И. Н. КОВАЛЕНКО
А. А. ФИЛИППОВА

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

И. Н. КОВАЛЕНКО,
А. А. ФИЛИППОВА

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Допущено Министерством высшего и среднего специального образования СССР в качестве учебного пособия для студентов высших технических учебных заведений, обучающихся по специальности «Автоматизированные системы управления» и «Прикладная математика»

МОСКВА «ВЫСШАЯ ШКОЛА» 1973

Коваленко И. Н., Филиппова А. А.

К 56 Теория вероятностей и математическая статистика. Учеб. пособие для втузов. М., «Высшая школа», 1973 г.

368 с. с илл.

В настоящем учебном пособии наряду с традиционными вопросами теории вероятностей излагаются основы теории марковских процессов, методы статистической оценки параметров и проверки статистических гипотез.

Строгость изложения сочетается с минимумом абстрактных математических понятий. Основные положения иллюстрированы рядом примеров из техники, играющих важную роль в развитии умения строить математические модели.

Предназначается для студентов втузов.

К $\frac{0223-449}{001(01)-73}$ БЗ—10—2—73 517.8

Рецензенты: кафедра вычислительной математики и теории вероятностей Университета дружбы народов им. П. Лумумбы и чл.-корр. АН СССР Бусленко Н. П.

Посвящаем светлой памяти Александра Николаевича Рублева, проректора Московского института электронного машиностроения.

ПРЕДИСЛОВИЕ

Настоящая книга представляет собой учебное пособие по теории вероятностей для студентов втузов. Она также будет полезна инженерам, повышающим свою математическую квалификацию.

Последние годы характеризуются интенсивным внедрением теоретико-вероятностных и статистических методов в практику анализа и синтеза сложных технических систем. В этом процессе участвуют тысячи исследователей, имеющих инженерное образование, и сотни математиков-теоретиков. Наряду с классическими задачами статистического анализа, решение которых основывается на биномиальном, нормальном и связанных с ними распределениях, приложения которых неуклонно развиваются, появилось множество задач, связанных с дискретными вероятностными автоматами, приложениями к анализу надежности и эффективности систем, исследованием устойчивости вычислительных процессов на ЭВМ, различными задачами исследования операций. Подобно тому, как математическая логика и теория конечных автоматов заняли свое законное место при изучении дискретных цифровых автоматов, соответствующее место в подготовке специалистов по приложениям теории вероятностей должны занять теория меры и функциональный анализ в абстрактных пространствах. В настоящей книге отражена идейная сторона современного понимания основных вероятностных объектов и приводятся некоторые основные понятия теории функций, служащие основанием вывода теорем теории вероятностей.

Подобная структура имеет ряд важных преимуществ. Выпускник втуза, творчески занимающийся тем или иным приложением математики, сталкивается с необходимостью значительно расширить свои знания за счет изучения научной литературы. Такое изучение будет успешным в том случае, если молодой специалист уже мыслит понятиями, принятыми в современной науке, так как изучение литературы, написанной на непривычном специалисту языке, — процесс довольно сложный.

Абстрактные же математические конструкции, принятые в современной теории вероятностей, в своей основе поразительно близки к конструкциям, применяемым при практическом анализе реальных объектов. Необходимо, чтобы инженер нашел общий язык с математиком, ибо только в этом случае их союз будет плодотворным.

Теоремы, доказательства которых сравнительно легко осуществимы средствами классического анализа, известными студентам из общих математических курсов, выводятся со всеми подробностями.

Основные положения иллюстрированы рядом примеров из техники. Однако эти примеры не следует рассматривать как ключ к пониманию материала, так как сущность математических зависимостей проще понять из математического анализа, чем из рассмотрения физических примеров. Зато в развитии интуиции специалиста, в развитии умения строить математические модели реальных процессов подобные примеры играют важную роль. Поэтому мы рекомендуем студентам строить свои собственные примеры интерпретации различных вероятностных объектов, используя полученные знания.

Книга состоит из 11 глав. В первых семи главах излагаются основы теории вероятностей, в восьмой — элементы теории случайных процессов, в последних трех главах — необходимые сведения из математической статистики.

Объем материала, изложенного в книге, шире, чем предусматривается обычными втузовскими программами. Параграф о больших уклонениях в гл. II, вся гл. VIII о случайных величинах общего вида, гл. VIII о случайных процессах, параграф о мизесовском подходе к предельным распределениям математической статистики выходят за рамки принятых программ. Этот материал может быть использован при изучении курса теории вероятностей на факультетах прикладной математики, готовящих инженеров-математиков.

В главе, посвященной случайным процессам, даются лишь цепи Маркова, ступенчатые, полумарковские и марковские процессы с конечным множеством состояний, т. е. классы случайных процессов, теория которых непосредственно связана с теорией конечномерных случайных величин.

В главы, посвященные математической статистике, включены теория построения точечных оценок и доверительных интервалов, проверки гипотез, а также теория корреляции и регрессии, служащая основой важных методов множественного анализа.

Гл. I—V и VIII написаны И. Н. Коваленко, гл. VI, VII, IX—XI — А. А. Филипповой. Авторы заранее благодарны за замечания и предложения, которые будут высказаны о содержании книги. Все пожелания просим направлять по адресу: Москва, Большой Вузовский переулок, 3/12, МИЭМ, факультет прикладной математики, кафедра теории вероятностей и математической статистики.

Авторы

ГЛАВА I

ОСНОВНЫЕ ПОНЯТИЯ И ФОРМУЛЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

§ 1. КЛАССИЧЕСКАЯ ФОРМУЛА ДЛЯ ВЕРОЯТНОСТИ СОБЫТИЯ

Одним из основных понятий теории вероятностей является понятие *случайного события**. Для объяснения понятия «событие» можно сказать, что под «событием» понимается всякий факт, который в результате эксперимента может произойти или не произойти. Событиями являются, например, появление герба при бросании монеты, получение выигрыша в лотерее, получение «счастливого» билета на экзамене, попадание в цель при выстреле и т. п.

Еще не владея понятием вероятности, а исходя лишь из повседневного опыта, каждый из нас согласится, что различные события можно сравнивать по степени их возможности. Так, попадание в цель с близкого расстояния более возможно, чем с далекого; если в лотерее разыгрывается 10 автомобилей и 100 мотоциклов, то выигрыш мотоцикла — более возможное событие, чем выигрыш автомобиля; при бросании симметричной монеты естественно считать, что оба возможных события — выпадение герба и выпадение надписи — равновозможны.

Нельзя ли ввести числовую характеристику, которая служила бы мерой объективной возможности наступления события? Оказывается, такая числовая характеристика существует, это и есть вероятность события.

Изложим классическое определение вероятности, служившее основой этой науки со времен ее зарождения в XVII столетии (Паскаль, Ферма, Я. Бернулли) вплоть до 20—30-х годов текущего столетия, когда теория вероятностей стала на фундамент аксиоматического метода А. Н. Колмогорова.

* Поскольку в теории вероятностей рассматриваются только случайные события, очень часто слово «случайное» отбрасывают и говорят просто «событие».

В повседневной практике можно сравнивать различные предметы и явления, еще не владея понятием их точной количественной меры. Так, мы часто употребляем выражения «более способный студент», «более интересный фильм», «более комфортабельный автомобиль», «одинаково скучные спектакли», «такое же самочувствие, как и вчера», хотя, конечно, ни один из нас не может точно охарактеризовать определенной числовой функцией показатель умственных способностей, увлекательность кинофильмов и спектаклей, самочувствие человека и т. п.

В геометрии вначале определяется понятие равенства отрезков и понятие «больше», «меньше», а затем уже строится понятие длины и площади фигур. Совершенно так же обстоит дело и в теории вероятностей. Опираясь на повседневную практику, можно принять гипотезу о том, что определенные исходы опыта в заданных условиях равновозможны, и часто этого уже оказывается достаточно для построения понятия вероятности.

Предположим, что в результате эксперимента может произойти одно и только одно из n равновозможных событий. Тогда полагают, что вероятность каждого из этих событий равна $\frac{1}{n}$. Так, выпадение герба при бросании симметричной монеты имеет вероятность $\frac{1}{2}$, выпадение определенного числа очков от 1 до 6 при бросании симметричной игральной кости — вероятность $\frac{1}{6}$; определенную карту из идеально перетасованной колоды из 36 карт можно извлечь с вероятностью $\frac{1}{36}$ и т. п. Таким образом, если в результате эксперимента может произойти одно и только одно из n событий, то следующие два утверждения эквивалентны: а) указанные события равновозможны; б) вероятность каждого из указанных событий равна $\frac{1}{n}$.

Назовем каждое из n равновозможных событий, о которых идет речь, *элементарным событием*.

Любое событие, факт наступления которого вполне определяется тем, какое из элементарных событий наступило, называется *случайным событием*.

Перенумеруем элементарные события, обозначив их $\omega_1, \omega_2, \ldots, \omega_n$. То из элементарных событий, которое происходит в результате эксперимента, обозначим ω. Случайные события будем обозначать большими буквами латинского алфавита (возможно, с индексами). Пусть A — любое случайное событие. По условию, факт наступления события A зависит от ω. При некоторых значениях ω из множества $\{\omega_1, \omega_2, \ldots, \omega_n\}$ событие A наступает, при остальных — не наступает. Элементарное событие, при котором событие A наступает, назовем *благоприятствующим событию* A. Остальные элементарные события назовем *не благоприятствующими событию* A. Следовательно, любое событие вполне задается указанием тех ω_i, скажем, $\omega_{i_1}, \ldots, \omega_{i_k}$, которые ему благоприятствуют. Это можно записать так: $A = \{\omega_{i_1}, \ldots, \omega_{i_k}\}$ или с помощью таблицы, отмечая благоприятствующие элементарные события символом «1», не благоприятствующие — символом «0». Пусть, например, событие A состоит в выпадении четного числа очков при бросании симметричной игральной кости, где в качестве элементарных событий взяты события ω_i, состоящие в выпадении i очков ($i = 1, 2, \ldots, 6$). Тогда событию A можно сопоставить таблицу

ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
0	1	0	1	0	1

Из сказанного следует, что данные n равновозможных событий определяют 2^n различных случайных событий — именно столько, сколько можно построить функций с двумя возможными значениями при n значениях аргумента.

Функция $I_A(\omega)$ элементарного события ω, равная единице при ω, благоприятствующих событию A, и равная нулю при ω, не благоприятствующих этому событию, называется *индикатором* события A.

Вероятность $\mathrm{P}(A)$ любого события A, по определению, есть отношение числа m элементарных событий, благоприятствующих этому событию, к общему числу n равновозможных событий:

$$\mathrm{P}(A) = \frac{m}{n}. \tag{1.1}$$

Нетрудно видеть, что

$$\mathrm{P}(A) = \frac{1}{n}\sum_{i=1}^{n} I_A(\omega_i). \tag{1.2}$$

Среди случайных событий выделяются достоверное событие Ω — такое, которому благоприятствуют все элементарные события, и невозможное событие $\varnothing$ — событие, которому не благоприятствует ни одно элементарное событие. По определению, $I_\Omega(\omega_i) = 1$, $I_\varnothing(\omega_i) = 0$ для любого i, $1 \leqslant i \leqslant n$, откуда

$$P(\Omega) = 1, \quad P(\varnothing) = 0. \tag{1.3}$$

Заметим, что любое событие можно определять различными наборами условий. Пусть, например, A — событие, состоящее в том, что число выпавших очков при бросании игральной кости равно либо 2, либо 3, либо 4. Тогда это условие можно записать с помощью формулы $(\omega - 3)^2 \leqslant 1$ или $(\omega - 3)^2 \leqslant 3$.

Из принятого определения вероятности вытекает, что всегда

$$0 \leqslant P(A) \leqslant 1. \tag{1.4}$$

Подчеркнем следующий важный факт. Допустим, что мы хотим применить теорию вероятностей для расчета вероятности какого-нибудь интересующего нас физического события. Тогда расчет, основанный на классическом определении вероятности, будет точным, если нами правильно определены элементарные события и их вероятности. Последняя задача не может быть решена в рамках теории вероятностей как математической науки. Так, невозможно математически доказать, что вероятность выпадения герба равна 1/2 — это эмпирический факт; но коль скоро этот факт уже принят как рабочая гипотеза, вступают в силу вероятностные законы, и тогда можно найти вероятность любого события, зависящего от элементарных событий.

Сделать заключение о том, правильно ли выбраны элементарные события, можно на основании изучения физической стороны явления, а также используя методы математической статистики, элементы которой приводятся в соответствующих главах данной книги.

Рассмотрим некоторые примеры определения вероятностей событий, основанного на классическом определении вероятности.

Пример 1. Электронное устройство состоит из 5 элементов и функционирует нормально, если исправны все элементы. При сборке устройства элементы выбираются из партии объема 1000; любой из возможных C_{1000}^5 способов выбора имеет одну и ту же вероятность. В партии 950 исправных и 50 неисправных элементов. Событие A

состоит в том, что устройство работает исправно. Требуется найти $P(A)$.

Число элементарных событий равно C_{1000}^5. Элементарное событие благоприятствует событию A в том и только том случае, если все элементы выбраны из числа исправных; таким образом, число благоприятствующих событию A элементарных событий равно C_{950}^5 (число возможных способов выбора 5 элементов из 950). Следовательно,

$$P(A)=\frac{C_{950}^5}{C_{1000}^5}=\frac{950\cdot 949\cdot 948\cdot 947\cdot 946}{1000\cdot 999\cdot 998\cdot 997\cdot 996}\approx 0{,}77.$$

Пример 2. Все автомобили имеют четырехзначные номера от 00—00 до 99—99. Считая, что каждый из них может встретиться с одной и той же вероятностью, найти вероятность того, что первый встречный автомобиль обладает «счастливым» номером: сумма первых двух цифр номера равна сумме последних двух.

Общее число элементарных событий равно 10 000. Пусть $r(n)$ — количество двузначных чисел* с суммой цифр, равных n. Тогда число благоприятствующих элементарных событий равно $\sum_n r^2(n)$ [при данной сумме цифр, равной n, как первую, так и вторую половину записи «счастливого» номера можно выбрать $r(n)$ способами; всего же способов будет $r^2(n)$]. Нетрудно подсчитать, что

$$r(n)=\begin{cases} n+1 & \text{при } 0\leqslant n\leqslant 9,\\ 19-n & \text{при } 10\leqslant n\leqslant 18.\end{cases}$$

[В справедливости этой формулы легко убедиться, представив двузначное число $10x+y$ в виде точки $(x;\ y)$ координатной плоскости].

Таким образом, число m благоприятствующих элементарных событий

$$m=\sum_{n=0}^{9}(n+1)^2+\sum_{n=10}^{18}(19-n)^2.$$

Заменив в первой сумме $n+1$ на k, а во второй $19-n$ на k, получим

$$m=\sum_{n=1}^{10}k^2+\sum_{n=1}^{9}k^2=10^2+2\sum_{n=1}^{9}k^2=670.$$

Таким образом, вероятность «счастливого» номера равна $\frac{670}{10\,000}=$ $=0{,}067$.

Пример 3. Пусть задана произвольная функция $f(x)$, $0\leqslant f(x)\leqslant 1$, $0\leqslant x\leqslant 1$. Результатом эксперимента является пара целых чисел $(i,\ j)$, причем все возможные пары, подчиняющиеся

---

* В данном случае числа 00, 01 и т. д. также считаются двузначными.

условию $1 \leqslant i \leqslant N$, $1 \leqslant j \leqslant N$, равновероятны. Найти вероятность события A, состоящего в том, что $\frac{j}{N} < f\left(\frac{i}{N}\right)$.

В данном случае число элементарных событий $n = N^2$. По формуле (1.2) находим

$$\mathrm{P}(A) = \frac{1}{N^2} \sum_{i,\, j=1}^{N} I_A(i, j),$$

где

$$I_A(i, j) = \begin{cases} 1, \text{ если } \frac{j}{N} < f\left(\frac{i}{N}\right), \\ 0 \text{ — в противном случае.} \end{cases}$$

Если начертить график функции $y = f(x)$ и отметить в квадрате плоскости (x, y) с вершинами (0, 0), (1, 0), (1, 1) и (0, 1) точки вида $\left(\frac{i}{N}, \frac{j}{N}\right)$, где i и j — целые числа, то последняя формула для $\mathrm{P}(A)$ будет представлять число точек, лежащих под кривой $y = f(x)$, умноженное на $\frac{1}{N^2}$. Легко видеть, что если функция $f(x)$ интегрируема в смысле Римана, то при $N \to \infty$ эта величина сходится к $\int_0^1 f(x)\,dx$.

§ 2. СТАТИСТИЧЕСКАЯ УСТОЙЧИВОСТЬ ЧАСТОТ

При классическом подходе определение вероятности сводится к более простому понятию — равновозможности событий. Как было указано выше, в рамках теории вероятностей невозможно доказать равновозможность тех или иных событий: для этого нужно обращаться к изучению конкретных физических, а не только математических объектов.

Основанием для применения методов теории вероятностей к изучению реальных явлений является объективное существование реальных событий, обладающих свойством устойчивости частот. В природе, технике и т. п. имеется возможность наблюдать повторяющиеся опыты, в результате каждого из которых может произойти или не произойти некоторое событие A. Ясно, что наступление события зависит от условий, в которых производится опыт. Так, окончание химической реакции в заданное время зависит от температуры воздуха, массы реагентов, способа их смешивания и т. д. Падение монеты на определенную сторону зависит, например, от высоты подбрасывания. Попадание снаряда в цель зависит от температуры воздуха, силы и направления ветра, веса заряда и т. д.

Некоторые из условий опыта поддаются физическому измерению; эти условия могут быть выражены одним или несколькими числовыми параметрами, характеризующими опыт; другие же условия, называемые неконтролируемыми, связаны с проявлением хаотических, быстро меняющихся физических процессов, обычно не поддающихся точному измерению. Допустим, что для данной последовательности опытов (испытаний) выполнены следующие два условия:

1) числовые параметры во всех испытаниях одни и те же;

2) неконтролируемые условия, соответствующие различным испытаниям, определяются физически не связанными процессами.

Пусть $\nu_i = 1$, если в i-м испытании событие A произошло, и $\nu_i = 0$ в противном случае. Отношение

$$\bar{p}_n = \frac{1}{n}(\nu_1 + \ldots + \nu_n) \tag{1.5}$$

называется *частотой* события A в первых n испытаниях.

Если при условиях 1) и 2) для бесконечной последовательности испытаний

$$\bar{p}_n \xrightarrow[n\to\infty]{} p,$$

где p — постоянное число, то, по определению, событие A при данных числовых параметрах, характеризующих испытание, обладает вероятностью, равной p: $\mathrm{P}(A) = p$. Данное определение называется *статистическим определением вероятности.*

Рассмотрим для проведения параллели математическое определение плотности тела в точке x:

$$\rho(x) = \lim_{r\to 0} \frac{m(x, r)}{V(r)}, \tag{1.6}$$

где $m(x, r)$ — масса части тела, попадающая в шар радиуса r с центром x, $V(r)$ — объем этого шара. В буквальном смысле это определение понимать нельзя вследствие дискретности строения материи, а также различного рода движений частиц вещества. Точно так же нельзя реально представить себе возможность бесконечной последовательности опытов, да еще с одними и теми же значениями параметров.

На практике мы имеем дело с конечными совокупностями опытов, однако число их может быть настолько большим, что отклонение частоты от вероятности пренебрежимо мало. Поэтому важен лишь следующий практический вывод: объективную возможность осуществления события можно измерить принципиально со сколь угодно высокой точностью, осуществив достаточно большое число испытаний и подсчитав частоту наступления события в этой совокупности испытаний.

Мы пришли бы к абсурду, к противоречию с данными о строении материи, если бы считали, что существует предел отношения $\frac{m(x, r)}{V(r)}$ при $r \to 0$; если же это понятие использовать в определенных рамках, не пытаясь абсолютизировать его, мы получим выводы, согласующиеся с практикой. Подобно этому, понятие вероятности как предела частоты многократно используется в случаях, когда имеются лишь конечные множества опытов и частота имеет тенденцию к приближению к некоторому числу при увеличении числа опытов, хотя точно измерить это число и невозможно.

Рассмотрим второе требование — условие физической несвязанности процессов, определяющих неконтролируемые условия испытаний. Здесь также недопустимо буквальное понимание определения: все явления в природе, как известно, находятся во взаимосвязи. Однако взаимосвязь различных явлений может быть сильной или слабой. Свойством человеческого мышления, без которого, вероятно, научное познание мира было бы невозможным, является способность видеть главные факторы, влияющие на то или иное явление, отбрасывая второстепенные. Эта особенность есть результат объективного отражения природы реального мира, где существуют слабо связанные качественно устойчивые предметы и явления. Поэтому понятие физически не связанных испытаний хотя и является абстрактным, легко согласуется с нашими интуитивными представлениями, а выводы, получаемые на основании использования статистического определения вероятности, прекрасно согласуются с практикой.

Приведем пример использования понятия статистической вероятности для получения практических выводов.

Пример. Завод выпускает массовую продукцию. Если изделие, выпущенное заводом и поступившее в продажу, выходит из строя в течение года, то требуется заменить его запасным. Спрашивается: сколько необходимо запасных изделий, если в течение года продается N изделий, произведенных заводом? (Для простоты предположим, что выход из строя запасного изделия невозможен.)

Для практического решения сформулированной задачи поступают следующим образом. Считают, что событие «отказ изделия» обладает некоторой вероятностью p. Тогда отношение числа вышедших из строя изделий к N есть частота события при N испытаниях: $\overline{p}_N = \frac{m}{N}$. Если N достаточно велико, то частота приближенно равна вероятности: $\overline{p}_N \approx p$. Отсюда можно сделать вывод, что $m \approx \approx Np$ (приближенное равенство в смысле малой относительной погрешности). Таким образом, при известном p можно дать приблизительную оценку числа вышедших из строя изделий, а значит, и требуемого числа запасных изделий. Предположим, однако, что p — неизвестная величина. Тогда следует поставить некоторое число (n) опытов, каждый из которых состоит в испытании некоторого изделия в течение года. Пусть из этих n изделий отказано k. Вычисляем частоту: $\overline{p}_n = \frac{k}{n}$. Если n достаточно велико, то, как и ранее, можно считать, что $\overline{p}_n \approx p$. Таким образом, находим, что $m \approx N\overline{p}_n$*.

По такой же схеме оценивают урожай на плантации по урожаю на некотором числе небольших участков, удельный вес минерала по результатам взвешивания некоторого числа образцов, а также множество других физических, биологических, экономических и т. п. характеристик.

§ 3. СЛУЧАЙНЫЕ СОБЫТИЯ

Уточним данное в начале главы чисто описательное определение случайного события.

В каждой вероятностной задаче задается некоторое множество Ω элементов или точек ω. Ω называется основным (выборочным) пространством (пространством элементарных событий), ω — элементарными событиями. В классической схеме, как мы видели, Ω — конечное множество: $\Omega = \{\omega_1, \ldots, \omega_n\}$, где ω_i — равновозможные события. Однако в ряде задач множество Ω бесконечно. Так, при стрельбе по мишени в качестве множества Ω можно взять совокупность всех точек мишени и еще одну дополнитель-

---

* Уточнение этих приближенных соотношений является одной из задач математической статистики. В частности, оказывается, что запасных изделий нужно брать несколько больше, чем Np, даже при известном p.

ную «точку» на случай отсутствия попадания в мишень. Для такого случайного события, как отказ физического элемента, основное пространство представляет собой числовой луч $\{\omega \geqslant 0\}$, а именно, ω — момент отказа элемента.

Итак, должно быть задано множество Ω. Любое случайное событие отождествляется с некоторым подмножеством A множества Ω и интерпретируется как попадание элементарного события ω в множество A. Ввиду такого тесного соответствия между событием и множеством они могут обозначаться одним и тем же символом. Если событие A определяется некоторым условием Γ, наложенным на ω, то пишут $A = \{\Gamma\}$. Так, если ω — точки прямой, то можно определить события $\{a < \omega < b\}$, $\{\omega \leqslant b\}$, $\{|\omega| \geqslant c\}$, $\{\omega = \omega^2\}$ и т. п.

Перейдем к действиям над событиями. Пусть $A_1, \ldots, A_n$ — множества. *Объединением* этих множеств называется множество C, состоящее из тех и только тех элементов, которые принадлежат хотя бы одному из A_i, $1 \leqslant i \leqslant n$. Объединение обозначается так: $C = A_1 \cup \ldots \cup A_n = \bigcup\limits_{i=1}^{n} A_i$. *Пересечением* $A_1, \ldots, A_n$ называется множество D, содержащее те и только те элементы, которые принадлежат всем A_i, $1 \leqslant i \leqslant n$, одновременно. Пересечение обозначается так: $D = A_1 \cap \ldots \cap A_n = \bigcap\limits_{i=1}^{n} A_i = A_1 \ldots A_n$.

Если все рассматриваемые множества подмножества некоторого множества Ω, которое считается фиксированным, то *дополнением* множества A называется множество тех элементов Ω, которые не принадлежат A. Дополнение множества A обозначается символом $\overline{A}$. В теории множеств используется также символ разности «$\setminus$». Так, $A \setminus B$ есть множество всех элементов A, не являющихся элементами B.

Эти операции над множествами позволяют определить операции над случайными событиями. Сформулируем соответствующие определения.

Пусть A_α — события при $\alpha \in I$, где I — заданное множество. Тогда $\bigcup\limits_{\alpha \in I} A_\alpha$ — событие, состоящее в наступлении хотя бы одного из событий A_α, называется объединением (суммой) событий A_α, $\alpha \in I$. Событие $\bigcap\limits_{\alpha \in I} A_\alpha$, состоящее в том, что произойдут все события A_α, $\alpha \in I$, называется *пересечением* (произведением) событий A_α, $\alpha \in I$. Если

$I = \{1, 2, \dots, n\}$, то обозначают:

$$\bigcup_{\alpha\in I} A_\alpha = \bigcup_{i=1}^{n} A_i = A_1 \cup \dots \cup A_n,$$

$$\bigcap_{\alpha\in I} A_\alpha = \bigcap_{i=1}^{n} A_i = A_1 \cap \dots \cap A_n = A_1 \dots A_n.$$

Если $I = \{1, 2, \dots\}$, то обозначают:

$$\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \dots; \quad \bigcap_{i=1}^{\infty} A_i = A_1 A_2 \dots .$$

Событие $\Omega = \{\omega \in \Omega\}$ называется *достоверным* событием. Событие $\varnothing = \{\omega \in \varnothing\}$ (здесь $\varnothing$ — символ пустого множества) называется *невозможным* событием. Событие $A \setminus B$, состоящее в том, что происходит A, но не происходит B, называется *разностью событий* A и B. Событие $\Omega \setminus A$ называется событием, *противоположным* A, или *дополнением* к A, и обозначается $\bar{A}$:

$$\bar{A} = \{\omega \bar{\in} A\} = \{\omega \in \Omega \setminus A\}. \tag{1.7}$$

Можно установить соответствие между событиями, с одной стороны, и логическими высказываниями — с другой. Так, событию A можно поставить во взаимно однозначное соответствие высказывание A^*: «A произошло». A^* истинно в том и только в том случае, если событие A происходит. Если обозначить высказывания указанного вида теми же символами, что и соответствующие события, со значком «$*$» вверху, то имеют место следующие соотношения:

$$\left(\bigcup_{\alpha\in I} A_\alpha\right)^* = \exists_{\alpha\in I} A_\alpha^*, \quad \left(\bigcap_{\alpha\in I} A_\alpha\right)^* = \forall_{\alpha\in I} A_\alpha^*, \quad (\bar{A})^* = \sim A^*,$$

где $\exists$ — квантор существования, $\forall$ — квантор общности $\sim$ — символ логического отрицания.

Напомним еще символ включения $\subset$, используемый в теории множеств: $A \subset B$, если любой элемент A входит в B (в этом случае также пишут $B \supset A$). Мы договорились считать, что события отождествляются с подмножествами некоторого множества Ω. Пусть $A = \{\omega \in A\}$, $B = \{\omega \in B\}$. В случае соотношения $A \subset B$ между соответствующими подмножествами Ω говорят, что A есть частный случай B или что из A следует B (A влечет за собой B).

Можно, очевидно, записать

$$A = \bigcup_{\omega \in A} \omega,$$

откуда

$$A^* = \mathop{\exists}_{\omega \in A} \omega.$$

Если $A \subset B$, то

$$A^* = \mathop{\exists}_{\omega \in A} \omega \Rightarrow \mathop{\exists}_{\omega \in B} \omega = \mathop{\exists}_{\omega \in A \cup (B \setminus A)} \omega = B,$$

где $\Rightarrow$ — символ логического следования.

Если $A \subset B$ и $B \subset A$, то события A и B называют *равными* и пишут $A = B$.

Пример. Техническая система состоит из двух элементов, которые могут находиться либо в исправном, либо в неисправном состоянии. Система отказывает, если оба элемента неисправны. Требуется записать событие A, состоящее в том, что до момента t включительно система не откажет.

В качестве элементарного события возьмем пару $\omega = (\omega_1, \omega_2)$, где ω_i= момент отказа i-го элемента ($i = 1, 2$). Тогда событие $\overline{A}$ состоит в том, что система откажет до момента t. Однако $\overline{A} = \{\omega_1 \leqslant t, \ \omega_2 \leqslant t\} = \{\max\{\omega_1, \ \omega_2\} \leqslant t\}$. Отсюда $A = \{\max\{\omega_1, \omega_2\} > t\}$.

Наличие соответствия между событиями и множествами позволяет применять к действиям над событиями соответствующие действия над множествами. Выпишем некоторые основные соотношения такого рода:

1. $A \cup A = A$.
2. $A \cap A = A$.
3. $A \cup \overline{A} = \Omega$.
4. $A \cap \overline{A} = \varnothing$.
5. $A \cup \Omega = \Omega$.
6. $A \cap \Omega = A$.
7. $A \cup \varnothing = A$.
8. $A \cap \varnothing = \varnothing$.
9. Если $A \subset B$, $B \subset C$, то $A \subset C$.
10. Если $A \subset B$, то $B = A \cup (B \setminus A)$.
11. $A \setminus B = A\overline{B}$.

12. $A \cup B = A \cup (B \setminus A)$.

13. $A \cup (B \cup C) = (A \cup B) \cup C$.

14. $A \cap (B \cap C) = (A \cap B) \cap C$.

15. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

16. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

17. $\bigcup\limits_{\alpha \in I} \overline{A}_\alpha = \overline{\bigcap\limits_{\alpha \in I} A_\alpha}$.

18. $\bigcap\limits_{\alpha \in I} \overline{A}_\alpha = \overline{\bigcup\limits_{\alpha \in I} A_\alpha}$*.

Каждая из этих формул устанавливается непосредственным путем.

Для примера выведем семнадцатую формулу. Пусть $\omega \in \bigcup\limits_{\alpha \in I} \overline{A}_\alpha$. Тогда $\exists_{\alpha \in I} \{\omega \in \overline{A}_\alpha\}$, или, что то же, $\exists_{\alpha \in I} \{\omega \,\overline{\in}\, A_\alpha\}$. Но в таком случае $\omega \,\overline{\in} \bigcap\limits_{\alpha \in I} A_\alpha$, т. е. $\omega \in \overline{\bigcap\limits_{\alpha \in I} A_\alpha}$. Мы доказали, что левая часть формулы представляет собой множество, входящее в множество, записанное в правой части. Пусть теперь $\omega \in \overline{\bigcap\limits_{\alpha \in I} A_\alpha}$. Тогда $\omega \,\overline{\in} \bigcap\limits_{\alpha \in I} A_\alpha$, а следовательно, $\exists_{\alpha \in I} \{\omega \,\overline{\in}\, A_\alpha\}$, или, что то же, $\exists_{\alpha \in I} \{\omega \in \overline{A}_\alpha\}$. Окончательно $\omega \in \bigcup\limits_{\alpha \in I} \overline{A}_\alpha$. Утверждение доказано.

Заметим, что подобного рода рассуждения становятся громоздкими при оперировании со сложными выражениями, как, например,

$$\left(\left(\overline{(A \cap B) \cup (C \cap \overline{D})}\right) \cap (A \cup (D \cap \overline{C}))\right) \cup (\overline{F} \cap A \cap (\overline{B} \cup C)).$$

Поэтому приведем универсальный способ проверки соотношений между событиями.

Пусть имеется любое событие A, записанное в виде выражения, включающего события $A_1, A_2, \ldots, A_n$ и символы вида $\cup$, $\cap$, $\overline{}$. Рассмотрим высказывание A^*: {событие A произошло}. Если A_i^* — высказывание {событие A_i произошло} $(i = 1, 2, \ldots, n)$, то A^* можно записать в виде логической функции, причем эту запись можно получить непосредственно из записи выражения A, если заменить A_i на A_i^* и символы $\cup$, $\cap$, $\overline{}$ — логическими сим-

* Соотношения 17 и 18 называются правилами де-Моргана.

волами $\vee$ (или), $\wedge$ (и), $\sim$ (не). Так, высказывание, соответствующее последнему примеру, имеет вид

$$(\sim((A^* \wedge B^*) \vee (C^* \wedge \sim D^*)) \wedge (A^* \vee (D^* \wedge \sim C))) \vee \vee (\sim F^* \wedge A^* \wedge (\sim B^* \vee C^*)).$$

Каждой из переменных A_1^*, A_2^*, ..., A_n^* можно придавать числовые значения: $A_i^* = 1$ (высказывание A_i^* истинно) и $A_i^* = 0$ (высказывание A_i^* ложно). Если все эти символы заменены числами, то можно вычислить значение выражения $A^* = A^*(x_1, x_2, \ldots, x_n)$, где x_i — конкретное числовое значение, подставленное вместо A_i^*. При этом надлежит пользоваться следующими соотношениями: $1 \vee 1 = 1 \wedge 1 = 1$, $1 \vee 0 = 0 \vee 1 = 1$, $1 \wedge 0 = 0 \wedge 1 = 0$, $0 \vee 0 = 0$, $0 \wedge 0 = 0$, $\sim 1 = 0$, $\sim 0 = 1$ и производить последовательные действия аналогично правилам обычной арифметики. Перебрав все 2^n наборов $(x_1, \ldots, x_n)$, получим таблицу истинности высказывания A^*:

№	$(x_1, x_2, \ldots, x_n)$	$A^*(x_1, x_2, \ldots, x_n)$
1	$(0, 0 \ldots 0, 0)$	$A^*(0, 0 \ldots 0, 0)$
2	$(0, 0 \ldots 0, 1)$	$A^*(0, 0 \ldots 0, 1)$
.	. . . .	. . . .
.	. . . .	. . . .
.	. . . .	. . . .
2^n	$(1, 1 \ldots 1, 1)$	$A^*(1, 1 \ldots 1, 1)$

Из таблицы видно, произойдет ли событие A, если известно, какие из событий A_i произошли, а какие — нет. Например, пусть события A_1, A_3, A_5, ... произошли, а A_2, A_4, A_6, ... не произошли. Тогда событие A происходит, если $A^*(1, 0, 1, 0, 1, \ldots) = 1$, и не происходит в противном случае.

Пусть A и B — два события, записанные в виде выражений, включающих события A_1, A_2, ..., A_n. Составим соответствующие таблицы истинности $A^*(x_1, \ldots, x_n)$ и $B^*(x_1, \ldots, x_n)$, или сокращенно $A^*(x)$ и $B^*(x)$, где $x = (x_1, \ldots, x_n)$. Обозначим через X множество возможных значений x, т. е. множество n-мерных векторов со значениями координат 1 и 0. Тогда справедливы следующие утверждения:

$$\underset{x \in X}{\forall}\{A^*(x) \leqslant B^*(x)\} \Longleftrightarrow \{A \subset B\};$$

$$\underset{x \in X}{\forall}\{A^*(x) = B^*(x)\} \Longleftrightarrow \{A = B\}.$$

Этим способом можно проверить выполнение равенства или соотношения следования для двух случайных событий.

(Заметим, что замена теоретикомножественных символов логическими и отметка символов событий звездочками

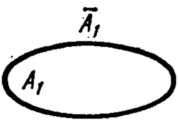

Рис. 1.1

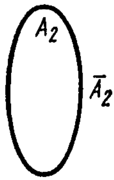

Рис. 1.2

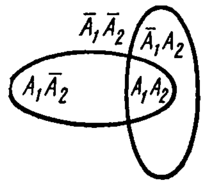

Рис. 1.3

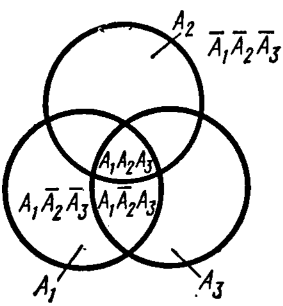

Рис. 1.4

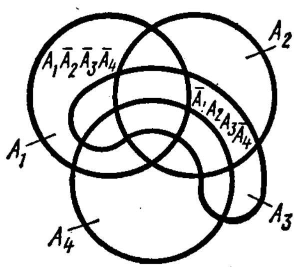

Рис. 1.5

сделаны лишь для разъяснения смысла описанной конструкции; на практике же можно оперировать непосредственно теоретикомножественными символами.)

Для вычислений по данному алгоритму существуют программы, реализованные на ЭВМ.

При небольшом числе событий A_i (до 4—5) теоретикомножественные выражения, включающие их, анализируются с помощью диаграмм Венна, которые строятся следующим образом.

Начертим замкнутую кривую (рис. 1.1) и предположим, что событие A_1 представляется множеством точек, ограниченных этой кривой. Проведем еще одну замкнутую кривую (рис. 1.2) и будем интерпретировать событие A_2 множеством точек, ограниченных этой кривой. Тогда четыре части плоскости, на которые она разбита двумя кривыми, соответствуют событиям A_1A_2, $\overline{A}_1A_2$, $A_1\overline{A_2}$, $\overline{A}_1\overline{A_2}$ (рис. 1.3). Точно так же можно представить три события A_1, A_2, A_3 (рис. 1.4) и четыре события (рис. 1.5). Важно, чтобы при этом n кривых разбили плоскости на 2^n частей, причем так, чтобы нашлась область, соответствующая любому событию вида $A_1A_2 \ldots A_n$, $\overline{A}_1A_2 \ldots A_n$, ..., $\overline{A}_1 \; \overline{A}_2 \ldots \overline{A}_n$.

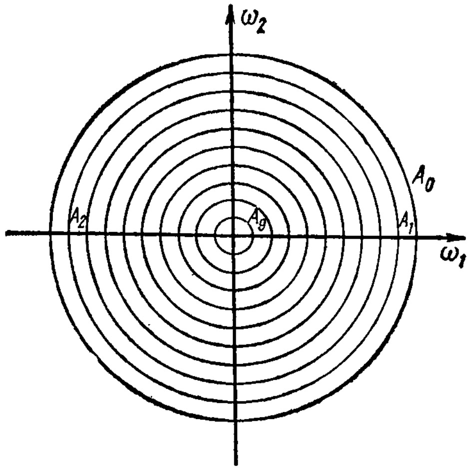

Рис. 1.6

По диаграмме Венна можно непосредственно выводить соотношения между событиями, комбинируя различные части плоскости. Так, из рис. 1.3 видно, что $\overline{A}_2 = A_1\overline{A}_2 \cup \overline{A}_1\overline{A}_2$.

Диаграмма Венна — это лишь графическая альтернатива изложенной выше логической интерпретации соотношений между событиями. Однако им можно придать и более непосредственный смысл. Так, допустим, что элементарные события отождествлены с точками плоскости: $\omega = (\omega_1, \omega_2)$, где ω_1, ω_2 — координаты точки. (В этом случае говорят: на плоскость случайно бросается точка). Очертив какую-нибудь фигуру, можно говорить о событии, состоящем в попадании случайной точки внутрь этой фигуры.

Так, пусть производится стрельба по мишени (рис. 1.6). Обозначим через A_i событие, состоящее в наборе стрелком i очков ($i = 0, 1, 2, \ldots, 10$). Если $\omega = (\omega_1, \omega_2)$ — точка попадания пули в мишень, то событие A_i отождествляется с множеством точек ω, принадлежащих соответствующей области плоскости. Если A — событие, состоящее в попадании в мишень, то можно записать

$$A = \bigcup_{i=1}^{10} = \Omega \setminus A_0 = \overline{A}_0.$$

Предположим, что стрелок делает 3 выстрела, и обозначим через A_{ki} событие, состоящее в наборе i очков при k-м выстреле. Рассмотрим событие B, состоящее в наборе стрелком не менее 28 очков. Имеем

$$B = \bigcup_{i_1+i_2+i_3\geqslant 28} A_{1i_1} A_{2i_2} A_{3i_3}.$$

Допустим, что при некотором выстреле стрелок набрал менее 8 очков; тогда не может быть $i_1 + i_2 + i_3 \geqslant 28$, поскольку $i_1 \leqslant 10$, $i_2 \leqslant 10$, $i_3 \leqslant 10$. Значит, каждый из индексов i_1, i_2, i_3 может принимать только значения 8, 9, 10. Возможные сочетания индексов i_1, i_2, i_3, при которых $i_1 + i_2 + i_3 \geqslant 28$, следующие:

№	i_1	i_2	i_3	№	i_1	i_2	i_3
1	8	10	10	6	10	10	9
2	10	8	10	7	10	9	9
3	10	10	8	8	9	10	9
4	9	10	10	9	9	9	10
5	10	9	10	10	10	10	10

Таким образом,

$$B = A_{1,8} A_{2,10} A_{3,10} \cup A_{1,10} A_{2,8} A_{3,10} \cup \ldots \cup A_{1,10} A_{2,10} A_{3,10},$$

т. е. B — объединение 10 событий. Это же событие можно записать так:

$$B = A_{1,8} A_{2,10} A_{3,10} \cup A_{1,9} (A_{2,9} A_{3,10} \cup A_{2,10} (A_{3,9} \cup A_{3,10})) \cup$$

$$\cup A_{1,10}(A_{2,8}A_{3,10} \cup A_{2,9}(A_{3,9} \cup A_{3,10}) \cup A_{2,10}(A_{3,8} \cup A_{3,9} \cup A_{3,10})).$$

Обозначим через C событие, состоящее в том, что при втором выстреле набрано большее число очков, чем при первом. Имеем

$$C = \bigcup_{i=0}^{10} A_{1i} \left(\bigcup_{j=i+1}^{10} A_{2,\,j} \right).$$

В заключение параграфа введем следующие важные понятия.

События A *и* B *называются несовместными, если* $AB = \varnothing$. Так, события A и $\overline{A}$ всегда несовместны.

События A_α, $\alpha \in I$, *называются несовместными, если несовместны* A_α *и* A_β *при* $\alpha \neq \beta$, $\alpha \in I$, $\beta \in I$.

События A_α, $\alpha \in I$, *образуют полную систему событий, если*

$$\bigcup_{\alpha\in I} A_\alpha = \Omega;\ A_\alpha A_\beta = \varnothing,\ \alpha \neq \beta,\ \alpha \in I,\ \beta \in I.$$

Записанные соответствия можно объяснить так: в рассматриваемом опыте произойдет одно и только одно из событий A_α.

Пусть имеется n равновозможных событий, каждое из которых имеет вероятность $1/n$. Эти события несовместны и образуют полную систему событий. Обозначим указанные события $\omega_1, \omega_2, \dots, \omega_n$. Любая полная система $\{A_\alpha\}$, $\alpha \in I$, событий в данном случае может быть представлена следующим образом. Любому $\alpha \in I$ ставится в соответствие множество индексов $I_\alpha \subset \{1, 2, \dots, n\}$: при $i \in I_\alpha$ элементарное событие ω_i благоприятствует A_α, а при $i \overline{\in} I_\alpha$ — не благоприятствует.

§ 4. АКСИОМАТИЧЕСКОЕ ОПРЕДЕЛЕНИЕ ВЕРОЯТНОСТИ

В настоящем параграфе изложено аксиоматическое определение вероятности, предложенное А. Н. Колмогоровым; это определение служит логическим фундаментом всей современной теории вероятностей.

Пусть имеется произвольное множество $\Omega = \{\omega\}$ элементарных событий. Вначале определяется некоторая система $\mathfrak{A}$ подмножеств Ω, называемых (случайными) событиями; для каждого $A \in \mathfrak{A}$ определяется числовая функция $\mathrm{P}(A)$; $\mathrm{P}(A)$ называется вероятностью события A. Множество $\mathfrak{A}$ должно быть σ-алгеброй. *σ-алгеброй* называется непустая совокупность $\mathfrak{A}$ подмножеств A некоторого множества Ω со следующими свойствами:

1) $\Omega \in \mathfrak{A}$;

2) если $A_n \in \mathfrak{A}$, $n = 1, 2, \dots$, то $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$;

3) если $A \in \mathfrak{A}$, $B \in \mathfrak{A}$, то $A \setminus B \in \mathfrak{A}$.

Из свойств 1—3 вытекают еще такие свойства:

4) если $A \in \mathfrak{A}$, $B \in \mathfrak{A}$, то $A \cup B \in \mathfrak{A}$;

5) если $A \in \mathfrak{A}$, $B \in \mathfrak{A}$, то $A \cap B \in \mathfrak{A}$;

6) если $A \in \mathfrak{A}$, то $\bar{A} \in \mathfrak{A}$.

Докажем четвертое, пятое и шестое свойства. Пусть $A \in \mathfrak{A}$. Тогда поскольку также $\Omega \in \mathfrak{A}$, то вследствие свойства 3 $\bar{A} = \Omega \setminus A \in \mathfrak{A}$, что доказывает справедливость свойства 6. Далее, $\Omega \setminus \Omega = \varnothing \in \mathfrak{A}$. Пусть теперь $A \in \mathfrak{A}$, $B \in \mathfrak{A}$. Положив $A_1 = A$, $A_2 = B$, $A_3 = A_4 = \ldots$ $\ldots = \varnothing$, найдем $\bigcup\limits_{n=1}^{\infty} A_n = A \cup B \in \mathfrak{A}$, что доказывает справедливость свойства 4. Наконец, по правилу де-Моргана, $\overline{A \cap B} = \bar{A} \cup \bar{B}$. Если $A \in \mathfrak{A}$, $B \in \mathfrak{A}$, то $\bar{A} \in \mathfrak{A}$, $\bar{B} \in \mathfrak{A}$, по доказанному свойству 6. Отсюда на основании свойства 4 $\bar{A} \cup \bar{B} \in \mathfrak{A}$, т. е. $\overline{A \cap B} \in \mathfrak{A}$. Однако событие $A \cap B$ противоположно событию $\overline{A \cap B}$; значит, вследствие свойства 6 $A \cap B \in \mathfrak{A}$. Свойство 5 также установлено.

Таким образом, σ-алгебра $\mathfrak{A}$ обладает тем свойством, что операции $\cup$, $\setminus$, $\cap$, —, производимые над множествами из $\mathfrak{A}$, не выводят за ее пределы; операцию $\cup$ можно при этом производить бесконечное (счетное) число раз*.

Вспомним следующее свойство. Для каждого $A \in \mathfrak{A}$ определена вероятность $\mathrm{P}(A)$. Значит, если некоторые события A_n принадлежат $\mathfrak{A}$, то с помощью всевозможных теоретикомножественных операций из $\{A_n\}$ можно получить только события, имеющие определенные вероятности. Таким образом, не нужно всякий раз обосновывать существование вероятности события, скажем, $A \cup B$ — подобно тому, как в обычной арифметике не нужно доказывать существование суммы или разности двух чисел.

Итак, вероятность $\mathrm{P}(A)$ определяется для всех A из некоторой σ-алгебры $\mathfrak{A}$. Функция множества $\mathrm{P}(A)$ должна удовлетворять следующим аксиомам:

1) $\mathrm{P}(A) \geqslant 0$, $A \in \mathfrak{A}$;
2) $\mathrm{P}(\Omega) = 1$;
3) если A_α, $\alpha \in I$ — непересекающиеся множества, где

* Напомним понятие счетного множества. Бесконечное множество элементов называется счетным, если его элементы можно пронумеровать целыми числами 1, 2, ... Множества рациональных чисел, а также конечномерных векторов с рациональными координатами счетны. Объединение конечного или счетного множества конечных или счетных множеств либо конечно, либо счетно.

I — конечное или счетное множество, то

$$P\left(\bigcup_{\alpha\in I} A_\alpha\right) = \sum_{\alpha\in I} P(A_\alpha)$$

(последнее свойство называется *расширенной аксиомой сложения*, или *аксиомой счетной аддитивности*);

4) функция P (A) определена для любых подмножеств множества нулевой вероятности: если для $A \subset \Omega$ определена P (A), причем P $(A) = 0$, а B — любое множество, входящее в A (хотя, возможно, и не принадлежащее $\mathfrak{A}$), тогда P $(B) = 0$ (сформулированное свойство называется *аксиомой полноты*).

Тройка $(\Omega, \mathfrak{A}, P)$, где Ω — произвольное множество, $\mathfrak{A}$ — σ-алгебра подмножеств Ω, P = P (A) — вероятность, определенная на $\mathfrak{A}$ и удовлетворяющая перечисленным выше условиям, называется *вероятностным пространством*.

Функция P (A) называется также *вероятностной мерой*. По терминологии, принятой в теории меры, вероятностная мера — это любая полная мера, нормированная условием P $(\Omega) = 1$. Установление соответствия между вероятностью и мерой позволяет использовать в теории вероятностей результаты, полученные в теории меры; в то же время сама теория меры обогатилась за счет глубокого развития теории вероятностей.

Событие $A = \Omega$ называется достоверным, событие $\varnothing$ — невозможным. Так как $\Omega = \Omega \cup \varnothing$, $\Omega\varnothing = \varnothing$, то

$$P(\Omega) = P(\Omega) + P(\varnothing), \text{ откуда } P(\varnothing) = 0.$$

При классическом определении вероятности аксиомы 1—4 легко проверяются.

В самом деле, P $(A) \geqslant 0$, P $(\Omega) = 1$. Возьмем любые несовместные события A_α, $\alpha \in I$, где I — конечное или счетное множество. Пусть событию A_α благоприятствуют элементарные события ω_i, $i \in I_\alpha$. Обозначим через m_α мощность (число элементов) множества I_α. Тогда можно сказать, что событию A_α благоприятствует m_α элементарных событий. Событию $A = \bigcup A_\alpha$ благоприятствуют те и только те элементарные события ω_i, для которых $i \in \bigcup_{\alpha\in I} I_\alpha$. Поскольку для различных α множества I_α не пересекаются, то в множество $\bigcup_{\alpha\in I} I_\alpha$ входит $\sum_{\alpha\in I} m_\alpha$ элементов. Следовательно, событию A благоприятствует $\sum_{\alpha\in I} m_\alpha$ элементарных событий.

Таким образом,

$$P(A_\alpha)=\frac{m_\alpha}{n},\quad \alpha\in I;\quad P\Bigl(\bigcup_{\alpha\in I}A_\alpha\Bigr)=\sum_{\alpha\in I}\frac{m_\alpha}{n}.$$

Сравнив эти формулы, приходим к соотношению

$$P\Bigl(\bigcup_{\alpha\in I}A_\alpha\Bigr)=\sum_{\alpha\in I}P(A_\alpha).$$

Аксиома 3 в данном случае является следствием классического определения вероятности.

Наконец, аксиома 4 выполняется тривиальным образом. В самом деле, в классической схеме $P(A)=0$ только в том случае, если событию A не благоприятствует ни одно из элементарных событий; иначе говоря, $A=\varnothing$. Если $B\subset A$, то, конечно, $B=\varnothing$, так что и $P(B)=0$.

Приведем пример вероятностного пространства, не сводящийся к классическому определению вероятности. Пусть $\omega=(\omega_1,\omega_2)$ — точки единичного квадрата Ω $(0\leqslant\omega_1\leqslant 1,$ $0\leqslant\omega_2\leqslant 1)$, $\mathfrak{A}$ — множество плоских фигур, входящих в этот квадрат и имеющих определенную площадь. Определим $P(A)$ как площадь фигуры A. Тогда, например, свойство $P(\bigcup_{\alpha\in I}A_\alpha)=\sum_{\alpha\in I}P(A_\alpha)$ для непересекающихся A_α легко проверяется, по крайней мере в том случае, когда I — конечное множество.

§ 5. ОСНОВНЫЕ СООТНОШЕНИЯ МЕЖДУ ВЕРОЯТНОСТЯМИ СОБЫТИЙ

Выведем из аксиом, приведенных в § 4, важнейшие соотношения между вероятностями событий, предполагая, что все рассматриваемые события обладают определенными вероятностями.

1. $P(\overline{A})=1-P(A)$.

Действительно, $A\cup\overline{A}=\Omega$, $A\overline{A}=\varnothing$, откуда

$$1=P(\Omega)=P(A\cup\overline{A})=P(A)+P(\overline{A}).$$

2. Если $A\subset B$, то $P(A)\leqslant P(B)$. В самом деле, $B=A\cup(B\setminus A)$, $A\cap(B\setminus A)=\varnothing$, так что $P(B)=P(A)+P(B\setminus A)$. Так как $P(B\setminus A)\geqslant 0$, то $P(B)\leqslant P(A)$.

3. $P(A)\leqslant 1$. Это неравенство следует из того, что $A\subset\Omega$, откуда $P(A)\leqslant P(\Omega)=1$.

4. Пусть A_n, $n = 1, 2, \ldots$, — любые множества из $\mathcal{A}$. Образуем множества

$$B_1 = A_1, \ B_2 = A_2 \setminus A_1, \ B_3 = A_3 \setminus (A_1 \cup A_2), \ldots,$$
$$\ldots, B_n = A_n \setminus (A_1 \cup \ldots \cup A_{n-1}), \ldots.$$

Тогда

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(B_n). \qquad (1.8)$$

Эта формула заменяет расширенную аксиому сложения в случае, если A_n — не обязательно непересекающиеся множества. Докажем формулу (1.8).

Прежде всего

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n.$$

Действительно, если $\omega \in \bigcup_{n=1}^{\infty} A_n$, то найдется такое $n \geqslant 1$, что $\omega \in A_n$, $\omega \bar{\in} A_i$, $i < n$. Тогда $\omega \in B_n = A_n \setminus \bigcup_{i<n} A_i$; следовательно, $\omega \in \bigcup_{n=1} B_n$. Пусть теперь $\omega \in \bigcup_{n=1}^{\infty} B_n$. Тогда найдется такое n, что $\omega \in B_n$. Так как $B_n \subset A_n$, то $\omega \in A_n$, откуда $\omega \in \bigcup_{n=1}^{\infty} A_n$. Для справедливости доказываемой формулы осталось удостовериться в том, что B_n, $n \geqslant 1$, — непересекающиеся множества. Пусть $m < n$, тогда

$$B_m = A_m \setminus (A_1 \cup \ldots \cup A_{m-1}) \subset A_m.$$

Следовательно, для любого $\omega \in B_m$ выполняется соотношение $\omega \in A_m$. В то же время $B_n = A_n \setminus (A_1 \cup \ldots \cup A_{n-1}) \subset A_n \setminus A_m \subset \Omega \setminus A_m = \bar{A}_m$. Следовательно, $\omega \bar{\in} B_n$. Таким образом, не существует ω, принадлежащих одновременно B_m и B_n, т. е. $B_m B_n = \varnothing$. Формула полностью обоснована. Так как всегда $B_n \subset A_n$, то $P(B_n) \leqslant P(A_n)$ и

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leqslant \sum_{n=1}^{\infty} P(A_n). \qquad (1.9)$$

Любое конечное или счетное множество I можно перенумеровать числами 1, 2, ... ; поэтому

$$P\left(\bigcup_{\alpha \in I} A_\alpha\right) \leqslant \sum_{\alpha \in I} P(A_\alpha),$$

каковы бы ни были события $A_\alpha \in \mathcal{A}$, $\alpha \in I$.

Напомним данное выше понятие равенства событий: $A = B$ в том и только том случае, если $A \subset B$ и $B \subset A$. Введем понятие эквивалентности событий.

События A и B называются *эквивалентными* ($A \sim B$), если $\mathrm{P}(A \setminus B) = 0$, $\mathrm{P}(B \setminus A) = 0$.

Представим события A и B плоскими фигурами (как показано на рис. 1.7).

Эквивалентность A и B обозначает, что вероятность события, соответствующего заштрихованной площади, равна нулю. Из рис. 1.7 непосредственно видно, что

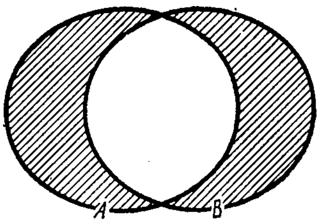

Рис. 1.7

$$\{A \sim B\} \Longleftrightarrow \{\mathrm{P}(A \cup B) = \mathrm{P}(AB)\}.$$

Если $A \sim B$, то $\mathrm{P}(A) = \mathrm{P}(B)$. Действительно, $A = AB \cup A\overline{B}$, причем $(AB) \cap (A\overline{B}) = \varnothing$. Отсюда $\mathrm{P}(A) = \mathrm{P}(AB) + \mathrm{P}(A\overline{B})$. Так как $A\overline{B} = A \setminus B$, то, по условию, $\mathrm{P}(A\overline{B}) = 0$. Следовательно, $\mathrm{P}(A) = \mathrm{P}(AB)$. Поменяв местами A и B, найдем, что $\mathrm{P}(B) = \mathrm{P}(AB)$. Итак, $\mathrm{P}(A) = \mathrm{P}(B)$. Если A — любое событие, B — событие с вероятностью 0, то

$$A \cup B \sim A, \quad A \setminus B \sim A.$$

Для доказательства достаточно заметить, что $(A \cup B) \setminus A \subset B$, $A \setminus (A \cup B) = \varnothing$, $(A \setminus B) \setminus A = \varnothing$, $A \setminus (A \setminus B) \subset B$.

Событие, эквивалентное Ω, называется *почти достоверным*. Вероятность почти достоверного события равна единице. Событие, эквивалентное $\varnothing$, называется *почти невозможным*. Вероятность такого события равна нулю. События A и B называются *почти несовместными*, если $\mathrm{P}(AB) = 0$. События $\{A_\alpha\}$, $\alpha \in I$, называются почти несовместными, если $\mathrm{P}(A_\alpha A_\beta) = 0$ при $\alpha \neq \beta$, $\alpha \in I$, $\beta \in I$.

Если $\{A_\alpha\}$, $\alpha \in I$, — почти несовместные события, где I — конечное или счетное множество, то

$$\mathrm{P}\Big(\bigcup_{\alpha \in I} A_\alpha\Big) = \sum_{\alpha \in I} \mathrm{P}(A_\alpha). \tag{1.10}$$

Для доказательства последней формулы введем событие

$$\Lambda = \bigcup_{\alpha \neq \beta} A_\alpha A_\beta.$$

Имеем

$$\mathrm{P}(\Lambda) \leqslant \sum_{\alpha \neq \beta} \mathrm{P}(A_\alpha A_\beta) = \sum_{\alpha \neq \beta} 0 = 0.$$

Введем события $B_\alpha = A_\alpha \setminus \Lambda$. Тогда, как легко видеть, B_α — непересекающиеся события. Следовательно,

$$\mathrm{P}\left(\bigcup_{\alpha \in I} B_\alpha\right) = \sum_{\alpha \in I} \mathrm{P}(B_\alpha).$$

В то же время

$$\bigcup_{\alpha \in I} A_\alpha = \left(\bigcup_{\alpha \in I} B_\alpha\right) \cup \Lambda,$$

откуда

$$\left(\bigcup_{\alpha \in I} B_\alpha\right) \setminus \left(\bigcup_{\alpha \in I} A_\alpha\right) = \varnothing,$$

$$\left(\bigcup_{\alpha \in I} A_\alpha\right) \setminus \left(\bigcup_{\alpha \in I} B_\alpha\right) \subset \Lambda,$$

так что

$$\bigcup_{\alpha \in I} A_\alpha \sim \bigcup_{\alpha \in I} B_\alpha.$$

Таким образом,

$$\mathrm{P}\left(\bigcup_{\alpha \in I} A_\alpha\right) = \mathrm{P}\left(\bigcup_{\alpha \in I} B_\alpha\right) = \sum_{\alpha \in I} \mathrm{P}(B_\alpha).$$

Остается заметить, что $B_\alpha \sim A_\alpha$, а значит, $\mathrm{P}(B_\alpha) = \mathrm{P}(A_\alpha)$. Окончательно имеем

$$\mathrm{P}\left(\bigcup_{\alpha \in I} A_\alpha\right) = \sum_{\alpha \in I} \mathrm{P}(A_\alpha).$$

В рассмотренном примере, где $\omega = (\omega_1, \omega_2)$, $0 \leqslant \omega_1 \leqslant 1$, $0 \leqslant \omega_2 \leqslant 1$ и вероятность численно равна площади фигуры, события $A \{\omega_1 + \omega_2 < 1\}$ и $B = \{\omega_1 + \omega_2 \leqslant 1\}$ эквивалентны. Действительно, $A \setminus B = \varnothing$, так что $\mathrm{P}(A \setminus B) = 0$; множество $B \setminus A$ представляет собой прямолинейный отрезок $\{\omega_1 + \omega_2 = 1,\ 0 \leqslant \omega_1 \leqslant 1\}$, имеющий нулевую площадь, откуда $\mathrm{P}(B \setminus A) = 0$.

В заключение параграфа приведем важную *аксиому непрерывности*, состоящую в следующем.

Пусть $\{A_n\}$, $n = 1, 2, \ldots,$ — убывающая последовательность событий (т. е. $A_{n+1} \subset A_n$, $n \geqslant 1$), причем $\bigcap\limits_{n=1}^{\infty} A_n = B$. Тогда

$$\lim_{n \to \infty} \mathrm{P}(A_n) = \mathrm{P}(B).$$

Аксиома непрерывности равносильна расширенной аксиоме сложения и выводится из нее следующим образом. Прежде всего заметим, что события $\Omega \setminus A_1$, $A_1 \setminus A_2$, $A_2 \setminus A_3$, ... несовместны и их объединение равно $\overline{B}$ (рис. 1.8). Отсюда

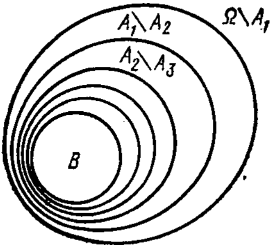

Рис. 1.8

$$P(\overline{B}) = P(\Omega \setminus A_1) + P(A_1 \setminus A_2) + P(A_2 \setminus A_3) + \dots .$$

Ряд, записанный в правой части равенства, сходится к $P(\overline{B})$, что можно записать так:

$$P(\overline{B}) = \lim_{n\to\infty} [P(\Omega \setminus A_1) + P(A_1 \setminus A_2) + \dots + P(A_{n-1} \setminus A_n)].$$

В то же время

$$(\Omega \setminus A_2) \cup (A_1 \setminus A_2) \cup \dots \cup (A_{n-1} \setminus A_n) = \overline{A}_n.$$

Следовательно,

$$P(\overline{B}) = \lim_{n\to\infty} P(\overline{A}_n).$$

Заметив, что

$$P(\overline{B}) = 1 - P(B),\ P(\overline{A}_n) = 1 - P(A_n),$$

приходим к требуемому равенству

$$P(B) = \lim_{n\to\infty} P(A_n).$$

Пусть теперь $\{A_n\}$ — возрастающая последовательность событий ($A_n \subset A_{n+1}$ для любого $n = 1, 2, \dots$) и $A = \bigcup_{n=1}^{\infty} A_n$. Тогда

$$P(A) = \lim_{n\to\infty} P(A_n).$$

Действительно, в данном случае

$$A = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \dots,$$

откуда

$$P(A) = P(A_1) + P(A_2 \setminus A_1) + \dots =$$
$$\lim_{n\to\infty} [P(A_1) + \dots + P(A_n \setminus A_1)] = \lim_{n\to\infty} P(A_n),$$

так как

$$A_n = A_1 \cup (A_2 \setminus A_1) \cup \dots \cup (A_n \setminus A_{n-1}).$$

Доказанную формулу также иногда называют аксиомой непрерывности.

Введем следующее определение.

Определение. *Множество событий* A_α, $\alpha \in I$, *где* I *— конечное или счетное множество, называется почти полной системой событий, если*

$$\mathrm{P}\left(\bigcup_{\alpha\in I} A_\alpha\right) = 1,$$

$$\mathrm{P}(A_\alpha A_\beta) = 0, \ \alpha \neq \beta, \ \alpha \in I, \ \beta \in I.$$

§ 6. НЕЗАВИСИМЫЕ СОБЫТИЯ

События A и B называются *статистически независимыми*, если

$$\mathrm{P}(AB) = \mathrm{P}(A)\,\mathrm{P}(B) \tag{1.11}$$

(обычно слово «статистически» опускается). Пусть, например, последовательно бросаются две монеты; A — выпадение герба при первом бросании, B — выпадение герба при втором бросании. Допустим, что $\mathrm{P}(A) = \mathrm{P}(B) = \frac{1}{2}$, $\mathrm{P}(AB) = \frac{1}{4}$. Тогда события A и B независимы.

Докажем следующее утверждение. *Если* A *и* B *— независимые события, то независимы также* $\bar{A}$ *и* B, A и $\bar{B}$, $\bar{A}$ и $\bar{B}$.

Все три свойства доказываются совершенно одинаково; поэтому приведем доказательство лишь первого из них. Имеем

$$B = AB \cup \bar{A}B, \ (AB)(\bar{A}B) = \varnothing,$$

откуда

$$\mathrm{P}(B) = \mathrm{P}(AB) + \mathrm{P}(\bar{A}B).$$

Значит,

$$\mathrm{P}(\bar{A}B) = \mathrm{P}(B) - \mathrm{P}(AB) = \mathrm{P}(B) - \mathrm{P}(A)\,\mathrm{P}(B) =$$

$$= \mathrm{P}(B)\,[1 - \mathrm{P}(A)] = \mathrm{P}(B)\,\mathrm{P}(\bar{A}).$$

Итак, $\bar{A}$ и B — независимые события.

Если A и B — не независимые события, то они называются *зависимыми*. Если A и B независимы, говорят также, что любое из этих событий независимо (не зависит) от другого. Независимость — свойство взаимное: не может быть, чтобы A было зависимо от B, а B — не зависело от A.

События A_α, $\alpha \in I$, где I — конечное или счетное множество, называются *независимыми* (в совокупности), если для любого конечного набора различных $\alpha_1, \alpha_2, \ldots, \alpha_n \in I$

$$P\left(A_{\alpha_1} A_{\alpha_2} \ldots A_{\alpha_n}\right) = P\left(A_{\alpha_1}\right) P\left(A_{\alpha_2}\right) \ldots P\left(A_{\alpha_n}\right).$$

События A_α, $\alpha \in I$, где I — любое множество, называются *попарно независимыми*, если при любых $\alpha_1 \neq \alpha_2$, $\alpha_1 \in I$, $\alpha_2 \in I$

$$P\left(A_{\alpha_1} A_{\alpha_2}\right) = P\left(A_{\alpha_1}\right) P\left(A_{\alpha_2}\right)$$

Справедливо следующее утверждение. Пусть A_α, $\alpha \in I$, где I — конечное или счетное множество, — независимые события. Если заменить некоторые из событий A_α противоположными им событиями, то полученные события также будут независимыми. (Так, пусть A_1, A_2, A_3, A_4 — независимые события. Тогда $A_1, \overline{A}_2, \overline{A}_3, A_4$ — независимые события.)

Чтобы доказать это утверждение, введем следующее символическое обозначение: если A — любое событие, то $A^1 = A$, $A^0 = \overline{A}$. Тогда требуемое утверждение можно перефразировать так. Пусть A_α, $\alpha \in I$ — независимые события, $\varepsilon_\alpha = 0$ или 1, $\alpha \in I$. Тогда события $A_\alpha^{\varepsilon_\alpha}$, $\alpha \in I$, независимы. Для доказательства достаточно установить, что

$$P\left(A_{\alpha_1}^{\varepsilon_{\alpha_1}} \ldots A_{\alpha_n}^{\varepsilon_{\alpha_n}}\right) = P\left(A_{\alpha_1}^{\varepsilon_{\alpha_1}}\right) \ldots P\left(A_{\alpha_n}^{\varepsilon_{\alpha_n}}\right)$$

для любых различных $\alpha_1, \ldots, \alpha_n$. При $n = 2$ данное утверждение уже доказано: достаточно лишь обозначить $A_{\alpha_1} = A$, $A_{\alpha_2} = B$. Предположим, что данное утверждение справедливо при некотором n, и докажем его при $n + 1$.

Фиксируем $\alpha_1, \ldots, \alpha_{n+1}$ и запишем для краткости $A_{\alpha_i}^{\varepsilon_{\alpha_i}} = A_i^{\varepsilon_i}$. Нужно показать, что для всех $\varepsilon_1, \ldots, \varepsilon_{n+1}$, равных нулю или единице,

$$P\left(A_1^{\varepsilon_1} \ldots A_{n+1}^{\varepsilon_{n+1}}\right) = P\left(A_1^{\varepsilon_1}\right) \ldots P\left(A_{n+1}^{\varepsilon_{n+1}}\right).$$

Обозначим через Γ множество векторов $(\varepsilon_1, \ldots, \varepsilon_{n+1})$, для которых данная формула выполняется. Пусть $(\varepsilon_1, \ldots, \varepsilon_{n+1}) \in \Gamma$. Покажем, что любой из индексов ε_i, равных единице, можно заменить нулем и при этом формула также будет справедливой. Пусть, например, $\varepsilon_{n+1} = 1$. Покажем, что можно взять $\varepsilon_{n+1} = 0$. Имеем

$$\left(A_1^{\varepsilon_1} \ldots A_n^{\varepsilon_n} A_{n+1}\right) \cup \left(A_1^{\varepsilon_1} \ldots A_n^{\varepsilon_n} \overline{A}_{n+1}\right) = A_1^{\varepsilon_1} \ldots A_n^{\varepsilon_n},$$

В то же время события, объединение которых здесь записано, очевидно, несовместны. Отсюда

$$P\left(A_1^{\varepsilon_1} \ldots A_n^{\varepsilon_n} \overline{A}_{n+1}\right) = P\left(A_1^{\varepsilon_1} \ldots A_n^{\varepsilon_n}\right) - P\left(A_1^{\varepsilon_1} \ldots A_n^{\varepsilon_n} A_{n+1}\right).$$

Использовав предположение индукции и то, что $(\varepsilon_1, \ldots, \varepsilon_n, 1) \in \Gamma$, найдем

$$P\left(A_1^{\varepsilon_1} \ldots A_n^{\varepsilon_n} \overline{A}_{n+1}\right) = P\left(A_1^{\varepsilon_1}\right) \ldots P\left(A_n^{\varepsilon_n}\right) -$$

$$- P\left(A_1^{\varepsilon_1}\right) \ldots P\left(A_n^{\varepsilon_n}\right) P\left(A_{n+1}\right) = P\left(A_1^{\varepsilon_1}\right) \ldots P\left(A_n^{\varepsilon_n}\right) [1 -$$

$$- P\left(A_{n+1}\right)] = P\left(A_1^{\varepsilon_1}\right) \ldots P\left(A_n^{\varepsilon_n}\right) P\left(\overline{A}_{n+1}\right).$$

Таким образом, $(\varepsilon_1, \ldots, \varepsilon_n, 0) \in \Gamma$. Далее, по условию, $P(A_1 \ldots A_{n+1}) = P(A_1) \ldots P(A_{n+1})$, т. е. $(1, 1, \ldots, 1) \in \Gamma$. Последовательно заменяя одну из единиц нулем, можно за конечное число шагов прийти от вектора $(1, 1, \ldots, 1)$ к любому вектору $(\varepsilon_1, \ldots, \varepsilon_{n+1})$ из нулей и единиц. Это и требовалось доказать.

Понятие независимости является одним из центральных в теории вероятностей. Для того чтобы показать, как его используют в практике, рассмотрим следующий пример.

Пример. На ЭВМ решается некоторая вычислительная задача. Нас интересует событие A, состоящее в правильности решения. Допустим, что справедливо соотношение $A = A_1A_2A_3A_4A_5$, где A_1 — отсутствие ошибок при программировании; A_2 — отсутствие ошибок в исходных данных; A_3 — правильный ввод информации в машину; A_4 — отсутствие сбоев при решении задачи; A_5 — отсутствие сбоев печатающего устройства при печатании результата.

Вероятность каждого из событий A_i можно оценить опытным путем. Приняв гипотезу в независимости этих событий, находим

$$P(A) = P(A_1) \ldots P(A_5).$$

Каким же образом обосновывается независимость реальных событий? Проверка формулы $P(A_{\alpha_1} A_{\alpha_2} \ldots) =$

$= P(A_{\alpha_1}) P(A_{\alpha_2}) \ldots$ в случае, подобном описанному, ничего не дает; если бы мы имели возможность вычислить $P(A_{\alpha_1} A_{\alpha_2} \ldots)$, то ничего большего и не требовалось бы.

Существует способ, основанный на гипотезе о физической независимости событий. Если какие-либо события обусловливаются пренебрежимо мало связанными физическими процессами, то их считают физически независимыми, а из физической независимости следует независимость статистическая. Это утверждение является эмпирическим фактом, но многовековой опыт человечества подтверждает возможность такой концепции и ее практическую пользу. Так, например, отсутствие ошибок при программировании не зависит от отсутствия сбоев ЭВМ при решении задачи: эти события — результат действия пренебрежимо мало связанных физических процессов.

Однако может быть и так, что события обусловливаются одним и тем же физическим процессом и тем не менее статистически независимы. В таких случаях статистическую независимость можно выявить с помощью анализа характера зависимости событий от исходного процесса.

Пусть, например, имеются две физические величины $x = \cos 2\pi\omega$, $y = \sin 2\pi\omega$, где ω — точка интервала $(0, 1)$, причем $P(\omega \in \Delta) = |\Delta|$ [Δ — любой интервал, входящий в $(0, 1)$, $|\Delta|$ — длина Δ]. ω описывает физический процесс, определяющий величины x и y. Рассмотрим события $A = \{x < 0\}$ и $B = \{y > 0\}$. Имеем:

$$P(A) = P(x<0) = P(\cos 2\pi\omega < 0) = P\left(\frac{\pi}{2} < 2\pi\omega < \frac{3\pi}{2}\right) =$$

$$= P\left(\frac{1}{4} < \omega < \frac{3}{4}\right) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2},$$

$$P(B) = P(y > 0) = P(\sin 2\pi\omega > 0) =$$

$$= P(0 < 2\pi\omega < \pi) = P\left(0 < \omega < \frac{1}{2}\right) = \frac{1}{2},$$

$$P(AB) = P(x < 0,\ y > 0) = P(\cos 2\pi\omega < 0,\ \sin 2\pi\omega > 0) =$$

$$= P\left(\frac{\pi}{2} < 2\pi\omega < \frac{3\pi}{2},\ 0 < 2\pi\omega < \pi\right) = P\left(\frac{\pi}{2} < 2\pi\omega < \pi\right) =$$

$$= P\left(\frac{1}{4} < \omega < \frac{1}{2}\right) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

Мы видим, что $P(AB) = P(A)P(B)$, т. е. A и B статистически независимы.

Статистическую, как и физическую независимость, не следует абсолютизировать: абсолютно независимых событий не существует. Обычно бывает так, что в одних задачах можно считать события независимыми, в других же играет роль даже незначительное отклонение от равенства $P(AB) = P(A)P(B)$.

Пусть имеется последовательность испытаний, занумерованных числами n, и для любого n задана почти полная система событий $\{A_{n\alpha}\}$, $\alpha \in I_\alpha$, которые могут произойти в результате n-го испытания (I_α — конечные или счетные множества). Если для любого набора $\{\alpha_n\}$ события $\{A_{n\alpha_n}\}$ независимы, то данная последовательность испытаний называется *последовательностью независимых испытаний.* Подчеркнем, что независимость испытаний определяется по отношению к системам событий $\{A_{n\alpha}\}$.

Пусть $(\Omega_1, \mathfrak{A}_1, P_1), \dots (\Omega_n, \mathfrak{A}_n, P_n)$ — вероятностные пространства. Если для любых $A_1 \in \mathfrak{A}_1, \dots, A_n \in \mathfrak{A}_n$ события $A_1, \dots, A_n$ независимы, элементарные события $\omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n$ называются *независимыми элементарными событиями.*

§ 7. УСЛОВНАЯ ВЕРОЯТНОСТЬ

Пусть A и B — случайные события, причем $P(B) > 0$. Тогда отношение $\frac{P(AB)}{P(B)}$ называется условной вероятностью события A при условии B и обозначается $P(A/B)$. Таким образом,

$$P(A/B) = \frac{P(AB)}{P(B)}. \qquad (1.12)$$

Умножив обе части этого равенства на $P(B)$, найдем

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

Так как левая часть этого равенства симметрична относительно A и B, то можно записать

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

Эта формула заменяет более простую формулу $P(AB) = P(A)P(B)$, справедливую для независимых событий A и B, и называется *правилом умножения вероятностей.*

Пусть теперь $A_1, A_2, \dots, A_n$ — произвольные события.

Обозначим $A_1A_2 \dots A_i = B_i$, $1 \leqslant i \leqslant n-1$, $A = A_1A_2 \dots A_n$ и предположим, что $P(B_i) > 0$, $1 \leqslant i \leqslant n-1$. Тогда

$$P(A) = P(B_{n-1}A_n) = P(B_{n-1})\,P(A_n/B_{n-1}).$$

В свою очередь

$$P(B_{n-1}) = P(B_{n-2}A_{n-1}) = P(B_{n-2})\,P(A_{n-1}/B_{n-2}).$$

так что

$$P(A) = P(B_{n-2})\,P(A_{n-1}/B_{n-2})\,P(A_n/B_{n-1}).$$

Применив то же преобразование к $P(B_{n-2})$, найдем

$$P(A) = P(B_{n-3})\,P(A_{n-2}/B_{n-3})\,P(A_{n-1}/B_{n-2})\,P(A_n/B_{n-1})$$

и т. д. Окончательно имеем

$$P(A) = P(A_1)\,P(A_2/B_1)\,P(A_3/B_2)\dots P(A_n/B_{n-1}),$$

или, что то же,

$$P(A) = P(A_1)\,P(A_2/A_1)\,P(A_3/A_1A_2)\dots P(A_n/A_1\dots A_{n-1}).$$

[Напомним, что если $A_1, A_2, \dots, A_n$ — независимые события, $A = A_1A_2 \dots A_n$, то $P(A) = P(A_1)\,P(A_2)\dots P(A_n)$.]

Выясним смысл условной вероятности. *Вероятность события A — это мера объективной возможности данного события при определенных условиях опыта.* Совокупность условий, определяющих опыт, обозначим γ (так, в схеме с подбрасыванием монеты совокупность γ определяется формой монеты, высотой подбрасывания и т. п.). Таким образом, вероятность события A можно записать так: $P(A) = P(A/\gamma)$, указывая этой записью зависимость вероятности от совокупности условий опыта.

Допустим, что при данных условиях γ произошло событие B. Тогда наступление события B можно считать дополнением к совокупности условий γ. Вероятность наступления события A при совокупности условий (γ, B) и есть условная вероятность $P(A/B) = P(A/\gamma, B)$.

Чтобы еще лучше разъяснить введенное понятие, обратимся к статистическому определению вероятности как пределу частоты при безгранично увеличивающемся числе наблюдений. Несколько перефразировав это определение и используя понятие последовательности независимых испытаний, сформулируем следующее утверждение. Пусть производятся независимые испытания, в каждом из которых выполняется совокупность условий γ. Обозначим через

$\bar{p}_n$ долю тех испытаний среди первых n, в которых произошло событие A. Тогда

$$\bar{p}_n \xrightarrow[n\to\infty]{} \mathrm{P}(A/\gamma).$$

Это определение естественно несколько расширить, а именно: допустим, что в каждом из испытаний совокупность условий γ может либо выполняться, либо не выполняться. Обозначим через $N = N(n)$ число испытаний среди первых n, в каждом из которых совокупность условий γ выполнена. Тогда

$$p = \mathrm{P}(A/\gamma) = \lim_{N\to\infty} \frac{m_N}{N} = \lim_{n\to\infty} \frac{m_{N(n)}}{N(n)},$$

где m_N — число испытаний, в которых произошло событие A, среди первых N испытаний, для которых выполнялась совокупность условий γ. [Предполагается, что $N(n) \to \infty$ при $n \to \infty$.]

Пусть теперь A и B — два события, причем в каждом испытании выполняется совокупность условий γ. Определим вероятность $\mathrm{P}(A/\gamma, B)$. Для этого обозначим через $N = N(n)$ число испытаний среди первых n, в которых произошло событие B. Пусть среди этих N испытаний в m_N произошло также событие A. Тогда, очевидно.

$$\mathrm{P}(A/\gamma,\ B) = \lim_{n\to\infty} \frac{m_{N(n)}}{N(n)},$$

если только $N(n) \xrightarrow[n\to\infty]{} \infty$.

Однако, по тому же статистическому определению вероятности,

$$\lim_{n\to\infty} \frac{N(n)}{n} = \mathrm{P}(B).$$

Предположив, что $\mathrm{P}(B) > 0$, найдем

$$\mathrm{P}(A/\gamma,\ B) = \lim_{n\to\infty} \left\{ \frac{m_{N(n)}}{n} \Big/ \frac{N(n)}{n} \right\} = \lim_{n\to\infty} \frac{m_{N(n)}}{n} \Big/ \mathrm{P}(B).$$

Здесь $m_{N(n)}$ — число тех испытаний среди первых n, в которых произошло и A, и B. Отсюда

$$\lim_{n\to\infty} \frac{m_{N(n)}}{n} = \mathrm{P}(AB).$$

Окончательно

$$P(A/\gamma, B) = \frac{P(AB)}{P(B)}.$$

Вероятности событий при фиксированной совокупности условий γ в отличие от условных вероятностей называются *безусловными* вероятностями. Как видно из вышеизложенного, разница между безусловными и условными вероятностями состоит лишь в различии совокупности условий.

Очевидно, безусловная вероятность есть частный случай условной, если условие B есть достоверное событие. Условная вероятность обладает теми же свойствами, что и безусловная:

1) $0 \leqslant P(A/B) \leqslant 1$, так как $0 \leqslant \frac{P(AB)}{P(B)} \leqslant 1$;

2) $P(A/B) \leqslant P(C/B)$, если $A \subset C$, так как $\frac{P(AB)}{P(B)} \leqslant \leqslant \frac{P(CB)}{P(B)}$;

3) $P(A/B) = 1$, если A — достоверное событие, так как в этом случае $AB = B$; $P(C/B) = 0$, если C — невозможное событие, так как в этом случае CB невозможно.

