поэтому и соответствующее нормальное распределение надо рассматривать в гиперплоскости $\sum y_i = 0$, где через y_i обозначена величина $y_i = \frac{x_i - np_i}{\sqrt{n}}$.

Пирсон сделал замечательное открытие, в силу которого удобно рассматривать не величины y_i , а величины

$$z_i = \frac{y_i}{\sqrt{p_i}} = \frac{x_i - np_i}{\sqrt{np_i}}, \quad i = 1, \dots, m.$$

Распределение величин z_i сосредоточено в гиперплоскости

$$L = \left\{ (z_1, \dots, z_m) : \sum_{i=1}^m z_i \sqrt{p_i} = 0 \right\}.$$

На современном математическом языке открытие К. Пирсона можно выразить следующей теоремой.

Теорема 11.1'. Если распределение вектора $z = (z_1, \dots, z_m)$ рассматривать в гиперплоскости L , то при $n \rightarrow \infty$ оно сходится к стан-

дартному нормальному распределению в L (т. е. к нормальному распределению с параметрами $(0, E)$, E — единичная матрица).

Доказательство. Достаточно показать, что в гиперплоскости L матрица ковариаций вектора z равна E . Пусть вектор $\alpha = (\alpha_1, \dots, \alpha_m)$ лежит в L и длина его равна 1, т. е.

$$\sum_{i=1}^m \alpha_i \sqrt{p_i} = 0, \quad \sum_{i=1}^m \alpha_i^2 = 1.$$

Вычислим

$$D(\alpha, z) = (C_z \alpha, \alpha),$$

где C_z есть матрица ковариаций вектора z , т. е.

$$C_z = \begin{pmatrix} \frac{1}{\sqrt{p_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{p_2}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\sqrt{p_m}} \end{pmatrix} C_\mu = \begin{pmatrix} \frac{1}{\sqrt{p_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{p_2}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\sqrt{p_m}} \end{pmatrix} =$$

$$= \begin{pmatrix} q_1 & -\sqrt{p_1 p_2} & \dots & -\sqrt{p_1 p_m} \\ -\sqrt{p_2 p_1} & q_2 & \dots & -\sqrt{p_2 p_m} \\ \dots & \dots & \dots & \dots \\ -\sqrt{p_m p_1} & -\sqrt{p_m p_2} & \dots & q_m \end{pmatrix}.$$

Имеем

$$(C_z \alpha, \alpha) = (1 - p_1) \alpha_1^2 - \sqrt{p_1 p_2} \alpha_1 \alpha_2 - \dots - \sqrt{p_1 p_m} \alpha_1 \alpha_m +$$

$$+ (1 - p_2) \alpha_2^2 - \sqrt{p_2 p_1} \alpha_2 \alpha_1 - \dots - \sqrt{p_2 p_m} \alpha_2 \alpha_m + \dots$$

$$\dots + (1 - p_m) \alpha_m^2 - \sqrt{p_m p_1} \alpha_m \alpha_1 - \dots - \sqrt{p_m p_{m-1}} \alpha_m \alpha_{m-1} =$$

$$= \alpha_1^2 + \dots + \alpha_m^2 - \alpha_1 \sqrt{p_1} \sum_{i=1}^m \alpha_i \sqrt{p_i} - \alpha_2 \sqrt{p_2} \sum_{i=1}^m \alpha_i \sqrt{p_i} -$$

$$- \dots - \alpha_m \sqrt{p_m} \sum_{i=1}^m \alpha_i \sqrt{p_i} = \alpha_1^2 + \dots + \alpha_m^2 = 1,$$

что и требовалось доказать.

Следствие. При $n \rightarrow \infty$ распределение величины

$$\sum_{i=1}^m \frac{(x_i - np_i)^2}{(\sqrt{np_i})^2} = \sum_{i=1}^m \frac{(x_i - np_i)^2}{np_i} = \sum_{i=1}^m z_i^2 = (z, z)$$

стремится к распределению χ_{m-1}^2 , т. е. к распределению χ^2 с $(m-1)$ степенями свободы.

Доказательство. Размерность гиперплоскости L равна $(m-1)$, распределение вектора z стремится (при больших n) к стандартному нормальному распределению в L в смысле слабой сходимости. Отсюда вытекает, что $\mathbf{P}\{(z, z) < R\}$ стремится к вероятности $\mathbf{P}\{\xi_1^2 + \dots + \xi_{m-1}^2 < R\}$, где ξ_i независимы и имеют распределение $N(0,1)$. Следствие доказано.

Критерий χ^2 . Пусть из каких-то соображений мы считаем, что вероятности p_1, \dots, p_m для данной полиномиальной схемы равны заданным числам p_1^0, \dots, p_m^0 , и хотим проверить эту гипотезу. Очевидно, с этой целью мы должны составить какую-то статистику с известным распределением, которая измеряла бы отклонение вектора результатов наблюдений (x_1, x_2, \dots, x_m) от его «теоретического» математического ожидания, т. е. вектора $np_1^0, np_2^0, \dots, np_m^0$. Только что доказанное следствие говорит нам, что если в качестве этой меры отклонения взять величину

$$\chi^2 = \sum_{i=1}^m \frac{(x_i - np_i^0)^2}{np_i^0},$$

то распределение этой величины есть приближенно χ_{m-1}^2 . Мы получили возможность проверять по экспериментальным данным гипотезу $p_1 = p_1^0, p_2 = p_2^0, \dots, p_m = p_m^0$.

11.2. Критерий χ^2 в случае зависящих от параметров вероятностей. Основные применения критерия χ^2 относятся к случаю, когда вероятности p_1, p_2, \dots, p_m заранее не известны, но определяются некоторым числом $k < m-1$ параметров: $p_i = p_i(a_1, \dots, a_k)$. Типичный путь возникновения такой ситуации состоит в следующем. Пусть дана выборка ξ_1, \dots, ξ_n из какого-то теоретического закона распределения $F(x; a_1, \dots, a_k)$, зависящего от параметров a_1, \dots, a_k . Разобьем вещественную ось на интервалы

$$(-\infty, c_1), [c_1, c_2), \dots, [c_{m-2}, c_{m-1}), [c_{m-1}, \infty)$$

(всего m интервалов) и обозначим через x_i число значений величин ξ_1, \dots, ξ_n , попавших в интервал с номером i : x_i есть число таких ξ_j , что $c_{i-1} \leq \xi_j < c_i$ (где $c_0 = -\infty, c_m = \infty$). Очевидно, что на вектор (x_1, \dots, x_m) можно смотреть как на результат n полиномиальных испытаний с вероятностями успеха

$$p_i = p_i(a_1, \dots, a_k) = F(c_i; a_1, \dots, a_k) - F(c_{i-1}; a_1, \dots, a_k).$$

Параметры a_1, \dots, a_k обычно неизвестны. Воспользуемся следующим способом их определения. Составим статистику

$$\chi^2 = \chi^2(a_1, \dots, a_k) = \sum_{i=1}^m \frac{[x_i - np_i(a_1, \dots, a_k)]^2}{np_i(a_1, \dots, a_k)}$$

и найдем числа $\hat{a}_1, \dots, \hat{a}_k$, реализующие минимум этой статистики:

$$\chi^2(\hat{a}_1, \dots, \hat{a}_k) = \min_{a_1, \dots, a_k} \chi^2(a_1, \dots, a_k).$$

Этот метод получения оценок $\hat{a}_1, \dots, \hat{a}_k$ для параметров a_1, \dots, a_k называется методом минимума χ^2 . Рассмотрим распределение случайной величины $\chi^2(\hat{a}_1, \dots, \hat{a}_k)$ в предположении, что семейство законов распределения $F(x; a_1, \dots, a_k)$ указано верно, т. е. существуют такие истинные значения параметров a_1^*, \dots, a_k^* , что выборка ξ_1, \dots, ξ_n в самом деле извлечена из распределения $F(x; a_1^*, \dots, a_k^*)$. Нетрудно показать (см. Б. Л. ван дер Варден [6], § 51), что распределение величины $\chi^2(\hat{a}_1, \dots, \hat{a}_k)$ при больших n близко к распределению величины

$$\tilde{\chi}^2 = \sum_{i=1}^m \frac{[x_i - np_i(\hat{a}_1, \dots, \hat{a}_k)]^2}{np_i(a_1^*, \dots, a_k^*)},$$

где в знаменатель подставлены истинные значения параметров. Имеем

$$\tilde{\chi}^2 = \sum_{i=1}^m \left[\frac{x_i - np_i(a_1^*, \dots, a_k^*)}{\sqrt{np_i(a_1^*, \dots, a_k^*)}} - \frac{np_i(\hat{a}_1, \dots, \hat{a}_k) - np_i(a_1^*, \dots, a_k^*)}{\sqrt{np_i(a_1^*, \dots, a_k^*)}} \right]^2.$$

По доказанному в пункте 11.1', величины

$$z_i = \frac{x_i - np_i(a_1^*, \dots, a_k^*)}{\sqrt{np_i(a_1^*, \dots, a_k^*)}}$$

имеют приблизительно стандартное нормальное распределение в гиперплоскости

$$L = \{z = (z_1, z_2, \dots, z_m) : \sum_{i=1}^m z_i \sqrt{p_i(a_1^*, \dots, a_k^*)} = 0\}.$$

