

ВЕРОЯТНОСТЬ. МЕРА

Одномерные и многомерные распределения вероятности; функции распределения; плотности; каноническое разложение неубывающей функции; дискретные, атомные, сингулярные, непрерывные и абсолютно непрерывные распределения вероятности; неубывающие функции нескольких переменных; среднее; математическое ожидание; моменты; стандартное отклонение; характеристическая функция; коэффициенты корреляции и корреляционные матрицы; меры; функции множеств; теорема о расширении и теорема Рисса о представлении для мер; выборка; выборочное среднее; выборочная дисперсия; маргинальная и условная вероятности; нормальное распределение; центральная предельная теорема; метод Монте-Карло; вероятность и мера в гильбертовом пространстве.

Предварительные сведения: гл. 1—3, интеграл Стильбеса.

Понятие функции распределения вероятностей создает основу изучения вероятностей в конечномерных пространствах. В большинстве случаев в физических приложениях вероятности или дискретны, или абсолютно непрерывны, или представляют собой смесь того и другого, но для полноты концептуальной основы требуется также понятие сингулярно непрерывных распределений. Теория вероятностей применяется главным образом в квантовой механике (что мы обсудим в следующей главе), в статистической механике, в анализе ошибок и в методах Монте-Карло. Замечательным явлением оказывается стремление в среднем к универсальному, так называемому нормальному распределению, составляющее суть центральной предельной теоремы. Функция распределения, маргинальная и условная вероятности используются в методах Монте-Карло, в которых моделируются путем вычислений естественные случайные явления, плохо поддающиеся анализу. Представления о вероятностях и мерах как функциях множеств необходимы для изучения вероятности в бесконечномерных и абстрактных пространствах, которые появляются в статистической механике и в теории стохастических процессов.

13.1. ОДНОМЕРНЫЕ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ.**ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ. ПЛОТНОСТЬ**

Результаты многих независимых повторений какого-либо наблюдения или эксперимента (например, над атомной системой) можно описать при помощи распределения вероятности. Результатом отдельного наблюдения является совокупность значений некоторых величин; скажем n величин, таких, как углы отклонения,

напряжения, энергии, показания счетчиков и т. п. Итак, этот результат может быть представлен точкой некоторого n -мерного пространства. Для большинства элементарных процессов характерно то, что, как бы тщательно ни контролировались условия опыта, эти результаты меняются значительно и случайным образом от одного повторения к другому. Полученное распределение точек в n -мерном пространстве описывается функцией распределения в этом пространстве. В данном рассмотрении вероятность представляется интуитивным понятием: например, утверждение, что нейтрон имеет вероятность, равную 0.316, пройти через слой фольги без столкновений, подразумевает следующее: (1) так случилось с нейтронами в 31.6% большого числа испытаний, (2) предполагается, что при последующих испытаниях нейтроны будут вести себя аналогично, причем соответствующее процентное отношение будет стремиться к некоторому значению, близкому к 31.6%, когда число испытаний неограниченно возрастает.

В теории вероятностей случаи $n = 1$ и $n > 1$ обычно называются *одномерным* и *многомерным* соответственно. Начнем с одномерного случая. Если ξ — величина, которая определяется в результате эксперимента, то функция F , определяемая из условия

$$F(x) = \mathbf{P}\{\xi \leq x\}, \quad (13.1.1)$$

где для любого вещественного x $\mathbf{P}\{\xi \leq x\}$ обозначает вероятность того, что измеряемая величина ξ меньше или равна x , называется *функцией распределения случайной переменной*¹⁾. Другие вероятности могут быть выражены через эту функцию F ; в частности, вероятность того, что ξ попадает в интервал $(x_1, x_2]$, имеет вид

$$\mathbf{P}\{x_1 < \xi \leq x_2\} = F(x_2) - F(x_1); \quad (13.1.2)$$

если функция F имеет скачки, то они описывают вероятность попадания в точки, именно

$$\mathbf{P}\{\xi = x_0\} = F(x_0 + 0) - F(x_0 - 0). \quad (13.1.3)$$

Члены в правой части этой формулы обозначают предельные значения, достигаемые функцией $F(x)$, когда $x \rightarrow x_0$ справа и слева соответственно.

Из определения (13.1.1) следует, что $F(x)$ — неубывающая и непрерывная справа функция, принимающая значения от 0 до 1, т. е.

$$F(b) \geq F(a), \quad \text{если } b \geq a, \quad (13.1.4)$$

$$F(a + \epsilon) \rightarrow F(a) \quad \text{при } \epsilon \downarrow 0, \quad (13.1.5)$$

$$F(-\infty) = 0, \quad F(+\infty) = 1. \quad (13.1.6)$$

¹⁾ Автор называет эту функцию кумулятивной вероятностью (cumulative probability), но мы будем придерживаться общепринятого термина. — *Прим. перев.*

Обратно, любая функция с такими свойствами описывает некоторое распределение вероятности. Например, если известна F , то можно сгенерировать с любой точностью множество случайных чисел, имеющих соответствующее распределение, используя для этой цели вычислительную машину.

Требование односторонней непрерывности (13.1.5) довольно произвольно, и мы будем часто записывать равенства, подобные (13.1.3), таким образом, чтобы они не зависели от этого требования. (Правая часть (13.1.3) не зависит от соглашения, что $F(x_0 + 0) = F(x_0)$.)

ПРИМЕР 1

Если ξ принимает лишь конечное число значений x_1, \dots, x_N , записанных, допустим, в порядке возрастания и если $P\{\xi = x_i\} = p_i$, то $F(x)$ — ступенчатая функция:

$$F(x) = \begin{cases} 0 & \text{при } -\infty < x < x_1, \\ \sum_{k=1}^{i-1} p_k & \text{при } x_{i-1} \leq x < x_i \quad (i=2, \dots, N), \\ 1 & \text{при } x_N \leq x < \infty. \end{cases} \quad (13.1.7)$$

Такого рода функции пригодны для рассмотрения бросаний монеты и игральной кости и вообще конечной игры. Они также удобны для описания распределения энергий фотонов, испускаемых при скачкообразном переходе атома из некоторого возбужденного состояния на более низкий энергетический уровень (см. рис. 13.1).

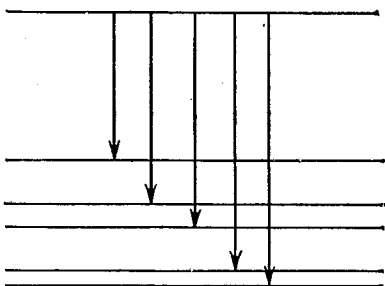


Рис. 13.1. Диаграмма уровней энергии.

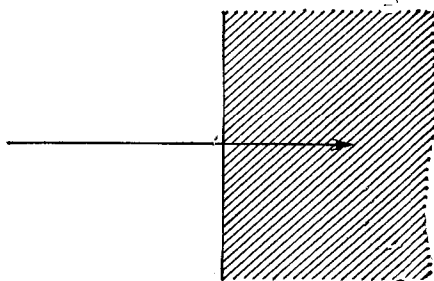


Рис. 13.2. Пробег нейтрона для примера 2.

ПРИМЕР 2

Нейтрон или фотон проникает в вещество равномерной плотности и проходит расстояние ξ до столкновения (рис. 13.2). Функция распределения случайной переменной ξ имеет вид

$$F(x) = \begin{cases} 0 & \text{при } x < 0, \\ 1 - e^{-x/\lambda} & \text{при } x > 0, \end{cases}$$

где λ — длина среднего свободного пробега (см. рис. 13.3). Здесь F имеет производную $f(x) = F'(x)$, которая непрерывна всюду, исключая скачок в точке

$x=0$. Эта производная называется *плотностью распределения вероятности*, так как для любого x_0

$$f(x_0) = \lim_{\Delta x \rightarrow 0} (1/\Delta x) P\{x_0 < \xi < x_0 + \Delta x\}. \quad (13.1.8)$$

Такое распределение вероятности называется *непрерывным*, или точнее *абсолютно непрерывным* (см. ниже определение).

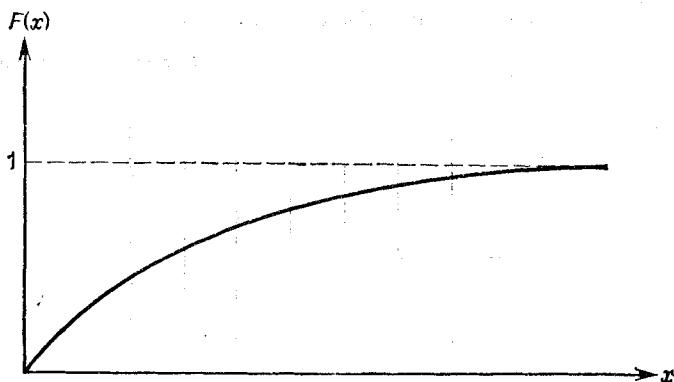


Рис. 13.3. Функция распределения для примера 2.

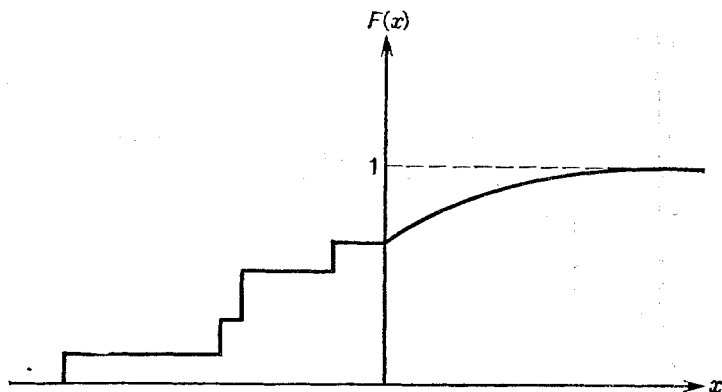


Рис. 13.4. Функция распределения для примера 3.

Пример 3

Рассмотрим атом в основном состоянии, находящийся в поле излучения с непрерывным спектром. После поглощения кванта атом может перейти в одно из различных возбужденных состояний или в состояние непрерывного спектра при энергии, большей энергии ионизации, которую примем за нуль энергии. Распределение вероятности энергии E после поглощения одного кванта частично дискретно, частично непрерывно, как на рис. 13.4. Функция распределения $F(x) = P\{E \leq x\}$ является ступенчатой функцией при $x < 0$ и непрерывной функцией при $x \geq 0$.

ПРИМЕР 4

Нейтрон или фотон проходит через бесконечную последовательность параллельных тонких поглощающих слоев фольги, расположенных на одинаковых расстояниях друг от друга, как на рис. 13.5. Пусть ξ — расстояние от первого слоя до того слоя, где происходит поглощение, и пусть α — вероятность поглощения в каждом слое ($0 < \alpha < 1$). Тогда функция распределения переменной ξ записывается так:

$$F(x) = 1 - \alpha^n \quad \text{при} \quad (n-1)d < x \leq nd, \quad n = 1, 2, \dots,$$

где d — расстояние между последовательными слоями. Эта ступенчатая функция с бесконечным числом ступеней изображена на рис. 13.6.

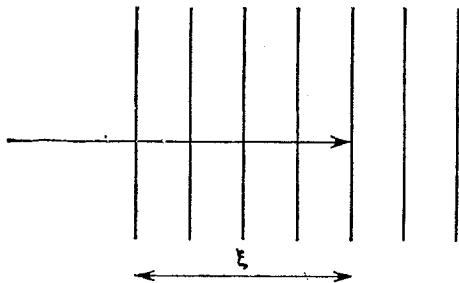


Рис. 13.5. Пробег нейтрона для примера 4.

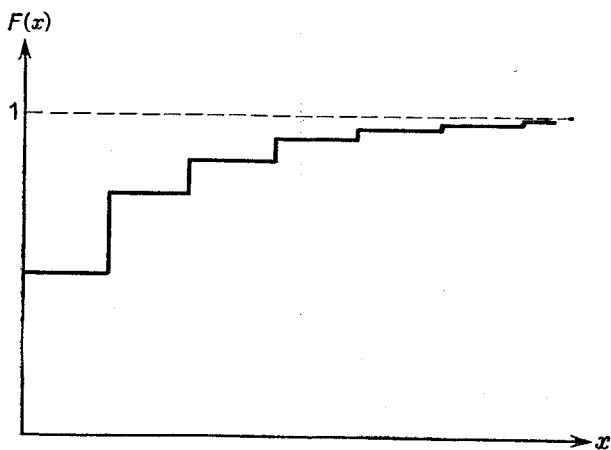


Рис. 13.6. Функция распределения для примера 4.

ПРИМЕР 5

Пусть в последнем примере $\varphi = \varphi(t) = \varphi_0 \sin \omega t$ — переменное напряжение, имеющееся в цепи, в то время как частица движется через последовательность слоев фольги; пусть τ — время, которое требуется для того, чтобы частица прошла от одного слоя до следующего, и пусть $\theta = \omega\tau$. Тогда в момент поглощения частицы напряжение φ имеет значение 0 с вероятностью $1 - \alpha$, зна-

чение $\varphi_0 \sin \theta$ — с вероятностью $\alpha - \alpha^2, \dots$, значение $\varphi_0 \sin n\theta$ — с вероятностью $\alpha^n (1 - \alpha)$ и т. д. Функция распределения $F(x) = P\{\varphi \leq x\}$ имеет вид

$$F(x) = \sum_{n=0}^{\infty} \alpha^n (1 - \alpha) \quad (\varphi_0 \sin n\theta \leq x);$$

здесь суммирование проводится по всем таким n , что $\varphi_0 \sin n\theta \leq x$. Если θ — иррациональное кратное π , то $F(x)$ имеет бесконечно много скачков, плотно распределенных в интервале $[-\varphi_0, \varphi_0]$.

ПРИМЕР 6 (функция Кантора)

Предположим, что цифровая вычислительная машина имеет приспособление (использующее радиоактивный распад, тепловой шум или что-либо подобное), которое бесконечно генерирует по требованию последовательность независимых случайных чисел x_1, x_2, \dots , равномерно распределенных на отрезке

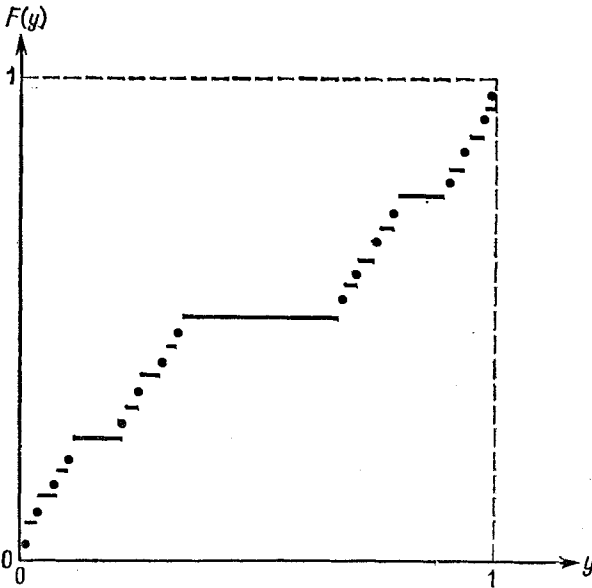


Рис. 13.7. Функция Кантора.

$\{0, 1\}$; эти числа берутся в качестве значений случайной переменной ξ . Допустим, что каждое число x выражено в виде бесконечного двоичного разложения

$$x = 0,abcd\dots, \quad (13.1.9)$$

где каждая цифра a, b, c, d, \dots имеет значение 0 или 1. Допустим также, что в вычислительной машине имеется подпрограмма, преобразующая каждое такое x в число

$$y = 0,aabbccdd\dots \quad (13.1.10)$$

путем дублирования каждой цифры; числа y принимаем в качестве значений другой случайной переменной η . Легко найти функцию распределения F для переменной η . Любое число y вида (13.1.10) обязательно либо меньше $0,01 = 1/4$, либо больше (или равно) $0,11 = 3/4$ в зависимости от того, меньше x $0,1 = 1/2$

либо больше этой величины (или равно ей). Поэтому $F(y)$ имеет постоянное значение $1/2$ при $1/4 \leq y \leq 3/4$. Аналогично $F(y)$ имеет значение $1/4$ при $1/16 \leq y \leq 3/16$; значение $3/4$ при $13/16 \leq y \leq 15/16$ и т. д. Если y имеет вид (13.1.10), где цифры попарно равны в бесконечном двоичном разложении, то для такого y $F(y) = 0.abc\dots$, т. е.

$$P\{\eta < y = 0.aabbcc\dots\} = 0.abc\dots \quad (13.1.11)$$

Если y не представляется в виде (13.1.10), то это значение лежит в одном из интервалов постоянства F , описанных выше.

Сумма длин всех интервалов постоянства F равна

$$\frac{1}{2} + 2 \cdot \frac{1}{8} + 4 \cdot \frac{1}{32} + \dots + \frac{1}{2^n} + \dots = 1,$$

т. е. эти интервалы заполняют интервал $[0, 1]$, а то, что отброшено, имеет меру нуль. С другой стороны, F непрерывна, поскольку: (1) если y_0 имеет вид (13.1.10), то утверждение $y \rightarrow y_0$ с очевидностью влечет за собой $x \rightarrow x_0$ в соответствии с (13.1.11), (2) любое другое y_0 принадлежит интервалу постоянства, а следовательно, тем более интервалу непрерывности функции F . График функции $F(y)$ изображен на рис. 13.7.

