П. Бикел
К. Доксам

МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ МЕТОДЫ • ЗА РУБЕЖОМ •

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

П. Бикел, К. Доксам

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Книга одобрена на заседании редколлегии серии «Математико-статистические методы за рубежом» 28.04.81

Зав. редакцией *А. В. Павлюков*
Редактор *К. М. Чижевская*
Мл. редактор *О. А. Ермилина*
Техн. редактор *Л. Г. Челышева*
Корректоры *Т. М. Васильева, Н. П. Сперанская* и *З. С. Кандыба*
Худож. редактор *О. Н. Поленова*
ИБ № 1248

Сдано в набор 10.09.82. Подписано в печать 04.02.83. Формат 60×90[1]/16. Бум. тип. № 2. Гарнитура «Литературная». Печать высокая. Уч.-изд. л. 19,92 Усл. п. л.17,5. Тираж 7000 экз. Заказ 1182
Цена 2 р. 40 к.

Издательство «Финансы и статистика», 101000, Москва, ул. Чернышевского, 7

Московская типография № 4 Союзполиграфпрома
при Государственном комитете СССР
по делам издательств, полиграфии и книжной торговли
129041, Москва, Б. Переяславская ул., д. 46.

МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ

МЕТОДЫ ЗА РУБЕЖОМ

MATHEMATICAL STATISTICS

Basic Ideas and Selected Topics

Peter J. Bickel

University of California
Berkeley

Kjell A. Doksum

University of California
Berkeley

Holden-Day, Inc.
San Francisco—Düsseldorf—Johannesburg—London—Panama
Singapore—Sydney

П. Бикел, К. Доксам

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

ВЫПУСК 1

Перевод с английского Ю. А. ДАНИЛОВА
Предисловие Ю. Н. ТЮРИНА

Москва «Финансы и статистика» 1983

ББК 22.172
Б 60

МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ МЕТОДЫ ЗА РУБЕЖОМ

ВЫШЛИ ИЗ ПЕЧАТИ

1. Ли Ц., Джадж Д., Зельнер А. Оценивание параметров марковских моделей по агрегированным временным рядам.
2. Райфа Г., Шлейфер Р. Прикладная теория статистических решений.
3. Клейнен Дж. Статистические методы в имитационном моделировании. Вып. 1.
4. Клейнен Дж. Статистические методы в имитационном моделировании. Вып. 2.
5. Бард Й. Нелинейное оценивание параметров.
6. Болч Б., Хуань К. Д. Многомерные статистические методы для экономики.
7. Иберла К. Факторный анализ.
8. Зельнер А. Байесовские методы в эконометрии.
9. Хейс Д. Причинный анализ в статистических исследованиях.
10. Пуарье Д. Эконометрия структурных изменений.
11. Драймз Ф. Распределенные лаги.
12. Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия. Вып. 1.
13. Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия. Вып. 2.

ГОТОВИТСЯ К ПЕЧАТИ

Лимер Э. Статистический анализ неэкспериментальных данных. Выбор формы связи.

$Б \frac{0702060000-048}{010(01)-83}$ 30-83

● ПРЕДИСЛОВИЕ К РУССКОМУ ПЕРЕВОДУ

Предлагаемая вниманию советского читателя книга П. Бикела и К. Доксама «Математическая статистика» представляет собой доступное, в лучшем смысле популярное изложение главных вопросов современной математической части статистической науки. Она открывает теперь уже широко известную серию по теории вероятностей и математической статистике под редакцией Э. Лемана, выпускаемую американским издательством «Holden-Day».

Авторы этой книги профессор Калифорнийского университета в Беркли (США) Питер Дж. Бикел и доцент того же университета Куэлл Доксам, поставив перед собой цель — дать сжатое и в то же время доходчивое введение в круг понятий и методов современной математической статистики, успешно справились с этой задачей.