Рассмотрим поверхность

$$w(a_1, \dots, a_k) = (w_1(a_1, \dots, a_k), \dots, w_m(a_1, \dots, a_k)),$$

заданную параметрическим уравнением

$$w_i(a_1, \dots, a_k) = \frac{np_i(a_1, \dots, a_k) - np_i(a_1^*, \dots, a_k^*)}{\sqrt{np_i(a_1^*, \dots, a_k^*)}}.$$

Очевидно, что эта поверхность лежит в L . Заметим, что можно считать, что оценки $\hat{a}_1, \dots, \hat{a}_k$ выбираются из условия минимума

$$\tilde{\chi}^2(a_1, \dots, a_k) = \sum_{i=1}^m (z_i - w_i(a_1, \dots, a_k))^2,$$

т. е.

$$\tilde{\chi}^2 = \min_{a_1, \dots, a_k} \tilde{\chi}^2(a_1, \dots, a_k).$$

Таким образом, точка $\omega(\hat{a}_1, \dots, \hat{a}_k)$ есть точка поверхности $\omega = \omega(a_1, \dots, a_k)$, лежащая ближе всего к точке $z = (z_1, \dots, z_m)$. В предположении, что поверхность можно заменить линейным многообразием той же размерности k , получаем: *величина $\tilde{\chi}^2$ имеет распределение χ^2_{m-k-1} .*

Таким образом, если мы определили k неизвестных параметров методом минимума χ^2 , то мера расхождения

$$\chi^2(\hat{a}_1, \dots, \hat{a}_k) = \sum_{i=1}^m \frac{(x_i - np_i(\hat{a}_1, \dots, \hat{a}_k))^2}{np_i(\hat{a}_1, \dots, \hat{a}_k)}$$

между теоретическим и выборочным распределением имеет асимптотическое распределение χ^2_{m-k-1} , в котором число степеней свободы уменьшено на k единиц.

Этот результат замечательно прост и удобен для приложений.

Заметим, что мы не дали его полного доказательства в основном потому, что не доказали возможности замены поверхности $\omega = \omega(a_1, \dots, a_k)$ линейным многообразием. Для доказательства возможности такой замены нам пришлось бы прежде всего доказать, что оценки $\hat{a}_1, \dots, \hat{a}_k$ близки к истинным значениям параметров a_1^*, \dots, a_k^* . Это завело бы нас слишком далеко. Интересующийся читатель может познакомиться с этим доказательством по книге Б. Л. ван дер Вардена [6]. Полное же доказательство вида асимптотического распределения $\chi^2(\hat{a}_1, \dots, \hat{a}_k)$ приведено в книге Г. Крамера [14].

Практически оказывается, что применение метода минимума χ^2 для нахождения оценок $\hat{a}_1, \dots, \hat{a}_k$ неудобно. Однако замена этих оценок обычно применяемыми оценками по моментам распределения может привести к нарушению распределения χ^2_{m-k-1} для меры расхождения между теоретическим и выборочным распределением. Об этом необходимо помнить при практической работе. Интересующийся читатель вновь отсылается к книгам [14] и [6].

Заметим, что только что изложенные свойства меры расхождения позволяют, наконец, понять, почему индекс n у распределения χ^2_n называется «числом степеней свободы»: определение k параметров сокращает число степеней свободы на k единиц.

11.3'. Замечания к п. 11.2. Здесь на языке линейной алгебры выражена чрезвычайно общая модель, пригодная в самых различных случаях. В данной книге мы оставляем в стороне дисперсионный анализ, также основанный на линейной модели (изложение дисперсионного анализа с такой точки зрения имеется в книге Г. Шеффе «Дисперсионный анализ» [27]).

11.4'. Замечания об обработке нормальных выборок. При способе изложения, принятом в § 11, модель выборки является частным случаем линейной модели. Основное утверждение о независимости \bar{x} и s^2 оказывается частным случаем общей теории. Однако можно было бы его доказать отдельно.

Через \bar{x} , s^2 выражаются все статистики, применяемые при оценке параметров и проверке гипотез для нормальных выборок. Таким образом, в предположении нормальности, вся информация, содержащаяся в выборке (x_1, x_2, \dots, x_n) , может быть редуцирована к трем числам: n (число наблюдений), \bar{x} , s . Статистики \bar{x} и s называются «достаточными статистиками». Общее определение этого понятия читатель может найти в любом подробном учебнике математической статистики.

На практических занятиях должны быть рассмотрены примеры использования таблиц распределений Пирсона, Стьюдента и Фишера для обработки нормальных выборок.

Проверка тех или иных статистических гипотез имеет смысл тогда, когда она делается много раз на независимом экспериментальном материале. Допустим, что мы проверили гипотезу $a=a_0$ на материале m независимых выборок. В результате выбора уровня значимости α мы каждый раз отвергаем или не отвергаем проверяемую гипотезу. Спрашивается, как истолковать результаты m независимых проверок одной и той же гипотезы. Если m велико, то даже при малом α и верной гипотезе мы будем иногда ее отвергать. Наоборот, если гипотеза неверна, то вполне возможно, что при большом m мы несколько раз ее не отвергнем. Для объединения результатов m независимых проверок одной и той же гипотезы применяется следующий прием.

Как известно, для проверки гипотезы составляется некоторая статистика ζ , распределение которой $F_\zeta(x)$ известно. Критическая область обычно имеет вид $\{\zeta > x_\alpha\}$, где $P\{\zeta > x_\alpha\} \leq \alpha$, α — заранее назначенный уровень значимости. Например, при проверке гипотезы $a=a_0$ при односторонней альтернативе $\zeta = \frac{1}{s}(\bar{x} -$

$-a_0)V\bar{n}$, а при двусторонней альтернативе $\zeta = \frac{1}{s}|\bar{x} - a_0|V\bar{n}$.

Назовем предельным уровнем значимости β тот уровень значимости, при котором еще можно было бы принять гипотезу при данном значении статистики ζ . Иначе говоря,

$$\beta'_\zeta = \beta'_\zeta(\zeta) = 1 - F_\zeta(\zeta).$$

Пусть $F_\zeta(x)$ — непрерывная функция. Тогда случайная величина $\beta = \beta(\zeta)$ имеет, как легко видеть, равномерное распределение на отрезке $[0,1]$. Если в результате n независимых проверок мы получили значения $\beta_1, \beta_2, \dots, \beta_n$, то (при верной проверяемой гипотезе) эти значения будут образовывать выборку из равномерного зако-

на на отрезке $[0,1]$. Таким образом, соединение результатов нескольких независимых проверок гипотезы (или даже разных гипотез) сводится к проверке того, что $\beta_1, \beta_2, \dots, \beta_n$ образуют выборку из равномерного закона. Ошибочность проверяемой гипотезы выразится в том или ином отклонении от равномерности распределения.

Поскольку теоретический закон распределения β_1, \dots, β_n точно известен, можно применить критерий Колмогорова (см. [8]). Другой возможный способ проверки равномерности основан на том, что $M\beta_i = \frac{1}{2}$, $D\beta_i = \frac{1}{12}$ и, следовательно, величина

$$\sqrt{\frac{12}{n}} \left(\beta_1 + \beta_2 + \dots + \beta_n - \frac{1}{2} n \right)$$

имеет при больших n стандартное нормальное распределение $N(0,1)$. Заметим, что в случае равномерного закона распределения хорошая близость к нормальному закону достигается уже при n порядка 4—5. Третий способ основан на том, что, как нетрудно проверить, явно вычисляя плотность распределения, величина

$$-2 \ln \beta_i$$

имеет распределение χ^2_2 . Следовательно, сумма

$$-2 \sum_{i=1}^n \ln \beta_i$$

имеет распределение χ^2_{2n} .

Какой способ проверки равномерности распределения выбрать — зависит от того, каких именно отклонений от равномерности мы ждем, если проверяемая гипотеза неверна (а верна одна из альтернативных гипотез). Обычно в таком случае распределение предельного уровня значимости изменится так, что малые значения (близкие к нулю) станут более вероятными, чем полагалось бы при равномерном распределении. Такое отклонение лучше чувствуют критерии, основанные на нормальном законе или на распределении χ^2_{2n} , чем критерий, основанный на распределении Колмогорова. Зато при работе с критерием Колмогорова полагается вычертить эмпирическую функцию распределения для β_1, \dots, β_n , содержащую всю ту информацию, которая есть в выборке β_1, \dots, β_n , в то время как при работе с двумя другими критериями мы имеем дело лишь с результатами вычисления статистик.

Например, если данные были подогнаны так, чтобы все время принимать гипотезу на 5%-ном уровне значимости, мы увидим неразумно большой скачок эмпирической функции распределения

в узкой области около абсциссы 0,05. При использовании только статистик мы можем не заметить таких особенностей.

Рассмотрим на этом же примере вопрос об отбраковке резко выделяющихся результатов. Статистическая однородность — феномен редкий и тонкий, и его легко испортить. Поэтому отбраковка резко выделяющихся результатов, которые (предположительно) явились следствием временного нарушения статистической однородности, является делом необходимым. Правда, при этом резко понижается доверие к тому, что оставшиеся результаты образуют статистически однородную совокупность. Ситуация, в которой доверие отчасти восстанавливается, должна быть следующей: до отбраковки применение статистических критериев указывает высоко значимое нарушение основной гипотезы; после отбраковки применение тех же критериев с учетом отбрасывания дает совершенно незначимый результат.