Отступление по поводу множеств меры нуль

Множество S на \mathbb{R} имеет меру нуль, если оно может быть покрыто совокупностью интервалов, сумма длин которых произвольно мала. Например, пусть S состоит из рациональных чисел в $(0, 1)$. Они могут быть записаны в виде последовательности $\alpha_1, \alpha_2, \alpha_3, \dots$, например, так: $1/2, 1/3, 2/3, 1/4, 3/4, 1/5, 2/5, 3/5, \dots$. Для любого заданного $\varepsilon > 0$ α_1 может быть заключено в интервал длины $\varepsilon/2$, α_2 — в интервал длины $\varepsilon/4$, \dots , α_n — в интервал длины $\varepsilon/2^n$ и т. д. Эти интервалы покрывают S , а сумма их длин равна ε . Следовательно, рациональные числа образуют множество меры нуль.

Если в примере 6 отбросить первые $1 + 2 + 4 + \dots + 2^{n-1}$ интервалов постоянства функции $F(y)$ (взятых в описанном выше порядке), то остаток интервала $[0, 1]$ будет состоять из интервалов с суммарной длиной, равной $1/2^n$, что может быть сделано произвольно малым при достаточно большом n . Таким образом, множество значений y , в которых F не является постоянной, имеет меру нуль. Иначе говоря, производная $F'(y)$ существует и равна нулю везде, исключая множество меры нуль.

Если некоторое соотношение справедливо на всем \mathbb{R} , исключая, возможно, множество меры нуль, то говорят, что это соотношение справедливо почти всюду (или почти везде). При этом предполагается, что интервалы, о которых шла речь выше, не являются вырожденными (интервалами $[a, a]$, состоящими из единственной точки); поэтому можно допустить, что все интервалы являются открытыми. В \mathbb{R}^n множество меры нуль определяется как множество, которое может быть заключено в открытое множество произвольно малого объема.

Рассмотренная функция $F(y)$ представляет собой известный пример Кантора непрерывной отличной от постоянной функции, производная которой $f(y) = F'(y)$ равна нулю почти всюду.

Если функция распределения $F(x)$ независимо от того, является ли она непрерывной, имеет производную, равную нулю почти везде, то распределение вероятности называется *сингулярным*. Распределение в последнем примере является как непрерывным, так и сингулярным.

Используя две теоремы о декомпозиции, одна из которых принадлежит Жордану, а другая — Лебегу (см. книгу Феллера [1966]), любую неубывающую функцию $F(x)$ можно записать в виде суммы трех неубывающих функций:

$$F(x) = F_1(x) + F_2(x) + F_3(x).$$

Здесь $F_1(x)$ — чисто ступенчатая функция со скачками p_1, p_2, \dots в точках x_1, x_2, \dots (не обязательно упорядоченных), т. е. имеющая вид

$$F_1(x) = \sum_{x_i \leq x} p_i \quad (\forall p_i > 0).$$

Далее, $F_2(x)$ непрерывна и является интегралом от своей производной $f(x) \geq 0$ (в смысле Лебега):

$$F_2(x) = \int_0^x f(y) dy, \quad f(x) = F_2'(x).$$

Наконец, $F_3(x)$ сингулярна и непрерывна. Если $F(x)$ — чисто ступенчатая функция, как в рассмотренных выше примерах 1, 4 и 5, то распределение вероятности называется *атомным*; в примерах 1 и 4 распределения называют также *дискретными* (скачки происходят в изолированных точках). Функция $F_2(x)$ принадлежит к классу функций, называемых *абсолютно непрерывными*; см. § 13.10.

13.2. СРЕДНИЕ И МАТЕМАТИЧЕСКИЕ ОЖИДАНИЯ

Допустим, что $F(x)$ — функция распределения случайной переменной ξ и что желательно найти среднее значение ξ для некоторой продолжительной последовательности измерений. Сначала рассмотрим случай, в котором ξ — ограниченная случайная переменная, т. е. предположим, что $F(x) = 0$ при $x < a$ и $F(x) = 1$ при $x \geq b$. Таким образом, все измеренные значения ξ находятся в интервале $[a, b]$. Чтобы получить приближенно среднее значение, разобьем интервал $[a, b]$ на N подынтервалов точками x_i (i от 0 до N) так, что

$$a = x_0 < x_1 < \dots < x_N = b.$$

Обозначим через x'_i произвольную точку в подынтервале $[x_{i-1}, x_i]$ ($i = 1, 2, \dots, N$). Доля измерений, в которых ξ находится между x_{i-1} и x_i , составляет $F(x_i) - F(x_{i-1})$. Поэтому приближенно среднее значение равно

$$\sum_{i=1}^N x'_i [F(x_i) - F(x_{i-1})].$$

В пределе, когда разбиение бесконечно измельчается ($N \rightarrow \infty$), эта сумма стремится к интегралу Стильтеса

$$\int_a^b x dF(x). \quad (13.2.1)$$

Теория интеграла Стильтеса от непрерывной функции почти совпадает с соответствующей теорией интеграла Римана (см. книгу Натансона [1950, гл. 8]). Интеграл Стильтеса широко используется в физике. В частности, в рассмотренном выше примере 3 такое использование позволяет включить в один член суммирование по дискретным состояниям и интегрирование по непрерывным состояниям. Для сумм Римана—Стильтеса обычно используются следующие обозначения: интервал $x_{i-1} < x \leq x_i$ обозначается через Δ_i , а $F(x_i) - F(x_{i-1})$ — через $F(\Delta_i)$. Тогда рассматриваемая сумма записывается как $\sum x'_i F(\Delta_i)$, а соответствующая сумма Римана—Стильтеса для непрерывной функции $\varphi(x)$ — как $\sum \varphi(x'_i) F(\Delta_i)$. Обозначения такого вида особенно удобны в многомерном случае, который будет обсуждаться в следующем параграфе.

Интеграл (13.2.1) называется *средним значением* или *ожидаемым значением* или *математическим ожиданием* величины ξ и обозначается через $E(\xi)$ или μ . В общем случае неограниченной случайной переменной математическое ожидание имеет вид

$$\mu = E(\xi) = \int_{-\infty}^{\infty} x dF(x) = \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_a^b x dF(x) \quad (13.2.2)$$

при условии, что этот предел существует. Если предел не существует, то среднее значение измерений величины не стремится к какому-либо фиксированному предельному значению при неограниченном росте числа повторений эксперимента.

Пусть теперь φ — непрерывная функция. Каждому измеренному значению x величины ξ соответствует некоторое число $\varphi(x)$, и эти числа являются значениями случайной переменной, обозначаемой через $\varphi(\xi)$. Ожидаемое значение $\varphi(\xi)$ представляет собой

$$E(\varphi(\xi)) = \int_{-\infty}^{\infty} \varphi(x) dF(x) \quad (13.2.3)$$

снова при условии, что интеграл существует. Отметим некоторые важные случаи.

Случай 1. $\varphi(x)$ — степень x . Величины $E(\xi^k)$, $k=1, 2, \dots$, если они существуют, называются *моментами* данного распределения. Первый из них есть $E(\xi) = \mu$. Важной комбинацией первых двух моментов является *дисперсия* $\sigma^2 = E((\xi - \mu)^2)$. Так как $E(1) = 1$ и $E(\xi) = \mu$, дисперсию σ^2 можно записать также в виде

$$\sigma^2 = E(\xi^2) - 2\mu E(\xi) + \mu^2 E(1) = E(\xi^2) - \mu^2;$$

величина σ называется *стандартным отклонением* ξ . Если ξ — ограниченная случайная переменная, то моменты всех порядков существуют и конечны. Важнейшей классической задачей является *проблема моментов*, состоящая в вычислении F , когда все моменты известны (см. книгу Феллера, т. 2).

Случай 2. φ — ограниченная непрерывная функция. Тогда $E(\varphi(\xi))$ всегда существует. Наиболее важный пример дает $\varphi(x) = e^{-i\lambda x}$ (λ вещественно). Комплекснозначная функция

$$\chi(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda x} dF(x) \quad (13.2.4)$$

вещественного параметра λ называется *характеристической функцией* распределения вероятности или случайной переменной ξ . Эта функция играет важную роль в доказательстве центральной предельной теоремы в § 13.6.

Случай 3. φ — пробная функция ($\varphi \in C_0^\infty$). Тогда распределение f (в смысле Шварца) можно определить как линейный функционал

$$\langle f, \varphi \rangle = E(\varphi(\xi)) = \int_{-\infty}^{\infty} \varphi(x) dF(x) \quad \forall \varphi \in C_0^\infty. \quad (13.2.5)$$

Интегрирование по частям (по поводу интегрирования по частям в интегралах Стильтьеса см. книгу Натансона [1950]) дает

$$\langle f, \varphi \rangle = \varphi(x) F(x) \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} F(x) \varphi'(x) dx,$$

но проинтегрированный член обращается в нуль, и, следовательно,

$$\langle f, \varphi \rangle = -\langle F, \varphi' \rangle,$$

где F — распределение, определяемое через функцию $F(x)$ формулой

$$\langle F, \psi \rangle = \int_{-\infty}^{\infty} F(x) \psi(x) dx \quad \forall \psi \in C_0^\infty.$$

Таким образом, $f = F'$. Распределение f называется *вероятностной мерой* случайной величины ξ (f является обычной функцией тогда и только тогда, когда $F(x)$ — абсолютно непрерывная функция, и в таком случае $f(x)$ представляет собой плотность вероятности $F'(x)$ величины ξ).

Функция $F(x)$ является распределением медленного роста, поскольку она ограничена; поэтому $f = F'$ тоже распределение медленного роста, а характеристическая функция $\chi(\lambda)$ равна умноженному на $\sqrt{2\pi}$ преобразованию Фурье от f .

13.3. ДВУМЕРНЫЕ И МНОГОМЕРНЫЕ РАСПРЕДЕЛЕНИЯ. НЕУБЫВАЮЩИЕ ФУНКЦИИ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ

Если каждое повторение эксперимента дает два числа x, y , являющихся значениями случайных переменных ξ и η , то распределение полученных точек на плоскости x, y называется *двумерным*. Сейчас мы будем рассматривать эти двумерные распределения; многомерный случай будет очевидным обобщением, и мы кратко обсудим его в конце данного параграфа.

Совместная функция распределения $F(\cdot, \cdot)$ задается в виде

$$F(x, y) = \mathbf{P} \{ \xi \leq x, \eta \leq y \} \quad (13.3.1)$$

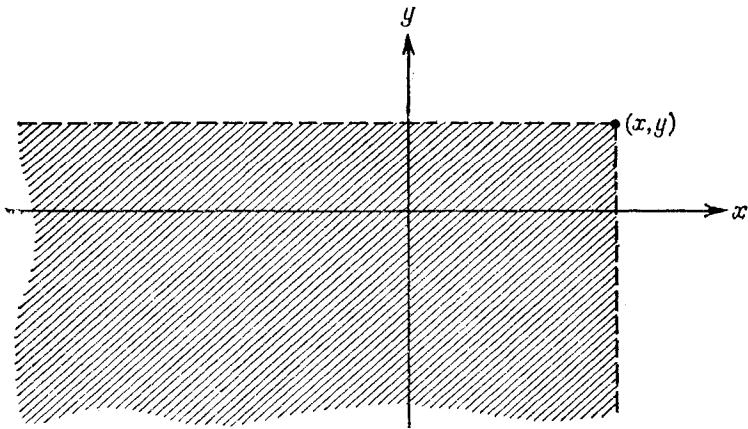


Рис. 13.8. Область плоскости, указанная в определении (13.3.1),

и представляет собой долю тех экспериментов, для которых результирующие точки лежат в квадранте, расположенном левее и ниже точки (x, y) (см. рис. 13.8). Другие вероятности могут быть выражены через функцию $F(\cdot, \cdot)$. Если $x_1 < x_2$, то $F(x_2, y) - F(x_1, y)$ есть вероятность того, что $x_1 < \xi \leq x_2$, тогда как η

имеет любое значение, не превышающее y . Если к тому же $y_1 < y_2$, то

$$\mathbf{P}\{x_1 < \xi \leq x_2, y_1 < \eta \leq y_2\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1). \quad (13.3.2)$$

Если указанную прямоугольную область плоскости обозначить через \square :

$$\square = \{(x, y): x_1 < x \leq x_2, y_1 < y \leq y_2\},$$

то (13.3.2) можно записать в виде

$$\mathbf{P}\{(\eta, \xi) \in \square\} = F(\square), \quad (13.3.3)$$

где $F(\square)$ стоит вместо правой части (13.3.2).

Ясно, что $F(\infty, \infty) = 1$, а $F(x, -\infty) = 0$ для любого x и $F(-\infty, y) = 0$ для любого y . Требование о неубывании $F(\cdot)$ в одномерном случае заменяется здесь требованием, чтобы $F(\square) \geq 0$ для любого прямоугольника с условием $x_1 \leq x_2$ и $y_1 \leq y_2$. Это требование включает также предельный случай $F(x_2, y_2) - F(x_1, y_2) \geq 0$ при $y_1 \rightarrow -\infty$ и предельный случай $F(x_2, y_2) - F(x_2, y_1) \geq 0$ при $x_1 \rightarrow -\infty$. При этих условиях $F(x, y)$ называется *неубывающей функцией* двух переменных x и y .

Для того чтобы найти среднее значение непрерывной функции $\varphi(\xi, \eta)$, допустим сначала, что ξ и η ограничены так, что получаемые из эксперимента точки (x, y) лежат в прямоугольнике $a < x < b$, $c < y < d$. Этот прямоугольник разбивается на множество малых прямоугольников \square_{jk} горизонтальными и вертикальными прямыми $x = x_0, x_1, x_2, \dots, x_N$ и $y = y_0, y_1, y_2, \dots, y_M$. Если (x'_j, y'_k) — точка в \square_{jk} , то среднее значение $\varphi(\xi, \eta)$ приближенно представляется двойной суммой Римана—Стилтьеса

$$\sum_{j,k} \varphi(x'_j, y'_k) F(\square_{jk}), \quad (13.3.4)$$

которая сходится к двойному интегралу Стилтьеса $\int_a^b \int_c^d \varphi(x, y) \times \times d^2 F(x, y)$, когда разбиение бесконечно измельчается ($N, M \rightarrow \infty$). Если для неограниченного случая этот интеграл имеет предел при $b, d \rightarrow \infty$ и $a, c \rightarrow -\infty$, то математическое ожидание $\varphi(\xi, \eta)$ определяется как

$$\mathbf{E}(\varphi(\xi, \eta)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) d^2 F(x, y). \quad (13.3.5)$$

Если вторая производная $f(x, y) = \partial^2 F / \partial x \partial y$ существует и является кусочно непрерывной, то она называется *плотностью* данного распределения, а само распределение тогда называется *абсолютно непрерывным*; в этом случае

$$\mathbf{E}(\varphi(\xi, \eta)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) f(x, y) dx dy. \quad (13.3.6)$$

В другом крайнем случае распределение может быть дискретным относительно обеих переменных, т. е. вероятности могут концентрироваться в изолированных точках плоскости x, y . Между указанными крайними случаями существует столько возможностей, что мы не будем проводить полную их классификацию.

Если соответствующие интегралы существуют, то числа $E(\xi^k \eta^l)$ представляют собой *моменты* двумерного распределения. Первые моменты суть средние:

$$\mu_1 = E(\xi), \quad \mu_2 = E(\eta); \quad (13.3.7)$$

из вторых моментов получается *ковариационная матрица* с элементами

$$\begin{aligned} \rho_{11} &= E((\xi - \mu_1)^2) = E(\xi^2) - \mu_1^2, \\ \rho_{22} &= E((\eta - \mu_2)^2) = E(\eta^2) - \mu_2^2, \\ \rho_{12} = \rho_{21} &= E((\xi - \mu_1)(\eta - \mu_2)) = E(\xi\eta) - \mu_1\mu_2. \end{aligned} \quad (13.3.8)$$

Величина $\rho = \rho_{12} / \sqrt{\rho_{11}\rho_{22}}$ называется *коэффициентом корреляции*; согласно неравенству Шварца, ρ лежит в интервале $[-1, 1]$. Если $\rho = \pm 1$, то ξ и η полностью коррелированы, а все точки (x, y) , получаемые в эксперименте, лежат на некоторой прямой, проходящей через точку (μ_1, μ_2) . Это имеет место и в том случае, когда ρ не определено, т. е. когда $\rho_{11} = 0$ или $\rho_{22} = 0$. Если ξ и η являются *независимыми* случайными переменными (а это значит, что $F(x, y)$ имеет вид $F_1(x)F_2(y)$, то $\rho = 0$ и переменные не коррелированы (из того, что ξ велико, не следует, что η велико или мало). Однако ρ может быть равным нулю и в том случае, когда ξ и η не являются независимыми; такой пример дает распределение с плотностью

$$f(x, y) = \begin{cases} (4\pi)^{-1} & \text{при } x^2 + y^2 < 1, \\ 0 & \text{при } x^2 + y^2 > 1. \end{cases}$$

Здесь $\rho_{12} (= \rho_{21})$ обращается в нуль в силу симметрии f , но f нельзя записать в виде $f_1(x)f_2(y)$.