4) $P(A \cup C/B) = P(A/B) + P(C/B)$, если A и C несовместны, так как в таком случае AB и CB также несовместны и поэтому по правилу сложения для безусловных вероятностей имеем

$$P(A \cup C/B) = \frac{P((A \cup C)B)}{P(B)} = \frac{P(AB \cup CB)}{P(B)} =$$

$$= \frac{P(AB) + P(CB)}{P(B)} = P(A/B) + P(C/B).$$

5) $P(A \cup C/B) = P(A/B) + P(C/B) - P(AC/B)$ — проверяется так же, как 4.

6) $P(\bar{A}/B) = 1 - P(A/B)$, так как $\frac{P(\bar{A}B)}{P(B)} + \frac{P(AB)}{P(B)} =$

$$= \frac{P((A \cup \bar{A})B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Рассмотрим пример.

Пример. Какова вероятность того, что оба ребенка — мальчики (событие A), если известно, что в семье с двумя детьми есть мальчик (событие B)? Вероятность рождения мальчика считаем равной 1/2. Нас интересует вероятность $P(AB)=\frac{P(AB)}{P(B)}$. Заметим, что $A\subset B$, поэтому $AB=A$ и $P(AB)=P(A)$. Рассмотрим распределение детей по полу в семьях с двумя детьми. Оно может быть следующим:

1) первый ребенок мальчик, второй — мальчик;
2) первый — мальчик, второй — девочка;
3) первая — девочка, второй — мальчик;
4) первая — девочка, вторая — девочка.

Все эти четыре исхода равновозможны, каждый имеет вероятность 1/4. Событию A благоприятствует один исход, поэтому $P(A)=\frac{1}{4}$, событию B благоприятствуют три исхода, поэтому $P(B)=\frac{3}{4}$ и $P(A/B)=\frac{P(AB)}{P(B)}=\frac{P(A)}{P(B)}=\frac{1}{3}$.

§ 8. ФОРМУЛА ПОЛНОЙ ВЕРОЯТНОСТИ И ФОРМУЛА БАЙЕСА

Пусть $H_1, H_2, \ldots, H_n$ — полная система событий; будем называть эти события *гипотезами*. Пусть A — некоторое событие. Докажем, что

$$P(A)=\sum_{k=1}^{n} P(H_k)\cdot P(A/H_k), \qquad (1.13)$$

т. е. *вероятность события A есть сумма произведений вероятностей каждой гипотезы на условную вероятность события A при этой гипотезе.* (Приведенную формулу называют *формулой полной вероятности*.)

Доказательство. Событие A может произойти только одновременно с некоторым из событий $H_1, H_2, \ldots, H_n$, так как последние образуют полную систему событий. Поэтому событие A есть сумма: $A=AH_1\cup AH_2\cup\ldots\cup AH_n$. События $AH_1, AH_2, \ldots, AH_n$ попарно несовместны; так как гипотезы $H_1, H_2, \ldots, H_n$ попарно несовместны, то применима аксиома сложения вероятностей, по которой имеем $P(A)=\sum_{k=1}^{n} P(AH_k)$. Применив к вероятностям $P(AH_k)$ правило умножения, имеем

$$P(AH_k)=P(H_k)\,P(A/H_k),$$

откуда получаем

$$P(A) = \sum_{k=1}^{n} P(H_k) \cdot P(A/H_k), \text{ что и требовалось.}$$

Пример 1. В урне 5 белых и 4 черных шаров. Наудачу извлекают из урны один шар, а затем другой. Найти вероятность того, что во втором случае был вынут белый шар (шары в урну не возвращаются). Пусть A состоит в извлечении в первый раз черного шара, B — в извлечении во второй раз белого шара. Требуется найти $P(B)$. Рассмотрим полную систему событий A и $\bar{A}$, где $\bar{A}$ — событие, противоположное A, состоящее, очевидно, в извлечении в первый раз белого шара. Повторяя рассуждения, проведенные в § 6, получаем

$$P(A) = \frac{4}{9}, \quad P(\bar{A}) = 1 - P(A) = \frac{5}{9}, \quad P(B/A) = \frac{5}{8},$$

$$P(B/\bar{A}) = \frac{4}{8} = \frac{1}{2}.$$

По формуле полной вероятности имеем

$$P(B) = P(A) \cdot P(B/A) + P(\bar{A}) \cdot P(B/\bar{A}) = \frac{4}{9} \cdot \frac{5}{8} + \frac{5}{9} \cdot \frac{4}{8} = \frac{5}{9}.$$

Таким образом, вероятность вынуть белый шар при втором извлечении та же, что и при первом.

Пример 2. На фабрике машины a, b и c производят соответственно 25, 35 и 40% всех изделий. В их продукции брак составляет соответственно 5, 4 и 2%. Какова вероятность того, что случайно выбранное изделие, произведенное на фабрике, дефектно?

Пусть событие A состоит в том, что случайно выбранное изделие дефектно, и пусть H_1, H_2, H_3 — события, состоящие в том, что изделие произведено машинами a, b и c соответственно. Очевидно, события H_1, H_2, H_3 образуют полную систему событий. Числа 0,25; 0,35; 0,40 (25%, 35%, 40%) являются вероятностями этих событий, т. е. $P(H_1) = 0{,}25$; $P(H_2) = 0{,}35$; $P(H_3) = 0{,}40$. Аналогично, числа 0,05; 0,04; 0,02 (5%, 4% и 2%) являются условными вероятностями события A при выполнении гипотез H_1, H_2, H_3 соответственно, т. е.

$$P(A/H_1) = 0{,}05; \quad P(A/H_2) = 0{,}04; \quad P(A/H_3) = 0{,}02.$$

Применив формулу полной вероятности, найдем

$$P(A) = \sum_{k=1}^{3} P(H_k) \cdot P(A/H_k) = 0{,}25 \cdot 0{,}05 + 0{,}35 \cdot 0{,}04 + 0{,}40 \cdot 0{,}02 = 0{,}0345.$$

Пусть $H_1, H_2, \ldots, H_n$ — полная система событий (гипотез). Пусть A — некоторое событие, для которого $P(A) \neq 0$. Докажем, что

$$P(H_k/A) = \frac{P(H_k) \cdot P(A/H_k)}{\sum_{i=1}^{n} P(H_i) P(A/H_i)}. \qquad (1.14)$$

Эта формула носит название *формулы Байеса*. Докажем справедливость этой формулы.

Доказательство. Согласно определению условной вероятности имеем

$$P(H_k/A) = \frac{P(AH_k)}{P(A)}, \qquad (1.15)$$

но $P(AH_k) = P(H_k) \cdot P(A/H_k)$ (по правилу умножения вероятностей), а $P(A) = \sum_{i=1}^{n} P(H_i) P(A/H_i)$ (по формуле полной вероятности). Отсюда, учитывая формулу (1.12), имеем

$$P(H_k/A) = \frac{P(A/H_k) \cdot P(H_k)}{\sum_{i=1}^{n} P(H_i) P(A/H_i)},$$

что и требовалось доказать.

Пример 3. Пусть выполнены условия примера 2 настоящего параграфа и пусть известно, что случайно выбранное изделие оказалось дефектным. Какова вероятность того, что оно было произведено машинами a, b, c соответственно?

Пусть A, H_1, H_2, H_3 означают то же, что и в примере 2. Тогда задача состоит в нахождении условных вероятностей гипотез H_1, H_2, H_3 при условии, что событие A произошло, т. е. вероятностей $P(H_1/A)$, $P(H_2/A)$ и $P(H_3/A)$. В условиях задачи заданы вероятности: $P(H_1) = 0{,}25$; $P(H_2) = 0{,}35$;

$P(H_3) = 0{,}40$; $P(A/H_1) = 0{,}05$; $P(A/H_2) = 0{,}04$; $P(A/H_3) = 0{,}02$.

Применяя формулу Байеса, имеем

$$P(H_1/A) = \frac{P(H_1) \cdot P(A/H_1)}{\sum_{i=1}^{3} P(H_i) P(A/H_i)} =$$

$$= \frac{0{,}25 \cdot 0{,}05}{0{,}25 \cdot 0{,}05 + 0{,}35 \cdot 0{,}04 + 0{,}40 \cdot 0{,}02} = \frac{125}{345} = \frac{25}{69}.$$

Аналогично получаем

$$\mathrm{P}(H_2/A) = \frac{28}{69} \text{ и } \mathrm{P}(H_3/A) = \frac{16}{69}.$$

Упражнения к главе I

1. Пусть A, B, C — произвольные события. Найти выражения для событий, состоящих в том, что:
а) произошло только событие A;
б) произошли события A и B, но событие C не произошло;
в) произошло по крайней мере одно из этих событий;
г) произошло по крайней мере два события;
д) ни одного события не произошло;
е) произошло не больше двух событий.

2. Проверить следующие соотношения:

а) $\overline{A \cup B} = \overline{A}\,\overline{B}$;

б) $\overline{\overline{A}\,\overline{B}} = A \cup B$;

в) $\overline{A_1 \cup A_2 \cup \ldots \cup A_n} = \overline{A}_1\,\overline{A}_2 \ldots \overline{A}_n$;

г) $\overline{A} \cup \overline{B} = \overline{AB}$;

д) $AB = \overline{\overline{A} \cup \overline{B}}$;

е) $A_1A_2 \ldots A_n = \overline{\overline{A}_1 \cup \overline{A}_2 \ldots \cup \overline{A}_n}$;

ж) $\overline{A_1A_2 \ldots A_n} = \overline{A}_1 \cup \overline{A}_2 \ldots \cup \overline{A}_n$;

з) $(A \cup B)\,C = AC \cup BC$.

3. В партии n деталей. Событие A_i заключается в том, что i-я деталь дефектна. Записать событие, состоящее в том, что:
а) ни одна из деталей не имеет дефектов;
б) хотя бы одна деталь имеет дефект;
в) только бы одна деталь имеет дефект;
г) не более двух деталей дефектны;
д) по крайней мере два изделия имеют дефекты;
е) ровно два изделия дефектны.

4. Показать, что для любых двух событий A и B соотношения $A \subset B$, $\overline{A} \supset \overline{B}$, $A \cup B = B$, $A\overline{B} = \varnothing$ равносильны.

5. В лифт девятиэтажного дома на первом этаже вошли 5 человек. Предположим, что каждый из них независимо от других может выйти на любом из этажей, начиная со второго. Найти вероятность того, что все пятеро выйдут на разных этажах.

6. Бросают две игральные кости. Пусть A — событие, состоящее в том, что сумма выпавших очков нечетна, B — событие, состоящее в том, что хотя бы на одной из костей выпала единица. Найти вероятность событий AB, $A \cup B$, считая все 36 возможных исходов равновероятными.

7. Из последовательности чисел 1,2, ..., n наудачу выбирают сначала одно число, затем другое. Найти вероятность того, что одно из этих чисел окажется меньше k, а другое больше k, где $1 < k < n$ — произвольное целое число.

8. В лотерее n билетов, из них m выигрышных. Найти вероятность выигрыша для лица, имеющего k билетов.

9. В партии, состоящей из N изделий, имеется M бракованных. Наудачу выбраны $n \leqslant N$ изделий. Найти вероятность того, что среди них окажется ровно $m \leqslant M$ бракованных.

10. Точка брошена наудачу внутрь круга радиуса R. Вероятность попадания точки в любую область, расположенную внутри круга, пропорциональна площади этой области. Найти вероятность того, что точка находится от центра на расстоянии, меньшем r, где $r < R$.

11. Студент пришел на экзамен, зная лишь 20 из 25 вопросов программы, и получил два вопроса, наудачу выбранных из 25. Найти вероятность того, что студент знает оба этих вопроса: а) используя правило умножения; б) по классической формуле.

12. Известно, что 5% всех мужчин и 0,25% всех женщин дальтоники. Наугад выбранное лицо страдает дальтонизмом. Какова вероятность того, что это мужчина? (Считать, что мужчин и женщин одинаковое число.)

13. Бросают 3 игральные кости. Найти вероятность того, что хотя бы на одной из них выпадет одно очко, при условии, что на всех трех костях выпали разные грани.

14. События $A_1, A_2, \ldots, A_n$ независимы в совокупности, причем $\mathrm{P}(A_k) = p_k$. Найти вероятность:

а) появления хотя бы одного из этих событий;

б) непоявления всех этих событий;

в) появления ровно одного из них.

15. Из урны, содержащей 3 белых и 2 черных шара, переложены два вынутых наудачу шара в урну, содержащую 4 белых и 4 черных шара. Найти вероятность того, что шар, наудачу вынутый из второй урны, окажется белым.

16. Изделия некоторого производства удовлетворяют стандарту с вероятностью 0,96. Предлагается упрощенная система испытаний, дающая положительный результат с вероятностью 0,98 для изделий, удовлетворяющих стандарту, а для изделий, которые не удовлетворяют стандарту, с вероятностью 0,05. Какова вероятность того, что изделие, выдержавшее испытание, удовлетворяет стандарту?

17. Вероятность попадания в цель при одном выстреле равна p, а вероятность поражения цели при $k \geqslant 1$ попаданиях в нее равна $1 - q^k$. Найти вероятность того, что цель поражена, если было произведено n выстрелов.

18. Из урны, в которой было $m \geqslant 3$ белых и n черных шаров, потеряли один шар неизвестного цвета. Для того чтобы определить состав шаров в урне, из нее наудачу вынуты два шара. Найти вероятность того, что был потерян белый шар, если известно, что вынутые шары оказались белыми.

ГЛАВА II

ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§ 1. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

В гл. I было дано понятие о случайных событиях как исходах некоторого испытания. Представим себе, что производится испытание, в результате которого происходит одно из несовместных случайных событий A_i, где i принадлежит конечному или счетному множеству индексов I.

Таким образом,

$$A_iA_j = \varnothing, \quad i \neq j, \tag{2.1}$$

$$\bigcup_{i \in I} A_i = \Omega, \tag{2.2}$$

де $\varnothing$ — невозможное событие, Ω — достоверное событие. Пусть каждому исходу A_i испытания поставлено в соответствие некоторое действительное число x_i. В этом случае говорят, что задана дискретная случайная величина.

В гл. III будут изучены случайные величины общего вида. И в самом общем случае случайная величина — это функция исхода случайного испытания; термин «дискретная» показывает, что такая случайная величина может принимать лишь конечное или счетное множество значений,

Поскольку A_i, по условию, — случайные события, им соответствуют определенные вероятности $\mathrm{P}(A_i)$, $i \in I$. Из формул (2.1) и (2.2) в силу расширенной аксиомы сложения вероятностей следует, что

$$\sum_{i \in I} \mathrm{P}(A_i) = 1. \tag{2.3}$$

Обозначим случайную величину символом ξ. Очевидно, значение ξ зависит от исхода случайного испытания.

Основной интерес представляет вероятности, с которыми случайная величина принимает различные числовые значения.

Пусть x — произвольное число. Обозначим через $p_\xi(x)$ вероятность того, что $\xi = x$. Очевидно, событие $\{\xi = x\}$

равносильно тому, что в результате случайного испытания произошло случайное событие A_j, причем $x_j = x$. Таким образом,

$$p_\xi(x) = \sum_{j:x_j=x} \mathrm{P}(A_j) = \sum_{x_j=x} \mathrm{P}(A_j). \qquad (2.4)$$

Мы видим, что $p_\xi(x)$ может принимать положительные значения лишь при значениях x, принадлежащих конечному или счетному множеству $\{x_i\}_{i \in I}$; при всех остальных x $p_\xi(x) = 0$.

В дальнейшем будем предполагать, что все x_i попарно различны. Тогда, по формуле (2.4),

$$p_\xi(x_i) = \sum_{x_j=x_i} \mathrm{P}(A_j) = \mathrm{P}(A_i). \qquad (2.5)$$

О п р е д е л е н и е. *Законом распределения дискретной случайной величины называется совокупность пар чисел* (x_i, p_i), *где* x_i *— возможные значения случайной величины, а* p_i *— вероятности, с которыми она принимает эти значения. При этом*

$$\sum_{i \in I} p_i = 1. \qquad (2.6)$$

Заметим, что задание (x_i, p_i) при $i \in I$ равносильно заданию функции $p_\xi(x)$ при $-\infty < x < \infty$. Действительно, при $x = x_i$ $p_\xi(x) = p_i$, а при $x \notin \{x_i\}_{i \in I}$ $p_\xi(x) = 0$.

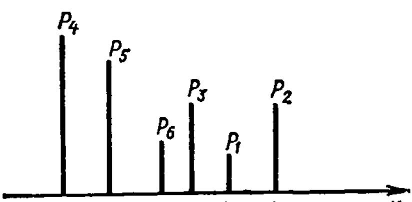

Рис. 2.1

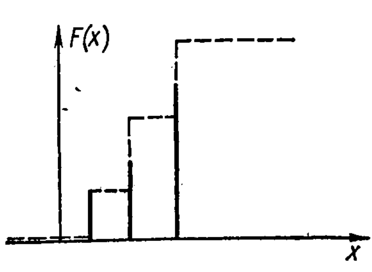

Рис. 2.2

Выше было сказано, что дискретная случайная величина есть функция исхода испытания. Однако дискретную случайную величину можно задавать и непосредственно законом распределения (x_i, p_i), $i \in I$, считая, что события A_i, появляющиеся в результате случайного испытания, тождественны соответствующим значениям случайной величины. Это можно объяснить так. Множество исходов ω испытания совпадает с множеством чисел x_i, $i \in I$. Если $\omega = x_i$, то будем считать, что происходит событие A_i. Закон распределения дискретной случайной величины можно изобразить графически (рис. 2.1). Нанесем на ось абсцисс прямоугольной системы координат точки x_i, $i \in I$, и от каждой из них отложим в положительном направлении оси ординат отрезок, по длине равный соответствующему значению p_i. Такой график оказывается полезным при подсчете различных вероятностей, связанных со случайной величиной. Пусть, например, требуется найти вероятность $F_\xi(x)$ того, что $\xi < x$, где x — некоторое заданное число. Тогда $F_\xi(x) = \mathrm{P}(\xi < x) = \sum\limits_{x_i < x} p_i$. На графике эта вероятность равна сумме длин вертикальных отрезков, расположенных левее прямой, параллельной оси ординат и проходящей через точку $(x, 0)$ (рис. 2.2).

Определение. *Функция $F_\xi(x)$ как функция переменной $x \in (-\infty, \infty)$ называется функцией распределения случайной величины ξ.*

Свойства функций распределения будут рассмотрены в главе, посвященной случайным величинам общего вида. Тем не менее для лучшего их понимания мы выведем их и в частном случае дискретных случайных величин.

1. $F_\xi(x)$ — неубывающая функция. Действительно, если $x < y$, то

$$F_\xi(y) = \sum_{x_i < y} p_i = \sum_{x_i < x} p_i + \sum_{x \leqslant x_i < y} p_i \geqslant \sum_{x_i < x} p_i = F_\xi(x),$$

так как слагаемое $\sum p_i$, очевидно, неотрицательно.

2. Функция $F_\xi(x)$ непрерывна слева, т. е. если $\{\alpha_n\}$ — любая последовательность положительных чисел, убывающая к нулю, то

$$\lim_{n\to\infty} F_\xi(x - \alpha_n) = F_\xi(x). \qquad (2.7)$$

Докажем это свойство.

Поскольку сумма ряда $\sum\limits_{i \in I} p_i$ равна единице, то по любому заданному $\varepsilon > 0$ можно выделить конечное число его членов (скажем, $\sum\limits_{i \in I_0} p_i$) такое, что $\sum\limits_{i \in I_0} p_i > 1 - \varepsilon$ и, следовательно,

$$\sum_{i \bar{\in} I_0} p_i < \varepsilon. \qquad (2.8)$$

Ввиду того, что множество I_0 содержит лишь конечное число элементов, для некоторого $\delta > 0$ в интервале $(x - \delta, x)$ вовсе не будет точек этого множества. Выберем теперь столь большое n, что $\alpha_n < \delta$. Тогда

$$F(x) = \sum_{x_i < x} p_i = \sum_{x_i < x - \alpha_n} p_i + \sum_{x - \alpha_n < x_i < x} p_i =$$
$$= F(x - \varepsilon_n) + \sum_{x - \alpha_n \leqslant x_i < x} p_i, \tag{2.9}$$

откуда

$$F(x) - F(x - \alpha_n) = \sum_{x - \alpha_n \leqslant x_i < x} p_i.$$

На основании предыдущего, из неравенства $x - \alpha_n \leqslant x_i < x$, вытекает, что $x - \delta < x_i < x$, т. е. $i \,\overline{\in}\, I_0$. Таким образом,

$$0 \leqslant \sum_{x - \alpha_n \leqslant x_i < x} p_i \leqslant \sum_{i \overline{\in} I_0} p_i < \varepsilon \tag{2.10}$$

на основании формулы (2.8). Отсюда же в силу формулы (2.9) находим, что

$$0 \leqslant F(x) - F(x - \alpha_n) < \varepsilon. \tag{2.11}$$

Справедливость свойства 2 доказана.

3. $\lim\limits_{x \to -\infty} F_\xi(x) = 0, \quad \lim\limits_{x \to \infty} F_\xi(x) = 1.$ (2.12)

Это свойство доказывается аналогично предыдущему. Пусть $\varepsilon > 0$ — фиксированное число; I_0 — конечное подмножество I, для которого $\sum\limits_{i \in I_0} p_i > 1 - \varepsilon$, т. е. $\sum\limits_{i \overline{\in} I_0} p_i < \varepsilon$.

Пусть $a = \min\limits_{i \in I_0} x_i, \quad b = \max\limits_{i \in I_0} x_i.$

Тогда, если $x < a$, то

$$F(x) = \sum_{x_i < x} p_i \leqslant \sum_{i \overline{\in} I_0} p_i, \tag{2.13}$$

а следовательно,

$$F(x) < \varepsilon. \tag{2.14}$$

Если же $x > b$, то (поскольку для всех $i \in I_0$ $x_i < b$, а значит, и $x_i < x$)

$$F(x) = \sum_{x_i < x} p_i \geqslant \sum_{i \in I_0} p_i. \tag{2.15}$$

Из неравенства (2.15) следует, что

$$F(x) > 1 - \varepsilon. \tag{2.16}$$

Неравенства (2.14) и (2.16) доказывают (2.12), что и требовалось.

§ 2. СХЕМА ИСПЫТАНИЙ БЕРНУЛЛИ

Предположим, что производится n независимых испытаний, каждое из которых может иметь два исхода: «успех» с вероятностью p и «неудачу» с вероятностью $q = 1 - p$. Такая схема называется *схемой Бернулли*. (Термины «успех» и «неудача» употребляются лишь в силу традиции. Важно лишь наличие двух различных исходов испытания.)

Обозначим через ν_n число успехов в n испытаниях. ν_n есть дискретная случайная величина, заданная на вероятностном пространстве. Чтобы убедиться в этом, положим $\omega_i = 1$, если в i-м испытании был успех, $\omega_i = 0$ в противном случае. Затем введем векторы $\omega = (\omega_1, \omega_2, \dots, \omega_n)$. Это и будут точки вероятностного пространства. Задание точки ω полностью определяет исходы всех n испытаний (и наоборот). Теперь можно записать

$$\nu_n = \sum_{i=1}^{n} \omega_i, \tag{2.17}$$

т. е. ν_n — функция исхода случайного испытания.

Найдем вероятность $\mathrm{P}(\omega) = \mathrm{P}(\omega_1, \omega_2, \dots, \omega_n)$ определенного исхода n независимых испытаний. Так как испытания независимы и вероятность успеха в каждом из них одна и та же, то

$$\mathrm{P}(\omega) = \mathrm{P}(\omega_1)\,\mathrm{P}(\omega_2)\dots\mathrm{P}(\omega_n), \tag{2.18}$$

где $\mathrm{P}(\omega_i)$ — вероятность исхода ω_i в i-м испытании $(1 \leqslant i \leqslant n)$.

Далее имеем

$$\mathrm{P}(1) = p, \quad \mathrm{P}(0) = q. \tag{2.19}$$

Таким образом, в правой части формулы (2.18) столько сомножителей, равных p, сколько среди ω_i единиц, и столько сомножителей, равных q, сколько среди ω_i нулей, т. е.

$$\mathrm{P}(\omega) = p^{\sum_{i=1}^{n}\omega_i}\, q^{\,n - \sum_{i=1}^{n}\omega_i}. \tag{2.20}$$

Найдем закон распределения случайной величины ν_n. Пусть k — любое целое число от 0 до n. Тогда

$$P(\nu_n = k) = P\left(\sum_{i=1}^{n} \omega_i = k\right) = \sum P(\omega_1) P(\omega_2) \ldots P(\omega_n),$$

$$\omega_1 + \omega_2 + \ldots + \omega_n = k. \tag{2.21}$$

В силу равенства (2.20) любое из слагаемых последней суммы равно $p^k q^{n-k}$. Всего же слагаемых C_n^k, ибо именно столько существует векторов $(\omega_1, \omega_2, \ldots, \omega_n)$, у каждого из которых k компонент равны 1, а $n-k$ равны 0 (любой такой вектор определяется местами, на которых расположены единицы, а эти места можно выбрать C_n^k различными способами).

Итак,

$$p_{\nu_n}(k) = P(\nu_n = k) = C_n^k p^k q^{n-k}. \tag{2.22}$$

Для любых $x \overline{\in} \{0, 1, 2, \ldots, n\}$ $p_{\nu_n}(x) = 0$. В этом можно убедиться и непосредственно на основании определения данной случайной величины, и на основании равенства

$$\sum_{k=0}^{n} p_{\nu_n}(k) = 1, \tag{2.23}$$

которое следует из формулы разложения бинома Ньютона:

$$1 = (p+q)^n = \sum_{k=0}^{n} C_n^k p^k q^{n-k}. \tag{2.24}$$

Используем формулу (2.22) для вычисления вероятности π_n того, что число успехов в n испытаниях по схеме Бернулли четно. Очевидно,

$$\pi_n = \sum_{k_{\text{чет}}} C_n^k p^k q^{n-k}, \tag{2.25}$$

где $k_{\text{чет}}$ означает, что суммирование производится по всем четным k от 0 до n. По формуле бинома Ньютона,

$$(q-p)^n = \sum_{k=0}^{n} C_n^k (-1)^k p^k q^{n-k} = \sum_{k_{\text{чет}}}^{k=0} C^k p^k q^{n-k} -$$

$$- \sum_{k_{\text{неч}}} C_n^k p^k q^{n-k}, \tag{2.26}$$

где $k_{\text{неч}}$ означает, что суммирование производится по всем нечетным k от 0 до n. Складывая почленно равенство (2.26) с равенством

$$1 = \sum_{k_{\text{чет}}} C_n^k p q^{n-k} + \sum_{k_{\text{неч}}} C_n^k p^k q^{n-k}, \tag{2.27}$$

следующим из формулы (2.24), найдем

$$2 \sum_{k_{\text{чет}}} C_n^k p^k q^{n-k} = 1 + (q - p)^n. \tag{2.28}$$

Учитывая формулу (2.25), окончательно находим

$$\pi_n = \frac{1}{2} + \frac{1}{2} (q - p)^n. \tag{2.29}$$

Пример. Предположим, что требуется реализовать устройство, вырабатывающее случайную величину ξ, которая бы удовлетворяла следующим условиям. Возможные значения ξ — 0 и 1, причем $\left| P(\xi = 0) - \frac{1}{2} \right| \leqslant 0{,}001$. Для этого можно использовать датчики, функционирующие независимо и вырабатывающие случайные величины ξ_i (i — номер датчика), где $P(\xi_i = 0) + P(\xi_i = 1) = 1$, $P(\xi_i = 1) = 0{,}4$. Взяв n датчиков, мы полагаем, что ξ есть сумма $\xi_1, \xi_2, \ldots, \xi_n$ по модулю два, т. е. $\xi = 0$, если $\xi_1 + \xi_2 + \ldots + \xi_n$ четно, и $\xi = 1$, если $\xi_1 + \xi_2 + \ldots + \xi_n$ нечетно. Требуется рассчитать минимальное значение n, при котором удовлетворяется заданное условие, касающееся $P(\xi = 0)$.

Имеем $q - p = 1 - 2p = 0{,}2$. Из формулы (2.29) находим

$$\left| P(\xi = 0) - \frac{1}{2} \right| = \frac{1}{2} (0,\ 2)^n. \tag{2.30}$$

Таким образом, должно выполняться неравенство $\frac{1}{2} \cdot 0{,}2^n \leqslant 0{,}001$ или, что то же, $5^n \geqslant 500$. Наименьшее целое n, удовлетворяющее этому неравенству, есть 4 ($5^4 = 625$). Таким образом, устройство с требуемым свойством можно реализовать с помощью четырех датчиков.

Приведем еще примеры непосредственного использования формулы (2.22).

1. Имеется некоторое устройство, состоящее из 5 элементов. В течение фиксированного времени каждый из элементов может выйти из строя с вероятностью $p = 0{,}1$. Устройство функционирует нормально, если число вышедших из строя элементов не более двух. Требуется найти вероятность нормального функционирования устройства.

Число вышедших из строя элементов есть число успехов в 5 независимых испытаниях с вероятностью успеха 0,1 в отдельном испытании. Интересующая нас вероятность

$$P(\nu_5 \leqslant 2) = P(\nu_5 = 0) + P(\nu_5 = 1) + P(\nu_5 = 2) =$$

$$= C_5^0 \cdot 0{,}1^0 \cdot 0{,}9^5 + C_5^1 \cdot 0{,}1 \cdot 0{,}9^4 + C_5^2 \cdot 0{,}1^2 \cdot 0{,}9^3 =$$

$$= 0{,}9^5 + 5 \cdot 0{,}1 \cdot 0{,}9^4 + 10 \cdot 0{,}1^2 \cdot 0{,}9^3 = 0{,}99144.$$

2. Устройство состоит из 100 элементов, каждый из которых независимо от прочих может отказать с вероятностью p. Какому условию должно удовлетворять p, чтобы число отказавших элементов было не более 1 с вероятностью, не меньшей 0,99?

Имеем

$$C_{100}^0 p^0 q^{100} + C_{100}^1 p q^{99} \geqslant 0{,}99,$$

или

$$q^{100} + 100 p q^{99} \geqslant 0{,}99.$$

Решив это уравнение приближенным методом, находим, что p должно быть приближенно не больше 0,0015.

3. Производится n выстрелов по некоторой цели. Вероятность поражения цели при каждом выстреле составляет 0,15. Каким должно быть n, чтобы вероятность поражения цели была не меньше 0,7?

Вероятность непоражения цели n выстрелами составляет $(1-0{,}15)^n = 0{,}85^n$. Нужно, чтобы $0{,}85^n \leqslant 0{,}3$. Этому условию удовлетворяют $n \geqslant 8$ и не удовлетворяют $n < 8$.

4. В партии изделий каждое из изделий независимо от других может оказаться дефектным с вероятностью p. Из партии берется выборка, состоящая из 15 изделий, и эти изделия проверяются на годность. Если число дефектных изделий в выборке не более 2, то партия принимается, в противном случае подвергается сплошному контролю. Какова вероятность того, что партия, для которой $p = 0{,}2$, будет принята?

Искомая вероятность есть вероятность не более 2 успехов в 15 испытаниях по схеме Бернулли с вероятностью успеха p в каждом испытании, т. е.

$$C_{15}^0 p^0 q^{15} + C_{15}^1 p q^{14} + C_{15}^2 p^2 q^{13} = 0{,}8^{15} + 15 \cdot 0{,}2 \cdot 0{,}8^{14} +$$

$$+ 105 \cdot 0{,}2^2 \cdot 0{,}8^{13} = 0{,}8^{13}(0{,}64 + 2{,}4 + 4{,}2) = 7{,}24 \cdot 0{,}8^{13} \approx 0{,}4.$$

§ 3. ПРЕДЕЛЬНОЕ ПОВЕДЕНИЕ ВЕРОЯТНОСТЕЙ $p_{\nu_n}(k)$ ПРИ БОЛЬШИХ n

Несмотря на элементарность формулы $p_{\nu_n}(k) = C_n^k p^k q^{n-k}$, при больших n непосредственное вычисление по ней связано с большой вычислительной работой, в особенности если требуется не просто вычислить $p_{\nu_n}(k)$ при конкретных значениях n и k, а решить какую-либо экстремальную задачу. Поэтому широкое применение нашли приближенные формулы, которые и будут приведены в настоящем параграфе.

Вначале проанализируем выражение C_n^k. Имеем.

$$C_n^k = \frac{n!}{k!\,(n-k)!}. \qquad (2.31)$$

Так как

$$C_n^{n-k} = C_n^k, \qquad (2.32)$$

то можно ограничиться случаем, когда $k \leqslant \frac{n}{2}$. Запишем следующую формулу:

$$C_n^k = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} = \frac{n^k}{k!}\left(1-\frac{1}{n}\right)\times$$

$$\times\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k-1}{n}\right). \qquad (2.33)$$

Отсюда

$$\frac{C_n^k}{n^k/k!} = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k-1}{n}\right). \qquad (2.34)$$

Если k остается ограниченным, а $n \to \infty$, то правая часть формулы (2.34) стремится к 1, т. е. в пределе можно вместо C_n^k использовать выражение $\frac{n^k}{k!}$

Выведем оценку погрешности такой аппроксимации для любых конечных n и k.

Вначале выведем следующую оценку. Если $0 \leqslant a_i \leqslant 1$, $1 \leqslant i \leqslant k$, $k \geqslant 1$, то

$$(1-a_1)(1-a_2)\ldots(1-a_k) \geqslant 1-(a_1+a_2+\ldots+a_k). \qquad (2.35)$$

При $k = 1$ соотношение (2.35) очевидно. Предположим, что это неравенство выполняется при некотором k, и докажем, что оно справедливо также при $k+1$. Имеем

$$(1-a_1)\ldots(1-a_k)(1-a_{k+1}) \geqslant [1-(a_1+\ldots+a_k)](1-$$

$$-a_{k+1}) = 1-(a_1+\ldots+a_k+a_{k+1})+(a_1+\ldots+a_k)a_{k+1} \geqslant$$

$$\geqslant 1-(a_1+\ldots+a_{k+1}),$$

что и требовалось доказать.

Применив неравенство (2.35) к произведению, записанному в правой части (2.34), найдем

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k-1}{n}\right) \geqslant 1-$$

$$-\left(\frac{1}{n}+\frac{2}{n}+\ldots+\frac{k-1}{n}\right) = 1-\frac{k(k-1)}{2n}. \qquad (2.36)$$

Очевидно также, что это произведение не больше 1. Итак, можно записать

$$1-\frac{k(k-1)}{2n}\leqslant\frac{C_n^k}{n^k/k!}\leqslant 1,\quad 0\leqslant k\leqslant n,\quad n\geqslant 1. \quad (2.37)$$

Отсюда можно вывести следующее практическое правило аппроксимации закона распределения числа успехов в схеме Бернулли.

Если $k\leqslant 0{,}14\sqrt{n}$, то имеет место приближенная формула

$$C_n^k p^k q^{n-k}\approx\frac{1}{k!}\left(\frac{np}{q}\right)^k q^n \quad (2.38)$$

с относительной погрешностью, меньшей 1%. Действительно, при $k\leqslant 0{,}14\sqrt{n}$ $k^2/n\leqslant 0{,}0196$ и тем более $k(k-1)/2n<0{,}01$. Из формулы (2.37) можно вывести важное следствие — предельную теорему Пуассона о вероятностях редких событий.

Теорема 2.1. *Если $n\to\infty$, $p\to 0$, $np\to a$, где a — фиксированное положительное число, то вероятность k успехов в схеме независимых испытаний с вероятностью p успеха в отдельном испытании стремится к $e^{-a}\frac{a^k}{k!}$ при любых $k=0, 1, 2, \ldots$.*

Доказательство. Имеем $p_{\nu_n}(k)=C_n^k p^k q^{n-k}$. Использовав оценку (2.37), получим

$$\left[1-\frac{k(k-1)}{2n}\right]\frac{n^k}{k!}p^k q^{n-k}\leqslant p_{\nu_n}(k)\leqslant\frac{n^k}{k!}p^k q^{n-k}. \quad (2.39)$$

Далее, при любом фиксированном k и $n\to\infty$

$$1-\frac{k(k-1)}{2n}\to 1, \quad (2.40)$$

$$(np)^k\to a^k, \quad (2.41)$$

$$q^{-k}=(1-p)^{-k}\to 1 \quad (2.42)$$

(так как $p\to 0$) и, наконец,

$$q^n=\left(1-\frac{np}{n}\right)^n\to e^{-a} \quad (2.43)$$

[из математического анализа известно, что если φ_n — любая последовательность, сходящаяся к числу a, то при

$n \to \infty$ $(1 + \frac{\varphi_n}{n})^n \to e^a$. В нашем случае $\varphi_n = -np$].

Подставив соотношения (2.40)—(2.43) в оценку (2.39), найдем, что

$$p_{\nu_n}(k) \underset{n\to\infty}{\longrightarrow} e^{-a} \frac{a^k}{k!}.$$

Это и требовалось доказать.

Дискретная случайная величина ξ с законом распределения

$$p_\xi(x) = e^{-a} \frac{a^x}{x!}, \quad x = 0, 1, 2, \ldots, \tag{2.44}$$

называется *распределенной по закону Пуассона*. Для того чтобы убедиться, что это действительно закон распределения, достаточно проверить, что сумма вероятностей значений 0, 1, 2, ... равна единице. Имеем

$$e^{-a} \sum_{x=0}^{\infty} \frac{a^x}{x!} = e^{-a}e^a = 1, \tag{2.45}$$

так как фигурирующий здесь ряд есть разложение экспоненциальной функции.

Приведем примеры применения закона Пуассона.

1. Система состоит из 10 000 элементов, каждый из которых в течение заданного времени может независимо от других отказать с вероятностью 0,00005. 1) Сколько нужно взять запасных элементов, чтобы все вышедшие из строя элементы заменить новыми с вероятностью, не меньшей 0,95? 2) Оценить снизу вероятность того, что ни один из запасных элементов не выйдет из строя, если вероятность отказа каждого из них такая же, как и у основных элементов.

Имеем: $n = 10\,000$, $p = 0{,}00005$, $np = 0{,}5$.

1) Если ν_n — число отказавших элементов, то приближенно будем считать, что ν_n распределено по закону Пуассона. Тогда

$$P(\nu_n = 0) = e^{-0,5}, \qquad P(\nu_n = 1) = e^{-0,5} \cdot 0{,}5,$$

$$P(\nu_n = 2) = e^{-0,5} \frac{0{,}5^2}{2}, \ldots, P(\nu_n = m) = e^{-0,5} \frac{0{,}5^m}{m!}.$$

Если m — число запасных элементов, то m следует выбрать из условия

$$P(\nu_n = 0) + P(\nu_n = 1) + \ldots + P(\nu_n = m-1) <$$
$$< 0{,}95 \leqslant P(\nu_n = 0) + P(\nu_n = 1) + \ldots + P(\nu_n = m).$$

С помощью таблицы находим, что этому условию удовлетворяет $m = 2$.

2) Вероятность того, что ни один из запасных элементов не выйдет из строя в течение заданного времени, равна

$$(1-0{,}0001)^m=(1-0{,}0001)^2>0{,}9998.$$

2. Телефонная станция обслуживает 1000 абонентов. В данном интервале времени любой абонент независимо от остальных может сделать вызов с вероятностью 0,005. Требуется найти вероятность того, что в данном интервале было не более 7 вызовов.

Имеем $np=5$,

$$P(\nu_n\leqslant 7)\approx e^{-5}\left(1+5+\frac{5^2}{2}+\frac{5^3}{6}+\frac{5^4}{24}+\frac{5^5}{120}+\right.$$
$$\left.+\frac{5^6}{720}+\frac{5^7}{5040}\right)\approx 0{,}867.$$

Таким образом, теорема Пуассона и оценка (2.39) позволяют приближенно оценивать $p_{\nu_n}(k)=C_n^kp^kq^{n-k}$ при больших n и относительно малых k (в том смысле, что отношение k^2/n должно быть малым).

Решим аналогичную задачу для случая, когда $k\to\infty$ произвольным образом. Как было отмечено, в силу равенства $C_n^{n-k}=C_n^k$ можно ограничиться случаем, когда $k\leqslant\frac{n}{2}$).

Обозначим

$$\alpha=\frac{k}{n},\quad \beta=\frac{n-k}{n}=1-\alpha \tag{2.46}$$

и используем известную в анализе формулу Стирлинга

$$n!=\sqrt{2\pi}\,n^{n+\frac{1}{2}}\,e^{-n+\frac{\theta_n}{12n}}, \tag{2.47}$$

где $0<\theta_n<1$, $n=1, 2, 3, \ldots$. (Следует отметить, что с помощью этой формулы можно оценивать вероятностные характеристики во многих практических задачах, для которых нет готовых предельных теорем.)

Использовав формулу (2.47), находим

$$C_n^k=C_n^{\alpha n}=$$

$$=\frac{\sqrt{2\pi}\,n^ne^{-n+\frac{\theta_n}{12n}}}{\sqrt{2\pi}\,(\alpha n)^{\alpha n+\frac{1}{2}}e^{-\alpha n+\frac{\theta_k}{12k}}\sqrt{2\pi}\,(\beta n)^{\beta n+\frac{1}{2}}e^{-\beta n+\frac{\theta_{n-k}}{12(n-k)}}}=$$

$$=\frac{1}{\sqrt{2\pi n\alpha\beta}}\left(\frac{1}{\alpha^{\alpha}\beta^{\beta}}\right)^{n} e^{\frac{\theta_n}{12n}-\frac{\theta_k}{12k}-\frac{\theta_{n-k}}{12(n-k)}}. \qquad (2.48)$$

Таким образом,

$$C_n^k p^k q^{n-k} = C_n^k p^{\alpha n} q^{\beta n} =$$

$$=\frac{1}{\sqrt{2\pi n\alpha\beta}}\left(\frac{p^{\alpha} q^{\beta}}{\alpha^{\alpha}\beta^{\alpha}}\right)^{n} e^{\frac{\theta_n}{12n}-\frac{\theta_k}{12k}-\frac{\theta_{n-k}}{12(n-k)}}. \qquad (2.49)$$

Заменив единицей сомножитель $\exp\left\{\frac{1}{12}\left(\frac{\theta_n}{n}-\frac{\theta_k}{k}-\frac{\theta_{n-k}}{n-k}\right)\right\}$, приходим к приближенной формуле

$$C_n^k p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi n\alpha\beta}}\left(\frac{p^{\alpha} q^{\beta}}{\alpha^{\alpha}\beta^{\beta}}\right)^{n}. \qquad (2.50)$$

Если $\frac{1}{k}+\frac{1}{n-k}\leqslant 0,1$, то относительная погрешность формулы (2.50) меньше 1%.

Действительно, отношение левой части формулы (2.50) к правой ее части равно

$$\exp\left\{\frac{1}{12}\left(\frac{\theta_n}{n}-\frac{\theta_k}{k}-\frac{\theta_{n-k}}{n-k}\right)\right\}.$$

Так как $0\leqslant\theta_n\leqslant 1$ при любом $n\geqslant 1$, то

$$-\left(\frac{1}{k}+\frac{1}{n-k}\right)\leqslant\frac{\theta_n}{n}-\frac{\theta_k}{k}-\frac{\theta_{n-k}}{n-k}\leqslant\frac{1}{n}.$$

Если $\frac{1}{k}+\frac{1}{n-k}\leqslant 0,1$, то и $\frac{1}{n}\leqslant 0,1$, а значит,

$$\exp\left\{-\frac{1}{120}\right\}\leqslant\exp\left\{\frac{1}{12}\left(\frac{\theta_n}{n}-\frac{\theta_k}{k}-\frac{\theta_{n-k}}{n-k}\right)\right\}\leqslant\exp\left\{\frac{1}{120}\right\} \qquad (2.51)$$

ввиду монотонности экспоненциальной функции.
По таблице находим

$$\exp\left\{-\frac{1}{120}\right\}\approx 0,992, \quad \exp\left\{\frac{1}{120}\right\}\approx 1,008,$$

т. е. относительная погрешность действительно меньше 1%.

Таким образом, вероятности $p_{\nu_n}(k)$ могут быть оценены либо по формуле (2.38) с оценкой погрешности (2.37), либо по формуле (2.50) с оценкой погрешности (2.51).

§ 4. ЛОКАЛЬНАЯ ТЕОРЕМА МУАВРА-ЛАПЛАСА

Рассмотрим формулу (2.49), считая, что p и q — фиксированные числа, а $n \to \infty$. Выражение $\frac{p^\alpha q^\beta}{\alpha^\alpha \beta^\beta}$ всегда меньше 1 при $0 < \alpha < 1$, $\alpha + \beta = 1$ и равно 1 только при $\alpha = p$, $\beta = q$. Действительно, логарифм этого выражения равен $\alpha(\ln p - \ln \alpha) + (1 - \alpha)[\ln q - \ln(1 - \alpha)]$. Продифференцировав по α, получим

$$\ln p - \ln \alpha - \ln q + \ln(1 - \alpha) = \ln \frac{p(1-\alpha)}{q\alpha}.$$

Производная положительна при $\frac{p(1-\alpha)}{q\alpha} > 1$, т. е. при $p(1 - \alpha) > q\alpha = (1 - p)\alpha$. Приведя подобные члены, получим $\alpha < p$; аналогично производная отрицательна при $\alpha > p$. Таким образом, в точке $p = \alpha$ функция $\frac{p^\alpha q^\beta}{\alpha^\alpha \beta^\beta}$ имеет минимум, равный 1. Если предположить, что $k = pn$ — целое число, то из формулы (2.49) имеем

$$p_{\nu_n}(pn) = \frac{1}{\sqrt{2\pi npq}}(1 + o(1)) \quad (n \to \infty), \tag{2.52}$$

т. е. вероятность значения $\nu_n = np$ при $n \to \infty$ имеет порядок убывания $\frac{1}{\sqrt{n}}$. Если же α фиксировано и не равно p, то, как было установлено, $\frac{p^\alpha q^\beta}{\alpha^\alpha \beta^\beta} < 1$, откуда следует, что $p_{\nu_n}(k)$ при $n \to \infty$ убывает к нулю со скоростью геометрической прогрессии. Однако сумма всех $p_{\nu_n}(k)$ равна единице. Естественно ожидать, что в окрестности $k = np$ существует область значений k, для которых $p_{\nu_n}(k)$ имеют порядок $\frac{1}{\sqrt{n}}$. Это предположение подтверждается следующей локальной предельной теоремой Муавра — Лапласа.

Теорема 2.2. *Обозначим* $x = \frac{k - np}{\sqrt{npq}}$. *Если* $n \to \infty$, *p фиксировано и не равно ни* 0, *ни* 1, *а k изменяется таким* *образом, что* $|x| \leqslant T$, *где* T — *фиксированное число, то равномерно по* x *в этом отрезке*

$$p_{\nu_n}(k) : \left(\frac{1}{\sqrt{2\pi npq}} e^{-\frac{x^2}{2}}\right) \xrightarrow[n \to \infty]{} 1. \qquad (2.53)$$

Это означает следующее. Для любого T и любого $\varepsilon > 0$ можно указать такое $N = N(T, \varepsilon)$, что при $n \geqslant N$ отношение $p_{\nu_n}(k)$ к $\frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{x^2}{2}\right\}$ больше $1 - \varepsilon$ и меньше $1 + \varepsilon$, если только $|x| \leqslant T$.

Доказательство. По условию, $n \to \infty$, $k = np + x\sqrt{npq} \geqslant np - T\sqrt{npq} \to \infty$, $n - k = nq - x\sqrt{npq} \geqslant nq - T\sqrt{npq} \to \infty$; следовательно, относительная погрешность формулы (2.50) стремится к 0 равномерно по x, $|x| \leqslant T$.

Далее, $\alpha = \frac{k}{n} = p + x\sqrt{\frac{pq}{n}}$ $\beta = 1 - \alpha = q - x\sqrt{\frac{pq}{n}}$. Таким образом, $\alpha \to p$, $\beta \to q$ равномерно по x, $|x| \leqslant T$, т. е. сомножитель $\frac{1}{\sqrt{2\pi n\alpha\beta}}$ в правой части формулы (2.50) может быть в пределе заменен выражением $\frac{1}{\sqrt{2\pi npq}}$. Второй сомножитель запишем в виде

$$\left(\frac{p^\alpha q^\beta}{\alpha^\alpha \beta^\beta}\right)^n = \exp\left\{-n\left(\alpha \ln \frac{\alpha}{p} + \beta \ln \frac{\beta}{q}\right)\right\}. \qquad (2.54)$$

Исследуем выражения $\ln\frac{\alpha}{p}$ и $\ln\frac{\beta}{q}$, воспользовавшись разложением логарифмической функции по формуле Тейлора:

$$\ln(1 + z) = z - \frac{z^2}{2} + O(z^3) \quad (z \to 0), \qquad (2.55)$$

$$\ln\frac{\alpha}{p} = \ln\frac{p + x\sqrt{\frac{pq}{n}}}{p} = \ln\left(1 + x\sqrt{\frac{q}{np}}\right) =$$

$$= x\sqrt{\frac{q}{np}} - \frac{x^2 q}{2np} + O\left(\frac{1}{n\sqrt{n}}\right) \qquad (2.56)$$

$$\ln\frac{\beta}{q}=\ln\frac{q-x\sqrt{\frac{pq}{n}}}{q}=\ln\left(1-x\sqrt{\frac{p}{nq}}\right)=$$

$$=-x\sqrt{\frac{q}{nq}}-\frac{x^2p}{2nq}+O\left(\frac{1}{n\sqrt{n}}\right). \qquad (2.57)$$

Причем заметим, что в формулах (2.56) и (2.57) выражение $O\left(\frac{1}{n\sqrt{n}}\right)$ равномерно по x, коль скоро $|x|\leqslant T$. Подставив (2.56) и (2.57) в формулу (2.54), найдем

$$\left(\frac{p^{\alpha}q^{\beta}}{\alpha^{\alpha}\beta^{\beta}}\right)=\exp\left\{-\left[\left(p+x\sqrt{\frac{pq}{n}}\right)\left(x\sqrt{\frac{q}{np}}-\frac{x^2q}{2np}\right)-\right.\right.$$

$$\left.\left.-\left(q-x\sqrt{\frac{pq}{n}}\right)\left(x\sqrt{\frac{p}{nq}}+\frac{x^2p}{2nq}\right)+O\left(\frac{1}{n\sqrt{n}}\right)\right]\right\}.$$

После приведения подобных членов правая часть этой формулы принимает вид

$$\left(\frac{p^{\alpha}q^{\beta}}{\alpha^{\alpha}\beta^{\beta}}\right)^{n}=\exp\left\{-\frac{x^2}{2}+O\left(\frac{1}{\sqrt{n}}\right)\right\}. \qquad (2.58)$$

Оба сомножителя правой части формулы (2.50) исследованы. Окончательно можно записать

$$p_{\nu_n}(k)\sim\frac{1}{\sqrt{2\pi npq}}\,e^{-\frac{x^2}{2}} \qquad (2.59)$$

с относительной погрешностью, равномерно стремящейся к 0 при $n\to\infty$, $|x|\leqslant T$.

На рис. 2.3 изображена кривая $y=\frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}$ (кри-

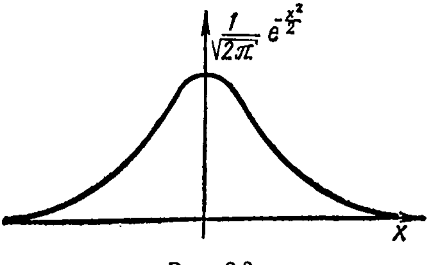

Рис. 2.3

вая Гаусса). Увеличенные в $\sqrt{npq}$ раз значения вероятностей $p_{\nu_n}(k)$ при $n \to \infty$ приближаются к соответствующим ординатам кривой Гаусса.

§ 5. ИНТЕГРАЛЬНАЯ ТЕОРЕМА ЛАПЛАСА

Локальная теорема Муавра — Лапласа позволяет оценивать отдельные значения $p_{\nu_n}(k)$, т. е. локальное поведение $p_{\nu_n}(k)$ как функции k, при больших n. Рассматриваемая в настоящем параграфе интегральная предельная теорема Лапласа позволяет оценивать функцию распределения случайной величины ν_n.

Обозначим через $F_n(x)$ функцию распределения случайной величины $\dfrac{\nu_n - np}{\sqrt{npq}}$; обозначим также

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}}\, dz, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{2.60}$$

Теорема 2.3. При $n \to \infty$ *равномерно по* x $(-\infty < x < \infty)$

$$F_n(x) \to \Phi(x). \tag{2.61}$$

Смысл теоремы 2.3 состоит в следующем. Для любого $\varepsilon > 0$ найдется такое $N = N(\varepsilon)$, что при $n \geqslant N$ будет выполняться неравенство

$$|F_n(x) - \Phi(x)| < \varepsilon, \quad -\infty < x < \infty. \tag{2.62}$$

Доказательству теоремы 2.3 предпошлем такое ее следствие.

Следствие. При $n \to \infty$ равномерно по a и b

$$P\left(a \leqslant \frac{\nu_n - np}{\sqrt{npq}} < b\right) \to \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{z^2}{2}}\, dz. \tag{2.63}$$

Докажем, что соотношение (2.63) действительно следует из соотношения (2.62). Для этого заметим, что

$$P\left(a \leqslant \frac{\nu_n - np}{\sqrt{npq}} < b\right) = P\left(\frac{\nu_n - np}{\sqrt{npq}} < b\right) -$$

$$-\mathrm{P}\left(\frac{\nu_n-np}{\sqrt{npq}}<a\right). \qquad (2.64)$$

Уменьшаемое в правой части (2.64), по теореме 2.3, равномерно сходится к $\Phi(b)$, а вычитаемое — к $\Phi(a)$. Следовательно, разность равномерно сходится к $\Phi(b)-\Phi(a)=$

$=\frac{1}{\sqrt{2\pi}}\int\limits_a^b e^{-\frac{z^2}{2}}dz$, что и требовалось доказать.

Докажем следующую лемму.

Лемма 1. Справедливость теоремы 2.3 следует из соотношения (2.63), если оно выполняется равномерно относительно a и b, $-T\leqslant a<b\leqslant T$, где T — любое фиксированное число.

Доказательство. Вначале докажем равенство

$$\int\limits_{-\infty}^{\infty}\varphi(z)\,dz=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}e^{-\frac{z^2}{2}}dz=1. \qquad (2.65)$$

Для этого заметим, что

$$\left(\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}e^{-\frac{z^2}{2}}dz\right)^2=\frac{1}{2\pi}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}e^{-\frac{x^2}{2}-\frac{y^2}{2}}dx\,dy. \qquad (2.66)$$

В двойном интеграле сделаем замену переменных:

$$x=\rho\cos\varphi, \quad y=\rho\sin\varphi.$$

Тогда $x^2+y^2=\rho^2$, и данный интеграл сводится к следующему:

$$\frac{1}{2\pi}\int\limits_0^{\infty}\rho e^{-\frac{\rho^2}{2}}d\rho\int\limits_0^{2\pi}d\varphi=\frac{1}{2\pi}\int\limits_0^{\infty}\rho e^{-\frac{\rho^2}{2}}d\rho\, 2\pi=1, \qquad (2.67)$$

так что и исходный интеграл равен 1.

Фиксируем некоторое $\varepsilon>0$ и найдем такое T, что

$$\int\limits_{-\infty}^{-T}\varphi(z)\,dz=\int\limits_{T}^{\infty}\varphi(z)\,dz=\varepsilon, \quad \int\limits_{-T}^{T}\varphi(z)\,dz=1-2\varepsilon. \qquad (2.68)$$

Положим $a = -T$, $b = T$. Тогда из соотношения (2.63) следует, что при достаточно большом n

$$P\left(-T \leqslant \frac{\nu_n - np}{\sqrt{npq}} < T\right) \geqslant \int_{-T}^{T} \varphi(z)\,dz - \varepsilon = 1 - 3\varepsilon. \quad (2.69)$$

Это означает, что при достаточно большом n

$$P\left(\frac{\nu_n - np}{\sqrt{npq}} \overline{\in} [-T, T)\right) \leqslant 3\varepsilon. \quad (2.70)$$

Возьмем теперь произвольное x. Тогда возможны три случая: 1) $x \leqslant -T$; 2) $x \geqslant T$; 3) $-T < x < T$. В случае 1) при достаточно большом n

$$F_n(x) \leqslant F_n(-T) \leqslant P\left(\frac{\nu_n - np}{\sqrt{npq}} \overline{\in} [-T, T)\right) \leqslant 3\varepsilon \quad (2.71)$$

вследствие неравенства (2.70) и

$$\Phi(x) \leqslant \Phi(-T) = \varepsilon$$

вследствие формулы (2.68). Так как $F_n(x) \geqslant 0$, $\Phi(x) > 0$, то

$$|F_n(x) - \Phi(x)| \leqslant 3\varepsilon, \quad x < -T. \quad (2.72)$$

В случае 2) при достаточно большом n

$$1 - F_n(x) \leqslant 1 - F_n(T) = P\left(\frac{\nu_n - np}{\sqrt{npq}} \geqslant T\right) \leqslant$$

$$\leqslant P\left(\frac{\nu_n - np}{\sqrt{npq}} \overline{\in} [-T, T)\right) \leqslant 3\varepsilon$$

и

$$1 - \Phi(x) = \int_{-\infty}^{\infty} \varphi(z)\,dz - \int_{-\infty}^{x} \varphi(z)\,dz =$$

$$= \int_{x}^{\infty} \varphi(z)\,dz \leqslant \int_{T}^{\infty} \varphi(z)\,dz = \varepsilon. \quad (2.73)$$

Так как $1 - F_n(x) \geqslant 0$, $\Phi(x) \geqslant 0$, то

$$|F_n(x) - \Phi(x)| = |[1 - F_n(x)] - [1 - \Phi(x)]| \leqslant 3\varepsilon. \quad (2.74)$$

В случае 3) при достаточно большом n

$$|F_n(x)-\Phi(x)| = |[F_n(x)-F_n(-T)]-[\Phi(x)-\Phi(-T)]+$$
$$+[F_n(-T)-\Phi(-T)]| = \left|P\left(-T \leqslant \frac{\nu_n-np}{\sqrt{npq}} < x\right)-\right.$$
$$\left.-\int_{-T}^{x}\varphi(z)\,dz+[F_n(-T)-\Phi(-T)]\right| \leqslant$$
$$\leqslant \left|P\left(-T \leqslant \frac{\nu_n-np}{\sqrt{npq}} < x\right)-\int_{-T}^{x}\varphi(z)\,dz\right|+$$
$$+F_n(-T)+\Phi(-T). \tag{2.75}$$

При достаточно большом n первое слагаемое будет меньше ε вследствие соотношения (2.63) (достаточно положить $a=-T$, $b=x$). Второе слагаемое не больше $P\left(\frac{\nu_n-np}{\sqrt{npq}}\,\bar{\in}\,[-T,\,T]\right)$, а следовательно, при достаточно большом n не больше 3ε по формуле 2.69. Наконец, третье слагаемое равно ε. Итак, при достаточно большом n

$$|F_n(x)-\Phi(x)| \leqslant \varepsilon+3\varepsilon+\varepsilon=5\varepsilon. \tag{2.76}$$

Объединив оценки (2.72), (2.74) и (2.76), найдем, что при достаточно большом n $|F_n(x)-\Phi(x)| \leqslant 5\varepsilon$. Так как $\varepsilon>0$ произвольно, то отсюда и следует справедливость утверждения леммы 1.

Итак, для того чтобы доказать теорему 2.3, достаточно установить справедливость соотношения (2.63) равномерно относительно a и b, подчиненных условию $-T \leqslant a < b \leqslant T$, где $T>0$ — произвольное фиксированное число. Для этого воспользуемся локальной предельной теоремой Муавра — Лапласа. Обозначим

$$x_k = \frac{k-np}{\sqrt{npq}} \tag{2.77}$$

(ранее эта величина обозначалась x; теперь же нам понадобится зависимость от k).

Имеем

$$P\left(a \leqslant \frac{k-np}{\sqrt{npq}} < b\right) = \sum_{a<\frac{k-np}{\sqrt{npq}}<b} p_{\nu_n}(k) =$$

$$= \sum_{a < x_k < b} p_{\nu_n}(k). \tag{2.78}$$

По теореме 3.2 при достаточно большом n

$$1 - \varepsilon \leqslant p_{\nu_n}(k) : \left(\frac{1}{\sqrt{2\pi npq}} \exp\left\{ -\frac{x_k^2}{2} \right\} \right) \leqslant 1 + \varepsilon, \tag{2.79}$$

где ε — любое фиксированное положительное число. Двойное неравенство (2.79) можно переписать так:

$$\frac{1}{\sqrt{2\pi npq}} \exp\left\{ -\frac{x_k^2}{2} \right\} (1 - \varepsilon) \leqslant p_{\nu_n}(k) \leqslant$$

$$\leqslant \frac{1}{\sqrt{2\pi npq}} \exp\left\{ -\frac{x_k^2}{2} \right\} (1 + \varepsilon). \tag{2.80}$$

Просуммировав по всем k, для которых $a \leqslant x_k < b$, найдем

$$\frac{1}{\sqrt{2\pi npq}} \sum_{a < x_k < b} \exp\left\{ -\frac{x_k^2}{2} \right\} (1 - \varepsilon) \leqslant \mathrm{P}\left(a \leqslant \frac{k - np}{\sqrt{npq}} < b \right) \leqslant$$

$$\leqslant \frac{1}{\sqrt{2\pi npq}} \sum_{a < x_k < b} \exp\left\{ -\frac{x_k^2}{2} \right\} (1 + \varepsilon). \tag{2.81}$$

Заметим, что

$$x_{k+1} - x_k = \frac{k + 1 - np}{\sqrt{npq}} - \frac{k - np}{\sqrt{npq}} = \frac{1}{\sqrt{npq}}. \tag{2.82}$$

Отсюда

$$\frac{1}{\sqrt{2\pi npq}} \sum_{a < x_k < b} \exp\left\{ -\frac{x_k^2}{2} \right\} = \sum_{a < x_k < b} \varphi(x_k)(x_{k+1} - x_k), \tag{2.83}$$

что представляет собой интегральную сумму для интеграла $\int_a^b \varphi(z)\, dz$. Очевидно, при $n \to \infty$ $x_{k+1} - x_k \to 0$, так что

$$\sum_{a < x_k < b} \varphi(x_k)(x_{k+1} - x_k) \to \int_a^b \varphi(z)\, dz. \tag{2.84}$$

Это предельное соотношение выполняется равномерно относительно a и b $(-T \leqslant a < b \leqslant T)$, поскольку

$$\left| \int_{x_k}^{x_{k+1}} \varphi(z)\, dz - \varphi(x_k)(x_{k+1} - x_k) \right| = \left| \int_{x_k}^{x_{k+1}} [\varphi(z) - \varphi(x_k)]\, dz \right| =$$

$$= \left| \int_{x_k}^{x_{k+1}} \varphi'(\overline{x_k})(z - x_k)\, dz \right| \leqslant \max_{-\infty < x < \infty} |\varphi'(x)| \int_{x_k}^{x_{k+1}} (z -$$

$$- x_k)\, dz = \frac{1}{2}(x_{k+1} - x_k)^2 \max_{-\infty < x < \infty} |\varphi'(x)| =$$

$$= \frac{1}{2npq} \max_{-\infty < x < \infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} |1 - x^2| = O\left(\frac{1}{n}\right); \quad (2.85)$$

в то же время число слагаемых в сумме (2.83) имеет порядок $\sqrt{n}$.*

Вследствие оценки (2.81)

$$\lim_{n \to \infty} \mathrm{P}\left(a \leqslant \frac{k - np}{\sqrt{npq}} < b\right) = \int_a^b \varphi(z)\, dz \qquad (2.86)$$

равномерно по a и b, $-T \leqslant a < b \leqslant T$, что и требовалось доказать.

§ 6. ВЕРОЯТНОСТИ БОЛЬШИХ УКЛОНЕНИЙ

Пусть ν_n — число успехов в n независимых испытаниях с вероятностью p успеха в отдельном испытании. np называется средним значением ν_n; разность $\nu_n - np$ — уклонением ν_n от среднего значения; выражение $\frac{\nu_n - np}{\sqrt{npq}}$ называется нормированным уклонением (ν_n от среднего значения). В теории вероятностей принята такая терминология. Если $n \to \infty$, а нормированное отклонение попадает в заранее фиксированный конечный интервал, то говорят, что имеет место нормальное уклонение. Если же $\left|\frac{\nu_n - np}{\sqrt{npq}}\right|$

* Строго говоря, это выполняется лишь в том случае, когда точки a и b совпадают с некоторыми из x_k. Однако если это не так, выражение (2.83) отличается от интегральной суммы лишь двумя слагаемыми: $\varphi(a)(x_i - a)$ и $-\varphi(x_j)(x_{j+1} - b)$, где x_i и x_j — соответственно минимальное и максимальное из x_k, по которым производится суммирование.

рассматривается как бесконечно большая величина, то говорят, что имеет место большое уклонение (ν_n от np). Большие уклонения имеют бесконечно малую вероятность Действительно, согласно интегральной предельной теореме Муавра — Лапласа, при $n \to \infty$

$$\mathrm{P}\left\{-x \leqslant \frac{\nu_n - np}{\sqrt{npq}} < x\right\} = \mathrm{P}\left\{\frac{\nu_n - np}{\sqrt{npq}} < x\right\} -$$

$$-\mathrm{P}\left\{\frac{\nu_n - np}{\sqrt{npq}} < -x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} qz - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-\frac{z^2}{2}} dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-\frac{z^2}{2}} dz.$$

При достаточно большом x

$$\frac{1}{\sqrt{2\pi}} \int_{-x}^{x} e^{-\frac{z^2}{2}} dz > 1 - \varepsilon,$$

где $\varepsilon > 0$ — любое заданное число. Таким образом, с вероятностью, стремящейся к 1 при $n \to \infty$, имеют место лишь нормальные уклонения. Тем не менее для многих вопросов, в том числе прикладных, важно знать вероятности больших уклонений.

В качестве примера рассмотрим задачу о расчете вероятностей, связанных с радиолокационным обнаружением объектов. Представим себе, что принимается n квантованных радиосигналов $\xi_1, \xi_2, \ldots, \xi_n \cdot \xi_i = 1$, если в i-м приемнике имело место превышение напряжением порога, $\xi_i = 0$ в противном случае. Решение же о наличии объекта принимается в случае, если $\nu_n = \xi_1 + \ldots + \xi_n \geqslant C$. Допустим, что объект, от которого могли бы поступить сигналы, отсутствует. Тогда ξ_i зависят от различных шумов, связанных с радиолокационным приемом. Понятно, что нормальная работа радиолокатора возможна лишь при условии, что $\mathrm{P}(\nu_n \geqslant C)$ (при отсутствии объекта) — величина весьма малая. Однако поскольку частота принятия решения в свою очередь весьма высока, то и при малой вероятности события $\{\nu_n \geqslant C\}$ «ложные тревоги» могут оказаться довольно частыми. Поэтому указанную вероятность следует оценивать весьма точно.

В настоящем параграфе будет решена следующая задача. Пусть заданы некоторые $C>0$ и δ, $\frac{1}{2}<\delta\leqslant 1$. Требуется асимптотически оценить следующие вероятности:

$$\pi_+=\pi_+(n, p, C, \delta)=\mathrm{P}\left(\nu_n-np\geqslant Cn^\delta\right) \tag{2.87}$$

и

$$\pi_-=\pi_-(n, p, C, \delta)=\mathrm{P}\left(\nu_n-np\leqslant -Cn^\delta\right), \tag{2.88}$$

считая, что $n\to\infty$, p фиксировано и не равно ни 0, ни 1.

Используем формулу (2.49)

$$p_{\nu_n}(k)=\frac{1}{\sqrt{2\pi n\alpha\beta}}\left(\frac{p^\alpha\ q^\beta}{\alpha^\alpha\ \beta^\beta}\right)^n\exp\left\{\frac{\Theta_n}{12n}-\frac{\Theta_k}{12k}-\frac{\Theta_{n-k}}{12(n-k)}\right\},$$

где $\alpha=\frac{k}{n}$, $\beta=1-\alpha$, $0\leqslant\Theta_n\leqslant 1$, $0\leqslant\Theta_k\leqslant 1$, $0\leqslant\Theta_{n-k}\leqslant 1$. Далее, заметим, что

$$\frac{p_{\nu_n}(k-1)}{p_{\nu_n}(k)}=\frac{C_n^{k-1}\,p^{k-1}\,q^{n-k+1}}{C_n^k\,p^k\,q^{n-k}}=\frac{kq}{(n-k+1)\,p}. \tag{2.89}$$

В дальнейшем будем считать, что $k_0=np-Cn^\delta$ — целое число.

Если $k<k_0$, то из равенства (2.89) следует, что

$$\frac{p_{\nu_n}(k-1)}{p_{\nu_n}(k)}<\frac{k_0q}{(n-k_0+1)\,p}=\frac{npq-Cqn^\delta}{npq+Cpn^\delta+p}=$$

$$=\frac{1-\frac{Cq}{n^{1-\delta}pq}}{1+\frac{Cp}{n^{1-\delta}pq}+\frac{1}{nq}}=1-\frac{C}{n^{1-\delta}pq}+$$

$$+O\left(\frac{1}{n^{2-2\delta}}\right)=1-\gamma. \tag{2.90}$$

Введем новую переменную Ψ по формуле

$$\pi_-(n, p, C, \delta)=p_{\nu_n}\left(np-Cn^\delta\right)\Psi. \tag{2.91}$$

Поскольку $p_{\nu_n}\left(np-C^\delta n\right)$ может быть оценено по формуле (2.50), в которую следует подставить $\alpha=p-Cn^{\delta-1}$,

$\beta = q + Cn^{\delta-1}$, наша задача сводится к оценке Ψ; ее решением мы и ограничимся.

Имеем

$$\Psi = \sum_{k<k_0} \frac{p_{\nu_n}(k)}{p_{\nu_n}(k_0)} \tag{2.92}$$

и

$$\frac{p_{\nu_n}(k)}{p_{\nu_n}(k_0)} = \frac{p_{\nu_n}(k)}{p_{\nu_n}(k+1)} \frac{p_{\nu_n}(k+1)}{p_{\nu_n}(k+2)} \cdots \frac{p_{\nu_n}(k_0-1)}{p_{\nu_n}(k_0)}. \tag{2.93}$$

В силу неравенства (2.90), которое справедливо для каждой из дробей в правой части (2.93), имеем

$$\frac{p_{\nu_n}(k)}{p_{\nu_n}(k_0)} < (1-\gamma)^{k_0-k}. \tag{2.94}$$

Просуммировав во всем $k \leqslant k_0$, получим

$$\Psi = \sum \frac{p_{\nu_n}(k)}{p_{\nu_n}(k_0)} < \sum_{-\infty<k_0} (1-\gamma)^{k_0-k} = \frac{1}{1-(1-\gamma)} =$$

$$= \frac{1}{\gamma} = \frac{n^{1-\delta} pq}{C} + O(1). \tag{2.95}$$

При $\delta = 1$ можно использовать другую оценку:

$$\Psi < \frac{1}{\gamma} = p + \frac{pq}{C} + O\left(\frac{1}{n}\right). \tag{2.96}$$

С другой стороны, если $N = \dfrac{Kn^{1-\delta} pq}{C}$, то

$$\Psi \geqslant \sum \frac{p_{\nu_n}(k)}{p_{\nu_n}(k)}. \tag{2.97}$$

Если $k_0 - N < k \leqslant k_0$, то из формул (2.93), (2.89) находим

$$\frac{p_{\nu_n}(k)}{p_{\nu_n}(k_0)} \geqslant \left[\frac{\left(np - C_n^{\delta} - N\right) q}{\left(nq + C_n^{\delta} + 1 + N\right) p}\right]^{k_0-k} = (1-\gamma)^{k_0-k}, \tag{2.98}$$

где

$$1-\gamma' = \frac{1-\dfrac{Cq}{n^{1-\delta}pq}-\dfrac{Kq}{Cn^{\delta}}}{1+\dfrac{Cp}{n^{1-\delta}pq}+\dfrac{Kp}{Cn^{\delta}}+\dfrac{1}{no}} =$$

$$= 1-\frac{C}{n^{1-\delta}pq}+o\left(\frac{1}{n^{1-\delta}}\right). \qquad (2.99)$$

Следовательно, в силу неравенства (2.97)

$$\Psi \geqslant \sum_{k_0-N<k<k_0}(1-\gamma')^{k_0-k} = \frac{1-(1-\gamma')^N}{\gamma'} =$$

$$= \left[1-(1-\gamma')^N\right]\frac{n^{1-\delta}pq}{C}(1+o(1)). \qquad (2.100)$$

Далее,

$$(1-\gamma')^N < e^{-N\gamma'} = \exp\left\{-\frac{Kn^{1-\delta}pq}{C}\,\frac{C}{n^{1-\delta}pq}\right\}. \qquad (2.101)$$

Выбрав K так, чтобы $e^{-K}<\varepsilon$ ($\varepsilon>0$ — заданное число), находим

$$\Psi \geqslant (1-\varepsilon)\frac{n^{1-\delta}pq}{C}(1+o(1)). \qquad (2.102)$$

Объединив эту оценку с оценкой (2.95), получим

$$(1-\varepsilon)\frac{n^{1-\delta}pq}{C}(1+o(1)) \leqslant \Psi \leqslant \frac{n^{1-\delta}pq}{C} +$$

$$+O(1)\left(\frac{1}{2}<\delta<1\right). \qquad (2.103)$$

Так как $\varepsilon>0$ произвольно, то

$$\Psi = \frac{n^{1-\delta}pq}{C}(1+o(1)) \leqslant \Psi \leqslant \frac{n^{1-\delta}pq}{C} +$$

$$+O(1)\left(\frac{1}{2}<\delta<1\right). \qquad (2.104)$$

Рассмотрим случай $\delta=1$. Положим

$$N = \frac{Kp(q+C)}{C}. \qquad (2.105)$$

Если $k>k_0-N$, то из равенств (2.93) и (2.89) имеем

$$\frac{p_{\nu_n}(k)}{p_{\nu_n}(k_0)} \geqslant \left[\frac{(np - C_n - N)\,q}{(np + C_n + 1 + N)\,p}\right]^{k_0 - k} = (1-\gamma')^{k_0-k}, \quad (2.106)$$

где

$$1-\gamma' = \frac{npq - C_n - \dfrac{Kp\,(q+C)\,q}{C}}{npq + Cnp + p + \dfrac{Kp^2\,(q+C)}{C}} =$$

$$= \frac{q\,(p-C)}{p\,(q+C)} + O\left(\frac{1}{n}\right). \quad (2.107)$$

Отсюда в силу неравенства (2.97) следует, что

$$\Psi \geqslant \sum_{k_0 - N < k < k_0} (1-\gamma')^{k_0-k} = \frac{1-(1-\gamma')^N}{\gamma'} =$$

$$= \left[1-(1-\gamma')^N\right]\left(p + \frac{pq}{C} + O\left(\frac{1}{n}\right)\right) \quad (2.108)$$

Далее,

$$(1-\gamma')^N < e^{-N\gamma'} = \exp\left\{-\frac{Kp\,(q+C)}{C}\,\frac{C}{p\,(q+C)} + \right.$$

$$\left. + O\left(\frac{1}{n}\right)\right\} = e^{-K}\,(1 + o\,(1)). \quad (2.109)$$

Выберем K из условия $e^{-K} < \varepsilon$, ($\varepsilon > 0$ — заданное число); тогда из выражения (2.108), учитывая выражение (2.109), найдем

$$\Psi \geqslant (1-\varepsilon)\left(p + \frac{pq}{C} + O\left(\frac{1}{n}\right)\right). \quad (2.110)$$

Объединив формулы (2.96) и (2.110), получим

$$(1-\varepsilon)\left(p + \frac{pq}{C} + O\left(\frac{1}{n}\right)\right) \leqslant \Psi \leqslant p + \frac{pq}{C} +$$

$$+ O\left(\frac{1}{n}\right). \quad (2.111)$$

Ввиду произвольности $\varepsilon > 0$

$$\lim_{n\to\infty} \Psi = p + \frac{pq}{C}. \quad (2.112)$$

Таким образом, мы исследовали выражение $\Psi = \sum\limits_{k<k_0} \frac{p_{\nu_n}(k)}{p_{\nu_n}(k_0)}$,

и задача оценки $\pi_-(n, p, C, \delta)$ решена.

Рассмотрим теперь $\pi_+(n, p, C, \delta)$. Имеем

$$\pi_+(n, p, C, \delta) = \mathrm{P}\left(\nu_n - np \geqslant C^\delta n\right) = \mathrm{P}\left(np - \nu_n \leqslant -C^\delta n\right) =$$
$$= \mathrm{P}\left(n - \nu_n - nq \leqslant -C^\delta n\right). \qquad (2.113)$$

Однако $n - \nu_n$ есть число неудач в n испытаниях, или, что то же, число успехов в n испытаниях с вероятностью успеха q. Следовательно,

$$\pi_+(n, p, C, \delta) = \pi_-(n, q, C, \delta). \qquad (2.114)$$

§ 7. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ ДИСКРЕТНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Пусть ξ — дискретная случайная величина, принимающая значения x_i, $i \in I$, с вероятностями p_i $\left(\sum\limits_{i \in I} p_i = 1\right)$.

Определение. *Если ряд* $\sum\limits_{i \in I} x_i p_i$ *имеет определенную сумму (конечную или бесконечную, но определенного знака), остающуюся неизменной при любых перестановках членов ряда, то эта сумма называется математическим ожиданием случайной величины* ξ *и обозначается* $\mathrm{M}\xi$*. *Если данный ряд не обладает определенной суммой в указанном смысле, то говорят, что случайная величина* ξ *не имеет математического ожидания.*

Примеры 1. ξ принимает значения $1, 2, \ldots, i, \ldots$ с вероятностями $\frac{1}{2}, \frac{1}{2^2}, \ldots, \frac{1}{2^i}, \ldots$.

Имеем

$$\sum_{i \in I} x_i p_i = \sum_{i=1}^{\infty} \frac{i}{2^i} = \sum_{i=1}^{\infty} i z^i \Big|_{z=\frac{1}{2}} = z \sum_{i=1}^{\infty} i z^{i-1} \Big|_{z=\frac{1}{2}} =$$
$$= z \sum_{i=1}^{\infty} \frac{d}{dz} z^i \Big|_{z=\frac{1}{2}} = z \frac{d}{dz} \sum_{i=1}^{\infty} z^i = z \frac{d}{dz} \frac{z}{1-z} \Big|_{z=\frac{1}{2}} =$$
$$= \frac{z}{(1-z)^2} \Big|_{z=\frac{1}{2}} = 2.$$

---

* Альтернативные обозначения: $E\xi$, $\bar{\xi}$. Альтернативные названия: среднее значение, среднее, ожидаемое значение.

Таким образом,

$$M\xi = 2.$$

2. ξ принимает значения $2,\ 2^2,\ \ldots,\ 2^i,\ \ldots$ с вероятностями

$$\frac{1}{2},\ \frac{1}{2^2},\ \ldots,\ \frac{1}{2^i},\ \ldots$$

Имеем

$$\sum_{i \in I} x_i p_i = \sum_{i=1}^{\infty} 2^i 2^{-i} = \sum_{i=1}^{\infty} 1 = \infty.$$

Данная случайная величина обладает математическим ожиданием ∞ : $M\xi = \infty$.

3. ξ принимает значения $-2,\ 2^2,\ -2^3,\ \ldots,\ (-1)^i 2^i,\ \ldots$ с вероятностями $\frac{1}{2},\ \frac{1}{2^2},\ \ldots,\ \frac{1}{2^i},\ \ldots$ В этом случае

$$\sum_{i \in I} x_i p_i = \sum_{i=1}^{\infty} (-1)^i 2^i 2^{-i} = \sum_{i=1}^{\infty} (-1)^i = -1 + 1 - 1 + 1 \mp \ldots$$

Этот ряд не обладает определенной суммой. Таким образом, данная случайная величина не имеет математического ожидания.

Определение математического ожидания можно объяснить еще так. Пусть I^+ — множество тех $i \in I$, для которых $x_i > 0$, I^- — множество тех $i \in I$, для которых $x_i < 0$. Образуем выражения

$$M^+ = \sum_{i \in I} x_i p_i, \quad M^- = \sum_{i \in I} x_i p_i. \tag{2.115}$$

В каждом из этих рядов (или конечных сумм) все слагаемые одного знака, следовательно, суммы M^+, M^- всегда существуют. Все возможные случаи представляются следующей таблицей.

№	M^+	M^-	$M\xi$
1	Конечное	Конечное	$M^+ + M^-$ (конечное)
2	∞	»	∞
3	Конечное	$-\infty$	$-\infty$
4	∞	$-\infty$	Не существует

В главе, посвященной закону больших чисел, будет приведена теорема о законе больших чисел, полностью раскрывающая роль математического ожидания. В ней доказывается, что если производится n независимых наблюдений $x_1, x_2, \dots, x_n$ над случайной величиной с конечным математическим ожиданием, то при любом фиксированном $\varepsilon > 0$

$$P\left(\left|\frac{1}{n}(x_1 + x_2 + \dots + x_n) - M\xi\right| \leqslant \varepsilon\right) \underset{n\to\infty}{\to} 1. \tag{2.116}$$

Если же $M\xi = \infty$ или $M\xi = -\infty$, то при любом сколь угодно большом E

$$P\left(\frac{1}{n}(x_1 + x_2 + \dots + x_n) > E\right) \underset{n\to\infty}{\to} 1, \tag{2.117}$$

соответственно

$$P\left(\frac{1}{n}(x_1 + x_2 + \dots + x_n) < -E\right) \underset{n\to\infty}{\to} 1. \tag{2.118}$$

Пусть на вероятностном пространстве $(\Omega, \mathfrak{A}, P)$ заданы две случайные величины: ξ_1 и ξ_2. Предположим, что ξ_1 принимает значения x_i, $i \in I$, с вероятностями p_i, а ξ_2 — значения y_j, $j \in J$, с вероятностями q_j. Множество пар (i, j), где $i \in I$, $j \in J$, является конечным или счетным множеством. Поскольку $\{\omega: \xi_1(\omega) = x_i\}$ и $\{\omega: \xi_2(\omega) = y_j\}$ входят в $\mathfrak{A}$, то и их пересечение, т. е. $\{\omega: \xi_1(\omega) = x_i, \xi_2(\omega) = y_j\}$, входит в $\mathfrak{A}$, а следовательно, обладает некоторой вероятностью. Обозначим эту вероятность p_{ij}. Рассмотрим функцию элементарного события ω вида

$$\eta = \xi_1 + \xi_2 = \eta(\omega) = \xi_1(\omega) + \xi_2(\omega), \quad \omega \in \Omega.$$

В силу только что сказанного η принимает конечное или счетное множество возможных значений, причем каждому из этих значений соответствует некоторая определенная вероятность. Следовательно, η есть дискретная случайная величина. Ее возможные значения $z_{(i,j)} = x_i + y_j$, где (i, j) пробегает множество значений $K = \{(i, j): i \in I, j \in J\}$.