Пусть, например, $\beta_{(1)}, \beta_{(2)}, \dots, \beta_{(n)}$ — значения предельных уровней значимости, записанные в порядке возрастания. Тогда функция распределения для $\beta_{(1)} = \min(\beta_1, \dots, \beta_n)$ вычисляется следующим образом:

$$\begin{aligned} P\{\beta_{(1)} < x\} &= 1 - P\{\beta_{(1)} \geq x\} = \\ &= 1 - P\{\beta_1 \geq x, \beta_2 \geq x, \dots, \beta_n \geq x\} = 1 - (1 - x)^n. \end{aligned}$$

Если значение $\beta_{(1)}$ слишком маловероятно при такой функции распределения, то мы можем считать, что $\beta_{(1)}$ получено в тот момент, когда нарушилась статистическая однородность, так что должно быть отброшено. Однако, отбросив $\beta_{(1)}$, мы должны проверить равномерность распределения оставшихся значений $\beta_{(2)}, \dots, \beta_{(n)}$ не на отрезке $[0, 1]$, а на отрезке $[\beta_{(1)}, 1]$ (например, путем линейного преобразования последнего отрезка в отрезок $[0, 1]$). Действительно, значение $\beta_{(1)}$ могло оказаться малым в силу двух причин: 1) в соответствующих опытах чисто случайно появилось малое значение; 2) имелось отклонение от статистической однородности.

Если имела место первая причина, то, как легко видеть, для выборки $\beta_{(2)}, \dots, \beta_{(n)}$ теоретическим законом распределения является равномерный закон на отрезке $[\beta_{(1)}, 1]$.

Если имела место вторая причина, то мы были бы вправе проверять равномерность на отрезке $[0, 1]$. Поскольку первой причины нельзя полностью исключить, то проверка равномерности на $[\beta_{(1)}, 1]$ является возможной. Но тогда из двух критериев исследователь должен выбрать более жесткий. При линейном преобразовании отрезка $[\beta_{(1)}, 1]$ в отрезок $[0, 1]$ все значения $\beta_{(2)}, \dots, \beta_{(n)}$ уменьшаются. Поэтому более жестким критерием является критерий равномерности на отрезке $[\beta_{(1)}, 1]$.

Во всяком случае, если мы объявляем числа $\beta_{(2)}, \dots, \beta_{(n)}$ выборкой из равномерного закона на отрезке $[0, 1]$, то числа $\beta_{(3)}, \dots$,

..., $\beta_{(n)}$ должны образовывать выборку из равномерного закона на отрезке $[\beta_{(2)}, 1]$.

Иными словами, если отбрасывается наименьшее значение $\beta_{(1)}$, мы рекомендуем применять более жесткие приемы контроля равномерности распределения оставшихся элементов выборки. В противном случае, отбросив достаточное число членов $\beta_{(1)}, \beta_{(2)}, \dots, \beta_{(k)}$, принимающих слишком малые значения, мы всегда сможем принять гипотезу о равномерности распределения.

Мы не будем рассматривать вопрос об отбраковке резко выделяющихся элементов выборки из нормального закона. Соответствующие приемы можно найти в таблицах Л. Н. Большева и С. В. Смирнова [4] вместе с практическими рекомендациями. При этом речь идет обычно о выделении не более чем одной трубой ошибки в выборке. Подробное и элементарное изложение приемов отбраковки можно найти в книге Арлея и Буха [2].

Замечания к § 12

12.1'. Замечания к пунктам 12.1 и 12.2. В этих пунктах рассматриваются две классические задачи — уравнивания измерений и определения параметров по наблюдениям. Метод решения этих задач предложен по существу Гауссом, но для оценки точности получаемого решения полезно применение достижений математической статистики нашего века, в частности распределений χ^2 Пирсона и Фишера.

Простейшей задачей уравнивания измерений является следующая. Пусть x_1, x_2, x_3 — три измерения углов треугольника a_1, a_2, a_3 , причем $a_1 + a_2 + a_3 = \pi$. Поскольку с вероятностью 1 $x_1 + x_2 + x_3 \neq \pi$, то считать x_1, x_2, x_3 истинными значениями a_1, a_2, a_3 невозможно и требуется уравнивать измерения x_1, x_2, x_3 . Лаплас полагал, что в качестве приближения к истинным значениям a_1, a_2, a_3 надо взять такие значения y_1, y_2, y_3 , что $y_1 + y_2 + y_3 = \pi$ и сумма $|x_1 - y_1| + |x_2 - y_2| + |x_3 - y_3|$ минимальна. Гаусс показал, что в случае нормального распределения ошибок измерений метод максимального правдоподобия приводит к минимизации

$$(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2,$$

т. е. к методу наименьших квадратов. Аргументы Гаусса в пользу метода наименьших квадратов подчеркивают большую гармоничность этого метода по сравнению с методом Лапласа, в частности простоту и удобство вычислений. Однако простота вычислений не является единственным аргументом Гаусса, так что вряд ли есть смысл пересматривать эту старую дискуссию теперь, когда применение ЭВМ могло бы позволить численно реализовать метод Лапласа (впоследствии сам Лаплас принял метод Гаусса).

В геодезии уравнивают измерения углов большого количества треугольников (с учетом того, что эти треугольники приблизительно сферические, а не плоские). При этом измерения проводятся при различных условиях, в то время как в модели метода наименьших квадратов предполагается равенство дисперсий отдельных наблюдений, вряд ли совместимое с различными условиями наблюдений. Неправильная оценка весов наблюдений, естественно, может поставить под сомнение результаты уравнивания.

Вся теория уравнивания измерений основана на предположении, что нелинейные связи между переменными можно заменить линейными. Иными словами, предполагается, что квадратами ошибок измерения можно беспрепятственно пренебречь. В предположении линейности связей применяется общая линейная модель (§ 11), и доверительная область для вектора параметров $a = (a_1, a_2, \dots, a_n)$ получается применением распределения Фишера. Действительно, в обозначениях п. 12.1 имеем: оценка для вектора a есть вектор $x' = \text{proj}_M x$,

$$\|x' - a\|^2 = \|\text{proj}_M \delta\|^2,$$

где $\delta = (\delta_1, \dots, \delta_n)$ — вектор ошибок наблюдения. Рассмотрим величину

$$\begin{aligned} \|x - x'\|^2 &= \|a + \delta - \text{proj}_M(a + \delta)\|^2 = \\ &= \|\delta - \text{proj}_M \delta\|^2 = \|\text{proj}_{M'} \delta\|^2, \end{aligned}$$

где M' есть ортогональное дополнение к M .

В силу следствий основной леммы из § 11, отношение

$$\frac{\frac{1}{\dim M} \|x' - a\|^2}{\frac{1}{n - \dim M} \|x - x'\|^2} = \frac{\frac{1}{n - m} \|\text{proj}_M \delta\|^2}{\frac{1}{m} \|\text{proj}_{M'} \delta\|^2} = F_{n-m, m}$$

имеет распределение Фишера $F_{n-m, m}$. Поэтому доверительная область для a имеет вид шара:

$$\mathbf{P} \left\{ \|x' - a\| \leq f_\alpha \frac{\dim M}{n - \dim M} \|x - x'\| \right\} = 1 - \alpha,$$

где f_α определяется из соотношения

$$\mathbf{P}\{F_{n-m, m} \leq f_\alpha\} = 1 - \alpha.$$

Однако доверительная область в виде шара хороша только в том случае, когда все параметры a_1, \dots, a_n равноправны. Это не обязательно так; например, это не так в примере 12.1, когда нас интересуют веса грузов a_1 и a_2 и не интересует сумма их весов

$a_3 = a_1 + a_2$. Оценки для параметров a_1 и a_2 оказываются зависимыми случайными величинами, и для построения доверительной области для (a_1, a_2) следовало бы привлечь общую теорию многомерного нормального распределения. Покажем, как это следует сделать в общем случае.

Пусть $x' = (x'_1, \dots, x'_n) = \text{proj}_M x$ и мы желаем построить доверительную область для параметров a_1, a_2, \dots, a_k , где $k < n$. Легко видеть, что вектор (x'_1, \dots, x'_k) имеет k -мерное нормальное распределение с вектором математических ожиданий (a_1, a_2, \dots, a_k) и с некоторой матрицей ковариаций вида $\sigma^2 C$, где $\sigma^2 = D x_i$. Действительно, при любом $j = 1, \dots, k$ величина

$$x'_j = b_{j0} + \sum_{s=1}^n b_{js} x_s = b_{j0} + \sum_{s=1}^n b_{js} (a_s + \delta_s)$$

является линейной формой от $\delta_1, \dots, \delta_n$, откуда вытекает, что элементы матрицы ковариаций пропорциональны σ^2 . Следовательно, полагая $\tilde{a} = (a_1, \dots, a_k)$, $\tilde{x} = (x'_1, \dots, x'_k)$, имеем: вектор

$$C^{-\frac{1}{2}} (\tilde{x} - \tilde{a})$$

имеет нормальное распределение с параметрами $(0, \sigma^2 E)$. В таком случае величина

$$C^{-\frac{1}{2}} (\tilde{x} - \tilde{a}), C^{-\frac{1}{2}} (\tilde{x} - \tilde{a}) = (C^{-1} (\tilde{x} - \tilde{a}), \tilde{x} - \tilde{a}),$$

имеет распределение $\sigma^2 \chi_k^2$, причем она не зависит от величины

$$\|x - x'\|^2 = \|\text{proj}_{M^\perp} \delta\|^2.$$

Таким образом, получаем, что выражение

$$\frac{1}{k} (C^{-1} (\tilde{x} - \tilde{a}), \tilde{x} - \tilde{a}) \left[\frac{1}{n - \dim M} \|x - x'\|^2 \right]^{-1}$$

имеет распределение Фишера $F_{k, m}$. Это дает возможность построить доверительную область для \tilde{a} вида

$$(C^{-1} (\tilde{x} - \tilde{a}), \tilde{x} - \tilde{a}) \leq f_\alpha \frac{k}{n - \dim M} \|x - x'\|^2.$$

Таким образом, доверительная область имеет вид эллипсоида.