Используя функцию распределения $F(x, y)$, мы определим *вероятностную меру* на \mathbb{R}^2 как распределение f , именно

$$\langle f, \varphi \rangle = \iint \varphi(x, y) d^2F(x, y) \quad \forall \varphi \in C_0^\infty(\mathbb{R}^2). \quad (13.3.9)$$

Здесь f — производная распределения $\partial^2 F / \partial x \partial y$ и является обычной функцией тогда и только тогда, когда F абсолютно непрерывна. Двумерная *характеристическая функция*

$$\chi(\lambda) = \iint e^{-i\lambda \cdot x} d^2F(x, y) \quad (13.3.10)$$

равна умноженному на 2π преобразованию Фурье от f .

Теперь сделаем несколько замечаний о многомерном случае, который является непосредственным обобщением двумерного случая.

(Совместной) функцией распределения случайных переменных $\xi_1, \xi_2, \dots, \xi_n$ является функция

$$F(x_1, x_2, \dots, x_n) = \mathbf{P}\{\xi_1 \leq x_1, \dots, \xi_n \leq x_n\}. \quad (13.3.11)$$

Обозначим через \square прямоугольный параллелепипед:

$$\square = \{x: a_1 < x_1 \leq b_1, \dots, a_n < x_n \leq b_n\} \quad (13.3.12)$$

и введем обозначение

$$F(\square) = \mathbf{P}(\xi \in \square), \quad (13.3.13)$$

где ξ — векторнозначная случайная переменная, компоненты которой суть ξ_1, \dots, ξ_n . Явная формула для $F(\square)$ представляет собой обобщение (13.3.2). Мы определим *вершину* v параллелепипеда как точку x , для которой каждая компонента x_j равна или a_j , или b_j , и обозначим количество a среди компонент вершины v через $N_a(v)$. Тогда

$$F(\square) = \sum_v (-1)^{N_a(v)} F(v), \quad (13.3.14)$$

причем сумма берется по всем 2^n вершинам \square . Из вероятностной интерпретации (13.3.11) ясно, что

$F(x)$ — функция *неубывающая*: $F(\square) \geq 0$ для каждого \square , (13.3.15)

$F(x)$ *нормирована*: $F(\infty, \dots, \infty) = 1$ и $F(x_1, \dots, x_n) = 0$,

если любое $x_j = -\infty$. (13.3.16)

Если $\varphi(x)$ — непрерывная функция, то интеграл Стильеса

$$\int_{\square} \varphi(x) d^n F(x) \quad (13.3.17)$$

является пределом сумм Римана—Стильеса: \square разбивают на большое число малых параллелепипедов \square_j гиперплоскостями $x_j = x_{j,p}$ ($p=0, 1, \dots, N$), где

$$a_j = x_{j,0} < x_{j,1} < \dots < x_{j,N} = b_j,$$

а затем полагают интеграл (13.3.7) равным

$$\lim \sum_j \varphi(x_j') F(\square_j),$$

где для каждого j x_j' — точка внутри \square_j , а \lim означает предел при бесконечном измельчении разбиения \square . Если $\varphi(x)$ ограничена во всем \mathbb{R}^n , то в силу свойств (13.3.15), (13.3.16) функции F интеграл (13.3.17) имеет предел, обозначаемый через $\int_{\mathbb{R}^n}$, когда

независимо b_j стремятся к $+\infty$, а a_j к $-\infty$.

Вероятностной мерой величин ξ , называется распределение f , определенное как

$$\langle f, \varphi \rangle = \int_{\mathbb{R}^n} \varphi(\mathbf{x}) d^n F(\mathbf{x}) \quad \forall \varphi \in C_0^\infty(\mathbb{R}^n), \quad (13.3.18)$$

а характеристическая функция определяется как

$$\chi(\lambda) = \int_{\mathbb{R}^n} e^{-i\lambda \cdot \mathbf{x}} d^n F(\mathbf{x}). \quad (13.3.19)$$

УПРАЖНЕНИЯ

1. Найдите функцию распределения $F(x, y)$ случайных переменных ξ, η , значения которых равномерно распределены на единичной окружности $\xi^2 + \eta^2 = 1$.
2. Покажите, что характеристической функцией распределения из упражнения 1 является

$$\chi(\lambda) = J_0(|\lambda|).$$

13.4. НОРМАЛЬНЫЕ РАСПРЕДЕЛЕНИЯ

Наиболее выдающимся положением элементарной теории вероятностей является центральная предельная теорема (см. следующий параграф), которая утверждает, что в результате осреднения все

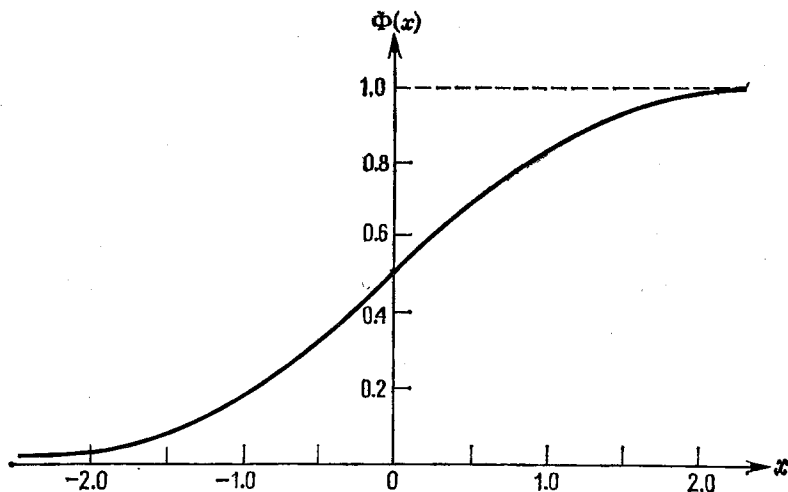


Рис. 13.9. Функция нормального распределения.

распределения вероятности стремятся к нормальным или гауссовым распределениям. Эти распределения мы сейчас и рассмотрим, начиная с одномерного случая. *Нормальным* или *гауссовым*

распределением на R называется распределение с функцией распределения

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (13.4.1)$$

и плотностью

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad (13.4.2)$$

как показано на рис. 13.9 и 13.10.

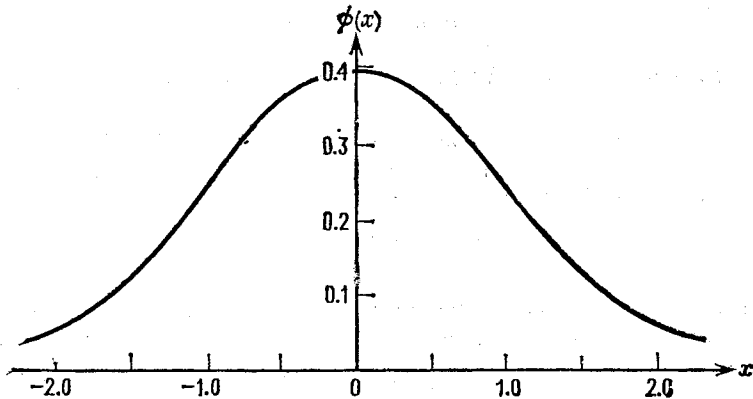


Рис. 13.10. Плотность нормального распределения.

УПРАЖНЕНИЕ

1. Покажите, что $\Phi(\infty) = 1$ и что это нормальное распределение имеет нулевое математическое ожидание и единичную дисперсию и вообще моменты всех порядков имеют вид

$$E(\xi^{2k-1}) = 0, \quad E(\xi^{2k}) = 1 \cdot 3 \cdot 5 \cdots (2k-1), \quad (13.4.3)$$

где k — любое положительное целое, а ξ — случайная переменная, подчиняющаяся нормальному распределению. Покажите, что характеристическая функция равна

$$\chi(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda x} d\Phi(x) = e^{-\lambda^2/2},$$

Подробные таблицы $\Phi(x)$ и родственных функций даны, например, в справочнике Абрамовица и Стиган [1964]. В частности, установлено, что

$$\Phi(-0.67449) = 1/4, \quad \Phi(+0.67449) = 3/4, \quad (13.4.4)$$

откуда следует для нормально распределенной величины ξ

$$P(-0.67449 \leq \xi \leq 0.67449) = 1/2. \quad (13.4.5)$$

В более общей формулировке говорят, что ξ нормально распределена со средним значением μ и дисперсией σ^2 , если

$$P\{\xi \leq x\} = F(x) = \Phi((x - \mu)/\sigma), \quad (13.4.6)$$

ибо тогда $E(\xi) = \mu$, $E((\xi - \mu)^2) = \sigma^2$.

Двумерное нормальное распределение случайных переменных ξ , η с данными средними μ_1 , μ_2 и с данной ковариационной матрицей $\rho = (\rho_{kl})$ (см. § 13.3) имеет плотность вида

$$f(x, y) = \exp[-(a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + b_1x + b_2y + c)], \quad (13.4.7)$$

где

$$a_{11}a_{22} - a_{12}^2 > 0, \quad (13.4.8)$$

а c выбрано так, что $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$. Чтобы найти соотноше-

ния между μ , ρ , a , b , заметим, что в силу (13.4.8) показатель экспоненты можно записать в виде $-1/2(u^2 + v^2)$ при помощи линейного преобразования

$$\begin{aligned} u &= c_{11}x + c_{12}y + u_0, \\ v &= c_{21}x + c_{22}y + v_0. \end{aligned} \quad (13.4.9)$$

В векторно-матричных обозначениях эти уравнения и их решения можно записать в виде

$$\mathbf{u} = C\mathbf{x} + \mathbf{u}_0, \quad \mathbf{x} = D\mathbf{u} + \mathbf{x}_0, \quad (13.4.10)$$

где $CD = I$ и $D\mathbf{u}_0 = -\mathbf{x}_0$. Пусть α и β — новые случайные переменные, полученные из ξ и η путем преобразования (13.4.9), т. е.

$$\begin{aligned} \alpha &= c_{11}\xi + c_{12}\eta + u_0, \\ \beta &= c_{21}\xi + c_{22}\eta + v_0. \end{aligned}$$

Тогда плотность вероятности $g(u, v)$ для α , β имеет вид

$$g(u, v) = f(x, y) |\partial(x, y)/\partial(u, v)|,$$

но якобиан есть константа, так что $g(u, v)$ пропорциональна $\exp\{-1/2u^2 - 1/2v^2\}$. Отсюда после нормировки имеем

$$g(u, v) = \frac{1}{2\pi} e^{-(u^2 + v^2)/2} = f(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right|.$$

Нетрудно видеть, что средние μ_1 и μ_2 для ξ и η являются компонентами вектора \mathbf{x}_0 , а ковариационная матрица имеет вид $\rho = DD^T$ (T означает транспонирование).

13.5. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

Пусть a — среднее значение большого числа n независимых изменений случайной переменной ξ ; его можно представлять себе как значение другой случайной переменной α , которая получается из

составного эксперимента, включающего n повторений исходного эксперимента с последующим вычислением среднего значения. Переменная α имеет то же математическое ожидание, что и ξ , но меньшую дисперсию (в \sqrt{n} раз). Более того, мы покажем, что если n велико, то α имеет распределение, близкое к нормальному. Пусть ξ имеет функцию распределения F ; тогда ее характеристической функцией будет

$$\chi_{\xi}(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda x} dF(x).$$

Поскольку между распределениями медленного роста и их преобразованиями Фурье существует взаимно однозначное соответствие, функция $\chi_{\xi}(\lambda)$ (λ вещественно) полностью определяет меру $f = F'$.

Пусть теперь F_1 и F_2 — функции распределения двух независимых случайных переменных ξ и η . Очевидно, что $\xi + \eta$ также можно рассматривать как случайную переменную. Будет показано, что характеристическая функция для $\xi + \eta$ есть произведение характеристических функций для ξ и для η . Пусть F_3 — функция распределения случайной переменной $\zeta = \xi + \eta$. Тогда $E(\varphi(\xi + \eta))$ задается в виде

$$E(\varphi(\xi + \eta)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x + y) dF_1(x) dF_2(y), \quad (13.5.1)$$

но, с другой стороны,

$$E(\varphi(\xi + \eta)) = E(\varphi(\zeta)) = \int_{-\infty}^{\infty} \varphi(\omega) dF_3(\omega). \quad (13.5.2)$$

Если $\varphi(x) = e^{-i\lambda x}$, то

$$\chi_{\xi+\eta}(\lambda) = \chi_{\xi}(\lambda) \chi_{\eta}(\lambda), \quad (13.5.3)$$

что и требовалось доказать.

Отметим, что абсолютное значение характеристической функции не может превышать единицу, ибо

$$|\chi(\lambda)| \leq \int_{-\infty}^{\infty} |e^{-i\lambda x}| dF(x) = F(\infty) - F(-\infty) = 1$$

(F — неубывающая функция).

Пусть далее ξ — случайная переменная, а F — ее функция распределения. Добавляя, если необходимо, подходящую константу к каждому измеренному значению ξ , можно сделать так, что среднее полученных значений будет равно нулю:

$$\mu = E(\xi) = \int_{-\infty}^{\infty} x dF(x) = 0; \quad (13.5.4)$$

затем, умножая каждое значение на другую константу, можно получить единичную дисперсию

$$\sigma^2 = E(\xi^2) = \int_{-\infty}^{\infty} x^2 dF(x) = 1. \quad (13.5.5)$$

Допустим, что мы уже осуществили указанные действия. При этом подразумевалось, что интегралы сходятся. Сделаем еще одно допущение:

$$\rho \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} |x|^3 dF(x) < \infty. \quad (13.5.6)$$

Пусть ξ_1, \dots, ξ_n — независимые случайные переменные, каждая из которых имеет функцию распределения F . Например, значения этих переменных могут получаться в результате n независимых повторений измерения ξ , так что $\eta \stackrel{\text{def}}{=} \xi_1 + \dots + \xi_n$ — случайная переменная как результат составного измерения, которое в свою очередь может повторяться неопределенно часто (всегда с тем же n), чтобы дать распределение значений η . Дисперсия η равна n , так как

$$E(\eta^2) = \sum_{j=1}^n \sum_{k=1}^n E(\xi_j \xi_k),$$

а

$$E(\xi_j \xi_k) = \begin{cases} \iint xy dF(x) dF(y) = 0, & j \neq k, \\ \int x^2 dF(x) = 1, & j = k. \end{cases}$$

Следовательно, случайная переменная

$$\zeta_n = (\xi_1 + \dots + \xi_n) / \sqrt{n} \quad (13.5.7)$$

имеет нулевое среднее и единичную дисперсию.

Любопытным и фундаментальным фактом теории вероятностей является то, что при $n \rightarrow \infty$ распределение ζ_n стремится к универсальному распределению вероятности, а именно к рассмотренному в предыдущем параграфе нормальному распределению, которое совершенно не зависит от первоначального распределения переменной ξ , пока выполняются условия (13.5.4)—(13.5.6). Этот результат составляет знаменитую центральную предельную теорему. Мы сформулируем и докажем простой вариант этой теоремы, а затем сформулируем без доказательства теорему Берри—Эссена.

Центральная предельная теорема. Пусть ξ и ζ_n имеют указанный выше смысл. Тогда распределение вероятности переменной ζ_n при $n \rightarrow \infty$ сходится (как распределение) к нормальному распре-

делению. То есть если F_n — функция распределения ξ_n , то при $n \rightarrow \infty$

$$\int_{-\infty}^{\infty} \varphi(x) dF_n(x) \rightarrow \int_{-\infty}^{\infty} \varphi(x) d\Phi(x) \quad \text{для любой пробной функции } \varphi, \quad (13.5.8)$$

причем Φ имеет вид (13.4.1).

Замечания. (1) Правый член в (13.5.8) можно записать как

$$\int \varphi(x) (2\pi)^{-1/2} \exp(-x^2/2) dx,$$

поскольку нормальное распределение имеет плотность

$$(2\pi)^{-1/2} \exp(-x^2/2).$$

(2) Согласно теореме Берри — Эссена (см. ниже) $F_n(x) \rightarrow \Phi(x)$ поточечно равномерно.