Рассмотрим выражение

$$\sum z_{(i,j)} p_{ij}.$$

Это выражение можно преобразовать так:

$$\sum_{(i,j)\in K} z_{(i,j)} p_{ij} = \sum_{i\in I,\ j\in J} (x_i + y_j) p_{ij}.$$

Предположим, что $M\xi_1$ и $M\xi_2$ существуют и конечны. Тогда, поскольку $\sum\limits_{j\in J} p_{ij} = p_i$, по формуле полной вероятности имеем

$$M\xi_1 = \sum_{i \in I} x_i p_i = \sum_{i \in I,\ j \in J} x_i p_{ij}.$$

Аналогично,

$$M\xi_2 = \sum_{j \in J} y_j q_j = \sum_{i \in I,\ j \in J} y_j p_{ij}.$$

Тогда существует и конечна сумма $\sum_{i \in I,\ j \in J} (x_i + y_j) p_{ij}$. По определению, она равна $M\eta = M(\xi_1 + \xi_2)$. Итак, если $M\xi_1$ и $M\xi_2$ конечны, то

$$M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2. \qquad (2.119)$$

Пусть теперь одно из чисел $M\xi_1$, $M\xi_2$ равно ∞ или $M\xi_1 = M\xi_2 = +\infty$ или, наконец, $M\xi_1 = M\xi_2 = -\infty$. Тогда

$$\sum_{i \in I,\ j \in J} x_i p_{ij} = \sum_{x_i > 0} x_i p_{ij} + \sum_{x_i < 0} x_i p_{ij}.$$

Подчеркнем, что каждый из рядов в правых частях последних двух формул знакопостоянен. Поскольку знакопостоянные ряды можно почленно суммировать (если только их суммы не равны бесконечности разных знаков), подобно предыдущему получаем

$$M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2.$$

Докажем следующее свойство математического ожидания.

Пусть C — постоянная, отличная от нуля. Тогда

$$M(c\xi) = cM\xi. \qquad (2.120)$$

Действительно, $c\xi$ принимает значения cx_i с вероятностями p_i, если $P(\xi = x_i) = p_i$. Поэтому остается использовать формулу

$$\sum cx_i p_i = c \sum x_i p_i.$$

$M(c\xi)$ и $M\xi$ при $c \neq 0$ либо одновременно существуют, либо одновременно не существуют. Между тем при $c = 0$ $M(c\xi)$ всегда существует и равно 0 независимо от того, существует ли $M\xi$.

Отметим следующее свойство математического ожидания. Пусть a — конечное число и пусть известно, что $P(\xi \geqslant a) = 1$. Тогда $M\xi$ существует и $M\xi \geqslant a$. Аналогично, если $b < \infty$, $P(\xi \leqslant b) = 1$, то $M\xi$ существует и $M\xi \leqslant b$.

Докажем для определенности последнее утверждение. Если $P(\xi \leqslant b) = 1$, то все $x_i \leqslant b$ при $i \in I$. Поэтому

$$\sum_{i \in I} x_i p_i \leqslant b \sum_{i \in I} p_i = b.$$

В частности, если $P(\xi = a) = 1$, то $M\xi = a$. Действительно, в этом случае и $\xi \leqslant a$, и $\xi \geqslant a$ с вероятностью 1.

§ 8. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ КАК ИНТЕГРАЛ

Предположим, что Ω — отрезок $[0, 1]$, ω — точки этого отрезка, $\mathfrak{A}$ — некоторая σ-алгебра множеств, содержащая все отрезки $[a, b]$, где $0 \leqslant a \leqslant b \leqslant 1$; $P(A)$ — функция множества, задаваемая равенством $P([a, b]) = b - a$. Таким образом, вероятность попадания ω на некоторый отрезок $[a, b]$ равна длине этого отрезка.

Отложим на отрезке $[0, 1]$ последовательность непересекающихся отрезков Δ_i длины p_i, $i \in I$. Если $\sum_{i \in I} p_i = 1$, то эти отрезки заполнят весь отрезок $[0, 1]$. Определим на отрезке $[0, 1]$ функцию $f(\omega)$ следующим образом. $f(\omega) = x_i$, если ω — внутренняя точка отрезка Δ_i; на стыках отрезков определим $f(\omega)$ либо по непрерывности справа, либо по непрерывности слева. Построенная таким образом функция $f(\omega)$ элементарного события ω есть дискретная случайная величина. Действительно, множество ее возможных значений конечно или счетно; если x — точка этого множества, то множество тех ω, для которых $\xi(\omega) = x$, есть объединение интервалов Δ_i, для которых $x_i = x$, т. е. вероятность события $\{\xi = x\}$ определена:

$$P(\xi = x) = \sum_{i : x_i = x} |\Delta_i|,$$

где $|\Delta_i|$ — длина отрезка Δ_i. В частности, если все x_i попарно различны, то

$$P(\xi = x_i) = |\Delta_i|; \qquad \sum_{i \in I} P(\xi = x_i) = 1.$$

Отметим следующее свойство математического ожидания:

$$M\xi = \int_0^1 f(\omega)\, d\omega. \tag{2.121}$$

Действительно, последний интеграл есть сумма длин отрезков Δ_i, умноженных на соответствующие ординаты кривой $y = f(\omega)$, т. е. на x_i. Таким образом, $\int\limits_0^1 f(\omega)\,d\omega = \sum x_i p_i$.

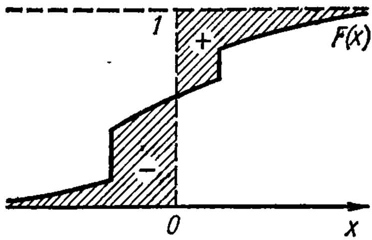

Рис. 2.4

Дискретную случайную величину ξ можно представить ступенчатой функцией $f(\omega)$ точки ω отрезка $[0, 1]$. В этом случае теорема о математическом ожидании суммы случайных величин и математическом ожидании $c\xi$ следуют из равенства

$$\int\limits_0^1 [c_1 f_1(\omega) + c_2 f_2(\omega)]\,d\omega = c_1 \int\limits_0^1 f_1(\omega)\,d\omega + c_2 \int\limits_0^1 f_2(\omega)\,d\omega. \qquad (2.122)$$

Аналогично интерпретируется теорема о том, что $M\xi \leqslant a$, если $\xi \leqslant a$ с вероятностью 1: если $f(\omega) \leqslant a$ при всех $\omega \in [0, 1]$, то и $\int\limits_0^1 f(\omega)\,d\omega \leqslant a$. Важно также выражение математического ожидания через функцию распределения случайной величины ξ. Имеем

$$M\xi = \sum_{i \in I} x_i p_i.$$

В то же время, по определению,

$$F_\xi(x) = \sum_{x_i < x} p_i.$$

Рассмотрим интеграл

$$I = -\int\limits_{-\infty}^{0} F_\xi(x)\,dx + \int\limits_0^\infty \left[1 - F_\xi(x)\right] dx. \qquad (2.123)$$

Этот интеграл изображен на рис. 2.4 (на рисунке показаны площади, составляющие первое и второе слагаемое данной формулы, и отмечены знаки «—» и «+», которые соответствуют этим площадям).

Пусть $x_i > 0$, $i \in I$. Рассмотрим прямоугольник, координаты вершин которого соответственно равны $(0, F(x_i))$, $(0, F(x_i+0))$, $(x_i, F(x_i))$, $(x_i, F(x_i+0))$. Поскольку $F(x_i+0) - F(x_i) = \mathrm{P}(\xi = x_i) = p_i$, то площадь прямоугольника составляет $x_i p_i$. Так как прямоугольники, соответствующие различным i, не пересекаются, то

$$\int_0^\infty [1 - F_\xi(x)]\,dx \geqslant \sum_{x_i > 0} x_i p_i.$$

С другой стороны, эти прямоугольники исчерпывают всю площадь, ограниченную прямой $y = 1$, кривой $y = F_\xi(x)$ и прямой $x = 0$. В самом деле, сумма высот прямоугольников равна $\sum_{x_i > 0} p_i = \mathrm{P}(\xi > 0) = 1 - F(+0)$, т. е. длине отрезка прямой $x = 0$, лежащего между прямой $y = 1$ и кривой $y = F_\xi(x)$. Следовательно,

$$\int_0^\infty [1 - F_\xi(x)]\,dx = \sum_{x_i > 0} x_i p_i.$$

Совершенно аналогично,

$$\int_{-\infty}^0 F_\xi(x)\,dx = \sum_{x_i < 0} |x_i|\, p_i = -\sum_{x_i < 0} x_i p_i.$$

Итак, $I = \mathrm{M}\xi$, т. е. математическое ожидание дискретной случайной величины можно выразить через функцию распределения этой случайной величины:

$$\mathrm{M}\xi = -\int_{-\infty}^0 F_\xi(x)\,dx + \int_0^\infty [1 - F_\xi(x)]\,dx. \qquad (2.124)$$

Отметим еще одно представление математического ожидания дискретной случайной величины.

В различных приложениях используется понятие δ-функции. δ-функция определяется посредством задания интеграла $\int_a^b f(x)\,\delta(x)\,dx$. Для любой непрерывной функции $f(x)$ он определяется так:

$$\int_a^b f(x)\delta(x)\,dx = \begin{cases} f(0), & \text{если } 0 \in [a, b), \\ 0, & \text{если } 0 \overline{\in} [a, b). \end{cases} \qquad (2.125)$$

Таким образом, при $a < 0 < b$

$$\int_a^b f(x)\,\delta(x)\,dx = \lim_{n\to\infty} \int_a^b f(x)\,\delta_n(x)\,dx, \qquad (2.126)$$

где $\delta_n(x)$ — любая последовательность функций со следующими свойствами.

1. $\delta_n(x) = 0$ при $|x| \geqslant \varepsilon_n$, где ε_n — бесконечно малая при $n \to \infty$ величина.
2. $\delta_n(x) \geqslant 0$ при $|x| < \varepsilon_n$.
3. $\int\limits_{|x|<\varepsilon_n} \delta(x)\,dx \xrightarrow[n\to\infty]{} 1.$

Подобным же образом

$$\int_a^b f(x)\delta(x-\tau)\,dx = \begin{cases} f(\tau), & \text{если } \tau \in [a, b), \\ 0, & \text{если } \tau \overline{\in} [a, b). \end{cases} \qquad (2.127)$$

Формально определим выражение $p_\xi(x)$:

$$p_\xi(x) = \sum_{i\in I} p_i \delta(x - x_i).$$

Тогда для любой непрерывной функции $f(x)$ можно определить

$$\int_a^b f(x)\,p_\xi(x)\,dx = \sum_{x_i} f(x_i)\,p_i,$$

если ряд справа сходится абсолютно, т. е. $\sum |f(x_i)|\,p_i < \infty$.

Пусть $f(x) = 1$ $(-\infty < x < \infty)$, тогда

$$\int_a^b p_\xi(x)\,dx = \sum_{x_i\in[a, b)} p_i = \mathrm{P}(a \leqslant \xi < b) = F_\xi(b) - F_\xi(a)$$

Устремляя a к $-\infty$, получим

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

Пусть $f(x) = x$, $a = -\infty$, $b = \infty$.
Тогда

$$\int_{-\infty}^{\infty} f(x)\,p_\xi(x)\,dx = \int_{-\infty}^{\infty} x p_\xi(x)\,dx = \sum_{i\in I} x_i\,p_i = \mathrm{M}\xi, \qquad (2.128)$$

если это математическое ожидание существует.

Математическое ожидание функции дискретной случайной величины. Пусть ξ — дискретная случайная величина, принимающая значения $x_i (i \in I)$ с вероятностями p_i; $f(x)$ — произвольная непрерывная функция. Тогда имеет место следующая важная формула:

$$\mathrm{M} f(\xi) = \int f(x)\,p_\xi(x)\,dx. \qquad (2.129)$$

По определению,

$$\int f(x)\,p_\xi(x)\,dx = \sum_{i\in I} f(x_i)\,p_i.$$

Пусть y_j — попарно различные возможные значения $f(x_i)$, $(i \in I)$. Тогда можно записать

$$\sum_{i\in I} f(x_i)\,p_i = \sum y_j \sum_{f(x_i)=y_j} p_i.$$

Однако $\sum\limits_{f(x_i)=y_j} p_i = \mathrm{P}\,(f(\xi) = x_i)$. Итак, рассматриваемая сумма равна сумме произведений возможных значений случайной величины $f(\xi)$ на соответствующие вероятности. По определению, это и есть $\mathrm{M} f(\xi)$. Формула доказана.

§ 9. МНОГОМЕРНЫЕ ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Пусть на вероятностном пространстве $(\Omega, \mathcal{A}, \mathrm{P})$ заданы $n \geqslant 1$ дискретных случайных величин $\xi_1 = \xi_1(\omega)$, $\xi_2 = \xi_2(\omega)$, .. , $\xi_n = \xi_n(\omega)$ с множеством возможных значений, соответственно $\{x_{1i}\}$, $i \in I_1$, $\{x_{2i}\}$, $i \in I_2$, ... , $\{x_{ni}\}$, $i \in I_n$. Будем считать, что при любом j все x_{ji} различны $(i \in I_j)$.

Определение. *Векторная функция* ω *вида* $\xi = \xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$ *называется n-мерной случайной величиной* (*случайным вектором*). Если $i_1, i_2, \ldots, i_n$ — произвольные элементы множеств $I_1, I_2, \ldots, I_n$, то можно определить вероятность:

$$p_\xi(i_1, i_2, \ldots, i_n) = \mathrm{P}(\xi_1 = x_{1i_1}, \xi_2 = x_{2i_2}, \ldots, \xi_n = x_{ni_n}). \tag{2.130}$$

Действительно, это вероятность пересечения n событий вида $\{\xi_j = x_{ji_j}, 1 \leqslant j \leqslant n\}$, которые входят в $\mathfrak{A}$. По свойству σ-алгебры, пересечение также входит в $\mathfrak{A}$. Функция $p_\xi(i_1, i_2, \ldots, i_n)$, заданная при произвольных $i_j \in I_j$, называется *законом распределения многомерной случайной величины* ξ.

Если для произвольных $i_1 \in I_1, i_2 \in I_2, \ldots, i_n \in I_n$ выполняется равенство

$$p_\xi(i_1, i_2, \ldots, i_n) = \prod_{j=1}^{n} \mathrm{P}(\xi_j = x_{ji_j}), \tag{2.131}$$

то дискретные случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ называются *независимыми в совокупности* (для краткости: независимыми).

Теорема 2.4. *Для независимости в совокупности случайных величин* $\xi_1, \xi_2, \ldots, \xi_n$ *необходимо и достаточно, чтобы для любых функций* $f_1(x), f_2(x), \ldots, f_n(x)$ *таких, что* $Mf_j(\xi_j)| < \infty$, $1 \leqslant j \leqslant n$, *выполнялось соотношение*

$$|M f_1(\xi_1) f_2(\xi_2) \ldots f_n(\xi_n) = Mf_1(\xi_1)\, Mf(\xi_2) \ldots Mf(\xi_n). \tag{2.132}$$

Доказательство. 1. Необходимость. Можно записать

$$\prod_{j=1}^{n} Mf_j(\xi_j) = \prod_{j=1}^{n} \sum_{i_j \in I_j} f_j(x_{ji_j}) \mathrm{P}(\xi_j = x_{ji_j}) =$$

$$= \sum_{i_1 \in I_1, \ldots, i_n \in I_n} f_1(x_{1i_1}) f_2(x_{2i_2}) \ldots f_n(x_{ni_n}) \mathrm{P}(\xi_1 = x_{1i_1}) \times$$

$$\times \mathrm{P}(\xi_2 = x_{2i_2}) \ldots \mathrm{P}(\xi_n = x_{ni_n}) =$$

$$= \sum f_1(x_{1i_1}) f_2(x_{2i_2}) \ldots f_n(x_{ni_n}) p_\xi(i_1, i_2, \ldots, i_n).$$

Последнее выражение представляет собой сумму произведений возможных значений $f_1(\xi_1) f_2(\xi_2) \ldots f_n(\xi_n)$ на их вероятности, что и есть $Mf_1(\xi_1) f_2(\xi_2) \ldots f_n(\xi_n)$.

II. Достаточность. Фиксируем индексы $i_1^0 \in I_1$, $i_2^0 \in I_2, \dots, i_n^0 \in I_n$ и рассмотрим функции

$$f_j(x) = \begin{cases} 1, \text{ если } x = x_{ji_j^0}, \\ 0, \text{ если } x \neq x_{ji_j^0}. \end{cases}$$

Тогда

$$Mf_j(\xi_j) = \mathrm{P}(\xi_j = x_{ji_j^0}), \qquad Mf_1(\xi_1)f_2(\xi_2)\dots f_n(\xi_n) =$$

$$= \mathrm{P}(\xi_1 = x_{1i_1^0}, \ \xi_2 = x_{2i_2^0}, \ \dots, \ \xi_n = x_{ni_n^0}) = p_\xi(i_1^0, i_2^0, \dots, i_n^0).$$

Таким образом,

$$p_\xi(i_1^0, i_2^0, \dots i_n^0) = \prod_{j=1}^{n} \mathrm{P}(\xi_j = x_{ji_j^0}).$$

В силу произвольности $i_1^0, i_2^0, \dots, i_n^0$ можно сделать вывод, что случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы в совокупности.

В качестве следствий выведем такие важные теоремы.

Теорема 2.5. *Если $\xi_1, \xi_2, \dots, \xi_n$ — независимые в совокупности дискретные случайные величины с конечными математическими ожиданиями, то*

$$M\,\xi_1\xi_2\dots\xi_n = M\,\xi_1 M\xi_2 \dots M\,\xi_n. \tag{2.133}$$

Это следствие непосредственно очевидно.

Теорема 2.6. *Если $\xi_1, \xi_2, \dots, \xi_n$ — назависимые в совокупности дискретные случайные величины, $g_1(x)$, $g_2(x), \dots, g_n(x)$ — произвольные функции, то случайные величины $\eta_1 = g(\xi_1)$, $\eta_2 = g_2(\xi_2), \dots, \eta_n = g_n(\xi_n)$ также независимы в совокупности.*

Доказательство. Пусть $f_1(y), f_2(y), \dots, f_n(y)$ — произвольные функции такие, что

$$|M\,f_j(\eta_j)| < \infty$$

Имеем

$$f_j(\eta_j) = f_j(g_j(\xi_j)) = \bar{f}_j(\xi_j).$$

Вследствие независимости $\xi_1, \xi_2, \dots, \xi_n$ в совокупности можно записать

$$Mf_1(\eta_1)f_2(\eta_2)\dots f_n(\eta_n) = M\bar{f}_1(\xi_1)\bar{f}_2(\xi_2)\dots\bar{f}_n(\xi_n) =$$

$$= \prod_{j=1}^{n} M\overline{f}_j(\xi_j) = \prod_{j=1}^{n} Mf_j(\eta_j). \qquad (2.134)$$

Отсюда и следует, что $\eta_1, \eta_2, \ldots, \eta_n$ независимы в совокупности. Теорема доказана.

Дадим следующее определение.

Определение. *Математическим ожиданием n-мерной дискретной случайной величины $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ называется вектор*

$$M\xi = (M\xi_1, \ M\xi_2, \ldots, \ M\xi_n). \qquad (2.135)$$

Условные распределения и условные математические ожидания. Пусть имеется n-мерная случайная величина $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$, где ξ_j — случайные величины с множеством возможных значений $\{x_{ji}\}$, $i \in I_j$, $1 \leqslant j \leqslant n$, и пусть задан закон распределения $p_\xi(i_1, i_2, \ldots, i_n)$ этой n-мерной случайной величины. Обозначим $\xi' = (\xi_1, \xi_2, \ldots, \xi_m)$, $\xi'' = (\xi_{m+1}, \xi_{m+2}, \ldots, \xi_n)$, где $1 \leqslant m < n$. Тогда функция

$$p_{\xi'/\xi''}(i_1, i_2, \ldots, i_m/i_{m+1}, \ldots i_n) = P(\xi_1 = x_{1i_1}, \ldots, \xi_m =$$

$$= x_{mi_m} / \xi_{m+1} = x_{m+1, i_{m+1}}, \ldots, \xi_n = x_{ni_n}) \qquad (2.136)$$

называется *условным законом распределения* m-мерной случайной величины ξ' при фиксированном значении $n - m$-мерной случайной величины ξ''. По общей формуле

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

находим

$$p_{\xi'/\xi''}(i_1, i_2, \ldots, i_m/i_{m+1}, \ldots, i_n) =$$

$$= \frac{p_\xi(i_1, i_2, \ldots, i_n)}{p_{\xi''}(i_{m+1} i_{m+2}, \ldots, i_n)}. \qquad (2.137)$$

Это определение имеет смысл лишь при условии, что $p_{\xi''}(i_{m+1}, i_{m+2}, \ldots, i_n) > 0$. Просуммировав по всем $i_1, i_2, \ldots, i_m$, получим

$$\sum_{i_1, \ldots, i_m} p_{\xi'/\xi''}(i_1, \ldots i_m/i_{m+1}, \ldots, i_n) =$$

$$= \frac{1}{p_{\xi''}(i_{m+1}, \ldots, i_n)} \sum_{i_1, \ldots, i_m} p_\xi(i_1, \ldots, i_m, i_{m+1}, \ldots, i_n) =$$

$$= \frac{1}{p_{\xi''}(i_{m+1}, \ldots, i_n)} \sum_{i_1, \ldots, i_m} P(\xi_{m+1} = x_{m+1,\, i_{m+1}}, \ldots, \xi_n =$$

$$= x_{ni_n}) = 1.$$

Таким образом, условное распределение представляет собой некоторое распределение вероятностей в обычном смысле.

Введем m-мерную случайную величину $\xi'(i_{m+1}, \ldots, i_n)$ с распределением $p_{\xi'/\xi''}(i_1, \ldots, i_m/i_{m+1}, \ldots, i_n)$. Эта величина рассматривается при фиксированных значениях $i_{m+1}, \ldots, i_n$. Пусть $x''(i_{m+1}, \ldots) = (x_{m+1,\, i_{m+1}}, \ldots, x_{ni_n})$. Обозначим

$$M\xi'(i_{m+1}, \ldots, i_n) = M(\xi'/\xi'' = x''(i_{m+1}, \ldots)) \tag{2.138}$$

и назовем этот m-мерный вектор условным математическим ожиданием ξ' при условии $\xi'' = x''(i_{m+1}, \ldots)$.

Справедлива следующая важная формула полного математического ожидания:

$$M\xi' = \sum_{i_{m+1}, \ldots, i_n} M\xi'(i_{m+1}, \ldots, i_n)\, p''(i_{m+1}, \ldots, i_n), \tag{2.139}$$

смысл которой состоит в том, что $M\xi'$ есть математическое ожидание условного математического ожидания ξ'.

Эту формулу следует понимать в таком смысле. Если $M\xi'$ существует (т. е. все компоненты этого вектора определены), то $M\xi'(i_{m+1}, \ldots, i_n)$ определено для всех $i_{m+1}, \ldots$, для которых $p_{\xi''}(i_{m+1}, \ldots, i_n) > 0$ и данная формула справедлива. Если же найдется $(i_{m+1}, \ldots, i_m)$, для которого $p_{\xi''}(i_{m+1}, \ldots, i_n) > 0$ и $M\xi'(i_{m+1}, \ldots, i_n)$ не существует, то и $M_{\xi'}$ не существует. Обозначим

$$\xi'(i_{m+1}, \ldots, i_n) = (\xi_1(i_{m+1}, \ldots, i_n), \ldots, \xi_m(i_{m+1}, \ldots, i_n)).$$

Формула, которую надлежит установить, равносильна системе равенств

$$M\xi_j = \sum_{i_{m+1}, \ldots, i_n} M\xi_j(i_{m+1}, \ldots, i_n)\, p_{\xi''}(i_{m+1}, \ldots, i_n),$$

$$1 \leqslant j \leqslant m. \tag{2.140}$$

Докажем эти равенства, ограничившись для простоты случаем, когда ξ_j может принимать лишь неотрицательные значения (из этого следует законность перестановки порядка суммирования для бесконечных рядов). Имеем

$$M\xi_j = \sum_{i_1, \ldots, i_n} x_{ji_j} \, p_\xi(i_1, \ldots, i_n) =$$

$$= \sum_{i_{m+1}, \ldots, i_n} p_{\xi''}(i_{m+1}, \ldots, i_n) \left[\frac{1}{p_{\xi''}(i_{m+1}, \ldots, i_n)} \times \right.$$

$$\left. \times \sum_{i_1, \ldots, i_m} x_{ji_j} \, p_\xi(i_1, \ldots, i_n) \right] = \sum_{i_{m+1}, \ldots, i_n} p_{\xi''}(i_{m+1}, \ldots, i_n) \times$$

$$\times \sum_{i_1, \ldots, i_m} x_{ji_j} \, p_{\xi'/\xi''}(i_1, \ldots, i_m / i_{m+1}, \ldots, i_n) =$$

$$= \sum_{i_{m+1}, \ldots, i_n} p_{\xi''}(i_{m+1}, \ldots, i_n) \, M\xi_j(i_{m+1}, \ldots, i_n),$$

что и требовалось доказать.

Рассмотрим задачу на вычисление условного распределения дискретной случайной величины и условного математического ожидания.

В счетчик Гейгера — Мюллера попадает поток космических частиц. Число ξ_1 частиц, попадающих в счетчик в течение времени $t > 0$, распределено по закону Пуассона с параметром λt, где $\lambda > 0$ — постоянная:

$$P(\xi_1 = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

Каждая из частиц независимо от других может быть зарегистрирована с вероятностью p $(0 < p < 1)$. Обозначим число частиц, зарегистрированных за время t, через ξ_2. Требуется вычислить условное распределение случайной величины ξ_1 при условии, что $\xi_2 = j$, а также условное математическое ожидание ξ_1 при том же условии.

Очевидно, с вероятностью 1 $\xi_1 \geqslant \xi_2$. Поэтому $P(\xi_1 = k/\xi_2 = j) = 0$ при $k < j$. Пусть $k \geqslant j$. Запишем формулу

$$P(\xi_1 = k/\xi_2 = j) = \frac{P(\xi_1 = k, \xi_2 = j)}{P(\xi_2 = j)}.$$

Соотношение $\{\xi_1 = k, \xi_2 = j\}$ означает, что в счетчик поступило k частиц и из них j зарегистрировано. По формуле для распределения числа успехов в k независимых испытаниях

$$P(\xi_1 = k, \quad \xi_2 = j) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} C_k^j p^j q^{k-j} =$$

$$= e^{-\lambda t} (\lambda t q)^k \frac{1}{j!\,(k-j)!} \left(\frac{p}{q}\right)^j,$$

где $q = 1 - p$. По формуле полной вероятности,

$$P(\xi_2 = j) = \sum_{i=j}^{\infty} P(\xi_1 = i, \ \xi_2 = j) = e^{-\lambda t} \sum_{i=j}^{\infty} \frac{(\lambda t q)^i}{j!\,(i-j)!} \left(\frac{p}{q}\right)^j =$$

$$= e^{-\lambda t p} \frac{(\lambda t p)^j}{j!}, \quad j > 0.$$

Отсюда

$$P(\xi_1 = k/\xi_2 = j) = e^{-\lambda t q} \frac{(\lambda t q)^{k-j}}{(k-j)!}, \quad k \geqslant j.$$

Подобным же образом

$$M\{\xi_1/\xi_2 = j\} = e^{-\lambda t q} \sum_{k=j}^{\infty} \frac{k(\lambda t q)^{k-j}}{(k-j)!} = e^{-\lambda t q} \sum_{k=j}^{\infty} \frac{(k-j)(\lambda t q)^{k-j}}{(k-j)!} +$$

$$+ e^{-\lambda t q}\, j \sum_{k=j}^{\infty} \frac{(\lambda t q)^{k-j}}{(k-j)!} = e^{-\lambda t q} \lambda t q e^{\lambda t q} + e^{-\lambda t q} j e^{\lambda t q} = \lambda t q + j.$$

Поставленная задача решена полностью. Рассмотрим теперь случайную величину $\xi_3 = \xi_1 - \xi_2$. Найдем условное распределение ξ_3 при фиксированной ξ_2. Имеем

$$P(\xi_3 = l/\xi_2 = j) = P(\xi_1 = l + j/\xi_2 = j) = e^{-\lambda t q} \frac{(\lambda t q)^l}{l!}, \quad l \geqslant 0.$$

Это означает, что число незарегистрированных частиц за время t распределено по закону Пуассона с параметром $\lambda t q$ независимо от числа зарегистрированных частиц. Отсюда, в частности, следует, что

$$M\{\xi_3/\xi_2 = j\} = \lambda t q.$$

ГЛАВА III

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ ОБЩЕГО ВИДА

§ 1. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

В предыдущей главе были изучены дискретные случайные величины. Класс дискретных случайных величин недостаточен для описания реальных величин, зависящих от случая. Действительно, таким величинам, как размеры любых физических объектов, температура, давление, длительность тех или иных физических процессов, совершенно неестественно приписывать дискретное множество возможных значений: напротив, естественно считать, что их возможные значения в принципе могут быть любыми числами в некоторых пределах (например, длина всегда положительна). В настоящей главе будет дано понятие случайной величины, не стесненное ограничением множества ее возможных значений.

Пусть $(\Omega, \mathcal{A}, P)$ — вероятностное пространство. Будем рассматривать числовые функции $\xi = \xi(\omega)$, заданные при всех $\omega \in \Omega$.

О п р е д е л е н и е. *Числовая функция* $\xi = \xi(\omega)$ *называется случайной величиной, если для любого* x, $-\infty < x < \infty$, *множество тех* ω, *для которых* $\xi(\omega) < x$, *принадлежит* σ*-алгебре* $\mathcal{A}$ *и*

$$P\{-\infty < \xi(\omega) < \infty\} = 1.$$

Первое требование имеет следующий простой смысл. Поскольку σ-алгебра $\mathcal{A}$ — это класс событий, для каждого из которых определена вероятность, то данное требование означает, что для любого x должна быть определена вероятность события $\{\xi < x\}$. Второе требование означает, что бесконечные значения принимаются случайной величиной ξ лишь с нулевой вероятностью.

Обозначим существующую в силу принятого определения вероятность события $\{\xi < x\}$ через $F_\xi(x)$:

$$F_\xi(x) = P\{\xi < x\}.$$

Определение. *$F_\xi(x)$ как функция x называется функцией распределения случайной величины ξ.*

Если в какой-либо задаче рассматривается единственная случайная величина, то обычно вместо $F_\xi(x)$ пишут $F(x)$; допускаются и другие обозначения.

§ 2. СВОЙСТВА ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

1. $F(x)$ — неубывающая функция x, $-\infty < x < \infty$ Действительно, при $x \leqslant y$

$$\{\xi < y\} = \{\xi < x\} \cup \{x \leqslant \xi < y\},$$

причем события $\{\xi < x\}$ и $\{x \leqslant \xi < y\}$ несовместны. Отсюда, по теореме сложения вероятностей, имеем

$$P(\xi < y) = P(\xi < x) + P(x \leqslant \xi < y),$$

или, что то же,

$$F(y) = F(x) + P(x \leqslant \xi < y). \tag{3.1}$$

Поскольку второе слагаемое правой части как некоторая вероятность неотрицательно, $F(y) \geqslant F(x)$, что и требовалось доказать.

2. Формулу (3.1) можно переписать так:

$$P(x \leqslant \xi < y) = F(y) - F(x), \quad x \leqslant y. \tag{3.2}$$

Таким образом, вероятность попадания случайной величины в полузамкнутый интервал $[x, y)$ есть разность значений функции распределения данной случайной величины в точках y и x.

3.

$$F(-\infty) = \lim_{x\to-\infty} F(x) = 0; \; F(\infty) = \lim_{x\to\infty} F(x) = 1. \tag{3.3}$$

Предположим, что $\{x_n\}$ и $\{y_n\}$ — две монотонные последовательности; $x_n \to -\infty$, $y_n \to \infty$. Пусть событие A_n равносильно попаданию случайной величины ξ в полузамкнутый интервал $[x_n, y_n)$. Очевидно,

$$\{-\infty < \xi < \infty\} = \bigcup_{n=1}^{\infty} A_n.$$

Так как $A_n \subset A_{n+1}$, $n \geqslant 1$, то вследствие аксиомы непрерывности

$$P(-\infty < \xi < \infty) = \lim_{n\to\infty} P(A_n).$$

С другой стороны, по определению случайной величины, $P(-\infty < \xi < \infty) = 1$ и по формуле (3.2) $P(A_n) = F(y_n) - F(x_n)$. Следовательно,

$$\lim_{n\to\infty} [F(y_n) - F(x_n)] = 1. \tag{3.4}$$

Значит, если $\varepsilon > 0$ фиксировано, то, начиная с некоторого номера N, для всех n

$$F(y_n) - F(x_n) > 1 - \varepsilon.$$

Тогда тем более $F(y_n) > 1 - \varepsilon$; $F(x_n) < F(y_n) - 1 + \varepsilon \leqslant \leqslant 1 - 1 + \varepsilon = \varepsilon$. Отсюда вытекает, что $F(x_n) \to 0$, $F(y_n) \to 1$ при $n \to \infty$. Требуемое доказано.

4. $P(\xi \geqslant x) = 1 - F(x)$. (3.5)

Действительно,

$$\{-\infty < \xi < \infty\} = \{\xi < x\} \cup \{\xi \geqslant x\},$$

откуда

$$1 = F(x) + P(\xi \geqslant x),$$

что равносильно формуле (3.5).

5. Функция $F(x)$ непрерывна слева в любой точке x, $-\infty < x < \infty$.

Пусть $\{x_n\}$ — любая возрастающая последовательность чисел, сходящаяся к x. Тогда можно записать

$$\{\xi < x\} = \{\xi < x_1\} \cup \{x_1 \leqslant \xi < x_2\} \cup \ldots \cup \{x_{n-1} \leqslant \xi < x_n\} \ldots \tag{3.6}$$

По расширенной аксиоме сложения вероятностей,

$$P(\xi < x) = P(\xi < x_1) + P(x_1 \leqslant \xi < x_2) + \ldots + P(x_{n-1} \leqslant \xi < x_n) + \ldots . \tag{3.7}$$

Поскольку правая часть соотношения (3.7) есть ряд из неотрицательных членов, сходящийся к $P(\xi < x)$, остаток ряда после некоторого n-го члена будет меньше ε, где $\varepsilon > 0$ — любое заранее фиксированное число. Итак,

$$P(\xi < x) \leqslant P(\xi < x_1) + P(x_1 \leqslant \xi < x_2) + \ldots + P(x_{n-1} \leqslant \xi < x_n) + \varepsilon. \tag{3.8}$$

Применив формулу (3.2), найдем

$$F(x) \leqslant F(x_1) + [F(x_2) - F(x_1)] + \ldots + [F(x_n) - F(x_{n-1})] + \varepsilon$$

или после сокращения подобных членов

$$F(x) \leqslant F(x_n) + \varepsilon. \tag{3.9}$$

Поскольку к тому же $F(x) \geqslant F(x_n)$, то

$$|F(x) - F(x_n)| \leqslant \varepsilon. \tag{3.10}$$

Последнее соотношение вследствие произвольности $\varepsilon > 0$ эквивалентно утверждению, что $F(x_n) \to F(x)$; это и требовалось доказать.

6. Справедлива формула

$$P(\xi \leqslant x) = F(x+0) \equiv \lim_{x_n \downarrow x} F(x_n)^*. \tag{3.11}$$

Действительно,

$$\{\xi > x\} = \{\xi \geqslant x_1\} \cup \{x_2 \leqslant \xi < x_1\} \cup \ldots \{x_n \leqslant \xi < x_{n-1}\} \cup \ldots, \tag{3.12}$$

если только $x_1 > x$. Аналогично доказательству предыдущего свойства находим, что

$$P(\xi > x) = \lim_{n\to\infty} P(\{\xi \geqslant x_1\} \cup \{x_2 \leqslant \xi < x_1\} \cup \ldots \cup \{x_n \leqslant \xi < x_{n-1}\}) = \lim_{n\to\infty} P(\xi \geqslant x_n) = 1 - \lim_{n\to\infty} P(\xi < x_n) = 1 - \lim_{n\to\infty} F(x_n) = 1 - F(x+0).$$

Отсюда

$$P(\xi \leqslant x) = 1 - P(\xi > x) = F(x+0).$$

7. Имеет место формула

$$P(\xi = x) = F(x+0) - F(x). \tag{3.13}$$

Это соотношение — следствие из формулы (3.11) и соотношения (3.10).

Введем следующее понятие. Пусть $\mathfrak{A}$ — минимальная σ-алгебра числовых множеств, содержащая все множества

* Формула $x_n \downarrow x$ означает, что $\{x_n\}$ — невозрастающая последовательность, сходящаяся к x.

вида $(-\infty, x)$, $-\infty < x < \infty$*. Элементы этой σ-алгебры называются *борелевскими множествами*.

8. Все элементы $\mathfrak{N}$ имеют определенные вероятности.

Доказательство. Обозначим через γ совокупность всех числовых множеств A, для которых определена $P(A)$ в том смысле, что множество $\{\omega : \xi(\omega) \in A\}$ принадлежит σ-алгебре $\mathfrak{A}$. Прежде всего теоретико-множественным операциям над множествами числовой прямой соответствуют те же операции над подмножествами Ω. Именно, если A_i — числовые множества, то

$$\{\omega : \xi(\omega) \in A_1 \cup A_2\} =$$
$$= \{\omega : \xi(\omega) \in A_1\} \cup \{\omega : \xi(\omega) \in A_2\},$$
$$\{\omega : \xi(\omega) \in A_1 \setminus A_2\} =$$
$$= \{\omega : \xi(\omega) \in A_1\} \setminus \{\omega : \xi(\omega) \in A_2\},$$
$$\left\{\omega : \xi(\omega) \in \bigcup_{n=1}^{\infty} A_n\right\} = \bigcup_{n=1}^{\infty} \{\omega : \xi(\omega) \in A_i\}.$$

Если $A_i \in \gamma$, то $\{\omega : \xi(\omega) \in A_i\} \in \mathfrak{A}$, а следовательно, множества, представляющие левые части выписанных соотношений, также входят в $\mathfrak{A}$. Отсюда же непосредственно следует, что множества $A_1 \cup A_2$, $A_1 \setminus A_2$, $\bigcup_{i=1}^{\infty} A_i$ входят в γ. Таким образом, γ есть σ-алгебра. Содержа множества вида $(-\infty, x)$, она тем самым должна содержать минимальную σ-алгебру числовых множеств, в которую входят такие множества, т. е. σ-алгебру борелевских множеств $\mathfrak{N}$. Требуемое доказано.

9. Функция распределения $F_\xi(x)$ однозначно определяет вероятности событий $\{\xi \in A\}$, где A — любые борелевские множества, а также множества, входящие в борелевские множества, которым соответствует нулевая вероятность.

Это доказать несложно, однако доказательство требует привлечения ряда понятий теории меры и поэтому здесь не приводятся.

Таким образом, если $A \subset B$, где B — борелевское множество, причем $P(B) = 0$, то $P(A) = 0$ независимо от того, является ли A борелевским множеством.

---

* Минимальной σ-алгеброй множеств с данным свойством называется σ-алгебра входящая в любую σ-алгебру множеств с данным свойством.

Сформулированное в данном пункте свойство функции распределения существенно используется лишь тогда, когда вероятностное пространство $(\Omega, \mathcal{A}, \mathrm{P})$ не задается заранее, а строится по функции распределения.

10. Пусть $F(x)$ — неубывающая, непрерывная слева функция, $F(-\infty) = 0$, $F(\infty) = 1$. Тогда существует вероятностное пространство $(\Omega, \mathcal{A}, \mathrm{P})$ и определенная на нем случайная величина ξ, для которой $F_\xi(x) = F(x)$.

Доказательство справедливости этого свойства осуществляется методами теории меры и основано на том, что в качестве Ω следует выбрать числовую прямую $(-\infty < \omega < < \infty)$, так что $\xi = \omega$ — тождественная функция элементарного события пространства Ω, и сначала определяются вероятности событий $\{\xi < x\}$ формулой $\mathrm{P}(\xi < x) = F(x)$, а затем определение $\mathrm{P}(A)$ распространяется на все борелевские множества A с помощью покрытия любого борелевского множества A последовательностями полуинтервалов $[a \leqslant \xi < b)$. При этом также определяется вероятность любого события $\{\xi \in A\}$, для которого $A \subset B$, где B — борелевское множество со свойством $P(B) = \mathrm{P}(\xi \in B) = 0 : \mathrm{P}(A) = 0$.

11. Если при некотором $a \neq -\infty$ $F(a) = 0$, то значения случайной величины, меньшие a, возможны лишь с нулевой вероятностью: $\mathrm{P}(\xi < a) = 0$. В этом случае говорят, что случайная величина ограничена снизу (числом a). Если при некотором $b \neq \infty$ $F(b + 0) = 1$ [тем более если $F(b) = 1$], то случайная величина может принимать значения, большие b, с вероятностью 0. Такая случайная величина называется ограниченной сверху (числом b).

Определение. *Если случайная величина ограничена снизу и сверху, она называется ограниченной случайной величиной.*

Пусть x — любое число, $-\infty < x < \infty$. Обозначим через $U(x, \varepsilon)$ полуоткрытый интервал $[x - \varepsilon, x + \varepsilon)$. Возможны два случая: либо при некотором $\varepsilon > 0$ $F(x+\varepsilon) = F(x - \varepsilon)$, либо при любых $\varepsilon > 0$ $F(x + \varepsilon) > F(x - \varepsilon)$. В последнем случае x есть точка роста функции $F(x)$, в первом — x не является точкой роста. Во втором случае случайная величина ξ может с положительной вероятностью принимать значения ξ, сколь угодно близкие к x, в первом для некоторого $\varepsilon > 0$ $\mathrm{P}(\xi \in U(x, \varepsilon)) = 0$. Обозначим через S множество точек роста функции $F(x)$.

12. Справедливо соотношение

$$P(\xi \in S)=1 \qquad (3,14)$$

Доказательство. Пусть Λ — множество точек x, не являющихся точками роста $F(x)$. Каждой точке $x \in \Lambda$ соответствует содержащий ее интервал $(a(x), b(x))$, сплошь занятый точками из Λ, причем либо $a(x) \in S$, либо $a(x) = -\infty$, либо $b(x) \in S$, либо $b(x) = \infty$. Очевидно, Λ покрывается множеством непересекающихся интервалов вида $[a(x), b(x]$. Среди этих интервалов не более счетного множества различных. (Известно,что на прямой можно выбрать не более счетного множества непересекающихся интервалов.) Обозначим эти различные интервалы: (a_1, b_1), (a_2, b_2), ... , (a_n, b_n), ... Тогда имеем

$$\Lambda = \bigcup_{n=1}^{\infty} (a_n, b_n),$$

откуда

$$P(\Lambda) = \sum_{n=1}^{\infty} P(a_n < \xi < b_n). \qquad (3.15)$$

Однако каждый член последнего ряда равен 0, так что и $P(\Lambda) = 0$ или, иначе, $P(S) = 1$; это и требовалось доказать.

13. Пусть на вероятностном пространстве $(\Omega, \mathfrak{A}, P)$ заданы две случайные величины: $\xi = \xi(\omega)$ и $\eta = \eta(\omega)$. Пусть, далее, известно, что $P(\xi(\omega) \leqslant \eta(\omega)) = 1$. Тогда для всех x выполняется неравенство

$$F_\xi(x) \geqslant F_\eta(x). \qquad (3.16)$$

т. е. неравенству для случайных величин соответствует неравенство противоположного знака для функций распределения.

Для доказательства формулы (3.16) введем следующие события:

$$A = \{\xi < x\},\ B = \{\eta < x\},\ C = \{\xi > \eta\}.$$

Можно записать

$$B \subset A \cup C. \qquad (3.17)$$

Действительно, для любого $\omega \in B$ либо $\xi(\omega) > \eta(\omega)$, и тогда $\omega \in C$, либо $\xi(\omega) \leqslant \eta(\omega)$, а следовательно, $\xi(\omega) < x$, т. е. $\omega \in A$. Из (3.17) выводим, что $P(B) \leqslant P(A) + P(C)$, но так как, по условию, $P(C) = 0$, то $P(B) \leqslant P(A)$, или, что то же, $F_\eta(x) \leqslant F_\xi(x)$,. Требуемое доказано.

§ 3. СИНГУЛЯРНЫЕ РАСПРЕДЕЛЕНИЯ

Распределение случайной величины ξ называется *сингулярным*, если выполнены следующие два условия:

1) функция $F(x) = F_\xi(x)$ непрерывна. В силу формулы (3.13) это означает, что каждое отдельное значение случайной величины ξ имеет вероятность 0;

2) для любого $\varepsilon > 0$ можно найти такую конечную или счетную последовательность интервалов $\Delta_1, \Delta_2, \ldots, \Delta_N, \ldots$ с суммарной длиной, меньшей ε, что S (см. п. 11 предыдущего параграфа) покрывается объединением $\bigcup_N \Delta_N$. Приведем пример синулярного распределения. Пусть

$$\xi = \frac{\xi_1}{10} + \frac{\xi_2}{10^2} + \cdots + \frac{\xi_n}{10^n} + \cdots,$$

где $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — символы десятичной записи числа ξ. Предположим, что для некоторой последовательности $1 \leqslant k_1 < k_2 < \ldots < k_m < \ldots$ символы $\xi_{k_1}, \xi_{k_2}, \ldots, \xi_{k_m}, \ldots$ фиксированы, а остальные ξ_n, которых также бесконечное множество, независимы, и каждая из них принимает значения $0, 1, \ldots, 9$ с вероятностью $\frac{1}{10}$. Докажем, что ξ обладает сингулярным распределением.

Обозначим через $r(n)$ число элементов последовательности $\{k_m\}$ среди чисел $1, 2, \ldots, n$. Таким образом, $r(n)$ — число фиксированных символов среди первых n символов записи дробного числа ξ. Множество дробных чисел с фиксированными первыми n десятичными знаками заполняет отрезок длины 10^{-n}. (Например, все числа, запись которых начинается с цифр 0,11, расположены в отрезке 0,11; 0,12.) Таким образом, ξ может принадлежать лишь одному из $10^{n-r(n)}$ отрезков длины 10^{-n}; длина этих отрезков в сумме составляет $10^{-r(n)}$ Ясно, что $r(n) \to \infty$ при $n \to \infty$. Взяв такое n, чтобы $10^{-r(n)} < \varepsilon$, найдем, что свойство 2) выполняется. В то же время, вероятность принадлежности ξ любому из этих $10^{n-r(n)}$ интервалов одна и та же и составляет $\frac{1}{10^{n-r(n)}}$.

Предположим, что функция $F(x)$ имеет хотя бы один скачок величины $\varepsilon > 0$. Тогда для любого n точка скачка должна была бы принадлежать некоторому отрезку длины 10^{-n}, на котором приращение $F(x)$ составляет $\frac{1}{10^{n-r(n)}}$. Таким образом, всегда было $\frac{1}{10^{n-r(n)}} \geqslant \varepsilon$; поскольку, очевидно, $n - r(n) \to \infty$ при $n \to \infty$ (по условию, нефиксированных символов бесконечно много), то это неравенство противоречиво. Отсюда можно сделать вывод, что условие 1 также выполнено.

§ 4. НЕПРЕРЫВНЫЕ РАСПРЕДЕЛЕНИЯ

Определение. *Распределение случайной величины ξ называется непрерывным, если существует такая интегрируемая функция $p(x)$, $-\infty < x < \infty$, что в любой точке x*

$$F_\xi(x) = \int_{-\infty}^{x} p_\xi(t)\,dt. \tag{3.18}$$

Функция $p_\xi(x)$, $-\infty < x < \infty$, называется плотностью вероятности случайной величины ξ или *плотностью распределения $F_\xi(x)$* (короче говорят: *плотность случайной величины* ξ, или *плотность* ξ). Вместо $p_\xi(x)$ часто пишут $p(x)$; возможны и другие обозначения.

Заметим, что в общем случае интеграл (3.18) понимается как интеграл Лебега. (Понятие интеграла Лебега определяется в конце настоящей главы.) Как известно, если функция интегрируема в смысле Римана, то ее интеграл Лебега существует и совпадает с интегралом Римана. Используемые на практике плотности обычно интегрируемы в смысле Римана. Поэтому интересующийся лишь приложениями теории вероятностей может понимать интеграл в формуле (3.18) в обычном смысле, как интеграл Римана. Очевидно, если распределение случайной величины ξ непрерывно, то функция $F_\xi(x)$ непрерывна, и тогда для любого x $P(\xi = x)=0$; для любых x, y, $x < y$,

$$P(x \leqslant \xi \leqslant y) = P(x < \xi < y) = P(x \leqslant \xi < y) = \\ = P(x < \xi \leqslant y).$$

Обратное, вообще говоря, неверно: в случае сингулярного распределения функция $F_\xi(x)$ непрерывна, но плотность не существует.

Пусть x — точка непрерывности функции $p_\xi(t)$. Возьмем неотрицательные числа h_1 и h_2 (в сумме равные Δx) и рассмотрим вероятность события $\{x - h_1 < \xi < x + h_2\}$. Имеем

$$P(x - h_1 < \xi < x + h_2) = F(x + h_2) - F(x - h_1) =$$

$$= \int\limits_{x-h_1}^{x+h_2} p_\xi(t)\,dt = p_\xi(x)\int\limits_{x-h_1}^{x+h_2} dt + \int\limits_{x-h_1}^{x+h_2} [p_\xi(t) - p_\xi(x)]\,dt =$$

$$= p_\xi(x)\,\Delta x + \int\limits_{x-h_1}^{x+h_2} [p_\xi(t) - p_\xi(x)]\,dt. \tag{3.19}$$

В силу непрерывности функции $p_\xi(t)$ в точке x

$$\sup_{x-h_1<t<x+h_2} |p_\xi(t) - p_\xi(x)| = o(1).$$

Отсюда

$$\mathrm{P}(x - h_1 < \xi < x + h_2) = p_\xi(x)\,\Delta x + o(1)\int\limits_{x-h_1}^{x+h_2} dt =$$

$$= p_\xi(x)\,\Delta x + \Delta x\, o(1). \tag{3.20}$$

Следовательно, $p_\xi(x)\,dx$ есть дифференциал вероятности попадания случайной величины ξ в бесконечно малый интервал длины dx [в точках, где $p_\xi(x)$ непрерывна].

Заметим, что плотность распределения определяется неоднозначно: ее можно произвольным образом изменить в отдельных точках (более точно, на любом множестве лебеговой меры 0), не нарушив равенства (3.18). В то же время любые две плотности, удовлетворяющие этому равенству при всех x, $-\infty < x < \infty$, различаются лишь значениями на множестве лебеговой меры 0, т. е. на множестве, которое можно покрыть счетной совокупностью интервалов со сколь угодно малой суммарной длиной.

Простейшим примером непрерывного распределения является равномерное распределение, задаваемое плотностью вида

$$p(x) = \begin{cases} \dfrac{1}{b-a} & \text{при } x \in (a,\ b), \\ 0 & \text{при } x \,\overline{\in}\, (a,\ b), \end{cases} \tag{3.21}$$

где $(a,\ b)$ — интервал положительной длины. Интервал $(a,\ b)$ называется *интервалом концентрации равномерного распределения*. Значение плотности внутри этого интервала обратно пропорционально его длине. Плотность равномерного распределения изображена на рис. 3.1.

Важнейшее из распределений — нормальное распределение $N(a, \sigma)$, плотность которого задается формулой

$$p(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < \infty \tag{3.22}$$

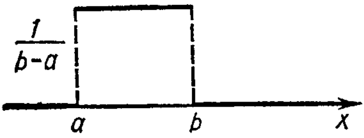

Рис. 3.1

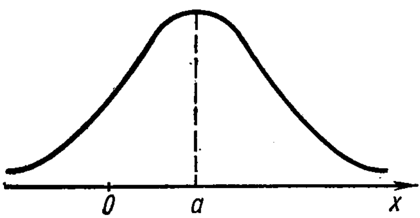

Рис. 3.2

также является непрерывным распределением (a и σ^2 — параметры нормального распределения, смысл которых будет выяснен несколько позже).

График функции $y = p(x)$ изображен на рис. 3.2. Это кривая Гаусса, уже встречавшаяся в гл. II, где она использовалась для аппроксимации вероятностей $p_{\nu_n}(k)$ для последовательности независимых испытаний. Эта кривая симметрична относительно прямой $x = a$.

Пусть случайная величина ξ имеет непрерывное распределение, причем $p_\xi(x) = 0$ при $x < 0$.

Рассмотрим функцию

$$\lambda_\xi(x) = \frac{p_\xi(x)}{1 - F_\xi(x)}, \qquad x > 0, \tag{3.23}$$

называемую *интенсивностью случайной величины*. Это название оправдывается следующим примером прикладного характера.

Пусть имеется элемент со случайным временем ξ безотказной работы, причем существует непрерывная плотность $p_\xi(x)$. Вычислим вероятность $\mathrm{P}(x, \Delta x)$ того, что элемент откажет в интервале времени $(x, x + \Delta x)$ при условии, что он проработал без отказа до момента x. Имеем

$$\mathrm{P}(x, \Delta x) = \mathrm{P}(x < \xi < x + \Delta x / \xi \geqslant x) = \frac{F_\xi(x + \Delta x) - F_\xi(x)}{1 - F_\xi(x)}$$

Далее,

$$\frac{P(x,\Delta x)}{\Delta x}=\frac{1}{1-F_\xi(x)}\;\frac{F_\xi(x+\Delta x)-F_\xi(x)}{\Delta x}\xrightarrow[\Delta x\to 0]{}\frac{p_\xi(x)}{1-F_\xi(x)}=$$

$$=\lambda_\xi(x).$$

Таким образом, $\lambda_\xi(x)\,dx$ есть вероятность попадания ξ в бесконечно малый интервал $(x, x+dx)$ при условии, что $\xi > x$. В теории надежности функция $\lambda_\xi(x)$ называется *интенсивностью отказа элемента.*

Пусть функция $p_\xi(x)$ непрерывна. Тогда равенство (3.23) можно записать в виде дифференциального уравнения

$$\frac{F'_\xi(x)}{1-F_\xi(x)}=\lambda_\xi(x). \tag{3.24}$$

Интегрируем это уравнение, считая, что $t > 0$:

$$\int\limits_0^t \frac{F'(x)\,dx}{1-F_\xi(x)}=-\ln[1-F_\xi(x)]\Big|_0^t=-\ln[1-F_\xi(t)]=$$

$$=\int\limits_0^t \lambda_\xi(x)\,dx.$$

Отсюда

$$1-F_\xi(t)=\exp\{-\int\limits_0^t\lambda_\xi(x)\,dx\},\quad x>0,$$

или

$$F_\xi(t)=1-\exp\{-\int\limits_0^t\lambda_\xi(x)\,dx\},\quad x>0. \tag{3.25}$$

Рассмотрим следующий пример. Пусть $\lambda_\xi(x)=\lambda$ при всех $x > 0$. Тогда из уравнения (3.25) находим

$$F_\xi(t)=1-e^{-\lambda t}. \tag{3.26}$$

Распределение случайной величины ξ, для которого

$$F_\xi(x)=\begin{cases}1-e^{-\lambda x} & \text{при } x>0,\\ 0 & \text{при } x\leqslant 0,\end{cases}$$

называется *показательным* (*экспоненциальным*). Постоянная $\lambda > 0$ называется *параметром показательного распределения.*

Плотность показательного распределения имеет вид экспоненты (рис. 3.3):

$$p_\xi(x) = F'_\xi(x) = \lambda e^{-\lambda x} \tag{3.27}$$

Распределение характеризуется тем, что при $x > 0$ и $t > 0$

$$P(\xi - x > t/\xi > x) = \frac{P(\xi > t + x)}{P(\xi > x)} = \frac{e^{-\lambda(t+x)}}{e^{-\lambda x}} = e^{-\lambda t}.$$

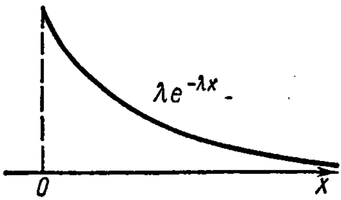

Рис. 3.3

Это означает, что если распределение времени безотказной работы какого-либо элемента — показательное, то распределение остатка времени безотказной работы не зависит от того, сколько времени элемент уже проработал (разумеется, если он до этого не отказал).

§ 5. ДИСКРЕТНЫЕ СМЕСИ РАСПРЕДЕЛЕНИЙ

Пусть имеется набор функций распределения $F_1(x)$, $F_2(x), \ldots, F_k(x), \ldots$. Пусть, далее, $p_1, p_2, \ldots, p_k \ldots$, — неотрицательные числа, в сумме равные 1. Определим функцию $F(x)$ с помощью формулы

$$F(x) = p_1F_1(x) + p_2F_2(x) + \ldots + p_kF_k(x) + \ldots . \tag{3.28}$$

Очевидно, что $F(x)$ — функция неубывающая как сумма ряда, составленного из неубывающих функций. Выберем теперь такое n, чтобы $\sum_{i=1}^{n} p_i > 1 - \frac{\varepsilon}{2}$, где $\varepsilon > 0$ — заранее фиксированное число. Можно записать:

$$F(x) = \sum_{i=1}^{n} p_iF_i(x) + R(x),$$

где

$$R(x) = \sum_{i=n+1}^{\infty} p_iF_i(x).$$

Поскольку $0 \leqslant F_i(x) \leqslant 1$, $i \geqslant 1$, то для любых x, $-\infty < x < \infty$, имеем

$$0 \leqslant \sum_{i=n+1}^{\infty} p_i F_i(x) \leqslant \sum_{i=n+1}^{\infty} p_i < \frac{\varepsilon}{2}.$$

Выберем теперь такое A, чтобы для всех i, $1 \leqslant i \leqslant n$, одновременно выполнялись неравенства

$$F_i(-A) < \frac{\varepsilon}{2}, \quad F_i(A) > 1 + \frac{\varepsilon}{2}.$$

Тогда

$$F(-A) \leqslant \sum_{i=1}^{n} p_i \frac{\varepsilon}{2} + R(-A) \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon;$$

$$F(A) \geqslant \sum_{i=1}^{n} p_i F_i(A) \geqslant \sum_{i=1}^{n} p_i \left(1 - \frac{\varepsilon}{2}\right) \geqslant \left(1 - \frac{\varepsilon}{2}\right)^2 \geqslant 1 - \varepsilon.$$

Так как $\varepsilon > 0$ произвольно, то заключаем, что

$$F(-\infty) = 0, \quad F(\infty) = 1. \tag{3.29}$$

Пусть теперь некоторая последовательность $x_j \uparrow x$, где x — фиксированное число, n и ε — те же, что и прежде. Так как функция $F_i(x)$ непрерывна слева, то найдется такое N, что при $j > N$ и $i = 1, 2, \ldots, n$

$$0 \leqslant F_i(x) - F_i(x_j) \leqslant \frac{\varepsilon}{2}$$

Тогда

$$0 \leqslant F(x) - F(x_j) = \sum_{i=1}^{\infty} p_i [F_i(x) - F_i(x_j)] \leqslant$$

$$\leqslant \sum_{i=1}^{n} p_i [F_i(x) - F_i(x_j)] + \sum_{i=n+1}^{\infty} p_i \leqslant$$

$$\leqslant \sum_{i=1}^{n} p_i \frac{\varepsilon}{2} + \sum_{i=n+1}^{\infty} p_i \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{3.30}$$

Отсюда вытекает, что $F(x)$ непрерывна слева. Таким образом, $F(x)$ — неубывающая, непрерывная слева функция, причем $F(-\infty) = 0$, $F(\infty) = 1$. Поэтому (см. п. 10, § 2) $F(x)$ можно рассматривать как функцию распределения некоторой случайной величины η.

Определение. *Функция распределения вида* $F(x) = \sum_{k=1}^{\infty} p_k F_k(x)$ *называется дискретной смесью функций рас-*

пределения $F_1(x), F_2(x), \ldots, F_k(x), \ldots$ *с весами* $p_1, p_2, \ldots, p_k, \ldots$. В частности, если $p_i = 0$ при всех $i > k$, смесь называется конечной. Пусть ν — случайная величина, принимающая значения $1, 2, \ldots, k, \ldots$ с вероятностями соответственно $p_1, p_2, \ldots, p_k, \ldots$, и пусть имеется набор случайных величин $\xi_1 = \xi_1(\omega), \xi_2 = \xi_2(\omega), \ldots, \xi_k = \xi_k(\omega), \ldots$, определенных на вероятностном пространстве $(\Omega, \mathcal{A}, \mathrm{P})$ и имеющих функции распределения $F_1(x), F_2(x), \ldots$ соответственно (причем для любых $i \geqslant 1$ и $A \in \mathcal{A}$ события $\{\nu = i\}$ и $\{\omega \in A\}$ независимы). Тогда случайная величина $\eta = \xi_\nu$ будет иметь распределение $F(x)$, задаваемое формулой (3.29). Действительно, по формуле полной вероятности,

$$\mathrm{P}(\eta < x) = \sum_{i=1}^{\infty} \mathrm{P}(\nu = i \quad \eta < x) = \sum_{i=1}^{\infty} \mathrm{P}(\nu = i,\ \xi_\nu < x) =$$

$$= \sum_{i=1}^{\infty} \mathrm{P}(\nu = i,\ \xi_i < x) = \sum_{i=1}^{\infty} \mathrm{P}(\nu = i)\,\mathrm{P}(\xi_i < x) =$$

$$= \sum_{i=1}^{\infty} p_i F_i(x).$$

Если существуют плотности распределений $F_1(x), F_2(x), \ldots, F_k(x)$, т. е.

$$F_i(x) = \int_{-\infty}^{x} p_i(t)\,dt, \tag{3.31}$$

то

$$F(x) = \sum_{i=1}^{\infty} p_i \int_{-\infty}^{x} p_i(t)\,dt = \int_{-\infty}^{x} \left(\sum_{i=1}^{\infty} p_i p_i(t) \right) dt;^{*} \tag{3.32}$$

последнее означает, что распределение $F(x)$ обладает плотностью

$$p(x) = \sum_{i=1}^{\infty} p_i p_i(x). \tag{3.33}$$

Пусть

$$E(x) = \begin{cases} 1 & \text{при } x > 0, \\ 0 & \text{при } x \leqslant 0. \end{cases}$$

---

* Обоснование возможности перестановки символов Σ и $\int$ дается теоремой Лебега из теории функций действительного переменного.

Тогда $E(x)$ является функцией распределения вырожденной случайной величины, принимающей с вероятностью 1 значение 0.

Подобным же образом при любом конечном τ

$$E(x-\tau)=\begin{cases}1 \text{ при } x>\tau,\\ 0 \text{ при } x\leqslant\tau\end{cases}$$

есть функция распределения вырожденной случайной величины, принимающей с вероятностью 1 значение τ.

Докажем, что если для некоторого конечного или счетного множества I $\sum_{i\in I} p_i=1$, $p_i\geqslant 0$, $i\in I$, а x_i — различные конечные числа, $i\in I$, то

$$F(x)=\sum_{i\in I} p_i E(x-x_i)$$

представляет собой функцию распределения дискретной случайной величины ξ с возможными значениями x_i, которые принимаются данной величиной с вероятностями p_i. Действительно,

$$\sum_{i\in I} p_i E(x-x_i)=\sum_{x_i<x} p_i=\mathrm{P}\{\xi<x\}=F_\xi(x).$$

Таким образом, $F(x)$ является смесью функций распределения вырожденных случайных величин.

Приведем практический пример, в котором используется понятие смеси распределений.

Пример. В ящике имеется 1000 радиодеталей, характеризующихся некоторым параметром ξ (например, величиной емкости для конденсатора). Допустим, что 700 деталей изготовлены на одном заводе и 300 — на другом. На заводах применяется различная технология; поэтому в первом случае распределение параметра ξ есть $F_1(x)$, во втором — $F_2(x)$. Допустим, что из ящика извлекается наудачу одна деталь. Тогда функция распределения параметра для этой детали составляет $0{,}7\,F_1(x)+0{,}3F_2(x)$.

Справедлива следующая теорема.

Теорема 3.1. *Пусть $F(x)$ — произвольная функция распределения. Тогда существуют функция дискретного распределения $F_1(x)$, функция непрерывного распределения $F_2(x)$ и функция сингулярного распределения $F_3(x)$, а также неотри-*

цательные постоянные p_1, p_2 и p_3, в сумме равные 1 и такие, что тождественно относительно x выполняется равенство

$$F(x) = p_1F_1(x) + p_2F_2(x) + p_3F_3(x); \qquad (3.34)$$

при этом p_1, p_2 и p_3 определяются однозначно. Если $p_i > 0$, то и $F_i(x)$ определяется однозначно ($i = 1, 2, 3$).

Доказательство этой теоремы требует знания аппарата теории меры и поэтому опускается.

Заметим, что если $F(x)$ — функция дискретного распределения, то $p_1 = 1$, $p_2 = p_3 = 0$. Подобным же образом для непрерывного распределения $p_2 = 1$, $p_1 = p_3 = 0$, для сингулярного $p_3 = 1$, $p_1 = p_2 = 0$. Дискретные, непрерывные и сингулярные распределения, т. е. все такие распределения, которым в представлении (3.34) соответствует только одно ненулевое p_i, называются *чистыми*, все остальные — *смешанными* распределениями.

Приведем пример смешанного распределения.

Представим себе техническое устройство, которое периодически используется для какой-либо операции в моменты $T, 2T, \ldots, nT, \ldots$. В начальный момент устройство исправно. Если устройство безотказно проработало до момента $t > 0$, то вероятность его отказа в интервале (t, dt) составляет λdt $(\lambda > 0)$ в том случае, если $t \neq nt$, $n = 0, 1, 2, \ldots$.

В момент nT устройство может отказать с вероятностью p_0, $0 < p_0 < 1$. Пусть ξ — момент отказа устройства, $p(t) = 1 - F_\xi(t) = \mathrm{P}\{\xi \geqslant t\}$. Рассмотрим два момента времени t и $t + dt$ $(dt > 0)$. Пусть $t \neq nT$, $n = 0, 1, 2, \ldots$. Можно записать

$$p(t + dt) = p(t)(1 - \lambda dt), \qquad (3.35)$$

откуда

$$p'(t) = -\lambda p(t). \qquad (3.36)$$

Это уравнение выполняется во всех интервалах $(nT, (n + 1)T)$ $(n \geqslant 0)$. Таким образом, имеем

$$p(t) = e^{-\lambda\left(t - T\left[\frac{t}{T}\right]\right)} p\left(T\left[\frac{t}{T}\right] + 0\right), \qquad (3.37)$$

где $[x]$ — целая часть числа x, т. е. $T\left[\frac{t}{T}\right]$ — последний момент использования устройства, предшествующий моменту t (если t не является кратным T).

Если t кратно T, то, по теореме умножения вероятностей,

$$p(t + 0) = p(t)(1 - p_0). \qquad (3.38)$$

Пусть $\left[\frac{t}{T}\right] = n$, $t - T\left[\frac{t}{T}\right] = \tau > 0$. Тогда, применяя последовательно формулы (3.37) и (3.38), получим

$$p(t)=e^{-\lambda t}p(nT+0)=e^{-\lambda\tau}(1-p_0)\,p(nT)=$$
$$=e^{-\lambda\tau}(1-p_0)\,e^{-\lambda T}p((n-1)\,T+0)=e^{-\lambda\tau}(1-p_0)\;e^{-\lambda T}(1-$$
$$-p_0)\;p((n-1)\,T)\ldots=e^{-\lambda(\tau+nT)}(1-p_0)^n=e^{-\lambda t}(1-p_0)^{\left[\frac{t}{T}\right]}. \tag{3.39}$$

Отсюда при t, не являющихся кратными T, $F_\xi(t)=1-e^{-\lambda t}(1-p_0)^{\left[\frac{t}{T}\right]}$. Поскольку $F_\xi(t)$ как функция распределения должна быть непрерывной, то получим

$$F_\xi(t)=1-e^{-\lambda t}(1-p_0)^{\left[\frac{t}{T}-0\right]},\quad t\geqslant 0. \tag{3.40}$$

Функция $F_\xi(t)$ изменяется непрерывно в интервалах $(nT,\,(n+1)T)$, $n\geqslant 0$. В точках $t=nT$, $n\geqslant 1$, эта функция имеет скачки:

$$F_\xi(nT+0)-F_\xi(nT)=\left[1-e^{-\lambda nT}(1-p_0)^n\right]-$$
$$-\left[1-e^{-\lambda nT}(1-p_0)^{n-1}\right]=(1-p_0)^{n-1}p_0e^{-\lambda nT}. \tag{3.41}$$

В интервалах между скачками функция $F_\xi(t)$ дифференцируема и ее производная равна $\lambda e^{-\lambda t}(1-p_0)^{\left[\frac{t}{T}\right]}$. Таким образом, если положить

$$p(t)=\begin{cases}\lambda e^{-\lambda t}(1-p_0)^{\left[\frac{t}{T}\right]} & \text{при } t>0,\\ 0 & \text{при } t<0,\end{cases} \tag{3.42}$$

то имеем

$$F(x)=\int_{-\infty}^{x}p(t)\,dt+\sum_{n=1}^{\infty}(1-p_0)^{n-1}p_0e^{-\lambda nT}E(x-nT), \tag{3.43}$$

т. е. $F_\xi(x)$ представлена в виде смеси функции непрерывного и функции дискретного распределения. Веса, соответствующие данным распределениям, соответственно равны

$$\int_{-\infty}^{\infty}p(t)\,dt=\frac{e^{\lambda T}-1}{e^{\lambda T}-1+p_0}$$

и

$$\sum_{n=1}^{\infty}(1-p_0)^{n-1}p_0e^{-\lambda nT}=\frac{p_0}{e^{\lambda T}-1+p_0},$$

[В формуле (3.34) эти величины обозначены через p_2 и p_1.]

§ 6. СХОДИМОСТЬ СЛУЧАЙНЫХ ВЕЛИЧИН И ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ

Пусть на вероятностном пространстве $(\Omega, \mathfrak{A}, P)$ задана последовательность случайных величин $\xi_n = \xi_n(\omega)$ $(n = 1, 2, \ldots)$ и случайная величина $\xi = \xi(\omega)$.

Определение. *Последовательность* $\{\xi_n\}$ *сходится к* ξ *с вероятностью* 1, *если множество тех* ω, *для каждого из которых числовая последовательность* $\{\xi_n(\omega)\}$ *не сходится к числу* $\xi(\omega)$, *имеет вероятность* 0.

Можно ли вообще говорить о вероятности события $\{\xi_n(\omega) \underset{n\to\infty}{\longrightarrow} \xi(\omega)\}$? Покажем, что такая вероятность всегда определена, для чего докажем несколько вспомогательных предложений.

1. Пусть ξ — случайная величина, a — постоянная. Тогда $\xi + a$ — случайная величина.

В самом деле, по определению случайной величины,

$$\{\xi + a < x\} = \{\xi < x - a\} \in \mathfrak{A}.$$

Далее, очевидно, $\{|\xi| < \infty\} = \{|\xi + a| < \infty\}$. Требуемое доказано.

2. Докажем, что если ξ и η — две случайные величины на вероятностном пространстве $(\Omega, \mathfrak{A}, P)$, то $\{\xi < \eta\}$ — случайное событие.

Если $\xi < \eta$, то найдется такая рациональная точка r, что $\xi < r < \eta$, и наоборот: если такая r существует, то $\xi < \eta$. Следовательно,

$$\{\xi < \eta\} = \bigcup_r \{\xi < r < \eta\} = \bigcup_r (\{\xi < r\} \cap \{r < \eta\}).$$

Так как $\{\xi < r\}$ и $\{r < \eta\}$ — случайные события, то их пересечение также есть случайное событие. Счетное же объединение случайных событий является случайным событием. Итак, $P(\xi < \eta)$ всегда определена, а следовательно, $\{\xi < \eta\}$ — случайное событие.

3. Докажем, что если $\varepsilon > 0$ — заданное число, а ξ и η — случайные величины, то $\{|\xi - \eta| < \varepsilon\}$ — случайное событие. Действительно,

$$\{|\xi - \eta| < \xi\} = \{-\varepsilon < \xi - \eta < \varepsilon\} = \\ = \{\eta < \xi + \varepsilon\} \cap \{\xi < \eta + \varepsilon\}.$$

4. Докажем следующее утверждение. Пусть $\varepsilon > 0$ — фиксированное число. Тогда событие $A_{N\varepsilon}$, состоящее в том, что $|\xi_n - \xi| < \varepsilon$ для всех $n \geqslant N$, есть случайное событие.

Это утверждение очевидно из соотношения

$$\{|\xi_n - \xi| < \varepsilon,\ n \geqslant N\} = \bigcap_{n=N}^{\infty} \{|\xi_n - \xi| < \varepsilon\}.$$

5. Пусть $\varepsilon > 0$ — фиксированное число. Пусть событие A_ε состоит в следующем. Найдется такое N, что для всех $n \geqslant N$, $|\xi_n - \xi| < \varepsilon$ A_ε есть случайное событие. Действительно,

$$A_\varepsilon = \bigcup_{N=1}^{\infty} A_{N\varepsilon}.$$

Пусть теперь A — событие, состоящее в том, что $\xi_n(\omega) \underset{n\to\infty}{\to} \xi(\omega)$ эквивалентно тому, что для некоторой последовательности $\varepsilon_n \to 0$ при любом $n \geqslant 1$ выполняется событие A_{ε_n}. Таким образом,

$$A = \bigcap_{n=1}^{\infty} A_{\varepsilon_n}.$$

Следовательно, $\{\xi_n(\omega) \underset{n\to\infty}{\to} \xi(\omega)\}$ — случайное событие, и определение сходимости с вероятностью 1 имеет логическое основание.

З а м е ч а н и е. Определение сходимости с вероятностью 1 распространяется и на тот случай, когда $\xi(\omega)$ — неслучайная величина, но $\{\omega : \xi(\omega) < x\} \in \mathcal{A}$ при любом x [т. е. $\xi(\omega)$ отличается от случайной величины тем, что может принимать значения $\pm\infty$ с положительной вероятностью]. В этом случае также обычным образом определено понятие сходимости числовой последовательности $\xi_n(\omega)$ к числу $\xi(\omega)$ (хотя бы и равному $\pm\infty$). Можно показать, что $\{\omega : \xi_n(\omega) \underset{n\to\infty}{\to} \xi(\omega)\}$ есть случайное событие.

Приведем пример сходимости с вероятностью 1. Пусть ξ — любая случайная величина, $\xi_n = \min\{\xi, n\}$. Тогда

$$P(\xi_n \underset{n\to\infty}{\longrightarrow} \xi) = 1 \tag{3.44}$$

Докажем справедливость соотношения (3.44). Имеем

$$P(\xi_n = \xi,\ n \geqslant N) = P\{\xi \leqslant N\}.$$

Вместе с тем если $\xi_n = \xi$, $n \geqslant N$, то, очевидно, $\xi_n \underset{n\to\infty}{\longrightarrow} \xi$.

Поэтому

$$P(\xi_n \xrightarrow[n\to\infty]{} \xi) \geqslant P(\xi \leqslant N) \geqslant F_\xi(N).$$

Так как $\lim\limits_{N\to\infty} F_\xi(N) = 1$, то $P(\xi_n \xrightarrow[n\to\infty]{} \xi) = 1$, что и требовалось доказать.

Докажем следующую теорему.

Теорема 3.2. *Если $\xi_n \to \xi$ с вероятностью 1, то $F_{\xi_n}(x) \xrightarrow[n\to\infty]{} F_\xi(x)$ в любой точке x, в которой $F_\xi(x)$ непрерывна.*

Доказательство. Коль скоро $\xi_n \to 1$ с вероятностью 1, для любого $\varepsilon > 0$

$$P(|\xi_n - \xi| < \varepsilon) \geqslant P(\bigcup_{N=1}^{n} A_{N\varepsilon}) \xrightarrow[n\to\infty]{} 1. \tag{3.45}$$

По формуле полной вероятности имеем

$$F_{\xi_n}(x) = P(\xi_n < x) = P(\xi_n < x,\ |\xi_n - \xi| < \varepsilon) + P(\xi_n < x,\ |\xi_n - \xi| \geqslant \varepsilon).$$

Очевидно,

$$P(\xi_n < x,\ |\xi_n - \xi| \geqslant \varepsilon) \xrightarrow[n\to\infty]{} 0.$$

Следовательно,

$$\lim_{n\to\infty} [F_{\xi_n}(x) - P(\xi_n < x,\ |\xi_n - \xi| < \varepsilon)] = 0.$$

Однако

$$\{\xi < x - \varepsilon,\ |\xi_n - \xi| < \varepsilon\} \subset \{\xi_n < x,\ |\xi_n - \xi| < \varepsilon\} \subset \subset \{\xi < x + \varepsilon,\ |\xi_n - \xi| < \varepsilon\}.$$

Таким образом,

$$P(\xi < x - \varepsilon,\ |\xi_n - \xi| < \varepsilon) \leqslant P(\xi_n < x,\ |\xi_n - \xi| < \varepsilon) \leqslant \leqslant P(\xi < x + \varepsilon,\ |\xi_n - \xi| < \varepsilon).$$

Средний член этого двойного неравенства, как было показано ранее, представляет собой величину, отличающуюся от $F_{\xi_n}(x)$ на величину бесконечно малую. Для крайних членов справедлива оценка

$$|P(\xi < x \pm \varepsilon),\ |\xi_n - \xi| < \varepsilon) - P(\xi < x \pm \varepsilon)| \leqslant$$

$$\leqslant P(|\xi_n - \xi| \geqslant \varepsilon),$$

или

$$P(\xi < x \pm \varepsilon, \quad |\xi_n - \xi| < \varepsilon) - F_\xi(x \pm \varepsilon) \underset{n\to\infty}{\longrightarrow} 0.$$

Отсюда находим, что для произвольно малого $\delta > 0$ при достаточно больших n

$$F_\xi(x - \varepsilon) - \delta \leqslant F_{\xi_n}(x) \leqslant F_\xi(x + \varepsilon) + \delta.$$

Если x — точка непрерывности функции $F_\xi(x)$, то разность между последним и первым членами этого двойного неравенства может быть сделана сколь угодно малой. Отсюда вытекает, что во всех точках непрерывности функции $F(x)$

$$F_{\xi_n}(x) \to F_\xi(x). \tag{3.46}$$

Будем говорить, что последовательность случайных величин $\xi_n(\omega)$, определенных на некотором вероятностном пространстве, является неубывающей, если для любых $m < n$

$$P(\xi_m \leqslant \xi_n) = 1, \tag{3.47}$$

или, что то же,

$$P(\xi_m > \xi_n) = 0. \tag{3.48}$$

При этом

$$P(\xi_1 \leqslant \xi_2 \leqslant \ldots \leqslant \xi_n \leqslant \ldots) = 1. \tag{3.49}$$

Для доказательства этого утверждения введем событие

$$A = \{\xi_1 \leqslant \xi_2 \leqslant \ldots \leqslant \xi_n \leqslant \ldots\}.$$

Событие, противоположное A, состоит в том, что хотя бы для одной пары индексов $m < n$ $\xi_m > \xi_n$. Таким образом,

$$\overline{A} = \bigcup_{m<n} \{\xi_m > \xi_n\} \quad \text{и} \quad P(\overline{A}) \leqslant \sum_{m<n} P(\xi_m > \xi_n),$$

но, по формуле (3.48), каждое из слагаемых последней суммы равно нулю; следовательно, и вся бесконечная сумма равна нулю. Утверждение доказано.

Теорема 3.3. *Если неубывающая последовательность случайных величин $\{\xi_n(\omega)\}$ сходится к случайной величине ξ с* *вероятностью* 1, *то в каждой точке непрерывности* $F_\xi(x)$

$$F_{\xi_n}(x) \downarrow F_\xi(x). \tag{3.50}$$

Выше было доказано, что $F_{\xi_n}(x) \to F_\xi(x)$. То, что последовательность $\{F_{\xi_n}(x)\}$ — убывающая, является следствием формулы (3.16).

Введем следующие определения.

Определение. *Последовательность функций распределения* $\{F_n(x)\}$, $n \geqslant 1$, *называется слабо сходящейся или сходящейся в основном к функции распределения* $F(x)$, *если в любой точке* x, *где функция* $F(x)$ *непрерывна,*

$$F_n(x) \xrightarrow[n\to\infty]{} F(x).$$

Определение. *Последовательность случайных величин* $\{\xi_n\}$, $n \geqslant 1$ *называется сходящейся к случайной величине* ξ *по распределению, если последовательность* $\{F_{\xi_n}(x)\}$ *слабо сходится к* $F_\xi(x)$.

С помощью введенного понятия сходимости по распределению теорему 2.3 можно сформулировать так.

Пусть ν_n — число успехов в n независимых испытаниях ξ — случайная величина с нормальным распределением:

$$F_\xi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz. \tag{3.51}$$

Введем последовательность случайных величин

$$\xi_n = \frac{\nu_n - np}{\sqrt{npq}}, \tag{3.52}$$

где p и q — соответственно вероятность успеха и неудачи в отдельном испытании. Тогда $\xi_n \xrightarrow[n\to\infty]{} \xi$ по распределению, т. е. последовательность $\{F_{\xi_n}(x)\}$ слабо сходится к $F_\xi(x)$*.

§ 7. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ СЛУЧАЙНОЙ ВЕЛИЧИНЫ ОБЩЕГО ВИДА

В гл. II было дано понятие математического ожидания дискретной случайной величины:

$$M\xi = \sum_{i \in I} x_i p_i, \tag{3.53}$$

* Заметим, однако, что в данной формулировке отсутствует свойство равномерности предельного соотношения.

где x_i — возможные значения случайной величины, p_i — вероятности, с которыми они принимаются, I — некоторое конечное или счетное множество индексов.