Задача об определении параметров из эксперимента, рассмотренная в п. 12.2, очень похожа на задачу уравнивания измерений. Можно было бы даже сформулировать ее в точности так же, но для применений это неестественно.

Принципиальная сторона этих задач исчерпывается применением общей линейной модели. Однако для освоения практических

приемов работы необходимы соответствующие упражнения. Практические приемы здесь могут быть довольно сложными, так что конкретизация решения, записанного у нас в терминах линейных подпространств и проекций, может представить большие затруднения. Среди имеющихся руководств по методу наименьших квадратов можно рекомендовать книгу Ю. В. Линника [15].

12.2'. Общая задача об определении одной величины по другой. Проведение многочлена методом наименьших квадратов интерпретировалось в п. 12.3 как отделение «истинной», или «регулярной», закономерности от «случайных», или «нерегулярных», ошибок измерений. Важно заметить, что та же вычислительная процедура применяется и для других целей в задачах, основанных на других вероятностных моделях. При этом результат вычислений нужно, естественно, интерпретировать по-другому. Сейчас мы и рассмотрим эти другие возможности.

В п. 12.3 принималась следующая модель:

$$x_i = P_m(t_i) + \delta_i,$$

где x_i — значение наблюдаемой переменной, а t_i — значение переменной, определяющей условия опыта. Значения t_1, t_2, \dots, t_n могли быть взяты произвольно, в частности (но не обязательно) образовывать выборку из некоторого распределения, но, вообще говоря, в этих значениях не предполагалось ничего случайного: это просто некоторые известные числа.

Сейчас мы рассмотрим вопрос о совместной обработке n пар наблюдений

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), \quad (12.1')$$

где возможны другие вероятностные предположения, а потому мы изменяем обозначения.

Мы постараемся поставить вопрос о том, что можно сказать об одной из переменных (x, y) , если известно значение другой.

Простейший случай — когда переменные x и y связаны функциональной связью: $y=f(x)$ или $F(x, y)=0$. В этом случае, восстановив (в результате более или менее трудной работы) вид функций f или F , мы можем по значению одной переменной точно или с некоторым выбором восстановить значение другой. Очевидно, что операция такого восстановления есть решение соответствующего уравнения (либо прямое нахождение y по формуле $y=f(x)$). Для восстановления функции f или F могут потребоваться наблюдения (12.1'). Читатель легко представит себе пример, когда x есть сила, y — ускорение, а для определения массы нужен опыт.

Этот случай есть случай полной устойчивости эксперимента, рассмотренный в § 1. Возможен случай, когда в эксперименте нет никакой устойчивости, но этим случаем неинтересно заниматься,

так как тогда ничего определенного сказать нельзя. Нас интересует только случай, когда в том или ином виде предполагается статистическая устойчивость, т. е. имеет смысл говорить о тех или иных распределениях вероятностей.

Возможно симметричное предположение о наличии совместного распределения x и y . Однако это предположение не охватывает только что указанного случая

$$x_i = P_m(t_i) + \delta_i. \quad (12.2')$$

Поэтому мы введем менее ограничительное условие, согласно которому имеет смысл говорить о распределении переменного y для каждого известного значения x (это распределение может быть или не быть¹ условным распределением). Например, в случае модели (12.2') распределение x при известном t есть нормальное распределение $N(P_m(t), \sigma)$, где

$$\sigma^2 = D\delta_i.$$

Можно было бы требовать существования условного распределения y вектора (y_1, \dots, y_n) при условии, что задан вектор (x_1, \dots, x_n) , но в простейших случаях это есть просто распределение независимых случайных величин, а более сложные случаи требуют привлечения теории случайных процессов, которая здесь не рассматривается.

Если распределения y при различных x нам известны, то для каждого α , $0 \leq \alpha \leq 1$ мы можем найти величину $y_\alpha(x)$ из соотношения

$$P\{y < y_\alpha(x)/x\} = \alpha.$$

Величина $y_\alpha(x)$ называется α -квантилью (для простоты мы будем считать, что все эти уравнения относительно $y_\alpha(x)$ имеют единственное решение).

Представим себе график (рис. 12.1'), на котором нанесены кривые $y_\alpha(x)$ для достаточного числа значений α . Как можно использовать этот график для получения информации об y , если известно значение x ? Ответ очень прост: с заданной вероятностью $1-2\alpha$ имеет место равенство

$$y_\alpha(x) \leq y < y_{1-\alpha}(x).$$

Например, при $x=x_0$ значение y с вероятностью 0,98 лежит в заштрихованном интервале AB .

Сложнее сообразить, как по известному $y=y_0$ указать доверительный интервал или доверительную область для $x=x_0$. Однако и в этом случае ответ прост: доверительной областью является

¹ Оно будет условным распределением только в том случае, когда имеет смысл говорить о совместном распределении x и y .

$$U_{1-2\alpha} = \{x: y_{\alpha}(x) \leq y_0 < y_{1-\alpha}(x)\}.$$

Действительно (при $\alpha < \frac{1}{2}$),

$$\begin{aligned} P\{x_0 \notin U_{1-2\alpha}\} &= P\{(y_0 < y_{\alpha}(x_0) \cup (y_0 \geq y_{1-\alpha}(x_0)))\} = \\ &= P\{y_0 < y_{\alpha}(x_0)\} + P\{y_0 \geq y_{1-\alpha}(x_0)\} = 2\alpha \end{aligned}$$

(в этих формулах знак вероятности P обозначает вероятность при известном $x=x_0$).

Следовательно,

$$P\{x_0 \in U_{1-2\alpha}\} = 1 - 2\alpha.$$

Если функции $y_{\alpha}(x)$ монотонны, то доверительной областью для x_0 является интервал, концы которого $x_0^{(1)}$ и $x_0^{(2)}$ суть решения уравнений

$$y_0 = y_{\alpha}(x_0^{(1)}), \quad y_0 = y_{1-\alpha}(x_0^{(2)}).$$

Так, 98% доверительной областью для x_0 на рисунке 12.1' будет интервал CD .

Пример 12.1'. Пусть случайная величина μ имеет показательное распределение с параметром λ . Известно, что в опыте μ приняла значение y . Дать доверительный интервал для λ .

Решение. Очевидно, что существует

$$P\{\mu < z | \lambda\} = 1 - e^{-\lambda z}.$$

Поэтому концами $\lambda^{(1)}, \lambda^{(2)}$ доверительного интервала для λ будут решения уравнений

$$1 - e^{-\lambda^{(1)}y} = \alpha, \quad 1 - e^{-\lambda^{(2)}y} = 1 - \alpha.$$

Замечание. Слегка видоизменяя указанный способ решения для дискретных распределений, можно построить доверительные интервалы для параметров любых распределений, зависящих от одного параметра (например, для биномиального и пуассоновского распределений, а также для выборочного коэффициента корреляции¹). Соответствующие таблицы и графики имеются в таблицах Я. Янко [29] и Л. Н. Большева и С. В. Смирнова [4].

Практически вряд ли можно требовать чего-либо лучшего для восстановления одной из переменных x или y по известной другой, чем график вида, приведенного на рис. 12.1'. Вопрос состоит

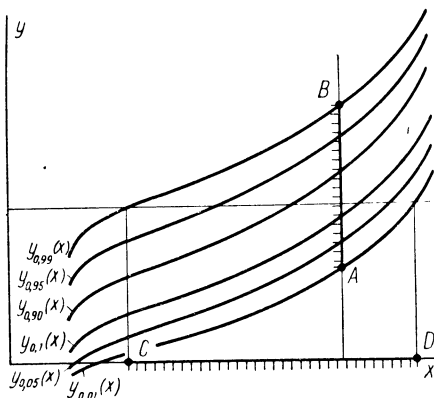


Рис. 12.1'

¹ Определение выборочного коэффициента корреляции см. ниже.

в том, как построить подобный график, не зная заранее законов распределения y при каждом x , а имея лишь наблюдения $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Эффективно это может быть выполнено (кроме уже рассмотренного случая), когда x является параметром распределения вероятностей для y , лишь в нескольких частных случаях. Конечно, речь идет лишь о приближенном построении чертежа 12.1'.