Чтобы показать, какое место занимает эта книга в системе современного преподавания математической статистики, напомним основные черты сложившейся в этом преподавании традиции. К ним относится изложение материала на основе теории меры и интеграла Лебега, а также большое внимание к прикладным вопросам, в частности тщательный анализ реальных проблем, вместе с числовыми расчетами. Такая традиция была заложена известной книгой Г. Крамера «Математические методы статистики», на которой воспитано не одно поколение специалистов. Ее продолжали многие последующие книги, среди которых выделяется книга С. Р. Рао «Линейные статистические методы и их применение».

Несомненно, эти черты правильно отражают особенности самой науки. Но справедливо и то, что читателя, приступающего к изучению математической статистики, это ставит в нелегкое положение. Ему одновременно приходится знакомиться со статистическими идеями, преодолевать трудности математического аппарата, учиться применять статистические методы на деле. Помимо прочего читателя затрудняет разносторонность стоящих перед ним задач.

Многочисленные краткие, вводные и т. п. курсы математической статистики находят выход из этого затруднительного положения в том, что говорят обо всем понемногу. Авторы предлагаемой читателям книги отыскали иной путь. В обучении всякому новому делу, будь то наука или ремесло, полезно расчленить его на элементы, которым можно учиться отдельно. Разумеется, в конце концов все эти элементы должны слиться в сознании в нечто целое. П. Бикел и К. Доксам объектом своего основного внимания избрали главные определившиеся к настоящему времени в области математической статистики идеи, методы и подходы, обходясь элементарными средствами. К математической подготовке читателя они предъявляют скромные требования, опираясь лишь на классический анализ (с элементами функционального анализа), линейную алгебру и теорию вероятностей в объеме, примерно соответствующем уровню математической подготовки в наших технических и экономических вузах. Приложения к реальным проблемам также отнесены на второй план, хотя каждый раздел книги сопровождается условными упражнениями и задачами, различными по сложности и разнообразными по тематике.

Необходимые сведения из теории вероятностей авторы сообщают на двух уровнях: перечень наиболее важных понятий и теорем (без доказательств) приведен в приложении, содержащем более традиционный материал, а менее традиционные результаты изложены в гл. 1.

Большой педагогический опыт авторов позволил им избежать упрощенчества почти при неосторожном проецировании конструкций математической статистики на элементарный уровень. Там, где доказательство ведется на эвристическом уровне, читатель получает ясные указания, как восполнить соответствующие пробелы.

Комментарии и дополнения повышают полноту охвата материала, показывая различные его взаимосвязи с другими концепциями и теориями. Большое число задач (их в книге более 400) позволяет не только уверенно овладеть методами, но и узнать много нового, о чем авторы не сказали в основном тексте.

Работы по математической статистике обычно содержат таблицы. Эта традиция не нарушена. К книге приложен небольшой набор таблиц, необходимых для выполнения упражнений, проведения расчетов и иллюстрирующих текст. В статистической практике приходится, конечно, обращаться к более полным сборникам статистических таблиц.

Как можно заметить по оглавлению, помимо традиционного материала книга П. Бикела и К. Доксама включает и такие относительно новые вопросы, как теория статистических решений, устойчивость статистических правил, логарифмический линейный анализ и др.

Разумеется, изучить математическую статистику по одной этой книге нельзя. Она на это и не претендует. За ее чтением должна последовать более углубленная специализация в тех или иных разделах, в ходе которой можно будет научиться прилагать статистические идеи к практике, что и составляет настоящую цель этой науки. Но и для учащихся, и для преподавателей книга П. Бикела и К. Доксама будет хорошим подспорьем. Полезна она будет и тем практикам, которые, имея опыт работы в одной области математической статистики, пожелают овладеть и другими ее разделами.

Ю. Н. ТЮРИН

● ПРЕДИСЛОВИЕ

Эта книга отражает наши представления о том, каким должно быть введение в математическую статистику для студентов с хорошей математической подготовкой, означающей в нашем понимании свободное владение линейной алгеброй и теорией матриц, математическим анализом и избранными главами функционального анализа, к числу которых мы не относим теорию меры. Так как наша книга представляет собой введение в математическую статистику, мы основательно опираемся на теорию вероятностей и ожидаем, что читатель знаком с ней в объеме, например, «Введения в теорию вероятностей» П. Хоела, С. Порта и Ч. Стоуна. Все необходимые сведения по теории вероятностей приведены в приложении. Следует подчеркнуть, что теория вероятностей изложена в приложении кратко, с небольшим числом доказательств и без примеров или задач.