Доказательство теоремы. Если $P\{\xi \leq x\} = F(x)$, то $P\{a\xi \leq y\} = F(y/a)$. Отсюда следует, что если $\chi(\lambda)$ — характеристическая функция ξ , то $\chi(a\lambda)$ — характеристическая функция $a\xi$. Согласно (13.5.3), характеристическая функция переменной

$$\xi_n = n^{-1/2} (\xi_1 + \dots + \xi_n)$$

имеет вид

$$\chi_{\xi_n}(\lambda) = [\chi_{\xi}(\lambda/\sqrt{n})]^n. \quad (13.5.9)$$

Для заданного λ из (13.5.4) — (13.5.6) следует, что

$$\chi_{\xi}(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda x} dF(x) = \int_{-\infty}^{\infty} [1 - i\lambda x - 1/2 \lambda^2 x^2 + O(\lambda^3 x^3)] dF(x) = 1 - 1/2 \lambda^2 + O(\lambda^3).$$

Поэтому

$$\chi_{\xi}(\lambda/\sqrt{n}) = 1 - \lambda^2/(2n) + O(\lambda^3/n^{3/2}),$$

что можно записать как

$$\chi_{\xi}(\lambda/\sqrt{n}) = (1 - \lambda^2/(2n)) [1 + O(\lambda^3/n^{3/2})]; \quad (13.5.10)$$

следовательно,

$$\chi_{\xi_n}(\lambda) = (1 - \lambda^2/(2n))^n [1 + O(\lambda^3/\sqrt{n})]. \quad (13.5.11)$$

[Для n -й степени выражения в квадратных скобках из (13.5.10) использовано биномиальное разложение.] Мы видим, что

$$\chi_{\xi_n}(\lambda) \rightarrow e^{-\lambda^2/2} \quad \text{при } n \rightarrow \infty, \quad (13.5.12)$$

причем сходимость равномерна по λ в любом конечном интервале. Теперь покажем, что если φ — любая функция из класса \mathcal{S} пробных функций для распределений медленного роста, то

$$\int_{-\infty}^{\infty} \varphi(\lambda) [\chi_{\xi_n}(\lambda) - e^{-\lambda^2/2}] d\lambda \rightarrow 0 \quad \text{при } n \rightarrow \infty. \quad (13.5.13)$$

Иначе говоря, если f_n и f обозначают соответственно распределение случайной переменной ξ_n и нормальное распределение, то $\langle \hat{\varphi}, \hat{f}_n \rangle \rightarrow \langle \hat{\varphi}, \hat{f} \rangle$ для любой $\varphi \in \mathcal{S}$, где величины «с крышкой» означают преобразования Фурье. Вспомним, что $\varphi \rightarrow \hat{\varphi}$ представляет собой взаимно однозначное отображение класса \mathcal{S} на себя и что характеристическая функция нормального распределения равна $\exp(-\lambda^2/2)$. Так как для распределений медленного роста $\langle \hat{\varphi}, \hat{f} \rangle$ всегда равно $\langle \hat{\varphi}, \hat{f} \rangle$, то $\langle \hat{\varphi}, \hat{f}_n \rangle \rightarrow \langle \hat{\varphi}, \hat{f} \rangle$ для любой пробной функции $\hat{\varphi}$, а это в точности требуемый результат (13.5.8). Для того чтобы доказать (13.5.13), разобьем интервал интегрирования на две части: $|\lambda| < a$ и $|\lambda| > a$. Так как $|\chi(\lambda)| \leq 1$ для любой характеристической функции, вклад в интеграл (13.5.13) от $|\lambda| > a$ можно сделать сколь угодно малым (независимо от n), выбирая a достаточно большим, поскольку $\varphi \in \mathcal{S}$. Затем в силу равномерной сходимости (13.5.12) при $|\lambda| < a$ вклад от $|\lambda| < a$ можно сделать произвольно малым, выбирая n достаточно большим. Это завершает доказательство.

В данной теореме ничего не говорится ни о типе, ни о скорости сходимости $F_n(x)$ к $\Phi(x)$. [Теоремы о характеристиках сходимости см. в книге Феллера, т. 2.] Вообще говоря, быстрая сходимость при $n \rightarrow \infty$ требует существования моментов распределения ξ выше третьего, которым мы ограничились в нашем случае. Требуется также некоторая минимальная степень гладкости $F(x)$, которая обычно выражается через поведение характеристической функции $\chi(\lambda)$ при больших λ . Чтобы получить равномерность относительно x для этих высоких скоростей сходимости, необходимо постулировать еще более высокую степень гладкости $F(x)$. Замечательная теорема, полученная независимо Берри в 1941 г. и Эссеном в 1942 г. (см. книгу Феллера [1966]), дает равномерную сходимость со скоростью $O(n^{-1/2})$, не требуя никаких допущений, кроме (13.5.4)–(13.5.6). (Вместо единичной дисперсии в ней фигурирует явно дисперсия σ^2 .)

Теорема (Берри—Эссен). При допущениях данного параграфа

$$\left| F_n(x) - \Phi\left(\frac{x}{\sigma}\right) \right| \leq C \frac{\rho}{\sigma^3 \sqrt{n}} \quad (\forall x, \forall n), \quad (13.5.14)$$

где C — универсальная постоянная.

В первоначальной статье Эссен установил, что

$$0.410 \leq C \leq 7.59. \quad (13.5.15)$$

Здесь имелось в виду, что при $C = 7.59$ неравенство (13.5.14) справедливо для всех случаев, тогда как при $C < 0.410$ это неравенство нарушается хотя бы в одном случае. Этот результат последовательно улучшался (см. книгу Феллера [1966]), и дальнейшие исследования привели к следующим пределам:

$$0.410 \leq C \leq 0.800 \quad (13.5.16)$$

(частное сообщение проф. Эссена, 1971 г.).

Далее, ρ не может быть меньше σ^3 (см. упражнение ниже), но в некоторых ситуациях можно ожидать, что ρ и σ^3 будут

сравнимы. Затем из (13.5.16) мы видим, что F_n отличается от функции нормального распределения не более чем на 0.05 для $n \approx 256$ и не более чем на 0.01 для $n \approx 6400$.

Во многих вариантах центральной предельной теоремы (включая вариант Берри—Эссена) не требуется, чтобы переменные $\xi_1, \xi_2, \dots, \xi_n$ имели одинаковые распределения, а достаточно, чтобы их распределения отличались друг от друга не слишком сильно в том смысле, который определен в этих теоремах. И в этом случае результат тот же: распределение переменной $(\xi_1 + \dots + \xi_n)/\sqrt{n}$ стремится к нормальному распределению при $n \rightarrow \infty$.

УПРАЖНЕНИЕ

1. Используя неравенство Шварца, покажите, что $\int |x| dF(x) \leq \sigma$. Путем дальнейшего использования этого неравенства установите, что

$$\sigma^4 \leq \int |x| dF(x) \int |x|^3 dF(x),$$

и выведите отсюда, что $\sigma^3 \leq \rho$. Обобщите этот результат,

13.6. ВЫБОРКА

Иногда распределение вероятности может быть вычислено из физических закономерностей, но чаще заключение о физических законах делается из наблюдаемого распределения вероятности. Например, наблюдаемую ненулевую корреляцию между двумя случайными переменными (когда коэффициент корреляции не равен нулю) можно принять как указание существования причинной связи между этими переменными. В качестве другого примера допустим, что два различных предположения предсказывают различные значения x_1 и x_2 величины ξ , которая имеет случайные осцилляции, связанные с экспериментальными или измерительными ошибками. Если большое число наблюдаемых значений ξ показывает, что математическое ожидание $E(\xi)$ согласуется с каким-то из чисел x_1, x_2 , то это воспринимается как подтверждение соответствующего предположения.

Конечное (но большое) число n наблюдаемых значений случайной переменной ξ или набора коррелированных переменных ξ, η, \dots , получающихся при n независимых повторениях некоторого эксперимента или наблюдаемого явления, называется *выборкой* распределения переменной ξ или переменных ξ, η, \dots . Теория выборки имеет отношение к получению информации путем проб, когда неизвестно лежащее в основе явления распределение вероятности.

Пусть требуется найти математическое ожидание $E(\xi) = \mu$ некоторой единственной случайной переменной ξ , и пусть n неза-

висимых измерений ξ дали значения x_1, \dots, x_n . Тогда в качестве *выборочного среднего значения* принимают

$$a = (x_1 + \dots + x_n)/n \quad (13.6.1)$$

и рассматривают его как приближение к μ . Мы хотим получить оценку для «вероятной ошибки» этого приближения. Ясно, что a является измеренным значением случайной переменной

$$\alpha = (\xi_1 + \dots + \xi_n)/n, \quad (13.6.2)$$

где ξ_1, \dots, ξ_n независимы и имеют то же распределение, что и ξ . Можно представить себе n независимых повторений эксперимента как некоторый составной эксперимент, дающий значение α , и можно спросить, как распределено наблюдаемое значение α , если много раз повторяется этот составной эксперимент. Математическое ожидание переменной α равно $(1/n)[E(\xi_1) + \dots + E(\xi_n)] = \mu$, но нетрудно видеть, что дисперсия $E((\alpha - \mu)^2)$ меньше в n раз, чем дисперсия σ^2 переменной ξ , поскольку

$$E((\alpha - \mu)^2) = E\left(\left(\frac{1}{n} \sum \xi_i - \mu\right)^2\right) = (1/n^2) E\left(\left(\sum [\xi_i - \mu]\right)^2\right) = (1/n^2) E\left(\sum [\xi_i - \mu] \sum [\xi_j - \mu]\right) = (1/n^2) \sum E((\xi_i - \mu)^2) = \sigma^2/n. \quad (13.6.3)$$

Здесь предпоследнее равенство получено благодаря обращению в нуль корреляций $E((\xi_i - \mu)(\xi_j - \mu))$ при $i \neq j$, что следует из независимости переменных ξ_i .

Если бы σ была известна, то в качестве меры вероятной ошибки выборки можно было бы принять стандартное отклонение σ/\sqrt{n} переменной α . Однако поскольку распределение вероятности неизвестно, σ также является искомой величиной. Поэтому вводится *выборочная дисперсия*

$$V = (1/n) \sum_{j=1}^n (x_j - a)^2. \quad (13.6.4)$$

Интуитивно ясно, что V довольно близка к σ^2 . Из (13.6.1) следует, что V можно записать в виде $(1/n) \sum x_j^2 - a^2$. Значит, V есть измеренное значение случайной переменной

$$\beta = (1/n) \sum \xi_i^2 - \alpha^2, \quad (13.6.5)$$

и простые вычисления дают

$$E(\beta) = [(n-1)/n] \sigma^2. \quad (13.6.6)$$

Поэтому разумно допустить, что $\sigma^2 \approx Vn/(n-1)$, и считать $\sqrt{V/(n-1)}$ приближенно равным стандартному отклонению переменной α .

Теперь из центральной предельной теоремы следует, что если n велико, то функция распределения α приближенно описывается как

$$F(a) \approx \Phi\left(\frac{a - \mu}{\sqrt{V/(n-1)}}\right),$$

где $\mu = E(\alpha) = E(\xi)$. Следовательно, если мы определим вероятную ошибку в виде

$$\delta a = 0.6745 \sqrt{V/(n-1)}, \quad (13.6.7)$$

то формула (13.4.4) дает

$$P\{|a - \mu| < \delta a\} = 0.5. \quad (13.6.8)$$

Этот результат можно интерпретировать следующим образом: допустим, мы производим не одну серию n измерений, а много серий из n измерений. Тогда a и δa будут иметь различные численные значения после проведения различных серий, в то время как μ имеет всегда одно и то же неизвестное значение, и неравенство $|a - \mu| < \delta a$ будет верным примерно для половины этих серий и неверным для другой половины.

Переходя к более надежной оценке ошибки, можно положить

$$\delta' a = 1.6449 \sqrt{V/(n-1)}; \quad (13.6.9)$$

тогда

$$P\{|a - \mu| < \delta' a\} = 0.9. \quad (13.6.10)$$

Хотя обычно используются эти формулы для ошибки, следовало бы отметить, что подобные формулы могут быть получены из более элементарного неравенства Чебышева (см. ниже упражнение 1). Действительно, если константы в (13.6.7) и (13.6.9) увеличить примерно вдвое, то из неравенства Чебышева при $\sigma = \sqrt{V/(n-1)}$ получатся аналоги формул (13.6.8 и 13.6.10), в которых знак равенства будет заменен на \geq .

УПРАЖНЕНИЯ

1. Пусть $F(x)$ — функция распределения случайной переменной ξ со средним значением μ и дисперсией $\sigma^2 < \infty$. Рассматривая интеграл

$$\int_{|x - \mu| > t} (x - \mu)^2 dF(x),$$

получите неравенство Чебышева

$$P\{|x - \mu| > t\} \leq \sigma^2/t^2.$$

2. Проверьте равенство (13.6.6.).

В случае выборки распределения двух случайных переменных ξ , η измеряемыми значениями являются пары (x_i, y_i) , $i = 1, 2, \dots, n$. *Выборочной ковариационной матрицей* называется (2×2) -матрица, элементами которой являются

$$\begin{aligned} V_{11} &= (1/n) \sum (x_i - a)^2, & V_{22} &= (1/n) \sum (y_i - b)^2, \\ V_{12} &= V_{21} = (1/n) \sum (x_i - a)(y_i - b), \end{aligned} \quad (13.6.11)$$

где a и b — выборочные средние переменных ξ и η соответственно.

Выборочный коэффициент корреляции определяется как $W = V_{12}/\sqrt{V_{11}V_{22}}$. Из неравенства Коши следует, что $-1 \leq W \leq 1$, точно так же как из неравенства Шварца следовало $-1 \leq \rho \leq 1$.

УПРАЖНЕНИЕ

3. Найдите случайные переменные (связанные с составными экспериментами из n измерений ξ и η), для которых V_{11} , V_{12} и V_{22} представляют собой ожидаемые значения. Найдите приближенные вероятные ошибки и придумайте способ грубого установления того, является ли существенным заданный ненулевой выборочный коэффициент корреляции W .

13.7. МАРГИНАЛЬНАЯ И УСЛОВНАЯ ВЕРОЯТНОСТИ

Пусть каждое повторение эксперимента дает значения x и y случайных переменных ξ , η , и пусть совместная функция распределения ξ , η определяется так же, как в (13.3.1):

$$F(x, y) = P\{\xi \leq x, \eta \leq y\}.$$

Тогда распределение так полученных значений x при полном игнорировании значений y имеет функцию распределения

$$F(x, \infty) = \text{Маргинальная функция распределения } \xi; \quad (13.7.1)$$

аналогично

$$F(\infty, y) = \text{Маргинальная функция распределения } \eta. \quad (13.7.2)$$

Возьмем другую крайность: отбросим все эксперименты, кроме тех, в которых обнаружено, что η лежит в некотором малом интервале ($y \leq \eta < y + \Delta y$). Тогда получаемое распределение значений x переменной ξ называется *условным распределением* переменной ξ , причем условие состоит в том, что η лежит в указанном интервале.

Чтобы обсудить далее условные вероятности, сделаем упрощающее предположение о том, что $F(x, y)$ имеет непрерывную плотность

$$f(x, y) = \partial^2 F(x, y) / \partial x \partial y.$$

Позднее мы вернемся к общему случаю.

Маргинальное распределение ξ имеет плотность

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy, \quad (13.7.3)$$

которая равна производной функции $F(x, \infty)$, потому что в общем случае

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x', y') dx' dy'$$

вследствие граничных условий $F(-\infty, y) = F(x, -\infty) = 0$. Аналогично маргинальное распределение переменной η имеет плотность

$$f_2(x) = \int_{-\infty}^{\infty} f(x, y) dx. \quad (13.7.4)$$

Для того чтобы найти условные вероятности, заметим, что доля экспериментов, в которых $x \leq \xi \leq x + \Delta x$ и $y \leq \eta \leq y + \Delta y$, равна приблизительно $f(x, y) \Delta x \Delta y$, тогда как полная доля экспериментов, в которых $y \leq \eta \leq y + \Delta y$ независимо от значений ξ ,

равна $\int_{-\infty}^{\infty} f(x, y) dx \Delta y$. Если отбросить все те эксперименты, в которых $\eta \notin [y, y + \Delta y]$, то распределение ξ в остающихся экспериментах имеет плотность (в пределе $\Delta x, \Delta y \rightarrow 0$)

$$\frac{1}{\Delta x} \frac{f(x, y) \Delta x \Delta y}{\int_{-\infty}^{\infty} f(x, y) dx \Delta y} = \frac{f(x, y)}{\int_{-\infty}^{\infty} f(x, y) dx} \stackrel{\text{def}}{=} f(x|y) \quad (13.7.5)$$

при условии, что знаменатель не равен нулю. Величина $f(x|y)$ называется *плотностью условной вероятности* переменной ξ при условии, что $\eta = y$. Если знаменатель обращается в нуль, то и числитель также обращается в нуль, а значит, $f(x|y)$ не определена. Однако часто условная вероятность имеет физический смысл, и ее можно найти из модифицированного эксперимента. Например, в эксперименте по определению распределения рассеянных частиц для данной первоначальной энергии E_0 может случиться, что спектр первичных частиц не содержит частиц с энергиями, близкими к E_0 , и тогда потребуется другой источник первичных частиц.

Физически все вероятности условны, поскольку любой результат любого эксперимента зависит от любого из условий, при которых он осуществляется. Лишь обращаясь к физике явления, можно предсказать, например, будет ли влиять фаза Луны на ядерно-физический эксперимент. [Тем, кому этот пример кажется крайнейностью, нужно указать на то, что фаза Луны оказывает слабое воздействие на магнитное поле Земли через влияние атмосферных приливов на ионосферные токи.]