Один из этих случаев, задаваемый моделью (12.2'), был рассмотрен в п. 12.3. Действительно, там был указан способ приближенного нахождения коэффициентов многочлена $P_m(t)$ и дисперсии σ^2 , т. е. параметров распределения $M(P_m(t), \sigma^2)$.

Второй случай — случай двумерного нормального распределения x и y — рассмотрим сейчас.

12.3. Случай двумерного нормального закона. Если совместное распределение x и y — двумерное нормальное, то имеют место равенства

$$y - My = r \sqrt{\frac{Dy}{Dx}} (x - Mx) + \delta, \quad (12.3')$$

$$x - Mx = r \sqrt{\frac{Dx}{Dy}} (y - My) + \zeta, \quad (12.4')$$

где δ и ζ — случайные величины, не зависящие соответственно от x и от y , r — коэффициент корреляции между x и y . Если нам известны параметры

$$Mx, My, Dx, Dy, r, \quad (12.5')$$

то лучшее, что можно сделать для предсказания значения величины y при известном значении x — это положить приблизительно $\delta=0$ в уравнении (12.3'), т. е. считать

$$\hat{y} = My + r \sqrt{\frac{Dy}{Dx}} (x - Mx)$$

приближенным значением величины y . Доверительный интервал для y , середина которого есть точка \hat{y} , легко находится с помощью нормального распределения для δ с параметрами

$$M\delta = 0, \quad D\delta = (1 - r^2) Dy.$$

Легко могут быть построены и α -квантили $y_\alpha(x)$, введенные в предыдущем пункте.

Не следует, однако, пользоваться способом, изложенным в предыдущем пункте, для отыскания x при известном y . Дело в том, что этот способ был основан на предположении, что каждому x отвечает распределение вероятностей для y , но не наоборот. Сейчас же роль переменных x и y совершенно симметрична. Поэтому имеет место равенство (12.4'); его и надо использовать для

отыскания x при известном y . Тогда будет получаться более короткий доверительный интервал, чем при использовании для этой цели α -квантилей $y_\alpha(x)$, найденных из уравнения (12.3').

Действительно, из уравнения (12.3') имеем

$$x - Mx = \frac{1}{r} \sqrt{\frac{Dx}{Dy}} (y - My) - \frac{1}{r} \sqrt{\frac{Dx}{Dy}} \delta,$$

так что ширина доверительного интервала для x определяется дисперсией величины $\frac{1}{r} \sqrt{\frac{Dx}{Dy}} \delta$, равной $\frac{1-r^2}{r^2} Dx$. В то же время ширина доверительного интервала для x , найденного из (12.4'), определяется дисперсией

$$D\zeta = (1 - r^2) Dx.$$

В силу сложившейся терминологии, уравнения (12.3') и (12.4') называются уравнениями регрессии, соответственно y на x и x на y , хотя мы бы предпочли употреблять этот термин лишь для уравнения (12.2') во избежание путаницы между двумя разными моделями. Мы видим, что когда имеется совместное распределение (x, y) , то для оценки одной переменной при известной второй надо употреблять подходящее уравнение регрессии. Очевидно, что в модели (12.2') уравнение регрессии t на x вовсе не имеет смысла, а потому для оценки t по x надо пользоваться доверительными интервалами, построенными в предыдущем пункте.

В случае ненормального распределения пары x и y также справедливы равенства (12.3') и (12.4'), однако δ не обязательно должно быть независимым от x , а ζ от y . Верно лишь, что $\text{cov}(\delta, x) = \text{cov}(\zeta, y) = 0$. В этом случае нет хороших способов для оценки одной переменной по другой, так как неизвестен закон распределения δ и ζ . Единственное, что остается — оценка δ и ζ с помощью неравенства Чебышева, но это обычно слишком грубо.

Посмотрим теперь, что произойдет, если параметры (12.5') теоретически не известны, а требуют оценки по выборке

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n).$$

Мы уже хорошо знаем вид оценок Mx , My , Dx , Dy :

$$Mx \approx \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad My \approx \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i;$$

$$Dx \approx s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad Dy \approx s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2.$$

Для $\text{cov}(x, y)$ несмещенной оценкой будет величина

$$\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

а потому для коэффициента корреляции $r = \frac{\text{cov}(x, y)}{\sqrt{Dx Dy}}$ естественно в качестве оценки взять так называемый выборочный коэффициент корреляции

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Подставляя в уравнение (12.3') вместо параметров (12.5') их оценки, получим эмпирический аналог (12.3') следующего вида:

$$\hat{y} - \bar{y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} (x - \bar{x}) + \delta^*,$$

где δ^* имеет распределение

$$N\left(0, \sqrt{(1-\rho^2)s_y^2}\right),$$

а на зависимость δ^* от x можно не обращать внимания.

В точности то же самое уравнение линейной регрессии мы бы получили, если бы стали обрабатывать пары $(x_1, y_1), \dots, (x_n, y_n)$, не предполагая наличия совместного распределения, а принимая модель, рассмотренную в п. 12.4. Таким образом, две разные вероятностные модели ведут к одним и тем же арифметическим действиям. Это часто приводит к их смещению, а тем самым и к ошибкам в интерпретации результата.

Например, в случае двумерной нормальной совокупности для определения величины x по известной величине y надо применить регрессию (12.4') величины x на y . Что получится, если на самом деле верна модель

$$y = x + \delta,$$

в которой x нельзя рассматривать как случайную величину, если по выборке (12.1') мы построим эмпирический аналог линии регрессии x на y и применим его для определения x при известном y ? Выражение « x нельзя рассматривать как случайную величину» означает, что мы не предполагаем устойчивого закона распре-

деления вероятностей для x . Будем предполагать, например, что наши наблюдения $(x_1, y_1), \dots, (x_n, y_n)$ таковы, что x_1, \dots, x_n есть выборка из некоторого распределения на отрезке $[-1, 1]$, а то новое значение $x = x_{n+1}$, которое мы попытаемся узнать по известному $y = y_{n+1}$, не принадлежит отрезку $[-1, 1]$. Имеем

$$y_i = x_i + \delta_i, \quad (12.7')$$

где δ_i не зависят от x_i и друг от друга. Пусть число наблюдений n велико. Тогда

$$s_x^2 \approx Dx, \quad s_y^2 \approx Dy = Dx + \sigma^2, \quad \sigma^2 = D\delta_i,$$

$$\rho \approx \frac{\text{cov}(x, y)}{\sqrt{Dx Dy}} = \frac{Dx}{\sqrt{Dx Dy}} = \sqrt{\frac{Dx}{Dy}}.$$

Поэтому эмпирический аналог линии регрессии y на x имеет вид

$$y - My \approx (x - Mx) + \delta^*,$$

что, конечно, разумно в свете модели (12.7').

Эмпирический аналог линии регрессии x на y имеет вид

$$x - Mx \approx \frac{Dx}{Dy} (y - My) + \zeta^* = \frac{Dx}{Dx + \sigma^2} (y - My) + \zeta^*,$$

где коэффициент $\frac{Dx}{Dx + \sigma^2} < 1$. Поэтому при большом значении $x = x_{n+1} \gg 1$ результаты определения x_{n+1} по $y_{n+1} = x_{n+1} + \delta_{n+1}$ с помощью двух разных уравнений регрессии будут резко различны. Ясно, что результат, вычисленный по уравнению регрессии x на y , будет совершенно неверным, если для y_{n+1} сохраняется действие модели (12.7').

Таким образом, не имея правильной модели, мы не сможем правильно восстановить x по y . Например, если мы подозреваем, что ни модель с двумерным нормальным распределением, ни модель со случайной ошибкой (12.2') на самом деле не имеет места, наше положение является затруднительным, а выводы не заслуживают доверия.

Сделаем одно замечание о распределении выборочного коэффициента корреляции, которое легко может быть получено из результатов п. 12.4 и из совпадения арифметических приемов в двух разных моделях. Пусть $(x_1, y_1), \dots, (x_n, y_n)$ — выборка из двумерного нормального распределения, и мы желаем проверить гипотезу о том, что x_i не зависит от y_i , т. е. что теоретический коэффициент корреляции $r=0$. С этой целью мы должны уметь решить вопрос о значимости полученного эмпирического коэффициента корреляции ρ , иначе говоря, знать распределение ρ при гипотезе $r=0$.

Рассмотрим условное распределение значений y_1, \dots, y_n при заданных x_1, \dots, x_n . При гипотезе о независимости распределений

оно совпадает с безусловным распределением. Следовательно, условное распределение может быть представлено моделью

$$y_i - a = 0 \cdot x_i + \delta_i, \quad a = My_i,$$

где δ_i независимы и имеют нормальное распределение $N(0, \sigma)$, где $\sigma^2 = D\delta_i = Dy_i$.

Мы имеем частный случай модели, рассмотренный в п. 12.4 (в котором $b_1 = c_1 = 0$). В пункте 12.4 было показано, что статистика

$$\frac{(\beta_1 - b_1) \sqrt{W^1, W^1}}{\Delta} \sqrt{n-2} = \frac{\beta_1 \sqrt{W^1, W^1}}{\Delta} \sqrt{n-2}$$

имеет распределение Стьюдента t_{n-2} .