Любое введение в математическую статистику, по нашему глубокому убеждению, должно непременно выполнять следующие задачи:

1) давать описание основных понятий математической статистики с указанием связи теории с практикой;

2) приводить тщательные доказательства большинства «элементарных» результатов (таких, как лемма Неймана — Пирсона, теорема Лемана—Шеффе, неравенство информации и теорема Гаусса — Маркова);

3) обсуждать на эвристическом уровне более сложные результаты (такие, как оценки максимума правдоподобия в теории больших выборок и структура байесовских и допустимых решений в теории решений), причем должно быть четко указано, где именно в доказательствах имеются пробелы и до какой степени их можно восполнить;

4) показывать, каким образом излагаемые идеи и результаты находят применение в различных областях математической статистики (например, в гауссовых линейных моделях, в мультиномиальных моделях и в непараметрических моделях).

Хотя хороших вводных курсов математической статистики существует немало, по нашему мнению, ни один из них не обладает необходимым сочетанием ширины охвата с глубиной изложения. Книга С. Р. Рао «Линейные статистические методы и их применения» (русский перевод: М., Наука, 1968) содержит значительную часть материала, включенного в нашу книгу, и многие другие результаты, но изложение гораздо более абстрактно и использует теорию меры. На другом конце шкалы трудностей для книг выбранного нами уровня—«Введение в математическую статистику» Р. Хогга и Ч. Ч. Крэга (3-е издание). Эти авторы рассматривают большинство затронутых нами проблем, но во многих случаях не считают нужным подробно обсудить вопросы, которые мы считаем важными, например теоремы существования, вычислительную сторону предлагаемых методов и поведение больших выборок.

Наша книга содержит больше материала, чем курс, рассчитанный на один семестр. В полугодовых курсах математической статистики, предназначенных для студентов, специализирующихся по математике, математической статистике, различным разделам физики, и будущих инженеров, мы читали основное ядро гл. 2—7 (от моделирования через оценивание и проверку гипотез до линейных моделей). Ощущая настоятельную необходимость материала, изложенного в гл. 10 (теория решений), мы включали в наши курсы по меньшей мере первые два раздела из этой главы. Кроме того, мы затрагивали ряд тем из гл. 8 по дискретным данным и гл. 9 по непараметрическим моделям.

Гл. 1 нашей книги посвящена не столько математической статистике, сколько теории вероятностей. К сожалению, многие из необходимых нам сведений не входят в стандартные учебники по теории вероятностей, но мы вынуждены отложить более основательное знакомство с ними до конца книги. При чтении лекционного курса гл. 1 можно не излагать отдельно, а объединять с материалом гл. 2—7 или включать в пропедевтический курс теории вероятностей, обычно предшествующий курсу математической статистики.

Особенностью нашей книги является большое число задач. Они различны по уровню — от тривиальных численных упражнений и элементарных задач, предназначенных для лучшего усвоения студентами излагаемого материала, до более трудных проблем, решаемых в тексте. Отбирая задачи, мы стремились, с одной стороны, предоставить возможность студенту проверить, насколько активно он овладел тем или иным разделом, а с другой стороны, продемонстрировать неисчерпаемое богатство идей и результатов, не включенных в текст книги по вполне понятным соображениям объема.

При написании книги мы придерживались следующих соглашений.

1) Для сокращения числа сносок мы поместили в конце каждой главы, перед задачами, примечания, расположенные по разделам, к которым они относятся. В пределах каждого раздела на соответствующее примечание в конце главы указывает его номер, например [1] означает первое примечание, [2] — второе и т. д. В примечаниях содержатся различного рода отступления, оговорки и дополнительные библиографические ссылки.