Определение (13.7.5) можно записать в виде

$$f(x, y) = f_2(y) f(x|y), \quad (13.7.6)$$

где f_2 — маргинальная плотность переменной η , задаваемая (13.7.4). Разложения такого вида очень важны при вычислениях по методу Монте-Карло. В трехмерном случае случайных переменных ξ, η, ζ можно записать

$$f(x, y, z) = f(z) g(y|z) h(x|y, z).$$

Здесь $g(y|z)$ — плотность значений η (при $\eta = y$) при условии, что $\zeta = z$ и игнорировании значений ξ . Метод Монте-Карло использует соответствующие функции распределения:

$$F(z) = \int_{-\infty}^z f(z') dz' = \mathbf{P} \{ \xi \leq z \}$$

(ξ и η игнорируются), (13.7.7)

$$G(y|z) = \int_{-\infty}^y g(y'|z) dy' = \mathbf{P} \{ \eta \leq y | \zeta = z \}$$

(ξ игнорируется), (13.7.8)

$$H(x|y, z) = \int_{-\infty}^x h(x'|y, z) dx' = \mathbf{P} \{ \xi \leq x | \eta = y | \zeta = z \}. \quad (13.7.9)$$

Ясно, что три функции F , G , H полностью определяют совместное распределение переменных ξ , η , ζ . Однако такое описание не является достаточно общим, так как существование функций G и H зависит от некоторой гладкости распределения относительно η и ζ (в настоящем рассмотрении мы даже допустили существование плотностей $f(x, y, z)$), хотя для большинства применений методов Монте-Карло это описание удовлетворительно. Более общие формулировки см. в книге Феллера [1966].

Как указывалось ранее, из-за возможного обращения в нуль числителя и знаменателя в (13.7.5) теряется единственность функций G и H . Если z_0 принадлежит интервалу постоянства $F(z)$, то распределение не содержит никаких троек ξ , η , ζ с $\zeta = z_0$ и поэтому $G(y|z_0)$ и $H(x|y, z_0)$ не определены, хотя это означает лишь непригодность данного представления. Аналогично если для некоторого z_1 y_0 принадлежит интервалу, в котором $G(y|z_1)$ постоянна, то это означает лишь неопределенность $H(x|y_0, z_1)$.

В следующем параграфе мы объясним использование этих функций в методе Монте-Карло.

13.8. МОДЕЛИРОВАНИЕ. МЕТОД МОНТЕ-КАРЛО

Первоначально метод Монте-Карло разрабатывался в связи с изучением нейтронной цепной реакции в системе с размножением, и этот метод легче всего описать, имея в виду такое применение. Пусть система состоит из нескольких фиксированных областей пространства, заполненных веществами с известными свойствами. Отдельный эксперимент (пользуясь терминологией предыдущих параграфов) заключается в следующем: один нейтрон вводится в данную систему в точке пространства x_0 со скоростью v_0 в момент времени t_0 и нейтронной цепочке разрешается ветвиться до некоторого момента «переписи» $T > t_0$. Тогда случайными пере-

менными ξ, η, \dots будут положения и скорости x_j, v_j ($j = 1, \dots, \nu$) нейтронов, имеющих в момент времени T . Число нейтронов ν также является одной из случайных переменных.

Описанный эксперимент следует осуществить повторно и независимо большое число n раз. Хотелось бы получить средние значения различных величин $\varphi(\xi, \eta, \dots)$, таких, как полная кинетическая энергия нейтронов в момент T , некоторые моменты их пространственных распределений и т. п.

Однако этот эксперимент не проводится в лаборатории, а моделируется на ЭВМ с использованием случайных чисел. Здесь мы сталкиваемся с ситуацией, промежуточной между двумя случаями, рассмотренными в предыдущих параграфах. В первом случае была известна функция распределения $F(x, y, \dots)$ случайных переменных ξ, η, \dots и мы хотели лишь вычислить некоторые средние. Во втором случае была неизвестна функция распределения $F(x, y, \dots)$ и нам хотелось получить сведения о ней из большой выборки измеряемых значений.

В рассматриваемом случае вероятностные законы элементарных процессов (нейтронно-ядерного взаимодействия) нам полностью известны, но для ветвящихся цепочек эти законы комбинируются столь сложно, что практически почти невозможно записать и тем более использовать некую формулу для функции распределения $F(x, y, \dots)$ результата. Однако знания элементарных процессов вполне хватает для достаточно точного моделирования ветвящихся цепочек с помощью ЭВМ, что не только дешевле и надежнее, но и гораздо удобнее с точки зрения измерений, чем изучение цепной реакции в лаборатории.

Программа, реализующая метод Монте-Карло, содержит подпрограмму, называемую генератором случайных чисел. Каждый раз эта подпрограмма активизируется (при помощи оператора CALL или ему подобного) для получения числа r в интервале $0 < r < 1$. Числа r_1, r_2, \dots , порожденные таким образом, в практическом отношении ведут себя подобно независимым значениям некоторой случайной переменной ρ , равномерно распределенной на $(0, 1)$, т. е. имеющей функцию распределения

$$F(r) = \begin{cases} 0, & r < 0, \\ r, & 0 \leq r < 1, \\ 1, & 1 \leq r. \end{cases}$$

Строго говоря, эти числа не являются ни случайными, ни независимыми, поскольку каждое из них как-то вычисляется из предшествующих, но, как правило, они удовлетворяют всем стандартным статистическим тестам случайности, равномерности и независимости с точностью, значительно превышающей требуемую

в методе Монте-Карло. В одной из первых таких подпрограмм использовалась простая формула

$$r_{k+1} \equiv 7^{13} r_k \pmod{1}, \quad (13.8.1)$$

где r — десятичные дроби, содержащие 11 цифр, причем произведение образуется с 22 знаками, но затем производится усечение по модулю 1. Эта схема производит около 10^{10} различных чисел до повторения, если начать, скажем, с $r_0 = 10^{-11}$ (очевидно, r_0 не должно быть равным нулю или иметь нуль в качестве своей наименьшей значащей цифры).

По изучению генераторов случайных чисел была проделана большая работа, вероятно, даже бóльшая, чем это было необходимо, поскольку уже простейшие генераторы типа (13.8.1) оказались вполне удовлетворительными для практических целей.

В процедуре моделирования каждая ветвящаяся цепочка (т. е. каждое повторение «эксперимента») строится последовательными шагами с использованием случайных чисел и известных вероятностных законов для элементарных процессов.

Первый шаг после введения первоначального нейтрона в точку x_0 со скоростью v_0 состоит в определении точки x_1 первого столкновения. Длиной первого свободного пробега, т. е. расстоянием $|x_1 - x_0|$, которое нейтрон пройдет до первого столкновения, является случайная переменная, обозначенная через ξ в примере 2 § 13.1, причем ее функция распределения $F(x)$ была показана на рис. 13.3. Легко видеть, что если мы, получив случайное число r при помощи генератора, приравняем это расстояние $-\lambda \ln r$, где λ — средняя длина свободного пробега, то мы получим правильное распределение вероятности для расстояния. Так как движение происходит в направлении единичного вектора $v_0/|v_0|$, мы полагаем

$$x_1 = x_0 + ((-\lambda \ln r)/|v_0|) v_0. \quad (13.8.2)$$

Момент времени первого столкновения определяется как

$$t_1 = t_0 + (-\lambda \ln r)/|v_0|. \quad (13.8.3)$$

Этот пример иллюстрирует общий принцип, состоящий в том, что если $F(x)$ — функция распределения случайной переменной ξ и F^{-1} обозначает функцию, обратную к F , так что

$$r = F(x) \Leftrightarrow x = F^{-1}(r),$$

и если, кроме того, ρ имеет равномерное распределение на $(0, 1)$, то случайная переменная

$$\xi = F^{-1}(\rho) \quad (13.8.4)$$

имеет распределение, определенное при помощи F , ибо

$$\mathbf{P}\{\xi \leq x\} = \mathbf{P}\{\rho \leq r\} = r = F(x).$$

Поэтому правило выработки одномерного распределения заключается в подстановке случайного числа r в функцию, обратную данной функции распределения.

Правила выборки для многомерного распределения можно получить аналогичным образом при помощи разложения по маргинальной, смешанной и условной вероятностям, рассмотренным в § 13.7. Например, если в трехмерном случае функции F, G, H , заданные в (13.7.7) — (13.7.9), описывают распределение переменных ξ, η, ζ , а F^{-1}, G^{-1}, H^{-1} — функции, обратные этим функциям относительно первого аргумента, т. е. если

$$F(z) = r \Leftrightarrow z = F^{-1}(r),$$

$$G(y|z) = r \Leftrightarrow y = G^{-1}(r|z),$$

$$H(x|y, z) = r \Leftrightarrow x = H^{-1}(r|y, z),$$

то выборочные значения x, y, z переменных ξ, η, ζ определяются как

$$z = F^{-1}(r_1), \quad y = G^{-1}(r_2|z), \quad x = H^{-1}(r_3|y, z),$$

где r_1, r_2, r_3 — независимые случайные числа, полученные при помощи генератора.

Второй шаг моделирования цепочки заключается в выяснении того, сколько нейтронов появилось в результате столкновения в точке x_i . Число появившихся нейтронов является случайной переменной, функция распределения которой представляет собой ступенчатую функцию, причем она предполагается известной из лабораторных измерений, а ее значение определяется путем подстановки случайного числа в функцию, обратную этой ступенчатой функции. Направления движения и значения энергии новых нейтронов затем определяются путем выборки других элементарных распределений, также известных из лабораторных измерений, и т. д. Это продолжается до тех пор, пока для всех ветвей данной цепочки не наступит момент переписи T .

После того, как таким образом будет смоделировано большое число независимых цепочек, скажем от 1000 до 10 000, искомые статистические свойства цепной реакции получаются путем усреднения, а вероятные ошибки средних можно вычислить по формулам § 13.6.

Искусство использования метода Монте-Карло на практике базируется на многих методиках, которые в течение многих лет разрабатывались в целях упрощения процедур выборки и уменьшения дисперсии; см. книгу Спанье и Гелбарда [1969]. Хотя точность этого метода по существу ограничена (ошибка составляет примерно 1%), он часто дает полезные ответы для задач статистической физики, которые полностью неразрешимы аналитическими методами из-за сложной природы физических систем.

13.9. МЕРЫ

Хотя описание через функции распределения является наиболее приемлемым для конечномерных случаев, вероятностные распределения могут быть также получены в рамках общей теории распределений или классической теории меры. Такие описания позволяют перейти к бесконечномерным случаям, к абстрактным выборочным пространствам и к современной теории стохастических процессов.

Вероятностные распределения в \mathbb{R}^n принадлежат классу распределений на \mathbb{R}^n , называемых *мерами*, которые мы сейчас обсудим большей частью без доказательств.

Классы $C_0^\infty = \mathcal{D}$ и \mathcal{S} пробных функций часто оказываются более узкими, чем это необходимо для определения частного распределения $\langle f, \cdot \rangle$, а типы сходимости $\varphi_j \xrightarrow{\mathcal{D}} \psi$ и $\varphi_j \xrightarrow{\mathcal{S}} \psi$ пробных функций часто оказываются сильнее, чем это необходимо для непрерывности линейного функционала. Например, распределение $f(x) = \delta(x - x_0)$ можно рассматривать как функционал, определяемый в виде $\langle f, \varphi \rangle = \varphi(x_0)$ для всех непрерывных φ ; в этом случае поточечная сходимость φ_j к ψ гарантирует сходимость $\langle f, \varphi_j \rangle$ к $\langle f, \psi \rangle$.

В последующих рассуждениях тип сходимости пробных функций в \mathbb{R}^n будет ослаблен так, чтобы выделить класс распределений, называемых мерами, и окажется, что любое распределение этого класса может быть представлено в виде (13.3.18), где $F(x)$ — произвольная функция локально ограниченной вариации. Тогда область определения функционала будет расширена, чтобы включить все непрерывные функции $\varphi(x)$ с ограниченным носителем.

Определение 1. Запись $\varphi_j \xrightarrow{C_0} \psi$ означает, что, во-первых, функции $\{\varphi_j(x)\}$ имеют носители в некоторой общей ограниченной области в \mathbb{R}^n и, во-вторых, $\varphi_j(x) \rightarrow \psi(x)$ равномерно в \mathbb{R}^n .

Замечание. Это определение совпадает с определением сходимости $\varphi_j \xrightarrow{\mathcal{D}} \psi$, за исключением того, что не требуется сходимость производных $\varphi_j(x)$. Далее это определение будет применено вообще к пробным функциям из пространства C_0 , которые непрерывны, но не обязательно дифференцируемы.

Определение 2. *Мерой* на \mathbb{R}^n называется распределение f на \mathbb{R}^n , такое, что функционал $\langle f, \cdot \rangle$ непрерывен по отношению к только что определенному типу сходимости, т. е. такое распределение, что из $\varphi_j \xrightarrow{C_0} \psi$ следует $\langle f, \varphi_j \rangle \rightarrow \langle f, \psi \rangle$.

В частности, вероятностная мера f , определенная с помощью (13.3.18), является мерой, поскольку если последовательность $\varphi_j(x)$ сходится равномерно к $\psi(x)$, то

$$|\langle f, \varphi_j \rangle - \langle f, \psi \rangle| \leq \sup_x |\varphi_j(x) - \psi(x)| \int_{\mathbb{R}^n} d^n F(x), \quad (13.9.1)$$

но правая часть неравенства стремится к нулю при $j \rightarrow \infty$, так как $\int_{\mathbb{R}^n} d^n F(x) = 1$. Таким образом, f есть мера.

Распределение f на \mathbb{R}^n , такое, что $f \geq 0$ на всем \mathbb{R}^n , называется *положительным*.

Теорема. *Положительное распределение f на \mathbb{R}^n есть мера (следовательно, $\delta(x)$ — мера, тогда как $\delta'(x)$, $\delta''(x)$ и т. д. не являются мерами).*

Доказательство. Допустим, что $\{\varphi_k\}$ и ψ принадлежат $\text{real}C_0^\infty$ и так, что $\varphi_k \xrightarrow{C_0} \psi$. Нужно показать, что $\langle f, \varphi_k \rangle \rightarrow \langle f, \psi \rangle$; тогда то же будет справедливо и для последовательностей в $\text{срх}C_0^\infty$. Пусть χ — неотрицательная функция в C_0^∞ , причем $\chi(x) = 1$ в области, содержащей носители всех φ_k и ψ . Зададим произвольное $\varepsilon > 0$. Для достаточно большого k

$$-\varepsilon\chi(x) < \varphi_k(x) - \psi(x) < \varepsilon\chi(x),$$

откуда следует, что обе функции $\varepsilon\chi \pm (\varphi_k - \psi)$ неотрицательны и

$$\varepsilon \langle f, \chi \rangle \pm [\langle f, \varphi_k \rangle - \langle f, \psi \rangle] \geq 0,$$

так как f — положительное распределение; поэтому $\langle f, \varphi_k \rangle \rightarrow \langle f, \psi \rangle$.

Более того, распределение f , определенное через интеграл Стильтеса (13.3.18), является мерой, даже если F не удовлетворяет требованиям (13.3.15) и (13.3.16) для функции распределения, а лишь представляет собой любую (вообще говоря, комплексную) функцию локально ограниченной вариации. Согласно приложению к этой главе, F можно записать как $F_1 - F_2 + iF_3 - iF_4$, где каждая $F_k(x)$ — неубывающая функция. Тогда любое из $\langle f, \varphi_j \rangle$, $\langle f, \psi \rangle$ можно разложить на четыре члена, для каждого из которых выполняется (13.9.1) после замены $\int_{\mathbb{R}^n}$ на \int_{\square} , где \square — параллелепипед, содержащий носители всех φ_j . Таким образом, f является мерой.

Согласно теореме Рисса о представлении, которую мы сейчас сформулируем без доказательства, любая мера может быть приведена к виду (13.3.18) путем надлежащего выбора функции $F(x)$. Поэтому в одном измерении любая мера f на \mathbb{R} является производной F' (в смысле теории распределений) функции $F(x)$ локально ограниченной вариации. В n измерениях $f = \partial^n F / \partial x_1 \dots \partial x_n$.

Теорема (Рисс). *Если f — мера на \mathbb{R}^n , т. е. если функционал $\langle f, \cdot \rangle$ непрерывен относительно типа сходимости $\xrightarrow{C_0}$, то суще-*

стает функция $\sigma(x)$ локально ограниченной вариации, такая, что

$$\langle f, \varphi \rangle = \int_{\mathbb{R}^n} \varphi(x) d^n \sigma(x) \quad \forall \varphi \in C_0(\mathbb{R}^n). \quad (13.9.2)$$

Более того, если σ нормирована условиями

$$\sigma(0) = 0, \quad \sigma(x) = \lim_{\substack{\varepsilon \rightarrow 0 \\ \varepsilon > 0}} \sigma(x + \varepsilon)$$

(или каким-либо другим способом), то σ — единственная для заданной меры f ($\varepsilon > 0$ означает, что $\varepsilon_j > 0$ для $j = 1, \dots, n$).