Напомним, что для существования математического ожидания дискретной случайной величины необходимо и достаточно, чтобы хотя бы одна из сумм $\sum\limits_{x_i>0} x_i p_i, \sum\limits_{x_i<0} x_i p_i$ была конечной. При этом если конечны обе эти суммы, то $M\xi$ конечно; если первая сумма бесконечна, а вторая — конечна, то $M\xi = \infty$; наконец, если первая из сумм конечна, а вторая бесконечна, то $M\xi = -\infty$. Формула для математического ожидания дискретной случайной величины имеет вид

$$M\xi = -\int_{-\infty}^{0} F_\xi(x)\,dx + \int_{0}^{\infty} [1 - F_\xi(x)]\,dx. \tag{3.54}$$

Математическое ожидание существует в том и только том случае, если хотя бы один из интегралов $\int\limits_{-\infty}^{0} F_\xi(x)\,dx$, $\int\limits_{\infty}^{0} [1 - F_\xi(x)]\,dx$ конечен.

Дадим понятие математического ожидания $M\xi$ случайной величины ξ общего вида.

Определение. *Пусть ξ — случайная величина, $\{\xi_n\}$ — последовательность дискретных случайных величин, удовлетворяющих условию*

$$P\{|\xi_n - \xi| > \varepsilon_n\} = 0, \quad n = 1, 2, \ldots,$$

где ε_n — бесконечно малая величина при $n \to \infty$. Тогда если для некоторого $n \geqslant 1$ существует $M\xi_n$, то

$$M\xi = \lim_{n\to\infty} M\xi_n. \tag{3.55}$$

Если же для некоторого n $M\xi_n$ не существует, то $M\xi$ также не существует.

Покажем, что в условиях данного определения либо все $M\xi_n$ существуют, либо все $M\xi_n$ не существуют.

Так как $\{\varepsilon_n\}$ — сходящаяся последовательность, то найдется такое число z, что $\varepsilon_n \leqslant z$ при всех n. Имеем

$$|\xi_n - \xi_m| \leqslant |\xi_n - \xi| + |\xi_m - \xi| \leqslant 2z.$$

Таким образом, $\xi_n-\xi_m$ — дискретная случайная величина, ограниченная по абсолютной величине числом $2z$. Следовательно, $-2z \leqslant M(\xi_n - \xi_m) \leqslant 2z$. Пусть $M\xi_m$ существует; тогда, по теореме о математическом ожидании суммы случайных величин, $M\xi_m = M(\xi_n-\xi_m) + M\xi_m$ также существует.

Покажем, что если $M\xi_m$ существует при некотором m, то $\lim\limits_{n\to\infty} M\xi_n$ существует и не зависит от выбора последовательности случайных величин $\{\xi_n\}$; необходимо лишь, чтобы выполнялось следующее условие: для некоторой последовательности $\varepsilon_n \underset{n\to\infty}{\to} 0$ $P(|\xi_n - \xi| > \varepsilon_n) = 0$. Если $M\xi_m = +\infty$ или $M\xi_m = -\infty$, то из последней формулы следует, что для всех n $M\xi_n = +\infty$ (соответственно $M\xi_n = -\infty$); поэтому предел существует. Если же для некоторого n (а следовательно, по доказанному, и для всех n) $|M\xi_n| < \infty$, то

$$|M\xi_n - M\xi_m| = |M(\xi_n - \xi_m)| \leqslant \varepsilon_n + \varepsilon_m,$$

поскольку

$$|\xi_n - \xi_m| \leqslant |\xi_n - \xi| + |\xi_m - \xi| \leqslant \varepsilon_n + \varepsilon_m.$$

Следовательно, числовая последовательность $\{M\xi_n\}$ фундаментальна: для любого $\varepsilon > 0$ найдется такое N, что $|M\xi_n - M\xi_m| < \varepsilon$ при $n \geqslant N$, $m \geqslant N$. Отсюда, по известной из анализа* теореме, вытекает, что последовательность $\{M\xi_n\}$ имеет предел. Докажем, что значение этого предела не зависит от выбора последовательности $\{\xi_n\}$. Пусть $M\xi_n \to a$, $M\xi'_n \to b$, $P(|\xi_n - \xi| > \varepsilon_n) = P(|\xi'_n - \xi| > \varepsilon'_n) = 0$, где $\varepsilon_n \to 0$, $\varepsilon'_n \to 0$ $(n \to \infty)$. Тогда $P(|\xi_n - \xi'_n| > \varepsilon_n + \varepsilon'_n) = 0$, откуда $|M\xi_n - M\xi'_n| \leqslant \varepsilon_n + \varepsilon'_n \underset{n\to\infty}{\to} 0$. В то же время при $a \neq b$ для любого $\delta < |a - b|$ было бы $|M\xi_n - M\xi'_n| > \delta$ при $n \geqslant N_0$. Следовательно, $a = b$, что и требовалось доказать.

* Если $\{x_n\}$ — числовая последовательность, причем $\lim\limits_{m,n\to\infty} |x_m - x_n| = 0$, то существует $\lim\limits_{n\to\infty} x_n = x^*$. Это доказывается следующим образом. Так как найдется такое N, что $|x_n - x_N| < 1$ при всех $n > N$, то $\{x_n\}$ — ограниченная последовательность. Значит, существует последовательность $\{m_k\}$, для которой $x_{m_k} \to x^*$. Поскольку $|x_n - x_{m_k}| < \varepsilon$ при $n \geqslant N$, $m_k \geqslant N$, то $|x_n - x^*| < \varepsilon$ при $n \geqslant N$. Это и означает, что $x_n \to x^*$ при $n \to \infty$.

Найдем выражение математического ожидания случайной величины общего вида через ее функцию распределения.

Пусть ξ — случайная величина с функцией распределения $F_\xi(x)$. Разобьем числовую прямую точками $\frac{k}{n}$, $k = 0, \pm 1, \pm 2, \ldots$ на отрезки равной длины $\frac{1}{n}$. Введем две случайные величины: ξ_n и ξ_n', где ξ_n — наибольшее из чисел $\frac{k}{n}$, не превосходящее ξ, $\xi_n' = \xi_n + \frac{1}{n}$. Так как $\xi_n \leqslant \xi < \xi_n'$, то имеет место неравенство

$$F_{\xi_n'}(x) \leqslant F_\xi(x) \leqslant F_{\xi_n}(x).$$

Отсюда

$$-\int_{-\infty}^{0} F_{\xi_n}(x)\,dx + \int_{0}^{\infty} [1 - F_{\xi_n}(x)]\,dx \leqslant -\int_{-\infty}^{0} F_\xi(x)\,dx +$$

$$+\int_{0}^{\infty} [1 - F_\xi(x)]\,dx \leqslant -\int_{-\infty}^{0} F_{\xi_n'}(x)\,dx + \int_{0}^{\infty} \left[1 - F_{\xi_n'}(x)\right]dx.$$

Поскольку крайние члены этого двойного неравенства при $n \to \infty$ сходятся к $M\xi$ (если оно существует), то

$$M\xi = -\int_{-\infty}^{0} F_\xi(x)\,dx + \int_{0}^{\infty} [1 - F_\xi(x)]\,dx. \qquad (3.56)$$

Легко видеть, что условием существования математического ожидания является конечность хотя бы одного из интегралов

$$\int_{-\infty}^{0} F_\xi(x)\,dx, \quad \int_{0}^{\infty} [1 - F_\xi(x)]\,dx.$$

Предположим, что случайная величина ξ непрерывна, т. е. $F_\xi(x) = \int_{-\infty}^{x} p_\xi(u)\,du$, где $p_\xi(x)$ — плотность данной случайной величины. Тогда, подставив это выражение $F_\xi(x)$ в формулу для математического ожидания и проинтегрировав по частям, получим

$$M\,\xi = \int\limits_{-\infty}^{\infty} x p_{\xi}(x)\,dx. \qquad (3.57)$$

Свойства математического ожидания в общем случае. Основные свойства математического ожидания, установленные нами для дискретных случайных величин, справедливы и для случайных величин общего вида.

Так, если $M\xi_1$ и $M\xi_2$ существуют и к тому же $c_1M\xi_1$ и $c_2M\xi_2$ не равны бесконечности противоположных знаков, то

$$M\,(c_1\xi_1 + c_2\xi_2) = c_1M\xi_1 + c_2M\xi_2.$$

Докажем это свойство. По заданному $\varepsilon > 0$ выберем такие дискретные случайные величины η_1 и η_2, что $|\eta_1 - \xi_1| \leqslant \varepsilon$, $|\eta_2 - \xi_2| \leqslant \varepsilon$ с вероятностью 1. Тогда

$$|(c_1\eta_1 + c_2\eta_2) - (c_1\xi_1 + c_2\xi_2)| \leqslant \varepsilon\,(|c_1| + |c_2|).$$

Таким образом,

$$|M\,(c_1\eta_1 + c_2\eta_2) - M\,(c_1\xi_1 + c_2\xi_2)| \leqslant \varepsilon\,(|c_1| + |c_2|),$$

$$|M\eta_1 - M\xi_1| \leqslant \varepsilon, \; |M\eta_2 - M\xi_2| \leqslant \varepsilon.$$

В то же время

$$M\,(c_1\eta_1 + c_2\eta_2) = c_1M\eta_1 + c_2M\eta_2.$$

Поэтому левая часть доказываемого равенства отличается от правой части не более чем на $2\varepsilon\,(|c_1| + |c_2|)$. Так как $\varepsilon > 0$ произвольно, то равенство доказано.

Определение. *Случайные величины* $\xi_1(\omega)$, $\xi_2(\omega)$, ..., $\xi_n(\omega)$ *называются независимыми, если для любых* x_1, x_2, ..., x_n

$$P\,(\xi_1(\omega) < x_1\, \xi_2(\omega) < x_2, \ldots, \xi_n(\omega) < x_n) =$$
$$= F_{\xi_1}(x_1)\, F_{\xi_2}(x_2) \ldots F_{\xi_n}(x_n).$$

Справедливо следующее предложение. Если $\varphi_1(x)$, $\varphi_2(x)$, ..., $\varphi_n(x)$ — функции для которых $|M\varphi_i(\xi_i)| < \infty$, то

$$M\,\varphi_1(\xi_1)\,\varphi_2(\xi_2)\ldots\varphi_n(\xi_n) = M\,\varphi_1(\xi_1)\,M\,\varphi_2(\xi_2)\ldots M\,\varphi_n(\xi_n) \qquad (3.58)$$

Доказательство этого предложения проводится аналогично доказательству теоремы о математическом ожидании $c_1\xi_1 + c_2\xi_2$.

§ 8. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ МОНОТОННОЙ ФУНКЦИИ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Вначале предположим, что случайная величина $\xi = \xi(\omega)$ задана на пространстве $\Omega = (0, 1)$, на котором вероятность $P(A)$ определена так, что $P(\omega \in \Delta) = |\Delta|$, где Δ — любой подынтервал интервала $(0, 1)$; $|\Delta|$ — длина Δ. Иначе можно сказать, что ω есть случайная величина с равномерным распределением в интервале $(0,1)$. Предположим, что $\xi(\omega)$ является неубывающей функцией ω. В этом случае можно связать $\xi(\omega)$ с функцией распределения случайной величины ξ. Временно предположим, что функция $\xi(\omega)$ строго возрастает в интервале $(0, 1)$. Тогда существует единственная функция $\omega = \omega(\xi)$, обратная к функции $\xi = \xi(\omega)$. Найдем функцию распределения случайной величины $\xi = \xi(\omega)$. Так как $\xi(\omega)$ — возрастающая функция, то $\xi(\omega) < x$ в том и только том случае, если $0<\omega<\omega(x)$. Здесь — равномерно распределенная в интервале $(0, 1)$ случайная величина, поэтому $P(0 < \omega < \omega(x)) = \omega(x)$; отсюда

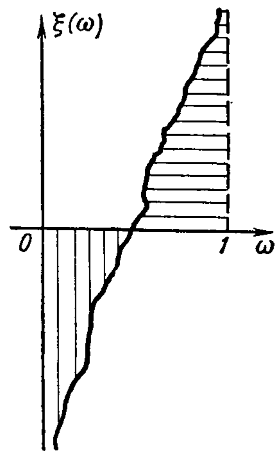

Рис. 3.4

$$P(\xi(\omega) < x) = \omega(x).$$

Следовательно, $\omega(x)$ является функцией распределения случайной величины $\xi(\omega)$. На рис. 3.4 разность между площадью, заштрихованной горизонтальными линиями, и площадью, заштрихованной вертикальными линиями, можно представить, с одной стороны, в виде

$$-\int_{-\infty}^{0} \omega(x)\,dx + \int_{0}^{\infty} [1-\omega(x)]\,dx,$$

а с другой стороны — как

$$\int_{0}^{1} \xi(\omega)\,d\omega.$$

Но первое выражение есть $M\xi$, так как $\omega(x) = F_{\xi}(x)$.

Итак,

$$M\xi = \int_0^1 \xi(\omega)\, d\omega. \tag{3.59}$$

Пусть теперь $\xi = \xi(\omega)$ — возрастающая функция ω, $a < \omega < b$, где $-\infty \leqslant a < b \leqslant \infty$; ω есть непрерывная случайная величина с плотностью $p(x)$, причем $\int_a^b p(x)\, dx = 1$ $p(x) > 0$ при $a < x < b$. Найдем вероятностный смысл

$$I = \int_a^b \xi(\omega)\, p(\omega)\, d\omega. \tag{3.60}$$

Введем новую переменную t по формуле

$$t = F(\omega), \tag{3.61}$$

где

$$F(\omega) = \int_{-\infty}^{\omega} p(u)\, du. \tag{3.62}$$

Поскольку $t = F(\omega)$ — строго возрастающая функция, то существует обратная к ней функция $\omega = F^{-1}(t)$. При этом $dt = F'(\omega)\, d\omega = p(\omega)\, d\omega$. Поэтому наш интеграл преобразуется к виду $I = \int_0^1 \xi(F^{-1}(t))\, dt$. Очевидно, $\xi(F^{-1}(t))$ — неубывающая функция. Если $t = t(\xi)$ — обратная к ней функция, то

$$\int_0^1 \xi(F^{-1}(t))\, dt = -\int_{-\infty}^{0} t(x)\, dx + \int_0^{\infty} [1 - t(x)]\, dx. \tag{3.63}$$

Найдем $t(x)$. Имеем

$$\xi(F^{-1}(t)) = x,$$

откуда

$$F^{-1}(t) = \xi^{-1}(x), \quad t = t(x) = F(\xi^{-1}(x)). \tag{3.64}$$

Однако $F(z)$ есть функция распределения случайной величины ω, следовательно,

$$F(\xi^{-1}(x)) = \mathrm{P}(\omega < \xi^{-1}(x)) = \mathrm{P}(\xi(\omega) < x) = F_{\xi(\omega)}(x). \tag{3.65}$$

Итак, доказана формула

$$\int_{-\infty}^{\infty} \xi(\omega) p(\omega) d\omega = M\xi(\omega), \tag{3.66}$$

где ω — случайная величина с плотностью $p(\omega)$. Однако метод доказательства требовал предположения, что функция $\xi(\omega)$ является возрастающей*.

Теория математического ожидания случайной величины, заданной на произвольном пространстве $(\Omega, \mathfrak{A}, P)$, может быть развита на основании теории абстрактного интеграла Лебега.

§ 9. ПОНЯТИЕ ОБ АБСТРАКТНОМ ИНТЕГРАЛЕ ЛЕБЕГА

Пусть имеется пространство Ω точек ω с заданной на нем σ-алгеброй подмножества $\mathfrak{A}$. *Мерой* μ на пространстве Ω называется функция множества $\mu(A)$, обладающая следующими свойствами:

1) $\mu(A)$ задана при любых $A \in \mathfrak{A}$;
2) $\mu(A) \geqslant 0$;
3) для любой последовательности $A_1, A_2, \ldots, A_n, \ldots$ непересекающихся множеств из $\mathfrak{A}$ справедливо равенство

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A_n). \tag{3.67}$$

Если $\mu(\Omega) < \infty$, то мера μ называется *конечной*.

Функция $f(\omega)$ с вещественными значениями, заданная при всех $\omega \in \Omega$, называется *измеримой*, если для любых $a < b$ множество ω, для которых $a \leqslant f(\omega) < b$, принадлежит σ-алгебре $\mathfrak{A}$.

Пусть μ — конечная мера, $f(\omega)$ — измеримая функция.

---

* Если $\xi(\omega) = \xi_1(\omega) - \xi_2(\omega)$, где $\xi_1(\omega)$ и $\xi_2(\omega)$ — монотонно возрастающие функции [что всегда возможно, если длина отрезка кривой $y = \xi(x)$, $|x| < A$, конечна при любом A], то

$$M\xi(\omega) = M\xi_1(\omega) - M\xi_2(\omega) = \int_{-\infty}^{\infty} \xi_1(\omega) p(\omega) d\omega -$$

$$- \int_{-\infty}^{\infty} \xi_2(\omega) p(\omega) d\omega = \int_{-\infty}^{\infty} \xi(\omega) p(\omega) d\omega,$$

если только $|M\xi_i(\omega)|$, $i = 1, 2$, — конечные числа.

Интеграл функции f по мере μ определяется следующим образом. Возьмем полузамкнутый интервал $[A, B)$ и разобьем его на n полузамкнутых подынтервалов $\Delta_1, \Delta_2, \ldots, \Delta_n$. Пусть h — максимальная из длин $\Delta_1, \Delta_2, \ldots, \Delta_n$. Образуем сумму

$$S_n = \sum_{i=1}^{n} y_i \mu(f^{-1}(\Delta_i)),$$

где $f^{-1}(\Delta_i)$ — множество тех ω, для которых $f(\omega) \in \Delta_i$, y_i — точка Δ_i. При $h \to 0$ существует предел S_n, который не зависит от выбора интервалов Δ_i и точек y_i. Докажем это утверждение.

Пусть $\Delta_1, \Delta_2, \ldots, \Delta_n$ — некоторое разбиение полуинтервала $[A, B)$, $\{\Delta_{i,j}\}$ — более мелкое разбиение, т. е. Δ_i разбивается на полуинтервалы $\Delta_{i,1}\Delta_{i,2}, \ldots, \Delta_{i,m_i}$. Пусть S_n — сумма, составленная по разбиению $\{\Delta_i\}$ и точкам $y_i \in \Delta_i$, S_N — сумма, составленная по разбиению $\{\Delta_{i,j}\}$ и точкам $y_{i,j} \in \Delta_{i,j}$ $(N = \sum_{i=1}^{n} m_i)$. Тогда, поскольку

$$f^{-1}(\Delta_i) = \bigcup_{j=1}^{m_i} f^{-1}(\Delta_{i,j}), \tag{3.68}$$

имеем

$$S_n = \sum_{i,j} y_i \mu(f^{-1}(\Delta_{i,j})), \tag{3.69}$$

$$S_N = \sum_{i,j} y_{i,j} \mu(f^{-1}(\Delta_{i,j})). \tag{3.70}$$

Отсюда

$$|S_n - S_N| = \left| \sum_{i,j} (y_i - y_{i,j}) \mu(f^{-1}(\Delta_{i,j})) \right| \leqslant$$

$$\leqslant \max_{i,j} |y_i - y_{i,j}| \sum_{i,j} \mu(f^{-1}(\Delta_{i,j})).$$

Однако y_i и $y_{i,j}$ принадлежит одному и тому же полуинтервалу Δ_i, откуда следует, что $|y_i - y_{i,j}| \leqslant h$. Далее,

$$\sum_{i,j} \mu(f^{-1}(\Delta_{i,j})) = \mu\Big(f^{-1}\Big(\bigcup_{i,j} \Delta_{i,j}\Big)\Big) \leqslant \mu(\Omega) < \infty.$$

Отсюда

$$|S_n - S_N| \leqslant h\mu(\Omega)$$

Если теперь $\{\Delta_i\}$ и $\{\Delta_i'\}$ — два разбиения $[A, B)$, для каждого из которых максимальная длина подынтервалов не превосходит h, то, взяв разбиение $\{\Delta_i''\}$, которое образуется совокупностью точек первого и второго разбиений, получим

$$|S_n - S_{n''}| \leqslant h\mu(\Omega), \qquad |S_{n'} - S_{n''}| \leqslant h\mu(\Omega),$$

где S_n, $S_{n'}$, $S_{n''}$ — суммы, соответствующие разбиениям $\{\Delta_i\}$, $\{\Delta_i'\}$, $\{\Delta_i''\}$, откуда следует

$$|S_n - S_{n'}| \leqslant 2h\mu(\Omega).$$

Итак, последовательность $\{S_n\}$ удовлетворяет свойству $S_n - S_m| \xrightarrow[m,n\to\infty]{} 0$, так что существует предел

$$\lim_{n\to\infty} S_n = I(A, B).$$

Если существует предел $\lim\limits_{\substack{A\to-\infty\\ B\to\infty}} I(A, B)$, то этот предел называется *интегралом Лебега функции* $f(\omega)$ *по мере* μ и обозначается так:

$$\int f(\omega)\,\mu(d\omega) = \int_\Omega f(\omega)\,\mu(d\omega). \qquad (3.71)$$

§ 10. ПРЕДСТАВЛЕНИЕ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ АБСТРАКТНЫМ ИНТЕГРАЛОМ ЛЕБЕГА

Пусть μ — вероятностная мера P, т. е. $\mu(A)$ есть вероятность попадания ω в множество A. Пусть, далее, $f(\omega) = \xi(\omega)$, где ξ — некоторая случайная величина. Предположим сначала, что для всех ω и некоторых A, B, $A \leqslant \xi(\omega) < B$. Возьмем Δ_i, y_i те же, что и в определении S_n, и определим случайную величину $\xi_n = \xi_n(\omega)$ следующей формулой:

$$\xi_n(\omega) = y_i, \text{ если } \omega \in \xi^{-1}(\Delta_i),\ 1 < i < n.$$

Тогда непосредственно очевидно, что

$$S_n = \sum_{i=1}^n y_i\,\mathrm{P}(\omega \in \xi^{-1}(\Delta_i)) = \sum_{i=1}^n y_i\,\mathrm{P}(\xi_n(\omega) = y_i) = M\xi_n.$$

Далее, если $\xi_n = y_i$, то $\omega \in \xi^{-1}(\Delta_i)$, т. е. $\xi(\omega) \in \Delta_i$. Поскольку y_i — точка Δ_i, то $|\xi_n - \xi| \leqslant h$. Отсюда $|M\xi_n - M\xi| \leqslant h$, а значит, и $|S_n - M\xi| \leqslant h$. Так как $S_n \xrightarrow[h\to 0]{}$

$I(A, B)$, то $I(A, B) = M\xi$. Перейдя в этом равенстве к пределу при $A \to -\infty$, $B \to \infty$, найдем, что

$$\int \xi(\omega)\, P(d\omega) = M\xi(\omega). \tag{3.72}$$

Данная формула без особого труда распространяется и на случай неограниченных случайных величин ξ.

Известно, что если $\Omega = (-\infty, \infty)$, а мера μ задается формулой

$$\mu(A) = \int\limits_A p(x)\, dx, \tag{3.73}$$

где $p(x)$ — некоторая интегрируемая функция, то

$$\int\limits_\Omega f(\omega)\, \mu(d\omega) = \int\limits_{-\infty}^{\infty} f(x)\, p(x)\, dx. \tag{3.74}$$

Отсюда можно сделать следующий вывод. Пусть ξ — случайная величина, обладающая плотностью $p_\xi(x)$, $f(x)$ — произвольная функция, для которой существует интеграл $\int\limits_{-\infty}^{\infty} f(x)\, p(x)\, dx$. Тогда

$$Mf(\xi) = \int\limits_{-\infty}^{\infty} f(x)\, p(x)\, dx. \tag{3.75}$$

если существуют интегралы

$$\int\limits_{x:\, f(x) > 0} f(x)\, p(x)\, dx, \qquad \int\limits_{x:\, f(x) < 0} f(x)\, p(x)\, dx$$

и хотя бы один из этих интегралов конечен.

ГЛАВА IV

НЕПРЕРЫВНЫЕ И НЕПРЕРЫВНО-ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§ 1. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

Любая случайная величина ξ характеризуется функцией распределения $F_\xi(x) = \mathrm{P}(\xi < x)$. $F_\xi(x)$ — неубывающая функция x, $-\infty < x < +\infty$, удовлетворяющая условиям

$$F_\xi(-\infty) = \lim_{x\to-\infty} F_\xi(x) = 0, \quad F_\xi(+\infty) = \lim_{x\to+\infty} F_\xi(x) = 1.$$

Напомним, что дискретная случайная величина ξ характеризуется множеством возможных значений $\{x_i\}$, $i \in I$, где I — некоторая конечная или бесконечная последовательность индексов, и вероятностями p_i, с которыми принимаются эти значения:

$$p_i = \mathrm{P}(\xi = x_i), \quad i \in I; \text{ при этом } \sum_{i\in I} p_i = 1.$$

Непрерывная случайная величина ξ — это такая случайная величина, функция распределения которой удовлетворяет следующему условию. Для любого x $(-\infty < x < +\infty)$

$$F_\xi(x) = \int_{-\infty}^{x} p_\xi(t)\,dt, \tag{4.1}$$

где $p_\xi(t)$ — неотрицательная функция, удовлетворяющая условию

$$\int_{-\infty}^{\infty} p_\xi(t)\,dt = 1 \tag{4.2}$$

и носящая название *плотности вероятности* (плотности распределения).

Для дискретных случайных величин также можно определить плотность $p_\xi(x)$, однако не как обычную функцию, а как линейную комбинацию δ-функций:

$$p_\xi(x) = \sum_{i \in I} p_i \delta(x - x_i), \tag{4.3}$$

где $\delta(x)$ (дельта-функция) — объект, для которого можно определить значение интеграла формулой

$$\int_{-\infty}^{\infty} f(x) \delta(x - \tau) dx = f(\tau), \tag{4.4}$$

какова бы ни была функция $f(x)$, непрерывная в точке $x = \tau$ (подробнее см. § 8 гл. II).

Определение. *Случайная величина ξ называется непрерывно-дискретной, если ее функция распределения $F_\xi(x)$ представима в виде*

$$F_\xi(x) = \int_{-\infty}^{x} p_\xi(t) dt, \tag{4.5}$$

где

$$p_\xi(x) = p_c(x) + p_d(x),$$

где в свою очередь $p_c(x)$ — обычная интегрируемая функция, $p_d(x)$ — линейная комбинация δ-функций;

$$p_d(x) = \sum_{i \in I} p_i \delta(x - x_i)^*. \tag{4.6}$$

При этом $p_i \geqslant 0$, $i \in I$; однако $0 \leqslant \sum_{i \in I} p_i \leqslant 1$. В частном случае, когда $\sum_{i \in I} p_i = 0$, непрерывно-дискретная случайная величина превращается в непрерывную, а в случае, когда эта сумма равна единице, — в дискретную. Заметим, что

$$\int_{-\infty}^{\infty} p_c(x) dx = 1 - \sum_{i \in I} p_i. \tag{4.7}$$

$p_c(x)$ называется *непрерывной компонентой* $p_\xi(x)$; $p_d(x)$ — *дискретной компонентой.*

---

* Говорят также, что ξ имеет непрерывно-дискретное распределение.

В настоящей главе будут изучаться только случайные величины данного класса; поэтому термин «случайная величина» следует понимать в смысле «непрерывно-дискретная случайная величина» (если не оговорено противное). Функцию $p_\xi(x)$ будем называть плотностью вероятности (плотностью распределения).

Условимся считать, что если ξ может принять с положительной вероятностью значение x, то $\int\limits_{-\infty}^{x} f(t)\, p_\xi(t)\, dt$ понимается как предел

$$\int\limits_{-\infty}^{x} f(t)\, p_\xi(t)\, dt = \lim_{z \uparrow x} \int\limits_{-\infty}^{z} f(t)\, p_\xi(t)\, dt.$$

В частности, положив $f(t) = 1$, найдем

$$\int\limits_{-\infty}^{x} p_\xi(t)\, dt = \lim_{z \uparrow x} \int\limits_{-\infty}^{z} p_\xi(t)\, dt = \mathrm{P}(\xi < x) = F_\xi(x).$$

При $x < y$ имеем

$$F_\xi(y) - F_\xi(x) = \int\limits_{-\infty}^{y} p_\xi(t)\, dt - \int\limits_{-\infty}^{x} p_\xi(t)\, dt = \int\limits_{x}^{y} p_\xi(t)\, dt.$$

С другой стороны,

$$F_\xi(y) - F_\xi(x) = \mathrm{P}(\xi < y) - \mathrm{P}(\xi < x) = \mathrm{P}(x \leqslant \xi < y).$$

Таким образом,

$$\mathrm{P}(x \leqslant \xi < y) = \int\limits_{x}^{y} p_\xi(t)\, dt. \tag{4.8}$$

Для непрерывных случайных величин справедливо равенство $\mathrm{P}(\xi = x) = 0$, откуда при $x \leqslant y$ имеем

$$\mathrm{P}(x \leqslant \xi \leqslant y) = \mathrm{P}(x \leqslant \xi < y) = \mathrm{P}(x < \xi \leqslant y) = \mathrm{P}(x < \xi < y) = \int\limits_{x}^{y} p_\xi(t)\, dt.$$

Математическое ожидание дискретной случайной величины определяется как $\sum\limits_{i \in I} x_i p_i$, если хотя бы одна из двух сумм $\sum\limits_{x_i > 0} x_i p_i$, $\sum\limits_{x_i < 0} x_i p_i$ конечна.

Математическое ожидание непрерывной случайной величины определяется формулой

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

если хотя бы один из двух интегралов $\int_{-\infty}^{0} x p_{\xi}(x)\, dx$, $\int_{0}^{\infty} x p_{\xi}(x) dx$ конечен.

Для непрерывно-дискретной случайной величины математическое ожидание определяется формулой

$$M\xi = \int_{-\infty}^{\infty} x p_{\xi}(x)\, dx, \tag{4.9}$$

где

$$\int_{-\infty}^{\infty} x p_{\xi}(x)\, dx = \int_{-\infty}^{\infty} x p_c(x)\, dx + \sum_{i \in I} x_i p_i.$$

$M\xi$ существует, если конечно хотя бы одно из следующих двух выражений:

$$\int_{-\infty}^{0} x p_{\xi}(x)\, dx = \int_{-\infty}^{0} x p_c(x)\, dx + \sum_{x_i < 0} x_i p_i,$$

$$\int_{0}^{\infty} x p_{\xi}(x)\, dx = \int_{0}^{\infty} x p_c(x)\, dx + \sum_{x_i > 0} x_i p_i.$$

Одним из основных свойств математического ожидания, доказанным в гл. II для дискретных случайных величин и затем перенесенного в гл. III на произвольные случайные величины, является равенство

$$M\,(c_1\xi_1 + c_2\xi_2) = c_1 M\xi_1 + c_2 M\xi_2. \tag{4.10}$$

Как было показано в гл. III, математическое ожидание любой случайной величины, если оно существует, определяется формулой

$$M\xi = -\int_{-\infty}^{0} F_{\xi}(x)\, dx + \int_{0}^{\infty} \left[1 - F_{\xi}(x)\right] dx. \tag{4.11}$$

Обратно, если хотя бы один из двух интегралов, состав ляющих правую часть формулы (4.11), конечен, то $M\xi$ существует и задается этой формулой.

Если $f(x)$ — произвольная непрерывная функция такая, что существует $Mf(\xi)$, то имеет место формула

$$Mf(\xi) = \int_{-\infty}^{\infty} f(x)\, p_\xi(x)\, dx. \tag{4.12}$$

Эта формула — следствие общей формулы (3.72). Мы, однако, докажем ее независимо от указанной формулы. Разобьем интервал $(-\infty, +\infty)$ на бесконечное множество полуинтервалов $\Delta_i = [a_i, b_i)$ таким образом, чтобы на каждом из них функция $f(x)$ изменялась не более, чем на заданное $\varepsilon > 0$: $|f(x) - f(y)| \leqslant \varepsilon$, если $a_i \leqslant x < y < b_i$. Определим функцию $\bar{f}(x)$ формулой $\bar{f}(x) = f(a_i)$, если $a_i \leqslant x < b_i$. Тогда

$$\left| \int_{-\infty}^{\infty} \bar{f}(x)\, p_\xi(x)\, dx - \int_{-\infty}^{\infty} f(x)\, p_\xi(x)\, dx \right| =$$

$$= \left| \int_{-\infty}^{\infty} [\bar{f}(x) - f(x)]\, p_\xi(x)\, dx \right| \leqslant \int_{-\infty}^{\infty} |\bar{f}(x) - f(x)|\, p_\xi(x)\, dx \leqslant$$

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

Определим далее дискретную случайную величину η формулой $\eta = \bar{f}(\xi)$. Тогда

$$M\eta = \sum_i f(a_i)\, \mathrm{P}(\eta = f(a_i)) = \sum_i f(a_i)\, \mathrm{P}(\xi \in \Delta_i) =$$

$$= \sum_i \int_{\Delta_i} f(a_i)\, p_\xi(x)\, dx = \int_{-\infty}^{\infty} \bar{f}(x)\, p_\xi(x)\, dx.$$

В сочетании с предыдущей эта формула приводит к неравенству

$$\left| M\eta - \int_{-\infty}^{\infty} f(x)\, p_\xi(x)\, dx \right| \leqslant \varepsilon.$$

Так как $|f(x) - \bar{f}(x)| \leqslant \varepsilon$, то $|f(\xi) - \eta| = |f(\xi) - \bar{f}(\xi)| \leqslant \varepsilon$.

Отсюда

$$|M[f(\xi)-\bar{f}(\xi)]| = \left|\int_{-\varepsilon}^{\varepsilon} x p_{f(\xi)-\bar{f}(\xi)}(x)\,dx\right| \leqslant \varepsilon,$$

а следовательно, $|Mf(\xi)-M\bar{f}(\xi)| \leqslant \varepsilon$, откуда

$$\left|Mf(\xi)-\int_{-\infty}^{\infty} f(x)\,p_{\xi}(x)\,dx\right| \leqslant |Mf(\xi)-M\bar{f}(\xi)| +$$

$$+\left|M\bar{f}(\xi)-\int_{-\infty}^{+\infty} f(x)\,p_{\xi}(x)\,dx\right| \leqslant \varepsilon+\varepsilon=2\varepsilon.$$

Ввиду того что $\varepsilon > 0$ — произвольное число, приходим к формуле (4.12).

§ 2. РАСПРЕДЕЛЕНИЕ ФУНКЦИЙ ОТ СЛУЧАЙНЫХ ВЕЛИЧИН

Пусть ξ — случайная величина с плотностью $p_{\xi}(x)$, $f(x)$ — некоторая функция. Образуем случайную величину вида

$$\eta = f(\xi).$$

Требуется найти распределение η.

Рассмотрим сначала простейший случай. Пусть $f(x) = 1$, если x принадлежит какому-либо из полуинтервалов $\Delta_1 = [a_1, b_1)$, $\Delta_2 = [a_2, b_2)$, ..., $\Delta_n = [a_n, b_n)$, где $-\infty \leqslant a_1 < b_1 < a_2 < b_2 < \ldots < a_n < b_n \leqslant +\infty$; $f(x) = 0$ — в противном случае. Тогда η — дискретная случайная величина, принимающая значения 1 и 0. Имеем

$$P(\eta=1) = P(f(\xi)=1) = P\left(\xi \in \bigcup_{i=1}^{n} [a_i, b_i)\right) =$$

$$= \sum_{i=1}^{n} P(\xi \in [a_i, b_i)) = \sum_{i=1}^{n} \left[F_{\xi}(b_i) - F_{\xi}(a_i)\right] =$$

$$= \sum_{i=1}^{n} \int_{a_i}^{b_i} p_{\xi}(x)\,dx.$$

Приведем следующий пример. Известно, что при определенных значениях параметров динамических систем может наступить резонанс, приводящий к разрушению системы. Пусть параметр ξ является случайным и имеет нормальное распределение с плотностью

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < +\infty,$$

а точки резонанса равны kh ($k = 0, \pm 1, \pm 2 \ldots$). Нужно найти вероятность того, что значение параметра удалено от точек резонанса более чем на ρ, где $\rho < \frac{h}{2}$. Обозначим искомую вероятность через p. Тогда

$$p = \mathrm{P}\left(\bigcup_{k=-\infty}^{\infty} \{kh + \rho < \xi < (k+1)h - \rho\}\right) =$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \sum_{k=-\infty}^{+\infty} \int_{kh+\rho}^{(k+1)h-\rho} e^{-\frac{(x-a)^2}{2\sigma^2}}\, dx.$$

Пусть теперь $f(x)$ — монотонно возрастающая непрерывно дифференцируемая функция; $\eta = f(\xi)$. Найдем распределение η. Имеем

$$F_\eta(x) = \mathrm{P}(f(\xi) < x) = \mathrm{P}(\xi < f^{-1}(x)), \tag{4.13}$$

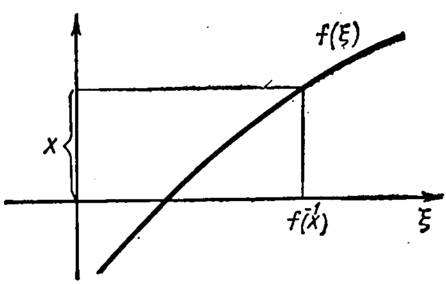

Рис. 4.1

где $f^{-1}(x)$ — функция, обратная к $f(x)$ (рис. 4.1). Далее,

$$F_\eta(x) = \int_{-\infty}^{f^{-1}(x)} p_\xi(t)\, dt. \tag{4.14}$$

Пусть $p_\xi(x) = p_c(x) + p_d(x)$. Тогда

$$\frac{d}{dx} \int\limits_{-\infty}^{f^{-1}(x)} p_c(t)\,dt = p_c(f^{-1}(x)) \frac{d}{dx} f^{-1}(x), \tag{4.15}$$

так как $f(x)$ — непрерывно дифференцируемая функция. Отсюда

$$\int\limits_{-\infty}^{f^{-1}(x)} p_c(t)\,dt = \int\limits_{-\infty}^{x} p_c(f^{-1}(t)) \frac{d}{dt} (f^{-1}(t))\,dt. \tag{4.16}$$

В то же время

$$\int\limits_{-\infty}^{f^{-1}(x)} p_d(t)\,dt = \sum_{x_i < f^{-1}(x)} p_i = \int\limits_{-\infty}^{f^{-1}(x)} \sum_{i \in I} p_i \delta(t - x_i)\,dt. \tag{4.17}$$

С помощью подстановки $t = f^{-1}(u)$ этот интеграл приводится к виду $\int\limits_{-\infty}^{x} \sum\limits_{i \in I} p_i \delta(f^{-1}(u) - x_i \frac{d}{du} f^{-1}(u))\,du$. Таким образом, случайная величина η является непрерывно-дискретной, и ее плотность

$$p_\eta(x) = \left[p_c(f^{-1}(x)) + p_d(f^{-1}(x))\right] \frac{d}{dx} (f^{-1}(x)) =$$

$$= p_\xi(f^{-1}(x)) \frac{d}{dx} (f^{-1}(x)). \tag{4.18}$$

Приведем пример применения этой формулы. Пусть ξ — нормальная случайная величина с плотностью $p_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $f(x) = x^3$. Требуется найти плотность вероятности случайной величины $\eta = f(\xi) = \xi^3$.

В данном случае

$$f^{-1}(x) = x^{1/3}, \quad \frac{d}{dx} (f^{-1}(x)) = \frac{1}{3} x^{-2/3};$$

поэтому

$$p_\eta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^{2/3}}{2}} \cdot \frac{1}{3} x^{-\frac{2}{3}}.$$

Пусть $f(x)$ — монотонно убывающая, непрерывно дифференцируемая функция. Найдем распределение $\eta = f(\xi)$. Обозначим $-f(\xi) = g(\xi)$. Тогда $g(\xi)$ будет монотонно возрастающей непрерывно дифференцируемой. Имеем

$$p_{g(\xi)}(x) = p_\xi(g^{-1}(x)) \frac{d}{dx}(g^{-1}(x)).$$

Очевидно, $p_{f(\xi)}(x) = p_{g(\xi)}(-x)$. С другой стороны, из равенства $-f(t) = g(t)$ следует, что $g^{-1}(x) = f^{-1}(-x)$; отсюда $\frac{d}{dx}(g^{-1}(x)) = -\frac{d}{dx}(f^{-1}(-x))$.

Окончательно

$$p_{f(\xi)}(x) = p_\xi(g^{-1}(-x)) \frac{d}{dx}\ (g^{-1}(-x)) =$$

$$= -p_\xi(f^{-1}(x)) \frac{d}{dx} f^{-1}(x) = p_\xi(f^{-1}(x)) \left| \frac{d}{dx} f^{-1}(x) \right|.$$

Оба случая — монотонно возрастающей и монотонно убывающей функции $f(x)$ — можно объединить одной формулой

$$p_{f(\xi)}(x) = p_\xi(f^{-1}(x)) \cdot \left| \frac{d}{dx} f^{-1}(x) \right|. \tag{4.19}$$

Рассмотрим, например, распределение случайной величины $\eta = \frac{1}{\xi}$, считая, что $P(\xi > 0) = 1$. Имеем

$$f(x) = \frac{1}{x}, \quad f^{-1}(x) = \frac{1}{x},$$

откуда

$$\left| \frac{d}{dx} f^{-1}(x) \right| = \left| -\frac{1}{x^2} \right| = \frac{1}{x^2} \ (x > 0).$$

Поэтому

$$p_\eta(x) = \frac{1}{x^2} p_\xi\left(\frac{1}{x}\right) \qquad x > 0. \tag{4.19'}$$

Отметим важный частный случай формулы (4.19). Пусть $\eta = a\xi + b$, где a и b — константы, причем $a \neq 0$. В данном случае

$$f(x) = ax + b, \quad f^{-1}(x) = \frac{x-b}{a}, \quad \frac{d}{dx} f^{-1}(x) = \frac{1}{a},$$

так что

$$p_{a\xi+b}(x) = \frac{1}{|a|} p_\xi\left(\frac{x-b}{a}\right) \tag{4.20}$$

Пусть теперь $f(x)$ имеет более сложный вид: в некоторых полуинтервалах Δ_i функция $f(x)$ постоянна:

$$f(x) = y_i, \quad x \in \Delta_i, \tag{4.21}$$

в некоторых полуинтервалах Δ_i' функция $f(x)$ монотонно возрастает и непрерывно дифференцируема; наконец, в некоторых полуинтервалах Δ_i'' функция $f(x)$ монотонно убывает и непрерывно дифференцируема. Число полуинтервалов каждого типа может быть как конечным, так и бесконечным. Полуинтервалы будем считать замкнутыми слева и открытыми справа, т. е. полуинтервал — множество точек x, для которых $a \leqslant x < b$ (в частности, может быть $a = -\infty$ или $b = \infty$).

Изучим распределение случайной величины $\eta = f(\xi)$. Рассмотрим событие $\{\eta < x\}$.

Возможны следующие случаи:

1) $\xi \in \Delta_i, \ y_i < x$;

2) $\xi \in \left(\bigcup \Delta_i'\right) \cup \left(\bigcup \Delta_i''\right); \quad \xi = x_i \quad i \in I; \quad f(x_i) < x$;

3) $\xi \in \left(\bigcup \Delta_i'\right) \cup \left(\bigcup \Delta_i''\right); \quad \xi \bar{\in} \{x_i, i \in I\}; f(\xi) < x$.

Поэтому, по формуле полной вероятности.

$$F_\eta(x) = \sum_i \int_{-\infty}^{x} \delta(t - y_i)\,dt \int_{\Delta_i} p_\xi(u)\,du + \sum_{\substack{j \in I \\ x_j \in \left(\cup \Delta_i'\right) \cup \left(\cup \Delta_i''\right)}} p_j \int_{-\infty}^{x} \delta(t -$$

$$- f(x_j))\,dt + \sum_i \int_{\substack{t \in \Delta_i' \\ f(t) < x}} p_c(t)\,dt + \sum_i \int_{\substack{t \in I \Delta_i'' \\ f(t) < x}} p_c(t)\,dt. \tag{4.22}$$

Обозначим функцию, обратную к $f(t)$ на полуинтервале Δ_i', через $a_i(t)$; на полуинтервале Δ_i'' — через $b_i(t)$. Таким образом, если $y = f(t)$, $t \in \Delta_i'$, то $t = a_i(y)$; если $y = f(t)$, $t \in \Delta_i''$, то $t = b_i(y)$. Поскольку на полуинтервалах $\Delta_i' = [a_i', b_i')$ функция $f(t)$ возрастает, то условие $f(t) < x$, $t \in \Delta_i''$ равносильно условию $a_i' \leqslant t < a_i(x)$. Аналогично, поскольку на полуинтервале $\Delta_i'' = [a_i'', b_i'')$ функция $f(t)$ убывает, то $f(t) < x$, $t \in \Delta_i''$ в том и только том случае, если $b_i(x) < t < b_i''$. Следовательно,

$$\int\limits_{\substack{t\in\Delta_i' \\ f(t)<x}} p_c(t)\,dt = \int\limits_{a_i'}^{a_i(x)} p_c(t)\,dt;$$

заменой $t = a_i(u)$ этот интеграл приводится к виду

$$\int\limits_{f(a_i')}^{x} p_c(a_i)(u))\,a_i'(u)\,du,$$

но только при $f(a_i') < x < f(b_i')$, в противном случае исходное выражение равно 0. Итак, можно записать

$$\int\limits_{\substack{t\in\Delta_i' \\ f(t)<x}} p_c(t)\,dt = \int\limits_{f(a_i')}^{f(b_i')} \delta(t-x)\,dt \int\limits_{f(a_i')}^{x} p_c(a_i(u))\,a_i'(u)\,du. \quad (4.23)$$

Аналогично имеем

$$\int\limits_{\substack{t\in\Delta_i'' \\ f(i)<x}} p_c(t)\,dt = -\int\limits_{f(b_i'')}^{f(a_i'')} \delta(t-x)\,dt \int\limits_{f(b_i'')}^{x} p_c(b_i(u))\,b_i'(u)\,du. \quad (4.24)$$

Окончательно приходим к следующему выводу. При сформулированных условиях случайная величина $\eta = f(\xi)$ является непрерывно-дискретной. Ее плотность $p_\eta(x)$ имеет дискретную компоненту

$$p_{d\eta}(x) = \sum_i \delta(t-y_i)\int\limits_{\Delta_i} p_\xi(u)\,du + \sum_{\substack{i\in I \\ x_i\in(\cup\Delta_i')\cup(\cup\Delta_i'')}} p_i\,\delta(x-f(x_i)) \quad (4.25)$$

и непрерывную компоненту

$$p_{c\eta}(x) = \sum_i p_c(a_i^{*}(x))\,a_i'(x)\int\limits_{f(a_i')}^{f(b_i')} \delta(t-x)\,dt -$$

$$-\sum_i p_c(b_i(x))\, b_i'(x)\cdot\int\limits_{f(b_i'')}^{f(a_i'')}\delta(t-x)\,dx. \qquad (4.26)$$

Рассмотрим следующие примеры.

1. Пусть ξ — случайная величина с плотностью $p_\xi(x)$, причем $\mathrm{P}(\xi=0)=0$. Требуется найти распределение случайной величины $|\xi|$.

Имеем $\Delta_1'=(0,\infty]$, $a_1(x)=x$, $(x>0)$, $\Delta_1''=(-\infty,0)$, $b_1(x)=-x$, $(x<0)$.

Отсюда при $x>0$

$$p_{|\xi|}(x)=p_\xi(x)\,a_i'(x)+p_\xi(-x)\,|b_1'(x)|=p_\xi(x)+p_\xi(-x).$$

Так, если ξ имеет нормальное распределение

$$p_\xi(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}(-\infty<x<+\infty),$$

то $|\xi|$ обладает плотностью

$$p_{|\xi|}(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\left[\exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}+\exp\left\{-\frac{(x+a)^2}{2\sigma^2}\right\}\right],\ x>0.$$

В частности, при $a=0$ получаем так называемое *сдвоенное нормальное распределение*:

$$p_{|\xi|}(x)=\frac{2}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{x^2}{2\sigma^2}\right\}, \qquad x>0.$$

При $x<0$ $\quad p_{|\xi|}(x)=0$.

2. Пусть случайная величина ξ распределена по показательному закону: $p_\xi(x)=\lambda e^{-\lambda x}$ при $x>0$; $p_\xi(x)=0$ при $x\leqslant 0$. Требуется найти распределение случайной величины $\eta=\sin\xi$. В данном случае $p_d(x)\equiv 0$; интервалов постоянства функции $f(x)=\sin x$ также не существует. Таким образом, $p_{d\eta}(x)\equiv 0$. Вычислим $p_{c\eta}(x)=p_\eta(x)$. Интервалы монотонности функции $\sin x$ таковы:

$$\Delta_i'=\left[2\pi i-\frac{\pi}{2},\ 2\pi i+\frac{\pi}{2}\right), \quad -\infty<i<\infty;$$

$$\Delta_i''=\left[2\pi i+\frac{\pi}{2},\ 2\pi i+\frac{3\pi}{2}\right), \quad -\infty\leqslant i<\infty.$$

Функции, обратные к $y = \sin x$, имеют вид $a_i(y) = \arcsin y + 2\pi i$, $b_i(y) = \pi - \arcsin y + 2\pi i$. Далее, $f(a'_i) = f(b''_i) = -1$, $f(b'_i) = f(a''_i) = 1$. Поэтому

$$\int\limits_{f(a'_i)}^{f(b'_i)} \delta(t - x)\,dt = \int\limits_{f(b''_i)}^{f(a''_i)} \delta(t - x)\,dt = 1,$$

если $0 < x < 1$; в противном случае эти интегралы равны 0.

Поскольку $a'_i(x) = -b'_i(x) = \dfrac{1}{\sqrt{1 - x^2}}$, то при $x \notin (-1, 1)$ $p_\eta(x) = 0$; при $x \in (-1, 1)$

$$p_\eta(x) = \sum_{i=-\infty}^{\infty} p_\xi(\arcsin x + 2\pi i)\frac{1}{\sqrt{1 - x^2}} + \sum_{i=-\infty}^{\infty} p_\xi(\pi - \arcsin y + 2\pi i)\frac{1}{\sqrt{1 - x^2}}. \quad (4.27)$$

При $i < 0$ общие члены обеих сумм равны 0; при $i > 1$ (а для второй суммы также при $i = 1$) они положительны. Член первой суммы, соответствующий случаю $i = 0$, положителен при $x > 0$ и равен 0 при $x \leqslant 0$. Окончательно имеем

$$p_\eta(x) = \int\limits_0^1 \delta(t - x)\,dt\,\lambda \exp\{-\lambda \arcsin x\}\frac{1}{\sqrt{1 - x^2}} + \int\limits_{-1}^{1} \delta(t - x)\,dt \left\{\sum_{i=1}^{\infty} \lambda \exp(\arcsin x + 2\pi i)\right\} + \left.\sum_{i=0}^{\infty} \lambda \exp\{-\lambda(\pi - \arcsin x + 2\pi i)\}\right\}\frac{1}{\sqrt{1 - x^2}}. \quad (4.28)$$

(Предлагаем читателю произвести дальнейшее упрощение самостоятельно, просуммировав геометрическую прогрессию со знаменателем $\exp\{-2\pi\lambda\}$.)

3. Пусть ξ — случайная величина с плотностью $p_\xi(x)$. Определим функцию $f(x)$ формулой

$$f(x) = \begin{cases} a, & \text{если } x < a, \\ x, & \text{если } a \leqslant x \leqslant b, \\ b, & \text{если } x > b. \end{cases} \tag{4.29}$$

Случайная величина $\eta = f(\xi)$ называется *срезкой* случайной величины ξ*. Срезки возникают в практических ситуациях, когда существуют некоторые физические ограничения возможных значений случайной величины. Найдем распределение η. По формулам (4.25), (4.26) или непосредственно получаем, что

$$p_\eta(x) = \int\limits_a^b \delta(t - x)\,dt \cdot p_\xi(x) + \delta(x - a) F_\xi(a) + \\ + \delta(x - b)\left[1 - F_\xi(b)\right]. \tag{4.30}$$

Так, если $p_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ $(-\infty < x < \infty)$, то

$$\mathrm{P}(\eta = a) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{a} e^{-\frac{t^2}{2}}\,dt,$$

$$\mathrm{P}(x < \eta < x + dx) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx. \quad (a < x < b), \tag{4.31}$$

$$\mathrm{P}(\eta = b) = \frac{1}{\sqrt{2\pi}} \int\limits_{b}^{\infty} e^{-\frac{x^2}{2}}\,dx.$$

Отметим еще одну операцию над случайной величиной [которую, впрочем, нельзя непосредственно выразить формулой $\eta = f(\xi)$] — усечение. Пусть $p_\xi(x)$ — плотность некоторого распределения. *Усечением данного распределения*

* Если $|a| < \infty$, $|b| < \infty$, срезку называют двусторонней; если $a = -\infty$ либо $b = \infty$ — односторонней.

называется распределение с плотностью

$$P_\eta(x) = \begin{cases} C \cdot p_\xi(x), & a < x < b, \\ 0 & , x \,\overline{\in}\, (a,b). \end{cases} \tag{4.32}$$

Постоянную C нужно определить так, чтобы $\int\limits_{-\infty}^{+\infty} p_\eta(x)dx = 1$. Таким образом,

$$C^{-1} = \int\limits_a^b p_\xi(x)\, dx^*. \tag{4.33}$$

Пример. Радиодетали, выпускаемые заводом, характеризуются некоторым основным параметром ξ (например, емкостью). Допустим, что этот параметр является случайной величиной с нормальным распределением

$$P_\xi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{\frac{(x-a)^2}{2\sigma^2}\right\}.$$

Продукция завода подвергается контролю, действующему по такому принципу. Если параметр детали находится в пределах (x_1, x_2), то деталь поставляется заказчику; если $\xi \,\overline{\in}\, (x_1, x_2)$, то деталь используется иным образом. Спрашивается: как распределено значение параметра в партии, поставленной заказчику?

Параметр является случайной величиной с распределением, отличным от $p_\xi(x)$; поэтому обозначим параметр в поставленной партии символом η. Имеем

$$\mathrm{P}(\eta \,\overline{\in}\, (x_1, x_2)) = 0; \quad \mathrm{P}(\eta < x) = \mathrm{P}(\xi < x / x_1 < \xi < x_2) =$$

$$= \frac{\mathrm{P}(\xi < x,\ x_1 < \xi < x_2)}{\mathrm{P}(x_1 < \xi < x_2)}.$$

При $x_1 < x < x_2$ числитель этой дроби равен $\mathrm{P}(\xi < x) - \mathrm{P}(\xi < x_1)$. Отсюда при $x_1 < x < x_2$ имеем

$$F_\eta(x) = \int\limits_{x_1}^{x} \frac{p_\xi(t)}{F_\xi(x_2) - F_\xi(x_1)}\, dt,$$

---

* Если $a = -\infty$, $b < \infty$ либо $a > -\infty$, $b = \infty$, усечение называется односторонним; если $-\infty < a$, $b < \infty$ — двусторонним.

откуда

$$p_\eta(x) = \frac{p_\xi(x)}{F_\xi(x_2) - F_\xi(x_1)} = \frac{e^{-\frac{(x-a)^2}{2\sigma^2}}}{\int\limits_{x_1}^{x_2} e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx}, \quad x_1 < x < x_2. \tag{4.34}$$

§ 3. НЕЗАВИСИМЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ФУНКЦИИ ОТ НИХ

Единственно возможный способ построения вероятностной модели реальных явлений состоит в том, что определенные события рассматриваются как независимые; исследователь пренебрегает возможной зависимостью случайных событий исходя из представлений о слабой физической связи явлений. Пусть, например, изучается среднее время простоя автомобилей при пересечении им перекрестка. Тогда прибытие автомобилей с различных сторон рассматривается как совокупность независимых событий. В то же время эти события физически связаны: если автомобиль подъезжает к перекрестку по одной улице, то *данный* автомобиль не может одновременно проезжать по другой улице или по той же, но в противоположном направлении.

Предположим, что на некотором вероятностном пространстве заданы n случайных величин: $\xi_1, \xi_2, \dots \xi_n$. Рассмотрим в n-мерном пространстве область D. Обозначим через $P(D)$ вероятность события $\{\xi \in D\}$, где $\xi = (\xi_1, \xi_2, \dots, \xi_n)$. Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ называются *независимыми в совокупности* в том и только том случае, если для любой области D

$$P(D) = \int\int \underset{D}{\dots} \int p_{\xi_1}(x_1)\, p_{\xi_2}(x_2) \dots$$

$$\dots p_{\xi_n}(x_n)\, dx_1\, dx_2 \dots dx_n. \tag{4.35}$$

Можно показать, что для независимости в совокупности $\xi_1, \xi_2, \dots, \xi_n$ необходимо и достаточно, чтобы для любых отрезков $\Delta_1, \Delta_2, \dots, \Delta_n$ выполнялось равенство

$$P(\xi_1 \in \Delta_1,\ \xi_2 \in \Delta_2, \dots, \xi_n \in \Delta_n) = \prod_{i=1}^{n} \int\limits_{\Delta_i} p_{\xi_i}(x)\, dx. \tag{4.36}$$

Рассмотрим следующий важный пример. Пусть случайные величины ξ_1, ξ_2 независимы и распределены по нормальному закону:

$$p_{\xi_1}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad p_{\xi_2}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-b)^2}{2\sigma^2}}.$$

Требуется найти вероятность $\mathrm{P}(r)$ того, что вектор (ξ_1, ξ_2) лежит в круге радиуса $r \geqslant 0$ с центром в точке (a, b).

Имеем

$$\mathrm{P}(r) = \mathrm{P}((\xi_1 - a)^2 + (\xi_2 - b)^2 \leqslant r^2) =$$

$$= \iint\limits_{(x_1-a)^2+(x_2-b)^2<r^2} p_{\xi_1}(x_1)\, p_{\xi_2}(x_2)\, dx_1\, dx_2 =$$

$$= \frac{1}{2\pi\sigma^2} \iint\limits_{(x_1-a)^2+(x_2-b)^2<r^2} \exp\left\{-\frac{1}{2\sigma^2}\left[(x_1-a)^2 + (x_2-b)^2\right]\right\} dx_1 dx_2.$$

Данный интеграл, по-видимому, проще всего вычислить, перейдя к полярным координатам: $x_1 - a = \rho\sigma \cos\varphi$, $x_2 - b = \rho\sigma \sin\varphi$. Тогда области $D = (x_1 - a)^2 + (x_2 - b)^2 \leqslant r^2$ в плоскости (x_1, x_2) в новых координатах будет соответствовать область $0 \leqslant \rho \leqslant \frac{r}{\sigma}$, $0 \leqslant \varphi \leqslant 2\pi$. Якобиан преобразования равен

$$\begin{vmatrix} \frac{\partial x_1}{\partial \rho} & \frac{\partial x_2}{\partial \rho} \\ \frac{\partial x_1}{\partial \varphi} & \frac{\partial x_2}{\partial \varphi} \end{vmatrix} = \rho\sigma^2.$$

Поэтому интеграл приводится к виду

$$\frac{1}{2\pi}\int\limits_0^{2\pi} d\varphi \int\limits_0^{r/\sigma} \exp\left\{-\frac{\rho^2}{2}\right\}\rho\, d\rho = \int\limits_0^{r/\sigma} \exp\left\{-\frac{\rho^2}{2}\right\}\rho\, d\rho =$$

$$= \int\limits_0^{r/\sigma} \exp\left\{-\frac{\rho^2}{2}\right\} d\left(\frac{\rho^2}{2}\right) = 1 - \exp\left\{-\frac{r^2}{2\sigma^2}\right\}.$$

Таким образом,

$$\mathrm{P}(r) = F_{\sqrt{\xi_1^2+\xi_2^2}}(r) = 1 - \exp\left\{-\frac{r^2}{2\sigma^2}\right\},$$

$$p_{\sqrt{\xi_1^2+\xi_2^2}}(r) = F'_{\sqrt{\xi_1^2+\xi_2^2}}(r) = \frac{r}{\sigma^2}\exp\left\{-\frac{r^2}{2\sigma^2}\right\}, \quad r > 0.$$

Положив $r^2 = x$, найдем

$$\mathrm{P}(\xi_1^2+\xi_2^2 < x) = F_{\xi_1^2+\xi_2^2}(x) = F_{\sqrt{\xi_1^2+\xi_2^2}}(\sqrt{x}) =$$

$$= 1 - \exp\left\{\frac{x}{2\sigma^2}\right\}, \quad x > 0.$$

Отметим следующий простой факт.

Лемма. *Пусть ξ_i — независимые случайные величины и определены случайные величины*

$$\eta_1 = f_1(\xi_{i_{11}}, \xi_{i_{12}}, \dots, \xi_{i_{1k_1}}), \quad \eta_2 = f_2(\xi_{i_{21}}, \xi_{i_{22}}, \dots, \xi_{i_{2k_2}}), \dots$$

$$\dots, \eta_n = f_m\ (\xi_{i_{m_1}}, \xi_{i_{m_2}}, \dots, \xi_{i_{m\,km}}),$$

где все индексы i_{jl} различны. Тогда $\eta_1, \eta_2, \dots, \eta_m$ независимы в совокупности.

Действительно,

$$\mathrm{P}(\eta_j \in \Delta_j,\ 1 \leqslant j \leqslant m) = \int\limits_D \dots \int p_{\xi_1}(x_1)\, p_{\xi_2}(x_2) \dots$$

$$\dots p_{\xi_n}(x_n)\, dx_1\, dx_2 \dots dx_n,$$

где область D задается условием

$$D = D_1 \cap D_2 \cap \dots \cap D_m, \quad D_j = \left\{f_j\left(\xi_{i_{j_1}} \xi_{i_{j_2}}, \dots, \xi_{i_{j_{k_j}}}\right) \in \Delta_j\right\},$$

$$1 \leqslant j \leqslant m.$$

Можно записать

$$\mathrm{P}(\eta_j \in \Delta_j,\ 1 \leqslant j \leqslant m) = \prod_{j=1}^{m} \int\limits_{D_j} \dots \int p_{\xi_{i_{j_1}}}\left(x_{i_{j_1}}\right) p_{\xi_{i_{j_2}}}\left(x_{i_{j_2}}\right) \dots$$

$$\dots p_{\xi_{i_{j_{k_j}}}}\left(x_{i_{j_{k_j}}}\right) dx_{i_{j_1}} \dots dx_{i_{j_{k_j}}} = \prod_{j=1}^{m} \mathrm{P}(\eta_j \in \Delta_j);$$

отсюда и вытекает в силу равенства (4.36) независимость в совокупности $\eta_1, \eta_2, \dots, \eta_m$.

В дальнейшем, как это обычно принято, мы будем для краткости говорить вместо «случайные величины $\xi_1, \xi_2, \dots, \xi_n$ независимы в совокупности» просто «случайные величины $\xi_1, \dots, \xi_n$ независимы».

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины с плотностями $p_{\xi_1}(x), p_{\xi_2}(x), \ldots, p_{\xi_n}(x)$. Наиболее важной функцией $\xi_1, \xi_2, \ldots, \xi_n$ является сумма

$$S_n = \xi_1 + \xi_2 + \ldots + \xi_n.$$

Поскольку в силу леммы S_{n-1} и ξ_n независимы, то достаточно вывести формулу для распределения суммы двух независимых случайных величин. Итак, пусть $\eta = \xi_1 + \xi_2$, где ξ_1 и ξ_2 — независимые случайные величины. По формуле (4.35) имеем

$$F_\eta(z) = p(\xi_1 + \xi_2 < z) = \iint\limits_{x_1+x_2<z} p_{\xi_1}(x_1)\, p_{\xi_2}(x_2) dx_1 dx_2. \quad (4.37)$$

Запишем

$$p_{\xi_j}(x) = p_{c\xi_j}(x) + \sum_{i\in I} p_{ji}\,\delta(x - x_{ji}), \quad j = 1,2.$$

Интеграл (4.37) можно представить следующим образом:

$$\begin{aligned} F_\eta(z) = &\iint\limits_{x_1+x_2<z} p_{c\xi_1}(x_1)\, p_{c\xi_2}(x_2)\, dx_1\, dx_2 + \\ &+ \sum_{i\in I} p_{1i} \iint\limits_{x_1+x_2<z} \delta(x_1 - x_{1i})\, p_{c\xi_2}(x_2)\, dx_1\, dx_2 + \\ &+ \sum_{i\in I} p_{2i} \iint\limits_{x_1+x_2<z} \delta(x_2 - x_{2i})\, p_{c\xi_1}(x_1)\, dx_1\, dx_2 + \\ &+ \sum_{i\in I}\sum_{k\in I} p_{1i}\, p_{2k} \iint\limits_{x_1+x_2<z} \delta(x_1 - x_{1i})\,\delta(x_2 - x_{2k})\, dx_1\, dx_2. \end{aligned} \quad (4.38)$$

Если $f_1(x)$ и $f_2(x)$ — обычные функции или δ-функции, причем $\int |f_j(x)|\, dx < \infty$, $j = 1, 2$, то можно записать

$$\iint\limits_{x_1+x_2<z} f_1(x_1)\, f_2(x_2)\, dx_1\, dx_2 = \int_{-\infty}^{z} dx \int_{-\infty}^{\infty} f_1(t)\, f_2(x-t)\, dt =$$

$$= \int_{-\infty}^{z} dx \int_{-\infty}^{\infty} f_1(x-t)\, f_2(t) dt.$$

Выражение $\int_{-\infty}^{\infty} f_1(t)\, f_2(x-t)\, dt$ *как функция* x *называется* *сверткой функций f_1 и f_2 и обозначается $f_1 * f_2$:*

$$(f_1 * f_2)(x) = \int_{-\infty}^{\infty} f_1(t) f_2(x-t)\,dt \qquad (4.39)$$

Свертка обычной функции f с функцией $\delta_\tau(x) = \delta\,(x-\tau)$ определяется равенством

$$(\delta_\tau * f)(x) = (f * \delta_\tau)(x) = \int_{-\infty}^{\infty} f(t)\,\delta(x - \tau - t)\,dt = f(x-\tau). \qquad (4.40)$$

Наконец, свертка $\delta_\tau * \delta_{\tau'}$ определяется формулой

$$(\delta_\tau * \sigma_{\tau'})(x) = \int_{-\infty}^{\infty} \delta(t-\tau)\delta(x-t-\tau')dt = \delta(x-\tau-\tau'). \qquad (4.41)$$

Свертка удовлетворяет соотношениям

$$(c_1 f_1 + c_2 f_2) * (c_3 f_3 + c_4 f_4) = c_1 c_3 f_1 * f_3 + c_2 c_3 f_2 * f_3 + c_1 c_4 f_1 * f_4 + c_2 c_4 f_2 * f_4. \qquad (4.42)$$

Таким образом, формулу (4.38) можно переписать так:

$$F_{\xi_1+\xi_2}(z) = \int_{-\infty}^{z} \left\{ \int_{-\infty}^{\infty} p_{\xi_1}(t)\, p_{\xi_2}(x-t)\,dt \right\} dx. \qquad (4.43)$$

Последнее означает, что $\xi_1 + \xi_2$ — непрерывно-дискретная случайная величина и что ее плотность

$$p_{\xi_1+\xi_2}(x) = \int_{-\infty}^{\infty} p_{\xi_1}(t)\, p_{\xi_2}(x-t)\,dt = (p_{\xi_1} * p_{\xi_2})(x). \qquad (4.44)$$

Теперь, воспользовавшись леммой, можно заключить, что

$$p_{S_n}(x) = p_{\xi_1+\xi_2+\ldots+\xi_n}(x) = (p_{\xi_1} * p_{\xi_2} * \ldots * p_{\xi_n})(x) \qquad (4.45)$$

где $p_{\xi_1} * p_{\xi_2} * \ldots * p_{\xi_n}$ при $n = 2$ уже определено, а при $n \geqslant 3$ определяется рекуррентной формулой

$$p_{\xi_1} * p_{\xi_2} * \ldots * p_{\xi_n} = (p_{\xi_1} * p_{\xi_2} * \ldots * p_{\xi_{n-1}}) * p_{\xi_n}.$$

Свертку функций f_1, f_2 можно определять приведенной выше формулой и в том случае, если одна из функций

f_1, f_2 не интегрируема в бесконечном интервале, но ограничена.

Поскольку

$$\iint\limits_{x_1+x_2<z} f_1(x_1)\, f_2(x_2)\, dx_1\, dx_2 = \int\limits_{-\infty}^{\infty} \left\{ \int\limits_{-\infty}^{x_2} f_1(x_1)\, dx_1 \right\} \times$$

$$\times\, f_2(z - x_2)\, dx_2, \qquad (4.46)$$

т. е. это выражение равно свертке функций $\int\limits_{-\infty}^{x} f_1(u)\, du$ и $f_2(x)$, то формулу (4.38) можно переписать еще и так:

$$F_{\xi_1+\xi_2}(x) = \left(F_{\xi_1} * f_{\xi_2}\right)(x) = \left(f_{\xi_1} * F_{\xi_2}\right)(x). \qquad (4.47)$$

Следовательно, функция распределения суммы независимых случайных величин равна свертке функции распределения одной из этих величин с плотностью другой. По индукции получаем, что для любого $n \geqslant 1$

$$F_{\xi_1+\xi_2} + \ldots + \xi_n(x) = \left(F_{\xi_1} * f_{\xi_2} * \ldots * f_{\xi_n}\right)(x). \quad (4.48)$$

Приведем примеры вычисления распределения суммы независимых случайных величин.

1. Пусть ξ_1, ξ_2 независимы и равномерно распределены в интервале $(0, 1)$. Требуется найти плотность вероятности $\eta = \xi_1 + \xi_2$.

Имеем

$$p_{\xi_1}(x) = p_{\xi_2}(x) = \begin{cases} 1 \text{ при } x \in (0, 1), \\ 0 \text{ при } x \,\overline{\in}\, (0, 1). \end{cases} \qquad (4.49)$$

Поэтому

$$p_{\xi_1+\xi_2}(x) = \int\limits_{-\infty}^{+\infty} p_{\xi_1}(t)\, p_{\xi_2}(x - t)\, dt = \int\limits_{\substack{0<t<1 \\ 0<x-t<1}} 1 \cdot dt.$$

Множество точек $\{0 < t < 1,\ 0 < x - t < 1\}$ представляет собой параллелограмм в плоскости (t, x) (рис. 4.2). Искомый интеграл есть длина отрезка, отсекаемого сторонами параллелограмма от прямой $x = \text{const}$. При $0 \leqslant x \leqslant 1$ эта длина равна x, а при $1 \leqslant x \leqslant 2$ она составляет $2 - x$. Таким образом,

$$p_{\xi_1+\xi_2}(x) = \begin{cases} x \text{ при } 0 \leqslant x \leqslant 1, \\ 2 - x \text{ при } 1 \leqslant x \leqslant 2, \\ 0 \text{ при } x \,\overline{\in}\, (0,2). \end{cases} \qquad (4.50)$$

Вспомнив формулу (4.20), приходим к следующему выводу. Если ξ_1 и ξ_2 — случайные величины, равномерно распределенные в интервалах $(a, a+h)$ и $(b, b+h)$, то $\xi_1+\xi_2$ обладает плотностью вида

$$p_{\xi_1+\xi_2}(x)=\begin{cases}\frac{1}{h^2}(h-|a+b+h-x|) & \text{при } |a+b+h-x|\leqslant h,\\ 0 & \text{при } |a+b+h-x|>h.\end{cases}\tag{4.51}$$

График плотности (4.50) изображен на рис. 4.3.

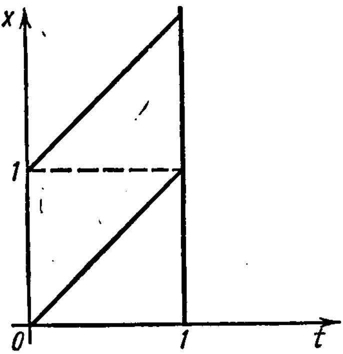

Рис. 4.2

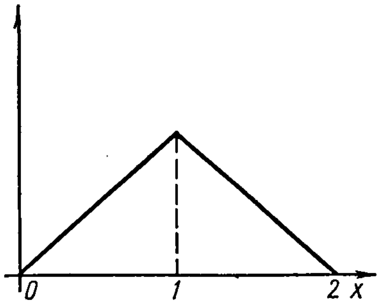

Рис. 4.3

2. Весьма часто плотности распределения, выводимые из эмпирических данных, задают в виде кусочно-постоянных функций (гистограмм). Пусть ξ_1 и ξ_2 — независимые случайные величины с плотностями:

$$p_{\xi_1}(x)=\begin{cases}\frac{1}{h}\cdot p_k & \text{при } a+(k-1)h<x<a+kh,\ 1\leqslant k\leqslant m,\\ 0 & \text{при } x\,\overline{\in}\,(a, a+mh),\end{cases}\tag{4.52}$$

где $p_1+p_2+\ldots+p_m=1$, $p_k\geqslant 0$, $1\leqslant k\leqslant m$; (рис. 4.4)

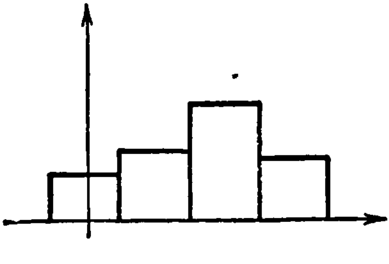

Рис. 4.4

$$p_{\xi_2}(x)=\begin{cases}\frac{1}{h}\cdot p'_k \text{ при } b+(k-1)h<x<b+kh,\ 1\leqslant k\leqslant n,\\ 0 \qquad \text{при } x\,\overline{\in}\,(b,\ b+nh),\end{cases} \tag{4.53}$$

где $p'_1+p'_2+\ldots+p'_n=1$.

Обозначим через $p_\tau(x)$ плотность равномерного распределения в интервале $(\tau,\ \tau+h)$, через $f_\tau(x)$ — плотность вида (4.51) при $a+b=\tau$. Тогда можно записать

$$\begin{gathered} p_{\xi_1}(x)=\sum_{k=1}^{m} p_k\, p_{a+(k-1)h}(x),\\ p_{\xi_2}(x)=\sum_{j=1}^{n} p'_j\, p_{b+(j-1)h}(x), \end{gathered} \tag{4.54}$$

откуда вследствие линейности операции свертки [формула (4.42)] имеем

$$\begin{aligned} p_{\xi_1+\xi_2}(x) &= \sum_{k=1}^{m}\sum_{j=1}^{n} p_k\, p_j\left(p_{a+(k-1)h} * p_{b+(j-1)h}\right)(x)=\\ &= \sum_{k=1}^{m}\sum_{j=1}^{n} p_k\, p'_j\, f_{a+b+(k+j-2)h}(x)=\\ &= \sum_{i=0}^{m+n-2} f_{a+b+ih}(x)\cdot\sum_{k+j=i} p_k\, p'_j=\sum_{i=0}^{m+n-2}\lambda_i\, f_{a+b+ih}, \end{aligned} \tag{4.55}$$

где $\lambda_i=\sum\limits_{k+j=i} p_k p'_j$. Таким образом, $p_{\xi_1+\xi_2}(x)$ есть сумма функций вида (4.51) (рис. 4.3). В точках $x=a+b+(i+1)\,h$ $p_{\xi_1+\xi_2}(x)=\frac{1}{h}\lambda_i$; в точках $x=a+b$ и $x=a+b+(m+n)\,h$ $p_{\xi_1+\xi_2}(x)=0$; в интервалах между перечисленными точками $p_{\xi_1+\xi_2}(x)$ линейна. [График функции $p_{\xi_1+\xi_2}(x)$ изображен на рис. 4.5.]

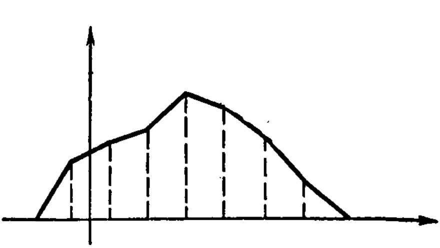

Рис. 4.5

Замечание. Если ξ_1 и ξ_2 — неотрицательные независимые случайные ве-

личины ($P(\xi_1 \geqslant 0) = P(\xi_2 \geqslant 0) = 1$),
то

$$p_{\xi_1+\xi_2}(x) = \int_0^x p_{\xi_1}(t)\, p_{\xi_2}(x-t)\, dt. \tag{4.56}$$

В самом деле, при $t < 0$ $p_{\xi_1}(t) = 0$, а при $t > x$ $p_{\xi_2}(x-t) = 0$. Рассмотрим пример с неотрицательными случайными величинами.

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые показательно распределенные случайные величины:

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

где λ — постоянный параметр. Требуется найти распределение $s_n = \xi_1 + \xi_2 + \ldots + \xi_n$.

Пусть $n = 2$. Имеем

$$p_{s_2}(x) = \int_0^x \lambda e^{-\lambda t} \lambda \cdot e^{-\lambda(x-t)}\, dt = \lambda^2 e^{-\lambda x} \int_0^x dt =$$

$$= \lambda^2 x e^{-\lambda x}, \quad x > 0.$$

Предположим, что для некоторого $n \geqslant 1$ справедлива формула

$$p_{s_n}(x) = \lambda \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x}, \quad x > 0 \tag{4.57}$$

Докажем, что она справедлива также для $n + 1$. По формуле (4.56) имеем

$$p_{s_{n+1}}(x) = \lambda^2 \int_0^x \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t - \lambda x + \lambda t}\, dt = \frac{\lambda^{n+1}}{(n-1)!} \times$$

$$\times e^{-\lambda x} \int_0^x t^{n-1}\, dt = \lambda \frac{(\lambda x)^n}{h!} e^{-\lambda x}.$$

Требуемое доказано. Итак, формула (4.57) выполняется при всех $n = 1, 2, \ldots$. Распределение с плотностью вида (4.57) называется *распределением Эрланга* n-го порядка.

В начале настоящего параграфа была приведена формула для вероятности попадания $(\xi_1, \xi_2, \ldots, \xi_n)$, где ξ_i —

независимые случайные величины, в область D:

$$P(\xi \in D) = \int \dots \int\limits_{D} p_{\xi_1}(x_1)\, p_{\xi_2}(x_2) \dots p_{\xi_n}(x_n)\, dx_1 dx_2 \dots dx_n.$$

Эту формулу можно интерпретировать и несколько иначе. Любую область можно представить в виде $f(x_1, \dots x_n) \in \Delta$, где $f(x_1, \dots, x_n)$ — некоторая непрерывная вещественная функция, Δ — некоторый отрезок*. Обратно, для любой непрерывной функции $f(x_1, x_2, \dots, x_n)$ и любого отрезка Δ множество тех $(x_1, x_2, \dots, x_n)$, для которых $f(x_1, x_2, \dots, x_n) \in \Delta$, представляет собой некоторую область n-мерного пространства. Значит,

$$P((\xi_1, \xi_2, \dots, \xi_n) \in D) = P(f(\xi_1, \xi_2, \dots \xi_n) \in \Delta) =$$
$$= \int \dots \int\limits_{f(x_1, x_2, \dots, x_n) \in \Delta} p_{\xi_1}(x_1)\, p_{\xi_2}(x_2) \dots p_{\xi_n}(x_n)\, dx_1 dx_2 \dots dx_n. \tag{4.58}$$

Возьмем произвольные m, $1 \leqslant m < n$; тогда последний интеграл можно записать так:

$$P(f(\xi_1, \xi_2 \dots \xi_n) \in \Delta) = \int\limits_{-\infty}^{\infty} \dots \int\limits_{-\infty}^{\infty} p_{\xi_1}(x_1) \dots p_{\xi_m}(x_m)\, dx_1 \dots$$
$$\dots dx_m \left\{ \int\int \dots \int\limits_{f(x_1, x_2, \dots, x_n) \in \Delta} p_{\xi_{m+1}}(x_{m+1}) \dots p_{\xi_n}(x_n)\, dx_{m+1} \dots dx_n \right\}.$$

Интеграл в фигурных скобках в соответствии с общей формулой (4.58) есть вероятность события $\{f(x_1, x_2, \dots, x_m, \xi_{m+1}, \dots, \xi_n) \in \Delta\}$. Следовательно, при любом m, $1 \leqslant m < n$,

$$P(f(\xi_1, \xi_2, \dots, \xi_n) \in \Delta) =$$
$$= \int\limits_{-\infty}^{\infty} \dots \int\limits_{-\infty}^{\infty} p_{\xi_1}(x_1) \dots p_{\xi_m}(x_m)\, P(f(x_1, x_2, \dots x_m,$$
$$\xi_{m+1}, \dots, \xi_n) \in \Delta)\, dx_1 \dots dx_m. \tag{4.59}$$

---

* Для того чтобы это сделать, достаточно построить функцию $f(x_1, x_2, \dots, x_n)$, которая во всех точках области D положительна, а во всех остальных точках меньше или равна нулю; в качестве Δ в этом случае достаточно взять интервал $(0, \infty)$.

Приведем примеры применения формулы (4.59) для нахождения распределения $\xi_2 - \xi_1$, $\xi_1\xi_2$, $\frac{\xi_2}{\xi_1}$, где ξ_1, ξ_2 — независимые случайные величины, причем в случае частного $P(\xi_1 = 0) = 0$.
Имеем:

$$F_{\xi_2-\xi_1}(x) = \int_{-\infty}^{\infty} p_{\xi_1}(x_1)\,P(\xi_2 - x_1 < x)\,dx_1 =$$

$$= \int_{-\infty}^{\infty} p_{\xi_1}(x_1)\,P(\xi_2 < x + x_1)\,dx_1 = \int_{-\infty}^{\infty} p_{\xi_1}(x_1)\,F_{\xi_2}(x + x_1)\,dx_1 \tag{4.60}$$

$$F_{\xi_1\xi_2}(x) = \int_{-\infty}^{\infty} p_{\xi_1}(x_1)\,P(\xi_2 x_1 < x)\,dx_1 =$$

$$= \int_{-\infty}^{0} p_{\xi_1}(x_1)\,P(\xi_2 x_1 < x)\,dx_1 + \int_{0}^{\infty} p_{\xi_1}(x_1)\,P(\xi_2 x_1 < x)\,dx_1 =$$

$$= \int_{-\infty}^{0} p_{\xi_1}(x_1)\left[1 - F_{\xi_2}\left(\frac{x}{x_1} + 0\right)\right]dx_1 +$$

$$+ \int_{0}^{\infty} p_{\xi_1}(x_1)\,F_{\xi_2}\left(\frac{x}{x_1}\right)dx_1. \tag{4.61}$$

$$F_{\xi_2/\xi_1}(x) = \int_{-\infty}^{\infty} p_{\xi_1}(x_1)\,P(\xi_2/x_1 < x)\,dx_1 =$$

$$= \int_{-\infty}^{0} p_{\xi_1}(x_1)\,P(\xi_2 > x x_1)\,dx_1 + \int_{0}^{\infty} p_{\xi_1}(x_1)\,P(\xi_2 < x \cdot x_1)\,dx_1 =$$

$$= \int_{-\infty}^{0} p_{\xi_1}(x_1)\,[1 - F_{\xi_2}(x x_1 + 0)]\,dx_1 +$$

$$+ \int_{0}^{\infty} p_{\xi_1}(x_1)\,F_{\xi_2}(x x_1)\,dx_1. \tag{4.62}$$

Вычислим в качестве примера распределение частного показательной и равномерной случайных величин:

$$p_{\xi_1}(x) = \begin{cases} \frac{1}{a}, & \text{если } x \in (0, a); \\ 0, & \text{если } x \overline{\in} (0, a), \end{cases}$$

$$p_{\xi_2}(x) = e^{-x}, \quad x > 0.$$

Согласно формуле (4.62), имеем

$$F_{\xi_2/\xi_1}(x) = \int_0^a \frac{1}{a}(1 - e^{-xx_1})\, dx_1 = 1 - \frac{1}{ax}(1 - e^{-ax}), \quad x > 0. \tag{4.63}$$

Очевидно, при $x < 0$ $p_{\xi_2/\xi_1}(x) = 0$.