2) Перечень условных обозначений и сокращений с указанием тех мест, где они впервые вводятся, приведен после предисловия.

3) Основные (традиционные) обозначения для теоретико-вероятностных объектов (случайных величин, векторов, плотностей вероятности, функций распределения и моментов) вводятся в приложении.

Мы хотели бы выразить нашу признательность коллегам, студентам и друзьям, помогавшим нам на различных стадиях работы над книгой (черновые наброски, пробное издание, окончательный вариант рукописи).

Беркли — *Питер Дж. Бикел,*
1976 г. — *Куэлл Доксам*

● ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

$BB(n, \vartheta)$	биномиальное распределение с параметрами n и ϑ, 2, 223
$B(z, s)$	бета-функция Эйлера, вычисленная при заданных r и s, 1, 23
$\beta(r, s)$	бета-распределение с параметрами r и s, 1, 23, 24
χ_n^2	распределение хи-квадрат с n степенями свободы, 1, 23, 25
$x_n(p)$	p-й квантиль распределения χ_n^2, 1, 28
$EE(\lambda)$	экспоненциальное распределение с параметром λ, 2, 227
$FF_{k,m}$	F-распределение с k и m степенями свободы, 1, 26
$\Gamma(p)$	гамма-функция Эйлера, вычисленная при заданном p, 1, 22
$\Gamma(p, \lambda)$	гамма-распределение с параметрами p и λ, 1, 22, 23
$HH(D, N, n)$	гипергеометрическое распределение с параметрами D, N и n, 2, 223
$MM(n, \vartheta_1, \ldots, \vartheta_q)$	мультиномиальное распределение с параметрами n, $\vartheta_1, \ldots, \vartheta_q$, 2, 224
$NN(\mu, \sigma^2)$	нормальное распределение со средним μ и дисперсией σ^2, 2, 226
$NN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$	двумерное нормальное распределение, 1, 34
$NN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	двумерное нормальное распределение, 1, 34
Φ	стандартная нормальная функция распределения, 2, 226
φ	стандартная нормальная плотность, 2, 226
$z(p)$	p-й квантиль стандартного нормального распределения, 1, 28, 167
$PP(\lambda)$	распределение Пуассона с параметром λ, 2, 224
TT_k	t-распределение (Стьюдента) с k степенями свободы, 1, 26
$t_k(p)$	p-й квантиль распределения T_k, 1, 28, 170
$UU(a, b)$	равномерное распределение на отрезке (a, b), 2, 227
$I(\vartheta)$	количество информации по Фишеру, 1, 138
$I_1(\vartheta)$	$I(\vartheta)$ для одного наблюдения, 1, 141
I_A	индикатор множества A, 2, 217
ψ_X	производящая функция моментов случайной величины X, 2, 222
$L(\vartheta, x)$	функция правдоподобия, 1, 91
$LL(\vartheta, x)$	логарифм правдоподобия, 1, 112
$L(x, \vartheta_0, \vartheta_1)$	отношение правдоподобия, 1, 207
$\lambda(x)$	статистики критерия отношения правдоподобия, 1, 225

$X \sim F$	X распределен по закону F, 1, 225
$X \sim Y$	X имеет такое же распределение, как Y
$X = Y$	X равен Y с вероятностью 1, 2, 216
$\approx$	приближенное равенство
$\simeq$	пропорциональность, 1, 89
$T_n \xrightarrow{LL} F$, $LL(T_n) \to F$	сходимость по распределению к T, где $T \sim F$, 2, 228
$\xrightarrow{P}$	сходимость по вероятности, 2, 228
R	вещественная прямая
R^k	k-мерное евклидово пространство
■	конец примера или доказательства
н. д. г.	нижняя доверительная граница, 1, 168
о. н. к.	оценка наименьших квадратов, 1, 107
о. м. п.	оценка максимума правдоподобия, 1, 111
н. м.	наиболее мощный (критерий), 1, 207
с. к. о.	среднеквадратическая ошибка, 1, 45
в. д. г.	верхняя доверительная граница, 1, 169
р. н. м.	равномерно наиболее мощный (критерий), 1, 207
н. о. р. м. д.	несмещенная оценка с равномерно минимальной дисперсией, 1, 130

Глава 1 ● НЕКОТОРЫЕ ВОПРОСЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

В этой главе приведены некоторые сведения из теории вероятностей, существенные для нашего изложения статистики и не всегда рассматриваемые достаточно подробно в курсах теории вероятностей.