Первоначальная формулировка теоремы Рисса, относящаяся к 1909 г. (см. книгу Рисса и Секефальви-Надя [1953, § 50]), охватывала случай конечного интервала $[a, b]$ в одном измерении. Приведенная здесь теорема рассматривается, например, в книге Лорана Шварца [1950, § 1.1] (см. также книгу Рисса и Секефальви-Надя [1953, § 59]). Соответствующая теорема для меры на абстрактном компактном хаусдорфовом пространстве приведена в книге Данфорда и Шварца [1966, § IV.6.3].

Теперь покажем, что если f — мера, то область определения функционала $\langle f, \cdot \rangle$ может быть расширена от C_0^∞ до класса C_0 всех непрерывных функций с ограниченным носителем на \mathbb{R}^n , причем функционал остается непрерывным относительно определенного выше типа сходимости $\xrightarrow{C_0}$.

Сначала доказываем, что f локально ограничена. Затем вспоминаем, что линейный функционал в банаховом или гильбертовом пространстве непрерывен (относительно сходимости по норме этого пространства) тогда и только тогда, когда этот функционал ограничен. Здесь мы имеем несколько более слабое утверждение:

Лемма. Если линейный функционал $\langle f, \cdot \rangle$ на \mathbb{R}^n непрерывен в смысле определения 2, т. е. если f является мерой, и если Ω — ограниченная область в \mathbb{R}^n , то найдется такая постоянная $K = K(\Omega)$, что

$$|\langle f, \varphi \rangle| \leq K(\Omega) \sup_x |\varphi(x)|$$

для всех пробных функций φ с носителем в Ω .

Доказательство от противного предоставляется читателю в качестве упражнения.

Теорема о расширении. Если f — мера, то область определения функционала $\langle f, \cdot \rangle$ может быть расширена, чтобы включить все функции φ из класса C_0 : расширенный функционал, скажем $\langle f_1, \cdot \rangle$, остается непрерывным в том смысле, что если $g_n(x)$ и $g(x)$ принадлежат C_0 и $g_n(x) \xrightarrow{C_0} g(x)$, то $\langle f_1, g_n \rangle \rightarrow \langle f_1, g \rangle$. Это расширение единственно.

Доказательство (частичное). Нужно для любой заданной функции $g = g(x) \in C_0$ определить $\langle f_i, g \rangle$. Пусть $\{\varphi_n(x)\}$ — последовательность функций из C^∞ с носителями в ограниченной области Ω (которая содержит носитель g). Эта последовательность такова, что $\varphi_n(x) \rightarrow g(x)$ равномерно для всех x , и построена, например, при помощи применения операторов сглаживания к g (см. § 2.6). Тогда по лемме

$$|\langle f, \varphi_k \rangle - \langle f, \varphi_l \rangle| \leq K(\Omega) \sup_x |\varphi_k(x) - \varphi_l(x)|,$$

откуда следует, что $\{\langle f, \varphi_n \rangle\}$ — числовая последовательность Коши. Определим

$$\langle f_i, g \rangle = \lim_{n \rightarrow \infty} \langle f, \varphi_n \rangle.$$

УПРАЖНЕНИЕ

1. Завершите доказательство, установив, что:

1) если $\{\varphi_n(x)\}$ — любая другая такая последовательность, тоже сходящаяся равномерно к $g(x)$, то последовательность $\{\langle f, \varphi_n \rangle\}$ имеет тот же предел, что и $\{\langle f, \varphi_n \rangle\}$;

2) определенный нами функционал $\langle f_i, \cdot \rangle$ линеен;

3) если $g(x) \in C_0^\infty$, то $\langle f_i, g \rangle = \langle f, g \rangle$;

4) если $g_n(x)$ и $g(x)$ принадлежат C_0 и $g_n \xrightarrow{C_0} g$, то $\langle f_i, g_n \rangle \rightarrow \langle f_i, g \rangle$;

5) если $\langle f_i, \cdot \rangle$ — любой другой функционал на C_0 , обладающий теми же свойствами, что и функционал $\langle f_i, \cdot \rangle$, то $f_i = f_i$.

Индекс у f_i обычно опускается, кроме тех случаев, когда необходимо отличать функционал $\langle f, \cdot \rangle$ от его расширения.

Для данного класса распределений имеется естественное пространство пробных функций. Для класса распределений Шварца этим пространством является $C_0^\infty = \mathcal{D}$, для распределений медленного роста — пространство \mathcal{S} , для мер — пространство C_0 , для распределений в L^2 (гл. 5) — само L^2 . В каждом из этих примеров мы имеем некоторое пространство X пробных функций с типом сходимости \xrightarrow{X} , причем C_0^∞ является всюду плотным вложением в X . Это означает следующее: 1) для функций из C_0^∞ $\varphi_j \xrightarrow{\mathcal{D}} \varphi$ влечет $\varphi_j \xrightarrow{X} \varphi$; 2) для любой функции χ из X существуют функции $\psi_j \in C_0^\infty$, такие, что $\psi_j \xrightarrow{X} \chi$. Непрерывный линейный функционал на X определяет ограничение на C_0^∞ и, с другой стороны, полностью определяется своим ограничением, которое является распределением в смысле гл. 2. Мы называем X *естественным* пространством пробных функций для таких распределений.

Мера f называется *ограниченной*, если существует такая постоянная K , что $|\langle f, \varphi \rangle| \leq K \sup |\varphi(x)|$ для всех $\varphi \in C_0$. [Согласно лемме, каждая мера локально ограничена.]

Следствие теоремы о расширении. Если f — ограниченная мера, то область определения функционала $\langle f, \cdot \rangle$ может быть расширена на все C (пространство ограниченных непрерывных функций), причем норма функционала $\langle f, \cdot \rangle$ сохраняется.

ДОКАЗАТЕЛЬСТВО для одного измерения. Дано $\psi \in C$, нужно определить $\langle f, \psi \rangle$. Пусть $\chi_n(x)$ — непрерывная функция, равная 1 при $|x| \leq n$, равная нулю при $|x| \geq n+1$, линейная при $-n-1 < x < -n$ и при $n < x < n+1$. Тогда для любого n $\chi_n(x)\psi(x) \in C_0$. Распределение f и функция ψ могут быть представлены как

$$f = f_1 - f_2 + if_3 - if_4, \quad \psi = \psi_1 - \psi_2 + i\psi_3 - i\psi_4,$$

где f_i и ψ_i неотрицательны. Для любых i, j $\langle f_i, \chi_n \psi_j \rangle$ является ограниченной неубывающей функцией n , и, следовательно, существует предел $\langle f, \chi_n \psi \rangle$, который мы и примем в качестве $\langle f, \psi \rangle$. Далее,

$$|\langle f, \chi_n \psi \rangle| \leq K \sup |\chi_n(x) \psi(x)| \leq K \sup |\psi(x)|,$$

откуда

$$|\langle f, \psi \rangle| \leq K \sup |\psi(x)|,$$

что и требовалось доказать.

Если $\langle f, \varphi \rangle$ рассматривается как значение интеграла $\int_{-\infty}^{\infty} f(x)\varphi(x)dx$ в обобщенном смысле, то расширение области определения функционала $\langle f, \cdot \rangle$ сводится к расширению понятия интеграла вида

$$\int_{-\infty}^{\infty} f(x)g(x)dx.$$

Если f — произвольное распределение Шварца, то этот интеграл определяется только для $g \in C_0^\infty$; если f — распределение медленного роста, то интеграл определяется для всех $g \in \mathcal{S}$; если f — мера, то интеграл определяется для всех $g \in C_0$; если f — ограниченная мера, то интеграл определяется для всех ограниченных непрерывных функций g ; если $f \in L^2$, то интеграл определяется для всех $g \in L^2$. Аналогичные замечания можно сделать для распределений на \mathbb{R}^n .

13.10. МЕРЫ КАК ФУНКЦИИ МНОЖЕСТВ

Математическая теория вероятностей обычно трактуется как ветвь теории меры. Хотя это и не обязательно для большинства приложений, включая рассматриваемые в этой книге (где выборочное пространство конечномерно), в этом параграфе предлагается краткое введение в классическое представление с точки зрения теории меры. Это делается из-за того, что, во-первых, читатель может встретиться с таким подходом в литературе, а во-вторых, это единственно возможный подход к некоторым бесконечномерным задачам, например когда каждое событие является некоторой траекторией в пространстве-времени и совокупность всех таких траекторий представляет собой бесконечномерное пространство.

Напомним, что неубывающая функция $F(x)$ на \mathbb{R} определяет некоторую (положительную) меру на \mathbb{R} . Классически мера определяется следующим образом: каждому множеству S из некоторого класса точечных множеств на \mathbb{R} ставится в соответствие неотрицательное число $\mu(S)$, и $\mu(S)$ является *мерой* множества S . Мера интервала определяется так:

$$\begin{aligned} \mu(\Delta) &= F(b-0) - F(a+0), & \text{если } \Delta = (a, b), \\ \mu(\Delta) &= F(b+0) - F(a-0), & \text{если } \Delta = [a, b], \\ \mu(\Delta) &= F(b+0) - F(a+0), & \text{если } \Delta = (a, b], \\ \mu(\Delta) &= F(b-0) - F(a-0), & \text{если } \Delta = [a, b), \end{aligned} \quad (13.10.1)$$

где $F(x \pm 0)$ означает предел $F(x \pm \epsilon)$ при $\epsilon \downarrow 0$.

Замечание 1. На самом деле необходимо лишь определить $\mu(\Delta)$ для открытых интервалов, поскольку меры дополнений измеримых множеств определяются ниже.

Замечание 2. Мера вырожденного интервала $[a, a]$, содержащего одну точку, вполне определена и может быть положительной. Если полное изменение функции F равно единице, так что F можно рассматривать как функцию распределения некоторой случайной переменной ξ , то $\mu(\Delta)$ представляет собой вероятность того, что ξ лежит в Δ . Мы хотим расширить это понятие так, чтобы $\mu(S)$ представляло вероятность того, что ξ лежит в S , где S — более общее множество. Если Δ_1 и Δ_2 — любые непересекающиеся интервалы, то ясно, что $\mu(\Delta_1 \cup \Delta_2)$ должна быть определена как $\mu(\Delta_1) + \mu(\Delta_2)$. Так как объединение любых двух интервалов всегда можно записать в виде объединения двух непересекающихся интервалов, то мера $\mu(\Delta_1 \cup \Delta_2)$ определена для любых Δ_1 и Δ_2 , а именно не превышает $\mu(\Delta_1) + \mu(\Delta_2)$. Отсюда следует, что $\mu(S)$ определена в случае, когда S — любое конечное объединение интервалов. Если S — счетное объединение $\bigcup_{j=1}^{\infty} \Delta_j$,

то $\mu\left(\bigcup_{j=1}^n \Delta_j\right)$ является неубывающей функцией n и определяется как

$$\mu(S) = \lim_{n \rightarrow \infty} \mu\left(\bigcup_{j=1}^n \Delta_j\right). \quad (13.10.2)$$

Замечание. Если полное изменение F на \mathbb{R} бесконечно, а S — неограниченное множество, то $\mu(S)$ может быть бесконечной. Тогда предположим, что множество $S = \bigcup_{j=1}^{\infty} S_j$ содержится в интервале Δ , и обозначим через S' дополнение S относительно Δ , т. е. множество всех точек, которые принадлежат Δ , но не при-

надлежат S . Вероятностное представление показывает, что должно быть $\mu(S') = \mu(\Delta) - \mu(S)$, даже если само S' не является счетным объединением интервалов. Например, S может быть плотным в Δ (так что S' не содержит никаких невырожденных интервалов), хотя S' все еще несчетное множество.

Описанная процедура, повторенная бесконечное число раз, определяет $\mu(S)$ для всех множеств S , принадлежащих *борелеву классу* (или *полю*) \mathbf{B} множеств на \mathbb{R} . Этот класс представляет собой наименьший класс множеств, таких, что 1) каждый интервал содержится в \mathbf{B} , 2) счетное объединение множеств из \mathbf{B} принадлежит \mathbf{B} , 3) дополнение множества из \mathbf{B} также содержится в \mathbf{B} .

Мера μ является функцией множеств, определенной для всех S из \mathbf{B} , причем $\mu(S) \geq 0$ для всех S из \mathbf{B} и

$$\mu\left(\bigcup_{j=1}^{\infty} S_j\right) \leq \sum_{j=1}^{\infty} \mu(S_j),$$

$$\mu\left(\bigcup_{j=1}^{\infty} S_j\right) = \sum_{j=1}^{\infty} \mu(S_j), \text{ если } S_j \text{ попарно не пересекаются.} \quad (13.10.3)$$

Множество S имеет *меру нуль*, если для любого $\epsilon > 0$ S может быть вложено в борелево множество S' меры, меньшей или равной ϵ , или, что эквивалентно, в открытое множество S' меры, меньшей или равной ϵ , как это утверждалось в § 13.1. Если S — любое множество, отличающееся от борелева множества S_0 на множество меры нуль (т. е. если множество S может быть получено из S_0 выбрасыванием точек множества S_1 и добавлением точек множества S_2 , причем S_1 и S_2 имеют меру нуль), то $\mu(S)$ определяется как $\mu(S_0)$. Этот последний шаг известен как *пополнение меры*, а множества S , получаемые таким образом, называются *измеримыми* или *μ -измеримыми* множествами.

Заметим, что класс борелевых множеств не зависит от выбора функции F , тогда как класс множеств, имеющих меру нуль, зависит от выбора F . Если F постоянна в интервале Δ , то $\mu(\Delta) = 0$, даже если длина интервала отлична от нуля. С другой стороны, если F имеет скачок при $x = \xi$ и если через $\{\xi\}$ обозначить множество, состоящее из единственной точки ξ , то $\mu(\{\xi\}) > 0$. Если $F(x) \equiv x$ (в этом случае мера любого интервала совпадает с его длиной), то мера μ называется *лебеговой мерой* и часто обозначается через m . Известно, что существуют множества S , которые неизмеримы в смысле Лебега. Если такое S лежит в интервале, в котором функция F , определяющая меру μ , постоянна, то S является μ -измеримым и $\mu(S) = 0$. Следовательно, класс измеримых множеств зависит от μ (т. е. от F), а класс борелевых множеств не зависит.

Если F — вещественная функция локально ограниченной вариации, причем не обязательно неубывающая, то μ — вещественная (не обязательно положительная) мера на \mathbb{R} . Если F — комплексная функция локально ограниченной вариации, то μ — комплексная мера на \mathbb{R} .

Общий n -мерный случай мы опишем для $n = 2$. Пусть $F = F(x, y)$ — неубывающая функция на плоскости \mathbb{R}^2 (см. § 13.3). Сначала при помощи F определим меру μ для открытых множеств. Пусть Ω — любое открытое множество на плоскости, и пусть $\mathcal{A}(\Omega)$ — множество допустимых функций для Ω , а именно вещественных функций $\varphi(x, y)$, таких, что 1) $\varphi(x, y)$ непрерывна на всей \mathbb{R}^2 , 2) $\varphi(x, y) = 0$ для $(x, y) \notin \Omega$, 3) $\varphi(x, y) \leq 1$ на всей \mathbb{R}^2 . Тогда $\mu(\Omega)$ определяется как

$$\mu(\Omega) = \sup_{\varphi \in \mathcal{A}(\Omega)} \iint_{\mathbb{R}^2} \varphi(x, y) d^2F(x, y). \quad (13.10.4)$$

[Такой двойной интеграл Стильеса непрерывной функции φ относительно неубывающей функции F мы обсуждали в § 13.3.] Функция φ может быть сколь угодно «близкой» в некотором смысле к характеристической функции множества Ω , т. е. к функции, равной 1 на Ω и равной нулю вне Ω . Это подсказывает, что (13.10.4) следовало бы записать в виде

$$\mu(\Omega) = \iint_{\Omega} d^2F(x, y). \quad (13.10.5)$$

Легко проверить, что это верно, если Ω — такая область (прямоугольник, круг, эллипс, треугольник, кольцо и т. п.), для которой известно, как определить данный интеграл. В противном случае (13.10.4) можно рассматривать в качестве определения для (13.10.5). Если $F(x, y) = x + y$, то μ является лебеговой мерой на \mathbb{R}^2 , а $\mu(\Omega)$ представляет собой площадь области Ω и может быть записана как $\iint_{\Omega} dx dy$. Если F — функция распределения

случайных переменных ξ, η , то $\mu(\Omega)$ — вероятность того, что точка (ξ, η) лежит в Ω .

Начиная с этого места, рассуждения совпадают с проведенными для \mathbb{R} , за исключением того, что «интервал» всюду заменяется «открытым множеством». Борелевы множества в \mathbb{R}^2 суть множества, которые могут быть получены из открытых множеств операциями дополнения и счетного объединения (а следовательно, также и счетного пересечения). Функция множества $\mu(S)$, определенная, как описано выше, для борелевых множеств, обладает свойствами (13.10.3) и называется положительной мерой на \mathbb{R}^2 .