§ 4. МОМЕНТЫ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Пусть ξ — случайная величина. Тогда ξ^n также будет случайной величиной при любом целом n. Если ξ^n обладает математическим ожиданием $M\xi^n$, то $M\xi^n$ называется *моментом порядка n случайной величины* ξ. Если $M\xi = a$ конечно, то при любом n можно рассмотреть случайную величину $(\xi - a)^n$.

Если $(\xi - a)^n$ обладает математическим ожиданием $M(\xi - a)^n$, то это число (конечное или бесконечное) называется *центральным моментом порядка n случайной величины* ξ.

Пример. Пусть ξ — случайная величина, распределенная по равномерному закону: $p_\xi(x) = 1$ при $0 < x < 1$; $p_\xi(x) = 0$ при $x \overline{\in} [0, 1]$. По формуле (4.11), для любой случайной величины η, принимающей с вероятностью 1 неотрицательные значения, $M\eta = \int_0^\infty [1 - F_\eta(x)]\, dx$. Положим $\eta = \xi^n$. Тогда при $x \geqslant 1$ и любом $n \geqslant 1$ $1 - F_\eta(x) = 0$, а при $0 < x < 1$

$$1 - F_\eta(x) = \mathrm{P}\{\xi^n \geqslant x\} = \mathrm{P}\left\{\xi \geqslant x^{\frac{1}{n}}\right\} = 1 - x^{\frac{1}{n}}.$$

Отсюда

$$M\xi^n = M\eta = \int_0^1 \left(1 - x^{\frac{1}{n}}\right) dx = 1 - \frac{n}{n+1} x^{\frac{n+1}{n}} \Big|_0^1 = \frac{1}{n+1}.$$

В частности, $M\xi$, или, согласно принятой терминологии, момент порядка 1 случайной величины ξ, равен $\frac{1}{2}$.

Основная формула для момента порядка n имеет следующий вид:

$$M\xi^n = \int_{-\infty}^{\infty} x^n p_\xi(x)\,dx, \tag{4.64}$$

а для центрального момента порядка n — вид

$$M(\xi - M\xi)^n = \int_{-\infty}^{\infty} (x - M\xi)^n p_\xi(x)\,dx. \tag{4.65}$$

Эти формулы следуют из общей формулы (3.72) для математического ожидания произвольной непрерывной функции от случайной величины. Приведем также непосредственное доказательство.

Очевидно, формулы (4.64) и (4.65) являются частными случаями формулы

$$M(\xi - a)^n = \int_{-\infty}^{\infty} (x - a)^n p_\xi(x)\,dx, \tag{4.66}$$

где a — произвольное число. Поэтому достаточно доказать формулу (4.66).

Найдем плотность случайной величины $(\xi - a)^n$. Имеем $(\xi - a)^n = f(\xi)$, где $f(x) = (x - a)^n$. Функция $f(x)$ имеет единственную вещественную обратную функцию $f^{-1}(x) = a + x^{\frac{1}{n}}$ при нечетном n и две обратных функции $a(x) = a + x^{\frac{1}{n}}$, $b(x) = a - x^{\frac{1}{n}}$ при n четном.

Далее, $a'(x) = \frac{1}{n} x^{\frac{1-n}{n}}$, $b'(x) = -\frac{1}{n} x^{\frac{1-n}{n}}$.

Отсюда по формуле (4.18) при нечетном n имеем

$$p_{(\xi-a)^n}(x) = \frac{1}{n} p_\xi\left(a + x^{\frac{1}{n}}\right) \cdot x^{\frac{1-n}{n}}, \quad -\infty < x < \infty,$$

при четном же n

$$p_{(\xi-a)^n}(x)=\frac{x^{\frac{1-n}{n}}}{n}\left[p_\xi\left(a+x^{\frac{1}{n}}\right)+p_\xi\left(a-x^{\frac{1}{n}}\right)\right].$$

Следовательно, при нечетном n

$$M(\xi-a)^n=\int_{-\infty}^{\infty}xp_{(\xi-a)^n}(x)\,dx=\frac{1}{n}\int_{-\infty}^{\infty}x^{\frac{1}{n}}p_\xi\left(a+x^{\frac{1}{n}}\right)dx,$$

а при n четном

$$M(\xi-a)^n=\frac{1}{n}\int_{0}^{\infty}x^{\frac{1}{n}}\left[p_\xi\left(a+x^{\frac{1}{n}}\right)+p_\xi\left(a-x^{\frac{1}{n}}\right)\right]dx.$$

Подстановками $a\pm x^{\frac{1}{n}}=t$ в обоих случаях интегралы приводятся к виду $\int\limits_{-\infty}^{\infty}(t-a)^np_\xi(t)\,dt$, что и требовалось доказать.

Пусть $M\xi^n$ конечен. Тогда вследствие формулы (4.64) $\int\limits_{-\infty}^{\infty}|x|^np_\xi(x)\,dx<\infty$. Возьмем теперь любое целое m $(0\leqslant m<n)$. Тогда

$$\pm\int_{-\infty}^{\infty}x^mp_\xi(x)\,dx=\pm\int_{-1}^{1}x^mp_\xi(x)\,dx\pm\int_{|x|>1}x^mp_\xi(x)\,dx\leqslant$$

$$\leqslant\int_{-1}^{1}p_\xi(x)\,dx+\int_{|x|>1}|x|^np_\xi(x)\,dx<\infty.$$

Таким образом, если у данной случайной величины конечен момент порядка n, то конечны также моменты порядков $1, 2, \ldots, n-1$. Отсюда вытекает, что если выражение $M(\xi-a)^n$ конечно при некотором a, то оно конечно при любых a. В самом деле, пусть $|M(\xi-b)^n|<\infty$. Тогда

$$M(\xi-a)^n=\int_{-\infty}^{\infty}(x-a)^np_\xi(x)\,dx=\int_{-\infty}^{\infty}(x-b-b-a)^n\times$$

$$\times p_\xi(x)\,dx=\sum_{k=0}^{n}C_n^k(b-a)^{n-k}\int_{-\infty}^{\infty}(x-b)^kp_\xi(x)\,dx.$$

В правой части последнего равенства — сумма конечных слагаемых; отсюда и в левой части — конечное число.

Наиболее важен для теории вероятностей центральный момент случайной величины ξ второго порядка, иначе называемый дисперсией ξ и обозначаемый $D\xi$. По определению,

$$D\xi = M(\xi - M\xi)^2,$$

т. е. дисперсия случайной величины — это математическое ожидание квадрата отклонения случайной величины от ее математического ожидания. Дисперсия, как и любой центральный момент, определена только для тех случайных величин, которые обладают конечным математическим ожиданием. Условимся, однако, считать, что если $M\xi = \pm\infty$ либо $M\xi$ не существует, то $D\xi = \infty$. При таком определении любая случайная величина обладает дисперсией, конечной или бесконечной. Дисперсия ξ конечна тогда и только тогда, когда конечен $M\xi^2$.

Докажем формулу

$$D\xi = M\xi^2 - (M\xi)^2. \tag{4.67}$$

Имеем

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

$$- 2M\xi \int_{-\infty}^{\infty} x p_\xi(x)\,dx + (M\xi)^2 \int_{-\infty}^{\infty} p_\xi(x)\,dx = M\xi^2 -$$

$$- 2M\xi M\xi + (M\xi)^2 = M\xi^2 - (M\xi)^2.$$

Важна также следующая формула:

$$D(a\xi + b) = a^2 D\xi \tag{4.68}$$

при любых постоянных a и b, за исключением случая, когда $a = 0$, $D\xi = \infty$. Действительно, если $M\xi$ не является конечной величиной, то $M(a\xi + b)$ также не будет конечным при $a \neq 0$. Следовательно, $D\xi = D(a\xi + b) = \infty$ и формула (68) выполняется. Если же $M\xi$ конечно, то

$$M(a\xi + b) = aM\xi + b$$

и

$$D(a\xi + b) = M(a\xi + b - aM\xi - b)^2 =$$
$$= M[a(\xi - M\xi)]^2 = a^2 M(\xi - M\xi)^2 = a^2 D\xi,$$

что и требовалось доказать. В частности, $D(\xi + b) = D\xi$.

Дисперсия служит характеристикой (хотя и не полной) рассеяния случайной величины около ее математического ожидания. Представим себе следующую схему образования случайных величин. Пусть производится стрельба из винтовки, закрепленной в станке, по мишени A, находящейся от конца ствола на расстоянии z_1 (рис. 4.6). Предположим, что за мишенью A расположена мишень B, находящаяся от конца ствола на расстоянии $z_2 > z_1$. Пусть мишени центрированы таким образом, что линия прицеливания проходит через центры обоих мишеней. Обозначим

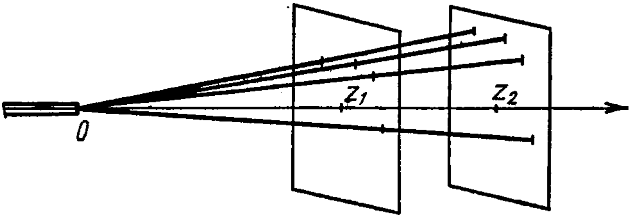

Рис. 4.6

через ξ_1 отклонение точки попадания пули от центра по вертикали для мишени A, ξ_2 — то же, для мишени B. Тогда (в предположении, что траектория пули есть прямая линия) справедливо соотношение

$$\xi_2/\xi_1 = z_2/z_1.$$

Отсюда вследствие формулы (4.68) $D\xi_2/D\xi_1 = z_2^2/z_1^2$. С другой стороны, чем дальше мишень, тем точки попадания более рассеяны. Поэтому дисперсия служит некоторой мерой рассеяния.

Рассмотрим примеры на вычисление моментов распределений.

1. Пусть ξ — равномерно распределенная случайная величина в интервале (a, b). Тогда $p_\xi(x) = \frac{1}{b-a}$, $x \in (a, b)$. Отсюда

$$M\xi^n = \int_{-\infty}^{\infty} x^n p_\xi(x)\,dx = \frac{1}{b-a}\int_a^b x^n\,dx = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}.$$

В частности,

$$M\xi = \frac{a+b}{2}, \quad M\xi^2 = \frac{a^2+ab+b^2}{3} \tag{4.69}$$

откуда

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{(b-a)^2}{12}. \tag{4.70}$$

2. Пусть ξ имеет нормальное распределение:

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty. \tag{4.71}$$

Тогда

$$M\xi^n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^n e^{-\frac{x^2}{2}} dx. \tag{4.72}$$

При нечетных n этот интеграл как интеграл от нечетной функции равен нулю. Таким образом, остается найти $M\xi^n$ при n четных. Пусть сначала $n = 2$, тогда

$$M\xi^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \sqrt{\frac{2}{\pi}} \int_0^{\infty} x^2 e^{-\frac{x^2}{2}} =$$

$$= \sqrt{\frac{2}{\pi}} \int_0^{\infty} x d\left(-e^{-\frac{x^2}{2}}\right) = -\sqrt{\frac{2}{\pi}}\, x e^{-\frac{x^2}{2}} \Big|_0^{\infty} +$$

$$+ \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} p_\xi(x)\, dx = 1. \tag{4.73}$$

Подобным же образом, проинтегрировав интеграл, выражающий $M\xi^n$, по частям, найдем

$$M\xi^n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^n e^{-\frac{x^2}{2}} dx = \sqrt{\frac{2}{\pi}} \int_0^{\infty} x^n e^{-\frac{x^2}{2}} dx =$$

$$= \sqrt{\frac{2}{\pi}} \int_0^{\infty} x^{n-1} d\left(e^{-\frac{x^2}{2}}\right) = -\sqrt{\frac{2}{\pi}}\, x^{n-1} e^{-\frac{x^2}{2}} \Big|_0^{\infty} +$$

$$+(n-1)\sqrt{\frac{2}{\pi}}\int\limits_0^{\infty} x^{n-2}e^{-\frac{x^2}{2}}\,dx =$$

$$=(n-1)\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} x^{n-2}e^{-\frac{x^2}{2}}\,dx = (n-1)\,M\xi^{n-2}.$$

Итак, доказана рекуррентная формула

$$M\xi^n = (n-1)\,M\xi^{n-2}. \tag{4.73'}$$

Отсюда

$$M\xi^n = \frac{M\xi^n}{M\xi^{n-2}}\cdot\frac{M\xi^{n-2}}{M\xi^{n-4}}\cdots\frac{M\xi^4}{M\xi^2}\,M\xi^2 = (n-1)(n-3)\ldots 3.1$$

Положим $n = 2k$. Тогда

$$(n-1)(n-3)\ldots 1 = (2k-1)(2k-3)\ldots 1 =$$

$$=\frac{(2k-1)\,!}{[2\,(k-1)]\,[2\,(k-2)]\ldots 2} = \frac{(2k-1)\,!}{2^{k-1}\,(k-1)\,!}.$$

Итак,

$$M\xi^{2k} = \frac{(2k-1)\,!}{2^{k-1}\,(k-1)\,!},\quad k = 1, 2, \ldots$$

Найдем теперь центральные моменты распределения случайной величины η с плотностью

$$p_\eta(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-\frac{(x-a)^2}{2\sigma^2}},\quad -\infty < x < \alpha \tag{4.74}$$

Если ввести случайную величину ξ по формуле

$$\xi = \frac{\eta - a}{\sigma}\ \text{ или }\ \eta = a - \sigma\xi, \tag{4.75}$$

то имеем

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}.$$

Из формулы (4.75) находим, что

$$M\eta = a + \sigma M\xi = a. \tag{4.76}$$

Итак, математическое ожидание случайной величины η с плотностью (4.74) равно a. По формуле (4.73), $M\xi^2 = 1$. Отсюда

$$D\eta = M(\eta - M\eta)^2 = M(\eta - a)^2 = M(\sigma\xi)^2 = \sigma^2 M\xi^2 = \sigma^2.$$

Значит, дисперсия данного распределения равна σ^2. Корень из дисперсии $\sigma \geqslant 0$ называется *среднеквадратическим отклонением случайной величины.*

Рассмотрим центральные моменты η любых порядков. Имеем

$$M(\eta - M\eta)^n = M(\sigma^n\xi^n)\,\sigma^n M\xi^n =$$

$$= \begin{cases} 0 & \text{при } n = 2k-1 \\ \dfrac{\sigma^{2k}(2k-1)!}{2^{k-1}(k-1)!} & \text{при } n = 2k \ (k = 1, 2, \ldots) \end{cases} \tag{4.77}$$

3. Пусть

$$p_\xi(x) = \lambda e^{-\lambda x}, \ x > 0.$$

Тогда

$$M\xi^n = \lambda \int_0^\infty x^n e^{-\lambda x}\,dx = \frac{1}{\lambda^n}\int_0^\infty t^n e^{-t}\,dt = \frac{1}{\lambda^n}\Gamma(n+1) =$$

$$= \frac{n!}{\lambda^n}{}^*. \tag{4.78}$$

Докажем следующее предложение. Если случайная величина ξ имеет при некотором четном $n \geqslant 2$ центральный момент n-го порядка, равный 0, то $\xi = M\xi$ с вероятностью 1. Пусть

$$\int_{-\infty}^{\infty} (x - M\xi)^n p_\xi(x)\,dx = 0.$$

Можно записать

$$0 = \int_{-\infty}^{\infty} (x - M\xi)^n p_\xi(x)\,dx \geqslant \int_{|x - M\xi| \geqslant \delta} (x - M\xi)^n p_\xi(x)\,dx \geqslant$$

$$\geqslant \delta^n \int_{|x - M\xi| \geqslant \delta} p_\xi(x)\,dx = \delta^n \mathrm{P}\{|\xi - M\xi| \geqslant \delta\}.$$

---

* Эту формулу можно найти по индукции, проинтегрировав по частям, подобно тому, как была доказана формула (4.73).

Таким образом, при произвольном $\delta > 0$

$$P\{|\xi - M\xi| \geqslant \delta\} = 0.$$

Это означает, что $P\{|\xi - M\xi| > 0\} = 0$, т. е. $\xi = M\xi$ с вероятностью 1. Обратно, если $\xi = M\xi$ с вероятностью 1, то $(\xi - M\xi)^n = 0$ с вероятностью 1 для любого n, т. е. $M(\xi - M\xi)^n = 0$. В частности, необходимым и достаточным условием того, чтобы случайная величина ξ сводилась к постоянной, является равенство $D\xi = 0$.

Пусть ξ_1, ξ_2 — случайные величины, обладающие конечными математическими ожиданиями. Допустим, что $(\xi_1 - M\xi_1) \times (\xi_2 - M\xi_2)$ также является случайной величиной, обладающей конечным математическим ожиданием. Тогда выражение $M(\xi_1 - M\xi_1) \times (\xi_2 - M\xi_2)$ называется *ковариацией случайных величин* ξ_1 и ξ_2 и обозначается $\operatorname{cov}(\xi_1, \xi_2)$.

Случайные величины ξ_1, ξ_2 называются *некоррелированными*, если $\operatorname{cov}(\xi_1, \xi_2) = 0$; в противном случае они называются *коррелированными*.

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — случайные величины, обладающие конечными дисперсиями и попарно некоррелированные:

$$\operatorname{cov}(\xi_i, \xi_j) = 0 \text{ при } 1 \leqslant i < j \leqslant n.$$

Тогда

$$D(\xi_1 + \xi_2 + \ldots + \xi_n) = D\xi_1 + D\xi_2 + \ldots + D\xi_n. \tag{4.79}$$

Для доказательства используем следующий полезный прием. Если ξ — любая случайная величина с конечным математическим ожиданием, то обозначим $\xi^0 = \xi - M\xi$. (Операцию перехода от ξ к ξ^0 называют центрированием.) Очевидно, $M\xi^0 = 0$ и $D\xi = D\xi^0 = M(\xi^0)^2$. Если ξ, η — любые случайные величины с конечными $M\xi$ и $M\eta$, то $\operatorname{cov}(\xi, \eta) = \operatorname{cov}(\xi^0, \eta^0)$. В самом деле,

$$\operatorname{cov}(\xi, \eta) = M(\xi - M\xi)(\eta - M\eta) = M\xi^0\eta^0 = M(\xi^0 - M\xi^0) \times (\eta^0 - M\eta^0) = \operatorname{cov} \xi^0\eta^0.$$

Таким образом, для доказательства равенства (4.79) достаточно установить аналогичное равенство для ξ_i^0 вместо ξ_i $(1 \leqslant i \leqslant n)$.

Поскольку, очевидно, $(\Sigma\, \xi_i)^0 = \Sigma\, \xi_i^0$, то

$$D(\xi_1^0 + \xi_2^0 + \ldots + \xi_n^0) = M(\xi_1^0 + \xi_2^0 + \ldots + \xi_n^0)^2 =$$

$$= M\left[\sum_{i=1}^{n} (\xi_i^0)^2 + \sum_{i \neq j} \xi_i^0\xi_j^0\right] = \sum_{i=1}^{n} M(\xi_i^0)^2 + \sum_{i \neq j} M\xi_i^0\xi_j^0 =$$

$$= \sum_{i=1}^{n} D\xi_i + \sum_{i \neq j} \operatorname{cov}(\xi_i \xi_j) = \sum_{i=1}^{n} D\xi_i,$$

что и требовалось доказать. Если ξ, η — независимые случайные величины, обладающие конечными математическими ожиданиями, то

$$M(\xi - M\xi)(\eta - M\eta) = M\xi^0\eta^0 = M\xi^0 M\eta^0 = 0,$$

т. е. ξ и η некоррелированы. Следовательно, формула (4.79) выполняется, в частности, для случая, когда любая ξ_i не зависит от любой ξ_j, $i \neq j$. В частности, если при этом ξ_i обладают одной и той же дисперсией σ^2, то

$$D\left(\frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}\right) = \frac{1}{n^2} D(\xi_1 + \xi_2 + \ldots + \xi_n) =$$
$$= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Среднеквадратическое отклонение случайной величины $\frac{1}{n}(\xi_1 + \xi_2 + \ldots + \xi_n)$ равно $\frac{\sigma}{\sqrt{n}}$. Пусть, например, производится измерение какой-либо физической величины, причем ошибка измерения есть случайная величина с математическим ожиданием a и дисперсией σ^2. Если произведено n независимых измерений и в качестве оценки физической величины взято среднее арифметическое результатов измерений, то это среднее будет иметь математическое ожидание a и среднеквадратическое отклонение $\frac{\sigma}{\sqrt{n}}$. Если $a = 0$ (или, как говорят, отсутствует систематическая ошибка измерения), то $M\eta_n^2 = \frac{\sigma^2}{n}$, где η_n — ошибка среднего арифметического. При $n \to \infty$ $M\eta_n^2 \to 0$. т. е. рассеяние ошибок стремится к нулю.

§ 5. НЕКОТОРЫЕ СПЕЦИАЛЬНЫЕ РАСПРЕДЕЛЕНИЯ

Пусть ξ — непрерывно-дискретная случайная величина с плотностью $p_\xi(x)$. Комплексную функцию

$$\varphi_\xi(t) = \int_{-\infty}^{\infty} \cos tx p_\xi(x)\,dx + i \int_{-\infty}^{\infty} \sin tx p_\xi(x)\,dx =$$
$$= \int_{-\infty}^{\infty} e^{itx} p_\xi(x)\,dx, \qquad (4.80)$$

определенную при любых t, $-\infty < t < \infty$, назовем *характеристической функцией случайной величины* ξ. Характеристическая функция — это преобразование Фурье плотности $p_\xi(x)$, свойства которого изучаются в курсе математического анализа. Для нас наиболее важны следующие свойства характеристической функции.

1. Если $\int\limits_{-\infty}^{\infty} |\varphi_\xi(t)|\, dt < \infty$, то случайная величина ξ обладает непрерывным распределением с непрерывной плотностью $p_\xi(x)$, которая при любом x определяется равенством

$$p_\xi(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx} \varphi_\xi(t)\, dt. \tag{4.80}$$

2. Если $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины, то

$$\varphi_{\xi_1+\xi_2+\ldots+\xi_n}(t) = \varphi_{\xi_1}(t)\, \varphi_{\xi_2}(t) \ldots \varphi_{\xi_n}(t). \tag{4.81}$$

3. Если случайная величина ξ обладает конечным моментом k-го порядка ($k \geqslant 1$), то

$$M\xi^k = (-i)^k \varphi_\xi^{(k)}(0). \tag{4.82}$$

В частности,

$$M\xi = -i\varphi'_\xi(0), \quad M\xi^2 = -\varphi''_\xi(0), \tag{4.83}$$

откуда

$$D\xi = [\varphi'_\xi(0)]^2 - \varphi''_\xi(0). \tag{4.84}$$

(Подробнее о характеристических функциях см. § 2 гл. VII.)

Наибольшую пользу аппарат характеристических функций приносит при исследовании распределения сумм независимых случайных величин. Рассмотрим важные примеры.

1. Вычислим характеристическую функцию нормально распределенной случайной величины. Имеем:

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} \qquad -\infty < x < \infty;$$

$$\varphi_\xi(t)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\exp\left\{itx-\frac{(x-a)^2}{2\sigma^2}\right\}dx=$$

$$=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty}\exp\left\{-\frac{(x-a-\sigma^2it)^2}{2\sigma^2}-\frac{\sigma^2t^2}{2}+ait\right\}dx=$$

$$=\exp\left\{iat-\frac{\sigma^2t^2}{2}\right\}\frac{1}{\sqrt{2\pi}}\int_{-\infty-it}^{\infty-it}\exp\left\{-\frac{u^2}{2}\right\}du.$$

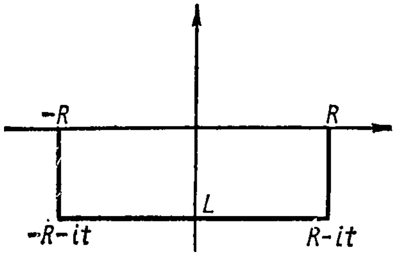

Рис. 4.7

По теореме Коши, известной из теории функций комплексного переменного, интеграл аналитической функции по замкнутому контуру равен нулю. Рассмотрим контур L, изображенный на рис. 4.7. Тогда получим

$$\int_{-R-it}^{R-it}e^{-\frac{u^2}{2}}du=\int_{-R}^{R}e^{-\frac{u^2}{2}}du+\int_{R}^{R-it}e^{-\frac{u^2}{2}}du+\int_{-R-it}^{-R}e^{-\frac{u^2}{2}}du.$$

Последние два интеграла стремятся к 0 при $R\to\infty$, поскольку

$$e^{-\frac{u^2}{2}}=\exp\left\{-\frac{1}{2}(R+iv)^2\right\},$$

где $|v|\leqslant|t|$.

Отсюда

$$\int_{-\infty-it}^{\infty-it}e^{-\frac{u^2}{2}}du=\int_{-\infty}^{\infty}e^{-\frac{u^2}{2}}du=\sqrt{2\pi}.$$

Итак,

$$\varphi_\xi(t)=\exp\left\{iat-\frac{\sigma^2t^2}{2}\right\}. \tag{4.85}$$

Докажем следующее утверждение. *Если* $\xi_1,\xi_2,\ldots,\xi_n$ — *независимые нормально распределенные случайные вели-*

чины с математическими ожиданиями $a_1, a_2, \ldots, a_n$ и дисперсиями $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, то $S_n = \xi_1 + \xi_2 + \ldots + \xi_n$ — нормально распределенная случайная величина с математическим ожиданием $a_1 + a_2 + \ldots + a_n$ и дисперсией $\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$.

Доказательство.

$$\varphi_{S_n}(t) = \prod_{j=1}^{n} \varphi_{\xi_j}(t) = \exp\Big\{i(a_1 + a_2 + \ldots + a_n)t - -\frac{1}{2}(\sigma_1^2 + \sigma_2^2 + \ldots \sigma_n^2)t^2\Big\}. \tag{4.86}$$

Поскольку правая часть равенства (4.86), очевидно, абсолютно интегрируема в интервале $(-\infty, \infty)$, то, согласно первому из перечисленных свойств характеристических функций, S_n обладает плотностью, которая задается обратным преобразованием Фурье (4.80). Однако, подставив в формулу (4.85) $a = a_1 + a_2 + \ldots + a_n$, $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2$, найдем, что $\varphi_{S_n}(t)$ есть преобразование Фурье плотности нормального распределения с математическим ожиданием a и дисперсией σ^2. Значит, обратное преобразование Фурье есть именно данная плотность.

Итак,

$$p_{S_n}(x) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2)}} \times \exp\left\{-\frac{1}{2}\frac{[x - (a_1 + a_2 + \ldots + a_n)]^2}{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2}\right\},$$

$$-\infty < x < \infty. \tag{4.87}$$

2. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые нормально распределенные случайные величины; $M\xi_i = 0$, $D\xi_i = 1$. Случайная величина $\chi_n^2 = \sum_{i=1}^{n} \xi_i^2$ называется случайной величиной хи-квадрат с n степенями свободы [ее распределение называется распределением χ^2 (хи-квадрат) с n степенями свободы]. Прежде всего найдем распределение χ_1^2. Имеем

$$p_{\chi_1^2}(x) = p_{\xi_1^2}(x) =$$

$$
= \begin{cases} \dfrac{1}{2\sqrt{x}}\left[p_\xi(x) + p_\xi(-x)\right] = \dfrac{1}{\sqrt{2\pi x}} e^{-\frac{x^2}{2}} & \text{при } x > 0, \\ 0 & \text{при } x < 0. \end{cases} \tag{4.88}
$$

Характеристическую функцию этой случайной величины можно вычислить с помощью следующего косвенного приема. В § 2 данной главы было показано, что $\xi_1^2 + \xi_2^2$ распределена по закону

$$
p_{\xi_1^2 + \xi_2^2}(x) = \frac{1}{2} e^{-\frac{x}{2}}, \quad x > 0. \tag{4.89}
$$

Отсюда

$$
\varphi_{\xi_1^2 + \xi_2^2}(t) = \frac{1}{2}\int_0^\infty e^{itx - \frac{x}{2}}\, dx = \frac{1}{1 - 2it}. \tag{4.90}
$$

В то же время

$$
\varphi_{\xi_1^2 + \xi_2^2}(t) = \varphi_{\xi_1^2}(t)\,\varphi_{\xi_2^2}(t) = \left[\varphi_{\xi_1^2}(t)\right]^2, \tag{4.91}
$$

откуда

$$
\varphi_{\xi_1^2}(t) = \frac{1}{\sqrt{1 - 2it}}^{\,*}. \tag{4.92}
$$

Теперь уже непосредственно очевидно, что

$$
\varphi_{\chi_n^2}(t) = \frac{1}{(1 - 2it)^{\frac{n}{2}}}. \tag{4.93}
$$

Обратное преобразование Фурье функции вида (4.93) есть

$$
p_{\chi_n^2}(x) = \frac{x^{\frac{n}{2} - 1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\, \Gamma\left(\frac{n}{2}\right)}. \tag{4.94}
$$

---

* Выбирается та ветвь корня, которая соответствует значению $\varphi_{\xi_1^2}(0) = 1$ (так как для любой случайной величины η $\varphi_\eta(0) = \int_{-\infty}^{\infty} p_\eta(x)\, dx = 1$).

Пусть $\eta_1, \eta_2, \dots, \eta_n$ — независимые и равномерно распределенные в интервалах $(0, a_1)$, $(0, a_2)$, ..., $(0, a_n)$ — случайные величины, где $a_i > 0$, $1 \leqslant i \leqslant n$. Зададимся целью найти распределение произведения

$$\xi = \eta_1\eta_2 \dots \eta_n. \tag{4.95}$$

Прежде всего можно записать $\eta_i = a_i\eta_i^0$, где η_i^0 независимы и равномерно распределены в интервале $(0, 1)$. Отсюда

$$\zeta = a\eta_1^{(0)}\eta_2^{(0)} \dots \eta_n^{(0)},$$

где $a = a_1a_2 \dots a_n$. Обозначим $\xi_i = -\ln \eta_i^{(0)}$, $1 \leqslant i \leqslant n$. Тогда

$$P\{\xi_i < x\} = P\{-\ln \eta_i^{(0)} < x\} = P\{\eta_i^{(0)} > e^{-x}\} =$$
$$= 1 - e^{-x} \, (0 < x < \infty);$$

последнее означает, что ξ_i распределены по показательному закону

$$p_{\xi_i}(x) = e^{-x}, \quad x > 0.$$

Далее,

$$-\ln \zeta = -\ln a + \xi_1 + \dots + \xi_n.$$

Следовательно, по формуле (4.57),

$$p_{-\ln \zeta}(x) = p_{\xi_1 + \dots + \xi_n}(x + \ln a) =$$
$$= \frac{(x + \ln a)^{n-1}}{a\,(n-1)!} e^{-x}, \quad x > -\ln a. \tag{4.96}$$

Наконец,

$$p_\zeta(x) = \frac{1}{x} p_{-\ln \zeta}(-\ln x) = \frac{(\ln a - \ln x)^{n-1}}{a\,(n-1)!}, \quad 0 < x < a. \tag{4.97}$$

В этом примере плотность вероятности произведения случайных величин вычислена по плотности суммы их логарифмов.

Рассмотрим случай, когда $\zeta = \eta_1\eta_2 \dots \eta_n$, где η_i независимы и могут быть как положительными, так и отрицательными.

Пусть $P\{\eta_i > 0\} = p_{i,0}$, $P\{\eta_i < 0\} = p_{i,1}$, $p_{i,0} +$

$+p_{i,1}=1, \quad 1 \leqslant i \leqslant n$. Тогда

$$p_{\eta_i}(x) = p_{i,0}\, p_{\eta_{i0}}(x) + p_{i,1}\, p_{\eta_{i1}}(x), \tag{4.98}$$

где при $p_{i,0} > 0$

$$\left.\begin{array}{l} p_{\eta_{i,0}}(x) = \dfrac{p_{\eta_i}(x)}{\int\limits_0^\infty p_{\eta_i}(t)\,dt}, \quad x > 0; \; p_{\eta_{i,0}}(x) = 0, \; x < 0 \\ \text{и при } p_{i,1} > 0 \\ p_{\eta_{i1}}(x) = 0, \quad x > 0; \; p_{\eta_{i,1}}(x) = \dfrac{p_{\eta_i}(x)}{\int\limits_{-\infty}^0 p_{\eta_i}(t)\,dt}, \; x < 0. \end{array}\right\} \tag{4.99}$$

[Если $p_i = 0$, то $p_{\eta_{i,0}}(x)$, а если $q_i = 0$, то $p_{\eta_{i,1}}(x)$ можно определить произвольным образом.] Отсюда

$$p_\zeta(x) = \sum_{\substack{j_s=0,1 \\ 1<s<n}} p_{1,j_1} p_{2,j_2} \ldots p_{n,j_n} p_{\eta_{1,j_1}\eta_{2,j_2}\ldots\eta_{n,j_n}}(x) \tag{4.100}$$

Обозначим

$$\xi_{i0} = \ln \eta_{i,0}, \quad \xi_{i,1} = \ln(-\eta_{i,1}) = \ln|\eta_{i,1}|, \tag{4.101}$$

или вообще $\xi_{i,j} = \ln|\eta_{ij}|, \quad j = 0, \; 1.$

Отсюда

$$\eta_{1,j_1}\eta_{2,j_2}\ldots\eta_{n,j_n} = (-1)^{j_1+j_2+\ldots+j_n}|\eta_{1,j_1}||\eta_{2,j_2}|\ldots|\eta_{n,j_n}|, \tag{4.102}$$

поскольку знак левой части равенства (4.102) есть $\prod\limits_{i=1}^n \operatorname{sign}\eta_{i,j_i}$, а $\operatorname{sign}\eta_{ij_i} = (-1)^{j_i}$. Следовательно.

$$p_{\eta_{1,j_1}\eta_{2,j_2}\ldots\eta_{n,j_n}}(x) = p_{|\eta_{1,j_1}\eta_{2,j_2}\ldots\eta_{n,j_n}|}\left((-1)^{j_1+\ldots+j_n}x\right) =$$

$$
= \begin{cases} \dfrac{1}{|x|} p_{\xi_{1,j_1}+\xi_{2,j_2}+\ldots+\xi_{n,j_n}}(\ln|x|), \text{ если} \\ \qquad (-1)^{j_1+j_2+\ldots+j_n} x > 0, \\ 0, \text{ если } (-1)^{j_1+j_2+\ldots+j_n} x < 0. \end{cases} \tag{4.103}
$$

Окончательно имеем

$$
p_\zeta(x) =
$$

$$
= \begin{cases} \sum\limits_{j_1+j_2+\cdots+j_n \text{ чет}} p_{1,j_1} p_{2,j_2} \ldots p_{n,j_n} \dfrac{1}{x} p_{\xi_1 j_1+\ldots+\xi_{n,j_n}}(\ln x) \\ \qquad\qquad\qquad\qquad\qquad \text{при } x > 0 \\ \sum\limits_{j_1+j_2+\cdots+j_n \text{ неч}} p_{1,j_1} p_{2,j_2} \ldots p_{n,j_n} \dfrac{1}{|x|} p_{\xi_1 j_1}+\ldots+\xi_{n,j_n}(\ln|x|) \\ \qquad\qquad\qquad\qquad\qquad \text{при } x < 0. \end{cases} \tag{4.104}
$$

Решим следующую задачу. Пусть η_1 и η_2 — независимые положительные случайные величины с кусочно-постоянной плотностью

$$
p_{\eta_i}(x) = h_{i,j} \quad \text{при} \quad x \in (a_{i,j}, a_{i,j+1}) \tag{4.105}
$$

$$
0 \leqslant j \leqslant n_i - 1, \quad i = 1, 2,
$$

где $a_{i,0} < a_{i,1} < \ldots < a_{i,n}, \quad \sum\limits_{j=0}^{n=1} (a_{i,j+1} - a_{i,j}) h_{i,j} = 1,$

$$
i = 1, 2.
$$

Требуется вычислить $p_{\eta_1\eta_2}(x)$. [Мы ограничиваемся случаем, когда случайные величины с вероятностью 1 положительны, так как общий случай вследствие формулы (4.104) приводится к этому]. Можно было бы вычислить $p_{\eta_1\eta_2}(x)$ непосредственно; однако мы используем следующий прием. Известно, что если имеется функция типа (4.105), то ее можно представить в виде

$$
p_{\eta_i}(x) = \sum_{j=1}^{n_i} \alpha_{i,j} f_{i,j}(x), \quad i = 1, 2;
$$

где $f_1(x)$ — плотность равномерного распределения в интервале $(0, a_{ij})$. Отсюда если η_{ij} — случайная величина, равномерно распределенная в интервале $(0, a_{ij})$, причем

η_{ij} независимы в совокупности, то

$$p_{\eta_1\eta_2}(x) = \sum_{j=1}^{n_1} \alpha_{1,j} \sum_{k=1}^{n_2} \alpha_{2,k}\, p_{\eta_{1j}\eta_{2k}}(x) =$$

$$= \begin{cases} \sum\limits_{j=1}^{n_1} \alpha_{1,j} \sum\limits_{k=1}^{n_2} \alpha_{2,k} \dfrac{\ln(a_{1,j}\, a_{2,k}) - \ln x}{a_{1,j}\, a_{2,k}} \mathrm{E}(a_{1,j}\, a_{2,k} - x) & \text{при } x > 0, \\ 0 & \text{при } x < 0, \end{cases} \quad (4.106)$$

где $E(x) = \int\limits_{-\infty}^{x} \delta(t)\, dt = \begin{cases} 1 & \text{при} \quad x > 0 \\ 0 & \text{при} \quad x \leqslant 0. \end{cases}$

Аналогичным образом можно найти плотность распределения произведения любого конечного числа независимых случайных величин с плотностями вида (4.105).

Рассмотрим функцию

$$f_{a,\nu}(x) = \begin{cases} \dfrac{1}{\Gamma(\nu)} a^{\nu} x^{\nu-1} e^{-ax} & \text{при} \quad x > 0, \\ 0 & \text{при} \quad x < 0, \end{cases} \quad (4.107)$$

где $a > 0$, $\nu > 0$ — параметры.

Тогда $f_{a\nu}(x) \geqslant 0$ и $\int\limits_{-\infty}^{\infty} f_{a,\nu}(x)\, d(x) = 1$, так что $f_{a,\nu}(x)$ можно рассматривать как плотность некоторой случайной величины $\gamma = \gamma_{a,\nu}$; $f_{a,\nu}(x)$ называется *плотностью гамма-распределения* с параметрами a, ν. Частный случай гамма-распределения — распределение χ_n^2, плотность которого определяется формулой (4.94). В этом случае $a = \frac{1}{2}$, $\nu = \frac{n}{2}$:

$$p_{\chi_n^2}(x) = f_{\frac{1}{2}, \frac{n}{2}}(x).$$

Найдем плотность распределения случайной величины $\xi_{a,\nu} = \ln \gamma_{a,\nu}$. Имеем

$$p_{\xi_{a,\nu}}(x) = e^x p_{\gamma_{a,\nu}}(e^x) = \frac{1}{\Gamma(\nu)} a^{\nu} e^{\nu x - ae^x}, \quad -\infty < x < \infty. \quad (4.108)$$

Найдем плотность случайной величины $\chi_n = \sqrt{\chi_n^2}$. Имеем

$$p_{\chi_n}(x) = \begin{cases} 2xp_{\chi_n^2}(x^2) & \text{при} \quad x > 0, \\ 0 & \text{при} \quad x < 0, \end{cases} \tag{4.109}$$

так что

$$p_{\chi_n}(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}-1}\,\Gamma\left(\dfrac{n}{2}\right)}\, x^{n-1} e^{-\frac{x^2}{2}} & \text{при} \quad x > 0, \\ 0 & \text{при} \quad x < 0. \end{cases} \tag{4.110}$$

Пусть теперь ξ — нормально распределенная случайная величина с плотностью $p_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $-\infty < x < \infty$ x_n — случайная величина с плотностью (4.110), независимая от ξ. Обозначим $\zeta_n = \frac{\xi}{\chi_n}$. Из формулы (4.62) следует, что

$$p_{\zeta_n}(x) = \int_0^\infty t p_{\chi_n}(t)\, p_\xi(xt)\, dt.$$

Таким образом,

$$p_{\zeta_n}(x) = \frac{1}{2^{\frac{n}{2}-1}\,\Gamma\left(\frac{n}{2}\right)\sqrt{2\pi}} \int_0^\infty t^n e^{-\frac{t^2}{2} - \frac{(xt)^2}{2}}\, dt.$$

С помощью подстановки $u = \frac{1}{2}(1 + x^2)t^2$, $du = (1 + x^2)\, tdt$ этот интеграл приводится к виду

$$p_{\zeta_n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)} (1 + x^2)^{\frac{n+1}{2}}, \quad -\infty < x < \infty. \tag{4.111}$$

Распределение случайной величины ζ_n называется *распределением Стьюдента с n степенями свободы*. Это распределение важно для решения задач математической статистики.

§ 6. РАСПРЕДЕЛЕНИЕ НЕКОТОРЫХ ВЫБОРОЧНЫХ ХАРАКТЕРИСТИК ЧЛЕНОВ ВАРИАЦИОННОГО РЯДА

Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины с общей непрерывной функцией распределения $F(x)$. Расположив ξ_i, $1 \leqslant i \leqslant n$, в порядке возрастания, получим случайные величины $\xi_{(1)} < \xi_{(2)} < \dots < \xi_{(n)}$. Пусть, например, $\xi_1 = 0{,}2$, $\xi_2 = -0{,}11$, $\xi_3 = 0{,}49$. Тогда $\xi_{(1)} = \xi_2 = -0{,}11$, $\xi_{(2)} = \xi_1 = 0{,}2$, $\xi_{(3)} = \xi_3 = 0{,}49$. Конечная последовательность $\{\xi_{(1)}, \xi_{(2)}, \dots, \xi_{(n)}\}$ называется *вариационным рядом*, а $\xi_{(r)}$, $1 \leqslant r \leqslant n$ — r-м членом вариационного ряда. $\xi_{(1)}$ называется минимальным, $\xi_{(n)}$ — максимальным членом вариационного ряда. Очевидно,

$$P(\xi_{(1)} > x) = P(\xi_i > x, \quad 1 \leqslant i \leqslant n) = [1 - F(x)]^n, \tag{4.112}$$

$$P(\xi_{(n)} < x) = P(\xi_i < x, \quad 1 \leqslant i \leqslant n) = [F(x)]^n. \tag{4.113}$$

Если $F(x) = \int\limits_{-\infty}^{x} p_\xi(t)\,dt$, то из формул (4.112) и (4.113) находим:

$$p_{\xi_{(1)}}(x) = np_\xi(x)[1 - F(x)]^{n-1}, \quad p_{\xi_{(n)}}(x) = np_\xi(x)[F(x)]^{n-1}. \tag{4.114}$$

Предположим, что $n \to \infty$, $F(0) = 0$, $p_\xi(x) = 0$ при $x < x_0$, $p_\xi(x) = a + o(1)$ при $x \downarrow x_0$, где $a > 0$.

Рассмотрим случайную величину $\eta = n\ (\xi_1 - x_0)$. Имеем (при $x > 0$)

$$p_\eta(x) = \frac{1}{n} p_{\xi_{(1)}}\left(x_0 + \frac{x}{n}\right) = p_\xi\left(x_0 + \frac{x}{n}\right)\left[1 - F\left(x_0 + \frac{x}{n}\right)\right]^{n-1} = [a + o(1)]\left[1 - \frac{ax}{n} + o\left(\frac{x}{n}\right)\right]^n \xrightarrow[n\to\infty]{} ae^{-ax}, \quad x > 0. \tag{4.115}$$

Таким образом, в принятых предположениях при $n \to \infty$ плотность распределения случайной величины $n\xi_{(1)}$ сходится к плотности показательного распределения с параметром a. Аналогично доказывается, что если $F(x_1) = 1$, $p_\xi(x) = b + o(1)$ при $x \uparrow x_1$, то

$$\lim_{n\to\infty} p_{n(x_1-\xi_{(n)})}(x) = be^{-bx}, \quad x > 0. \tag{4.116}$$

Пусть r — любое целое число, $1 \leqslant r \leqslant n$. Найдем функцию распределения $\xi_{(r)}$. Обозначим число тех ξ_i, $1 \leqslant i \leqslant n$, которые меньше x, через $\nu(x)$. Событие $\{\xi_{(r)} < x\}$ наступает в том и только в том случае, если $\nu(x) \geqslant r$. Однако

$$\nu(x) = \sum_{i=1}^{n} \nu_i(x),$$

где $\nu_i(x) = 1$, если $\xi_i < x$; $\nu_i(x) = 0$, если $\xi_i \geqslant x$. $\nu_1, \nu_2, \ldots, \nu_n$ — независимые случайные величины, причем

$$\mathrm{P}(\nu_i(x) = 1) = \mathrm{P}(\xi_i < x) = F(x).$$

Отсюда находим, что

$$\mathrm{P}(\nu(x) = k) = C_n^k [F(x)]^k [1 - F(x)]^{n-k}.$$

Окончательно

$$F_{\xi_{(r)}}(x) = \sum_{k=r}^{n} \mathrm{P}\{\nu(x) = k\} = \sum_{k=r}^{n} C_n^k [F(x)]^k [1 - F(x)]^{n-k}. \tag{4.117}$$

Предположим, что функция $F(x)$ дифференцируема. Тогда из последнего равенства найдем

$$p_{\xi_{(r)}}(x) = \sum_{k=r}^{n} C_n^k \{k [F(x)]^{k-1} [1 - F(x)]^{n-k} -$$

$$-(n-k)[F(x)]^k [1 - F(x)]^{n-k-1}\} p_\xi(x) = \sum_{k=r-1}^{n-1} (k+1) C_n^{k+1} \times$$

$$\times [F(x)]^k [1 - F(x)]^{n-k-1} p_\xi(x) -$$

$$-\sum_{k=r}^{n} (n-k) C_n^k [F(x)]^k [1 - F(x)]^{n-k-1} p_\xi(x)$$

(в первой сумме вместо k в качестве переменной суммирования взято $k-1$). Однако

$$\frac{(k+1)\,C_n^{k+1}}{(n-k)\,C_n^k}=\frac{(k+1)\,n\,!\,k\,!\,(n-k!)}{(k+1)\,!\,(n-k-1)\,!\,(n-k)\,n!}=1.$$

Затем вторую сумму в последнем выражении (4.118) можно рассматривать как $\sum\limits_{k=r}^{n-1}$, так как ее n-й член равен 0.

После сокращения подобных членов находим

$$p_{\xi_{(r)}}(x)=rC_n^r\,[F(x)]^{r-1}\,[1-F(x)]^{n-r}\,p_\xi(x) \qquad (4.119)$$

Пусть теперь $n\to\infty$, $\frac{r}{n}=p$, $0<p<1$. Обозначим через x_p корень уравнения $F_\xi(x)=p$ (он называется p-квантилью распределения случайной величины ξ) и предположим, что в окрестности точки x_p

$$p_\xi(x)=a+o(x-x_p)\,(x\to x_p),$$

где $a>0$.

Тогда можно записать

$$F(x)=F(x_p)+\int\limits_{x_p}^{x}p_\xi(t)\,dt=p+a(x-x_p)+o(x-x_p); \qquad (4.120)$$

соответственно если обозначить $q=1-p$, то

$$1-F(x)=q-a\,(x-x_p)+o\,(x-x_p). \qquad (4.121)$$

Пусть

$$x=x_p+\frac{z}{\sqrt{n}}, \qquad (4.122)$$

где z — ограниченная величина: $|z|\leqslant A$. Тогда на основании формул (4.120) и (4.121) имеем

$$p'=F(x_p)=p+\frac{az}{\sqrt{n}}+o\left(\frac{1}{\sqrt{n}}\right),\quad q'=1-F(x_p)=$$

$$=q-\frac{az}{\sqrt{n}}+o\left(\frac{1}{\sqrt{n}}\right).$$

Далее, $r = np = np' + n(p - p') = np' - az\sqrt{n} + o(\sqrt{n})$, так что

$$r = np' - a'z\sqrt{n}, \qquad (4.123)$$

где $a' \to a$ при $n \to \infty$. Отсюда на основании формулы (4.117)

$$F_{\xi_{(r)}}(x) = \sum_{k=np'-a'z\sqrt{n}}^{n} C_n^k (p')^k (1-p')^{n-k}. \qquad (4.124)$$

По интегральной предельной теореме Муавра — Лапласа,

$$F_{\xi_{(r)}}(x) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\frac{az}{\sqrt{pq}}}^{\infty} e^{-\frac{u^2}{2}}\,du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{az}{\sqrt{pq}}} e^{-\frac{u^2}{2}}\,du. \qquad (4.125)$$

Следовательно, функция распределения случайной величины $\frac{\xi_{(r)} - x_p}{\sqrt{n}}$ стремится при $n \to \infty$ к функции распределения нормального закона с математическим ожиданием 0 и дисперсией $\frac{pq}{p_\xi^2(x_p)}$.*

Размахом конечной последовательности независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ называется случайная величина $R = \xi_{(n)} - \xi_{(1)}$. Найдем функцию распределения случайной величины R, предположив, что функция $F_\xi(x)$ непрерывна. Обозначим

* Строго говоря, здесь используется несколько более сильный вариант теоремы Муавра — Лапласа: если p' — вероятность успеха в одном испытании, $q' = 1 - p'$, причем при $n \to \infty$ $p' \to p \in (0,1)$, то для распределения числа ν успехов в n независимых испытаниях имеет место формула

$$\mathrm{P}\left\{\frac{\nu - np'}{\sqrt{np'q'}} < x\right\} \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,du.$$

Для доказательства этого предложения достаточно заметить, что оценки отклонения допредельного распределения от предельного (см. гл. II), используемые при доказательстве интегральной предельной теоремы, равномерны относительно p в любом интервале $(\varepsilon, 1-\varepsilon)$, где $\varepsilon > 0$ фиксировано.

через D область n-мерного пространства, включающую те и только те точки $(x_1, x_2, \ldots, x_n)$, для которых при $\xi_i = x_i$, $1 \leqslant i \leqslant n$, $\xi_{(n)} - \xi_{(1)} < x$. Тогда

$$D = \bigcup_{i \neq j} \left\{ x_i = \min_{1<k<n} x_k, \quad x_j = \max_{1<k<n} x_k, \quad x_j - x_i < x \right\}.$$

Следовательно,

$$F_R(x) = \sum_{i \neq j} \int \ldots \int_{\substack{x_i = \min x_k \\ x_j = \max x_k \\ x_j - x_i < x}} p_\xi(x_1) p_\xi(x_2) \ldots p_\xi(x_n)\, dx_1\, dx_2 \ldots dx_n. \tag{4.126}$$

Легко видеть, что все $n(n-1)$ слагаемых в правой части формулы (4.126) равны друг другу. Поэтому

$$F_R(x) = n(n-1) \iint_{0<x_1-x_2<x} p_\xi(x_1) p_\xi(x_2) \int \ldots \int_{\substack{x_1<x_3<x_2 \\ \cdots\cdots \\ x_1<x_n<x_2}} p_\xi(x_3) \times$$

$$\times\, p_\xi(x_4) \ldots p_\xi(x_n)\, dx_3 \ldots dx_n =$$

$$= n(n-1) \iint_{0<x_1-x_2<x} [F_\xi(x_2) - F_\xi(x_1)]^{n-2} p_\xi(x_1)\, p_\xi(x_2) dx_1 dx_2. \tag{4.127}$$

В частноoти, при $n = 2$ находим

$$F_R(x) = F_{|\xi_2-\xi_1|}\, x = 2 \iint_{0<x_1-x_2<x} p_\xi(x_1)\, p_\xi(x_2)\, dx_1 dx_2 =$$

$$= 2 \int_{-\infty}^{\infty} p_\xi(x_2)\, d_{x_2} \int_{x_2}^{x_2+x} p_\xi(x_1)\, dx_1 = 2 \int_{-\infty}^{\infty} [F_\xi(x+t) -$$

$$- F_\xi(t)] \times p_\xi(t)\, dt. \tag{4.128}$$

§ 7. НЕКОТОРЫЕ ФУНКЦИОНАЛЫ ОТ ПОСЛЕДОВАТЕЛЬНОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН

1. Пусть $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — независимые случайные величины с общей плотностью $p_\xi(x)$, ν = независимая от $\{\xi_n\}$ случайная величина с возможными значениями 0, 1, 2, ... и распределением $p_n = \mathrm{P}\{\nu = n\}$, $n \geqslant 0$. Рас-

смотрим случайную величину

$$\zeta = \xi_1 + \xi_2 + \ldots + \xi_\nu,^* \tag{4.129}$$

т. е. сумму случайного числа случайных величин. В силу независимости $\{\xi_n\}$ и ν

$$F_\zeta(x) = \sum_{n=0}^{\infty} \mathrm{P}\{\nu = n\}\,\mathrm{P}\{\xi_1 + \xi_2 + \ldots + \xi_n < x\}. \tag{4.130}$$

Левую часть этого равенства можно представить как

$$\int_{-\infty}^{x} p_\xi(t)\,dt,$$

а правую — как

$$\sum_{n=0}^{\infty} p_n \int_{-\infty}^{x} p_{\xi_1+\ldots+\xi_n}(t)\,dt = \int_{-\infty}^{x} \sum_{n=0}^{\infty} p_n\, p_{\xi_1+\ldots+\xi_n}(t)\,dt. \tag{4.131}$$

(Перестановка знаков суммирования и интегрирования законна, поскольку

$$p_n \int_{-\infty}^{x} p_{\xi_1+\ldots+\xi_n}(t)\,dt \leqslant p_n \int_{-\infty}^{\infty} p_{\xi_1+\ldots+\xi_n}(t)\,dt = p_n,$$

а ряд $\Sigma\, p_n$ сходится.) Отсюда

$$p_\zeta(x) = \sum_{n=0}^{\infty} p_n\, p_{\xi_1+\ldots+\xi_n}(x). \tag{4.132}$$

Пусть $\varphi_\xi(t)$ — характеристическая функция любой из случайных величин ξ_n и $\varphi_\zeta(t)$ — характеристическая функция ζ. На основании равенства (4.132) находим, что

$$\varphi_\zeta(t) = \sum_{n=0}^{\infty} p_n\, \varphi_\xi^n(t)^{**}, \tag{4.133}$$

Пусть, например,

$$p_\xi(x) = e^{-x}, \quad x > 0; \quad p_n = \frac{\lambda^n}{n!}e^{-\lambda}, \quad n = 0, 1, 2, \ldots .$$

* Сумма нулевого числа слагаемых всегда определяется как 0.

** При $n = 0$ $\varphi_\xi^n(t) = 1$, что соответствует характеристической функции случайной величины, равной 0 с вероятностью 1.

Тогда

$$\varphi_\xi(t) = \int_0^\infty e^{itx-x}\,dx = \frac{1}{1-it},$$

откуда

$$\varphi_\zeta(t) = e^{-\lambda}\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{\lambda}{1-it}\right)^n = \exp\left\{\lambda\left(\frac{1}{1-it}-1\right)\right\} =$$

$$-\exp\left\{\frac{it}{1-it}\right\}. \qquad (4.134)$$

2. Решим следующую задачу. Пусть $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — независимые случайные величины с плотностью $p_\xi(x)$. Обозначим $S_k = \xi_1 + \xi_2 + \ldots + \xi_k$. Требуется найти вероятность $G_n(x)$ того, что одновременно $S_1 < x_1\, S_2 < < x_1 \ldots S_n < x$, или, что то же, $\max\limits_{1<k<n} S_k < x$.

Поставленная задача важна для многих прикладных вопросов. Пусть, например, рассматривается некоторое устройство периодического использования. Устройство характеризуется некоторым параметром θ, который в начальный момент равен 0. Если $\theta < x$, устройство исправно, если $\theta > x$ — неисправно. При каждом использовании θ смещается на случайную величину ξ, так что значение параметра после k-кратного использования устройства равно $\xi_1 + \xi_2 + \ldots + \xi_k$, где $\xi_1, \xi_2 \ldots$ — независимые случайные величины. При такой интерпретации $G_n(x)$ есть вероятность того, что устройство не выйдет из строя при n-кратном использовании.

Можно записать

$$G_n(x) = \int\limits_{\substack{x_1<x,\, x_1+x_2<x,\ldots \\ \ldots,\, x_1+\ldots+x_n<x}} \cdots \int p_\xi(x_1)\,p_\xi(x_2)\ldots p_\xi(x_n)\,dx_1\,dx_2\ldots dx_n =$$

$$= \int_{-\infty}^{x} p_\xi(x_1)dx_1 \int\limits_{\substack{x_2<x-x_1,\ldots \\ \ldots\, x_2+\ldots+x_n<x-x_1}} \cdots \int p_\xi(x_2)\ldots p_\xi(x_n)\,dx_2\ldots dx_n. \qquad (4.135)$$

Взяв $n-1$ вместо n, получим

$$\int\limits_{\substack{x_2<x-x_1\ldots \\ \ldots,\, x_2+\ldots+x_n<x-x_1}} \cdots \int p_\xi(x_2)\ldots p_\xi(x_n)\,dx_2\ldots dx_n = G_{n-1}(x-x_1) \qquad (4.136)$$

Отсюда

$$G_n(x) = \int_{-\infty}^{x} G_{n-1}(x-t)\, p_\xi(t)\, dt. \qquad (4.137)$$

По этой рекуррентной формуле можно последовательно вычислить интересующую нас вероятность при любом n, если использовать очевидное равенство

$$G_1(x) = F_\xi(x). \qquad (4.138)$$

Интересно сравнить формулу (4.137) с формулой для функции распределения $F_n(x)$ суммы n случайных величин $\xi_1, \xi_2, \ldots, \xi_n$:

$$F_n(x) = \int_{-\infty}^{\infty} F_{n-1}(x-t)\, p_\xi(t)\, dt. \qquad (4.139)$$

Формулы эти отличаются интервалом, по которому производится интегрирование. Если же $\xi_i > 0$ с вероятностью 1, то обе формулы совпадают. Это закономерно, поскольку для положительных ξ_i $\max\limits_{1<k<n} S_k = S_n$.

§ 8. АППРОКСИМАЦИЯ РАСПРЕДЕЛЕНИЯ ПОЛОЖИТЕЛЬНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ ГИПЕРЭРЛАНГОВСКИМИ РАСПРЕДЕЛЕНИЯМИ

Рассуждения этого параграфа справедливы для любых распределений, а не только непрерывно-дискретных.

Напомним, что если $\xi_1, \xi_2, \ldots, \xi_n$ — независимые случайные величины с плотностью $\lambda e^{-\lambda x}$, $x > 0$, то

$$p_{\xi_1+\ldots+\xi_n}(x) = \frac{\lambda(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x}, \quad x > 0$$

[формула (4.57)]. Пусть $p_1, p_2, \ldots, p_k$ — неотрицательные числа, в сумме составляющие единицу. Рассмотрим распределение с плотностью

$$p_\xi(x) = \sum_{i=1}^{k} p_i \frac{\lambda_i(\lambda_i x)^{n_i-1}}{(n_i-1)!} e^{-\lambda_i x}, \quad x > 0. \qquad (4.140)$$

Это распределение называется *гиперэрланговским распределением* (смесью эрланговских распределений). При решении различных теоретических и прикладных задач с гиперэрланговским распределением оперировать проще, чем с самыми общими распределениями. Одним из полезных свойств гиперэрланговского распределения является то, что характеристическая функция ξ в этом случае дробно-рациональна относительно t, т. е. представляет собой отношение двух многочленов. Действительно,

$$\varphi_\xi(t) = \int_0^\infty e^{itx} p_\xi(x)\,dx = \sum_{i=1}^{k} p_i \varphi_{\xi_{i,1}+\xi_{i,2}+\dots+\xi_{i,n_i}}(t) =$$
$$= \sum_{i=1}^{k} p_i \left[\varphi_{\xi_{i,1}}(t)\right]^{n_i}, \tag{4.141}$$

где $\xi_{i,j}$ — показательно распределенные случайные величины с параметром λ_i. Далее,

$$\varphi_{\xi_{i,j}}(t) = \lambda_i \int_0^\infty e^{itx-\lambda_i x}\,dx = \frac{\lambda_i}{\lambda_i - it}, \tag{4.142}$$

откуда

$$\varphi_\xi(t) = \sum_{i=1}^{k} p_i \left(\frac{\lambda_i}{\lambda_i - it}\right)^{n_i}. \tag{4.143}$$

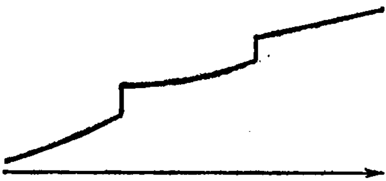

Рис. 4.8

Приведя эту сумму к общему знаменателю, приходим к требуемому выражению.

Рассмотрим вопрос о возможности аппроксимации произвольной функции распределения функцией распределения, соответствующей плотности вида (4.140).

Обычно аппроксимации функций строятся на том или ином понятии отклонения аппроксимирующей функции от аппроксимируемой. Пусть $F(x)$, $G(x)$ — любые две функции распределения. Построим график кривой $y = F(x)$, дополнив его в точках разрыва вертикальными отрезками (рис. 4.8). Затем от каждой точки этого графика отложим по обе стороны отрезки длины $\varepsilon\sqrt{2}$ в направлении биссектрисы второго и четвертого координатных углов. Построенные отрезки сплошь заполнят некоторую полосу.

Если график функции $y = G(x)$ целиком находится в данной полосе, будем говорить, что $\rho[F, G] \leqslant \varepsilon$; если хотя бы одна точка графика $y = G(x)$ выходит за пределы полосы, то $\rho[F, G] > \varepsilon$. $\rho[F, G]$ есть точная нижняя грань тех $\varepsilon \geqslant 0$, для которых $\rho[F, G] \leqslant \varepsilon$ ($\rho[F, G]$ называется расстоянием Леви между функциями распределения $F(x)$ и $G(x)$. Если $\rho[F, G] \leqslant \varepsilon$, то это равносильно тому, что для любого x, $-\infty < x < \infty$,

$$F(x-\varepsilon) - \varepsilon \leqslant G(x) \leqslant F(x+\varepsilon) + \varepsilon. \tag{4.144}$$

Можно показать, что если $F_n(x)$, $n \geqslant 1$, — функции распределения, то $F_n(x) \xrightarrow[n\to\infty]{} F(x)$ в любой точке непрерывности $F(x)$ в том и только том случае, когда $\rho[F_n, F] \xrightarrow[n\to\infty]{} 0$.

Лемма 1. Пусть

$$\tau > 0, \quad E_\tau(x) = \int_{-\infty}^{x} \delta(t-\tau)\,dt,$$

$$p_{n,\tau}(x) = \frac{1}{(n-1)!}\left(\frac{n}{\tau}\right)^n x^{n-1} e^{-\frac{nx}{\tau}}, \quad x > 0, \tag{4.145}$$

$$p_{n,\tau}(x) = 0, \quad x < 0, \quad F_{n,\tau}(x) = \int_{-\infty}^{x} p_{n,\tau}(t)\,dt.$$

Тогда

$$\rho\,[E_\tau,\ F_{n,\tau}] \underset{n\to\infty}{\longrightarrow} 0 \tag{4.146}$$

Доказательство. Обозначим через $\xi_{n,\tau}$ случайную величину с функцией распределения $F_{n,\tau}(x)$ и оценим вероятность события $\{|\xi_{n,\tau}-\tau| \geqslant \varepsilon\}$, где $\varepsilon > 0$ — фиксированное число. Пусть $x = \tau + z$.

Использовав формулу Стирлинга, можно записать

$$p_{n,\tau}(x) \backsim \frac{1}{\sqrt{2\pi(n-1)}}\left(\frac{e}{n-1}\right)^{n-1}\left(\frac{n}{\tau}\right)^n (\tau+z)^{n-1} e^{-\frac{n(\tau+z)}{\tau}} =$$

$$= \frac{1}{e}\left(\frac{n}{n-1}\right)^n \frac{1}{\tau+z}\sqrt{\frac{n-1}{2\pi}}\,R^n(z),$$

где $R(z) = e^{-\frac{z}{\tau}}\left(1+\frac{z}{\tau}\right)$. Заметим, что $e\left(\frac{n-1}{n}\right)^n \underset{n\to\infty}{\longrightarrow} 1$. Далее, дифференцированием легко проверить, что функция $R(z)$ в окрестности точки $z = 0$ допускает представление

$$R(z) = e^{-\frac{z^2}{2\tau^2} + O(z^3)}, \tag{4.147}$$

Пусть $|z| < \frac{1}{n^{3/8}}$. Тогда

$$e^{nO(z^3)} = e^{O\left(n^{-\frac{1}{8}}\right)} = 1 + o(1) \tag{4.148}$$

Следовательно,

$$\int_{\tau-n^{-3/8}}^{\tau+n^{-3/8}} p_{n,\tau}(x)\,dx \backsim \frac{1}{\tau}\sqrt{\frac{n}{2\pi}}\int_{n^{-3/8}}^{n^{-3/8}} e^{-\frac{nz^2}{2\tau^2}}\,dz =$$

$$= \frac{1}{\sqrt{2\pi}} \int\limits_{-\frac{1}{\tau} n^{1/8}}^{\frac{1}{\tau} n^{1/8}} e^{-\frac{u^2}{2}} du \underset{n \to \infty}{\longrightarrow} 1. \qquad (4.149)$$

Отсюда

$$P\left(\left| \xi_{n_1 \tau} - \tau \right| \geqslant n^{-1/8} \right) \leqslant \sqrt{\frac{2}{\pi}} \int\limits_{\frac{1}{\tau} n^{1/8}}^{\infty} e^{-\frac{u^2}{2}} du \underset{n \to \infty}{\longrightarrow} 0 \qquad (4.150)$$

и тем более при фиксированном $\varepsilon > 0$

$$P\left(\left| \xi_{n_1 \tau} - \tau \right| \geqslant \varepsilon \right) \leqslant \varepsilon, \quad n > n(\varepsilon). \qquad (4.151)$$

Из оценки (4.151) находим, что $F_{n_1 \tau}(x)$ заключена в полосе, показанной на рис. 4.9 штриховкой. Очевидно, что из этого следует соотношение $\rho[E_\tau, F_{n_1 \tau}] \leqslant \varepsilon$. Ввиду произвольности $\varepsilon > 0$ требуемое доказано.

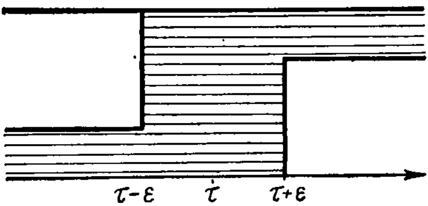

Рис. 4.9

(Заметим, что данную лемму можно доказать значительно проще на основании неравенства Чебышева, которое будет доказано в главе, посвященной закону больших чисел.)

Лемма 2. Пусть $F_i(x)$, $G_i(x)$, $1 \leqslant i < m$ — функции распределения,

$$F(x) = \sum_{i=1}^{m} p_i F_i(x), \quad G(x) = \sum_{i=1}^{m} p_i G_i(x),$$

где $p_i \geqslant 0$, $\sum\limits_{i=1}^{m} p_i = 1$. Тогда, если $\rho[F_i, G_i] < \varepsilon$, то также $\rho[F, G] < \varepsilon$.

Доказательство. Просуммировав двойные неравенства

$$F_i(x - \varepsilon) - \varepsilon < G_i(x) < F_i(x + \varepsilon) + \varepsilon$$

с коэффициентами p_i, придем к аналогичному неравенству для $F(x)$, $G(x)$. Лемма доказана.

Лемма 3. Для любой функции распределения $F(x)$ и любого $\varepsilon > 0$ найдется функция распределения $G(x)$ дискретной случайной величины, принимающей конечное число значений, и такая, что $\rho[F, G] < \varepsilon$. При этом, если $F(x_0) = 0$, то можно также удовлетворить условию $G(x_0 + \varepsilon) = 0$.

Доказательство очевидно из рис. 4.10: график требуемой функции $G(x)$ представлен в виде ломаной линии попеременно касающейся графиков $y = F(x - \varepsilon) - \varepsilon$ и $y = F(x + \varepsilon) + \varepsilon$. Величина скачка в каждой точке разрыва $G(x)$, кроме, возможно, крайней правой, не меньше 2ε, так что всего возможных значений данной случайной величины не более $\frac{1}{2\varepsilon} + 1$. В то же время $G(\varepsilon) = 0$.

Лемма 4. Пусть $F(x)$, $G(x)$, $H(x)$ — три любые функции распределения. Тогда

$$\rho[F, H] \leqslant \rho[F, G] + \rho[G, H]^{*} \qquad (4.152)$$

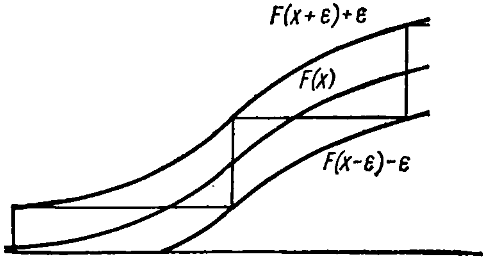

Рис. 4.10

Доказательство. Пусть $\rho[F, G] = h$, $\rho[G, H] = k$. Тогда

$$F(x - h) - h \leqslant G(x), \qquad (4.153)$$

$$G(x - k) - h \leqslant H(x). \qquad (4.154)$$

В неравенстве (4.153) положим $x - k$ вместо x. Тогда

$$F(x - h - k) - h \leqslant G(x - k). \qquad (4.155)$$

Использовав неравенство (4.154), найдем

$$F(x - h - k) - h \leqslant H(x) + k,$$

или, что то же,

$$F(x - h - k) - h - k \leqslant H(x). \qquad (4.156)$$

Аналогично доказывается и справедливость неравенства

$$H(x) \leqslant F(x + h + k) + h + k. \qquad (4.157)$$

Неравенства (4.156) и (4.157) в совокупности приводят к выводу, что $\rho[F, H] \leqslant h + k$. Последнее неравенство эквивалентно неравенству (1.152). Лемма доказана.

Теорема 4.1. *Пусть $F(x)$ — функция распределения положительной случайной величины, $\varepsilon > 0$ — любое число. Тогда найдется такое гиперэрланговское распределение $H(x)$, что $\rho[F, H] \leqslant \varepsilon$.*

Доказательство. Выберем произвольные положитель-

* Данное неравенство носит название «неравенства треугольника» и выполняется при любом естественном определении расстояния между объектами любой абстрактной природы.

ные h и k, в сумме составляющие ε. По лемме 3 построим функцию

$$G(x) = \sum_{i=1}^{m} p_i E_{\tau_i}(x),$$

где все $\tau_i > 0$, $\rho\,[F, G] \leqslant h$. По лемме 1, при достаточно большом n

$$\rho\left[E_{\tau_i}\, F_{n,\,\tau_i}\right] \leqslant k.$$

По лемме 2, тогда и $\rho\,[G, H] \leqslant k$, где $H(x) = \sum\limits_{i=1}^{m} p_i\, F_{n,\,\tau_i}(x)$.

Наконец, по лемме 4 имеем

$$\rho\,[F, H] \leqslant \rho\,[F, G] + \rho\,[G, H] \leqslant h + k = \varepsilon.$$

Теорема доказана.

В практике, однако, аппроксимация в смысле расстояния Леви не всегда удовлетворительна; в частности, это обстоятельство имеет место в случае аппроксимации случайных слагаемых некоторой суммы. Гиперэрланговские распределения можно использовать и в случае более жестких требований к аппроксимирующему распределению. Приведем соответствующую теорему.

Теорема 4.2. *Пусть $F(x)$ — функция распределения положительной случайной величины, первые r моментов которой конечны. Тогда по любому $\varepsilon > 0$ найдется такое гипергеометрическое распределение $H(x)$, что $\rho\,[F, H] \leqslant \varepsilon$ и при этом первые r моментов распределения $H(x)$ отличаются не более чем на ε от соответствующих моментов распределения $F(x)$.*

Доказательство. Построим функции $G(x)$ и $H(x)$ — те же, что и при доказательстве предыдущей теоремы. Для j-го момента распределения $F_{n,\,\tau}(x)$ находим выражение

$$M\xi^j_{n,\,\tau} = \int_0^\infty x^j\, p_{n,\,\tau}(x)\, dx = \frac{1}{(n-1)!}\left(\frac{n}{\tau}\right)^n \int_0^\infty x^{n+j-1}\, e^{-\frac{nx}{\tau}}\, dx =$$

$$= \frac{1}{(n-1)!}\left(\frac{\tau}{n}\right)^j \int_0^\infty z^{n+j-1}\, e^{-z}\, dz = \frac{(n+j-1)!}{(n-1)!}\left(\frac{\tau}{n}\right)^j =$$

$$= \frac{n}{n}\,\frac{n+1}{n}\,\cdots\,\frac{n+j-1}{n}\,\tau^j \underset{n\to\infty}{\longrightarrow} \tau^j. \qquad (4.158)$$

Отсюда j-й момент распределения

$$\int_0^\infty x^j \sum_{i=1}^{m} p_i\, p_{n,\,\tau_i}(x)\, dx = \sum_{i=1}^{m} p_i\, M\xi^j_{n,\,\tau_i} \underset{n\to\infty}{\longrightarrow} \sum_{i=1}^{m} p_i\, \tau_i^j. \qquad (4.159)$$

Этот предел и есть j-й момент распределения $G(x)$. Итак, при фиксированном r моменты распределения $H(x)$ первых r порядков можно сделать отличающимися не более чем на заданное $\varepsilon > 0$ от соответствующих моментов распределения $G(x)$.

Остается доказать, что можно сделать так, чтобы моменты распределения $G(x)$ первых r порядков сколь угодно мало отличались от соответствующих моментов $F(x)$.

Построим, как показано на рис. 4.11, ступенчатую функцию $G(x)$:

$$G(x) = F(k\Delta), \text{ если } (k-1)\Delta \leqslant x < k\Delta, \quad 2 \leqslant k \leqslant N;$$

$$G(x) = F(\Delta), \text{ если } 0 \leqslant x < \Delta; \quad G(x) = 1, \text{ если } x > N\Delta.$$

Пусть $p_\xi(x)$ — плотность, соответствующая распределению $F(x)$*. Тогда

$$M\xi^j = \int\limits_0^\infty x^j p_\xi(x)\,dx.$$

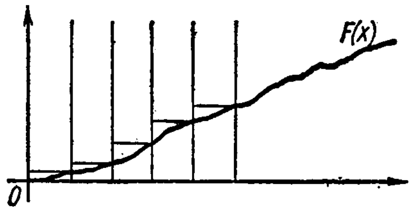

Рис. 4.11

Выберем столь большое $N\Delta$, чтобы $\int\limits_{N\Delta}^\infty x^j p_\xi(x)\,dx < \frac{\varepsilon}{2}$ ($j = 1, 2, \ldots, r$). Возьмем вспомогательную случайную величину η с плотностью

$$p_\eta(x) = p_\xi(x)\,E(N\Delta - x) + [1 - F_\xi(N\Delta)]\,\delta(x - N\Delta),$$

считая, что $P\{\xi = N\Delta\} = 0$. Тогда

$$M\eta^j = \int\limits_0^{N\Delta} x^j p_\xi(x)\,dx + (N\Delta)^j [1 - F_\xi(N\Delta)];$$

далее,

$$(N\Delta)^j [1 - F_\xi(N\Delta)] \leqslant \int\limits_{N\Delta}^\infty x^j p_\xi(x)\,dx.$$

Отсюда

$$|M\eta^j - M\xi^j| < \frac{\varepsilon}{2}. \tag{4.160}$$

Пусть ζ — случайная величина, определенная следующим образом:

$$\zeta = k\Delta, \text{ если } (k-1)\Delta \leqslant \eta < k\Delta, \quad 2 \leqslant k \leqslant N;$$

$$\zeta = \Delta, \text{ если } 0 \leqslant \eta < \Delta, \quad G(x) = F_\zeta(x).$$

---

*. Нижеследующее рассуждение справедливо в самом общем случае; предположение о существовании плотности используется лишь для того, чтобы не использовать обозначений, выходящих за рамки основного материала данной главы.

Тогда по теореме о среднем, известной из дифференциального исчисления,

$$|\zeta - \eta| \leqslant \Delta |\zeta^j - \eta^j| \leqslant j (N\Delta)^j \Delta, \qquad (4.161)$$

Выберем такое Δ, чтобы $j (N\Delta)^j \Delta < \dfrac{\varepsilon}{2}$, $1 \leqslant j \leqslant r$.

Тогда из второго неравенства (4.161) следует, что

$$|M\zeta^j - M\eta^j| < \frac{\varepsilon}{2}, \quad 1 \leqslant j \leqslant r$$

В сочетании с неравенством (4.160) это неравенство приводит к оценке

$$|M\zeta^j - M\xi^j| < \varepsilon, \quad 1 \leqslant j \leqslant r.$$

Утверждение теоремы доказано, поскольку неравенство $\rho\,[F, H] < \varepsilon$ при достаточно большом n составляет утверждение предыдущей теоремы.

ГЛАВА V

МНОГОМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§ 1. ОПРЕДЕЛЕНИЕ МНОГОМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Пусть $(\Omega, \mathfrak{A}, \mathrm{P})$ — вероятностное пространство. Пусть на этом пространстве заданы n случайных величин

$$\xi_1 = \xi_1(\omega),\ \xi_2 = \xi_2(\omega),\ \dots,\ \xi_n = \xi_n(\omega),\ \omega \in \Omega.$$

Совокупность этих величин $\xi(\omega) = \xi = (\xi_1, \xi_2, \dots, \xi_n)$ называется *многомерной* (n-мерной) *случайной величиной*. При этом ξ_i называется *компонентами* ξ.

Основной характеристикой многомерной случайной величины является ее функция распределения.

Функция n аргументов

$$F_\xi(x_1, x_2, \dots, x_n) = \mathrm{P}(\xi_1 < x_1,\ \xi_2 < x_2,\ \dots,\ \xi_n < x_n) \tag{5.1}$$

называется (многомерной) *функцией распределения n-мерной случайной величины* $\xi = (\xi_1, \xi_2, \dots, \xi_n)$.

Докажем, что функция распределения многомерной случайной величины всегда имеет определенный смысл, или, что то же, при любых $x_1, x_2, \dots, x_n$ событие вида $\{\xi_1 < x_1,\ \xi_2 < x_2,\ \dots,\ \xi_n < x_n\}$ принадлежит σ-алгебре событий $\mathfrak{A}$, для которых определены вероятности. Это утверждение доказывается весьма просто. Имеем

$$\{\xi_1 < x_1,\ \xi_2 < x_2,\ \dots,\ \xi_n < x_n\} = \bigcap_{i=1}^{n} \{\xi_i < x_i\}.$$

Так как при каждом i, $1 \leqslant i \leqslant n$, ξ_i есть случайная величина, то $\{\xi_i < x_i\} \in \mathfrak{A}$. Таким образом, интересующее нас событие представляет собой пересечение n событий из $\mathfrak{A}$. По свойству σ-алгебры оно принадлежит $\mathfrak{A}$.

Заметим следующее. Требование, чтобы все $\xi_i(\omega)$ были определены на одном и том же вероятностном пространстве, существенно.

Впрочем, существует прием, позволяющий конструировать пространство, соответствующее многомерной случайной величине, из пространств, соответствующих компонентам.

Пусть $(\Omega_1, \mathfrak{A}_1, P_1)$, $(\Omega_2, \mathfrak{A}_2, P_2)$ — два вероятностных пространства. Рассмотрим множество $\Omega = \Omega_1 \times \Omega_2$ элементов $\omega = (\omega_1, \omega_2)$, где $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$. Далее, рассмотрим класс K подмножеств Ω вида $\{(\omega_1, \omega_2) : \omega_1 \in A_1, \ \omega_2 \in A_2\}$ при всевозможных $A_1 \in \mathfrak{A}$, $A_2 \in \mathfrak{A}$. Пусть $\mathfrak{A}$ — минимальная σ-алгебра подмножеств Ω, содержащая все множества из класса K. Наконец, определим на всех $A \in \mathfrak{A}$ вероятность $P(A)$, задавшись условием, что если $A \in K$, а именно $A = \{(\omega_1, \omega_2) : \omega_1 \in A_1, \ \omega_2 \in A_2\}$, то $P(A) = P(\omega_1 \in A_1) P(\omega_2 \in A_2)$. Можно показать, что этим условием вероятности $P(A)$, $A \in \mathfrak{A}$ определяются однозначно. Вероятностное пространство $(\Omega, \mathfrak{A}, P)$ называется *прямым произведением вероятностных пространств* $(\Omega_1, \mathfrak{A}_1, P_1)$ и $(\Omega_2, \mathfrak{A}_2, P_2)$. Таким образом, если заданы случайные величины $(\xi_1, \xi_2, \ldots, \xi_n)$ и $(\eta_1, \eta_2, \ldots, \eta_m)$ на пространствах Ω_1 и Ω_2: $\xi_i = \xi_i(\omega_1)$, $1 \leqslant i \leqslant n$, $\eta_j = \eta_j(\omega_2)$, $1 \leqslant j \leqslant m$, то оказывается возможным построить $(n+m)$-мерную случайную величину

$$\zeta = (\xi_1, \xi_2, \ldots, \xi_n, \eta_1, \eta_2, \ldots, \eta_m),$$

где

$$\xi_i = \xi_i(\omega_1) \equiv \xi_i(\omega_1, \omega_2), \quad \eta_j = \eta_j(\omega_2) \equiv \eta_j(\omega_1, \omega_2).$$

Такая конструкция имеет содержательный смысл только при условии, что случайности, определяющие значения $(\xi_1, \xi_2, \ldots, \xi_n)$ и $(\eta_1, \eta_2, \ldots, \eta_m)$, могут рассматриваться, как физически не связанные.