Теорию меры мы обходим стороной, условившись раз и навсегда, что все рассматриваемые нами множества и функции измеримы.

1.1. ВВЕДЕНИЕ УСЛОВНОСТИ С ПОМОЩЬЮ СЛУЧАЙНОЙ ВЕЛИЧИНЫ ИЛИ СЛУЧАЙНОГО ВЕКТОРА

Понятие условности играет важную роль при изучении связей между случайными переменными или векторами. В этом разделе мы изложим некоторые результаты, полезные для теории предсказания, теории оценивания и регрессии.

1.1.A. Дискретный случай

Читатель, несомненно, знаком с понятием условной вероятности — вероятности события A при условии, что произошло событие B. Пусть $\mathbf{X}$ и $\mathbf{Y}$ — дискретные случайные векторы. Мы хотим изучить распределение условной вероятности для вектора $\mathbf{X}$ при $\mathbf{Y} = \mathbf{y}$.

Условную функцию частоты $p(\cdot \mid \mathbf{y})$ для вектора $\mathbf{X}$ при $\mathbf{Y} = \mathbf{y}$ определим как

$$p(\mathbf{x} \mid \mathbf{y}) = P[\mathbf{X} = \mathbf{x} \mid \mathbf{Y} = \mathbf{y}] = \frac{p(\mathbf{x}, \mathbf{y})}{p_{\mathbf{Y}}(\mathbf{y})}, \qquad (1.1.1)$$

где p и $p_{\mathbf{Y}}$ — функции частоты для $(\mathbf{X}, \mathbf{Y})$ и $\mathbf{Y}$. Условная функция частоты p определена лишь для таких значений $\mathbf{y}$, для которых $P_{\mathbf{Y}}(\mathbf{y}) > 0$. Из приведенного нами определения ясно, что $p(\cdot \mid \mathbf{y})$ — функция частоты некоторого распределения вероятности, так как из (А. 8. 11) следует, что

$$\sum_{\mathbf{x}} p(\mathbf{x} \mid \mathbf{y}) = \frac{\sum_{\mathbf{x}} p(\mathbf{x}, \mathbf{y})}{p_{\mathbf{Y}}(\mathbf{y})} = \frac{p_{\mathbf{Y}}(\mathbf{y})}{p_{\mathbf{Y}}(\mathbf{y})} = 1.$$

Такое распределение вероятности называется *условным распределением* для $\mathbf{X}$ при $\mathbf{Y} = \mathbf{y}$.

Пример 1.1.1. Пусть $\mathbf{X} = (X_1, \ldots, X_n)$, где X_i — индикаторы серий из n биномиальных испытаний с вероятностью благоприятного исхода p. Пусть $Y = \sum_{i=1}^{n} X_i$ — общее число благоприятных исходов Тогда Y имеет биномиальное распределение $BB\,(n, p)$, и

$$p(\mathbf{x}\,|y) = \frac{P(\mathbf{X}=\mathbf{x},\, \mathbf{Y}=y)}{\binom{n}{y} p^y (1-p)^{n-y}} = \frac{p^y (1-p)^{n-y}}{\binom{n}{y} p^y (1-p)^{n-y}} = \frac{1}{\binom{n}{y}}, \quad (1.1.2)$$

если все x_i принимают значения 0 или 1 и $\Sigma x_i = y$.

Итак, если нам сообщают, что в серии из n биномиальных испытаний k исходов оказались благоприятными, то речь с одинаковой вероятностью может идти о любой из серий. ■

Пример 1.1.2. Пусть X и Y имеют совместную функцию частоты, заданную табл. 1.1.1.