В абстрактной теории начинают с абстрактного выборочного пространства S (множества иначе неопределяемых точек x , кото-

рые называются *событиями* или *выборками* и мыслятся как возможные исходы пробы или эксперимента) и с так называемой σ -алгебры \mathbf{A} подмножеств \mathbf{S} ; это такая совокупность подмножеств, что дополнение любого S в \mathbf{A} также принадлежит \mathbf{A} и что если $\{S_n\}$ — любая счетная совокупность множеств в \mathbf{A} , то объединение и пересечение этих множеств $\{S_n\}$ принадлежат \mathbf{A} . Тогда на \mathbf{A} задается некоторая неотрицательная функция множества \mathbf{P} , для которо $\mathbf{P}(\mathbf{S})=1$ и которая является *счетно аддитивной*; это означает, что если $\{S_n\}$ — счетная совокупность непересекающихся множеств в \mathbf{A} , то $\mathbf{P}(\cup S_n) = \sum \mathbf{P}(S_n)$. (Если опустить слово «непересекающиеся», то « $=$ » следует заменить на « \leq ».) Тройка $\{\mathbf{S}, \mathbf{A}, \mathbf{P}\}$ называется *вероятностным пространством* (см. книгу Феллера [1966]).

Функция $\mathbf{P}(S)$ рассматривается как вероятность того, что исход пробы является одной из точек множества S . Если \mathbf{S} — топологическое пространство, а \mathbf{A} — наименьшая σ -алгебра, содержащая открытые множества из \mathbf{S} , то эти множества в совокупности \mathbf{A} называются *борелевыми множествами* пространства \mathbf{S} . Иногда используют иную терминологию и называют σ -алгебру «борелевой алгеброй», или « σ -полем», или «борелевым полем».

Если $\varphi(x)$ — вещественнозначная функция, определенная на \mathbf{S} , то ее математическое ожидание хотелось бы определить как интеграл Стилтеса

$$\mathbf{E}(\varphi(x)) = \int_{\mathbf{S}} \varphi(x) d\mathbf{P}(x). \quad (13.10.6)$$

Для того чтобы определить такой интеграл в абстрактном вероятностном пространстве, мы ограничимся следующим классом функций φ :

Определение. Функция $\varphi(x)$ называется *случайной переменной* на вероятностном пространстве $\{\mathbf{S}, \mathbf{A}, \mathbf{P}\}$, если для любого вещественного t множество всех x , для которых $\varphi(x) \leq t$, принадлежит σ -алгебре \mathbf{A} .

Замечание. x обозначает неопределяемую точку в абстрактном пространстве \mathbf{S} , но если в \mathbf{S} введена какая-либо *разумная* система координат (конечномерная или бесконечномерная), то каждая координата точки x является случайной переменной в полном соответствии с предыдущими параграфами.

Поскольку \mathbf{P} — положительная мера, в данном контексте нет необходимости использовать полную теорию интеграла Стилтеса в абстрактных пространствах с мерой для того, чтобы интерпретировать (13.10.6). Вместо этого мы поступим следующим образом (см. книгу Феллера [1966, § IV.4]): прежде всего $\varphi(x)$ — *простая* функция, если она принимает значения $\varphi^{(1)}$,

$\varphi^{(2)}, \dots$ на множествах S_1, S_2, \dots в данной алгебре. Тогда мы определяем

$$E(\varphi(x)) = \sum_{j=1}^{\infty} \varphi^{(j)} P(S_j) \quad (13.10.7)$$

при условии, что сумма сходится абсолютно; в противном случае математическое ожидание не определено. Любая случайная переменная $\varphi(x)$ является равномерным пределом последовательности $\{\varphi_k(x)\}$ простых функций, и мы полагаем

$$E(\varphi(x)) = \lim_{k \rightarrow \infty} E(\varphi_k(x)). \quad (13.10.8)$$

Это следует понимать так: может быть доказано, что или все величины $E(\varphi_k(x))$ существуют для достаточно больших k и их предел существует, или все они не определены. Кроме того, если этот предел существует, то он одинаков для всех последовательностей $\{\varphi_k\}$, сходящихся к φ .

На этом мы заканчиваем описание абстрактной концептуальной основы математической теории вероятностей; дальнейшее ее развитие см. в томе 2 книги Феллера [1966].

Теперь вернемся к конечномерному случаю и кратко рассмотрим вопрос о том, когда две различные меры определяют одинаковые измеримые множества. Такое равенство приводит к понятию абсолютно непрерывных функций и мер и к теореме Радона—Никодима. Мы обсудим в основном одномерный случай. Как было указано выше, борелевы множества на \mathbb{R} (или на \mathbb{R}^n) не зависят от выбора функции $F(x)$ (или $F(\mathbf{x})$), использованной в определении меры μ , но множества меры нуль, которые используются в пополнении меры, зависят от F . Поэтому интересующий нас вопрос можно поставить так: когда множества μ -меры нуль совпадают с множествами лебеговой меры нуль и когда при двух заданных мерах μ_1 и μ_2 , полученных при помощи функций F_1 и F_2 , множества μ_1 -меры нуль совпадают с множествами μ_2 -меры нуль?

Теорема 1. *Каждое множество на \mathbb{R} лебеговой меры нуль имеет μ -меру нуль тогда и только тогда, когда функцию $F(x)$ можно записать в виде интеграла Лебега*

$$F(x) = \int f(x') dx'. \quad (13.10.9)$$

Тогда производная $F'(x)$ существует почти всюду и совпадает почти всюду с $f(x)$.

Феллер [1966] определяет $F(x)$ как абсолютно непрерывную функцию, если она может быть выражена в указанной форме, а данную теорему в несколько более общей формулировке он

называет теоремой Радона—Никодима. Следуя этой точке зрения, мы назвали распределение вероятности \mathbf{P} абсолютно непрерывным (§ 13.1), если ее функция распределения может быть выражена в виде (13.10.9). Для такого распределения данная теорема показывает, что если S —множество лебеговой меры нуль, то вероятность $\mathbf{P}\{x \in S\} = 0$. Утверждение приведенной выше теоремы применимо, однако, и к любой вещественной или комплексной функции $F(x)$ локально ограниченной вариации. Другое и чаще используемое определение абсолютной непрерывности формулируется следующим образом.

Определение 1. Функция $F(x)$ называется *абсолютно непрерывной* на \mathbb{R} , если для любого $\varepsilon > 0$ существует такое $\delta > 0$, что каждый раз, когда объединение конечного числа непересекающихся интервалов (a_k, b_k) имеет лебегову меру, меньшую δ , т. е. когда

$$\sum (b_k - a_k) < \delta, \quad (13.10.10)$$

выполняется неравенство

$$\sum |F(b_k) - F(a_k)| < \varepsilon. \quad (13.10.11)$$

Абсолютная непрерывность в интервале (α, β) на \mathbb{R} определяется аналогично.

Эти два определения абсолютной непрерывности эквивалентны для конечного интервала, но не для всего \mathbb{R} . Например, функция $F(x) = x^2$ может быть записана в виде (13.10.9), но не является абсолютно непрерывной в смысле определения 1, поскольку для любого $\delta > 0$ мы можем найти такой интервал $(a, a + \delta)$, что величина $(a + \delta)^2 - a^2$ произвольно велика, просто положив a достаточно большим. Однако если дополнительно допустить, что полная вариация F на \mathbb{R} конечна, то из (13.10.9) следует абсолютная непрерывность. Обратное, если $F(x)$ абсолютно непрерывна на \mathbb{R} согласно приведенному определению, то, во-первых, легко доказать, что $F(x)$ имеет ограниченную вариацию на \mathbb{R} , а во-вторых, из теоремы Радона—Никодима (которая будет сформулирована ниже) следует, что $F(x)$ можно представить в виде (13.10.9). Наконец, теорема Банаха—Зарецкого (см. книгу Натансона [1950]) утверждает, что если $F(x)$ имеет ограниченную вариацию на \mathbb{R} и $\mu(S) = 0$ для любого множества S лебеговой меры нуль, то $F(x)$ абсолютно непрерывна в смысле определения 1. (Если $F(x)$ не имеет ограниченной вариации, то мера μ , разумеется, не определена.)

В случае когда $F(x)$ является функцией Кантора, описанной в примере 6 § 13.1 и изображенной на рис. 13.7, было показано, что для любого $\delta > 0$ можно найти конечную совокупность открытых интервалов на \mathbb{R} с общей длиной, меньшей δ , но та-

кую, что $F(x)$ возрастает только в этих интервалах. В этом случае сумма в (13.10.11) равна 1 и поэтому $F(x)$ не является абсолютно непрерывной. Было также указано, что $F'(x) = 0$ почти всюду. Таким образом, правая часть равенства (13.10.9), рассматриваемая как интеграл Лебега (а это единственно возможная интерпретация), является константой и, следовательно, не равна функции $F(x)$.

В теореме Радона—Никодима лебегова мера заменяется произвольной положительной мерой μ_1 на \mathbb{R} .

Определение 2. Мера μ_2 на \mathbb{R} называется *абсолютно непрерывной относительно* положительной меры μ_1 , если для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что каждый раз, когда объединение конечного числа непересекающихся интервалов Δ_k имеет меру μ_1 , меньшую δ , т. е. когда

$$\sum \mu_1(\Delta_k) = \sum F_1(\Delta_k) < \delta,$$

выполняется неравенство

$$\sum |\mu_2(\Delta_k)| = \sum |F_2(\Delta_k)| < \varepsilon.$$

(Заметим, что $\sum \mu_1(\Delta_k)$ можно было бы записать как $\sum |\mu_1(\Delta_k)|$ в силу положительности меры μ_1 .)

Теорема (Радон—Никодим). Если μ_2 абсолютно непрерывна относительно μ_1 , то F_2 можно записать в виде

$$F_2(x) = \int^x f(x') dF_1(x').$$

В § 14.8 мы встретимся с применением этой теоремы к одному из представлений некоторой физической системы, в которой данная наблюдаемая диагональна.

В многомерном случае имеет место все то же самое, за исключением того, что Δ_k представляет собой прямоугольный параллелепипед (также называемый *интервалом* в \mathbb{R}^n), $F(\Delta_k)$ определяется так же, как в § 13.3, а равенство в теореме Радона—Никодима имеет вид

$$F_2(x) = \int^{x_1} \dots \int^{x_n} f(x') d^n F_1(x').$$

Вариант этой теоремы для абстрактных пространств см. в книге Халмоша [1950] или в книге Данфорда и Шварца [1966].

Теперь мы можем ответить на вопрос, с которого мы начали данное рассмотрение: достаточное условие того, что две меры μ_1 и μ_2 определяют одинаковые множества меры нуль, состоит в их положительности и абсолютной непрерывности одной относительно другой.

УПРАЖНЕНИЯ

1. Пусть $f(x)$ и $F(y)$ — абсолютно непрерывные функции. Покажите, что $F(f(x))$ абсолютно непрерывна, если $f(x)$ монотонна или если $F(y)$ непрерывна по Липшицу.

2. Покажите, что если

$$f(x) = \begin{cases} (x \sin(1/x))^3 & \text{при } x \neq 0, \\ 0 & \text{при } x = 0, \end{cases}$$

$$F(y) = y^{1/3} \quad (\text{вещественный корень}),$$

то $f(x)$ и $F(y)$ абсолютно непрерывны в $(-1, 1)$, тогда как $F(f(x))$ таковой не является. Почему это не вступает в противоречие с упражнением 1? Заметьте, что интегралы

$$\int_{x_0}^1 \frac{d}{dx} F(f(x)) dx, \quad \int_{-1}^{x_0} \frac{d}{dx} F(f(x)) dx$$

расходятся при $x_0 = 0$ как интегралы Римана, причем не спасает положение и рассмотрение их в смысле Лебега.

13.11. ВЕРОЯТНОСТЬ В ГИЛЬБЕРТОВОМ ПРОСТРАНСТВЕ.

ЦИЛИНДРИЧЕСКИЕ МНОЖЕСТВА. ГАУССОВЫ МЕРЫ

Мы представим здесь некоторые простейшие соображения относительно мер в топологических векторных пространствах для того, чтобы проиллюстрировать идеи, высказанные в предшествующем параграфе, и чтобы показать нечто из того, что может происходить в бесконечномерном случае. Дальнейшие подробности см. в книге Гельфанда и Виленкина [1961].

В конечномерном пространстве \mathbb{R}^n лебегова мера m обладает тем свойством, что она одинакова для конгруэнтных множеств: если S измеримо, а S' получено из S сдвигом и вращением в \mathbb{R}^n , то $m(S') = m(S)$. Если Ω — открытое множество, то $m(\Omega)$ представляет его объем; поэтому если Ω также ограничено и непусто, то

$$0 < m(\Omega) < \infty. \quad (13.11.1)$$

В бесконечномерном случае нельзя найти меры, обладающей этими свойствами. Пусть $\{\varphi_k\}_1^\infty$ — полная ортонормированная система в сепарабельном вещественном гильбертовом пространстве H . Для любого $x \in H$ мы записываем

$$x = \sum_{k=1}^{\infty} x_k \varphi_k$$

и называем x_k координатами x . Рассмотрим ограниченные открытые множества

$$\begin{aligned} \Omega_1 &= \{x: 0 < x_k < 1/k, k=1, 2, \dots\}, \\ \Omega_{1/2} &= \{x: 0 < x_k < 1/(2k), k=1, 2, \dots\}. \end{aligned} \quad (13.11.2)$$

При помощи сдвигов множества $\Omega_{1/2}$ в различных направлениях мы можем породить бесконечное число непересекающихся копий этого множества в Ω_1 . Поэтому если $m(\Omega_{1/2}) > 0$, то $m(\Omega_1) = \infty$, а, с другой стороны, если $m(\Omega_1) < \infty$, то $m(\Omega_{1/2}) = 0$, и, таким образом, условие (13.11.1) нарушается. Отсюда следует, что в H нет понятия объема и понятия плотности вероятности. Однако существуют распределения вероятности, включая непрерывные, основанные на так называемых цилиндрических множествах, которые, согласно Гельфанду и Виленкину, были введены Колмогоровым в 1936 г.

Если M —любое конечномерное подпространство H , а S —любое борелево множество в M , то множество

$$Z = S + M^\perp, \quad (13.11.3)$$

т. е. множество всех точек $x + y$ из H , где $x \in S$, а $y \in M^\perp$, называется *цилиндрическим множеством*, S —его *основанием*, а M^\perp —*образующим подпространством*. (Если бы M было двумерным, а M^\perp —одномерным, то Z представляло бы собой обычный трехмерный цилиндр, но в приведенном выше определении M^\perp , конечно, обязательно является бесконечномерным подпространством.)

Говоря, что S —борелево множество в M , мы имеем в виду обычную топологию M , которое изоморфно \mathbb{R}^m для некоторого m . Борелевы множества порождаются операциями дополнения (относительно M) и счетного объединения, исходя из открытых множеств пространства M .

Основание S и образующее подпространство M^\perp цилиндрического множества Z неоднозначно определяются этим множеством. Например, мы всегда можем заменить подпространство M на большее подпространство M' (т. е. подпространство большей размерности), которое содержит M , а затем заменить множество S на множество S' в M' , задав это множество в виде $S + (M' \ominus M)$, где $M' \ominus M$ означает ортогональное дополнение M в M' . Ясно, что $S' + M'^\perp$ совпадает с $S + M^\perp$. Из этого следует, что любые два цилиндрических множества Z_1 и Z_2 (или любое конечное число таких множеств) могут быть описаны как имеющие основания в общем конечномерном подпространстве M и имеющие общее образующее подпространство M^\perp . А именно, если M_1 и M_2 являются подпространствами для Z_1 и Z_2 , то мы принимаем за подпространство M линейную оболочку M_1 и M_2 .

Таким образом мы видим, что цилиндрические множества образуют *алгебру* A_0 множеств, т. е. совокупность множеств, обладающих следующими свойствами: 1) объединение и пересечение двух любых цилиндрических множеств являются цилиндрическими множествами; 2) дополнение (в H) любого цилиндрического множества является цилиндрическим множеством.

Следующее свойство этой алгебры состоит в том, что если имеется счетная совокупность $\{Z_i\}_{i=1}^{\infty}$ цилиндрических множеств, каждое из которых имеет основание в общем конечномерном подпространстве, то их объединение $\bigcup_{i=1}^{\infty} Z_i$ и их пересечение $\bigcap_{i=1}^{\infty} Z_i$ представляют собой цилиндрические множества.

Алгебра A_0 не является σ -алгеброй, поскольку в общем случае счетное объединение $\bigcup_{i=1}^{\infty} Z_i$ не представляет собой цилиндрическое множество, если только все Z_i не имеют оснований в общем конечномерном подпространстве, но мы определяем A как наименьшую σ -алгебру, содержащую A_0 . Алгебра A получается при помощи операций дополнения и счетного объединения, исходя из цилиндрических множеств. Множества Ω_1 и $\Omega_{1/2}$, заданные равенствами (13.11.2), содержатся в A .

Допустим теперь, что P —счетно аддитивная функция множества, определенная на σ -алгебре A и удовлетворяющая аксиомам

$$0 \leq P(X) \leq 1 = P(H), \quad (13.11.4)$$

как в предыдущем параграфе. Тогда тройка $\{H, A, P\}$ представляет собой вероятностное пространство.