Однако в наших обозначениях имеем

$$\begin{aligned} \beta_1 \sqrt{W^1, W^1} &= \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}; \\ \Delta^2 &= \sum_{i=1}^n (y_i - \bar{y})^2 - \beta_1^2 \sum_{i=1}^n (x_i - \bar{x})^2 = \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{\left[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned}$$

Отсюда вытекает, что

$$\frac{\beta_1 \sqrt{W^1, W^1}}{\Delta} \sqrt{n-2} = \frac{\rho}{\sqrt{1-\rho^2}} \sqrt{n-2}.$$

Таким образом, условное распределение статистики $\frac{\rho \sqrt{n-2}}{\sqrt{1-\rho^2}}$ при условии, что x_1, \dots, x_n заданы, есть распределение Стьюдента t_{n-2} (следовательно, вовсе не зависит от условия x_1, \dots, x_n). Поэтому безусловное распределение статистики

$$\frac{\rho}{\sqrt{1-\rho^2}} \sqrt{n-2}$$

есть также t_{n-2} . Итак, для определения значимости выборочного коэффициента корреляции ρ нужно для указанной статистики назначить критическую область исходя из распределения t_{n-2} .

Мы имеем два разных способа выражения независимости. В модели с двумерным нормальным распределением гипотеза независимости имеет вид $r=0$. В модели

$$x_i = P_m(t_i) + \delta_i$$

гипотеза независимости имеет вид $P_m(t)=0$ (результат наблюдения не зависит от условий опыта t). В частности, коэффициент c_1 многочлена

$$P_m(t) = c_0 + c_1 t + \dots + c_m t^m$$

при гипотезе независимости равен нулю.

Совершенно разные (отвечающие разным моделям) гипотезы $r=0$ и $c_1=0$ проверяются при помощи одной и той же статистики, т. е. одних и тех же вычислений. Однако в модели двумерного нормального распределения нет никакого аналога гипотезам $c_2=0$, $c_3=0$, ..., $c_m=0$. Такие аналоги могли бы появиться в моделях двумерного, но не нормального распределения (нелинейная связь между переменными). Однако разговоры о нелинейных связях в такой модели совершенно пусты, поскольку неизвестно, как математически формулировать и проверять гипотезы о нелинейных связях между случайными величинами.

Обратим внимание также на часто встречающееся ошибочное истолкование коэффициента корреляции как меры близости двух наблюдаемых в опыте величин x и y . Степень ошибочности такого толкования зависит от того, какая вероятностная модель описывает на самом деле наблюдения. Если пара (x, y) — случайная величина (т. е. можно говорить о ее двумерном распределении вероятностей), то указанное толкование может (в случае двумерного нормального распределения) даже быть правильным.

Разберем этот вопрос несколько подробнее. Коэффициент корреляции

$$r = \frac{\text{cov}(x, y)}{\sqrt{Dx Dy}}$$

указывает (в случае значения r , близкого к единице) не на то, что величины x и y близки между собой, а на то, что линейно зависящее от x выражение

$$\hat{y} = M y + r \sqrt{\frac{Dy}{Dx}} (x - Mx)$$

отличается от y на величину $\delta = y - \hat{y}$ такую, что дисперсия $D\delta = (1-r^2) Dy$ мала (впрочем, мала не сама по себе, а лишь в сравнении с Dy). Переход от x к линейной форме \hat{y} может (в зависимости от структуры эксперимента, дающего x и y) быть как естественным, так и неестественным. Например, он весьма естествен, если выбор шкал измерения x и y произволен, т. е. сами x и y определены с точностью до линейного преобразования. Наоборот, он

весьма неестествен, если хотя бы нули шкал, в которых измеряются x и y , строго фиксированы. Пусть, например, x и y — наблюдения одной и той же величины разными экспериментаторами, и мы желаем проверить, хорошо ли согласуются эти наблюдения. Если мы с самого начала знаем, что нули отсчета и единицы измерения могли оказаться совершенно несогласованными (для примера укажем оценку умственных способностей с помощью двух разных наборов тестов), то для оценки близости между x и y естественно пользоваться коэффициентом корреляции. Наоборот, если шкалы должны быть согласованы, как бывает в экспериментах точных наук, то коэффициент корреляции (который, например, совершенно не чувствует систематического сдвига $y = x + a$, a — число) для оценки близости не годится.

Еще более важно заметить, что в том случае, когда нельзя говорить о двумерном распределении (x, y) , коэффициент корреляции никоим образом не годится для оценки близости между x и y . Конечно, речь в данном случае может идти лишь о выборочном коэффициенте корреляции:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}},$$

поскольку теоретический коэффициент не имеет смысла.

Например, если верна модель $y_i = x_i + \delta_i$, то близость между x и y определяется малостью дисперсии $\sigma^2 = D\delta_i$. Однако при любой дисперсии σ^2 выборочный коэффициент корреляции может быть сделан сколь угодно близким к 1, если значения независимого переменного x_i менять в достаточно широком интервале. Докажем это. Пусть x_i принимает значения $-N, -(N-1), \dots, 0, \dots, N-1, N$. Тогда

$$\bar{x} = 0, \quad \bar{y} = \frac{1}{2N+1} \sum_{i=-N}^N \delta_i.$$

Далее

$$\sum_{i=-N}^N (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=-N}^N i^2 + \sum_{i=-N}^N i \delta_i;$$

$$\sum_{i=-N}^N (x_i - \bar{x})^2 = \sum_{i=-N}^N i^2;$$

$$\sum_{i=-N}^N (y_i - \bar{y})^2 = \sum_{i=-N}^N i^2 + 2 \sum_{i=-N}^N i \delta_i + \sum_{i=-N}^N \delta_i^2 - (2N+1) (\bar{y})^2.$$

Величина $S_N = \sum_{i=-N}^N i^2$ имеет порядок N^3 . Путем подсчета дисперсий и применения неравенства Чебышева получаем: при $N \rightarrow \infty$

$$Q_N = \frac{1}{S_N} \sum_{i=-N}^N i \delta_i \rightarrow 0, \quad R_N = \frac{1}{S_N} \sum_{i=-N}^N \delta_i^2 \rightarrow 0,$$

$$T_N = \frac{1}{S_N} (2N+1) (\bar{y})^2 \rightarrow 0$$

в смысле сходимости по вероятности. Выражение для ρ имеет вид

$$\rho = \rho_N = \frac{S_N (1 + Q_N)}{S_N \sqrt{1 + 2Q_N + R_N - T_N}}.$$

Следовательно, $\rho = \rho_N \rightarrow 1$ при $N \rightarrow \infty$.

Наоборот, если значения независимого переменного x_i меняются в достаточно узких пределах, то при сколь угодно малой дисперсии $\sigma^2 = D\delta_i$ выборочный коэффициент корреляции будет мал. Математическое доказательство этого утверждения похоже на предыдущее, и мы предоставляем его читателю.

Таким образом, в случае модели $y_i = x_i + \delta_i$ выборочный коэффициент корреляции зависит не только от величины дисперсии σ^2 , но и от произвольно выбираемых значений x_i . В этом случае его истолкование как меры близости между x и y смысла не имеет.

Сказанное выше должно сделать понятным следующее замечание общего характера. При использовании средств математической статистики для исследования того или иного явления нужно представлять себе действительность в терминах более или менее адекватной вероятностной модели, в то время как использование одних арифметических приемов (в частности, так называемый «корреляционный анализ») будет давать результат, который неизвестно как интерпретировать. В свое время простое использование арифметических приемов было, вне всякого сомнения, прогрессивным. Однако это время прошло.

12.4'. Сколько чакрмов в километре и сколько километров в чакрме? Закончим эту книгу иллюстрацией того обычного обстоятельства, что для применения статистических методов недостаточно применить готовый рецепт, а нужно приспособить к рассматриваемому случаю рекомендованные для других случаев методы. Пример взят из книги Л. Н. Гумилева «Открытие Хазарии (историко-географический этюд)». Москва, «Наука», 1966, стр. 18—20. Цитируем подробно Л. Г. Гумилева.

«В один из весенних дней 1959 г. я вошел в читальный зал библиотеки Эрмитажа и увидел профессора М. И. Артамонова, рассматривающего карту калмыцких степей. «Сколько километров в фарсах?» — мрачно спросил он меня. Я припомнил общеприня-

тую величину — 5,5 км, но профессор буркнул: «Не выходит» и пригласил меня к карте. Дело заключалось в следующем. Хазарский царь Иосиф в письме к Хсдаи ибн-Шафруту описал ежегодную летнюю перекочевку своего двора. Весной он выезжал из своей столицы Итиль, расположенной на берегу Волги, и двигался на юг к реке В-д-шан. Затем он перекочевывал на север, очевидно, избегая летней жары в засушливых прикаспийских районах, но двигался не домой, а к реке Бузан, отождествляемой с Доном, и отсюда возвращался к себе в Итиль, находившийся в 20 фарсах от Бузана. Тут же царь Иосиф сообщает расстояния от своей столицы до границ своего царства: на восток до Гирканского, т. е. Каспийского, моря — 20 фарсахов, на юг до реки Уг-ру — 30 фарсахов и на север до уже упомянутой реки Бузан и «до склона нашей реки к морю Гирканскому», т. е. до сближения излучин Дона и Волги в современном месте Волго-Донского канала, — 20 фарсахов. Таким образом, все расстояния исчисляются от столицы Итиля. Следовательно, для того чтобы найти место столицы, М. И. Артамонов построил на карте треугольник, упиравшийся вершинами в реки Дон (Бузан), Волгу (Итиль) и Терек (Уг-ру), с длиной сторон, пропорциональной заданным расстояниям.