Таблица 1.1.1

x \ y	0	10	20	$p_X(x)$
0	0,25	0,05	0,05	0,35
1	0,05	0,15	0,05	0,25
2	0,05	0,10	0,25	0,40
$p_Y(y)$	0,35	0,30	0,35	1

Пусть, например, Y — число сигарет (округленное до ближайшего числа, кратного 10), выкуриваемых за день индивидуумом, выбранным наугад из некоторой генеральной совокупности, а X — общая оценка состояния здоровья того же индивидуума, колеблющаяся от 0 (хорошее состояние) до 2 (плохое состояние) (1 соответствует среднему — ни хорошему, ни плохому — состоянию здоровья). При $y = 20$ получаем:

x	0	1	2
$p(x\|20)$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{5}{7}$

Эти данные указывают на взаимосвязь между привычкой много курить и плохим состоянием здоровья, так как $p\,(2\,|\,20)$ почти вдвое больше, чем $p_X(2)$. ■

Условное распределение для $\mathbf{X}$ при $\mathbf{Y} = \mathbf{y}$ легко вычисляется в двух частных случаях:

1) если $\mathbf{X}$ и $\mathbf{Y}$ независимы, то $p(\mathbf{x} \mid \mathbf{y}) = p_{\mathbf{X}}(\mathbf{x})$ и условное распределение совпадает с частным распределением;

2) если $\mathbf{X}$ есть функция от $\mathbf{Y}$ ($\mathbf{X}=h(\mathbf{Y})$), то условное распределение для $\mathbf{X}$ вырождено: $\mathbf{X} = h(\mathbf{y})$ с вероятностью 1.

Оба утверждения следуют непосредственно из определения (1.1.1).

Из (1.1.1) и (П.4.5) следуют две важные формулы. Пусть $q(\mathbf{y} \mid \mathbf{x})$ — условная функция частоты для $\mathbf{Y}$ при $\mathbf{X} = \mathbf{x}$. Тогда

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x} \mid \mathbf{y})\, p_{\mathbf{Y}}(\mathbf{y}), \tag{1.1.3}$$

$$p(\mathbf{x} \mid \mathbf{y}) = \frac{q(\mathbf{y} \mid \mathbf{x})\, p_{\mathbf{X}}(\mathbf{x})}{\sum_{\mathbf{z}} q(\mathbf{y} \mid \mathbf{z})\, p_{\mathbf{X}}(\mathbf{z})} \quad \text{(правило Байеса)} \tag{1.1.4}$$

(знаменатель в правой части положителен).

Равенством (1.1.3) можно воспользоваться при построении моделей. Предположим, например, что Y — число дефектных изделий в партии из N изделий, изготовленных с помощью некоторого технологического процесса, обладает распределением $BB(N, \vartheta)$. Предположим, что из партии n раз производится выборка без возвращения, и пусть X — число дефектных изделий, обнаруженных в выборке. Известно, что X при заданном $Y = y$ обладает гипергеометрическим распределением $HH(y, N, n)$. Воспользуемся равенством (1.1.3) и запишем совместное распределение для X и Y:

$$P[X=x, Y=y] = \binom{N}{y} \vartheta^{y} (1-\vartheta)^{N-y} \frac{\binom{y}{x}\binom{N-y}{n-x}}{\binom{N}{n}}, \tag{1.1.5}$$

где комбинаторные коэффициенты $\binom{a}{b}$ обращаются в нуль, если a и b — не целые числа, удовлетворяющие неравенству $b \leqslant a$.