Если Z —цилиндрическое множество, то $P(Z)$ можно интерпретировать как маргинальную вероятность. Именно, пусть M —подпространство, которое содержит основание S множества Z , и пусть $\{\varphi_k\}_1^{\infty}$ —ортонормированный базис в H , причем такой, что $\{\varphi_1, \dots, \varphi_m\}$ является базисом в M , а $\{\varphi_{m+1}, \dots\}$ —базисом в M^{\perp} . Тогда координаты $\{x_k\}_1^{\infty}$ точки x в H относительно базиса $\{\varphi_k\}_1^{\infty}$ можно рассматривать как случайные переменные, которые описывают результат некоторого эксперимента, а $P(Z)$ есть вероятность того, что точка (x_1, \dots, x_m) лежит в S , тогда как значения x_{m+1}, \dots полностью игнорируются. В этом смысле $P(Z)$ —маргинальная вероятность. Определение $P(Z)$ для всех Z из A_0 сводится к определению всех возможных конечномерных маргинальных вероятностей. Предполагается, что это определение согласуется с вероятностной интерпретацией: именно, $P(Z)$ удовлетворяет (13.11.4), является конечно аддитивной, а также счетно аддитивной в том смысле, что для непересекающихся Z_i в A_0

$$P\left(\bigcup_1^{\infty} Z_i\right) = \sum_1^{\infty} P(Z_i), \quad (13.11.5)$$

когда $\bigcup_1^{\infty} Z_i$ содержится в A_0 . Мы называем P вероятностной мерой на A_0 . Следующая теорема является фундаментальной в теории вероятностей (см. книгу Феллера [1966, § IV.5]).

Теорема о расширении. Если A_0 — алгебра множеств, а P — вероятностная мера на A_0 , то P имеет единственное расширение до вероятностной меры на σ -алгебре A , порожденной алгеброй A_0 .

Если P ограничить подалгеброй алгебры A_0 , состоящей из всех Z с основанием S в некотором фиксированном конечномерном подпространстве M , то мы можем определить

$$P_M(S) = P(Z);$$

тогда ясно, что P_M есть вероятностная мера в M . Но M конечномерно, и поэтому P_M можно описать при помощи методов предыдущих параграфов, например, используя функцию распределения $F(x_1, \dots, x_m)$ или — в случае ее абсолютной непрерывности — плотность $f(x_1, \dots, x_m)$. Чтобы показать, что данная функция множества P является вероятностной мерой в H (после того, как уже показано, что P_M представляет собой вероятностную меру в каждом M), остается лишь показать, что счетная аддитивность (13.11.5) имеет место, когда не все Z_i имеют основания в общем M , хотя их объединение есть цилиндрическое множество. Такой пример будет приведен ниже в упражнении 4.

Гельфанд и Виленкин [1961, § IV.2] приводят различные условия, при которых P является счетно аддитивной, когда P_M — вероятностная мера в каждом M .

В том случае, когда P определена на A_0 , но предполагается лишь ее конечная аддитивность (и значит, она, строго говоря, не может быть вероятностью), эту функцию называют *мерой цилиндрических множеств*.

Основными примерами служат так называемые гауссовы меры в H , которые соответствуют нормальным распределениям в конечномерном пространстве. В § 13.4 было определено двумерное нормальное распределение, имеющее предписанные средние μ_1 и μ_2 случайных переменных ξ и η и предписанную ковариационную матрицу ρ . Было показано, что, переходя при помощи линейного преобразования от переменных ξ, η к переменным α, β , можно получить нормальное распределение переменных α, β с нулевыми средними и единичной ковариационной матрицей ρ . Плотность вероятности α, β тогда равна $\exp\{-(\alpha^2 + \beta^2)/2\}$. Сначала опишем обобщение этого случая на пространство H . Оказывается, что в этом случае нет счетной аддитивности, хотя некоторые другие гауссовы меры обладают таким свойством.

Пусть Z — цилиндрическое множество, основание которого S лежит в m -мерном подпространстве $M \in H$, и пусть x_1, \dots, x_m — декартовы координаты в M . Меру цилиндрических множеств определяем, положив

$$P(Z) = (2\pi)^{-m/2} \int_S \exp[-(x_1^2 + \dots + x_m^2)/2] dx_1 \dots dx_m. \quad (13.11.6)$$

Отсюда следует, что \mathbf{P}_M имеет плотность $(2\pi)^{-m/2} \exp[-(x_1^2 + \dots + x_m^2)/2]$ в M . (Напомним, что H — вещественное гильбертово пространство и поэтому координаты x_j вещественны.) Можно показать, что $\mathbf{P}(Z)$ не зависит от выбора M и S для данного Z (см. упражнение 2) и что \mathbf{P} не является счетно аддитивной (см. упражнение 3).

УПРАЖНЕНИЯ

1. Демонстрация того, что маргинальные вероятности могут давать большую информацию.) Пусть $F(x, y)$ — функция распределения случайных переменных ξ, η . Рассмотрим преобразования

$$\xi' = \xi \cos \theta + \eta \sin \theta,$$

$$\eta' = -\xi \sin \theta + \eta \cos \theta \quad (\theta \text{ вещественно}).$$

Допустим, что для каждого такого преобразования маргинальное распределение ξ' при игнорировании η' известно. Покажите, что тогда $F(x, y)$ определена.

2. Покажите, что при замене M и S на M' и S' для данного Z , как это было описано ранее, значение (13.11.6) для $\mathbf{P}(Z)$ не изменится.

3. Пусть x_j ($j=1, \dots$) — координаты относительно полной ортонормированной системы $\{\varphi_j\}_1^\infty$ в вещественном гильбертовом пространстве H . Рассмотрим цилиндрические множества

$$Z_1: |x_1| < a \quad (x_2, x_3, \dots \text{ произвольны}),$$

⋮

$$Z_k: \sum_{j=1}^k x_j^2 < a^2 \quad (x_{k+1}, \dots \text{ произвольны}),$$

⋮

где a — положительная постоянная. Покажите, что (13.11.6) дает

$$\mathbf{P}(Z_k) = \frac{\Gamma(k/2, a^2/2)}{\Gamma(k/2)},$$

где $\Gamma(x, y)$ — неполная гамма-функция, имеющая вид

$$\Gamma(x, y) = \int_0^y e^{-t} t^{x-1} dt$$

($\Gamma(x) = \Gamma(x, \infty)$). Используя очевидное неравенство $\Gamma(x, y) < y^{x-1}$ для положительных x и y и асимптотическую формулу Стирлинга для $\Gamma(x)$, покажите,

что для фиксированного $a > 0$ $\mathbf{P}(Z_k) \rightarrow 0$ при $k \rightarrow \infty$. Пересечение $\bigcap_{k=1}^\infty Z_k$

представляет собой шар B_a радиуса a в H и не является цилиндрическим множеством, но если бы функция \mathbf{P} могла быть расширена на σ -алгебру \mathcal{A} (которая содержит B_a), мы бы заключили, что $\mathbf{P}(B_a) = 0$ для любого a . Наконец, поскольку H является счетным объединением $\bigcup_{a=1}^\infty B_a$, мы видим, что гауссова мера \mathbf{P} не является счетно аддитивной.

4. Пусть x_j ($j=1, \dots$) снова координаты. Рассмотрим цилиндрические множества (во всех случаях неуказанные координаты произвольны):

$$Z_1: x_1 < 1,$$

$$Z_2: x_1 \geq 1, \quad x_2 < 1,$$

⋮

$$Z_k: x_j \geq 1 \quad (j=1, \dots, k-1), \quad x_k < 1.$$

⋮

⋮

Покажите, что Z_k не пересекаются, что их основания не лежат в каком-либо общем конечномерном подпространстве и что их объединение является цилиндрическим множеством.

Грубо говоря, причина того, что описанная выше гауссова мера не ведет себя как вероятность, заключается в следующем. Гауссова мера имеет единичную дисперсию в каждом из бесконечно многих направлений в \mathbf{H} , а это вызывает стремление «вытолкнуть» вероятность на большие расстояния до такой степени, что фактически вероятность нахождения x в любом конечном шаре равна нулю (см. упражнение 3). Разумеется, такой путь рассуждений носит чисто эвристический характер, поскольку в \mathbf{H} нет таких понятий, как плотность вероятности, а потому бессмысленно говорить о том, «где» локализована вероятность, но этот путь наводит на мысль, что, возможно, мы получим более разумную гауссову меру, если выберем фиксированный ортонормированный базис $\{\psi_l\}_l^\infty$ в \mathbf{H} и потребуем, чтобы дисперсия в направлении ψ_l стремилась к нулю с достаточной скоростью, когда $l \rightarrow \infty$. Конечно, полученная мера будет в высокой степени анизотропна.

С этой целью допустим, что B — положительно определенный компактный оператор, и обозначим $A = B^{-1}$. (Введенные выше ортонормированные векторы ψ_l будут собственными векторами B , а дисперсия в направлении ψ_l будет соответствующим собственным значением.) Если Z — любое цилиндрическое множество с основанием S в m -мерном подпространстве \mathbf{M} , то примем в качестве ортонормированного базиса в \mathbf{M} совокупность $\{\varphi_1, \dots, \varphi_m\}$ и определим эрмитову матрицу $A(\mathbf{M})$ с элементами

$$A(\mathbf{M})_{jk} = (\varphi_j, A\varphi_k) \quad (j, k = 1, \dots, m),$$

а затем положим

$$\mathbf{P}(Z) = \mathbf{P}_{\mathbf{M}}(S) = \frac{\sqrt{\det A(\mathbf{M})}}{(2\pi)^{m/2}} \int_S \exp \left\{ -\frac{1}{2} \sum_{j,k} x_j A(\mathbf{M})_{jk} x_k \right\} dx_1 \dots dx_m.$$

Мы утверждаем без доказательства, что это определяет меру цилиндрических множеств и что эта мера счетно аддитивна,

если B — ядерный оператор; тогда $\{H, A, P\}$ является вероятностным пространством.

Подробности см. в книге Гельфанда и Виленкина [1961].

ПРИЛОЖЕНИЕ К ГЛАВЕ 13.

ФУНКЦИИ ОГРАНИЧЕННОЙ ВАРИАЦИИ

Полной вариацией вещественной функции $f(x)$ одной переменной в интервале $[a, b]$ является, грубо говоря, сумма всех вертикальных смещений, необходимых при вычерчивании графика $f(x)$ от $x=a$ до $x=b$, если все смещения — и вверх, и вниз — берутся со знаком плюс. Если $f(x)$ имеет непрерывную производную, то полная вариация равна $\int_a^b |f'(x)| dx$. Однако чтобы придать этому понятию более общий смысл, не обязательно требовать, чтобы функция $f(x)$ была дифференцируемой или даже непрерывной. Полная вариация ступенчатой функции равна сумме модулей величин ее скачков.

Разбиение P_N отрезка $[a, b]$ представляет собой совокупность точек x_j , таких, что

$$a = x_0 < x_1 < \dots < x_N = b.$$

Тогда полная вариация f на $[a, b]$ определяется как

$$V_a^b = V_a^b(f) = \sup \left\{ \sum_{j=1}^N |f(x_j) - f(x_{j-1})| : \forall P_N, \forall N \right\}. \quad (13.A.1)$$

Ясно, что это определение согласуется с частными случаями, описанными выше. Функция $f(x)$ имеет ограниченную вариацию на $[a, b]$, если $V_a^b < \infty$. Функция имеет ограниченную вариацию на R , если V_a^b конечна и остается ограниченной, когда $a, b \rightarrow -\infty, +\infty$. Функция имеет локально ограниченную вариацию, если V_a^b конечна для любого конечного интервала $[a, b]$.

ПРИМЕР 1

Функция $f(x)$, принимающая значение 1 для рациональных x и значение 0 для иррациональных x , имеет неограниченную вариацию на любом интервале.

ПРИМЕР 2

Непрерывная функция

$$f(x) = \begin{cases} 0 & \text{при } x = 0, \\ x \sin(1/x) & \text{при } x \neq 0 \end{cases} \quad (13.A.2)$$

имеет неограниченную вариацию на любом интервале, включающем начало координат.

ПРИМЕР 3

Монотонная функция (неубывающая или невозрастающая) имеет ограниченную вариацию на любом интервале; в самом деле, $V_a^b = |f(b) - f(a)|$.

Частичное обращение утверждения примера 3 состоит в том, что любая функция $f(x)$ локально ограниченной вариации может быть представлена в виде разности двух неубывающих (или двух невозрастающих) функций:

$$f(x) = f_1(x) - f_2(x). \quad (13.A.3)$$

Чтобы это показать, разобьем V_a^b на две части — *положительную* и *отрицательную вариации* $\overset{+}{V}_a^b$ и \bar{V}_a^b , для чего введем обозначения

$$[y]^+ = \max \{0, y\}, \quad [y]^- = \max \{0, -y\} \quad (13.A.4)$$

для любых вещественных y и определим

$$\overset{+}{V}_a^b = \sup \left\{ \sum [f(x_j) - f(x_{j-1})]^+ : \forall P_N, \forall N \right\}, \quad (13.A.5)$$

$$\bar{V}_a^b = \sup \left\{ \sum [f(x_j) - f(x_{j-1})]^- : \forall P_N, \forall N \right\}.$$

Очевидно, что

$$\overset{+}{V}_a^b + \bar{V}_a^b = V_a^b, \quad \overset{+}{V}_a^b - \bar{V}_a^b = f(b) - f(a).$$

Наконец, в (13.A.3) f_1 и f_2 можно взять в виде

$$f_1(x) = \overset{+}{V}_a^x + \text{const}, \quad f_2(x) = \bar{V}_a^x + \text{const}, \quad (13.A.6)$$

где две константы выбраны так, что их разность равна $f(a)$.

Представление (13.A.3) весьма неоднозначно, поскольку f_1 и f_2 можно заменить на $f_1 + g$ и $f_2 + g$, где g — любая неубывающая функция. Функции, приведенные в (13.A.6) и называемые *восходящей* и *нисходящей частями* $f(x)$, имеют особое свойство, состоящее в том, что когда одна из них возрастает, другая остается постоянной. Таким образом, возрастание $f_1(x) + f_2(x)$ на интервале $[a, b]$ равно V_a^b , а не просто $\geq V_a^b$.

Если $f(x)$ — комплекснозначная функция, то определение (13.A.1) остается в силе, но разделение на восходящую и нисходящую части нужно проводить отдельно для $\text{Re } f(x)$ и $\text{Im } f(x)$; в таком случае мы имеем

$$f(x) = f_1(x) - f_2(x) + if_3(x) - if_4(x),$$

где каждая из f_k — неубывающая функция.

Рассмотрим теперь вещественную функцию $f(x, y)$ двух переменных. *Полная вариация* f в прямоугольнике $a \leq x \leq c, b \leq y \leq d$ есть

$$V_{ab}^{cd} = \sup \left\{ \sum_{j, k} |f(\square_{jk})| : \forall P_N, \forall N \right\}, \quad (13.A.7)$$

где P_N — разбиение данного прямоугольника на малые прямоугольники \square_{jk} вертикальными и горизонтальными прямыми с координатами x_j и y_k , такими, что

$$a = x_0 < x_1 < \dots < x_N = c,$$

$$b = y_0 < y_1 < \dots < y_N = d,$$

а обозначение $f(\square)$ было объяснено в § 13.3.

Положительные и отрицательные вариации $\overset{+}{V}_{ab}^{cd}$ и \bar{V}_{ab}^{cd} получаются заменой $|\cdot|$ в (13.A.7) на $[\cdot]^+$ и $[\cdot]^-$, как в одномерном случае. Когда величины $f(\square_{jk})$ суммируются по совокупности малых прямоугольников, покрывающих исходный прямоугольник, все члены, кроме тех, которые соответствуют угловым точкам исходного прямоугольника, взаимно уничтожаются; поэтому

$$\overset{+}{V}_{ab}^{cd} - \bar{V}_{ab}^{cd} = f(c, d) - f(c, b) - f(a, d) + f(a, b).$$

Если заменить c и d на x и y , а a и b рассматривать как постоянные, то можно записать

$$\begin{aligned} f(x, y) &= (\bar{V}_{ab}^{xy} + f(x, b) + f(a, y) + \text{const}) - (\bar{V}_{ab}^{xy} + \text{const}) = \\ &= f_1(x, y) - f_2(x, y), \end{aligned} \quad (13.A.8)$$

где f_1, f_2 — неубывающие функции x и y в том смысле, как это было определено в § 13.3.

Замечание. В этом равенстве члены $f(x, b)$ и $f(a, y)$, каждый из которых зависит от одной переменной, могут быть включены как в f_1 , так и в f_2 , поскольку для них $f(\square)$ всегда равна нулю.

Для n -мерного случая, описанного в конце § 13.3, все делается по тому же образцу. Используя обозначения § 13.3, мы определяем

$$V_a^b = \sup \left\{ \sum_j |f(\square_j)| : \forall P_N, \forall N \right\}. \quad (13.A.9)$$

Тогда комплекснозначную функцию $f(x)$ локально ограниченной вариации можно записать в виде

$$f = f_1 - f_2 + if_3 - if_4, \quad (13.A.10)$$

где каждая из $f_k(x)$ — неубывающая функция.