Однако установленная длина фарсак — 5,5 км противоречила его построению. Если принять эту длину за основу и опереть вершины треугольника на Дон и пусть даже не на Терек, а на Куму и Маныч, то столица Хазарского каганата должна оказаться в степи Северной Калмыкии, около Сарпинских озер. Это одно противоречило источникам, помещавшим Итиль на берегу Волги, а кроме того, пропадала большая река В-д-шан, находившаяся на 10 фарсахов севернее пограничной реки Уг-ру. Задача казалась неразрешимой, и именно это заставило моего учителя задуматься.

И тут у меня внезапно вспыхнула далекая ассоциация. В молодости, еще в 1932 г., мне довелось работать в Таджикистане малярийным разведчиком. Работа заключалась в том, что я находил болотца, где выводились комары, наносил их на план и затем отравлял воду «парижской зеленью». Количество комаров при этом несколько уменьшалось, но уцелевших вполне хватало для того, чтобы заразить малярией не только меня, но и все население района. Однако я извлек из этой работы максимальную пользу, потому что освоил глазомерную съемку и разговорный таджикский язык. Так как при определении расстояний мне неоднократно приходилось обращаться к местным жителям, то я волей-неволей усвоил среднеазиатскую меру длины — чакрым. Определить длину чакрыма в метрах было невозможно: он был до длинный, то короткий, но в вариациях наблюдалась строгая закономерность. Если путь в гору или по болоту — чакрым короткий, если с горы или по хорошей дороге — длинный, а все прочие величины располагались между этими лимитами. Собственно говоря, чакрым был мерой не длины, а усилий, которые человек должен был затратить, чтобы достигнуть цели. Нельзя не признать, что такая система отсчета

была очень удобна для местных жителей, хотя совершенно непригодна для картирования. И тут мне пришла в голову мысль, что таджикский «чакрым» не что иное, как персидский «фарсанг» (арабизированная форма — фарсах), и тогда следует учитывать не абстрактную длину, а проходимость путей перекочевков. Длина фарсаха вычислена европейцами в условиях пересеченного рельефа Иранского плоскогорья, а в прикаспийских степях, гладких как стол, она должна быть куда больше. Мы тут же прикинули расстояние, построили треугольник, и оказалось, что при длине хазарского фарсаха 10 км река Уг-ру — Терек, река Бузан — Дон, В-д-шан — Кума, а Итиль должен находиться на одном из берегов Волги между селами Енотаевским и Селитрянным».

Ситуация описана достаточно ясно. Попробуем научно поставить вопрос о переводе чакрымов в километры. В приведении различных данных укоренилась дурная традиция, когда приводятся лишь средние значения, и нельзя узнать, как велики бывают отклонения от средних. Допустим, что вы собираетесь в путешествие и узнаете по справочнику, что в районе вашего путешествия в июле в среднем 15 солнечных дней. Обычно при этом никак нельзя узнать, как часто в июле бывает 25 дождливых дней. Ясно, что путешественнику по Средней Азии совершенно недостаточно знать, что «в среднем» в чакрыме a километров (спросить бы еще, по какому «ансамблю» вычислено это среднее), но необходимо иметь какое-то представление о величине доверительного интервала для числа километров при заданном числе чакрымов.

Таким образом, необходимо нечто вроде графика 12.1', в котором по одной из осей отложено число чакрымов, а по другой — число километров. Но для получения такого графика нужно ввести одно из предположений — либо, что при данном числе чакрымов имеется распределение вероятностей для числа километров, либо, что при заданном числе километров имеется распределение вероятностей для числа чакрымов. В первоначальной постановке задачи обе меры — чакрымы и километры — совершенно симметричны.

Можно было бы попробовать даже ввести их совместное распределение. Однако это распределение, очевидно, должно быть распределением расстояний между какими-то пунктами (выбранными более или менее наудачу), измеренных в чакрымах и километрах, и заведомо не будет нормальным. Поэтому мы откажемся от модели с совместным распределением. Что же лучше — распределение чакрымов при известном числе километров или распределение километров при известном числе чакрымов? Число километров можно всегда узнать точно по карте. Что касается числа чакрымов — его придется спросить у местных жителей. Будет ли их ответ вполне точен? Автор книги вспоминает, что в одной подмосковной деревне ему сообщили, что до железнодорожной станции отсюда прежде было восемь километров, а как пошел автобус, так стало двенадцать (по той же дороге). Поэтому, ради получения

данных лучшего качества, лучше основываться на топографической карте.

Итак, предполагаем, что при заданном числе километров x число чакрымов y имеет некоторое распределение вероятностей. Каким должно быть это распределение? Разобьем дорогу длиной в x километров на отрезки x_1, x_2, \dots, x_k :

$$x = x_1 + x_2 + \dots + x_k.$$

Тогда число чакрымов y также представится в виде суммы

$$y = y_1 + y_2 + \dots + y_k,$$

где в отрезке дороги x_i километров содержится y_i чакрымов. Допустим, что отрезки x_1, \dots, x_k равны между собой и не слишком малы. Раз они равны, то y_1, y_2, \dots, y_k являются одинаково распределенными случайными величинами. А раз x_i не слишком малы, то величины y_1, y_2, \dots, y_k можно считать независимыми. Но тогда закон распределения суммы $y = y_1 + y_2 + \dots + y_k$ нормален с математическим ожиданием kMy_i и дисперсией kDy_i . Иными словами, при достаточно большом числе километров x соответствующее число чакрымов имеет нормальное распределение $N(ax, \sigma\sqrt{x})$, где a и σ — параметры, которые нужно найти из эксперимента (очевидно, a — «среднее» число чакрымов в километре, но нам нельзя удовлетвориться средним, а нужно знать еще параметр σ , характеризующий разброс). Итак, для перевода километров в чакрымы (а с помощью чертежа 12.1' и чакрымов — в километры) достаточно знать a и σ (а также минимальную величину расстояния в километрах, начиная с которой для числа чакрымов действует распределение $N(ax, \sigma\sqrt{x})$). Очевидно, что при малых расстояниях наш подход бесполезен, да там действительно может быть что угодно (в зависимости от крутизны горы или толкости болота).

Рассмотрим теперь эксперимент, с помощью которого можно определить a и σ . Самый наивный подход, впрочем, не требует применения распределения $N(ax, \sigma\sqrt{x})$, но зато требуется, чтобы при каждом числе километров x было произведено достаточно экспериментов для выяснения распределения числа чакрымов y . Этот способ теоретически годен и при малых x , но практически совершенно невозможен из-за слишком большого числа необходимых экспериментов. Действительно, для определения закона распределения при каждом x нужны сотни и тысячи экспериментов, так что если взять несколько значений x , то программа экспериментов становится невыполнимой.

Значение метода наименьших квадратов состоит в том, что он позволяет (за счет дополнительных предположений) совместно обрабатывать результаты экспериментов, проведенных в разных условиях.

Самый чистый эксперимент для определения длины чакрыма, который можно себе представить, заключается в следующем. Слу-

чайню выбирается точка в пределах Средней Азии, направление, в котором следует идти, и длина перехода x км. Вертолеты доставляют в выбранную точку группу местных жителей, особенно хорошо чувствующих «чакрымную меру» расстояния; они проходят выбранный отрезок, оценивают его длину в чакрымах, и их оценки усредняются для получения лучшей устойчивости результата. Другая группа проезжает тот же путь на ишаках и дает свою оценку длины пути в чакрымах. Затем вертолеты перебрасывают обе группы вместе с ишаками в другую наудачу выбранную точку и т. д. Таков современный психофизический эксперимент, но он, очевидно, безумно дорог.

Реален, очевидно, опрос местных жителей о расстояниях (в чакрымах) между n парами выбранных наудачу населенных пунктов (один из этих пунктов выбирается наудачу, а второй — так, чтобы расстояние от первого не было слишком малым или слишком большим). В результате измерения расстояний по карте получим совокупность наблюдений

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n).$$

Наша модель состоит в том, что y_i имеет нормальное распределение $N(ax_i, \sigma \sqrt{x_i})$, иначе говоря,

$$y_i = ax_i + \delta_i,$$

где δ_i имеет распределение $N(0, \sigma \sqrt{x_i})$ и $\delta_1, \delta_2, \dots, \delta_n$ независимы (для независимости $\delta_1, \dots, \delta_n$ не следует выбирать рядом пары населенных пунктов). Итак, веса отдельных наблюдений обратно пропорциональны числам x_i . Приведем нашу модель к модели с равными весами: положим

$$z_i = \frac{y_i}{\sqrt{x_i}}, \quad \xi_i = \frac{\delta_i}{\sqrt{x_i}}.$$

Тогда

$$z_i = a \sqrt{x_i} + \xi_i,$$

где ξ_i имеют распределение $N(0, \sigma)$.