Этой же моделью можно воспользоваться, чтобы проиллюстрировать (1.1.4). Поскольку обычно нам удается наблюдать X, может оказаться желательным узнать условное распределение для Y при $X = x$. Из (1.1.4) получаем

$$P[Y=y \mid X=x] = \binom{N}{y} \vartheta^{y} (1-\vartheta)^{N-y} \binom{y}{x}\binom{N-y}{n-x} \Big/ c(x),$$

где

$$c(x) = \sum_{y} \binom{N}{y} \vartheta^{y} (1-\vartheta)^{N-y} \binom{y}{x}\binom{N-y}{n-x}.$$

Эта формула упрощается (см. задачу 1.1.11) до биномиального распределения

$$P[Y=y \mid X=x] = \binom{N-n}{y-x} \vartheta^{y-x} (1-\vartheta)^{N-n-(y-x)}. \tag{1.1.6}$$

1.1.Б. Условное математическое ожидание для дискретных величин

Пусть X — случайная величина с $E(|X|) < \infty$. *Условное математическое ожидание* для X при $\mathbf{Y}=\mathbf{y}$ (обозначим его $E = (X|\mathbf{Y}=\mathbf{y})$) по определению есть

$$E(X|\mathbf{Y}=\mathbf{y}) = \sum_x xp(x|\mathbf{y}). \tag{1.1.7}$$

Заметим, что в силу (1.1.1), если $p_{\mathbf{Y}}(\mathbf{y}) > 0$,

$$\sum_x |x|\, p(x|\mathbf{y}) \leqslant \sum_x |x| \frac{p_X(x)}{p_{\mathbf{Y}}(\mathbf{y})} = \frac{E(|X|)}{P_{\mathbf{Y}}(\mathbf{y})}. \tag{1.1.8}$$

Пример 1.1.3. Предположим, что X и Y имеют совместную функцию частоты, заданную табл. 1.1.1. Тогда

$$E(X|Y=20) = 0\cdot\frac{1}{7} + 1\cdot\frac{1}{7} + 2\cdot\frac{5}{7} = \frac{11}{7} = 1{,}57.$$

Аналогичным образом, $E(X \mid Y = 10) = \frac{7}{6} = 1{,}17$ и $E(X \mid Y=0) = = \frac{3}{7} = 0{,}43$. Заметим, что в контексте разговора о вреде курения величину $E(X \mid Y = y)$ можно рассматривать как среднюю оценку состояния здоровья тех, кто выкуривает за день y сигарет. ■

Пусть $g(\mathbf{y}) = E(X \mid \mathbf{Y} = \mathbf{y})$. Случайную величину $g(\mathbf{Y})$ будем записывать в виде $E(X \mid \mathbf{Y})$ и называть *условным математическим ожиданием для* X *при заданном* $\mathbf{Y}$*.

В качестве примера вычислим $E(X_1| Y)$ для X_1 и Y из примера 1.1.1:

$$E(X_1, Y=i) = P[X_1=1|Y=i] = \frac{\binom{n-1}{i-1}}{\binom{n}{i}} = \frac{i}{n}, \text{ где } \binom{n-1}{-1} = 0. \tag{1.1.9}$$

Первое из этих равенств выполняется потому, что X_1 — характеристическая функция. Вто́рое равенство следует из (1.1.2), так как $\binom{n-1}{i-1}$ — число способов, которыми i благоприятных исходов могут распределиться среди n испытаний, если известно, что исход первого испытания благоприятен. Следовательно,

$$E(X_1|Y) = \frac{Y}{n}. \tag{1.1.10}$$

Условное распределение для случайного вектора $\mathbf{X}$ при $\mathbf{Y} = \mathbf{y}$ соответствует единственной вероятностной мере $P_{\mathbf{y}}$ на (Ω, AA). В частности, определим для $A \in AA$

$$P_{\mathbf{y}}(A) = P(A \mid [\mathbf{Y} = \mathbf{y}]), \text{ если } p_{\mathbf{Y}}(\mathbf{y}) > 0. \tag{1.1.11}$$

---

* Следуя традиции, мы будем также обозначать $E(X, \mathbf{Y})$ любую величину, равную $g(\mathbf{Y})$ с вероятностью 1.