Пусть $F_\xi(x_1, x_2, \ldots, x_n)$ — функция распределения многомерной случайной величины $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$. Возьмем любое m, $1 \leqslant m < n$ и положим $\eta = (\xi_1, \xi_2, \ldots, \xi_m)$. Тогда

$$F_\eta(x_1, x_2, \ldots, x_m) = \lim_{\substack{x_{m+1} \to \infty \\ \cdots\cdots \\ x_n \to \infty}} F_\xi(x_1, x_2, \ldots,$$

$$\ldots x_m, x_{m+1}, \ldots, x_n) = F_\xi(x_1, x_2, \ldots, x_m, \infty, \ldots, \infty). \tag{5.2}$$

[Пусть $x_j^{(N)} \to \infty$, $m+1 \leqslant j \leqslant n$, при $N \to \infty$. Поскольку ξ_i конечны с вероятностью 1, то

$$\bigcup_{N=1}^{\infty} \{\xi_1 < x_1, \ldots, \xi_m < x_m, \ \xi_{m+1} < x_{m+1}^{(N)}, \ldots, \xi_n < x_n^{(N)}\} =$$
$$= \{\xi_1 < x_1, \ldots, \xi_m < x_m, \ \xi_{m+1} < \infty, \ldots, \xi_n < \infty\} =$$
$$= \{\xi_1 < x_1, \ldots, \xi_m < x_m\},$$

так что равенство (5.2) справедливо вследствие аксиомы непрерывности.]

Очевидно, аналогичное свойство имеет место, если x_j устремить к ∞ по любой совокупности индексов $j \in J$. В частности,

$$F_{\xi_i}(x_i) = F_\xi(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty).$$

Таким образом, $F_\xi(x_1, \ldots, x_n)$ однозначно определяет $F_{\xi_i}(x_i)$; обратное, вообще говоря, неверно, что можно видеть на таком примере. Пусть ξ_1, ξ_2 — независимые случайные величины, равномерно распределенные в интервале (0,1). Образуем двумерные случайные величины $\xi = (\xi_1, \xi_2)$ и $\eta = (\eta_1, \xi_2)$, где $\eta_1 = \frac{\xi_1}{2}$, если $\xi_2 < \frac{1}{2}$; $\eta_1 = \frac{1+\xi_1}{2}$, если $\xi_2 \geqslant \frac{1}{2}$. В этом случае ξ_1 имеет то же распределение, что и η_1; в то же время F_ξ и F_η не совпадают тождественно. Так, $P\left(\eta_1 - \xi_2 > \frac{1}{2}\right) = 0$; $P\left(\xi_1 - \xi_2 > \frac{1}{2}\right) = \frac{1}{8}$.

Рассмотрим примеры многомерных случайных величин.

1. Пусть ξ_1 — равномерно распределенная в интервале (0, 1) случайная величина, $\xi_2 = \xi_1^2$. Тогда $\xi = (\xi_1, \xi_2)$ — двумерная случайная величина. Ее функция распределения

$$F_\xi(x_1, x_2) = P(\xi_1 < x_1, \ \xi_2 < x_2) = P(\xi_1 < x_1, \ \xi_1^2 < x_2).$$

При $x_1 \leqslant 0$ и при $x_2 \leqslant 0$ это выражение равно 0; при $x_1 > 0$, $x_2 > 0$

$$F_\xi(x_1, x_2) = P(\xi_1 < x_1, \ \xi_1 < \sqrt{x_2}) =$$
$$= P(\xi_1 < \min(x_1, \sqrt{x_2})) =$$

$$= \begin{cases} \min(x_1, \sqrt{x_2}), \text{ если } \min(x_1, x_2) \leqslant 1, \\ 1, \text{ если } \min(x_1, x_2) > 1 \end{cases} \tag{5.3}$$

2. Пусть вероятностное пространство состоит из пар $\omega = (\omega_1, \omega_2)$, где ω_1, ω_2 — независимые нормально распределенные случайные величины с математическим ожиданием 0 и дисперсией 1.

Рассмотрим случайные величины

$$\xi_1 = \omega_1 + \omega_2, \quad \xi_2 = \omega_1 - \omega_2. \tag{5.4}$$

Пусть $\xi = (\xi_1, \xi_2)$. Тогда

$$F_\xi(x_1, x_2) = \mathrm{P}(\xi_1 < x_1, \xi_2 < x_2) = \mathrm{P}(\omega_1 + \omega_2 < x_1, \omega_1 - \omega_2 < x_2) = \iint_D p_{\omega_1}(u)\, p_{\omega_2}(v)\, du dv,$$

где область плоскости (u, v) определяется условиями $u + v < x_1$, $u - v < x_2$. Подставив выражение для $p_{\omega_i}(\ldots)$, найдем

$$F_\xi(x_1, x_2) = \frac{1}{2\pi} \iint_D e^{-\frac{u^2}{2} - \frac{v^2}{2}}\, du dv. \tag{5.5}$$

Введем ортогональное преобразование переменных $\alpha = \frac{1}{\sqrt{2}}(u + v)$, $\beta = \frac{1}{\sqrt{2}}(u - v)$. Тогда получим $u^2 + v^2 = \alpha^2 + \beta^2$, а область D перейдет в область $\alpha < \frac{1}{\sqrt{2}} x_1$, $\beta < \frac{1}{\sqrt{2}} x_2$ плоскости (α, β).

Интеграл (5.5) преобразуется к виду

$$F_\xi(x_1, x_2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{1}{\sqrt{2}} x_1} e^{-\frac{\alpha^2}{2}}\, d\alpha \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{1}{\sqrt{2}} x_2} e^{-\frac{\beta^2}{2}}\, d\beta = \Phi_1(x_1)\, \Phi_2(x_2),$$

где $\Phi_i(x)$ — функция распределения нормальной случайной величины с математическим ожиданием 0 и дисперсией 2. Итак, ξ_1 и ξ_2 независимы и обладают плотностями вида

$$\frac{1}{2\sqrt{\pi}} e^{-\frac{x^2}{4}}, \quad -\infty < x < \infty.$$

3. Пусть ω — равномерно распределенная в интервале $(0, 2\pi)$ случайная величина, $\xi_1 = a \cos \omega$, $\xi_2 = a \sin \omega$, $\xi = (\xi_1, \xi_2)$, где a — постоянное число. (ξ_1, ξ_2) всегда лежит на окружности радиуса $|a|$; $F_\xi(x_1, x_2)$ равняется отношению длины части окружности $u^2 + v^2 = a^2$, попадающей внутрь квадранта $u < x_1$, $v < x_2$, к длине всей окружности.

4. Пусть ω_1, ω_2 — независимые случайные величины с плотностями $p_{\omega_1}(x)$, $p_{\omega_2}(x)$. Образуем двумерную случайную величину $\xi = (\xi_1, \xi_2)$, где $\xi_1 = \min(\omega_1, \omega_2)$, $\xi_2 = \max(\omega_1, \omega_2)$. Найдем вид функции $F_\xi(x_1, x_2) = \mathrm{P}(\xi_1 < x_1, \xi_2 < x_2)$

Пусть $x_2 \leqslant x_1$. Тогда

$$\mathrm{P}(\xi_1 < x_1, \xi_2 < x_2) = \mathrm{P}(\min(\omega_1, \omega_2) < x_1, \max(\omega_1, \omega_2) < x_2) = \mathrm{P}(\max(\omega_1, \omega_2) < x_2) = \mathrm{P}(\omega_1 < x_2, \omega_2 < x_2) = F_{\omega_1}(x_2) F_{\omega_2}(x_2).$$

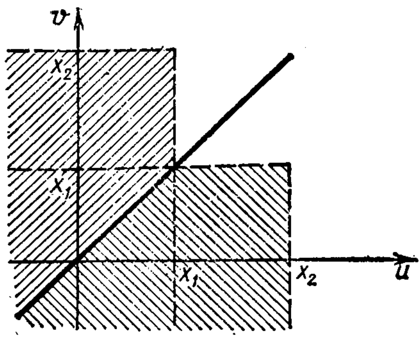

Рис. 5.1

Пусть $x_1 < x_2$. В этом случае справедливо равенство

$$\{\xi_1 < x_1, \xi_2 < x_2\} = \{\omega_1 < \omega_2, \omega_1 < x_1, \omega_2 < x_2\} \cup \{\omega_2 \leqslant \omega_1, \omega_2 < x_1, \omega_1 < x_2\}.$$

Отсюда

$$F_\xi(x_1, x_2) = \iint\limits_{\substack{u < v \\ u < x_1,\ v < x_2}} p_{\omega_1}(u) p_{\omega_2}(v)\, du\, dv + \iint\limits_{\substack{v < u \\ v < x_1,\ u < x_2}} p_{\omega_1}(u) p_{\omega_2}(v)\, du\, dv. \tag{5.6}$$

Области интегрирования, соответствующие первому и последнему слагаемому правой части (5.6), изображены на рис. 5.1 штриховкой. Перейдя от двойного интеграла к повторному, найдем

$$F_{\xi}(x_1, x_2) = \int_{-\infty}^{x_1} p_{\omega_1}(u)\,du \int_{u}^{x_2} p_{\omega_2}(v)\,dv +$$

$$+ \int_{-\infty}^{x_1} p_{\omega_2}(v)\,dv \int_{v}^{x_2} p_{\omega_1}(u)\,du = \int_{-\infty}^{x_1} \left[F_{\omega_2}(x_2) -\right.$$

$$\left. - F_{\omega_2}(u)\right] p_{\omega_1}(u)\,du + \int_{-\infty}^{x_1} p_{\omega_2}(v)\left[F_{\omega_1}(x_2) - F_{\omega_1}(v)\right] dv. \quad (5.7)$$

§ 2. НЕПРЕРЫВНЫЕ МНОГОМЕРНЫЕ РАСПРЕДЕЛЕНИЯ

Многомерное распределение случайной величины ξ называется *непрерывным*, а сама эта случайная величина непрерывной, если при произвольных x_i, $1 \leqslant i \leqslant n$,

$$F_{\xi}(x_1, x_2, \ldots, x_n) =$$

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} p_{\xi}(t_1, t_2, \ldots, t_n)\,dt_1\,dt_2 \ldots dt_n, \quad (5.8)$$

где $p_{\xi}(t_1, t_2, \ldots, t_n)$ — неотрицательная интегрируемая функция. Эта функция называется *плотностью вероятности* (плотностью распределения) случайной величины ξ.

Положив в формуле (5.8) $x_1 = x_2 = \ldots = x_n = \infty$, найдем

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p_{\xi}(t_1, t_2, \ldots, t_n)\,dt_1\,dt_2 \ldots dt_n = 1.$$

Допустим, что в точке $(x_1, x_2, \ldots, x_n)$ существует частная производная $\dfrac{\partial^n F_{\xi}(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}$. Тогда, по известной теореме дифференциального исчисления, имеем

$$\frac{\partial^n F_{\xi}(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n} = p_{\xi}(x_1, \ldots, x_n). \quad (5.9)$$

Пусть, например, ξ_i, $1 \leqslant i \leqslant n$ — непрерывные независимые случайные величины, $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$. Тогда

$$F_\xi(x_1, x_2, \ldots, x_n) = F_{\xi_1}(x_1)F_{\xi_2}(x_2)\ldots F_{\xi_n}(x_n) =$$

$$= \int\limits_{-\infty}^{x_1}\int\limits_{-\infty}^{x_2}\cdots\int\limits_{-\infty}^{x_n} p_{\xi_1}(t_1)\,p_{\xi_2}(t_2)\ldots p_{\xi_n}(t_n)\,dt_1\,dt_2\ldots dt_n$$

В этом случае плотность многомерной случайной величины ξ существует и равна

$$p_\xi(x_1, \ldots, x_n) = p_{\xi_1}(x_1)\,p_{\xi_2}(x_2)\ldots p_{\xi_n}(x_n). \qquad (5.10)$$

Фиксируем некоторое $h>0$ и обозначим через $\mathrm{P}_k(x_1, x_2, \ldots, x_n)$ вероятность события

$$A_k(x_1, \ldots, x_n) = \{x_1 \leqslant \xi_1 < x_1 + h, \ldots, x_k \leqslant \xi_k < x_k + h,\ \xi_{k+1} < x_{k+1}, \ldots, \xi_n < x_n\}, \quad 1 \leqslant k \leqslant n.$$

Тогда, поскольку

$$A_k(x_1, \ldots, x_n) = A_{k-1}(x_1, \ldots, x_{k-1},\ x_k + h,\ x_{k+1}, \ldots, x_n) \setminus A_{k-1}(x_1, \ldots, x_k, \ldots, x_n)^*,$$

то

$$\mathrm{P}_k(x_1, x_2, \ldots, x_n) = \mathrm{P}_{k-1}(x_1, x_2, \ldots, x_{k-1},\ x_k + h,\ x_{k+1}, \ldots, x_n) - \mathrm{P}_{k-1}(x_1, x_2, \ldots, x_k, \ldots, x_n).$$

Последовательно применяя последнюю формулу к выражению (8.11) при $k = 1, 2, \ldots, n$, получим:

$$\mathrm{P}_1(x_1\ x_2, \ldots, x_n) = \int\limits_{-\infty}^{x_1+h}\int\limits_{-\infty}^{x_2}\cdots\int\limits_{-\infty}^{x_1}\ldots dt_1\ldots -$$

$$- \int\limits_{-\infty}^{x_1}\int\limits_{-\infty}^{x_2}\cdots\int\limits_{-\infty}^{x_n}\ldots dt_1\ldots =$$

$$= \int\limits_{x_1}^{x_1+h}\int\limits_{-\infty}^{x_2}\cdots\int\limits_{-\infty}^{x_n} p_\xi(t_1, t_2, \ldots, t_n)\,dt_1\,dt_2\ldots dt_n, \qquad (5.11)$$

---

* Обозначив через B событие $\{x_i \leqslant \xi_i < x_i + h,\ 1 \leqslant i \leqslant k-1;\ \xi_i < x_i,\ k+1 \leqslant i \leqslant n\}$ найдем, что

$$A_{k-1}(x_1, \ldots, x_n) = B \cap \{\xi_k < x_k\},\quad A_{k-1}(x_1, \ldots, x_k + h, \ldots, x_n) = B \cap \{\xi_k < x_k + h\},\quad A_k(x_1, \ldots, x_n) = B \cap \{x_k \leqslant \xi_k < x_k + h\},$$

откуда и следует равенство (5.9).

$$\mathrm{P}_2(x_1, x_2, \ldots, x_n) = \int\limits_{x_1}^{x_1+h} \int\limits_{-\infty}^{x_2+h} \int\limits_{-\infty}^{x_3} \cdots \int\limits_{-\infty}^{x_n} \ldots dt_1 \ldots -$$

$$- \int\limits_{-\infty}^{x_1} \int\limits_{-\infty}^{x_2} \cdots \int\limits_{-\infty}^{x_n} \ldots dt_1 \ldots =$$

$$= \int\limits_{x_1}^{x_1+h} \int\limits_{x_2}^{x_2+h} \int\limits_{-\infty}^{x_3} \cdots \int\limits_{-\infty}^{x_n} p_\xi(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \ldots dt_n, \qquad (5.12)$$

. . . .

$$\mathrm{P}_k(x_1, x_2, \ldots, x_n) =$$

$$= \int\limits_{x_1}^{x_1+h} \cdots \int\limits_{x_k}^{x_k+h} \int\limits_{-\infty}^{x_{k+1}} \cdots \int\limits_{-\infty}^{x_n} p_\xi(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \ldots dt_n, \qquad (5.13)$$

. . . .

$$\mathrm{P}_n(x_1\, x_2, \ldots . x_n) =$$

$$= \int\limits_{x_1}^{x_1+h} \int\limits_{x_2}^{x_2+h} \cdots \int\limits_{x_n}^{x_n+h} p_\xi(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \ldots dt_n. \qquad (5.14)$$

Итак, вероятность попадания ξ в n-мерный гиперкуб

$$\mathrm{P}(x_i < \xi_i < x_i + h), \quad 1 \leqslant i \leqslant n =$$

$$= \int\limits_{x_1}^{x_1+h} \int\limits_{x_2}^{x_2+h} \cdots \int\limits_{x_n}^{x_n+h} p_\xi(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \ldots dt_n. \qquad (5.15)$$

Пусть теперь D — область n-мерного пространства, которую можно представить в виде объединения конечного или счетного числа гиперкубов. Так как интеграл по объединению множеств равен сумме интегралов по этим множествам, то

$$\mathrm{P}(\xi \in D) = \int \underset{D}{\ldots} \int p_\xi(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \ldots dt_n \qquad (5.16)$$

Известно, что любую область n-мерного пространства, ограниченную конечным множеством гладких поверхностей, можно представить в виде объединения конечного или счетного множества ориентированных гиперкубов (рис. 5.2).

Поэтому формула (5.16) справедлива для любой такой области D.

Предположим, что в точке $(x_1, x_2, \dots, x_n)$ функция $p_\xi(x_1, x_2, \dots, x_n)$ непрерывна. Пусть D — область объема $V(D)$, полностью входящая в гиперсферу радиуса r, описанную около точки $(x_1, x_2, \dots, x_n)$. Допустим, что $r \to 0$. Тогда

$$P(\xi \in D) = \int_D \dots \int p_\xi(t_1, \dots t_n)\, dt_1 \dots dt_n =$$

$$= p_\xi(x_1, \dots, x_n) \int_D \dots \int dt_1 \dots dt_n + \int_D \dots \int \left[p_\xi(t_1, \dots, t_n) - \right.$$

$$\left. - p_\xi(x_1, \dots, x_n) \right] dt_1 \dots dt_n. \tag{5.17}$$

Первое слагаемое в правой части формулы (5.17) равно $p_\xi(x_1, \dots, x_n)\, V(D)$, второе представляет собой величину $o(V(D))$ при $r \to 0$. Итак,

$$\frac{P(\xi \in D)}{V(D)} \xrightarrow[r \to 0]{} p_\xi(x_1, \dots, x_n) \tag{5.18}$$

если функция p_ξ непрерывна в точке $(x_1, \dots, x_n)$.

Отметим следующие полезные факты.

I. Если существует интегрируемая функция $f(x_1, x_2, \dots, x_n)$ такая, что для любой области D, содержащейся в сфере достаточно малого радиуса $r(x_1, x_2, \dots, x_n)$, описанной около точки $(x_1, x_2, \dots, x_n)$,

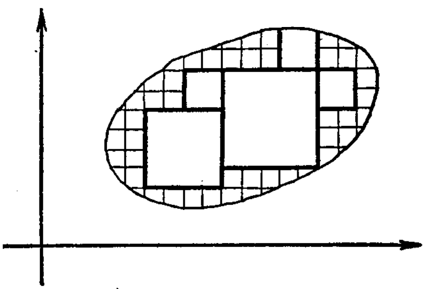

Рис. 5.2

$$P(\xi \in D) \leqslant f(x_1, x_2, \dots, x_n)\, V(D), \tag{5.19}$$

то распределение ξ непрерывно и $p_\xi(x_1, x_2, \dots, x_n) \leqslant \leqslant f(x_1, x_2, \dots, x_n)$.

II. Если при тех же условиях

$$P(\xi \in D) \leqslant \int_D \dots \int f(x_1, \dots, x_n)\, dx_1 \dots dx_n. \tag{5.20}$$

то распределение ξ непрерывно и $p_\xi(x_1, \dots, x_n) \leqslant \leqslant f(x_1, \dots, x_n)$.

Пусть $\xi = (\xi_1, \dots, \xi_n)$ обладает непрерывным распределением и пусть имеются случайные величины $\eta_i = f_i(\xi_1, \dots, \xi_n)$, $1 \leqslant i \leqslant n$, где $f_i(x_1, \dots, x_n)$ — функции, удовлетворяющие таким условиям. n-мерное пространство можно разбить на конечную или бесконечную последовательность областей $\Delta_1, \Delta_2, \dots$, в каждой из которых производные $\frac{\partial f_i}{\partial y_j}$ непрерывны и система уравнений

$$f_i(x_1, \dots, x_n) = y_i, \quad 1 \leqslant i \leqslant n, \tag{5.21}$$

имеет единственное решение

$$x_i = \psi_{ji}(y_1, \dots, y_n),\ 1 \leqslant i \leqslant n,\ (x_1, \dots, x_n) \in \Delta_j,\ j \geqslant 1, \tag{5.22}$$

$$(y_1, \dots, y_n) \in \{(f_1(x_1, \dots, x_n), \dots, f_n(x_1, \dots, x_n)), (x_1, \dots, x_n) \in \Delta_j.$$

Тогда случайная величина $\eta = (\eta_1, \dots, \eta_n)$ непрерывна и ее плотность определяется соотношением

$$\begin{aligned} p_\eta(y_1, \dots, y_n) = \sum_{j \geqslant 1} p_\xi(\psi_{j1}(y_1, \dots, y_n), \dots, \\ \psi_{jn}(y_1, \dots, y_n)) \, | J_j(y_1, \dots, y_n) |, \end{aligned} \tag{5.23}$$

где

$$J_j(y_1, \dots, y_n) = \begin{vmatrix} \frac{\partial \psi_{j1}}{\partial y_1} & \dots & \frac{\partial \psi_{j1}}{\partial y_n} \\ \cdot & \cdot & \cdot \\ \frac{\partial \psi_{jn}}{\partial y_1} & \dots & \frac{\partial \psi_{jn}}{\partial y_n} \end{vmatrix}, \quad j = 1, 2, \dots, \tag{5.24}$$

в точках $(y_1, \dots, y_n)$, для которых существуют функции $\psi_{ji}(y_1, \dots, y_n)$; $J_j(y_1, \dots, y_n) = 0$ в точках, где эти функции не существуют*.

Докажем формулу (5.23). Имеем

$$\mathrm{P}(\eta \in D) = \int\limits_{f(x_1, \dots, x_n) \in D} \cdots \int p_\xi(x_1, \dots, x_n)\, dx_1 \dots dx_n, \tag{5.25}$$

* Имеется в виду, что для данного $(y_1, \dots, y_n)$ в области Δ_i нет точек $(x_1, \dots, x_n)$, для которых $f_i(x_1, \dots, x_n) = y_i$, $1 < i \leqslant n$.

где $f(x_1, \dots, x_n)$ — векторная функция $(f_1(x_1, \dots, x_n), \dots, f_n(x_1, \dots, x_n))$. Можно разбить интеграл (5.25) на слагаемые:

$$P(\eta \in D) = \sum \underset{\substack{f(x_1, \dots, x_n) \in D, \\ (x_1, \dots, x_n) \in \Delta_j}}{\int \dots \int} p_\xi(x_1, \dots, x_n)\, dx_1 \dots dx_n. \tag{5.26}$$

Далее, можно записать

$$\underset{\substack{f(x_1, \dots, x_n) \in D \\ (x_1, \dots, x_n) \in \Delta_j}}{\int \dots \int} p_\xi(x_1, \dots, x_n)\, dx_1 \dots dx_n =$$

$$= \underset{\substack{f(x_1, \dots, x_n) \in D_j \\ (x_1, \dots, x_n) \in \Delta_j}}{\int \dots \int} p_\xi(x_1, \dots, x_n)\, dx_1 \dots dx_n, \tag{5.27}$$

где D_j — подобласть области D, включающая лишь те точки $(y_1, \dots, y_n)$, для которых существует $(x_1, \dots, x_n) \in \Delta_j$ такое, что $f(x_1, \dots, x_n) = (y_1, \dots, y_n)$. Другими словами, если $(y_1, \dots, y_n) \in D_j$, то существует решение (5.22) системы уравнений (5.21). Введем в интеграле (5.27) новые переменные $y_i = f_i(x_1, \dots, x_n)$. Тогда этот интеграл преобразуется к виду

$$\underset{(y_1, \dots, y_n) \in D_j}{\int \dots \int} p_\xi(\psi_{j1}(y_1, \dots, y_n), \dots, \psi_{jn}(y_1, \dots, y_n))\,| \times$$
$$\times |J_j(y_1, \dots, y_n)|\, dy_1 \dots dy_n.$$

Учитывая, что при $(y_1, \dots, y_n) \in D \setminus D_j$ функцию $J_j(y_1, \dots, y_n)$ мы условились определять как нуль, в последнем выражении можно заменить D_j на D. Следовательно,

$$P(\eta \in D) = \sum_j \underset{(y_1, \dots, y_n) \in D}{\int \dots \int} p_\xi(\psi_{j1}(y_1, \dots, y_n), \dots, \psi_{jn}(y_1, \dots, y_n)) \times$$
$$\times |J_j(y_1, \dots, y_n)|\, dy_1 \dots dy_n.$$

Ввиду того, что подынтегральные функции неотрицательны и сумма в правой части последнего равенства не превосходит 1, можно изменить порядок суммирования и интегрирования:

$$P(\eta \in D) = \underset{(y_1, \dots, y_n) \in D}{\int \dots \int} \Big[\sum_j p_\xi(\psi_{j1}(y_1, \dots, y_n), \dots, \psi_{jn}(y_1, \dots$$

$$\ldots, y_n)) \, | J_j \cdot (y_1, \ldots y_n) | \,] \, dy_1 \ldots dy_n$$

Очевидно, это равенство эквивалентно доказываемому равенству (5.23): выражение в квадратных скобках, с одной стороны, совпадает с правой частью (5.23), а с другой — является плотностью в силу равенства (5.16).

Отметим важный частный случай формулы (5.23). Пусть

$$\eta_i = \sum_{k=1}^{n} a_{ik} \xi_k \qquad 1 \leqslant i \leqslant n \tag{5.28}$$

где матрица $\|a_{ij}\|$ предполагается невырожденной. Тогда можно записать

$$\xi_i = \sum_{k=1}^{n} b_{ik} \eta_k \qquad 1 \leqslant i \leqslant n$$

где $\|b_{ik}\|$ — матрица, обратная к матрице $\|a_{ik}\|$. Отсюда

$$p_\eta (y_1, \ldots, y_n) = p_\xi \left(\sum_{k=1}^{n} b_{1k} y_k \ \ldots \ \sum_{k=1}^{n} b_{nk} y_k \right) \times \\ \times |\det \| b_{ik} \| \, | \tag{5.29}$$

где $\det \| b_{ik} \|$ — определитель матрицы $\| b_{ik} \|$.

Удобно записывать эту формулу в векторном виде. Обозначим

$$x = \begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix}; \quad y = \begin{Vmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{Vmatrix}; \quad A = \| a_{ik} \| \quad A^{-1} = \| b_{ik} \|.$$

Будем писать $p_\xi(x)$, $p_\eta(y)$ вместо $p_\xi(x_1, \ldots, x_n)$, $p_\eta(y_1, \ldots, y_n)$. Тогда, поскольку $\det (A^{-1}) = (\det A)^{-1}$, формула (5.29) принимает вид

$$p_\eta (y) = p_{A\xi} (y) = \frac{1}{|\det A|} p_\xi (A^{-1} y) \tag{5.30}$$

Рассмотрим еще более частный случай. Пусть преобразование переменных $\{\xi_i\}$ в переменные $\{\eta_i\}$, задаваемое формулой (5.28), сохраняет элемент объема n-мерного пространства (в частности, это имеет место в случае ортого нального преобразования). Тогда

$$p_\eta(y) = p_\xi(A^{-1} y). \tag{5.31}$$

Если преобразование имеет вид

$$\eta_i = \xi_i + c_i, \quad 1 \leqslant i \leqslant n, \tag{5.32}$$

то

$$p_\eta(x_1, \ldots, x_n) = p_\xi \; (x_1 - c_1, \ldots, \; x_n - c_n). \tag{5.33}$$

На основании формул (5.30) и (5.33) при неоднородном невырожденном преобразовании случайных величин вида $\eta = A\xi + C$,
где

$$\xi = \begin{Vmatrix} \xi_1 \\ \vdots \\ \xi_n \end{Vmatrix}; \quad \eta = \begin{Vmatrix} \eta_1 \\ \vdots \\ \eta_n \end{Vmatrix}; \quad C = \begin{Vmatrix} c_1 \\ \vdots \\ c_n \end{Vmatrix},$$

формулу преобразования плотности можно записать так:

$$p_\eta(y) = \frac{1}{|\det A|}\, p_\xi(A^{-1}(y - C)). \tag{5.34}$$

Отметим следующие полезные преобразования многомерных плотностей.

1. Пусть $a_1, a_2, \ldots, a_n$ — фиксированные положительные числа, $\eta = (\eta_1, \eta_2, \ldots, \eta_n)$ — результат приведения $\xi_1, \xi_2, \ldots, \xi_n$ по модулю $a_1, a_2, \ldots, a_n$. Иными словами,

$$\eta_i = a_i \left\{ \frac{\xi_i}{a_i} \right\},$$

где $\{x\}$ — символ дробной части числа x. Понятно, что если хотя бы при одном i $(1 \leqslant i \leqslant n)$ $y_i \,\overline{\in}\, [0, a_i]$, то $p_\xi(y_1, \ldots, \; y_n) = 0$.

Предположим, что $y_i \in (0, a_i)$, $1 \leqslant i \leqslant n$.
Тогда

$$p_\eta(y_1, \ldots \; y_n) = \sum_{y_1=-\infty}^{\infty} \ldots \sum_{j_n=-\infty}^{\infty} p_\xi(y_1 + j_1 a_1, \ldots \; y_n + j_n a_n) \tag{5.35}$$

Действительно, $\eta_i = y_i$ в том и только том случае, если $\xi_i - y_i$ является кратным a_i, или, что то же, для некоторого j_i $\xi_i = y_i + j_i a_i$.

2. Пусть ξ — многомерная случайная величина $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$, $\xi_{(k)}$ — k-е из $\{\xi_i\}$ по величине, так что

$\xi_{(1)} < \xi_{(2)} < \ldots < \xi_{(n)}$. Обозначим $\eta = (\xi_{(1)}, \xi_{(2)}, \ldots$ $\ldots, \xi_{(n)})$. Тогда справедлива формула

$$p_\eta(x_1, x_2, \ldots, x_n) = \begin{cases} \sum\limits_{j_1, \ldots, j_n} p_\xi(x_{j_1}, x_{j_2}, \ldots x_{j_n}), \text{ если } x_1 < x_2 < \ldots < x_n, \\ 0 \text{ — в противном случае.} \end{cases} \tag{5.36}$$

где $(j_1, j_2, \ldots, j_n)$ — всевозможные перестановки набора индексов $(1, 2, \ldots, n)$.

§ 3. ПРИМЕРЫ РАСПРЕДЕЛЕНИЙ БОЛЕЕ ОБЩЕГО ВИДА

Пусть в n-мерном пространстве имеется некоторое многообразие Γ размерности m, $1 \leqslant m < n$ (поверхность или линия) и пусть в Γ введены координаты $t_1, t_2, \ldots, t_m$, так что Γ можно рассматривать как совокупность $\{t \in D\}$, где $t = (t_1, \ldots, t_n)$; D — некоторая m-мерная область. Пусть $\eta = (\eta_1, \ldots, \eta_m)$ — m-мерная непрерывная случайная величина, с вероятностью 1 принадлежащая области D. Тогда имеет смысл плотность $p_\eta(t_1, t_2, \ldots, t_m)$, и если $A \subset D$, то

$$P(\eta \in A) = \int\limits_A \ldots \int p_\eta(t_1, \ldots, t_m)\, dt_1 \ldots dt_m. \tag{5.37}$$

В исходном n-мерном пространстве координатам $t_1, \ldots, t_m$ соответствуют координаты $x_1, x_2, \ldots, x_n$ в силу параметрических уравнений Γ:

$$x_i = f_i(t_1, \ldots, t_m), \quad 1 \leqslant i \leqslant n, \quad (t_1, \ldots, t_m) \in D. \tag{5.38}$$

В этом случае $\xi = (f_1(\eta_1, \ldots, \eta_m), \ldots, f_n(\eta_1, \ldots, \eta_m))$ является n-мерной случайной величиной. Рассмотрим вероятность попадания ξ в n-мерное множество A:

$$I(\xi \in A) = \int\limits_{\substack{(f_1(t_1 \ldots, t_m), \ldots, f_n(t_1, \ldots, t_m)) \in A, \\ (t_1, \ldots, t_m) \in D}} \ldots \int p_\xi(t_1, t_2, \ldots, t_m)\, dt_1 \ldots dt_m, \tag{5.39}$$

Пусть, например, имеется автоматическая установка, состояние которой характеризуется параметрами ξ_1, ξ_2. Имеется «устройство защиты», которое выключает установку в момент, когда либо $\xi = a$, либо $\xi_2 = b$. (Предполагается, что всегда $\xi_1 \leqslant a$, $\xi_2 \leqslant b$.) Обозначим через ξ вектор (ξ_1, ξ_2) в момент выключения установки. В данном случае Γ — совокупность двух лучей, исходящих из точки $A_0 = (a, b)$ и параллельных координатным осям. Γ можно задать в параметрическом виде:

$$x_1 = x_1(t) = a, \quad x_2 = x_2(t) = b - t, \quad \text{если} \quad t \geqslant 0;$$

$$x_1 = a + t, \quad x_2 = b, \quad \text{если} \quad t < 0.$$

Будем считать, что задана плотность $p_\eta(t)$ значения t, соответствующего вектору ξ. Используем формулу (5.39) для вычисления вероятности события $\{\xi_1 + \xi_2 < x\}$:

$$\mathrm{P}(\xi_1 + \xi_2 < x) = \int\limits_{x_1(t)+x_2(t)<x} p_\eta(t)\,dt =$$

$$= \begin{cases} \int\limits_{-(a+b-x)}^{a+b-x} p_\eta(t)\,dt & \text{при} \quad x \leqslant a + b, \\ 1 & \text{при} \quad x > a + b. \end{cases} \tag{5.40}$$

На практике встречаются случаи, когда в n-мерном пространстве необходимо выделить многообразия разных размерностей и ξ принадлежит каждому из них с положительной вероятностью. Рассмотрим пример, характерный для теории надежности и теории массового обслуживания. Пусть имеется два прибора, выполняющих какие-либо операции. Обозначим через ξ_i время, которое пройдет от заданного момента t_0 до момента окончания выполняемой i-м прибором операции; если в момент t_0 операция не выполняется, положим $\xi_i = 0$. Тогда при достаточно общих условиях $\xi = (\xi_1, \xi_2)$ имеет следующую структуру. С вероятностью p_0 $\xi = (0, 0)$ (p_0 есть вероятность того, что оба прибора не выполняют операций); с вероятностью $p_1(x)\,dx$ $x < \xi_1 < x + dx$, $\xi_2 = 0$ (первый прибор выполняет операцию с остаточным временем $\xi \in (x, x + dx)$, а второй не выполняет операции); с вероятностью $p_2(y)\,dy$ $\xi_1 = 0$, $y < \xi_2 < y + dy$ (случай, симметричный предыдущему); наконец, с вероятностью $p_3(x, y)\,dx dy$ $x < \xi_1 < x + dx$, $y < \xi_2 < y + dy$. Последний случай соответствует ситуации, когда оба прибора выполняют операции. Можно записать:

$$F_\xi(x, y) = p_0 + \int_0^x p_1(t)\,dt + \int_0^y p_2(u)\,du +$$

$$+ \int_0^x \int_0^y p_3(t, u)\,dt\,du, \quad x > 0, \quad y > 0,$$

или

$$F_\xi(x, y) = p_0 \int_{-\infty}^{x} \int_{-\infty}^{y} \delta(t)\,\delta(u)\,dt\,du + \int_{-\infty}^{x} \int_{-\infty}^{y} p_1(t)\,\delta(u)\,dt\,du +$$

$$+ \int_{-\infty}^{x} \int_{-\infty}^{y} \delta(t)\,p_2(u)\,dt\,du + \int_{-\infty}^{x} \int_{-\infty}^{y} p_3(t, u)\,dt\,du. \qquad (5.41)$$

где положено $p_1(t) = 0$, $t < 0$, $p_2(u) = 0$, $u < 0$, $p_3(t, u) = 0$, если $t < 0$ либо $u < 0$.

Унифицированным аппаратом, позволяющим анализировать характеристики многомерных случайных величин в самом общем случае, является аппарат интеграла Стильтьеса, который будет рассмотрен в следующем параграфе.

§ 4. ИНТЕГРАЛ СТИЛЬТЬЕСА

Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ — n-мерная случайная величина, D — некоторая ограниченная область n-мерного пространства, для которой определена вероятность $\mathrm{P}(\xi \in D)$, $f(x) = f(x_1, \ldots, x_n)$ — вещественная функция, ограниченная при $x \in D$. Разобьем D на N областей Δ_{N1}, $\Delta_{N2}, \ldots, \Delta_{NN}$ и в каждой из этих областей возьмем точку $x^{(N, j)} \in \Delta_{Nj}$, $1 \leqslant j \leqslant N$. Образуем сумму

$$S_N = \sum_{j=1}^{N} f\left(x^{(N, j)}\right) \mathrm{P}\left(\xi \in \Delta_{Nj}\right) \qquad (5.42)$$

Для областей Δ_{Nj} должны быть определены вероятности $\mathrm{P}\left(\xi \in \Delta_{Nj}\right)$. Эта сумма называется *интегральной суммой Стильтьеса*. Сделаем следующее важное замечание. Определим функцию $\overline{f}(x)$ равенством $\overline{f}(x) = f\left(x^{(N, j)}\right)$ при $x \in \Delta_j$,

$1 \leqslant j \leqslant N$; $\bar{f}(x) = 0$ при $x \overline{\in} D$. Тогда

$$S_N = M\bar{f}(\xi). \tag{5.43}$$

Действительно, правая часть равенства (5.42) есть сумма произведений возможных значений случайной величины $\bar{f}(\xi)$ на их вероятности.

Пусть $\rho(x, y)$ — расстояние между точками x, y n-мерного пространства:

$$\rho(x, y) = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2},$$

$$\text{если } x = (x_1, \ldots, x_n),\ y = (y_1, \ldots, y_n).$$

Для произвольного множества A обозначим

$$\alpha(A) = \sup_{x,\ y \in A} \rho(x, y) \tag{5.44}$$

и назовем $\alpha(A)$ *диаметром* множества A. Обозначим далее

$$\lambda_N = \max_{1 \leqslant j \leqslant N} d(\Delta_{Nj}).$$

Определение. *Если существует предел (конечный или бесконечный, но определенного знака)*

$$\lim_{\lambda_N \to 0} S_n = S \tag{5.45}$$

и его значение не зависит ни от последовательности $\{\Delta_{Nj}\}$, *ни от выбора точек* $x^{(N, j)} \in \Delta_{Nj}$, *то этот предел называется интегралом Стильтьеса функции* $f(x)$ *по области* D *(при заданном распределении случайной величины* ξ*) и обозначается так:*

$$S \doteq \int_D \ldots \int f(x_1, x_2, \ldots, x_n)\, dF_\xi(x_1, x_2, \ldots x_n) \tag{5.46}$$

или сокращенно

$$\int_D f(x)\, dF_\xi(x).$$

Сделаем замечание об одномерном интеграле Стильтьеса: хотя это и частный случай общего определения, но он используется наиболее часто, поэтому полезно дать и самостоятельное определение.

Пусть $f(x)$ — некоторая функция, $F(x)$ — функция распределения. Рассмотрим отрезок $[a, b)$ и разобьем его на отрезки $\Delta_1, \Delta_2, \ldots, \Delta_N$ точками $x_1 < x_2 < \ldots < x_{N-1}$. Обозначив $a = x_0$, $b = x_N$, имеем

$$\Delta_1 = [x_0, x_1), \quad \Delta_2 = [x_1, x_2), \ldots, \Delta_N = \left[x_{N-1}, x_N\right).$$

Возьмем произвольные точки $y_i \in \Delta_i$, $1 \leqslant i \leqslant N$ и образуем сумму

$$S_N = \sum_{i=1}^{N} f(y_i)\,[F(x_i) - F(x_{i-1})].$$

Обозначим

$$\lambda_N = \max_{1<i<N} \{x_i - x_{i-1}\}.$$

Если существует предел

$$S = \lim_{\substack{N \to \infty \\ \lambda_N \to 0}} S_N,$$

не зависящий от $\{x_i\}$ и $\{y_i\}$ при условии, что $\lambda_N \to 0$, $N \to \infty$, то этот предел называется *интегралом Стильтьеса функции* $f(x)$ *по функции распределения* $F(x)$ *и обозначается* $\int\limits_a^b f(x)\,dF(x)$.

Интеграл Стильтьеса по интервалу $(-\infty, \infty)$ определяется как предел предыдущего интеграла при $a \to -\infty$, $b \to \infty$ (если такой предел существует).

Для краткости обозначают

$$\int f(x)\,dF(x) = \int_{-\infty}^{\infty} f(x)\,dF(x).$$

Для неправильно-дискретного распределения

$$F(x) = \int_{-\infty}^{x} p(t)\,dt + \sum_i p_i E(x - x_i)$$

имеем

$$\int f(x)\,dF(x) = \int_{-\infty}^{\infty} f(x)\,p(x)\,dx + \sum_i p_i f(x_i).$$

Теорема 5.1. *Если функция* $f(x)$ *ограничена в ограниченной области* D, *для которой определена* $\mathrm{P}(\xi \in D)$, *и для любого* $\varepsilon > 0$ *область* D *можно разбить на две подобласти*

D_1 и D_0 такие, что в D_0 функция $f(x)$ равномерно непрерывна и $\mathrm{P}(\xi \in D_1) \leqslant \varepsilon$, *то интеграл* $\int\limits_D f(x)\,dF_\xi(x)$ *существует и конечен.*

Доказательство. Пусть $\left\{\Delta_{N_j}\right\}$ $(1 \leqslant j \leqslant N,$ $1 \leqslant N < \infty)$ — некоторая последовательность разбиений области D, для которой $\lambda_N = \max\limits_{1 \leqslant j \leqslant N} d\left(\Delta_{N_j}\right) \underset{N \to \infty}{\longrightarrow} 0$. Возьмем некоторые N и M и рассмотрим интегральные суммы Стильтьеса:

$$S_N = \sum_{j=1}^{N} f\left(x^{(N,\,j)}\right) \mathrm{P}\left(\xi \in \Delta_{Nj}\right),$$

$$S_M = \sum_{k=1}^{M} f\left(x^{(M,\,k)}\right) \mathrm{P}\left(\xi \in \Delta_{Mk}\right). \tag{5.47}$$

По формуле полной вероятности имеем:

$$\mathrm{P}\left(\xi \in \Delta_{Nj}\right) = \sum_{k=1}^{M} \mathrm{P}\left(\xi \in \Delta_{Nj}\Delta_{Mk}\right), \quad 1 \leqslant j \leqslant N,$$

$$\mathrm{P}\left(\xi \in \Delta_{Mk}\right) = \sum_{j=1}^{N} \mathrm{P}\left(\xi \in \Delta_{Nj}\Delta_{Mk}\right), \quad 1 \leqslant k \leqslant M;$$

поэтому

$$S_N = \sum_{j=1}^{N} \sum_{k=1}^{M} f\left(x^{(N,\,j)}\right) \mathrm{P}\left(\xi \in \Delta_{Nj}\Delta_{Mk}\right), \tag{5.48}$$

$$S_M = \sum_{j=1}^{N} \sum_{k=1}^{M} f\left(x^{(M,\,k)}\right) \mathrm{P}\left(\xi \in \Delta_{Nj}\Delta_{Mk}\right). \tag{5.49}$$

Вычтя (5.49) из (5.48), найдем

$$S_N - S_M = \sum_{j=1}^{N} \sum_{k=1}^{M} \left[f\left(x^{(N,\,j)}\right) - f\left(x^{(M,\,k)}\right)\right] \times \\ \times \mathrm{P}\left(\xi \in \Delta_{Nj}\Delta_{Mk}\right). \tag{5.50}$$

По заданному $\delta > 0$ построим область D_2 как множество точек, каждая из которых отстоит от некоторой точки области D_1 не более чем на δ. Таким образом, если область Δ содержит хотя бы одну точку, принадлежащую D_1 и $d(\Delta) \leqslant \delta$, то $\Delta \subset D_2$.

Введем множество R_1 пар индексов (j, k), для которых хотя бы одна точка множества $\Delta_j \cup \Delta_k$ принадлежит D_1, и обозначим через R_0 множество пар индексов (j, k), $1 \leqslant j \leqslant N$, $1 \leqslant k \leqslant M$, не входящих в R_1. Тогда на основании формулы (5.50) получим

$$|S_N - S_M| \leqslant \Big| \sum_{(j,k) \in R_0} [f(x^{(N,j)}) - f(x^{(M,k)}) \times$$
$$\times P(\xi \in \Delta_{Nj} \Delta_{Mk}) \Big| + \Big| \sum_{(j,k) \in R_1} [f(x^{(N,j)}) - f(x^{(M,k)})] \times$$
$$\times P(\xi \in \Delta_{Nj} \Delta_{Mk}) \Big|. \tag{5.51}$$

Выберем настолько большие N, M, чтобы $d(\Delta_{Nj}) \leqslant \delta$, $d(\Delta_{Mk}) \leqslant \delta$ для всех j и k. Тогда вторая сумма оценивается так:

$$\Big| \sum_{(j,k) \in R_1} [f(x^{(N,j)}) - f(x^{(M,k)})] P(\xi \in \Delta_{N} \Delta_{Mk}) \Big| \leqslant$$
$$\leqslant 2\max |f(x)| \sum_{(j,k) \in R_1} P(\xi \in \Delta_{Nj} \Delta_{Mk}) \leqslant$$
$$\leqslant 2 \max_{x \in D} |f(x)| P(\xi \in D_2) \tag{5.52}$$

Эта оценка следует из того, что если $\{\xi \in \Delta_{Nj} \Delta_{Mk}\} \times \times (j, k) \in R_1$, то хотя бы одна точка множества $\Delta_{Nj} \Delta_{Mk}$ принадлежит D_1, так что $\Delta_{Nj} \Delta_{Mk} \subset D_2$. Поскольку события $\Delta_{Nj} \Delta_{Mk}$ попарно несовместны, то сумма их вероятностей равна вероятности их объединения. Первую сумму в правой части (5.51) можно оценить так. Возьмем такое $\delta > 0$, чтобы из неравенства $\rho(x, y) \leqslant 2\delta$ следовало неравенство $|f(x) - f(y)| \leqslant \varepsilon$ для любых $x \in D_0$, $y \in D_0$. Это можно сделать, так как в области D_0 функция $f(x)$ равномерно непрерывна. Если $P(\xi \in \Delta_{Nj} \Delta_{Mk}) > 0$, то множества Δ_{Nj} и Δ_{Mk} имеют хотя бы одну общую точку z. Тогда

$$\rho(x^{(N,j)}, x^{(M,k)}) \leqslant \rho(x^{(N,j)}, z) + \rho(x^{(M,k)}, z) \leqslant$$

$$\leqslant d(\Delta_{Nj}) + d(\Delta_{Mk}) < 2\delta.$$

Отсюда

$$|f(x^{(N,\, j)}) - f(x^{(M,\, k)})| \leqslant \varepsilon,$$

и рассматриваемая сумма оценивается так:

$$\left| \sum_{(j,\, k) \in R_0} [f(x^{(N,\, j)}) - f(x^{(M,\, k)})] \, \mathrm{P}(\xi \in \Delta_{Nj} \Delta_{Mk}) \right| \leqslant$$

$$\leqslant \varepsilon \sum_{(j,\, k) \in R_0} \mathrm{P}(\xi \in \Delta_{Nj} \Delta_{Mk}) \leqslant \varepsilon. \qquad (5.53)$$

Итак, вследствие неравенств (5.52) и (5.33) имеем

$$|S_N - S_M| \leqslant 2 \max_{x \in D} |f(x)| \, \mathrm{P}(\xi \in D_2) + \varepsilon. \qquad (5.54)$$

По условию теоремы, $\mathrm{P}(\xi \in D_1) \leqslant \varepsilon$. Далее, можно записать: $D_1 = \lim\limits_{\delta \to 0} D_2$. По аксиоме непрерывности,

$$\mathrm{P}(\xi \in D_1) = \lim_{\delta \to 0} \mathrm{P}(\xi \in D_2).$$

Значит, при достаточно малом $\delta > 0$ $\mathrm{P}(\xi \in D_2) < 2\varepsilon$. Окончательно имеем

$$|S_N - S_M| \leqslant \varepsilon \left[4 \max_{x \in D} |f(x)| + 1\right], \quad M, N \geqslant N_0. \qquad (5.55)$$

Так как $\varepsilon > 0$ произвольно, то последовательность $\{S_N\}_{N \geqslant 1}$ является сходящейся, т. е. существует предел

$$S = \lim_{N \to \infty} S_N.$$

Этот предел не зависит от последовательности $\{\Delta_{Nj}\}$ и выбора точек $x^{(N,\, j)}$ при условии, что $\lambda_N \xrightarrow[N \to \infty]{} 0$. Действительно, пусть имеется две последовательности $\{\Delta_{Nj}\}$ и $\{\Delta'_{Nj}\}$, удовлетворяющие этому условию. Обозначим соответствующие интегральные суммы Стильтьеса через S_N и S'_N. Пусть $\Delta''_{Nj} = \Delta_{Nj}$ при четных N, $\Delta''_{Nj} = \Delta_{Nj}$ при нечетных N. Обозначим интегральную сумму, соответствующую Δ''_{Nj}, через S''_N. Тогда на основании доказанного существует $\lim\limits_{N \to \infty} S''$. Отсюда

$$\lim_{N \to \infty} S_N = \lim_{N \to \infty} S_{2N} = \lim_{N \to \infty} S''_{2N} = \lim_{N \to \infty} S''_N,$$

$$\lim_{N\to\infty} S'_N = \lim_{N\to\infty} S'_{2N+1} = \lim_{N\to\infty} S''_{2N+1} = \lim_{N\to\infty} S''_N .$$

Оба предела совпадают. Теорема доказана.

Пусть $f(x)$ — любая вещественная функция, определенная и конечная в области D, для которой имеет смысл выражение $\mathrm{P}(\xi \in D)$.

Обозначим

$$f^+(x) = \begin{cases} f(x) & \text{при} \quad f(x) > 0, \\ 0 & \text{при} \quad f(x) \leqslant 0; \end{cases}$$

$$f^-(x) = \begin{cases} 0 & \text{при} \quad f(x) \geqslant 0, \\ f(x) & \text{при} \quad f(x) < 0 \end{cases}$$

Если для любой последовательности областей $D_1 \subset \subset D_2 \subset \ldots \subset D_k \subset \ldots$, удовлетворяющей следующим условиям:

1) $\lim\limits_{k\to\infty} D_k = D$; 2) для любого k область D_k ограничена;

3) функция $f(x)$ ограничена в D_k и существуют

$$\lim_{k\to\infty} \int_D f^+(x)\, dF_\xi(x) = S^+, \tag{5.56}$$

и

$$\lim_{k\to\infty} \int_{D_k} f^-(x)\, dF_\xi(x) = S^-, \tag{5.57}$$

причем хотя бы один из этих пределов конечен, то $S = S^+ + S^-$ называется *интегралом Стильтьеса функции* $f(x)$ *по области* D и обозначается

$$S = \int_D f(x)\, dF_\xi(x).$$

В том случае, если D есть все n-мерное пространство, обозначают

$$\int_D f(x)\, dF_\xi(x) = \int f(x)\, dF_\xi(x).$$

Теорема 5.2. *Пусть* $f(\xi)$ — *одномерная случайная величина. Справедливо равенство*

$$\int f(x)\, dF_\xi(x) = Mf(\xi) \tag{5.58}$$

в том смысле, что если первое из этих выражений имеет смысл, то второе также имеет смысл и оба они совпадают.

Докажем данное утверждение в частном случае, когда $f(x)$ непрерывна и равна нулю вне ограниченной области D_1. Рассмотрим последовательность ограниченных областей $D_1 \subset D_2 \subset \ldots \subset D_k \subset \ldots$ и фиксируем какую-нибудь область D_k. Поскольку D_k ограничена, а $f(x)$ непрерывна во всем n-мерном пространстве, то по заданному $\varepsilon > 0$ можно найти такое $\delta > 0$, что $|f(x) - f(y)| \leqslant \varepsilon$ при $\rho(x, y) \leqslant \delta$, $x, y \in D_k$. Разобьем область D_k на области $\Delta_1, \Delta_2, \ldots, \Delta_N$ так, чтобы $d(\Delta_j) \leqslant \delta$, $1 \leqslant j \leqslant N$, и возьмем $x^{(j)} \in \Delta_j$, $1 \leqslant j \leqslant N$. Образуем сумму

$$S_N = \sum_{j=1}^{N} f(x^{(j)})\,\mathrm{P}(\xi \in \Delta_j).$$

Тогда

$$S_N = M\bar{f}(\xi), \tag{5.59}$$

где $\bar{f}(x) = f(x^{(j)})$ при $x \in \Delta_j$; $\bar{f}(x) = 0$ при $x \bar{\in} D_k$. С другой стороны, $|f(x) - \bar{f}(x)| \leqslant \varepsilon$ для всех x, так что

$$|Mf(\xi) - M\bar{f}(\xi)| \leqslant \varepsilon \tag{5.60}$$

[существование $Mf(\xi)$ следует из ограниченности функции $f(x)$].

Отсюда

$$|S_N - Mf(\xi)| \leqslant \varepsilon. \tag{5.61}$$

Однако при достаточно малом $\delta > 0$ сумма S_N сколь угодно близка к $\int\limits_{D_k} f(x)\,dF_\xi(x)$. Перейдя в неравенстве (5.61) к пределу при $\delta \to 0$, получим (в силу произвольности $\varepsilon > 0$)

$$\int\limits_{D_k} f(x)\,dF_\xi(x) = Mf(\xi).$$

Если теперь предположить, что при $k \to \infty$ области D_k покрывают все пространство, то придем к выводу, что $\int f(x)\,dF_\xi(x) = Mf(\xi)$.

В общем случае доказательство утверждения теоремы может быть существенно с помощью предельного перехода; однако мы эти выкладки опускаем.

Допустим, что ξ — непрерывная случайная величина.

Тогда для любой области D

$$\int\limits_D f(x)\,dF_\xi(x) = \int\underset{D}{\ldots}\int f(x_1, x_2, \ldots, x_n) \times$$

$$\times p_\xi(x_1, x_2, \ldots, x_n)\,dx_1 dx_2 \ldots dx_n \qquad (5.62)$$

если оба выражения имеют смысл.

Эта формула доказывается на основании следующей леммы, известной из теории интегрирования.

Лемма. Пусть $p(x_1, x_2, \ldots, x_n) \geqslant 0$ в некоторой области A; $f(x_1, \ldots, x_n)$ — произвольная функция, причем интегралы

$$\int\underset{A}{\ldots}\int p(x_1, \ldots, x_n)\,dx_1 \ldots dx_n, \quad \int\underset{A}{\ldots}\int f(x_1, \ldots, x_n)\,p(x_1, \ldots$$

$$\ldots, x_n)\,dx_1 \ldots dx_n$$

существуют и конечны. Тогда найдутся такие точки $(a_1, a_2, \ldots, a_n) \in A$, $(b_1, \ldots, b_n) \in A$, что

$$f(a_1, a_2, \ldots, a_n)\int\underset{A}{\ldots}\int p(x_1, x_2, \ldots, x_n)\,dx_1 dx_2 \ldots dx_n \leqslant$$

$$\leqslant \int\underset{A}{\ldots}\int f(x_1, x_2, \ldots, x_n)\,p(x_1, x_2, \ldots x_n)\,dx_1 dx_2 \ldots dx_n \leqslant$$

$$\leqslant f(b_1, b_2, \ldots, b_n)\int\underset{A}{\ldots}\int p(x_1, x_2, \ldots, x_n)\,dx_1 dx_2 \ldots dx_n. \qquad (5.63)$$

(Если $n = 1$, то $f(x)$ — непрерывная функция, $A = [c, d]$, и утверждение леммы следует из теоремы о среднем, известной из интегрального исчисления:

$$\int\limits_c^d f(x)\,p(x)\,dx = f(v)\int\limits_c^d p(x)\,dx,$$

где v — некоторая точка отрезка $[c, d]$).

Пусть область D ограничена и функция $f(x_1, x_2, \ldots, x_n)$ ограничена в ней.

Обозначим

$$J = \int\underset{D}{\ldots}\int f(x_1, \ldots, x_n)\,p_\xi(x_1, \ldots, x_n)\,dx_1 \ldots dx_n \qquad (5.64)$$

и разобьем область D на области $\Delta_1, \Delta_2, \ldots, \Delta_n$ с диаметром, меньшим $\delta > 0$. Тогда

$$I = \sum_{j=1}^{N} \int_{\Delta_j} \dots \int f(x_1, \dots, x_n)\, p_\xi(x_1, \dots, x_n)\, dx_1 \dots dx_n. \quad (5.65)$$

В дальнейшем для краткости будем писать $\int$, x, dx вместо $\int \dots \int$, $(x_1, \dots, x_n)$, $dx_1 dx_2 \dots dx_n$. В соответствии с приведенной леммой выберем такие $a_j \in \Delta_j$, $b_j \in \Delta_j$, $1 \leqslant j \leqslant N$, что

$$f(a_j) \int_{\Delta_j} p_\xi(x)\, dx \leqslant \int_{\Delta_j} f(x)\, p_\xi(x)\, dx \leqslant f(b_j) \int_{\Delta_j} p_\xi(x)\, dx. \quad (5.66)$$

Далее,

$$\int_{\Delta_j} p_\xi(x)\, dx = \mathrm{P}(\xi \in \Delta_j). \quad (5.67)$$

Таким образом,

$$\sum_{j=1}^{N} f(a_j)\, p(\xi \in \Delta_j) \leqslant J \leqslant \sum_{j=1}^{N} f(b_j)\, p(\xi \in \Delta_j) \quad (5.68)$$

Обе суммы в двойном неравенстве (5.68)представляют собой интегральные суммы Стильтьеса. При $\delta \to 0$ эти выражения сходятся к $\int_D f(x)\, dF_\xi(x)$. Справедливость формулы в данном случае установлена.

Для неограниченной области D и, возможно, неограниченной функции $f(x)$ как $\int f(x)\, p_\xi(x)\, dx$, так и $\int f(x)\, dF_\xi(x)$ определяются как пределы соответствующих выражений для ограниченных областей, в которых $f(x)$ ограничены. Поскольку допредельные выражения совпадают, то и пределы будут совпадать. Справедливость формулы (5.62) доказана в общем случае.

Пусть ξ дискретна, т. е. принимает значения x_i, $i \in I$, где I — конечное или счетное множество, с вероятностями p_i, $\sum_{i \in I} p_i = 1$. Тогда

$$\int f(x)\, dF(x) = \sum_{i \in I} f(x_i)\, p_i, \quad (5.69)$$

если первое выражение имеет смысл. Для доказательства формулы (5.69) достаточно вспомнить, что интеграл $\int f(x)\,dF(x)$ есть математическое ожидание $f(\xi)$.

Отметим основные свойства интеграла Стильтьеса.

1. Линейность:

$$\int_D [c_1 f_1(x) + c_2 f_2(x)]\, dF_\xi(x) = c_1 \int_D f_1(x)\, dF_\xi(x) + \\ + c_2 \int_D f_2(x)\, dF_\xi(x). \tag{5.70}$$

2. Аддитивность:

если $D = D_1 \cup D_2$, $D_1 \cap D_2 = \varnothing$, то

$$\int_D f(x)\, dF_\xi(x) = \int_{D_1} f(x)\, dF_\xi(x) + \int_{D_2} f(x)\, dF_\xi(x) \tag{5.71}$$

(предполагается, что все эти интегралы существуют).

3. Если $F(x_1, \ldots, x_n) = F_1(x_1, \ldots, x_m) F_2(x_{m+1}, \ldots, x_n)$, то

$$\int_D f(x)\, dF(x) = \int_{D_1} dF_1(x_1, \ldots, x_n) \int_{D_2(x_1, \ldots, x_m)} f(x)\, dF_2(x_{m+1}, \ldots, x_n), \tag{5.72}$$

где D_1 — множество точек $(x_1, \ldots, x_m)$, для каждой из которых существует хотя бы одна точка $(x_1, x_2, \ldots, x_n) \in D$; $D_2(x_1, x_2, \ldots, x_m)$ — множество тех $(x_{m+1}, \ldots, x_n)$, для которых $(x_1, \ldots, x_n) \in D$, если область D ограничена, а функция $f(x)$ равномерно непрерывна в ней.

Доказательство последнего утверждения аналогично доказательству возможности замены двойного интеграла повторным в случае ограниченной области интегрирования и равномерно непрерывной подынтегральной функции.

§ 5. ХАРАКТЕРИСТИКИ МНОГОМЕРНЫХ РАСПРЕДЕЛЕНИЙ

Равенство (5.58) позволяет выражать различные характеристики распределений многомерных случайных величин через интеграл Стильтьеса. Пусть $I_A(x)$ — характеристическая функция множества, т. е.

$$I_A(x) = \begin{cases} 1 & \text{если } x \in A, \\ 0, & \text{если } x \notin A. \end{cases} \tag{5.73}$$

Тогда $I_A(\xi)$ — так называемый *индикатор события* $\{\xi \in A\}$. $I_A(\xi) = 1$ в том и только том случае, если $\xi \in A$; в противном случае $I_A(\xi) = 0$. Имеем

$$\int I_A(x)\, dF(x) = \int_{A} \dots \int dF_\xi(x_1, x_2, \dots, x_n) =$$

$$= MI_A(\xi) = \mathrm{P}(I_A(\xi) = 1) = \mathrm{P}(\xi \in A). \qquad (5.74)$$

Положим, в частности, $A = \{x_i < z_i,\ 1 \leqslant i \leqslant n\}$. Тогда

$$\int I_A(x)\, dF(x) = \int_{\substack{x_i < z_i \\ 1 \leqslant i \leqslant n}} \dots \int dF_\xi(x_1, \dots, x_n) =$$

$$= \mathrm{P}(\xi_i < z_i, \quad 1 \leqslant i \leqslant n) = F_\xi(x_1, \dots, x_n). \qquad (5.75)$$

Подобным же образом,

$$\int_{\substack{a_i < x_i < b_i \\ 1 \leqslant i \leqslant n}} \dots \int dF_\xi(x_1, \dots, x_n) = \mathrm{P}(a_i \leqslant \xi_i < b_i, \quad 1 \leqslant i \leqslant n). \quad (5.76)$$

Положим $f(x_1, \dots, x_n) = x_i$, где i — любой фиксированный индекс, $1 \leqslant i \leqslant n$. Тогда

$$\int \dots \int x_i dF_\xi(x_1, \dots, x_n) = \mathrm{M}\xi_i, \quad 1 \leqslant i \leqslant n. \qquad (5.77)$$

Если ξ, η — одномерные случайные величины, обладающие конечными дисперсиями, то $\mathrm{M}(\xi - \mathrm{M}\xi)(\eta - \mathrm{M}\eta)$ называется *ковариацией случайных величин* ξ и η и обозначается $\operatorname{cov}(\xi, \eta)$. На основании формулы (5.58) при $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant n$ имеем

$$\operatorname{cov}(\xi_i, \xi_j) = \int \dots \int (x_i - m_i)(x_j - m_j)\, dF_\xi(x_1, \dots, x_n), \quad (5.78)$$

где $m_i = \mathrm{M}\xi_i$, $1 \leqslant i \leqslant n$.

Пусть $i = j$, тогда

$$\operatorname{cov}(\xi_i, \xi_i) = \mathrm{M}(\xi_i - m_i)^2 = \mathrm{D}\xi =$$

$$= \int \dots \int (x_i - m_i)^2\, dF_\xi(x_1, \dots, x_n). \qquad (5.79)$$

Матрица $\|\operatorname{cov}(\xi_i, \xi_j)\|$ порядка $n \times n$ называется *ковариационной матрицей* n-мерной случайной величины $(\xi_1, \dots, \xi_n)$. Заметим, что если к ξ_i прибавить любые постоянные числа a_i, то ковариационная матрица останется неизменной.

Пусть $k_1, k_2, \ldots, k_n$ — любые целые неотрицательные числа. Моментом многомерной случайной величины $\xi = (\xi_1, \ldots, \xi_n)$ порядка $k = (k_1, \ldots, k_n)$ называется величина $M\xi_1^{k_1}\xi_2^{k_2}\ldots\xi_n^{k_n}$, если этому математическому ожиданию можно придать смысл. Имеем

$$M\xi_1^{k_1}\xi_2^{k_2}\ldots\xi_n^{k_n} = \int\ldots\int x_1^{k_1}x_2^{k_2}\ldots x_n^{k_n}dF_\xi(x_1, x_2, \ldots, x_n). \tag{5.80}$$

В формулах (5.74)—(5.80) предполагается, что соответствующий интеграл Стильтьеса существует.

Пусть $\xi = (\xi_1, \ldots, \xi_n)$ рассматривается как вектор n-мерного пространства. Тогда вектор $a = (a_1, \ldots, a_n)$, где $a_i = M\xi_i$, $1 \leqslant i \leqslant n$, называется *вектором математических ожиданий* (или просто *математическим ожиданием*) ξ:

$$\xi = \xi_1 e_1 + \ldots + \xi_n e_n, \tag{5.81}$$

где e_i — единичный вектор i-й координатной оси n-мерного пространства, $1 \leqslant i \leqslant n$. Можно записать

$$M\xi = \sum_{i=1}^{n} M\xi_i e_i = \int \sum_{i=1}^{n} x_i e_i dF_\xi(x) = \sum_{i=1}^{n} e_i \int x_i dF_\xi(x). \tag{5.82}$$

Пусть, например, $n = 2$, $e_1 = 1$, $e_2 = i$, т. е. рассматривается случайная величина $\xi = \xi_1 + i\xi_2$, значения которой принадлежат пространству комплексных чисел, которое можно рассматривать как двумерное пространство (плоскость). Тогда

$$M\xi = \int (x_1 + ix_2)\, dF_\xi(x_1, x_2) =$$
$$= \int x_1 dF_\xi(x_1, x_2) + i\int x_2 dF_\xi(x_1, x_2) = \int x_1\, dF_{\xi_1}(x_1) +$$
$$+ i\int x_2 dF_{\xi_2}(x_2). \tag{5.83}$$

§ 6. УСЛОВНЫЕ РАСПРЕДЕЛЕНИЯ И УСЛОВНЫЕ МАТЕМАТИЧЕСКИЕ ОЖИДАНИЯ

Любая случайная величина ξ может рассматриваться как вещественная функция элементарного события ω, которое является элементом некоторого пространства Ω с определенной на σ-алгебре $\mathfrak{A}$ его подмножеств вероятностью $\mathrm{P}(A)$. Таким образом, $\xi = f(\omega)$, где $f(\omega)$ — функция, обладающая тем свойством, что для любого x $\{\omega : f(\omega) < x\} \in \mathfrak{A}$, т. е. всегда определена вероятность события $\{\xi < x\}$. Пусть теперь имеется два вероятностных пространства $(\Omega_1, \mathfrak{A}_1, \mathrm{P}_1)$ и $(\Omega_2, \mathfrak{A}_2, \mathrm{P}_2)$, причем ω_1 и ω_2 можно рассматривать как независимые элементарные события. Более точно, для любых $A_1 \in \mathfrak{A}_1$, $A_2 \in \mathfrak{A}_2$ события $\{\omega_1 \in A_1\}$ и $\{\omega_2 \in A_2\}$ независимы:

$$\mathrm{P}(\omega_1 \in A_1, \ \omega_2 \in A_2) = \mathrm{P}(\omega_1 \in A_1)\,\mathrm{P}(\omega_2 \in A_2).$$

Предположим, что случайная величина $\xi = (\xi_1, \ \xi_2)$ имеет следующую структуру: $\xi_1 = f_1(\omega_1)$, $\xi_2 = f_2(\xi_1, \ \omega_2)$. Таким образом, ξ_1 — функция ω_1, а ξ_2 — функция ω_2 и ξ_1. Следовательно, заданы две функции: $f_1(\omega_1)$ и $f_2(x_1, \ \omega_2)$. Допустим, что при фиксированном x_1 функция $f_2(x_1, \ \omega_2)$ может рассматриваться как случайная величина. Обозначим ее функцию распределения:

$$F(y/x) = \mathrm{P}(f_2(x, \ \omega_2) < y). \tag{5.84}$$

Эта функция, как функция y $(-\infty < y < \infty)$, называется *условной функцией распределения случайной величины* ξ_2 *при условии* $\xi_1 = x$.

Если существует математическое ожидание $f(x, \ \omega_2)$:

$$M(\xi_2/x) = Mf(x, \ \omega_2), \tag{5.85}$$

то оно называется *условным математическим ожиданием случайной величины* ξ_2 *при условии* $\xi_1 = x$.

Аналогично, если задано любое конечное множество вероятностных пространств $(\Omega_i, \mathfrak{A}_i, \ \mathrm{P}_i)$, $i \geqslant 1$, так, что $\omega_i \in \Omega_i$ рассматриваются, как независимые:

$$\mathrm{P}(\omega_i \in A_i, \ 1 \leqslant i \leqslant n) = \prod_{i=1}^{n} \mathrm{P}(\omega_i \in A_i), \quad A_i \in \mathfrak{A}_i; \quad 1 \leqslant i \leqslant n,$$

и заданы функции

$$\xi_1 = f_1(\omega_1), \ \ \xi_2 = f_2(x_1, \ \omega_2), \ \dots, \ \xi_i = f_i(x_1, \ x_2, \ \dots, \ x_{i-1}, \ \omega_i), \ \dots, \tag{5.86}$$

представляющие собой случайные величины при фиксированных $x_1, \ x_2, \ \dots$, то функция

$$F(x_i/x_1, \dots, x_{i-1}) = \mathrm{P}(f_i(x_1, \dots, x_{i-1}, \omega_i) < x_i) \tag{5.87}$$

называется *условной функцией распределения,* а величина

$$M(\xi_i/x_1, \ldots, x_{i-1}) = Mf_i(x_1, \ldots, x_{i-1}, \omega_i) \tag{5.88}$$

— *условным математическим ожиданием* случайной величины ξ_i при условии $\xi_1 = x_1, \ldots, \xi_{i-1} = x_{i-1}$.

Рассмотрим следующий пример. На заводе изготовляются снаряды. Случайные факторы, влияющие на вес снаряда, можно описать элементарным событием $\omega_1 \in \Omega_1$. Таким образом, вес снаряда $\xi_1 = f_1(\omega_1)$. Отклонение снаряда от цели при фиксированном его весе $\xi_1 = x_1$ зависит от случайных условий стрельбы (температуры воздуха, ветра и т. п.) $\omega_2 \in \Omega_2 : \xi_2 = f_2(x_1, \omega_2)$. Величина ущерба цели ξ_3 зависит от отклонения снаряда от цели, от веса снаряда и от прочности покрытия цели, которая также случайна и описывается элементарным событием $\omega_3 \in \Omega_3$. Таким образом, в конечном счете $\xi_3 = f_3(\xi_1, \xi_2, \omega_3)$. В данном примере $\mathrm{P}\,(f_2(x_1, \omega_2) < x_2)$ есть условная функция распределения отклонения снаряда от цели при фиксированном весе снаряда $\xi_1 = x_1$, $\mathrm{P}\,(f_3(x_1, x_2, \omega_3) < x_3)$ — условная функция распределения ущерба цели при условии, что вес снаряда равен x_1, а отклонение его от цели составляет x_2. Представляется очевидным, что $\omega_1, \omega_2, \omega_3$ можно рассматривать как физически не связанные, а следовательно, независимые случайные величины. Подобным же образом можно определить математическое ожидание ущерба, причиненного цели, при условии, что $\xi_1 = x_1$, $\xi_2 = x_2$:

$$M(\xi_3/x_1, x_2) = Mf_3(x_1, x_2, \omega_3).$$

В формуле (5.86) $\xi_1, \xi_2, \ldots$ могут быть и многомерными случайными величинами; тогда в формуле (5.87) x_i будет вектором, а неравенство $f_i < x_i$ заменится соответствующим числом неравенств для компонент вектора f_i.

Основные свойства условных распределений и условных математических ожиданий состоят в следующем. При $i \geqslant 2$

$$F_{\xi_i}(x_i) = \int \ldots \int F(x_i/x_1, \ldots, x_{i-1})\, dF_{(\xi_1, \ldots, \xi_{i-1})}(x_1, \ldots, x_{i-1}), \tag{5.89}$$

$$M\xi_i = \int \ldots \int M(\xi_i/x_1, \ldots, x_{i-1})\, dF_{\xi_1, \ldots, \xi_{i-1}}(x_1, \ldots, x_{i-1}), \tag{5.90}$$

если эти интегралы существуют.

Докажем вначале формулу (5.90), предположив, что можно использовать формулу (5.72) в силу независимости $(\xi_1, \dots, \xi_{i-1})$ и ω_i, и считая, что ω_i — также случайная величина. Имеем

$$M\xi_i = Mf_i(\xi_1, \xi_2, \dots, \xi_{i-1}, \omega_i) =$$

$$= \int \dots \int f_i(x_1, \dots, x_{i-1}, x_i)\, dF_{(\xi_1, \dots, \xi_{i-1}, \omega_i)}(x_1, x_2, \dots, x_{i-1}, x_i) = \int \dots \int dF_{(\xi_1, \dots, \xi_{i-1})}(x_1, \dots, x_{i-1}) \times$$

$$\times \int f_i(x_1, \dots, x_{i-1}, y)\, dF\omega_i(y) =$$

$$= \int \dots \int M(\xi_i/x_1, \dots, x_{i-1})\, dF_{(\xi_1, \dots, \xi_{i-1})}(x_1, \dots, x_{i-1}). \tag{5.91}$$

Теперь докажем формулу (5.89). Положим $\eta = 1$, если $\xi_i < x_i$, $\eta = 0$, если $\xi_i \geqslant x_i$. Тогда

$$F_{\xi_i}(x_i) = M\eta =$$

$$= \int \dots \int M(\eta/x_1, \dots, x_{i-1})\, dF_{(\xi_1, \dots, \xi_{i-1})}(x_1, \dots, x_{i-1}) =$$

$$= \int \dots \int F(x_i/x_1, \dots, x_{i-1})\, dF_{(\xi_1, \dots, \xi_{i-1})}(x_1, \dots, x_{i-1}), \tag{5.92}$$

что и требовалось доказать.

Справедливы и такие разновидности формул (5.89), (5.90):

$$F_{\xi_i}(x_i) = \int \dots \int E(x_i - f_i(x_1, \dots, x_{i-1}, y)) \times$$

$$\times dF_{(\xi_1, \dots, \xi_{i-1})}(x_1, \dots, x_{i-1})\, dF_{\omega_i}(y), \tag{5.93}$$

где $E(x) = 1$ при $x > 0$; $E(x) = 0$ при $x \leqslant 0$;

$$M\xi_i = \int \dots \int f_i(x_1, \dots, x_{i-1}, y) \times$$

$$\times dF_{(\xi_1, \dots, \xi_{i-1})}(x_1, \dots, x_{i-1})\, dF_{\omega_i}(y). \tag{5.94}$$

Подобным же образом можно было бы в рассмотренном выше примере вычислить функцию распределения величины ущерба, причиняемого цели снарядом, и математическое ожидание этой величины.

Пусть $\xi_1 = (\xi_{11}, \ldots, \xi_{1n})$, $\xi_2 = (\xi_{21}, \ldots, \xi_{2n})$, ..., $\xi_k = (\xi_{k1}, \ldots, \xi_{kn})$ — случайные величины с функциями распределения $F_{\xi_1}(x_1), \ldots, F_{\xi_k}(x_k)$, где $x_1, x_2, \ldots, x_k$ — векторы соответствующих размерностей. $\xi_1, \xi_2, \ldots, \xi_k$ называются независимыми (в совокупности), если для любых x_1, $x_2, \ldots, x_k$

$$F_{(\xi_1, \ldots, \xi_k)}(x_1, \ldots, x_k) = F_{\xi_1}(x_1) F_{\xi_2}(x_2) \ldots F_{\xi_k}(x_k). \quad (5.95)$$

Если распределения $\xi_1, \ldots, \xi_k$ непрерывны, то

$$p_{(\xi_1, \ldots, \xi_k)}(x_1, \ldots, x_k) = p_{\xi_1}(x_1) \ldots p_{\xi_k}(x_k). \quad (5.96)$$

Пусть $f_1(x_1), \ldots, f_k(x_k)$ — функции, для которых существуют $Mf_i(\xi_i)$, $1 \leqslant i \leqslant k$. Тогда

$$Mf_1(\xi_1) \ldots f_k(x_k) = Mf_1(\xi_1) \ldots Mf_k(\xi_k). \quad (5.97)$$

Эта формула является следствием того, что $f_1(\xi_1), \ldots, f_k(\xi_k)$ — независимые одномерные случайные величины.

Условные распределения и условные математические ожидания можно определить и в том случае, если ξ_i заданы не формулой, подобной (5.86), а просто функцией распределения.

Пусть $\xi_1, \ldots, \xi_l$ — случайные величины, заданные функцией распределения $F_{(\xi_1, \ldots, \xi_l)}(x_1, \ldots, x_l)$. Условной функцией распределения случайной величины ξ_l при условии $\xi_1 = x_1, \ldots, \xi_{l-1} = x_{l-1}$ называется такая функция $F(x_l/x_1, \ldots, x_{l-1})$, что для любых отрезков $\Delta_1, \Delta_2, \ldots, \Delta_{l-1}$ выполняется формула

$$P(\xi_1 \in \Delta_1, \ldots, \xi_{l-1} \in \Delta_{l-1}, \quad \xi_l < x_l) =$$
$$= \int\limits_{x_1 \in \Delta_1, \ldots, x_{l-1} \in \Delta_{l-1}} \!\!\!\ldots\! \int F(x_l/x_1, \ldots, x_{l-1}) dF_{(\xi_1, \ldots, \xi_{l-1})}(x_1, \ldots, x_{l-1}). \quad (5.98)$$

Если случайная величина $\xi = (\xi_1, \ldots, \xi_l)$ обладает плотностью $p_\xi(x_1, \ldots, x_l)$, то условию (5.98) удовлетворяет функция

$$F(x_l/x_1, \ldots x_{l-1}) = \frac{\int\limits_{-\infty}^{x_l} p_\xi(x_1, \ldots, x_{l-1}, t)\, dt}{\int\limits_{-\infty}^{\infty} p_\xi(x_1, \ldots, x_{l-1}, t)\, dt} \quad (5.99)$$

(если знаменатель правой части равен 0, определяем левую часть как 0).

Действительно,

$$\int_{-\infty}^{\infty} p_\xi(x_1, \ldots, x_{i-1}, t)\, dt = p_{(\xi_1, \ldots, \xi_{i-1})}(x_1, \ldots, x_{i-1}),$$

так что

$$\int_{x_1 \in \Delta_1, \ldots, x_{i-1} \in \Delta_{i-1}} \ldots \int F(x_1, \ldots, x_{i-1})\, dF_{(\xi_1, \ldots, \xi_{i-1})}(x_1, \ldots, x_{i-1}) =$$

$$= \int_{x_1 \in \Delta_1, \ldots, x_{i-1} \in \Delta_{i-1}} \ldots \int \int_{-\infty}^{x_i} p_\xi(x_1, \ldots, x_{i-1})\, dx_1, \ldots, dx_{i-1}, \quad dt =$$

$$= \mathrm{P}(\xi_1 \in \Delta_1, \ldots, \xi_{i-1} \in \Delta_{i-1}, \ \xi_i < x_i),$$

что и требовалось доказать.

Пусть $I_j = 1$, если $\xi_j \in \Delta_j$; $I_j = 0$, если $\xi_j \overline{\in} \Delta_j$ $(1 \leqslant j < i)$. Функция $M(\xi_i / x_1, \ldots, x_{i-1})$ называется *условным математическим ожиданием* ξ_i *при условии* $\xi_1 = x_1, \ldots, \xi_{i-1} = x_{i-1}$, *если для любых отрезков* Δ_1, $\Delta_2, \ldots, \Delta_{i-1}$ *выполняется равенство*

$$M\xi_i I_1 I_2 \ldots I_{i-1} =$$

$$= \int_{x_1 \in \Delta_1 \ldots x_{i-1} \in \Delta_{i-1}} \ldots \int M(\xi_i / x_1, \ldots, x_{i-1}) \times$$

$$\times\, dF_{(\xi_1, \ldots, \xi_{i-1})}(x_1, \ldots, x_{i-1}) \qquad (5.100)$$

Если случайная величина $\xi = (\xi_1, \ldots, \xi_i)$ обладает плотностью $p_\xi(x_1, \ldots, x_i)$, то условию (5.100) удовлетворяет функция

$$M(\xi_i / x_1, \ldots, x_{i-1}) = \frac{\int_{-\infty}^{\infty} t p_\xi(x_1, \ldots, x_{i-1}, t)\, dt}{\int_{-\infty}^{\infty} p_\xi(x_1, \ldots, x_{i-1}, t)\, dt}. \qquad (5.101)$$

§ 7. МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Рассмотрим совокупность n независимых нормальных случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, где

$$p_{\xi_i}(x) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left\{-\frac{x^2}{2\sigma_i^2}\right\}, \quad 1 \leqslant i \leqslant n.$$

Тогда, по формуле (5.10), n-мерная случайная величина $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ обладает плотностью

$$p_\xi(x_1, x_2, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}\,\sigma_1\sigma_2\ldots\sigma_n} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}\right\}. \tag{5.102}$$

Поверхности уровня данной функции имеют вид эллипсоидов $\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2} = c^2$, главные оси которых направлены вдоль координатных осей n-мерного пространства. Поскольку $\sum_{i=1}^{n}\frac{\xi_i^2}{\sigma_i^2}$ имеет распределение χ^2 с n степенями свободы, то мы можем вычислить вероятность попадания $(\xi_1, \xi_2, \ldots, \xi_n)$ внутрь указанного эллипсоида: $F_{\chi_n^2}(c^2)$ или $F_{\chi_n}(c)$.

Рассмотрим теперь ξ как вектор-столбец: $\xi = (\xi_1, \xi_2, \ldots, \xi_n)'$, где штрих — символ операции транспонирования матрицы, и введем новую n-мерную случайную величину $\eta = (\eta_1, \eta_2, \ldots, \eta_n)'$ по формуле

$$\xi = C\eta \quad \text{или} \quad \eta = C'\xi,$$

где $C = \|c_{ij}\|$ — произвольная ортогональная матрица.

По формуле (5.31) плотность $p_\eta(x)$ существует и равна

$$p_\eta(x) = p_\xi(Cx), \tag{5.103}$$

где $x = (x_1, \ldots, x_n)$.