Полагая $z = (z_1, \dots, z_n)$, $W = (\sqrt{x_1}, \dots, \sqrt{x_n})$, имеем оценку \hat{a} для a :

$$\hat{a} = \frac{(z, W)}{(W, W)} = \frac{\sum_{i=1}^n z_i \sqrt{x_i}}{\sum_{i=1}^n x_i} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i},$$

т. е. \hat{a} есть среднее число чакрымов, приходящихся на один километр, как естественно было и ожидать.

Оценка для σ^2 есть

$$s^2 = \frac{1}{n-1} \left\| z - \frac{(z, W)}{(W, W)} W \right\|^2 = \\ = \frac{1}{n-1} \sum_{i=1}^n (z_i - \hat{a} \sqrt{x_i})^2 = \frac{1}{n-1} \sum_{i=1}^n \frac{(y_i - \hat{a} x_i)^2}{x_i},$$

т. е. «кажущиеся ошибки» $y_i - \hat{a} x_i$ возводятся в квадрат и суммируются с множителями $1/x_i$.

При числе наблюдений n порядка нескольких десятков или сотен \hat{a} практически точно совпадает с a и s^2 с σ^2 . Возможны лишь погрешности за счет нарушения статистической однородности, которыми мы вынуждены пренебречь. Кривые $y_\alpha(x)$, задаваемые уравнениями

$$p\{y < y_\alpha(x) | x\} = \alpha, \quad 0 < \alpha < 1,$$

имеют вид

$$y_\alpha(x) = ax + \sigma k_\alpha \sqrt{x},$$

где k_α определяется из соотношения $\Phi(k_\alpha) = \alpha$, причем Φ — функция Лапласа. Аналог чертежа 12.1' следует, разумеется, рисовать только при достаточно больших x . Доверительный интервал с коэффициентом надежности $1-2\alpha$ для числа километров x_0 при известном числе чакрымов y_0 имеет вид $(x^{(1)}, x^{(2)})$, где $x^{(1)}$ и $x^{(2)}$ находятся из уравнений

$$ax^{(1)} + \sigma k_{1-\alpha} \sqrt{x^{(1)}} = y_0, \quad ax^{(2)} + \sigma k_\alpha \sqrt{x^{(2)}} = y_0.$$

Конечно, необходим чертеж для мгновенного графического решения этих уравнений.

Однако правильность расчетов прохождения маршрута зависит от справедливости положенного в основу всех расчетов предположения: при известном числе километров x число чакрымов y можно считать случайной величиной, притом имеющей не какое угодно распределение, а распределение $N(ax, \sigma\sqrt{x})$. Мы постеснялись предположить нормальное распределение для пары (x, y) по следующей причине.

Расстояние x между двумя выбранными наудачу населенными пунктами (в километрах) нельзя считать, с точки зрения предполагаемого путешественника, случайной величиной: у расстояния x есть какая-то неопределенность, но навряд ли есть статистическая устойчивость. Очевидно, что путешественник часть времени будет путешествовать почти сплошь по населенным пунктам (расстояние x будет принимать малые значения), а часть времени по малонаселенной местности (большие значения x). Наоборот, при заданном x у расстояния в чакрымах y (т. е. у трудности прохождения пути) должна быть большая статистическая устойчивость, связанная с

некоторой интуитивно ощущаемой устойчивостью свойств ландшафта в пределах данной природной зоны. Однако никому еще не удалось научно сформулировать это свойство устойчивости ландшафта.

В наших рассуждениях участвовала также независимость трудности прохождения отдельных (достаточно больших) участков пути. Это также есть свойство ландшафта, которое никто не в состоянии научно сформулировать. Таким образом, построенная нами модель представляет собой типичный пример вероятностных моделей, которые более или менее правдоподобны, но, строго говоря, до их экспериментальной статистической проверки ничего не стоят.

Хорошо то, что наша модель в принципе вполне допускает статистическую проверку. Возможно, что она когда-нибудь будет проверена, но, конечно, не для чакрымов, выражающих трудность пешего путешествия, а для меры трудности (например, времени) автомобильной поездки. Если модель выдержит проверку, то в результате может получиться аналог графика 12.1', который представляет очевидный хозяйственный интерес.

ЛИТЕРАТУРА

1. Александров П. С. Введение в общую теорию множеств и функций. М., Гостехиздат, 1948.
2. Арлей Н., Бух К. Введение в теорию вероятностей и математическую статистику. М., ИЛ, 1948.
3. Бернштейн С. Н. Теория вероятностей, изд. 4-е. М., Гостехиздат, 1946.
4. Большев Л. Н., Смирнов С. В. Таблицы математической статистики. М., «Наука», 1968.
5. Вальд А. Последовательный анализ. М., Физматгиз, 1960.
6. Ван дер Варден Б. Л. Математическая статистика. М., ИЛ, 1960.
7. Гельфанд И. М., Шилев Г. Е. Обобщенные функции, вып. I. Обобщенные функции и действия над ними. М., Физматгиз, 1959.
8. Гнеденко Б. В. Курс теории вероятностей, изд. 5-е. М., Физматгиз, 1969.
9. Гнеденко Б. В., Колмогоров А. Н. Предельные распределения для сумм независимых случайных величин. М., Гостехиздат, 1949.
10. Кендалл М., Стюарт А. Кривые распределения. М., «Наука», 1966.
11. Колмогоров А. Н. Основные понятия теории вероятностей. М.—Л., ОНТИ, 1936.
12. Колмогоров А. Н. Об одном новом подтверждении законов Менделя. ДАН СССР, 26, 6—9, 1940.
13. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа, изд. 2-е. М., «Наука». 1968.
14. Крамер Г. Математические методы статистики. М., ИЛ, 1948.
15. Линник Ю. В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. М., Физматгиз, 1962.
16. Лоев М. Теория вероятностей. М., ИЛ, 1962.
17. Мак Кинси Дж. Введение в теорию игр. М., Физматгиз, 1960.
18. Мешалкин Л. Д. Сборник задач по теории вероятностей. Изд-во МГУ, 1963.
19. Мизес Р. Вероятность и статистика. М., ГИЗ, 1930.
20. R. von Mises. Mathematical theory of probability and statistics. Edited and complemented by H. Geiringer. N. Y. and London, Acad. press, 1964.
21. Мостеллер Ф., Рурке Р., Томас Дж. Вероятность. М., «Мир», 1969.
22. Феллер В. Введение в теорию вероятностей и ее приложения, т. 1, изд. 2-е. М., «Мир», 1964; т. 2. М., «Мир», 1967.
23. Фихтенгольц Г. М. Курс дифференциального и интегрального исчисления, т. 1, изд. 7-е. М., 1970.
24. Хальд А. Математическая статистика с техническими приложениями. М., ИЛ, 1955.

25. Хинчин А. Я. Асимптотические законы теории вероятностей. М., ГТТИ, 1936.
26. Худсон Д. Статистика для физиков. М., «Мир», 1967.
27. Шеффе Г. Дисперсионный анализ. М., Физматгиз, 1963.
28. Эйнштейн А. Физика и реальность. М., «Наука», 1965.
29. Янко Я. Математико-статистические таблицы. М., Госстатиздат, 1961.

СОДЕРЖАНИЕ

Предисловие	3
-----------------------	---

ЧАСТЬ I. КРАТКИЙ КУРС ТЕОРИИ ВЕРОЯТНОСТЕЙ

§ 1. Дискретное пространство элементарных событий . .	5
§ 2. Условная вероятность. Независимость. Основные формулы	11
§ 3. Случайные величины и их основные характеристики	19
§ 4. Неравенство Чебышева. Закон больших чисел. Испытания Бернулли. Теорема Пуассона	31
§ 5. Статистическая проверка гипотез	41
§ 6. Аксиоматика Колмогорова. Интеграл Лебега	46
§ 7. Распределение случайных величин	58
§ 8. Центральная предельная теорема	76
§ 9. Применения центральной предельной теоремы . . .	93
§ 10. Выборка. Оценка параметров	99
§ 11. Общая линейная модель, связанная с нормальным распределением ошибок наблюдений	111
§ 12. Дальнейшие применения метода наименьших квадратов	121

ЧАСТЬ II. НАУЧНЫЕ И МЕТОДИЧЕСКИЕ ЗАМЕЧАНИЯ

Введение	142
Замечания к § 1	145
Замечания к § 2	149
Замечания к § 3	156
Замечания к § 4	164
Замечания к § 5	166
Замечания к § 6	169
Замечания к § 7	173
Замечания к § 8	182
Замечания к § 9	193
Замечания к § 10	195
Замечания к § 11	199
Замечания к § 12	208
Литература	228

Валерий Николаевич ТУТУБАЛИН

Теория вероятностей



Тематический план 1972 г. № 111

Редактор *Ю. И. Сионский*

Переплет художника

М. М. Носовой

Технический редактор

З. С. Кондрашова

Корректоры

М. И. Эльмус, С. Ф. Будаева

Сдано в набор 27.I 1972 г. Подписано к печати 27.X 1972 г. Л-109832. Формат 60×90/16. Бумага тип. № 2. Физ. печ. л. 14,5. Уч.-изд. л. 14,48. Изд. № 1550. Зак. 40. Тираж 22 780 экз. Цена 61 коп.

Издательство
Московского университета.
Москва, К-9, ул. Герцена, д. 5/7.
Типография Изд-ва МГУ.
Москва, Ленинские горы

Цена 61 коп.