Величина P_y есть не что иное, как условная вероятностная мера на (Ω, AA), задаваемая соотношением (П.4.2.). Но условное распределение для $\mathbf{X}$ при $\mathbf{Y}=\mathbf{y}$ — то же самое, что и распределение для $\mathbf{X}$, если $P_\mathbf{y}$ — вероятностная мера на (Ω, AA). Следовательно, условное математическое ожидание совпадает с обычным математическим ожиданием по вероятностной мере $P_\mathbf{y}$. Это означает, что все свойства математического ожидания, перечисленные в (П.10.3) — (П.10.8), остаются в силе и для условного математического ожидания при $\mathbf{Y} = \mathbf{y}$. Например, равенство

$$E(\alpha X_1 + \beta X_2 \mid \mathbf{Y} = \mathbf{y}) = \alpha E(X_1 \mid \mathbf{Y} = \mathbf{y}) + \beta E(X_2 \mid \mathbf{Y} = \mathbf{y}) \tag{1.1.12}$$

выполняется тождественно по $\mathbf{y}$ при любых X_1 и X_2, таких, что математические ожидания $E(|X_1|)$, $E(|X_2|)$ конечны. Так как это тождество справедливо при всех $\mathbf{y}$,

$$E(\alpha X_1 + \beta X_2 \mid \mathbf{Y}) = \alpha E(X_1 \mid \mathbf{Y}) + \beta E(X_2 \mid \mathbf{Y}). \tag{1.1.13}$$

Применяя этот прием к любому из соотношений (П.10.3) — (П.10.8), мы получаем аналогичные свойства условного ожидания. Важный пример такого рода рассуждений приведен в разд. 1.6.

В двух частных случаях условные математические ожидания удается вычислить непосредственно. Если X и $\mathbf{Y}$ независимы и $E(|X|) < \infty$, то

$$E(X \mid \mathbf{Y}) = E(X). \tag{1.1.14}$$

Это следует из (1).

С другой стороны, из (2) мы получаем, что

$$E(h(\mathbf{Y}) \mid \mathbf{Y}) = h(\mathbf{Y}). \tag{1.1.15}$$

В (1.1.15) неявно содержится утверждение о том, что если $\mathbf{Y} = \mathbf{y}$, то $\mathbf{Y}$ действует как константа. Продолжив это рассуждение, мы придем к соотношению, которое можно назвать *теоремой о подстановке для условных математических ожиданий*: равенство

$$E(q(\mathbf{X}, \mathbf{Y}) \mid \mathbf{Y} = \mathbf{y}) = E(q(\mathbf{X}, \mathbf{y}) \mid \mathbf{Y} = \mathbf{y}) \tag{1.1.16}$$

выполняется при всех $\mathbf{y}$, таких, что $p_\mathbf{Y}(\mathbf{y}) > 0$ и $q(\mathbf{X}, \mathbf{Y})$ имеет конечное математическое ожидание. Утверждение (1.1.16) следует непосредственно из определений, так как

$$P[q(\mathbf{X}, \mathbf{Y}) = a \mid \mathbf{Y} = \mathbf{y}] = P[q(\mathbf{X}, \mathbf{Y}) = a, \ \mathbf{Y} = \mathbf{y} \mid \mathbf{Y} = \mathbf{y}] = \\ = P[q(\mathbf{X}, \mathbf{y}) = a \mid \mathbf{Y} = \mathbf{y}] \tag{1.1.17}$$

при любом a.

Если $q(\mathbf{X}, \mathbf{Y}) = r(\mathbf{X})\, h(\mathbf{Y})$, где функция h ограничена, а $r(\mathbf{X})$ имеет конечное математическое ожидание, то из (1.1.16) получаем

$$E(r(\mathbf{X})\, h(\mathbf{Y}) \mid \mathbf{Y} = \mathbf{y}) = E(r(\mathbf{X})\, h(\mathbf{y}) \mid \mathbf{Y} = \mathbf{y}) = \\ = h(\mathbf{y})\, E(r(\mathbf{X}) \mid \mathbf{Y} = \mathbf{y}). \tag{1.1.18}$$

Следовательно,

$$E(r(\mathbf{X})\, h(\mathbf{Y}) \mid \mathbf{Y}) = h(\mathbf{Y})\, E(r(\mathbf{X}) \mid \mathbf{Y}). \tag{1.1.19}$$