Введем матрицу

$$D = \begin{Vmatrix} \frac{1}{\sigma_1^2} & & 0 \\ & \frac{1}{\sigma_2^2} \cdot \cdot & \cdot \\ 0 & & \frac{1}{\sigma_n^2} \end{Vmatrix}. \tag{5.104}$$

Тогда имеет место формула

$$\sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2} = x'Dx, \tag{5.105}$$

откуда

$$p_\xi(x) = \frac{1}{(2\pi)^{n/2}\sigma_1 \ldots \sigma_n} \exp\left\{-\frac{1}{2} x'Dx\right\}. \tag{5.106}$$

Подставив в эту формулу Cx вместо x, найдем

$$p_\eta(x) = \frac{1}{(2\pi)^{n/2}\sigma_1\sigma_2 \ldots \sigma_n} \exp\left\{-\frac{1}{2} x'C'DCx\right\}. \tag{5.107}$$

Обозначим

$$A = C'DC. \tag{5.108}$$

Тогда

$$\det A = \det C' \cdot \det D \cdot \det C = \det D = \frac{1}{\sigma_1^2\sigma_2^2 \ldots \sigma_n^2}.$$

На основании последнего равенства можно записать

$$p_\eta(x) = \frac{\sqrt{\det A}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2} x'Ax\right\}. \tag{5.109}$$

Если $A = \| a_{ij} \|$, то $x'Ax = \sum_{i,j=1}^{n} a_{ij}x_ix_j$.

Таким образом,

$$p_\eta(x_1, \ldots, x_n) = \frac{\sqrt{\det \| a_{ij} \|}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2} \sum_{i,j=1}^{n} a_{ij}x_ix_j\right\}. \tag{5.110}$$

Матрица A получена ортогональным преобразованием диагональной матрицы D с положительными элементами главной диагонали; поэтому A — положительно определенная матрица, т. е. $\sum a_{ij}x_ix_j > 0$ при любых $(x_1, \ldots, x_n)$, кроме $(0, \ldots, 0)$. Раскроем вероятностный смысл матрицы A. Для этого рассмотрим матрицу $\| M\xi_i\xi_j \|$.

Заметим, что $M\xi_i\xi_j = M\xi_iM\xi_j = 0$ при $i \neq j$ и $M\xi_i\xi_j =$

$= M\xi_i^2 = \sigma_i^2$ при $i = j$. Таким образом,

$$\| M\xi_i\xi_j \| = D^{-1}. \tag{5.111}$$

С другой стороны,

$$M\xi_i\xi_j = M\left(\sum_{k=1}^{n} c_{ik}\eta_k\right)\left(\sum_{l=1}^{n} c_{jl}\eta_l\right) =$$

$$= \sum_{k,\, l=1}^{n} c_{ik} M\eta_k\eta_l c'_{lj}, \tag{5.112}$$

где $c'_{jl} = c_{il}$. Последнее выражение представляет собой (i, j)-й элемент матрицы CRC', где $R = \| M\eta_k\eta_l \|$. Отсюда

$$D^{-1} = CRC', \quad D = CR^{-1}C'. \tag{5.113}$$

Подставив последнее выражение в формулу $A = C'DC$, найдем

$$A = C'CR^{-1}C'C = R^{-1}. \tag{5.114}$$

Поскольку $\eta = C^{-1}\xi$, т. е. η_i — линейные комбинации ξ_i, то из равенств $M\xi_j = 0$, $1 \leqslant j \leqslant n$, следует, что $M\eta_i = 0$, $1 \leqslant i \leqslant n$. Отсюда имеем

$$M\eta_i\eta_j = \operatorname{cov}(\eta_i, \eta_j) \quad R = \| \operatorname{cov}(\eta_i, \eta_j) \|. \tag{5.115}$$

Итак,

$$A = \| \operatorname{cov}(\eta_i, \eta_j) \|^{-1}; \quad \det A = [\det \| \operatorname{cov}(\eta_i, \eta_j) \|]^{-1}. \tag{5.116}$$

Окончательно можно записать

$$p_\eta(x) = \frac{1}{(2\pi)^{n/2}\sqrt{\det R}} \exp\left\{-\frac{1}{2} x'R^{-1}x\right\}, \tag{5.117}$$

где R — ковариационная матрица многомерной случайной величины $\eta = (\eta_1, \eta_2, \dots, \eta_n)'$.

Пусть $a = (a_1, a_2, \dots, a_n)'$ — некоторый вектор-столбец. Рассмотрим многомерную случайную величину

$$\zeta = \eta + a = (\eta_1 + a_1, \eta_2 + a_2, \dots, \eta_n + a_n)'. \tag{5.118}$$

Вследствие формулы (5.34) $p_\zeta(x)$ существует и задается формулой

$$p_\zeta(x) = p_\eta(x - a) = \frac{\sqrt{\det \| a_{ij} \|}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2} \sum_{i,\, j=1}^{n} a_{ij} \times\right.$$

$$\times (x_i - a_i)(x_j - a_j)\Bigg\} =$$

$$= \frac{\sqrt{\det A}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}(x' - a')A(x - a)\right\}. \quad (5.119)$$

Случайная величина ζ с плотностью вероятностей, задаваемой последней формулой при некоторой симметричной положительно определенной матрице A, называется *n-мерной нормальной случайной величиной*, а ее плотность — *n-мерной нормальной плотностью*.

Пусть $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_n)$. Тогда $M\zeta_i = a_i$, $1 \leqslant i \leqslant n$, а $\operatorname{cov}(\zeta_i, \zeta_j)$ являются элементами A^{-1}. Отсюда вытекает, что n-мерное нормальное распределение полностью определяется совокупностью $\{M\zeta_i, \operatorname{cov}(\zeta_i, \zeta_j)\}$. Поскольку любая симметричная положительно определенная матрица A может быть представлена в виде $A = C'DC$, где C — ортогональная матрица, а D — диагональная матрица с положительными диагональными элементами, то отсюда вытекает следующее утверждение.

Любая n-мерная нормальная случайная величина может быть представлена в виде

$$\zeta = B\xi + a, \quad (5.120)$$

где ξ — n-мерный вектор, составленный из n независимых нормальных случайных величин; B — ортогональная матрица; a — постоянный вектор*.

Отметим следующее замечательное свойство нормального распределения. Пусть имеется любое конечное число многомерных случайных величин $(\xi_1, \ldots, \xi_m)$, $(\xi_{m+1}, \ldots, \xi_n)$, ..., $(\xi_r, \ldots, \xi_s)$, причем $(\xi_1, \ldots, \xi_m, \xi_{m+1}, \ldots, \xi_s)$ — многомерная нормальная случайная величина. Пусть $\operatorname{cov}(\xi_i, \xi_j) = 0$, если ξ_i и ξ_j принадлежат различным совокупностям $(\xi_1, \ldots, \xi_m)$, $(\xi_{m+1}, \ldots, \xi_n)$ ... . Тогда многомерные случайные величины $(\xi_1, \ldots, \xi_m)$, $(\xi_{m+1}, \ldots, \xi_n)$ ... $(\xi_r, \ldots, \xi_s)$ независимы в совокупности. Короче можно сказать, что для нормальных случайных величин из некоррелированности следует независимость. [Случайные величины ξ, η называются некоррелированными, если $\operatorname{cov}(\xi, \eta) = 0$.]

Доказательство сформулированного утверждения весьма просто. В принятых предпосылках ковариационная мат-

---

*) В наших выкладках $B = C'$.

рица R случайной величины $\xi = (\xi_1, \xi_2, \dots, \xi_s)$ имеет вид

$$R = \begin{Vmatrix} R_{1,m} & & & 0 \\ & R_{m+1,n} & & \\ & & \ddots & \\ 0 & & & R_{r,s} \end{Vmatrix}, \tag{5.121}$$

где R_{ij} — ковариационная матрица $(\xi_i, \dots, \xi_j)$. Тогда

$$A = R^{-1} = \begin{Vmatrix} A_{1,m} & & & 0 \\ & A_{m+1,n} & & \\ & & \ddots & \\ 0 & & & A_{r,s} \end{Vmatrix}, \tag{5.122}$$

где $A_{ij} = R_{ij}^{-1}$. Отсюда следует, что квадратичная форма $\sum\limits_{i,j=1}^{n} a_{ij}(x_i - a_i)(x_j - a_j)$ может быть представлена в виде

$$\sum_{i,j=1}^{m} a_{ij}(x_i - a_i)(x_j - a_j) + \sum_{i,j=m+1}^{n} a_{ij}(x_i - a_i)(x_j - a_j) + \\ + \dots + \sum_{i,j=r}^{s} a_{ij}(x_i - a_i)(x_j - a_j). \tag{5.123}$$

Но тогда плотность многомерного случайного вектора $(\xi_1, \dots, \xi_s)$ представляется в виде произведения плотностей $(\xi_1, \dots, \xi_m)$, $(\xi_{m+1}, \dots, \xi_n)$, ..., $(\xi_r, \dots, \xi_s)$; последнее ее и означает, что $(\xi_1, \dots, \xi_m)$, $(\xi_{m+1}, \dots, \xi_n)$, ..., $(\xi_r, \dots, \xi_s)$ независимы в совокупности. Это и требовалось доказать.

Пусть $\xi = (\xi_1, \dots, \xi_n)$ — случайная величина. Случайная величина меньшей размерности $\eta = (\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_m})$, где $i_1, i_2, \dots, i_m$ — различные индексы от 1 до n, называется *редукцией случайной величины* ξ.

Теорема 5.3. *Если ξ — нормальная случайная величина, то ее редукция — также нормальная случайная величина.*

Без ограничения общности можно полагать, что $\eta = (\xi_1, \dots, \xi_m)$, где $1 \leqslant m \leqslant n$, и что $M\xi_i = 0$ $(1 \leqslant i \leqslant n)$. Введем новые переменные $z_{m+1}, \dots, z_n$ как линейные комбинации $x_{m+1}, \ldots, x_n$ такие, что

$$\sum_{i,j=m+1}^{n} a_{ij} x_i x_j = \sum_{i=m+1}^{n} b_i z_i^2.$$

Тогда

$$p_\eta(x_1, x_2, \ldots, x_m) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} p_\xi(x_1, \ldots, x_n)\, dx_{m+1} \ldots dx_n =$$

$$= c_1 \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \sum_{i,j=1}^{m} a_{ij} x_i x_j - \frac{1}{2} \sum_{i=m+1}^{n} b_i z_i^2 + \right.$$

$$\left. + \sum_{i=m+1}^{n} z_i \sum_{j=1}^{m} b_{ij} x_j\right\} dz_{m+1} \ldots dz_n,$$

где b_{ij} — некоторые постоянные.

Подстановкой

$$t_i = z_i - \frac{1}{\sqrt{b_i}} \sum_{j=1}^{n} b_{ij} x_j, \quad m+1 \leqslant i \leqslant n$$

данный интеграл приводится к виду

$$c_1 \exp\left\{-\frac{1}{2} Q(x_1, \ldots, x_m)\right\} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \sum b_i t_i^2\right\} \times$$

$$\times\, dt_{m+1} \ldots dt_n,$$

т. е. к виду

$$c_2 \exp\left\{-\frac{1}{2} Q(x_1, \ldots, x_m)\right\},$$

где c_1, c_2 — постоянные; Q — квадратическая форма. Теорема доказана.

Выше рассматривалось ортогональное преобразование многомерной нормальной случайной величины. Рассмотрим теперь линейные преобразования более общего вида.

Теорема 5.4. Пусть $\xi = (\xi_1, \ldots, \xi_2)'$ — *n-мерная нормальная случайная величина; A — матрица порядка $m \times n$, ранга m ($m \leqslant n$). Тогда случайная величина $\eta = A\xi$ имеет m-мерное нормальное распределение.*

Для доказательства дополним матрицу A до невырожденной матрицы A^* порядка $n \times n$, добавив к ней $n - m$ строк. Пусть $\eta^* = A^*\xi$. Тогда по формуле (5.90) имеем

$$p_{\eta^*}(x) = \frac{1}{|\det A^*|} p_\xi(A^{*-1}x). \tag{5.124}$$

Последнее выражение представляет собой постоянную, умноженную на экспоненту от некоторой квадратической формы. Поскольку $p_{\eta^*}(x)$ — плотность, то упомянутая квадратическая форма положительно определена. Отсюда следует, что η^* — n-мерная нормальная случайная величина. Так как η — редукция η^*, то η также имеет нормальное распределение. Теорема доказана.

Многомерное нормальное распределение полностью задается математическими ожиданиями и ковариационной матрицей, поэтому для нахождения распределения $\eta = A\xi$ [ξ — n-мерная нормальная случайная величина; A — матрица порядка $m \times n$, ранга m $(1 \leqslant m \leqslant n)$], достаточно уметь вычислять математическое ожидание и ковариационную матрицу случайной величины η.

Выведем формулы, позволяющие находить математические ожидания и ковариационные матрицы линейных функций любых многомерных случайных величин с конечными дисперсиями.

Если $A = \|a_{ij}\|$, то

$$\eta_i = \sum_{j=1}^{n} a_{ij}\xi_j, \quad 1 \leqslant i \leqslant m. \tag{5.125}$$

Отсюда

$$M\eta_i = \sum_{j=1}^{n} a_{ij} M\xi_j, \quad 1 \leqslant i \leqslant m. \tag{5.126}$$

В векторном виде последнее равенство записывается так:

$$\left\| \begin{matrix} M\eta_1 \\ \vdots \\ M\eta_m \end{matrix} \right\| = A \left\| \begin{matrix} M\xi_1 \\ \vdots \\ M\xi_n \end{matrix} \right\| \tag{5.127}$$

Для нахождения ковариационной матрицы η воспользуемся тем свойством, что она остается неизменной, если к случайным величинам добавить любые постоянные. Пусть $\eta_i^0 = \eta_i - M\eta_i$, $\xi_j^0 = \xi_j - M\xi_j$, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$. Тогда можно записать

$$\eta_i^0 = \sum_{j=1}^{n} a_{ij}\xi_j^0, \quad 1 \leqslant i \leqslant m, \tag{5.128}$$

откуда

$$\operatorname{cov}(\eta_i,\eta_k)=M\eta_i^0\eta_k^0=M\left(\sum_{j=1}^{n}a_{ij}\xi_j^0\right)\left(\sum_{l=1}^{n}a_{il}\xi_l^0\right)=$$

$$=\sum_{j,l=1}^{n}a_{ij}\operatorname{cov}(\xi_j,\xi_l)\,a'_{lk}, \tag{5.129}$$

где $a'_{lk}=a_{kl}$. Найденное выражение означает, что ковариационная матрица R_η случайной величины η выражается через ковариационную матрицу R_ξ случайной величины ξ следующим образом:

$$R_\eta=A\,R_\xi\,A'.$$

Заметим, что в приведенном рассуждении свойство нормальности распределения ξ не использовано. Найденные формулы и позволяют находить математические ожидания и ковариационные матрицы линейных функций любых многомерных случайных величин с конечными дисперсиями.

Рассмотрим частный случай многомерного нормального распределения при $n=2$. Пусть $D\xi_1=\sigma_1^2$, $D\xi_2=\sigma_2^2$. Ковариационная матрица случайной величины $\xi=(\xi_1,\xi_2)$ имеет вид

$$R=\begin{Vmatrix}\sigma_1^2 & \operatorname{cov}(\xi_1,\xi_2)\\ \operatorname{cov}(\xi_1,\xi_2) & \sigma_2^2\end{Vmatrix}. \tag{5.130}$$

Отношение

$$r=\frac{\operatorname{cov}(\xi_1,\xi_2)}{\sqrt{D\xi_1\,D\xi_2}} \tag{5.131}$$

называется *коэффициентом корреляции случайных величин* ξ_1, ξ_2. В соответствии с этим $\operatorname{cov}(\xi_1,\xi_2)=r\sigma_1\sigma_2$ и можно записать

$$R=\begin{Vmatrix}\sigma_1^2 & \sigma_1\sigma_2 r\\ \sigma_1\sigma_2 r & \sigma_2^2\end{Vmatrix}. \tag{5.132}$$

Матрицей, обратной к R, является матрица

$$A = R^{-1} = \left\| \begin{matrix} \frac{1}{\sigma_1^2 (1 - r^2)} & -\frac{r}{\sigma_1 \sigma_2 (1 - r^2)} \\ -\frac{r}{\sigma_1 \sigma_2 (1 - r^2)} & \frac{1}{\sigma_2^2 (1 - r^2)} \end{matrix} \right\|. \qquad (5.133)$$

Определитель этой матрицы

$$\det A = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - r^2)}. \qquad (5.134)$$

Отсюда

$$p_\xi (x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - r^2}} \exp \left\{ -\frac{1}{2(1 - r^2)} \left[\left(\frac{x_1 - a_1}{\sigma} \right)^2 - \right. \right.$$

$$\left. \left. -2r \left(\frac{x_1 - a_1}{\sigma_1} \right) \left(\frac{x_2 - a_2}{\sigma_2} \right) + \left(\frac{x_2 - a_2}{\sigma_2} \right)^2 \right] \right\}, \qquad (5.135)$$

где $a_1 = M\xi_1$, $a_2 = M\xi_2$.

При $r = 0$, т. е. при $\operatorname{cov}(\xi_1, \xi_2) = 0$, получаем $p_\xi (x_1, x_2)$ в виде произведения плотностей ξ_1 и ξ_2.

Выведем формулу для характеристической функции многомерной нормальной случайной величины.

Если $\xi = (\xi_1, \ldots, \xi_n)$ — любая многомерная случайная величина, то характеристической функцией этой величины называют функцию вещественного вектора $t = (t_1, \ldots, t_n)$ вида

$$\varphi_\xi (t) = \varphi_\xi (t_1, \ldots, t_n) = Me^{i(t, \xi)} = M \exp \left\{ i \sum_{i=1}^{n} t_i \xi_i \right\}.$$

Напомним, что если γ — одномерная нормальная случайная величина с математическим ожиданием 0 и дисперсией 1, то (см. § 5, гл. IV)

$$Me^{it\gamma} = \exp \left\{ -\frac{1}{2} \sigma^2 t^2 \right\}.$$

Пусть $\xi = (\xi_1, \xi_2, \ldots, \xi_n)'$ — вектор-столбец, составленный из независимых нормальных случайных величин с математическими ожиданиями, равными нулю, и дисперсиями $\sigma_1^2, \ldots, \sigma_n^2$. Тогда

$$\varphi_\xi(t) = Me^{i(t,\xi)} = \prod_{i=1}^{n} Me^{it_i\xi_i} = \prod e^{-\frac{\sigma_i^2 t_i^2}{2}} =$$

$$= \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sigma_i^2 t_i^2\right\}.$$

Пусть, далее, η — многомерная нормальная случайная величина, задаваемая соотношением $\xi = C\eta$, где C — ортогональная матрица. Тогда

$$\varphi_\eta(t) = Me^{i(t,\eta)} = Me^{i(t,C'\xi)} = Me^{i(Ct,\xi)} = \varphi_\xi^{(Ct)}.$$

Поскольку $\varphi_\xi(t)$ есть экспоненциальная функция от квадратической формы переменных $t_1, \ldots, t_n$, то $\varphi_\xi(Ct)$ — также экспоненциальная функция квадратической формы переменных $t_1, \ldots, t_n$:

$$\varphi_\eta(t) = \exp\left\{-\frac{1}{2}\sum_{i,j=1}^{n} f_{ij}t_i t_j\right\}$$

(без ограничения общности можно считать, что $f_{ij} = f_{ji}$).

Выясним смысл постоянных f_{ij}. Положим $t_i = x$; $t_j = 0$, $j \neq i$, тогда имеем

$$\varphi_\eta = (0, \ldots, 0, x, 0, \ldots, 0) = Me^{ix\eta_i}.$$

Отсюда

$$\frac{d^2}{dx^2}\varphi_\eta(0, \ldots, 0, x, 0, \ldots, 0)\Big|_{x=0} = \frac{\partial^2\varphi_\eta(t_1, \ldots, t_n)}{\partial t_i^2}\Big|_{t=0} =$$

$$= \frac{d^2}{dx^2}\int_{-\infty}^{\infty} e^{ix^2} p_{\eta_i}(z)\,dz\Big|_{x=0} = -\int_{-\infty}^{\infty} z^2 p_{\eta_i}(z)\,dz =$$

$$= -M\eta_i^2 = -D\eta_i.$$

Положим затем для некоторых фиксированных i и j ($i < j$) $t_i = x$, $t_j = y$; $t_k = 0$, $i \neq k \neq j$. Тогда

$$\varphi_\eta(0, \ldots, 0, x, 0, \ldots, 0, y, 0, \ldots, 0) = Me^{i(x\eta_i + y\eta_j)}.$$

Отсюда

$$\frac{\partial^2}{\partial x \partial y} \varphi_\eta (0, \ldots, x, \ldots, y, \ldots 0 \Big|_{x=y=0} =$$

$$= \frac{\partial^2 \varphi_\eta (t_1, \ldots, t_n)}{\partial t_i \partial t_j}\Bigg|_{t=0} = \frac{\partial^2}{\partial x \partial y} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} e^{i(xz+yu)} \times$$

$$\times\, p_{(\eta_i, \eta_j)} (z, u) dz du \Big|_{x=y=0} =$$

$$= -\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} zu p_{(\eta_i, \eta_j)} (z, u)\, dz du = -\operatorname{cov}(\eta_i, \eta_j).$$

Итак,

$$\frac{\partial^2 \varphi}{\partial t_i^2}\Bigg|_{t=0} = -D\eta_i, \quad \frac{\partial^2 \varphi}{\partial t_i \partial t_j}\Bigg|_{t=0} = -\operatorname{cov}(\eta_i, \eta_j), \quad i \neq j.$$

В то же время, как легко подсчитать

$$\frac{\partial^2}{\partial t_i^2}\left[\exp\left\{-\frac{1}{2}\sum f_{ij} t_i t_j\right\}\right]\Bigg|_{t=0} = -f_{ii}$$

$$\frac{\partial^2}{\partial t_i \partial t_j}\left[\exp\left\{-\frac{1}{2}\sum f_{ij} t_i t_j\right\}\right]\Bigg|_{t=0} = -f_{ij}, \ i < j.$$

Сравнив полученные формулы, найдем:

$$f_{ii} = D\eta_i, \quad f_{ij} = \operatorname{cov}(\eta_i, \eta_j) \quad i < j.$$

Итак,

$$\| f_{ij} \| = \| \operatorname{cov}(\eta_i, \eta_j \| = R.$$

Окончательно имеем

$$\varphi_\eta (t) = \exp\left\{-\frac{1}{2}\sum \operatorname{cov}(\eta_i, \eta_j)\, t_i t_j\right\} = e^{-(Rt,\, t)}.$$

Если $\zeta = (\zeta_1, \ldots, \zeta_n)'$ — произвольная n-мерная нормальная случайная величина с математическим ожиданием $a = (a_1, \ldots, a_n)'$, то

$$\varphi_\zeta (t) = Me^{i(t,\, \zeta)} = e^{i(t,\, a) - (Rt,\, t)}.$$

где

$$R = \| \operatorname{cov}(\zeta_i, \zeta_j) \|.$$

[Доказать, формулу, используя при этом формулу для $\varphi_\eta(t)$, предлагаем читателю.]

§ 8. ПРЕОБРАЗОВАНИЕ МНОГОМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ В ДРУГУЮ МНОГОМЕРНУЮ СЛУЧАЙНУЮ ВЕЛИЧИНУ С ЗАДАННЫМ РАСПРЕДЕЛЕНИЕМ

Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — n-мерная случайная величина с плотностью $p(x_1, \ldots, x_n)$; пусть $G(x_1, \ldots, x_n)$ — произвольная n-мерная функция распределения. Требуется найти такое функциональное преобразование

$$\begin{aligned} \eta_1 &= f_1(\xi_1), \\ \eta_2 &= f_2(\eta_1, \xi_2), \\ & \ldots\ldots\ldots\ldots \\ \eta_n &= f_n(\eta_1, \eta_2, \ldots, \eta_{n-1}, \xi_n), \end{aligned} \tag{5.136}$$

чтобы n-мерная случайная величина $(\eta_1, \eta_2, \ldots, \eta_n)$ имела в качестве многомерной функции распределения $G(x_1 \ldots, x_n)$.

Эта задача решается следующим образом.

Рассмотрим графики функций распределения ξ_1 и η_1. Первая из них имеет вид

$$F_1(x_1) = \int_{-\infty}^{x_1} dx_1 \int_{-\infty}^{\infty} dx_2 \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} p(x_1, \ldots, x_n)\, dx_n. \tag{5.137}$$

Последняя равна $G(x_1) = G(x_1, \infty, \infty, \ldots, \infty)$. В точках разрыва функции график дополним вертикальными отрезками до сплошной кривой. Функциональное соответствие между ξ_1 и η_1 установим из следующего условия:

$$G_1(\eta_1) \leqslant F_1(\xi_1) \leqslant G_1(\eta_1 + 0).$$

В частности, если $G_1(x_1)$ — непрерывная функция, то $G_1(\eta_1) = G_1(\eta_1 + 0)$, и мы получаем более простое соотношение:

$$G_1(\eta_1) = F_1(\xi_1). \tag{5.138}$$

Если прямая $y = F_1(x)$ пересекает график функции $y = G_1(x)$ по целому отрезку, *для данного* ξ определяем η_1

произвольным образом. (Такой случай может встретиться лишь с вероятностью 0.) Докажем, что η_1 имеет распределение $G_1(x)$. Действительно, пусть x — любое число; тогда $\eta_1 < x$ в том и только том случае, если $F_1(\xi_1) < G_1(x)$. В свою очередь

$$P(F_1(\xi_1) < y) = P(\xi_1 < F_1^{-1}(y)) = F_1(F_1^{-1}(y)) = y^{*)}.$$

Таким образом, подставив $y = G(x)$, найдем

$$P(\eta_1 < x) = P(F_1(\xi_1) < G_1(x)) = G_1(x), \tag{5.139}$$

т. е. η_1 действительно имеет функцию распределения $G_1(x)$.

Предположим, что $\eta_1, \eta_2, \ldots, \eta_{k-1}$ уже определены, причем $(k-1)$-мерная функция распределения случайной величины $(\eta_1, \eta_2, \ldots, \eta_{k-1})$ имеет вид

$$G_{k-1}(x_1, \ldots, x_{k-1}) = G(x_1, \ldots, x_{k-1}, \infty, \ldots, \infty).$$

Обозначим через $F_k(x_k/x_1, \ldots, x_{k-1})$ условную функцию распределения ξ_k при фиксированных $\eta_1, \ldots, \eta_{k-1}$, равных соответственно $x_1, \ldots, x_{k-1}$; через $G_k(x_k/x_1, \ldots, x_{k-1})$ — условную функцию распределения η_k при условии $\eta_1 = x_1, \ldots, \eta_{k-1} = x_{k-1}$. Установим соответствие между ξ_k и η_k с помощью формулы

$$G_k(x_k/x_1, \ldots, x_{k-1}) \leqslant F_k(x_k/x_1, \ldots, x_{k-1}) \leqslant \leqslant G_k(x_k + 0/x_1, \ldots, x_{k-1}) \tag{5.140}$$

[точно так же, как и между $F_1(x_1)$ и $G_1(x_1)$]; $x_1, x_2, \ldots, x_{k-1}$ играют роль параметров. Повторяя рассуждение, касающееся распределения η_1, находим, что при фиксированных $\eta_1, \ldots, \eta_{k-1}$ случайная величина η_k имеет условное распределение $G_k(x_k/x_1, \ldots, x_{k-1})$. Отсюда в силу предположения немедленно следует, что: 1) $(\eta_1, \ldots, \eta_{k-1})$ имеет $(k-1)$-мерное распределение $G_{k-1}(x_1, \ldots, x_{k-1})$; 2) $(\eta_1, \ldots, \eta_k)$ имеет распределение $G_k(x_1, \ldots, x_k) = G(x_1, \ldots, x_k, \infty, \ldots, \infty)$. По индукции, это имеет место для всех k, $1 \leqslant k \leqslant n$, т. е. в конечном счете $(\eta_1, \ldots, \eta_n)$ имеет распределение $G(x_1, \ldots, x_n)$, что и требовалось доказать.

Сделаем следующее замечание. Значениям $\eta_1 = x_1, \ldots, \eta_{k-1} = x_{k-1}$ соответствует, вообще говоря, множество значений $\xi_1, \xi_2, \ldots, \xi_{k-1}$ вида $a_1 \leqslant \xi_1 < b_1, \ldots, a_{k-1} \leqslant \xi_{k-1} < b_{k-1}$. В таком случае условное распределение

*) $F^{-1}(y)$ — функция, обратная к $F(x)$, т. е. если $y = F(x)$, то $x = F^{-1}(y)$.

$F_k(x_k/x_1, x_2, \ldots, x_{k-1})$ надлежит определять формулой

$$F_k(x_k/x_1, \ldots, x_{k-1}) =$$

$$= \frac{\underset{\substack{a_1 \leqslant y_1 \leqslant b_1 \\ \cdots\cdots\cdots \\ a_{k-1} \leqslant y_{k-1} \leqslant b_{k-1} \\ -\infty < y_k < x_k}}{\int \cdots \int} p_k(y_1, \ldots, y_{k-1}, y_k) \prod\limits_{\substack{1 \leqslant i \leqslant k-1 \\ b_i > a_i}} dy_i dy_k}{\underset{\substack{a_1 \leqslant y_1 \leqslant b_1 \\ \cdots\cdots\cdots \\ a_{k-1} \leqslant y_{k-1} \leqslant b_{k-1} \\ -\infty < y_k < \infty}}{\int \cdots \int} p_k(y_1, \ldots, y_{k-1}, y_k) \prod\limits_{\substack{1 \leqslant i \leqslant k-1 \\ b_i > a_i}} dy_i dy_k}. \quad (5.141)$$

Это выражение нужно понимать в том смысле, что интегрирование производится по переменным y_i, $1 \leqslant i \leqslant k-1$, для которых интервал изменения ξ_i при фиксированных $\eta_1, \eta_2, \ldots, \eta_i$ имеет положительную длину; в частности, если каждому набору $(y_1, \ldots, y_{k-1}) = (x_1, \ldots, x_{k-1})$ соответствует единственный набор $(\xi_1, \ldots, \xi_{k-1}) = (y_1, \ldots, y_{k-1})$, то

$$F_k(x_k/x_1, \ldots, x_{k-1}) = \frac{\int\limits_{-\infty}^{x_k} p_k(y_1, \ldots, y_{k-1}, y_k)\, dy_k}{\int\limits_{-\infty}^{\infty} p_k(y_1, \ldots, y_{k-1}, y_k)\, dy_k}. \quad (5.142)$$

Здесь $p_k(y_1, \ldots, y_k)$ — плотность распределения k-мерной случайной величины $(\xi_1, \ldots, \xi_k)$, т. е. при $k < n$

$$p_k(y_1, \ldots, y_k) =$$

$$= \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} p(y_1, y_2, \ldots, y_k, y_{k+1}, \ldots, y_n)\, dy_1 \ldots dy_n. \quad (5.143)$$

Принцип преобразования одних случайных величин в другие, изложенный в настоящем параграфе, разработан Н. В. Смирновым, эффективно применившим его к исследованию непараметрических статистических критериев.

Рассмотрим некоторые примеры.

1. Пусть ξ — случайная величина с произвольной функцией распределения $F(x)$. Требуется построить случайную величину η с функцией распределения $G(x) = F\left(\frac{x-a}{\sigma}\right)$, где a — произвольное число, σ — произвольное положительное число.

Имеем

$$F(\xi) = G(\eta), \tag{5.144}$$

или

$$F(\xi) = F\left(\frac{\eta - a}{\sigma}\right).$$

Достаточно положить $\xi = \frac{\eta - a}{\sigma}$ или $\eta = a + \sigma\xi$.

2. Пусть ξ — случайная величина с равномерным распределением в интервале (0,1). Требуется построить случайную величину $\eta = \eta(\xi)$ с заданной функцией распределения $G(x)$.

Так как $P(\xi < x) = x$, $0 \leqslant x \leqslant 1$, то получаем, что η определяется условием

$$G(\eta) \leqslant \xi \leqslant G(\eta + 0). \tag{5.145}$$

В частности, если $G(x)$ — непрерывная функция, то имеем $\xi = G(\eta)$ или $\eta = G^{-1}(\xi)$.

Это преобразование используется в методе Монте-Карло для выработки случайных чисел с заданным распределением из случайных чисел с равномерным распределением.

3. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — случайные величины с нормальным распределением, независимые в совокупности и обладающие математическими ожиданиями 0 и дисперсиями 1. Пусть, далее, $(b_1, \ldots, b_n)$ — заданный вектор, $\| a_{ij} \|$ — заданная положительно определенная матрица порядка $n \times n$. Требуется построить нормальный случайный вектор $\eta_1, \eta_2, \ldots, \eta_n$ с вектором математических ожиданий $(b_1, \ldots, b_n)$ и корреляционной матрицей $\| a_{ij} \|$.

Положим

$$\eta_1 = b_1 + \xi_1 \sqrt{a_{11}}. \tag{5.146}$$

Тогда, очевидно,

$$M\eta_1 = b_1 + M\xi_1 \sqrt{a_{11}} = b_1,$$

$$D\eta_1 = a_{11} D\xi_1 = a_{11}.$$

Будем искать η_k в виде

$$\eta_k = b_k + \eta_k^0, \quad 1 \leqslant k \leqslant n, \tag{5.147}$$

где

$$\eta_k^0 = c_{k1}\eta_1^0 + c_{k2}\eta_2^0 + \ldots + c_{k,\,k-1}\eta_{k-1}^0 + c_{k,\,k}\xi_k, \quad 1 \leqslant k \leqslant n. \tag{5.148}$$

При $1 \leqslant i \leqslant k - 1$ имеем

$$a_{ik} = M(\eta_i - M\eta_i)(\eta_k - M\eta_k) = M\eta_i^0\eta_k^0 =$$

$$= M\eta_i^0 \left(c_{k1}\eta_1^0 + \ldots + c_{k,\,k-1}\eta_{k-1}^0 + c_{k,\,k}\xi_k^0\right) =$$

$$= c_{k1} M\eta_i^0 \eta_1^0 + \ldots + c_{k,\,k-1} M\eta_i^0 \eta_{k-1}^0 =$$
$$= c_{k1} a_{i1} + \ldots + c_{k,\,k-1} a_{i,\,k-1} \tag{5.149}$$

(так как ξ_k не зависит от η_i^0, а, стало быть, $M\xi_k \eta_i^0 = M\xi_k M\eta_i^0 = 0$). Таким образом, для нахождения $c_{k,1}, \ldots, c_{k,\,k-1}$ нужно решить систему $k-1$ линейных уравнений:

$$\begin{gathered} c_{k1} a_{11} + c_{k2} a_{12} + \ldots + c_{k,\,k-1} a_{1,\,k-1} = a_{1k}, \\ c_{k1} a_{21} + c_{k2} a_{22} + \ldots + c_{k,\,k-1} a_{2,\,k-1} = a_{2k}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ c_{k1} a_{k-1,1} + c_{k2} a_{k-1,2} + \ldots + c_{k,\,k-1} a_{k-1,\,k-1} = a_{k-1,\,k}. \end{gathered} \tag{5.150}$$

Эта система имеет единственное решение $(c_{k1}, \ldots, c_{k,\,k-1})$. [В противном случае существовали бы такие $x_1, x_2, \ldots, x_{k-1}$, не все равные 0, что $\sum_{i=1}^{k-1} a_{ij} x_i = 0$ при всех j $(1 \leqslant j \leqslant k-1)$, т. е. $\sum_{j=1}^{k-1} x_j \sum_{i=1}^{k-1} a_{ij} x_i = 0$, что противоречит положительной определенности матрицы $\| a_{ij} \|$.]

Определив $c_{k1}, \ldots, c_{k,\,k-1}$, находим

$$a_{kk} = M(\eta_k^0)^2 = M(c_{k1}\eta_1^0 + \ldots + c_{k,\,k-1}\eta_{k-1}^0 + c_{kk}\xi_k)^2 =$$
$$= \sum_{\substack{1 \leqslant i \leqslant k-1 \\ 1 \leqslant j \leqslant k-1}} c_{ki} c_{kj} M\eta_i^0 \eta_j^0 + c_{kk}^2 M\xi_k^2 =$$
$$= \sum_{\substack{1 \leqslant i \leqslant k-1 \\ 1 \leqslant j \leqslant k-1}} c_{ki} c_{kj} a_{ij} + c_{kk}^2. \tag{5.151}$$

Следовательно, c_{kk} с точностью знака определяется равенством

$$c_{kk}^2 = a_{kk} - \sum_{\substack{1 \leqslant i \leqslant k-1 \\ 1 \leqslant j \leqslant k-1}} c_{ki} c_{kj} a_{ij}, \tag{5.152}$$

все члены в правой части которого уже известны.

Как следует из результатов, полученных в § 7, рассмотренную задачу можно также решить непосредственно с помощью линейного преобразования исходных случайных величин.

ГЛАВА VI
ЗАКОН БОЛЬШИХ ЧИСЕЛ

§ 1. ПРИНЦИП ПРАКТИЧЕСКОЙ ДОСТОВЕРНОСТИ И МАССОВЫЕ ЯВЛЕНИЯ

Рассматривая единичное случайное событие A, имеющее вероятность $\mathrm{P}(A)$, мы не можем, вообще говоря, высказать никаких суждений относительно того, наступит или не наступит это событие в данном опыте. Исключение составляют случаи, когда вероятность события A близка к единице или к нулю $[\mathrm{P}(A) \approx 1$ или $\mathrm{P}(A) \approx 0]$. Заметим, что если вероятность $\mathrm{P}(A)$ близка к единице (или близка к нулю), то вероятность противоположного события $\mathrm{P}(\overline{A}) = 1 - \mathrm{P}(A)$ близка к нулю (или близка к единице); таким образом, один случай легко сводится к другому. В случае, когда вероятность $\mathrm{P}(A)$ события A мала (или близка к единице), можно утверждать, что в данном единичном опыте событие A не произойдет (или произойдет). В соответствии с этим утверждением в практической деятельности поступают так, как если бы событие A было невозможным (или достоверным), т. е. считают событие A *практически невозможным* (или *практически достоверным*).

Таким образом, на практике можно руководствоваться следующим принципом:

если вероятность $\mathrm{P}(A)$ *события* A *мала* (*или близка к единице*), *то практически можно быть уверенным в том, что в данном единичном эксперименте событие* A *не произойдет* (*или произойдет*).

Насколько же малой должна быть вероятность $\mathrm{P}(A)$ события A, чтобы событие A можно было считать практически невозможным? Ответ на этот вопрос может дать только практика, причем ответ будет различен в разных конкретных ситуациях. Так, например, пусть вероятность $\mathrm{P}(A)$ выхода из строя некоторого прибора равна 0,01. Если этот прибор находится в наземной лаборатории (в случае поломки может быть заменен другим), то возможностью выхода его из строя можно пренебречь и считать событие A практически невозможным. Иное дело в условиях космического полета, когда прибор заменить нельзя и выход его из строя повлечет за собой серьезные последствия. В таких условиях вероятностью 0,01 пренебречь нельзя, и мы не можем считать выход из строя прибора практически невозможным событием.

Таким образом, в практических применениях теории вероятностей особую и притом важную роль играют события с вероятностями, близкими к нулю или к единице, или практически невозможные и практически достоверные события. Рассмотрим примеры подобных событий.

Пример 1. Пусть производится неограниченная последовательность испытаний Бернулли с вероятностью p успеха, и пусть μ_n — число успехов в первых n испытаниях. Тогда $\frac{\mu_n}{n}$ — частота успеха в этих испытаниях. При больших значениях n эта частота близка к вероятности p успеха:

$$\frac{\mu_n}{n} \approx p. \tag{6.1}$$

Заметим, что в левой части выражения (6.1) стоит случайная величина, в правой — неслучайное число p, а также что нет ничего невозможного в том, что все n испытаний приводили к успеху или только к неудаче (в таких случаях величина $\frac{\mu_n}{n}$ равна единице или нулю). Таким образом, встает вопрос о точном смысле приближенного равенства (6.1). Оставляя пока в стороне ответ на этот вопрос*), можем сказать, что событие A, состоящее в том, что частота успеха $\frac{\mu_n}{n}$ близка (в некотором смысле) к вероятности p, оказывается практически достоверным при больших n (это факт, известный из опыта, но никак не теорема).

Пример 2. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — результаты n независимых измерений некоторой физической постоянной a. Каждая из величин $\xi_1, \xi_2, \ldots, \xi_n$ есть сумма числа a и некоторой случайной ошибки измерения и поэтому является величиной случайной. На практике за приближенное значение величины a всегда принимается среднее арифметическое результатов n измерений: $\zeta_n = \frac{1}{n}\sum_{k=1}^{n} \xi_k$, т. е. полагается, что

$$a \approx \zeta_n. \tag{6.2**)}$$

На основании результатов, полученных из опыта, можно утверждать: событие A, состоящее в том, что ζ_n близко к постоянной

*) Ответ на этот вопрос дает теорема Бернулли (см. § 3 гл. VI).

**) Ответ на вопрос о том, какой точный смысл вкладывается в приближенное равенство (6.2), дает теорема Чебышева (см. § 3 гл. VI).

a (в некотором смысле), оказывается практически достоверным при больших n.

В рассмотренных примерах мы имели дело с массой случайных явлений. Таким образом, принцип практической достоверности применим, если мы имеем дело не с одним единичным явлением, а с массовыми явлениями. Математическая теория не может ссылаться на результаты эксперимента, она должна установить точный смысл утверждений типа (6.1) и (6.2), сформулировать общие условия, при которых эти утверждения справедливы, и обеспечить таким образом возможность их применения на практике. Этому посвящена настоящая глава.

§ 2. НЕРАВЕНСТВО ЧЕБЫШЕВА

В настоящем параграфе доказывается вспомогательное неравенство, называемое неравенством Чебышева.

Теорема 6.1. *Пусть ξ — случайная величина, имеющая конечные математическое ожидание $M\xi$ и дисперсию $D\xi$. Тогда вероятность того, что отклонение случайной величины ξ от ее математического ожидания $M\xi$ по абсолютной величине превзойдет наперед заданное положительное число* $\varepsilon > 0$, меньше, чем $\frac{D\xi}{\varepsilon^2}$, *т. е.*

$$P(|\xi - M\xi| > \varepsilon) < \frac{D\xi}{\varepsilon^2}. \tag{6.3}$$

Это неравенство и называется неравенством Чебышева.

Доказательство. Пусть ξ — случайная величина, имеющая функцию распределения вероятностей $F(x)$, тогда

$$D\xi = \int_{-\infty}^{+\infty} (x - M\xi)^2 \, dF(x) \geqslant \int_{|x - M\xi| > \varepsilon} (x - M\xi)^2 \, dF(x) >$$

$$> \varepsilon^2 \int_{|x - M\xi| > \varepsilon} dF(x) = \varepsilon^2 P\{|\xi - M\xi| > \varepsilon\}.$$

Рис. 6.1

Здесь область интегрирования $(-\infty, +\infty)$ мы вначале заменили на область $|x - M\xi| > \varepsilon$, изображенную на рис. 6.1, затем воспользовавшись тем, что в последней области выполнено неравенство $(x - M\xi)^2 > \varepsilon^2$, а интеграл $\int_D dF(x)$ равен вероятности $P\{\xi \in D\}$ попадания случайной величины в область D (в нашем случае область D является внешностью ε-окрестности точки $M\xi$). Разделив на ε^2 обе части полученного неравенства, имеем

$$P(|\xi - M\xi| > \varepsilon) < \frac{D\xi}{\varepsilon^2},$$

что и требовалось доказать.

Замечание. Если в неравенстве (6.3) положить $\eta = \xi - M\xi$, то $D\xi = M\eta^2$, и это неравенство принимает вид

$$P(|\eta| > \varepsilon) < \frac{M\eta^2}{\varepsilon^2}. \tag{6.4}$$

Проследив за доказательством неравенства Чебышева, можно убедиться, что неравенство (6.4) справедливо для любой величины η, имеющей конечную дисперсию (а следовательно и конечное математическое ожидание квадрата), а не только для величины η вида $\eta = \xi - M\xi$, где $D\xi < \infty$. Таким образом, получаем следующее обобщение неравенства Чебышева.

Теорема 6.2. *Пусть η — случайная величина, имеющая конечное математическое ожидание квадрата $M\eta^2 < \infty$ Тогда вероятность того, что величина η по абсолютной величине превзойдет наперед заданное положительное число $\varepsilon > 0$,* меньше, чем $\frac{M\eta^2}{\varepsilon^2}$, т. е. имеет место неравенство. (6.4).

Если в неравенстве (6.4) положить $\eta = \xi - M\xi$, где $D\xi < \infty$, то получим неравенство (6.3).

Неравенство (6.4) также будем называть неравенством Чебышева.

§ 3. ЗАКОН БОЛЬШИХ ЧИСЕЛ

В настоящем параграфе мы рассмотрим закон больших чисел в различных формах.

Теорема 6.3. (закон больших чисел Чебышева). *Пусть $\xi_1, \xi_2, \ldots$ — попарно независимые случайные величины* *с равномерно ограниченными дисперсиями, т. е.* $D\xi_k \leqslant C$, $k = 1, 2, \ldots,$ *где* C — *некоторая постоянная. Положим* $\zeta_n = \frac{1}{n}\sum_{k=1}^{n} \xi_k$. *Тогда последовательность* $(\zeta_n - M\zeta_n)$ *стремится по вероятности к нулю при* $n \to \infty$, *т. е.*

$$(\zeta_n - M\zeta_n) \xrightarrow{\text{по вер.}} 0 \quad \text{при} \quad n \to \infty.$$

[Смысл сходимости по вероятности $\alpha_n \xrightarrow{\text{по вер.}} \beta$ состоит в том, что

$$P(|\alpha_n - \beta| > \varepsilon) \xrightarrow[n\to\infty]{} 0 \text{ при любом } \varepsilon > 0.]$$

Доказательство. В силу попарной независимости величин $\xi_1, \xi_2, \ldots, \xi_n$ и условия $D\xi_k \leqslant C$ имеем

$$D\zeta_n = D\left(\frac{1}{n}\sum_{k=1}^{n} \xi_k\right) = \frac{1}{n^2} D\left(\sum_{k=1}^{n} \xi_k\right) = \frac{1}{n^2}\sum_{k=1}^{n} D\xi_k \leqslant$$

$$\leqslant \frac{1}{n^2} nC = \frac{C}{n}$$

(см. свойства дисперсии, приведенные в гл. IV).

Применяя к величине ζ_n неравенство Чебышева (6.3) и учитывая оценку, полученную для $D\zeta_n$, имеем

$$P(|\zeta_n - M\zeta_n| > \varepsilon) < \frac{D\zeta_n}{\varepsilon^2} < \frac{C}{n\varepsilon^2}.$$

Замечая, что при любом фиксированном $\varepsilon > 0$ величина $\frac{C}{n\varepsilon^2} \to 0$ при $n \to \infty$, из последнего неравенства получаем:

$$P(|\zeta_n - M\zeta_n| > \varepsilon) \to 0 \quad \text{при} \quad n \to \infty,$$

что и означает сходимость последовательности $(\zeta_n - M\zeta_n)$ к нулю по вероятности при $n \to \infty$.

Теорема 6.4 (частный случай теоремы Чебышева). *Пусть* $\xi_1, \xi_2, \ldots, \xi_n$ — *независимые, одинаково распределенные случайные величины, имеющие конечные математическое ожидание* $M\xi_k = a$ *и дисперсию* $D\xi_k = \sigma^2$. *Положим* $\zeta_n = \frac{1}{n}\sum_{k=1}^{n} \xi_k$. *Тогда последовательность* ζ_n *стремится по* *вероятности к a при $n \to \infty$*, т. е.

$$\zeta_n \xrightarrow{\text{по вер.}} a \text{ при } n \to \infty.$$

В самом деле, все условия теоремы 6.3 выполнены, $M\zeta_n = M\left(\frac{1}{n}\sum_{k=1}^{n} \xi_k\right) = \frac{1}{n} na = a$, поэтому последовательность $(\zeta_n - a) \xrightarrow{\text{по вер.}} 0$ при $n \to \infty$. Но последнее означает, что $\zeta_n \xrightarrow{\text{по вер.}} a$ при $n \to \infty$.

Теорема 6.5 (закон больших чисел Бернулли). *Пусть ν_n — число успехов в n испытаниях Бернулли с вероятностью успеха p. Тогда частота $\frac{\nu_n}{n}$ успеха в n испытаниях стремится по вероятности к вероятности p* успеха, когда $n \to \infty$, т. е.

$$\frac{\nu_n}{n} \xrightarrow{\text{по вер.}} p \text{ при } n \to \infty.$$

Доказательство. Пусть ξ_k — случайная величина, равная единице или нулю в зависимости от того, произошел при k-м испытании успех или была неудача. Тогда, очевидно,

$$\nu_n = \sum_{k=1}^{n} \xi_k, \quad \frac{\nu_n}{n} = \frac{1}{n}\sum_{k=1}^{n} \xi_k = \zeta_n, \quad M\xi_k = p, \quad D\xi_k = pq,$$

$$M\zeta_n = M\left(\frac{1}{n}\sum_{k=1}^{n} \xi_k\right) = p.$$

Величины $\xi_1, \xi_2, \ldots, \xi_n$ — независимые, одинаково распределенные случайные величины, имеют конечные математическое ожидание и дисперсию. Таким образом, все условия теоремы 6.4 оказываются выполненными, и в силу этой теоремы имеем

$$\zeta_n = \frac{\nu_n}{n} \xrightarrow{\text{по вер.}} p \text{ при } n \to \infty,$$

что и требовалось доказать.

Таким образом, закон больших чисел Бернулли (теорема 6.5) оказался частным случаем теоремы 6.4, а следовательно, и закона больших чисел Чебышева (теоремы 6.3). Имеет место следующая теорема Хинчина.

Теорема 6.6 (закон больших чисел Хинчина). *Пусть* $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — *независимые, одинаково распределенные случайные величины, имеющие конечное математическое ожидание* $M\xi_k = a$. *Положим* $\zeta_n = \frac{1}{n}\sum_{k=1}^{n}\xi_k$. *Тогда последовательность* ζ_n *стремится по вероятности к* a *при* $n \to \infty$, *т. е.*

$$\zeta_n \xrightarrow{\text{по вер.}} a \text{ при } n \to \infty.$$

Доказательство этой теоремы приведено в гл. VII. Заметим, что теорема 6.6 сходна с теоремой 6.4, но сильнее ее, так как в теореме 6.4 требуется дополнительно существование $D\xi_k$. Таким образом как из теоремы 6.3, так и из теоремы 6.6 вытекают теоремы 6.4 и 6.5. Теорема 6.5 является частным случаем теоремы 6.4, а теорема 6.4 — частным случаем теоремы 6.3 и частным (притом ослабленным) случаем теоремы 6.6.

Закон больших чисел имеет большое практическое значение, так как он позволяет придать точный смысл приближенным равенствам (6.1) и (6.2). Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — результаты n независимых измерений некоторой постоянной величины a. Измерения сопряжены со случайными ошибками, поэтому результаты измерений являются случайными величинами. Эти величины независимы и одинаково распределены. Предположим, что $M\xi_k = a$ (это гипотеза отсутствия систематической ошибки) и что $D\xi_k = \sigma^2 < \infty$ (теорема 6.6 делает это последнее требование излишним). Тогда к результатам измерений можно применить теорему 6.4 (или теорему 6.6), согласно которой при любом фиксированном $\varepsilon > 0$

$$P(|\zeta_n - a| > \varepsilon) \to 0 \quad \text{при} \quad n \to \infty,$$

или

$$P(|\zeta_n - a| < \varepsilon) \to 1 \quad \text{при} \quad n \to \infty.$$

Таким образом, каково бы ни было $\varepsilon > 0$, при достаточно большом n вероятность неравенства $|\zeta_n - a| < \varepsilon$ как угодно близка к единице, т. е. событие $\{|\zeta_n - a| < \varepsilon\}$ можно считать при достаточно больших n практически достоверным. Именно в этом смысле и следует понимать приближенное равенство (6.2).

Совершенно аналогично обстоит дело с приближенным равенством (6.1). Его надо понимать в том смысле (см.

теорему 6.5), что при любом $\varepsilon > 0$ событие $\left|\frac{\nu_n}{n} - p\right| < \varepsilon$ можно считать практически достоверным при достаточно больших n.

§ 4. УСИЛЕННЫЙ ЗАКОН БОЛЬШИХ ЧИСЕЛ

Закон больших чисел Бернулли устанавливает, что частота $\frac{\nu_n}{n}$ успеха в n испытаниях Бернулли с вероятностью успеха p стремится по вероятности к p при $n \to \infty$, т. е. $\frac{\nu_n}{n} \xrightarrow{\text{по вер.}} p$ при $n \to \infty$. Из сходимости по вероятности не следует , как известно, сходимость с вероятностью 1, поэтому из теоремы Бернулли нельзя сделать вывод о том, что $P\left(\frac{\nu_n}{n} \underset{n\to\infty}{\longrightarrow} p\right) = 1$, т. е. нельзя считать, что $\lim\limits_{n\to\infty} \frac{\nu_n}{n} = p$ с вероятностью 1. Сходимость по вероятности некоторой последовательности ξ_k к величине ξ при $n \to \infty$ не исключает даже того случая, когда $\lim\limits_{n\to\infty} \xi_n(\omega)$ не существует ни при каком ω. Однако в случае закона больших чисел дело обстоит иначе, а именно: теоремы 6.4, 6.5 и 6.6 остаются справедливыми, если заменить в них сходимость по вероятности сходимостью с вероятностью 1.

Полученные таким образом теоремы являются усилением теорем 6.4, 6.5 и 6.6, так как из этих теорем следуют теоремы 6.4, 6.5 и 6.6. В самом деле, как известно, из сходимости почти наверное вытекает сходимость по вероятности; обратное же утверждение неверно. Под *усиленным законом больших чисел* понимают группу теорем, являющихся усилением (в указанном смысле) теорем, носящих общее название закона больших чисел. Приведем без доказательства следующие теоремы.

Теорема 6.7. (усиленный закон больших чисел Колмогорова). *Пусть $\xi_1, \xi_2, \ldots$ — независимые, одинаково распределенные случайные величины, имеющие конечное математическое ожидание $M\xi_k = a$. Положим $\zeta_n = \frac{1}{n}\sum\limits_{k1}^{n} \xi_k$. Тогда последовательность ζ_n стремится к a с вероятностью 1, т. в.*

$$P(\zeta_n \to a \text{ при } n \to \infty) = 1.$$

Приведенная теорема является усилением теоремы 69, (теоремы Хинчина). Таким образом, теорема Хинчина является следствием приведенной теоремы, а потому и теоремы 6.4 и 6.5, как следствия теоремы Хинчина, также являются следствиями приведенной теоремы Колмогорова.

Теорема 6.8 (частный случай теоремы Колмогорова). *Пусть ν_n число успехов в n испытаниях Бернулли с вероятностью успеха p. Тогда частота $\frac{\nu_n}{n}$ успеха в n испытаниях стремится к вероятности p успеха с вероятностью* 1, *когда $n \to \infty$*, т. е.

$$P\left(\frac{\nu_n}{n} \xrightarrow[n\to\infty]{} p\right) = 1.$$

Теорема 6.8 выводится из теоремы 6.7 точно так же, как выводится теорема 6.5 из теоремы 6.4. Теорема 6.8 является усилением теоремы 6.5.

Заметим, что в настоящей главе предполагалось, что математическое ожидание $M\xi_i$, $(i = 1, 2, \ldots, n)$ существует и конечно. Можно, однако, отказаться от этого условия и рассмотреть последовательность независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ таких, что $M\xi_i = +\infty$ $(i = 1, 2, \ldots, n, \ldots)$. Тогда усиленный закон больших чисел Колмогорова (теорема 6.7) оказывается верным в следующей формулировке.

Теорема 6.9. *Пусть $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — независимые одинаково распределенные случайные величины, имеющие бесконечное математическое ожидание, т. е. $M\xi_i = +\infty$ $(i = 1, 2, \ldots, n)$. Положим $\zeta_n = \frac{1}{n}\sum_{k=1}^{n}\xi_k$. Тогда последовательность ζ_n стремится к ∞ с вероятностью* 1, *когда $n \to \infty$ т. е.*

$$P(\zeta_n \xrightarrow[n\to\infty]{} \infty) = 1.$$

Из приведенной теоремы вытекает сходимость последовательности ζ_n к ∞ при $n \to \infty$ также и по вероятности, т. е. закон больших чисел Хинчина (теорема 6.6) оказывается верным, если отказаться от требования $M\xi_i = a < \infty$ и считать $a = M\xi_i = +\infty$.

ГЛАВА VII

ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

§ 1. ПОСТАНОВКА ЗАДАЧИ

В предыдущей главе приведены теоремы, устанавливающие сходимость последовательностей случайных величин и носящие общее название закона больших чисел. В настоящей главе будет рассмотрена другая группа предельных теорем теории вероятностей, устанавливающих *сходимость законов распределения* для последовательностей сумм случайных величин. Эта группа теорем (ввиду ее особой важности как для теории, так и для приложений) носит название *центральной предельной теоремы.* С одной из теорем этой группы, именно с теоремой Муавра—Лапласа, мы уже знакомы (см. гл. II). Теорема Муавра — Лапласа является частным случаем гораздо более общих теорем, рассмотрению которых и посвящена настоящая глава.

Предварительно несколько преобразуем формулировку теоремы Муавра — Лапласа. Пусть ν_n — число успехов в n испытаниях Бернулли с вероятностью успеха p. Теорема Муавра — Лапласа утверждает, что при любом x

$$\mathrm{P}\{\nu_n^* < x\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt \quad \text{при } n \to \infty, \tag{7.1}$$

где $\nu_n^* = \dfrac{\nu_n - np}{\sqrt{npq}}$ — так называемое нормированное число успехов в n испытаниях Бернулли с вероятностью успеха p. Как известно (см. гл. II),

$$M\nu_n = np, \quad D\nu_n = npq, \quad \sigma_{\nu_n} = \sqrt{D\nu_n} = \sqrt{npq}.$$

Таким образом,

$$\nu_n^* = \frac{\nu_n - M\nu_n}{\sqrt{D\nu_n}}.$$

Пусть ξ_k — случайная величина, равная 1 или 0 в зависимости от того, произошел при k-м испытании успех или была неудача. Тогда имеем

$$\nu_n = \xi_1 + \xi_2 + \cdots + \xi_n,$$

и теорема Муавра — Лапласа может быть записана в виде

$$P\left(\frac{\sum\limits_{k=1}^{n} \xi_k - M\sum\limits_{k=1}^{n} \xi_k}{\sqrt{D\sum\limits_{k=1}^{n} \xi_k}} < x\right) \to \frac{1}{2\pi}\int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt = F(x)$$

$$\text{при } n \to \infty, \tag{7.2}$$

где $F(x)$ — функция распределения нормального закона с параметрами 0 и 1. Встает вопрос, справедливо ли соотношение (7.2) лишь для независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, принимающих только два значения: нуль и единица с вероятностями $q = 1 - p$ и p соответственно, или соотношение (7.2) остается справедливым и для случайных величин более общего вида. Положительный ответ на этот вопрос дает в различных условиях группа теорем, объединенных общим названием *центральная предельная теорема*. Точнее, эта группа теорем носит название *интегральной центральной предельной теоремы*, так как в этих теоремах идет речь о предельном поведении функции распределения (*интегральном законе распределения*).

Под локальной центральной предельной теоремой понимают группу теорем, устанавливающих асимптотическое поведение при $n \to \infty$ вероятностей $P\left\{\sum\limits_{k=1}^{n} \xi_k = j\right\}$ для дискретных независимых случайных величин ξ_k и плотностей сумм для непрерывных величин. Примером теорем этой группы служит локальная теорема Муавра — Лапласа.

В настоящей главе будет рассмотрена только интегральная предельная теорема. Предварительно познакомимся с характеристическими функциями.

§ 2. ХАРАКТЕРИСТИЧЕСКИЕ ФУНКЦИИ

Напомним определение характеристической функции случайной величины, данное в гл. IV.

Характеристической функцией $\varphi_\xi(t)$ *случайной величи-*

ны ξ называется математическое ожидание случайной величины $e^{it\xi}$, т. е.

$$\varphi_\xi(t) = Me^{it\xi} = M\cos t\xi + iM\sin t\xi.$$

Согласно свойству математического ожидания характеристическая функция вычисляется по формуле

$$\varphi_\xi(t) = \int e^{itx}\,dF(x). \tag{7.3}$$

В дискретном случае и в случае существования плотности распределения приходим соответственно к формулам

$$\varphi_\xi(t) = \sum_k e^{itx_k} p_k$$

и

$$\varphi_\xi(t) = \int_{-\infty}^{+\infty} e^{itx} p_\xi(x)\,dx.$$

Последняя из приведенных формул означает, что характеристическая функция есть преобразование Фурье плотности распределения вероятностей $p_\xi(x)$.

Характеристическая функция определена для всех значений t и удовлетворяет условию $|\varphi_\xi(t)| \leqslant 1$, какова бы ни была случайная величина ξ. В самом деле, учитывая, что $|e^{itx}| = 1$, имеем

$$\left|\int e^{itx}\,dF(x)\right| \leqslant \int |e^{itx}|\,dF(x) = \int dF(x) = 1.$$

Приведем важнейшие свойства характеристической функции.

Свойство 1. *Если $\eta = \sigma\xi + a$, где a и σ — постоянные, то*

$$\varphi_\xi(t) = e^{iat}\varphi_\xi(\sigma t).$$

В самом деле, имеем

$$\varphi_\eta(t) = Me^{it\eta} = Me^{it(\sigma\xi+a)} = e^{ita}Me^{it\sigma\xi} = e^{ita}\varphi_\xi(\sigma t).$$

Свойство 2. *Характеристическая функция суммы n независимых случайных величин равна произведению* *их характеристических функций, т. е.*

$$\varphi_{\xi_1+\xi_2+\ldots+\xi_n}(t)=\varphi_{\xi_1}(t)\,\varphi_{\xi_2}(t)\ldots\varphi_{\xi_n}(t),$$

если $\xi_1, \xi_2, \ldots, \xi_n$ *— независимые случайные величины.*

В самом деле, применяя теорему о математическом ожидании произведения независимых случайных величин, имеем

$$\varphi_{\xi_1+\xi_2+\ldots+\xi_n}(t)=Me^{it(\xi_1+\xi_2+\ldots+\xi_n)}=Me^{it\xi_1}\cdot e^{it\xi_2}\ldots e^{it\xi_n}=$$

$$=Me^{it\xi_1}\cdot Me^{it\xi_2}\ldots Me^{it\xi_n}=\varphi_{\xi_1}(t)\cdot\varphi_{\xi_2}(t)\ldots\varphi_{\xi_n}(t).$$

Свойство 3. *Если существует абсолютный момент n-го порядка $M|\xi|^n$, то характеристическая функция $\varphi_\xi(t)$ дифференцируема n раз при любом t и*

$$\varphi_\xi^{(l)}(0)=i^l M\xi^l$$

при $l=1, 2, \ldots, n$. В самом деле, формальное l-кратное дифференцирование равенства (7.3) дает

$$\varphi_\xi^{(l)}(t)=i^l\int_{-\infty}^{+\infty}x^l e^{itx}\,dF(x). \tag{7.4}$$

Но $|x^l e^{itx}|=|x^l|$. Поэтому интеграл в выражении (7.4) мажорируется не зависящим от t интегралом. При этом в силу существования $M|\xi|^n$ мажорирующий интеграл является сходящимся. (Существование $M|\xi|^n$ гарантирует существование $M|\xi|^l$ при $l<n$.) Отсюда следует сходимость интеграла, входящего в равенство (7.4), и законность проведенного формально дифференцирования. Таким образом, равенство (7.4) справедливо для любых t и $l=1, 2, \ldots, n$. Положив в равенстве (7.4) $t=0$, имеем

$$\varphi_\xi^{(l)}(0)=i^l\int_{-\infty}^{+\infty}x^l\,dF(x)=i^l M\xi^l.$$

Свойство 3 позволяет весьма просто вычислять моменты случайной величины, если известна ее характеристическая функция.

Свойство 4. *Если существует абсолютный момент n-го порядка $M|\xi|^n$, то характеристическая функция $\varphi_\xi(t)$ допускает разложение по формуле Тейлора:*

$$\varphi_\xi(t)=1+m_1it+m_2\frac{(it)^2}{2!}+\ldots+m_n\frac{(it)^n}{n!}+o(t^n),$$

где

$$m_k = M\,\xi^k,\ k = 1, 2, \ldots, n.$$

В самом деле, в силу свойства 3, функция $\varphi_\xi(t)$ n раз дифференцируема при любом t, поэтому она допускает разложение по формуле Тейлора с остаточным членом вида $o\,(t^n)$. Вид коэффициентов в формуле Тейлора также непосредственно следует из свойства 3. Первый член равен единице, так как $\varphi_\xi(0) = 1$.

Другие свойства характеристических функций приведем без доказательства*).

Свойство 5. **Теорема 7.1. (теорема единственности).** *Функция распределения однозначно определяется своей характеристической функцией.*

В случае существования плотности распределения вероятностей теорема 7.1 утверждает, что оригинал (т. е. плотность распределения) может быть однозначно восстановлен по его преобразованию Фурье.

В силу приведенной теоремы между функциями распределения и характеристическими функциями существует взаимно-однозначное соответствие: каждой функции распределения ставится в соответствие характеристическая функция, и, обратно, каждой характеристической функции ставится в соответствие ровно одна функция распределения. Формулы, выражающие функцию распределения через характеристическую функцию, нам не понадобятся; в дальнейшем будет использован тот факт, что из совпадения характеристических функций следует совпадение функций распределения.

Прежде чем формулировать очередное свойство, напомним следующее определение.

Определение. *Последовательность функций распределения $F_n(x)$ называется сходящейся в основном* (слабо сходящейся) *к функции распределения $F(x)$ при $n \to \infty$, если*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

в каждой точке x, где $F(x)$ непрерывна.

Таким образом, если $F(x)$ непрерывна всюду, то сходимость в основном совпадает со сходимостью в каждой точке.

Свойство 6. **Теорема 7.2 (предельная теорема).**

---

*) Доказательство можно найти в кн.: Б. В. Гнеденко. Курс теории вероятностей. М., «Наука», 1969.

Пусть $F_n(x)$, $n = 1, 2, \ldots$ и $F(x)$ — функции распределения, $\varphi_n(t)$ и $\varphi(t)$ — соответствующие характеристические функции. Тогда для сходимости в основном последовательности $F_n(x)$ функций распределения к функции распределения $F(x)$ необходимо и достаточно, чтобы последовательность $\varphi_n(t)$ характеристических функций сходилась при каждом t к характеристической функции $\varphi(t)$.

Основываясь на теоремах 7.1 и 7.2, можно изучать распределения вероятностей, включая их предельное поведение, не только с помощью функций распределения, но и с помощью характеристических функций.

Докажем с помощью этой теоремы закон больших чисел в форме Хинчина. Пусть ξ_i — независимые случайные величины с общей характеристической функцией $\varphi(t)$. Тогда характеристическая функция случайной величины $\xi_n = \frac{1}{n}(\xi_1 + \ldots + \xi_n)$ равна $\varphi^n\left(\frac{t}{n}\right)$. Если предположить, что $M\xi_n = a$, $|a| < \infty$, то

$$\varphi(u) = 1 + iua + o(u)(u \to 0).$$

Следовательно,

$$\varphi\left(\frac{t}{n}\right) = 1 + \frac{ita}{n} + o\left(\frac{1}{n}\right),$$

откуда

$$\varphi^n\left(\frac{t}{n}\right) = \left[1 + \frac{ita}{n} + o\left(\frac{1}{n}\right)\right]^n \underset{n\to\infty}{\to} e^{ita}.$$

С другой стороны, характеристическая функция величины, с вероятностью 1 равной постоянной a, составляет e^{ita}. Отсюда следует, что ζ_n сходится к a по вероятности.

Это утверждение и составляет содержание теоремы Хинчина.

Рассмотрим несколько примеров, представляющих самостоятельный интерес.

Пример 1. Пусть случайная величина ξ нормально распределена с параметрами a и σ, т. е.

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}},$$

тогда $\varphi_\xi(t) = e^{iat - \frac{\sigma^2 t^2}{2}}$.

При $a = 0$, $\sigma = 1$ имеем

$$\varphi_\xi(t) = e^{-\frac{t^2}{2}}.$$

(Этот результат уже был получен в гл. IV.)

Пример 2. Пусть случайная величина ξ распределена по биномиальному закону

$$P(\xi = k) = C_n^k p^k (1-p)^{n-k},$$

т. е. ξ можно рассматривать как число ν_n успехов в n испытаниях Бернулли с вероятностью успеха p. Величину ν_n можно представить так:

$$\nu_n = \xi_1 + \xi_2 + \ldots + \xi_n,$$

где ξ_k — число успехов в k-м испытании. Величины ξ_k, $k = 1, 2, \ldots, n$, независимы и одинаково распределены. Закон распределения каждой из величин ξ_k можно задать с помощью таблицы $\begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$, где $q = 1 - p$, поэтому $\varphi_{\xi_k}(t) = pe^{it} + q$. Согласно свойству 2 характеристических функций имеем

$$\varphi_\xi(t) = \varphi_{\mu_n}(t) = (pe^{it} + q)^n.$$

Характеристическая функция величины $\nu_n^* = \dfrac{\nu_n - np}{\sqrt{npq}}$, согласно свойству 1 характеристических функций, равна

$$\varphi_{\nu_n^*}(t) = e^{-it\sqrt{\frac{np}{q}}}\left(pe^{i\frac{t}{\sqrt{npq}}} + q\right)^n =$$

$$= \left(pe^{it\sqrt{\frac{q}{np}}} + qe^{-it\sqrt{\frac{p}{nq}}}\right)^n.$$

Пример 3. Пусть случайная величина ξ распределена по закону Пуассона с параметром λ, т. е.

$$\mathrm{P}(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots .$$

Имеем

$$\varphi_\xi(t) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{-\lambda} \cdot e^{\lambda e^{it}} = e^{\lambda(e^{it} - 1)}$$

Пример 4. Пусть случайная величина ξ равномерно распределена на отрезке $[-l, l]$, т. е. плотность распределения

$$p_\xi(x) = \begin{cases} \dfrac{1}{2l} & \text{при } -l \leqslant x \leqslant l, \\ 0 & \text{при } \quad x < -l \quad \text{или } x > l. \end{cases}$$

Имеем

$$\varphi_\xi(t) = \frac{1}{2l} \int_{-l}^{l} e^{itx}\, dx = \frac{\sin lt}{lt}.$$

§ 3. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — независимые одинаково распределенные случайные величины, имеющие конечное математическое ожидание и дисперсию:

$$M\,\xi_k = a,\ D\,\xi_k = \sigma^2,\ k = 1, 2, \ldots, n.$$

Пусть $\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n$, тогда

$$M\zeta_n = na,\ D\zeta_n = n\sigma^2.$$

Положим

$$\zeta_n^* = \frac{\zeta_n - M\,\zeta_n}{\sqrt{D\,\zeta_n}} = \frac{\zeta_n - na}{\sigma\sqrt{n}};$$

величину ζ_n^* будем называть *нормированной суммой* величин $\xi_1, \xi_2, \ldots, \xi_n$. Вообще, если для некоторой случайной величины ξ существуют конечные $M\xi$ и $D\xi$, то величину

$$\xi^* = \frac{\xi - M\,\xi}{\sqrt{D\xi}}$$

будем называть *нормированной величиной* ξ и обозначать той же буквой, но со звездочкой. Наша терминология и обозначения согласуются с принятыми ранее для числа ν_n успехов в n испытаниях Бернулли, так как

$$M\,\nu_n = np,\quad D\,\nu_n = npq,\quad \nu_n^* = \frac{\nu_n - np}{\sqrt{npq}}.$$

Нетрудно проверить, что $M\xi^* = 0$, $D\xi^* = 1$. В самом деле,

$$M\,\xi^* = M\frac{\xi - M\,\xi}{\sqrt{D\xi}} = \frac{1}{\sqrt{D\xi}}\,M(\xi - M\,\xi) = \frac{1}{\sqrt{D\xi}}\,(M\xi - M\xi) = 0,$$

$$D\,\xi^* = D\left(\frac{\xi - M\,\xi}{\sqrt{D\xi}}\right) = \frac{1}{D\,\xi}\cdot D\,(\xi - M\,\xi) = \frac{1}{D\,\xi}\cdot D\,\xi = 1.$$

Таким образом, для нормированной суммы ζ_n^* имеем

$$M\,\zeta_n^* = 0,\ D\zeta_n^* = 1.$$

Оказывается (в этом и состоит содержание центральной предельной теоремы для этого случая), что закон распределения нормированной суммы ζ_n^* стремится при $n \to \infty$ к нормальному закону с теми же параметрами 0 и 1. Таким образом, закон распределения нормированной суммы ζ_n^* n независимых одинаково распределенных случайных величин, имеющих конечные математическое ожидание и дисперсию, стремится к нормальному закону при $n \to \infty$, каков бы ни был закон распределения слагаемых. Это замечательное свойство нормального закона лежит в основе его многочисленных применений и делает понятной ту основную роль, которую он играет в теории вероятностей.

Теорема 7.3 (теорема Линдеберга—Леви). *Пусть* ξ_1, ξ_2, $\ldots$, $\xi_n, \ldots$ — *независимые одинаково распределенные случайные величины, имеющие конечные математическое ожидание* $M\xi_k = a$ *и дисперсию* $D\xi_k = \sigma^2$. *Положим* $\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n$. *Тогда закон распределения нормированной суммы* ζ_n^* *этих величин стремится при* $n \to \infty$ *к нормальному закону с параметрами* 0 *и* 1, *т. е.*

$$P(\zeta_n^* < x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \text{ при } n \to \infty.$$

Доказательство. В силу предельной теоремы (теоремы 7.2) и теоремы единственности (теоремы 7.1) о характеристических функциях для доказательства теоремы достаточно установить, что характеристическая функция $\varphi_{\zeta_n^*}(t)$ нормированной суммы ζ_n^* стремится при $n \to \infty$ к характеристической функции нормального закона с параметрами 0 и 1, т. е. к функции $\psi(t) = e^{-\frac{t^2}{2}}$ (см. пример 1 предыдущего параграфа). Поэтому доказательство теоремы будет состоять в нахождении предела характеристической функции $\varphi_{\zeta_n^*}(t)$ при $n \to \infty$.

Рассмотрим величину

$$\zeta_n - M\zeta_n = (\xi_1 - M\xi_1) + (\xi_2 - M\xi_2) + \ldots + (\xi_n - M\xi_n) = \\ = \eta_1 + \eta_2 + \ldots + \eta_n,$$

где $\eta_k = \xi_k - M\xi_k$, $k = 1, 2, \ldots, n$. Имеем

$$M\eta_k = M(\xi_k - M\xi_k) = 0, \quad D\eta_k = M\eta_k^2 = D\xi_k = \sigma^2.$$

Положим $\varphi(t) \doteq \varphi_{\eta_k}(t)$. Тогда, согласно свойству 4 характеристических функций, имеем

$$\varphi(t)=1+\frac{(it)^2}{2!}\sigma^2+o(t^2).$$

Для суммы $\zeta_n - M\zeta_n$ независимых слагаемых, согласно свойству 2, имеем

$$\varphi_{\zeta_n - M\zeta_n}(t)=[\varphi(t)]^n.$$

Согласно свойству 1 характеристических функций для величины

$\zeta_n^* = \frac{\zeta_n - M\zeta_n}{\sqrt{D\zeta_n}}$, учитывая, что $D\zeta_n = n\sigma^2$, получим

$$\varphi_{\zeta_n^*}(t)=\varphi_{\zeta_n - M\zeta_n}\left(\frac{t}{\sigma\sqrt{n}}\right)=\left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n=$$

$$=\left[1-\frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2+o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)\right]^n=$$

$$=\left[1-\frac{t^2}{2n}+o\left(\frac{t^2}{\sigma^2 n}\right)\right]^n.$$

Поэтому

$$\lim_{n\to\infty}\varphi_{\zeta_n^*}(t)=\lim_{n\to\infty}\left(1-\frac{t^2}{2n}+o\left(\frac{t^2}{\sigma^2 n}\right)\right)^n=$$

$$=e^{\lim\limits_{n\to\infty} n\left(-\frac{t^2}{2n}+o\left(\frac{t^2}{\sigma^2 n}\right)\right)}=e^{-\frac{t^2}{2}}.$$

Теорема доказана.

Из доказанной теоремы получаем приближенную формулу для функции распределения нормированной суммы независимых одинаково распределенных величин, справедливую при достаточно больших n:

$$P(\zeta_n^* < x)\approx\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt = F(x). \tag{7.5}$$

Через $F(x)$ здесь и в дальнейшем обозначается нормальная функция распределения с параметрами 0 и 1.

На практике число n, начиная с которого эта формула дает удовлетворительное приближение, весьма невелико (порядка нескольких десятков, а иногда даже и десятка).

Формулу (7.5) можно несколько преобразовать:

$$P(\alpha<\zeta_n^*<\beta)\approx\frac{1}{\sqrt{2\pi}}\int\limits_\alpha^\beta e^{-\frac{t^2}{2}}dt=F(\beta)-F(\alpha). \qquad (7.6)$$

Вероятность того, что сумма ζ_n независимых одинаково распределенных слагаемых лежит в заданных пределах, равна

$$P(A<\zeta_n<B)=P\left(\frac{A-na}{\sigma\sqrt{n}}<\frac{\zeta_n-na}{\sigma\sqrt{n}}<\frac{B-na}{\sigma\sqrt{n}}\right)=$$

$$=P\left(\frac{A-na}{\sigma\sqrt{n}}<\zeta_n^*<\frac{B-na}{\sigma\sqrt{n}}\right).$$

Отсюда и из формулы (7.6) получаем следующую приближенную формулу:

$$P(A<\zeta_n<B)\approx F\left(\frac{B-na}{\sigma\sqrt{n}}\right)-F\left(\frac{A-na}{\sigma\sqrt{n}}\right). \qquad (7.7)$$

Рассмотрим на примере использование доказанной теоремы.

Пример. На отрезке [0,1] случайным образом выберем $n=108$ чисел, точнее, рассмотрим $n=108$ независимых, равномерно распределенных на отрезке [0,1] случайных величин $\xi_1, \xi_2, \dots, \xi_n$. Найти вероятность того, что их сумма $\zeta_n=\sum\limits_{k=1}^n\xi_k$ заключена между 50 и 60, т. е. $P\{50<\zeta_n<60\}$.

Как нетрудно подсчитать, $M\xi_k=\frac{1}{2}$, $D\xi_k=\frac{1}{12}$, следовательно,

$$M\zeta_{108}=na=54,\ D\zeta_{108}=n\sigma^2=9,\ \sigma_{\zeta_{108}}=\sigma\sqrt{n}=3.$$

По формуле (7.7) имеем

$$P\{50<\zeta_{108}<60\}\approx F\left(\frac{60-54}{3}\right)-F\left(\frac{50-54}{3}\right)=$$

$$=F(2)-F\left(-\frac{4}{3}\right)=0{,}885.$$

(Последнее значение находим из таблиц.)

В доказанной теореме величины $\xi_1, \xi_2, \dots, \xi_n$ предполагаются одинаково распределенными. Встает вопрос о том, может ли эта теорема быть обобщена на случай различно распределенных слагаемых. Положительный ответ на этот вопрос дает *теорема Ляпунова*, согласно которой закон распределения нормированной суммы ζ_n^* независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ (вообще говоря, различно распределенных) стремится к нормальному закону при $n \to \infty$, если, грубо говоря, роль каждой из величин $\xi_1, \xi_2, \ldots, \xi_n$ в образовании суммы $\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n$ мала. Необходимость ограничения подобного рода станет понятной, если рассмотреть сумму n случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, первая из которых имеет какое-либо распределение, отличное от нормального, например равномерное на отрезке $[0, 1]$, а остальные величины равны нулю с вероятностью 1. Тогда сумма $\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n = \xi_1$ имеет при любом n равномерное распределение и закон распределения суммы при $n \to \infty$ не стремится к нормальному. Если же подобные случаи не рассматривать, то закон распределения нормированной суммы n независимых случайных величин должен быть близок к нормальному при больших n.

Теорема 7.4 (теорема Ляпунова). *Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — последовательность независимых случайных величин, и пусть существуют конечные $M\xi_k = a_k$, $D\xi_k = b_k^2$, $M|\xi_k - a_k|^3 = c_k^3$ для всех k. Положим*

$$\zeta_n = \xi_1 + \xi_2 + \ldots + \xi_n, \quad A_n = \sum_{k=1}^{n} a_k, \quad B_n^2 = \sum_{k=1}^{n} b_k^2,$$

$$C_n^3 = \sum_{k=1}^{n} c_k^3.$$

Пусть далее $\lim\limits_{n \to \infty} \dfrac{C_n}{B_n} = 0$. *Тогда*

$$P(\zeta_n^* < x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \quad \text{при} \quad n \to \infty.$$

Для доказательства теоремы используем следующую лемму.

Л е м м а. Для любой случайной величины ξ, имеющей третий абсолютный центральный момент $c^3 = M|\xi - M\xi|^3$, справедливо неравенство

$$\sqrt[3]{M|\xi-M\xi|^3} \geqslant \sqrt{M(\xi-M\xi)^2} \quad \text{или} \quad c \geqslant \sigma_\xi.$$

Доказательство леммы. Без ограничения общности предположим, что $M\xi = 0$. Рассмотрим величину

$$\int (1+u|x|)^2 dF(x) = 1 + 2ua + u^2\sigma_\xi^2,$$

где $a = M|\xi| = \int\limits_{-\infty}^{+\infty} |x|\,dF(x)$. Эта величина больше или равна нулю при любом u, поэтому дискриминант

$$a^2 - \sigma_\xi^2 \leqslant 0, \quad \text{или} \quad a^2 \leqslant \sigma_\xi^2. \tag{7.8}$$

Аналогично, рассматривая величину

$$\int \left(|x|^{\frac{1}{2}} + u|x|^{\frac{3}{2}}\right)^2 dF(x) = a + 2u\,\sigma_\xi^2 + u^2c^3 \geqslant 0,$$

получаем

$$\sigma_\xi^4 - ac^3 \leqslant 0, \quad \text{или} \quad \sigma_\xi^8 \leqslant a^2c^6. \tag{7.9}$$

Перемножая неравенства (7.8) и (7.9) почленно, имеем

$$\sigma_\xi^6 \leqslant c^6, \quad \text{или} \quad c \geqslant \sigma_\xi,$$

что и требовалось доказать.

Доказательство теоремы. Как и при доказательстве центральной предельной теоремы для одинаково распределенных слагаемых, достаточно доказать, что характеристическая функция $\varphi_{\zeta_n^*}(t)$ нормированной суммы ζ_n^* стремится при $n \to \infty$ к функции $\psi(t) = e^{-\frac{t^2}{2}}$. Полагая $\eta_k = \xi_k - M\xi_k = \xi_k - a_k$, $k = 1, 2, \ldots, n$, имеем

$$M\eta_k = 0, \quad D\eta_k = b_k^2 \quad M|\eta_k|^3 = c_k^3.$$

Обозначим через $\varphi_k(t)$ характеристическую функцию величины η_k:

$$\varphi_k(t) = Me^{it\eta_k}.$$

Согласно свойству 3 характеристических функций можем утверждать, что функция $\varphi_k(t)$ трижды дифференцируема при любом t. Поэтому $\varphi_k(t)$ можно разложить в ряд Тейлора в точке $t=0$ (см. свойство 4 характеристических функций):

$$\varphi_k(t) = 1 - \frac{b_k^2}{2} t^2 + \frac{\varphi_k^{(3)}(\theta)}{3!} t^3 \quad (0 < \theta < t). \qquad (7.10)$$

Величину $\varphi_k^{(3)}(\theta)$, учитывая формулу (7.4), можно оценить следующим образом:

$$\left|\varphi_k^{(3)}(\theta)\right| = \left| i^3 \int_{-\infty}^{+\infty} x^3 e^{i\theta x}\, dF(x) \right| \leqslant \int_{-\infty}^{+\infty} \left| x^3 e^{i\theta x} \right| dF(x) =$$

$$= \int_{-\infty}^{+\infty} |x^3|\, dF(x) = c_k^3.$$

В силу полученной оценки можем положить

$$\varphi_k^{(3)}(\theta)\, t^3 = c_k^3 \varkappa_k |t|^3 \quad (|\varkappa_k| \leqslant 1),$$

тогда формула (7.10) принимает вид

$$\varphi_k(t) = 1 - \frac{b_k^2}{2!} t^2 + \frac{c_k^3}{3!} |t|^3 \varkappa_k. \qquad (7.11)$$

Для характеристической функции $\varphi_{\zeta_n^*}(t)$ нормированной суммы

$$\zeta_n^* = \frac{\sum_{k=1}^n \xi_k - M \sum_{k=1}^n \xi_k}{\sqrt{D \sum_{k=1}^n \xi_k}} = \frac{\sum_{k=1}^n \xi_k - \sum_{k=1}^n a_k}{\sqrt{\sum_{k=1}^n b_k^2}} = \frac{\sum_{k=1}^n \xi_k - A_n}{B_n} =$$

$$= \frac{\sum_{k=1}^n \eta_k}{B_n},$$

согласно свойствам 1 и 2 характеристических функций, имеем

$$\varphi_{\zeta_n^*}(t) = \prod_{k=1}^n \varphi_k\left(\frac{t}{B_n}\right).$$

Рассмотрим величину

$$\ln \varphi_k\left(\frac{t}{B_n}\right) = \ln\left(1 - \frac{b_k^2}{B_n^2} \cdot \frac{t^2}{2!} + \frac{c_k^3}{B_n^3} \cdot \frac{|t|^3}{3!} \varkappa_k\right) = \ln(1 + z_k),$$

где

$$z_k = -\frac{b_k^2}{B_n^2} \frac{t^2}{2!} + \frac{c_k^3}{B_n^3} \frac{|t|^3}{3!} \varkappa_k. \qquad (7.12)$$

Согласно лемме $b_k \leqslant c_k$, поэтому $\alpha_k = \frac{b_k}{c_k} \leqslant 1$, и величину z_k можно представить в виде

$$z_k = -\frac{c_k^2}{B_n^2} \frac{t^2}{2!} \alpha_k^2 + \frac{c_k^3}{B_n^3} \cdot \frac{|t|^3}{3!} \varkappa_k = \frac{c_k^2}{B_n^2}\left[-\frac{t^2}{2!}\alpha_k^2 + \right.$$

$$\left. + \frac{c_k}{B_n} \frac{|t|^3}{3!} \varkappa_k\right]. \qquad (7.13)$$

Согласно условию теоремы $\frac{c_k}{B_n} \leqslant \frac{C_n}{B_n} \to 0$ при $n \to \infty$. поэтому при любом фиксированном t величины z_k стремятся к нулю при $n \to \infty$ и притом равномерно относительно k. Следовательно, для всех достаточно больших n $|z_k| \leqslant \frac{1}{2}$ при всех k и для $\ln(1 + z_k)$ получаем

$$\ln(1 + z_k) = \frac{z_k}{1} - \frac{z_k^2}{2}\left(1 - \frac{2}{3} z_k + \frac{2}{4} z_k^2 - \ldots\right) = z_k +$$

$$+ \frac{1}{2} z_k^2\left(1 + \frac{1}{2} + \frac{1}{2^2} + \ldots\right)\beta_k = z_k + \beta_k z_k^2,$$

где

$$\beta_k = -\frac{1 - \frac{2}{3} z_k + \frac{2}{4} z_k^2 - \ldots}{2} \quad \text{и} \quad |\beta_k| \leqslant 1.$$

Таким образом, учитывая формулы (7.12) и (7.13), имеем

$$\ln(1 + z_k) = -\frac{b_k^2}{B_n^2} \cdot \frac{t^2}{2!} + \frac{c_k^3}{B_n^3} \frac{|t|^3}{3!} \varkappa_k + \beta_k \frac{c_k^4}{B_n^4}\left[-\frac{t^2}{2!}\alpha_k^2 + \right.$$

$$\left. + \frac{c_k}{B_n} \cdot \frac{|t|^3}{3!} \varkappa_k\right]^2 = -\frac{b_k^2}{B_n^2} \cdot \frac{t^2}{2!} + \frac{c_k^3}{B_n^3}\left[\frac{|t|^3}{3!} \varkappa_k + \right.$$

$$+\beta_k \frac{c_k}{B_n}\left(-\frac{t^2}{2!}\alpha_k^2+\frac{c_k}{B_n}\frac{|t|^3}{3!}\varkappa_k\right)^2\Bigg].$$

Суммируя полученные равенства по k от 1 до n, находим

$$\ln \varphi_{\zeta_n^*}(t)=\ln\prod_{k=1}^{n}\varphi_k\left(\frac{t}{B_n}\right)=\sum_{k=1}^{n}\ln(1+z_k)=$$

$$=-\frac{\sum\limits_{k=1}^{n} b_k^2}{B_n^2}\cdot\frac{t^2}{2!}+\frac{C_n^3}{B_n^3}\gamma_n=-\frac{t^2}{2}+\frac{C_n^3}{B_n^3}\gamma_n,$$

где γ_n — величина ограниченная.

Переходя к пределу при $n\to\infty$ и учитывая, что $\frac{C_n}{B_n}\to 0$ при $n\to\infty$, имеем

$$\lim_{n\to\infty}\varphi_{\zeta_n^*}(t)=e^{-\frac{t^2}{2}}.$$

Теорема доказана.

§ 4. ПРИБЛИЖЕННАЯ НОРМАЛЬНОСТЬ СЛУЧАЙНОЙ ОШИБКИ ИЗМЕРЕНИЯ

Всякое измерение неизбежно сопряжено с ошибками; систематические ошибки можно устранить, случайные же ошибки измерения полностью никогда не могут быть устранены. Случайная ошибка измерения вызывается многими причинами, каждая из которых лишь незначительно влияет на результат. Каждая из причин порождает свою, так называемую элементарную ошибку измерения. Например, при взвешивании некоторого тела на точных весах случайная ошибка в определении его веса складывается из элементарных ошибок, вызываемых атмосферными причинами (колебаниями плотности, температуры, влажности воздуха, воздушными потоками); попаданием на чаши весов пылинок; неточностями, допущенными измерителем при снятии показаний со шкалы весов; незначительными вибрациями основания весов (которые в свою очередь могут вызываться многими причинами) и т. п. Реально наблюдаемая случайная ошибка измерения есть *сумма элементарных ошибок*. Количество элементарных ошибок велико, роль каждой из них в образовании случайной ошибки измерения мала, поэтому в силу теоремы Ляпунова случайная ошибка измерения должна быть распределена приближенно *по нормальному закону.*

Опыт показывает, что наблюдающиеся распределения вероятностей случайных ошибок измерения очень хорошо согласуются с нормальным законом. Итак, при *прямых измерениях случайная ошибка измерения распределена по закону, близкому к нормальному.*

Пусть a — измеряемая величина, ξ — результат измерения, η — ошибка измерения, тогда

$$\xi = a + \eta.$$

При отсутствии систематической ошибки, т. е. когда $M\xi = a$, имеем $M\eta = 0$, и, следовательно, случайная ошибка измерения η имеет плотность распределения вероятностей $p_\eta(x)$ вида

$$p_\eta(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}},$$

где σ — среднее квадратичное отклонение величины η, называемое также средней квадратичной ошибкой измерения. Результат измерения ξ имеет при этом следующую плотность распределения вероятностей:

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Упражнения к главе VII

1. На отрезке $(0,1)$ независимо друг от друга случайным образом выбираются n чисел. Подобрать n так, чтобы с вероятностью 0,99 сумма этих чисел была больше 100 [считать, что вероятность попадания каждого из чисел внутрь отрезка, целиком лежащего внутри $(0,1)$, равна длине этого отрезка].

2. Игральная кость бросается 1000 раз. Найти пределы, симметричные относительно среднего, в которых с вероятностью 0,95 будет лежать число выпавших очков.

3. Стрелок попадает при выстреле по мишени в десятку с вероятностью 0,5; в девятку — 0,3; в восьмерку — 0,1; в семерку — 0,05; в шестерку — 0,05. Сделано 100 выстрелов. Какова вероятность того, что выбито более 950 очков? Более 980 очков?

4. Случайная величина η является средней арифметической 10 000 независимых одинаково распределенных случайных величин, среднее квадратичное отклонение каждой из которых равно двум. Какое максимальное отклонение величины η от ее математического ожидания можно ожидать с вероятностью, не меньшей 0,9544?

ГЛАВА VIII

ЭЛЕМЕНТЫ ТЕОРИИ СЛУЧАЙНЫХ ПРОЦЕССОВ

§ 1. ОПРЕДЕЛЕНИЕ СЛУЧАЙНОГО ПРОЦЕССА

При изучении различных реальных явлений рассматривают всевозможные величины, зависящие от времени: $x = x(t)$. Такие величины принято называть *процессами*. Так, например, процессами являются температура воздуха в определенной точке пространства, давление в камере сгорания поршневого двигателя, скорость снаряда через время t после выстрела из орудия, число бактерий в популяции в момент t, уровень воды в водохранилище и т. п. Если значения $x(t)$ зависят от случая, т. е. при фиксированных t $x(t)$ — случайные величины, то $x(t)$ называется *случайным* процессом.

Строгое определение случайного процесса состоит в следующем.

Пусть $(\Omega, \mathfrak{A}, P)$ — вероятностное пространство, Θ — некоторое множество моментов времени. Пусть каждому $\omega \in \Omega$ поставлена в соответствие функция $\xi(t) = f(t, \omega)$, $t \in \Theta$, со значениями в m-мерном пространстве ($m \geqslant 1$) такая, что при каждом фиксированном $t \in \Theta$ $\xi(t) = f(t, \omega)$ является случайной величиной. Эта функция называется *случайным процессом*. Следовательно, случайный процесс — это совокупность случайных величин. В частности, обычную случайную величину ξ также можно рассматривать как случайный процесс, если определить Θ как одноэлементное множество ($\Theta = \{t_0\}$) и положить $\xi(t_0) = \xi$.

Итак, случайный процесс — функция двух переменных: t и ω. Как следует из определения, при любом фиксированном t $\xi(t)$ является m-мерной случайной величиной. Если же фиксировать ω, то мы получим функцию времени $\xi(t) = f(t, \omega)$, которая называется *реализацией* (траекторией, выборочной функцией) случайного процесса.

Приведем один из простейших примеров случайного процесса — так называемый веерный случайный процесс. Пусть ω — любая одномерная случайная величина. Определим случайный процесс формулой

$$\xi(t) = \omega(t - t_0) + b, \quad -\infty < t < \infty,$$

где t_0 и b — фиксированные числа. При фиксированном t $\xi(t)$ будет случайной величиной; при этом

$$F_{\xi(t)}(x) = P(\omega(t - t_0) + b < x) = \begin{cases} F_\omega\left(\dfrac{x-b}{t-t_0}\right) & \text{при } t > t_0, \\ \displaystyle\int_{-\infty}^{x} \delta(t-b)\,dt & \text{при } t = t_0, \\ 1 - F_\omega\left(\dfrac{x-b}{t-t_0} + 0\right) & \text{при } t < t_0. \end{cases}$$

Если же фиксировать ω, то $\xi(t)$, т. е. реализация процесса $\xi(t)$, представляет собой линейную функцию, график которой проходит через точку (t_0, b). Различные траектории процесса в данном случае представляют собой пучок прямых (рис. 8.1), откуда и название «веерный случайный процесс».

Если множество Θ является конечным, определение процесса равносильно определению многомерной случайной величины. Так,

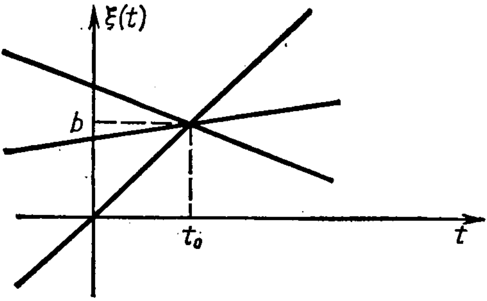

Рис. 8.1

пусть $\Theta = \{t_1, t_2, \dots, t_n\}$. Тогда можно положить $\xi_i = \xi(t_i)$, $1 < i \leqslant n$, и безразлично, что задавать: случайную величину $(\xi_1, \xi_2, \dots, \xi_n)$ или функцию $\xi(t)$, $t \in \{t_1, t_2, \dots, t_n\}$. Если множество Θ конечно или счетно: $\Theta = \{t_n\}$, то случайный процесс $\xi(t)$, $t \in \Theta$, называется *случайным процессом с дискретным временем* (с дискретным параметром) или *случайной последовательностью*. Если Θ — некоторый интервал положительной длины, конечный или бесконечный, то случайный процесс $\xi(t)$, $t \in \Theta$, называется *случайным процессом с непрерывным временем* (с непрерывным параметром). Конечно, Θ может быть любым числовым множеством; однако в приложениях основную роль играют случайные процессы с дискретным временем и с непрерывным временем в том смысле, в каком они нами определены.

Процесс $\xi(t)$ называется *дискретным случайным процессом* (иначе, случайным процессом с дискретным множеством значений), если существует такое конечное или счетное множество X, что функция $f(t, \omega)$ принимает значения только из множества X.

Рассмотрим некоторые основные классы случайных процессов, получившие широкое практическое применение.

§ 2. ЦЕПИ МАРКОВА ОБЩЕГО ВИДА

В соответствии с результатами, полученными в § 8 гл. V, если $\xi_1, \xi_2, \dots, \xi_n$ — произвольные случайные величины, заданные на одном и том же вероятностном пространстве, то можно построить такие независимые случайные величины $\omega_1, \omega_2, \dots, \omega_n, \dots$, что

$$\xi_1 = f_1(\omega_1), \quad \xi_2 = f_2(\xi_1, \omega_2), \dots, \xi_n = f_n(\xi_1, \dots, \xi_{n-1}, \omega_n), \dots,$$

т. е. n-я случайная величина ξ_n есть функция от n-го из независимых элементарных событий ω_n и случайных величин $\xi_1, \xi_2, \ldots, \xi_{n-1}$; при этом можно, например, полагать, что $\{\omega_n\}$ равномерно распределены в интервале $(0,1)$.

Таким образом, все разнообразие зависимостей между случайными величинами можно описать функциями $f_n(x_1, \ldots, x_{n-1}, y)$.

Зависимость между случайными величинами $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ называется марковской, если

$$f_3(x_1, x_2, y) = f_3(x_2, y), \ldots \quad f_n(x_1, \ldots, x_{n-1}, y) = f_n(x_{n-1}, y), \ldots, \tag{8.1}$$

т. е. если ξ_n полностью определяется значением ξ_{n-1} и n-м элементарным событием ω_n из последовательности независимых элементарных событий $\{\omega_n\}$. В этом случае последовательность случайных величин $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ называется *цепью Маркова*. Последовательность ξ_n называется *состоянием цепи Маркова* в момент n $(n = 1, 2, \ldots)$.

Рассмотрим некоторые примеры цепей Маркова.

1. Производится последовательность независимых испытаний, каждое из которых может быть «успешным» с вероятностью p и «неудачным» с вероятностью $q = 1-p$. Обозначим через ξ_n число успехов в первых n испытаниях $(n = 1, 2, \ldots)$. Тогда $\{\xi_n\}$ есть цепь Маркова. В этом можно убедиться следующим образом. Пусть $\omega_n = 1$, если n-е испытание успешно, $\omega_n = 0$ в противном случае. Тогда $\{\omega_n\}$ — последовательность независимых элементарных событий. В то же время $\xi_n = \xi_{n-1} + \omega_n$, т. е. выполняется определяющее свойство цепи Маркова. Вообще, если $\{\omega_n\}$ — произвольная последовательность независимых случайных величин, то последовательность сумм $S_n = \omega_1 + \ldots + \omega_n$, $n \geqslant 1$, образует цепь Маркова.

2. Имеется водохранилище, которое в любой момент $t \geqslant 0$ характеризуется некоторым уровнем наполнения $x(t)$. Минимальный уровень равен 0. Приток воды происходит в моменты $T, 2T, \ldots, nT, \ldots$, причем порции воды, поступающие в эти моменты, — независимые случайные величины $\omega_1, \omega_2, \ldots, \omega_n \ldots$. Сток воды равномерен, т. е. за время dt в интервале между nT и $(n+1)T$ уровень наполнения водохранилища уменьшается на величину αdt, $\alpha > 0$, если этот уровень еще не обратился в нуль. Обозначим через ξ_n уровень наполнения водохранилища в момент, непосредственно следующий за моментом поступления n-й порции воды:

$$\xi_n = x(nT + 0) = \lim_{\varepsilon \downarrow 0} x(nT + \varepsilon).$$

Между введенными случайными величинами существует следующая зависимость:

$$\xi_n = \begin{cases} \xi_{n-1} - \alpha T + \omega_n, & \text{если} \quad \xi_{n-1} - \alpha T > 0, \\ \omega_n, & \text{если} \quad \xi_{n-1} - \alpha T \leqslant 0. \end{cases}$$

Действительно, если $\xi_{n-1} > \alpha T$, то за время T уровень воды в водохранилище уменьшится на αT и в момент nT он увеличится на ω_n; если же $\xi_{n-1} \leqslant \alpha T$, то за время T уровень воды в водохранилище понизится до нуля; следовательно, в момент $nT + 0$ в водохранилище будет только та порция воды, которая поступила в момент nT.

3. Рассмотрим то же водохранилище с тем лишь отличием, что моменты поступления воды случайны: n-я порция поступает в момент T_n, где $(T_n - T_{n-1}) = \eta_n$ — независимые случайные величины. В этом случае имеем

$$\xi_n = \begin{cases} \xi_{n-1} - \alpha n_n + \omega_n, & \text{если} \quad \xi_{n-1} > \eta_n, \\ \omega_n, & \text{если} \quad \xi_{n-1} \leqslant \eta_n. \end{cases}$$

Последовательность $\{\xi_n\}$ является цепью Маркова, поскольку можно записать

$$\xi_n = f\left(\xi_{n-1},\ \omega_n'\right),$$

где $\omega_n' = (\eta_n,\ \omega_n)$ — независимые элементарные события.

4. Допустим, что изменение величины заряда $\xi \geqslant 0$ в электронной лампе описывается следующим законом. В интервалах между попаданием на сетку заряженных частиц заряд убывает по показательному закону с параметром α; в момент попадания частицы заряд увеличивается на случайную величину. Пусть η_n — время между попаданием на сетку n — 1-й и n-й частицы, ζ_n — величина, на которую возрастает заряд при попадании n-й частицы. Обозначим через ξ_n величину заряда после попадания на сетку n-й частицы. Тогда

$$\xi_n = \xi_{n-1} e^{-\alpha\eta_n} + \zeta_n$$

Здесь $\xi_n = f(\xi_{n-1},\ \omega_n)$, где $\omega_n = (\eta_n,\ \zeta_n)$. Если двумерные случайные величины $(\eta_n,\ \zeta_n)$ независимы в совокупности, то последовательность $\{\xi_n\}$ будет цепью Маркова.

Цепь Маркова можно определить и через условные распределения, не обращаясь к вероятностному пространству, на котором определены случайные величины. Согласно такому определению, последовательность случайных величин $\{\xi_n\}$, $n \geqslant 1$, образует цепь Маркова, если для любого $n \geqslant 3$ условное распределение ξ_n при фиксированных $\xi_1 = x_1, \ldots, \xi_{n-1} = x_{n-1}$ совпадает с условным распределением ξ_n при $\xi_{n-1} = x_{n-1}$:

$$\mathrm{P}\,(\xi_n \in A / \xi_1 = x_1, \ldots, \xi_{n-1} = x_{n-1}) = \mathrm{P}\,(\xi_n \in A / \xi_{n-1} = x_{n-1}) \quad (8.2)$$

при любых n, $x_1, \ldots, x_{n-1}$ и A, принадлежащих σ-алгебре множеств $\mathfrak{A}$. Равенство (8.2) следует из равенства (8.1), так как

$$\mathrm{P}\,(\xi_n \in A/\xi_1 = x_1, \ldots, \xi_{n-1} = x_{n-1}) = \mathrm{P}\,(f_n(\xi_{n-1},\ \omega_n) \in A/\xi_1 =$$
$$= x_1, \ldots, \xi_{n-1} = x_{n-1} = \mathrm{P}\,(f_n(x_{n-1},\ \omega_n) \in A) =$$
$$= \mathrm{P}\,(f_n(\xi_{n-1},\ \omega_n) \in A/\xi_{n-1} = x_{n-1}).$$

Предположим теперь, что случайные величины ξ_n заданы распределениями $F_n(x_1, \ldots, x_n) = \mathrm{P}\,(\xi_1 < x_1, \ldots, \xi_n < x_n)$, причем выполняется равенство (8.2). Покажем, как нужно строить пространства независимых элементарных событий $\Omega_n = \{\omega_n\}$, чтобы имело место равенство (8.1). Обозначим через $p_n(x, y)$ условную вероятность события $\{\xi_n < y\}$ при условии $\xi_{n-1} = x$. Теперь определим $\{\omega_n\}$ как последовательность независимых случайных величин, равномерно распределенных в интервале $(0,1)$, и положим $\xi_n =$

$= p_n^{-1}(\xi_{n-1}, \omega_n)$, $n \geqslant 2$, где $p_n^{-1}(x, z)$ — функция, обратная к функции $p_n(x, y)$ при фиксированном x: если $p_n(x, y) = z$, то $p_n^{-1}(x, z) = y$. [Если $p_n^{-1}(x, \omega_n)$ не существует, т. е. $p_n(x, y)$ не равно данному ω_n ни при каком y, положим $\xi_n = \inf\limits_{\omega > \omega_n} p_n^{-1}(x, \omega)$, где inf берется по тем ω, для которых $p_n^{-1}(x, \omega)$ существует. Это равносильно тому, что график функции $z = p_n(x,y)$ при фиксированном x дополняется отрезками, параллельными оси Oz, в точках разрыва этой функции и ξ_n определяется как значение y, при котором прямая $z = \omega_n$ пересекает построенный график.] Случайные величины имеют вид (8.1); в то же время они имеют заданное распределение $P(\xi_1 < x_1, \ldots, \xi_n < x_n) = F_n(x_1, \ldots, x_n)$. В самом деле,

$$P(\xi_1 \in \Delta_1, \ldots, \xi_n \in \Delta_n) = \int\limits_{\Delta_1} dF_{\xi_1}(x_1) \int\limits_{\Delta_2} dF_{\xi_2}(x_2/\xi_1 = x_1) \ldots \int\limits_{\Delta_n} dF_{\xi_n}(x/\xi_1 = x_1, \ldots . \xi_{n-1} = x_{n-1}), \tag{8.3}$$

т. е. распределение случайной величины $(\xi_1, \ldots, \xi_n)$ полностью определяется заданием $P(\xi_1 \in A_i/\xi_1 = x_1, \ldots, \xi_{i-1} = x_{i-1})$. В силу условия (8.2) случайные величины ξ_i будут иметь заданное распределение при выполнении следующих условий:

1) ξ_1 имеет заданное распределение;

2) $P(\xi_i \in A_i/\xi_1 = x_1, \ldots, \xi_{i-1} = x_{i-1}) = P(\xi_i \in A_i/\xi_{i-1} = x_{i-1})$;

3) $P(\xi_i \in A_i/\xi_{i-1} = x_{i-1})$ имеет заданный вид.

Первое из этих условий удовлетворяется, если выбрать соответствующим образом $F_{\xi_1}(x_1)$. Второе условие выполняется, поскольку построенные нами случайные величины удовлетворяют соотношению $\xi_i = f_i(\xi_{i-1}, \omega_i)$, где ω_i независимы в совокупности. Наконец, третье условие, как было показано в § 8 гл. V, выполняется в силу равенств

$$P(\xi_i < y/\xi_{i-1} = x) = P\left(P_i^{-1}(x, \omega_i) < y\right) = P(p_i(x, y) < \omega_i) = P_i(x, y)$$

[здесь использовано то, что ω_i обладает равномерным распределением в интервале $(0,1)$].

Следовательно, определения цепи Маркова равенствами (8.1) и (8.2) эквивалентны.

Равенство (8.3) можно переписать следующим образом:

$$P(\xi_1 \in \Delta_1, \ldots, \xi_n \in \Delta_n) = \int\limits_{\Delta_1} dF_{\xi_1}(x_1) \int\limits_{\Delta_2} dF_{\xi_2}(x_2/x_1) \ldots \int\limits_{\Delta_n} dF_{\xi_n}(x_n/x_{n-1}), \tag{8.4}$$

где $F_{\xi_i}(x_i/x_{i-1})$ — условная функция распределения случайной величины ξ_i при условии $\xi_{i-1} = x_{i-1}$, или, что то же,

$$F_{\xi_i}(x_i/x_{i-1}) = \mathrm{P}(f_i(x_{i-1}, \omega_i) < x_i), \quad i \geqslant 2. \tag{8.5}$$

Таким образом, распределение случайных величин $\{\xi_n\}$, образующих цепь Маркова, полностью определяется заданием функций $F_{\xi_1}(x_1)$ и $F_{\xi_i}(x_i/x_{i-1})$, $i \geqslant 2$. Первая из этих функций называется начальным распределением цепи Маркова, вторая — переходной вероятностной функцией цепи Маркова.

Цепь Маркова называется *однородной*, если функция $F_{\xi_i}(y/x)$ не зависит от $i \geqslant 2$.

Пусть $\{\xi_n\}$ — однородная цепь Маркова. Обозначим через $\mathrm{P}_n(y/x)$ условную вероятность события $\{\xi_{n+m} < y\}$ при условии $\xi_m = x$. Тогда справедливо уравнение

$$\mathrm{P}(y/x) = \int \mathrm{P}_{n-k}(y/z)\, d\,\mathrm{P}_k(z/x), \tag{8.6}$$

где k — любое число от 1 до $n-1$. Равенство (8.6) можно получить из равенства (8.4), положив в нем

$$\Delta_1 = \ldots \Delta_{n-1} = (-\infty, \infty), \quad \Delta_n = (-\infty, y),$$

$$F_{\xi_1}(x_1) = \int_{-\infty}^{x_1} \delta(t-x)\, dt, \quad x_k = z$$

и заметив, что

$$\int dF_{\xi_2}(x_2/x_1) \int dF_{\xi_3}(x_3/x_2) \cdots \int dF_{\xi_k}(x_k)\, F_{\xi_{k+1}}/x_k) = \\ = \mathrm{P}_k(x_{k+1}/x_1), \tag{8.7}$$

$$\int dF_{\xi_{k+2}}(x_{k+2}/x_{k+1}) \cdots \int dF_{\xi_n}(x_n/x_{n-1})\, F(x_{n+1}/x_n) = \\ = \mathrm{P}_{n-k}(x_{n+1}/x_n). \tag{8.8}$$

Если в формуле (8.6) считать $k = 1$, то получим

$$\mathrm{P}_n(y/x) = \int \mathrm{P}_{n-1}(y/z)\, d\,\mathrm{P}_1(z/x), \quad n \geqslant 2. \tag{8.9}$$

Эта формула позволяет последовательно найти все $\mathrm{P}_n(y/x)$ через $\mathrm{P}_1(y/x)$.

Распределение случайной величины ξ_n определяется по формуле

$$F_{\xi_n}(y) = \int \mathrm{P}_{n-1}(y/x)\, dF_{\xi_1}(y), \quad n \geqslant 2. \tag{8.10}$$

В качестве примера найдем переходную вероятностную функцию цепи Маркова, задаваемой равенством

$$\xi_n = \xi_{n-1}\, e^{-a\eta_n} + \zeta_n,$$

где η_n, ζ_n — независимые случайные величины с непрерывными функциями распределения $H(x)$, $G(x)$. Введем вспомогательную случайную величину $\gamma_n = \xi_{n-1}\, e^{-a\eta_n}$. Тогда

$$\mathrm{P}(\gamma_n < y/\xi_{n-1} = x) = \mathrm{P}\left(xe^{-\alpha\eta_n} < y\right) = \mathrm{P}\left(\eta_n > \frac{1}{\alpha}\ln\frac{x}{y}\right) =$$

$$= 1 - F_{\eta_n}\left(\frac{1}{\alpha}\ln\frac{x}{y}\right) = 1 - H\left(\frac{1}{\alpha}\ln\frac{x}{y}\right).$$

Так как при фиксированном $\xi_{n-1} = x$ ξ_n есть сумма независимых случайных величин ($\xi_n = \gamma + \zeta_n$), то по формуле для распределения суммы независимых слагаемых имеем

$$\mathrm{P}_1(y/x) = \int_0^y \mathrm{P}(\gamma_n < y - z/\xi_{n-1} = x)\, dF_{\zeta_n}(z) =$$

$$= \int_0^y \left[1 - H\left(\frac{1}{\alpha}\ln\frac{x}{y-z}\right)\right] dG(z). \tag{8.11}$$

Распределение вероятностей $F(y)$ называется *стационарным распределением* (вероятностей состояний) цепи Маркова $\{\xi_n\}$, если справедливо равенство

$$F(y) = \int \mathrm{P}_1(y/x)\, dF(x). \tag{8.12}$$

Допустим, что $F_{\xi_{n-1}}(y) = F(y)$, где $F(y)$ удовлетворяет равенству (8.12). Тогда

$$F_{\xi_n}(y) = \int \mathrm{P}_1(y/x)\, dF_{\xi_{n-1}}(x) = \int \mathrm{P}_1(y/x)\, dF(x) = F(y). \tag{8.13}$$

Следовательно, если ξ_{n_0} имеет стационарное распределение, то и все ξ_n, $n \geqslant n_0$ имеют то же распределение.

Распределение вероятностей $F(y)$ называется *предельным распределением* цепи Маркова $\{\xi_n\}$ при заданном начальном распределении $F_{\xi_1}(x)$, если

$$\lim_{n\to\infty} \int \mathrm{P}_n(y/x)\, dF_{\xi_1}(x) = F(y). \tag{8.14}$$

При достаточно общих условиях справедливо следующее утверждение: предельное распределение цепи Маркова может быть только стационарным.

Предельное распределение цепи Маркова $F(y)$ называется *эргодическим*, если оно не зависит от начального распределения $F_{\xi_1}(x)$. В частности, если

$$F_{\xi_1}(z) = \int_{-\infty}^{z} \delta(z - x)\, dz,$$

то

$$F_{\xi_{n+1}}(y) = \mathrm{P}(y/x),$$

откуда

$$F(x) = \lim_{n\to\infty} \mathrm{P}_n(y/x). \tag{8.15}$$

Это означает, что распределения $\mathrm{P}_n(y/x)$ при $n \to \infty$ сходятся к предельному распределению, не зависящему от x.

§ 3. КОНЕЧНЫЕ ОДНОРОДНЫЕ ЦЕПИ МАРКОВА

Цепь Маркова $\{\xi_n\}$ называется *конечной однородной*, если она однородна и множество ее состояний конечно. Без ограничения общности можно сказать, что возможные состояния цепи Маркова — числа 1, 2, ... , m. Задание начального распределения $F_{\xi_1}(x)$ равносильно заданию начальных вероятностей $p_i^{(0)} = P(\xi_1 = i)$, $1 \leqslant i \leqslant m$. Переходная вероятностная функция конечной однородной цепи Маркова будет полностью задана, если заданы постоянные

$$p_{ij} = P(\xi_n = j/\xi_{n-1} = i), \quad n \geqslant 2, \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant j \leqslant m. \tag{8.16}$$

Эти постоянные называются *вероятностями перехода* цепи Маркова (за один шаг). Постоянные

$$p_{ij}^{(n)} = P(\xi_{n+m} = j/\xi_m = i), \quad m \geqslant 1, \quad n \geqslant 1, \quad 1 \leqslant i \leqslant m,$$
$$1 \leqslant j \leqslant m \tag{8.17}$$

называются вероятностями перехода цепи Маркова за n шагов. Очевидно, $p_{ij}^{(1)} = p_{ij}$. Матрица вида

$$P = \begin{Vmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ \cdot & \cdot & \cdot & \cdot \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{Vmatrix} \tag{8.18}$$

называется *матрицей перехода* конечной однородной цепи Маркова.

По формуле полной вероятности имеем

$$p_{ij}^{(n)} = \sum_{k=1}^{m} p_{ik}\, p_{kj}^{(n-1)}. \tag{8.19}$$

Эта формула выражает тот факт, что переход из состояния i в момент m в состояние j в момент $n + m$ осуществляется следующим образом. Сначала цепь Маркова принимает некоторое состояние k в момент $m + 1$, а затем из состояния k в момент $m + 1$ переходит в состояние j в момент $n + m$.

Формула (8.19) в матричном виде имеет вид

$$P^{(n)} = P\, P^{(n-1)};$$

отсюда по индукции находим, что

$$P^{(n)} = \| p_{ij}^{(n)} \| = P^n, \tag{8.20}$$

т. е. матрица перехода за n шагов есть n-я степень матрицы перехода за один шаг. Можно записать еще так:

$$P^{(n)} = P^{(n-1)} P,$$

или

$$p_{ij}^{(n)}=\sum_{k=1}^{m} p_{ik}^{(n-1)}\, p_{kj}, \quad 1\leqslant i\leqslant m, \quad 1\leqslant j\leqslant m, \quad n\geqslant 2. \tag{8.21}$$

Рассмотрим способ вычисления $p_{ij}^{(n)}$. При небольших n можно просто возводить матрицу P в требуемую степень; при больших n используют следующий метод. Вначале вводят постоянные

$$p_{ij}^{(0)}=\delta_{ij}=\begin{cases} 1, & \text{если} \quad i=j, \\ 0, & \text{если} \quad i\neq j. \end{cases} \tag{8.22}$$

Тогда уравнения (8.19) и (8.21) выполняются и при $n=1$, что можно проверить непосредственной подстановкой. Введем теперь функции

$$\varphi_{ij}(z)=\sum_{n=0}^{\infty} z^n\, p_{ij}^{(n)}. \tag{8.23}$$

Ряд, записанный в правой части равенства, сходится по крайней мере при $|z|<1$, так как $0\leqslant p_{ij}^{(n)}\leqslant 1$.

Умножив обе части уравнения (8.21) на z^n и просуммировав по n от 1 до ∞, получим

$$\sum_{n=1}^{\infty} z^n\, p_{ij}^{(n)}=\sum_{n=1}^{\infty}\ \sum_{k=1}^{\infty} z^n\, p_{ik}^{(n-1)}\, p_{kj},$$

или

$$\sum_{n=0}^{\infty} z^n\, p_{ij}^{(n)}-p_{ij}^{(0)}=z\sum_{n=0}^{\infty}\ \sum_{k=1}^{m} z^n\, p_{ik}^{(n)}\, p_{kj}.$$

Отсюда

$$\varphi_{ij}(z)-p_{ij}^{(0)}=z\sum_{k=1}^{m} \varphi_{ik}(z)\, p_{kj}.$$

Подставив в последнее равенство (8.22), найдем

$$\varphi_{ij}(z)-z\sum_{k=1}^{m} \varphi_{ik}(z)\, p_{kj}=\delta_{ij}, \quad 1\leqslant i\leqslant m, \quad 1\leqslant j\leqslant m. \tag{8.24}$$

Мы получили систему m^2 уравнений с m^2 неизвестными. Однако как в левую, так и в правую часть равенств (8.24) входит одно и то же i, поэтому можно отдельно решить m уравнений с фиксированным i. Обозначим через $\Delta(z)$ определитель этой системы (он один и тот же для всех i). Имеем

$$\Delta(z)=\begin{vmatrix} (1-zp_{11}) & -zp_{21} & -zp_{m1} \\ -zp_{12} & (1-zp_{22}) & -zp_{m2} \\ \dots & \dots & \dots \\ -zp_{1m} & -zp_{2m} & (1-zp_{mm}) \end{vmatrix}. \tag{8.25}$$

При малом $|z|$ диагональные элементы близки к 1, а недиагональные—к 0, т. е. определитель $\Delta(z)$ близок к 1. Значит, в некоторой окрестности точки $z = 0$ $\Delta(z) \neq 0$, т. е. система уравнений (8.24) имеет единственное решение. Так как коэффициенты этих уравнений — линейные функции z, то $\varphi_{ij}(z)$ являются отношениями полиномов

$$\varphi_{ij}(z) = \frac{R_{ij}(z)}{S_{ij}(z)}, \tag{8.26}$$

где $R_{ij}(z)$, $S_{ij}(z)$ — некоторые полиномы. Выражение (8.26) после выделения целой части можно разложить на простейшие дроби вида $\frac{\lambda}{(1-\alpha z)^{r+1}}$, где α, λ — некоторые комплексные постоянные, $r = 0, 1, 2, \ldots$. Будем исходить из тождества

$$\sum_{n=0}^{\infty} \alpha^n z^n = \frac{1}{1-\alpha z}. \tag{8.27}$$

Продифференцировав его r раз, получим

$$\sum_{n=r}^{\infty} n(n-1)\ldots(n-r+1)\,\alpha^n z^{n-r} = \frac{\alpha^r r!}{(1-\alpha z)^{r+1}}$$

или, что то же,

$$\frac{1}{r!}\sum_{n=0}^{\infty}(n+r)(n+r-1)\ldots(n+1)\,\alpha^n z^n = \frac{1}{(1+\alpha z)^{r+1}}. \tag{8.28}$$

Таким образом, каждой элементарной дроби вида $\frac{\lambda}{(1-\alpha z)^{r+1}}$ $(r = 1, 2, \ldots)$ соответствует составляющая $p_{ij}^{(n)}$ вида $\frac{\lambda}{r!}(n+r)\ldots(n+1)\,\alpha^n$; дроби вида $\frac{\lambda}{1-\alpha z}$ соответствует составляющая $p_{ij}^{(n)}$ вида $\lambda\alpha^n$. Следовательно, $p_{ij}^{(n)}$ равно конечной сумме всех указанных составляющих.

Описанным образом находятся пределы $p_{ij}(n)$ при $n \to \infty$ (если они существуют).

Рассмотрим конечную однородную цепь Маркова с двумя состояниями, матрицу перехода которой можно задать в виде

$$\mathrm{P} = \begin{Vmatrix} 1-a & a \\ b & 1-b \end{Vmatrix}.$$

(Любая матрица перехода $\| p_{ij} \|$ обладает тем свойством, что $\sum_{j=1}^{m} p_{ij} = 1$ при любом i, $1 \leqslant i \leqslant m$; поэтому в рассматриваемом случае независимых параметров всего два.)

Для определенности положим $i = 1$. Тогда уравнения (8.24) принимают вид

$$\varphi_{11}(z) - z\,\varphi_{11}(z)\,p_{11} - z\,\varphi_{12}(z)\,p_{21} = 1,$$

$$\varphi_{12}(z) - z\,\varphi_{11}(z)\,p_{12} - z\,\varphi_{12}(z)\,p_{22} = 0.$$

Имеем

$$\Delta(z) = \begin{vmatrix} (1 - zp_{11}) & -zp_{21} \\ -zp_{12} & (1 - zp_{22}) \end{vmatrix} = (1 - zp_{11})(1 - zp_{22}) - z^2 p_{12} p_{21} =$$

$$= z^2(1 - a - b) - z(2 - a - b) + 1.$$

Один из корней уравнения $z^2(1 - a - b) - z(2 - a - b) + 1 = 0$ равен 1, другой равен $\frac{1}{1 - a - b}$. Поэтому $p_{1j}^{(n)}$ должны иметь следующий вид:

$$p_{1j}^{(n)} = c_j \cdot 1^n + \lambda_j (1 - a - b)^n = c_j + \lambda_j (1 - a - b)^n, \ j = 1,\ 2.$$

По правилу Крамера находим

$$\varphi_{11}(z) = \frac{\Delta_1(z)}{\Delta(z)},$$

$$\Delta_{12}(z) = \frac{\Delta_2(z)}{\Delta(z)},$$

где

$$\Delta_2(z) = \begin{vmatrix} (1 - zp_{11}) & 1 \\ -zp_{12} & 0 \end{vmatrix} = zp_{12} = az.$$

Отсюда

$$\varphi_{11}(z) = \frac{1 - (1 - b)^2 z}{z^2(1 - a - b) - z(2 - a - b) + 1} =$$

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

$$\varphi_{12}(z) = \frac{az}{z^2(1 - a - b) - z(2 - a - b) + 1} =$$

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

и

$$p_{11}^{(n)} = \frac{b}{a + b} + \frac{a}{a + b}(1 - a - b)^n, \quad n \geqslant 1,$$

$$p_{12}^{(n)} = \frac{a}{a + b} - \frac{a}{a + b}(1 - a - b)^n, \quad n \geqslant 1.$$

Мы видим, что $p_{11}^{(n)} + p_{12}^{(n)} = 1$ при любом $n \geqslant 1$; при $n \to \infty$

$$p_{11}^{(n)} \to \frac{b}{a + b}, \qquad p_{12}^{(n)} \to \frac{a}{a + b}.$$

Цепи Маркова являются адекватной моделью процессов в цифровых автоматах. Представим себе цифровой автомат, работающий по заданной программе и изменяющий свои состояния в дискретные моменты времени 0, 1, 2, ... Пусть ξ_n — состояние автомата в момент n. Если бы функционирование автомата не было подвержено влиянию случайных факторов, то ξ_n было бы неслучайной функцией ξ_{n-1}:

$$\xi_n = f(\xi_{n-1}).$$

Однако в результате действия случайных факторов ξ_n оказывается функцией ξ_{n-1} и некоторой случайной величины ω_n:

$$\xi_n = f(\xi_{n-1}, \ \omega_n).$$

Если $\{\omega_n\}$ независимы и одинаково распределены, то мы получим конечную однородную цепь Маркова $\{\xi_n\}$. Пусть, например, ξ_n — состояние в момент n счетчика по модулю m, возможные состояния которого — 0, 1, 2, ... , $m-1$. Тогда в случае, если случайные факторы на счетчик не действуют, ξ_n удовлетворяют уравнению

$$\xi_n = \begin{cases} \xi_{n-1} + 1, & \text{если} \quad 0 \leqslant \xi_{n-1} \leqslant m-2, \\ 0, & \text{если} \quad \xi_{n-1} = m-1. \end{cases}$$

Если на работу счетчика действуют случайные факторы, то последняя формула выполняется с некоторой вероятностью q, а с вероятностью $p = 1 - q$ счетчик «не срабатывает», т. е. $\xi_n = \xi_{n-1}$. Таким образом, матрица перехода данной цепи Маркова имеет вид

$$\mathrm{P} = \begin{Vmatrix} p & q & & & & \\ & p & q & & & \\ & & p & q & & \\ & & & p & \ddots & \\ & & & & \ddots & q \\ q & & & & & p \end{Vmatrix},$$

где все незаполненные места заняты нулями; состояния цепи Маркова перенумерованы следующим образом: 0, 1, 2, ... , $m-1$, так что элементами матрицы P являются p_{ij}, $0 \leqslant i \leqslant m-1$, $0 \leqslant j \leqslant m-1$.

Пусть $\xi_0 = 0$. Распределение ξ_n можно найти, возведя матрицу P в степень n; однако в данном случае можно поступить и проще. Именно, заметим, что число несрабатываний счетчика есть число успехов в n независимых испытаниях с вероятностью успеха p в каждом испытании. Допустим, что n кратно m. В этом случае если в моменты 1, 2, ... , n счетчик сработал, то $\xi_n = 0$. Вычислим вероятность этого события. Очевидно, $\xi_n = 0$ в том и только том случае, если число несрабатываний кратно m:

$$\mathrm{P}(\xi_n = 0) = \sum_{r=0}^{n/m} C_n^{rm} p^{rm} q^{n-rm}.$$

Во многих реальных системах, к анализу которых применимы цепи Маркова, процесс функционирования системы проходит через множество однородных циклов (как образно говорят инженеры, «почти бесконечное множество»), поэтому представляет большой интерес исследование предельного поведения конечной однородной цепи Маркова ξ_n при $n \to \infty$. В силу формулы

$$P(\xi_n = j) = \sum_{i=1}^{m} P(\xi_1 = i)\, p_{ij}^{(n-1)}$$

для решения этой задачи достаточно исследовать поведение при $n \to \infty$ вероятностей перехода $p_{ij}^{(n)}$. Вначале нужно произвести классификацию состояний цепи Маркова. Пусть i и — два состояния конечной однородной цепи Маркова. Скажем, что состояние j достижимо из состояния i, если для некоторого $n \geqslant 1$ $p_{ij}^{(n)} > 0$.

Если состояние i достижимо из состояния j, а состояние j достижимо из состояния i, то состояния i и j называются *сообщающимися.*

Множество C состояний цепи Маркова называется *замкнутым* классом, если выполнены следующие свойства:

1) любые состояния i, j из C являются сообщающимися;

2) если $i \in C$, $j \overline{\in} C$, то $p_{ij} = 0$.

Таким образом, если при некотором n_0 $\xi_{n_0} \in C$, то $\xi_n \in C$ при любых $n \geqslant n_0$. Для любой конечной однородной цепи Маркова существует по меньшей мере один замкнутый класс состояний. Пусть $C_1, C_2, \ldots, C_r$ — замкнутые классы состояний данной цепи Маркова. Состояния, не принадлежащие ни одному из замкнутых классов, называются *несущественными.* Используем следующую графическую интерпретацию.

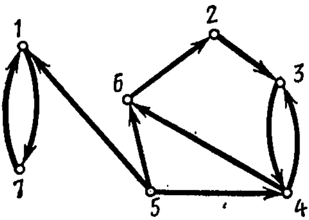

Рис. 8.2

Изобразим состояния цепи Маркова точками на плоскости. Если $p_{ij} > 0$, то от точки, изображающей состояние i, проведем дугу со стрелочкой к точке, изображающей состояние j. Таким образом, получим множество точек, некоторые из которых соединены направленными дугами (ориентированный граф) (рис. 8.2). Цепь Маркова, изображенная на этом рисунке, имеет два замкнутых класса состояний: $C_1 = \{1, 7\}$ и $C_2 = \{2, 3, 4, 6\}$. Действительно, из этих множеств состояний не выходит ни одна дуга. Состояние 5 является несущественным; из него можно попасть только в состояния 1, 4, 6, принадлежащие замкнутым классам C_1 и C_2.

Теорема 8.1. *Пусть C_0 — множество несущественных состояний. Пусть $\xi_1 = i$, где i — любое состояние цепи Маркова. Тогда существует такая случайная величина ν, что $\xi_n \overline{\in} C_0$ при $n \geqslant \nu$.*

Смысл этой теоремы в том, что в несущественных состояниях ξ_n может находиться лишь на протяжении конечного времени (если параметр n интерпретировать как время); выйдя из множества C_0, ξ_n попадает в один из замкнутых классов $C_1, C_2, \ldots, C_r$ и навсегда там остается. Обозначим через $f_{ij}^{(n)}$ вероятность того, что $\xi_2 \in C_0, \ldots, \xi_n \in C_0, \xi_{n+1} = j$ при условии $\xi_1 = i \overline{\in} C_0$. Если $j \overline{\in} C_0$, то $f_{ij}^{(n)}$ — вероятность того, что из несущественного состояния i цепь Маркова впервые выйдет из множества несущественных состояний через n шагов и при этом ее состояние будет j.

Можно записать

$$f_{ij}^{(n)} = p_{ij} \quad \text{при} \quad n = 1, \tag{8.29}$$

$$f_{ij}^{(n)} = \sum_{k \in C_0} f_{ik}^{(n-1)} p_{kj} \quad \text{при} \quad n \geqslant 2. \tag{8.30}$$

Решение этой системы уравнений аналогично решению системы (8.21).

Имеем

$$p_{ij}^{(n)} \xrightarrow[n\to\infty]{} 0 \quad \text{при} \quad j \in C_0, \tag{8.31}$$

$$p_{ij}^{(n)} = \sum_{k \overline{\in} C_0} \sum_{m=1}^{n-1} f_{ik}^{(m)} p_{kj}^{(n-m)} \quad \text{при} \quad j \overline{\in} C_0. \tag{8.32}$$

Из этой формулы следует, что для исследования предельного поведения $p_{ij}^{(n)}$ при любых i, j достаточно исследовать $p_{ij}^{(n)}$ при i, j, принадлежащих одному и тому же замкнутому классу состояний C_s $(1 \leqslant s \leqslant r)$, Так, например, если $p_{kj} \to \pi_j^{(s)}$ при $k \in C_s$, $j \in C_s$, $1 \leqslant s \leqslant r$, то из равенств (8.32) находим

$$p_{ij}^{(n)} \xrightarrow[n\to\infty]{} \sum_{s=1}^{r} \pi_j^{(s)} \sum_{m=1}^{\infty} \sum_{k \in C_s} f_{ik}^{(m)} \tag{8.33}$$

Однако $\sum\limits_{k \in C_s} f_{ik}^{(m)}$ есть вероятность ρ_{is} того, что ξ_n из состояния i попадает в замкнутый класс C_s $(1 \leqslant s \leqslant r)$. Таким образом,

$$p_{ij}^{(n)} \xrightarrow[n\to\infty]{} \sum_{s=1}^{r} \rho_{is} \pi_j^{(s)} \ . \tag{8.34}$$

Если $i \in C_k$ для некоторого $k \geqslant 1$, то $\rho_{ik} = 1$, $\rho_{is} = 0$ при $s \neq k$, и в этом случае

$$p_{ij}^{(n)} \xrightarrow[n\to\infty]{} \pi_j^{(s)}, \quad j \in C_k. \tag{8.35}$$

Рассмотрим теперь поведение $p_{ij}^{(n)}$ при i, j, принадлежащих одному и тому же замкнутому классу C_k. Чтобы несколько упростить обозначения, будем считать, что $C_k = \{1, 2, \ldots, m\}$; очевидно, это не ограничивает общности рассмотрения, а лишь упрощает обозначения.

Возьмем любое состояние i, принадлежащее данному замкнутому классу, и рассмотрим $p_{ii}^{(n)}$, $n \geqslant 1$. Пусть $(1 \leqslant)\, n_1 < n_2 < \ldots < n_N < \ldots$ — последовательность значений n, при которых $p_{ii}^{(n)} > 0$. Обозначим через l наибольший общий делитель чисел $n_1, n_2, \ldots, n_N, \ldots$. Возможно одно из двух: либо $l = 1$, либо $l > 1$.

Теорема 8.2. *Если* $l = 1$, *то*

$$p_{ij}^{(n)} \xrightarrow[n\to\infty]{} \pi_j, \quad i \in C_k, \tag{8.36}$$

где $\sum\limits_{j \in C_k} \pi_j = 1$.

Теорема 8.3. *Если* $l > 1$, *то*

$$p_{ij}^{(nl+t)} \xrightarrow[n\to\infty]{} \pi_{ij}^{(t)}, \quad i \in C_k, \tag{8.37}$$

где $\sum\limits_{j \in C_k} \pi_{ij}^{(t)} = 1,\ i \in C_k,\ 0 \leqslant t \leqslant l - 1.$

При $l = 1$ класс C_k называется апериодическим; при $l > 1$ — периодическим с периодом l. Таким образом, в случае апериодического замкнутого класса вероятности перехода при $n \to \infty$ сходятся к пределам, не зависящим от $i \in C_k$, при $l > 1$ существует предел $p_{ij}^{(n)}$ по любой возрастающей последовательности значений n, разности которых кратны l.

§ 4. ПУАССОНОВСКИЙ ПРОЦЕСС. ПРОСТЕЙШИЙ ПОТОК ОДНОРОДНЫХ СОБЫТИЙ

Пусть $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ — независимые положительные случайные величины с плотностью $p_{\xi_l}(x) = \lambda e^{-\lambda x}$, $x > 0$. Множество точек $\{z_n\}$, где $z_1 = \xi_1$, $z_2 = \xi_1 + \xi_2, \ldots, z_n = \xi_1 + \ldots + \xi_n, \ldots$, называется *простейшим потоком однородных событий*. Если для некоторого n $z_n = t$, то говорят, что в момент t произошло событие потока. Число тех n, для которых $a < z_n < b$, называется *числом событий потока* в интервале (a, b). Обозначим через $x(t)$, $t \geqslant 0$ число событий простейшего потока однородных событий в интервале $(0, t)$. Определенный таким образом случайный процесс $x(t)$ называется *пуассоновским процессом с параметром* λ.

Основные свойства процесса Пуассона и соответственно простейшего потока сводятся к следующему.

1. Стационарность: для любых интервалов (a_0, a_1), $(a_1, a_2), \ldots, (a_{n-1}, a_n)$, где $0 \leqslant a_0 \leqslant a_1 < a_2 < \ldots < a_n$, любого $\tau > 0$ и любых $k_1, k_2, \ldots, k_n$ вероятность того, что в интервале (a_0, a_1) произойдет k_1 событий потока, ..., в интервале (a_{n-1}, a_n) произойдет k_n событий потока, равны вероятности того, что в интервале $(a_0 + \tau, a_1 + \tau_n)$, произойдет k_n событий потока, ..., в интервале $(a_{n-1} + \tau, a_n + \tau)$ — k_1 событий потока. Это свойство можно еще сформулировать следующим образом. Для любых фиксированных $0 \leqslant a_0 < a_1 < \ldots < a_n$ и любого $\tau > 0$ многомерные случайные величины

$$(x(a_1) - x(a_0), \ldots, x(a_n) - x(a_{n-1}))$$

и

$$(x(a_1 + \tau) - x(a_0 + \tau), \ldots, x(a_n + \tau) - x(a_{n-1} + \tau))$$

имеют одинаковое распределение.

2. Ординарность: вероятность двух или большего числа событий потока в интервале $(t, t+h)$ есть $o(h)$ при $h \to 0$. Иначе, $P(x(t+h) - x(t) \geqslant 2) = o(h)$, $h \to 0$.

3. Отсутствие последействия: при $0 \leqslant a_0 < a_1 < \ldots < a_n$ случайные величины $x(a_1) - x(a_0), \ldots, x(a_n) - x(a_{n-1})$ независимы.

Докажем все перечисленные свойства.

Положим $a_0 = 0$. Тогда событие $\{x(a_1) - x(a_0) = k_1, \ldots, x(a_n) - x(a_{n-1}) = k_n\}$ равносильно сочетанию следующих условий:

$$z_{k_1} < a_1 < z_{k_1+1}, \quad z_{k_1+k_2} < a_2 < z_{k_1+k_2+1}, \ldots, z_{k_1+\ldots+k_n} < a_n < z_{k_1+\ldots+k_n+1}$$

Следовательно,

$$I = P(x(a_1) - x(a_0) = k_1, \ldots, x(a_n) - x(a_{n-1}) = k_n) =$$

$$= \lambda^{k_1+\ldots+k_n} \int \underset{D}{\ldots} \int \exp\{-\lambda[\tau_1 + (\tau_2 - \tau_1) + \ldots + \tau_{k_1+\ldots+k_n} - \tau_{k_1+\ldots+k_n-1} + \tau_{k_1+\ldots+k_n+1} - \tau_{k_1+\ldots+k_n}]\} d\tau_1 \ldots d\tau_{k_1+\ldots+k_n} \int_{a_n}^{\infty} d\tau_{k_1+\ldots+k_n+1}, \tag{8.38}$$

где область D определяется условиями

$$0 < \tau_1 < \ldots < \tau_{k_1} < a_1 < \tau_{k_1+1} < \ldots < \tau_{k_1+k_2} < a_2 < \tau_{k_1+k_2+1} < \ldots < \tau_{k_1+\ldots+k_n} < a_n < \tau_{k_1+\ldots+k_n+1} < \infty.$$

После сокращения подобных членов в показателе экспоненты находим

$$I = \lambda^{k_1+k_2+\ldots+k_n} \left\{ \int \underset{D}{\ldots} \int d\tau_1 \ldots d\tau_{k_1+\ldots+k_n} \right\} e^{-\lambda a_n}.$$

Интеграл

$$\int \underset{D}{\ldots} \int d\tau_1 \ldots d\tau_{k_1+\ldots+k_n}$$

можно представить в виде произведения $I_1 I_2 \ldots I_k$, где

$$I_1 = \int\limits_{0<\tau_1<\ldots<\tau_{k_1}<a_1} \ldots \int d\tau_1 \ldots d\tau_{k_1}, \quad I_2 = \int\limits_{a_1<\tau_{k_1+1}<\ldots<\tau_{k_1+k_2}<a_2} \ldots \int \times d\tau_{k_1+1} \ldots d\tau_{k_1+k_2}, \ldots, I_n =$$

$$= \int \cdots \int_{a_{n-1} < \tau_{k_1 + \ldots + k_{n-1} + 1} < \ldots < \tau_{k_1 + \ldots + k_n} < a_n} \times$$

$$\times d\tau_{k_1+\ldots+k_{n-1}+1} \ldots d\tau_{k_1+\ldots+k_n}.$$

Заменой переменных $\tau_1 = x_1, \ldots, \tau_{k1} = x_{k1}$ (тождественная замена), $\tau_{k_1+1} = x_1 + a_1, \ldots, \tau_{k_1+k_2} = x_{k_2} + a_1; \ldots; x_{k_1+\ldots+k_{n-1}+1} = x_1 + a_{n-1}, \ldots, x_{k_1+\ldots+k_n} = x_{k_n} + a_{n-1}$ эти интегралы приводятся к виду

$$I_j = \int \cdots \int_{0 < x_1 < \ldots < x_{k_j} < a_j - a_{j-1}} dx_1 \ldots dx_{k_j}, \quad 1 \leqslant j \leqslant n$$

Интеграл $\int\limits_0^a \ldots \int\limits_0^a dx_1 \ldots dx_m$, равный a^m, можно представить в виде суммы $m!$ равных между собой интегралов $\int \cdots \int_{0 < x_{i_1} < x_{i_2} < \ldots < x_{i_n} < a} dx_1 \ldots dx_m$. Отсюда

$$I_j = \frac{(a_j - a_{j-1})^{k_j}}{k_j!}, \quad 1 \leqslant j \leqslant n.$$

Окончательно имеем

$$I = \left\{ \prod_{j=1}^{n} \frac{(a_j - a_{j-1})^{k_j}}{k_j!} \right\} e^{-\lambda a_n} =$$

$$= \prod_{j=1}^{n} \left\{ \frac{(a_j - a_{j-1})^{k_j}}{k_j!} e^{-\lambda (a_j - a_{j-1})} \right\}. \tag{8.39}$$

Таким образом, случайные величины $x(a_1) - x(a_0), \ldots, x(a_n) - x(a_{n-1})$ независимы и распределены по закону Пуассона с параметрами $\lambda(a_1 - a_0), \ldots, \lambda(a_n - a_{n-1})$.

Пусть теперь $a_0 > 0$. Возьмем $a_0' = 0, \quad a_1' = a_0, \ldots, a_{n+1}' = a_n$. По доказанному, $x(a_1') - x(a_0'), \; x(a_2') - x(a_1'), \ldots, x(a_{n+1}') - x(a_n')$ независимы и распределены по закону Пуассона с параметрами $\lambda(a_1' - a_0'), \ldots, \lambda(a_{n+1}' - a_n')$. Последние n из этих величин также независимы.

Итак,

$$P(x(a_1) - x(a_0) = k_1, \ldots, x(a_n) - x(a_{n-1}) = k_n) =$$

$$= \prod_{j=1}^{n} \left\{ \frac{\lambda (a_j - a_{j-1})^{k_j}}{k_j!} e^{-\lambda (a_j - a_{j-1})} \right\}.$$

Отсюда и следует свойство отсутствия последействия простейшего потока однородных событий.

Свойство стационарности вытекает из того, что в последнюю формулу входят лишь разности $a_j - a_{j-1}$; таким образом, если ко всем a_j добавить некоторое τ, то вид вероятности не изменится.

Остается проверить свойство ординарности. Так как $x(t+h)-x(t)$ обладает распределением Пуассона с параметром λh, то

$$P(x(t+h)-x(t)\geqslant 2)=1-P(x(t+h)-x(t)=0)-$$

$$-P(x(t+h)-x(t)=1)=1-e^{-\lambda h}-\lambda h e^{-\lambda h}=$$

$$=e^{-\lambda h}(e^{\lambda h}-1-\lambda h)=\frac{\lambda^2h^2}{2}+O(h^3); \qquad (8.41)$$

полученное выражение, конечно, представляет собой величину порядка $o(h)$.

Можно показать, что справедливо и обратное утверждение: если имеется некоторая последовательность $\{z_n\}$ со свойствами стационарности, ординарности и отсутствия последействия, то случайные величины $z_1, z_2-z_1, \ldots, z_n-z_{n-1}, \ldots$ независимы к совокупности и распределены по показательному закону с некоторым общим параметром $\lambda > 0$.

Простейший поток однородных событий и процесс Пуассона имеют огромное число приложений. Этими математическими схемами описываются самые различные модели физических явлений. Рассмотрим одно из самых важных свойств простейшего потока.

Пусть на отрезок $(0, T)$ независимым образом случайно бросается N точек $\xi_1, \xi_2, \ldots, \xi_n$ так, что каждая ξ_i равномерно распределена в этом интервале. Тогда если $T\to\infty$, $\frac{N}{T}\to\lambda$, $x_N(t)$ — число точек ξ_i, попавших в интервал $(0, t)$, то для любых $n, a_0, a_1, \ldots, a_n$ выполняется соотношение

$$P(x_N(a_1)-x_N(a_0)=k_1, \ldots, x_N(a_n)-x_N(a_{n-1})=k_n)\xrightarrow[N\to\infty]{}$$

$$\xrightarrow[N\to\infty]{} P(x(a_1)-x(a_0)=k_1, \ldots, x(a_n)-x(a_{n-1})=k_n), \qquad (8.42)$$

где $x(t)$ — процесс Пуассона с параметром λ. Докажем сформулированное свойство. Фиксируем n групп индексов $I_1=\{i_{11}, i_{12}, \ldots, i_{1k_1}\}, \ldots, I_n=\{i_{n1}, \ldots, i_{nk_n}\}$. Подсчитаем вероятность p события B, состоящего в том, что при $i\in I_1$ $\xi_i\in(a_0, a_1)$, при $i\in I_2$ $\xi_i\in(a_1, a_2)$, ... при $i\in I_n$ $\xi_i\in(a_{n-1}, a_n)$, при $i\,\overline{\in}\, I_1\cup\ldots\cup I_n$ $\xi_i\,\overline{\in}\,(a_0, a_n)$. Поскольку ξ_i независимы, а вероятность попадания каждой из них в интервал длины t есть $\frac{t}{T}$, имеем

$$p=\left(\frac{a_1-a_0}{T}\right)^{k_1}\left(\frac{a_2-a_1}{T}\right)^{k_2}\cdots\left(\frac{a_n-a_{n-1}}{T}\right)^{k_n}\left(1-\right.$$

$$\left.-\frac{a_n-a_0}{T}\right)^{N-(k_1+\ldots+k_n)}. \qquad (8.43)$$

Очевидно, если происходит событие B, то также происходит и событие A, состоящее в том, что

$$x_N(a_1) - x_N(a_0) = k_1, \ \ldots, \ x_N(a_n) - x_N(a_{n-1}) = k_n.$$

Однако существуют еще другие события того же типа, что и B, при которых произойдет событие A. Всего таких событий столько, сколькими способами можно из множества индексов $1, \ldots, N$ выбрать группы индексов $I_1, I_2, \ldots, I_n$, т. е.

$$M = C_n^{k_1} C_{n-k_1}^{k_2} \ldots C_{n-(k_1+\ldots+k_{n-1})}^{k_n}. \tag{8.44}$$

Действительно, k_1 индексов, входящих в I_1, можно выбрать C_N^k способами; из оставшихся $N - k_1$ индексов можно выбрать k_2 индексов $C_{n-k_1}^{k_2}$ способами и т. д. Следовательно,

$$P(x_N(a_1) - x_N(a_0) = k_1, \ \ldots, \ x_N(a_n) - x_n(a_{n-1}) = k_n) =$$

$$= C_n^{k_1} C_{n-k_1}^{k_2} \ldots C_{n-(k_1+\ldots+k_{n-1})}^{k_n} \left(\frac{a_1 - a_0}{T}\right)^{k_1} \times$$

$$\times \left(\frac{a_2 - a_1}{T}\right)^{k_2} \ldots \left(\frac{a_n - a_{n-1}}{T}\right)^{k_n} \times \left(1 - \frac{a_n - a_0}{T}\right)^{N-(k_1+\ldots+k_n)}.$$

По формуле (2.34), $C_m^k \sim \dfrac{m^k}{k\,!}$ $(m \to \infty)$. Отсюда

$$C_n^{k_1} \sim \frac{1}{k_1\,!} n^{k_1} \sim \frac{1}{k!} (\lambda T)^{k_1}, \quad C_{n-k_1}^{k_2} \sim \frac{1}{k_2\,!} (n - k_1)^{k_2} \sim$$

$$\sim \frac{1}{k_2\,!} (\lambda T)^{k_2}, \ \ldots, \ C_{n-(k_1+\ldots+k_{n-1})}^{k_n} \sim \frac{1}{k_n\,!} [n -$$

$$- (k_1 + \cdots + k_{n-1})]^{k_n} \sim \frac{1}{k_n!} (\lambda T)^{k_n} \quad (N \to \infty). \tag{8.46}$$

Далее,

$$\left(1 - \frac{a_n - a_0}{T}\right)^{N-(k_1+\ldots+k_n)} =$$

$$= \left(1 - \frac{a_n - a_0}{T}\right)^{\lambda T \cdot \frac{N-(k_1+\ldots+k_n)}{\lambda T}} \sim e^{-\lambda(a_n - a_0)} \quad (N \to \infty). \tag{8.47}$$

Подставив соотношения (8.46) и (8.47) в формулу (8.45), получим

$$P(x_N(a_1) - x_N(a_0) = k_1, \ \ldots, \ x_N(a_n) - x_N(a_{n-1}) = k_n) \xrightarrow[N \to \infty]{}$$

$$\xrightarrow[N\to\infty]{} \prod_{i=1}^{n} [\lambda (a_i - a_{i-1})]^{k_i} e^{-\lambda (a_i - a_{i-1})} / k_i!,$$

что равносильно соотношению (8.42).

В качестве примера рассмотрим функционирование вычислительной машины, управляющей технологическими процессами N производственных объектов. На i-м объекте в случайный момент времени ξ_i, $0 < \xi_i < T$, возникает необходимость регулирования технологического процесса, и в этот момент посылается сигнал к вычислительной машине для решения задачи об оптимальном выборе параметров процесса. При проектировании такой системы возникает вопрос о выборе производительности машины, а это в свою очередь зависит от того, с какой вероятностью в некотором интервале в вычислительную машину поступит недопустимо большое число сигналов. Мы видим, что моменты поступления сигналов при $T \to \infty$, $N/T \to \lambda$, можно представить простейшим потоком однородных событий.

§ 5. ОБОБЩЕННЫЙ ПУАССОНОВСКИЙ ПРОЦЕСС

Пусть $z_1, z_2, \ldots, z_n, \ldots$ — моменты наступления событий простейшего потока однородных событий. (Таким образом, z_1, $z_2 - z_1, \ldots, z_n - z_{n-1}, \ldots$ — независимые случайные величины с плотностью $\lambda e^{-\lambda x}$, $x > 0$.) Пусть $\eta_1, \eta_2, \ldots, \eta_n, \ldots$ — случайные величины с общей функцией распределения $H(x)$. Допустим, что случайные величины $\{z_1, z_2, \ldots; \eta_1, \eta_2, \ldots\}$ независимы в совокупности.

Пусть

$$\xi(t) = \sum_{x_n < t} \eta_n. \tag{8.48}$$

Случайный процесс $\xi(t)$, определяемый этим равенством, называется *обобщенным пуассоновским процессом*. Иначе можно записать

$$\xi(t) = \sum_{n=1}^{x(t)} \eta_n, \tag{8.49}$$

где $x(t)$ — процесс Пуассона с параметром λ.

Основное свойство обобщенного пуассоновского процесса — независимость его приращений в непересекающихся интервалах времени: если $n \geqslant 2$ — любое число, $0 \leqslant a_0 < a_1 < \ldots < a_n$, то случайные величины

$$\zeta_1 = \xi(a_1) - \xi(a_0), \ \ldots, \ \zeta_n = \xi(a_n) - \xi(a_{n-1}) \tag{8.50}$$

независимы в совокупности.

Действительно,

$$\zeta_i = \sum_{a_{i-1} < z_n < a_n} \eta_n. \tag{8.51}$$

По формуле полной вероятности имеем

$$\mathrm{P}(\zeta_i < x_i, \quad 1 \leqslant i \leqslant n) = \sum_{k_0, k_1, \ldots, k_n} p(\nu_i = k_i,$$

$$1 \leqslant i \leqslant n)\, p(\zeta_i < x_i, \quad 1 \leqslant i \leqslant n / \nu_i = k_i, \quad 1 \leqslant i \leqslant n), \quad (8.52)$$

где $\nu_i = x(a_i) - x(a_{i-1})$, $i \geqslant 1$; $\nu_0 = x(a_0)$.

По свойству процесса Пуассона [см. формулу (8.39)],

$$\mathrm{P}(\nu_i = k_i, \quad 1 \leqslant i \leqslant n) = \prod_{i=0}^{n} \frac{[\lambda(a_i - a_{i-1})]^{k_i}}{k_i!} e^{-\lambda(a_i - a_{i-1})}, \quad (8.53)$$

где для определенности положено $a_{-1} = 0$.

Если теперь фиксированы $\nu_0 = k_0$, $\nu_1 = k_1, \ldots, \nu_n = k_n$, то

$$\xi(a_0) = \eta_1 + \cdots + \eta_{k_0}, \quad \xi(a_1) = \eta_1 + \cdots + \eta_{k_0+k_1}, \ldots, \xi(a_n) =$$

$$= \eta_1 + \cdots + \eta_{k_0+\ldots+k_n}. \quad (8.54)$$

Отсюда

$$\zeta_1 = \eta_{k_0+1} + \cdots + \eta_{k_0+k_1}, \quad \zeta_2 = \eta_{k_0+k_1+1} + \cdots +$$

$$+ \eta_{k_0+k_1+k_2}, \ldots, \zeta_n = \eta_{k_0+\ldots+k_{n-1}+1} + \eta_{k_0+\ldots+k_n}. \quad (8.55)$$

Таким образом, при фиксированных $k_0, \ldots, k_n$ ξ_i представляют собой суммы независимых случайных величин (разных для разных i).

Пусть $H^{(m)}(x) = \mathrm{P}(\eta_1 + \ldots + \eta_m < x)$, тогда из формулы (8.55) следует, что

$$\mathrm{P}(\zeta_i < x_i, \quad 1 \leqslant i \leqslant n / \nu_0 = k_0, \ldots, \nu_n = k_n) =$$

$$= H^{(k_1)}(x_1) \ldots H^{(k_n)}(x_n). \quad (8.56)$$

Подставив равенства (8.56), (8.53) в формулу (8.52), найдем

$$\mathrm{P}(\zeta_i < x_i, \quad 1 \leqslant i \leqslant n) =$$

$$= \sum_{k_0=0}^{\infty} \frac{(\lambda a_0)^{k_0}}{k_0!} e^{-\lambda a_0} \prod_{i=1}^{n} \sum_{k_i=0}^{\infty} \frac{[\lambda(a_i - a_{i-1})]^{k_i}}{k_i!} e^{-\lambda(a_i - a_{i-1})} \times$$

$$\times H^{(k_i)}(x_i).$$

Просуммировав по k_0 от 0 до ∞, приходим к формуле

$$\mathrm{P}(\zeta_i < x_i, \quad 1 \leqslant i \leqslant n) =$$

$$= \prod_{i=1}^{n} \frac{[\lambda(a_i - a_{i-1})]^{k_i}}{k_i!} e^{-\lambda(a_i - a_{i-1})} \cdot H^{(k_i)}(x_i), \quad (8.57)$$

откуда следует, что ζ_i независимы в совокупности. Если $a_i - a_{i-1} = h$, $1 \leqslant i \leqslant n$, то ζ_i одинаково распределены.

Положим $n = 1$, $a_0 = 0$, $a_1 = t$, $k_1 = k$. Тогда из формулы (8.57) найдем

$$\mathrm{P}(\xi(t) < x) = \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} H^{(k)}(x). \tag{8.58}$$

Пусть

$$\varphi(\tau) = \int_{-\infty}^{\infty} e^{i\tau x}\, dH(x), \quad \psi(\tau) = \int_{-\infty}^{\infty} e^{i\tau x}\, dF_{\xi(t)}(x). \tag{8.59}$$

Тогда

$$\varphi^k(\tau) = \int_{-\infty}^{\infty} e^{i\tau x}\, dH^{(k)}(x) \tag{8.60}$$

Умножив левую и правую части равенства (8.58) на $e^{i\tau x}$, проинтегрировав от $-\infty$ до ∞ и учтя равенство (8.60), найдем

$$\psi(\tau) = \sum_{k=0}^{\infty} \frac{[\lambda t \varphi(\tau)]^k}{k!} e^{-\lambda t} = \exp\{-\lambda t[1 - \varphi(\tau)]\}. \tag{8.61}$$

Пусть, например, моменты поступления на междугородный переговорный пункт заявок на телефонные разговоры образуют простейший поток однородных требований с параметром λ. Длительности разговоров случайны, независимы и имеют функцию распределения $H(x)$. Тогда характеристическая функция суммарной длины ζ разговоров, заявки на которые поступили за время t, определяется формулой (8.61). Вычислим математическое ожидание этой случайной величины Предположив, что $M\eta_1$ конечно, имеем

$$M\zeta = -i\psi'(0) = -i\lambda t\varphi'(0) = -i\lambda t i M\eta_1 = \lambda t M\eta_1. \tag{8.62}$$

§ 6. СТУПЕНЧАТЫЕ СЛУЧАЙНЫЕ ПРОЦЕССЫ

Пусть X — произвольное конечное или счетное множество. Условимся называть ступенчатой функцией в X функцию времени $x(t)$, $t \geqslant 0$, принимающую постоянные значения из X в каждом из полуинтервалов $[0, a_1)$, $[a_1, a_2)$, ..., $[a_{n-1}, a_n)$, ..., где $\{a_n\}$ — некоторая возрастающая до ∞ последовательность. Ступенчатая функция однозначно определяется заданием последовательности $\{a_n\}$ и последовательности $\{x_n\}$ значений в соответствующих полуинтервалах: $x(t) = x_n$ при $a_{n-1} \leqslant t \leqslant a_n$, $n \geqslant 1$ (для общности положено $a_0 = 0$).

Ступенчатый случайный процесс $\xi(t)$ с множеством значений X определяется как функция двух переменных

$$\xi(t) = \xi(t, \omega) \quad t \geqslant 0, \ \omega \in \Omega \tag{8.63}$$

если задано вероятностное пространство $(\Omega, \mathcal{A}, \mathrm{P})$ и при любом фиксированном $\omega \in \Omega$ функция $\xi(t, \omega_0)$ является ступенчатой функцией со значениями в X. Поскольку ступенчатая функция однозначно задается последовательностями $\{a_n\}$, $\{x_n\}$, вместо функции двух переменных $\xi(t, \omega)$ можно задать две последовательности: $a_n = a_n(\omega)$, $x_n = x_n(\omega)$. Пусть $a_n(\omega)$ — случайные величины, причем $a_n(\omega) \leqslant a_{n+1}(\omega)$ с вероятностью 1, а $x_n(\omega)$ — произвольные функции со значениями в X, с тем лишь условием, что события $\{x_n(\omega) = x\}$ имеют определенные вероятности при любых $x \in X$. Иначе это можно записать так: $\{\omega : x_n(\omega) = x\} \in \mathcal{A}$. Из сделанного предположения следует, что если $\Delta_1, \ldots, \Delta_n$ — любые интервалы, $x_1, x_2, \ldots, x_n$ — любые элементы множества X, то

$$S = \{a_i(\omega) \in \Delta_i,\ \ x_i(\omega) = x_i,\ \ 1 \leqslant i \leqslant n\} \in \mathcal{A}. \tag{8.64}$$

Действительно,

$$S = A_1 \cap \ldots \cap A_n \cap B_1 \cap \ldots \cap B_n,$$

где

$$A_i = \{a_i(\omega) \in \Delta_i\},\ \ B_i = \{x_i(\omega) = x_i\},\ \ 1 \leqslant i \leqslant n.$$

По условию, A_i и B_i имеют определенные вероятности; значит, и их пересечение будет обладать некоторой вероятностью.

Ступенчатый случайный процесс можно задать следующими равносильными способами:

1) заданием набора всех конечномерных функций распределения

$$F_n(t_1, \ldots, t_n;\ z_1, \ldots, z_n) = \mathrm{P}(a_i(\omega) < t_i,\ \ x_i(\omega) = z_i,\ \ 1 \leqslant i \leqslant n),$$
$$n \leqslant 1,\ \ 0 \leqslant t_i < \infty\ \ \ z_i \in X_1,\ \ 1 \leqslant i \leqslant n; \tag{8.65}$$

2) заданием начального распределения

$$F_1(t_1; z_1) = \mathrm{P}(a_1(\omega) < t_1,\ \ x_1(\omega) = z_1),\ \ 0 \leqslant t_1 < \infty,$$
$$z_1 \in X \tag{8.66}$$

и набора условных распределений

$$\Phi(t, z/t_1, \ldots, t_{n-1};\ z_1, \ldots, z_{n-1}) = \mathrm{P}(a_n(\omega) - a_{n-1}(\omega) < t,$$
$$x_n = z/a_i(\omega) = t_i,\ \ x_i = z_i,\ \ 1 \leqslant i \leqslant n-1) \tag{8.67}$$
$$n \geqslant z,\ \ 0 \leqslant t < \infty,\ \ z \in X,\ \ 0 \leqslant t_i < \infty,\ \ z_i \in X,\ \ 1 \leqslant i \leqslant n-1.$$

§ 7. ПОЛУМАРКОВСКИЕ ПРОЦЕССЫ

Полумарковский процесс — это ступенчатый процесс, для которого

$$\Phi(t, z/t_1, \ldots, t_{n-1};\ z_1, \ldots, z_{n-1}) = p_{z_{n-1}, z} F_{z_{n-1}, z}(t),\ \ n \geqslant 2. \tag{8.68}$$

Таким образом, полумарковский процесс задается набором функций $F_1(t, z)$, $F_{z_{n-1},z}(t)$ и набором вероятностей $p_{z'z}$. Обычно в качестве X рассматривают множества целых чисел. Поэтому можно писать $F_1(t,i)$ вместо $F_1(t,z)$, $F_{ij}(t)$ вместо $F_{z_{n-1},z}(t)$ и p_{ij} вместо $p_{z'z}$.

Полумарковские процессы используются для описания функционирования многих реальных систем. Рассмотрим в качестве примера одну схему из теории надежности.

Имеется некоторое устройство, характеризующееся в любой момент времени одним из двух состояний: 0, 1. Состояние 0 интерпретируется как исправное состояние устройства, 1 — как неисправное.

Пусть интервалы пребывания устройства в исправном и неисправном состояниях чередуются, причем все эти интервалы независимы в совокупности и имеют функции распределения $G(x)$ (для исправного состояния) и $H(x)$ (для неисправного). В момент $t = 0$ с вероятностью $G_0(x)$ устройство исправно и до его перехода в неисправное состояние осталось меньше x единиц времени; с вероятностью $H_0(x)$ устройство неисправно и перейдет в исправное состояние раньше, чем через x единиц времени. [Таким образом, вероятность исправного состояния устройства при $t = 0$ равна $G_0(\infty)$.] Обозначим через $\xi(t)$ состояние устройства (0 или 1) в момент t. Тогда $\xi(t)$ будет полумарковским процессом. Его определяющие характеристики в данном случае задаются формулами:

$$F_1(t, 0) = G_0(t), \quad F_1(t, 1) = H_0(t);$$

$$F_{01}(t) = G(t), \quad F_{10}(t) = H(t); \tag{8.69}$$

$$p_{00} = p_{11} = 0, \quad p_{01} = p_{10} = 1.$$

Метод полумарковских процессов является основным методом аналитического решения задач теории массового обслуживания и теории надежности. Для простоты положим $F_{ij}(t) = F_i(t)$.

Приведем способ нахождения одной из основных характеристик полумарковского процесса — вероятности $p_i(t)$ того, что $\xi(t) = i$. Для этого рассмотрим вспомогательный полумарковский процесс $\xi_k(t)$, для которого $F_i(t)$ и p_{ij} те же, что и для процесса $\xi(t)$, а $F_1(t, i)$ определяется формулой $F_1(t, i) = F_k(t)$ при $i = k$, при $i \neq k$ $F_1(t, i) = 0$. Пусть $p_{ki}(t)$ — вероятность события $\{\xi_k(t) = i\}$. Вначале найдем $\{p_{ki}(t)\}$, а затем с их помощью и $\{p_i(t)\}$.

Заметим, что поведение процесса $\xi(t)$ после попадания его в состояние k совпадает с поведением процесса $\xi_k(t)$.

Событие $\{\xi_k(t) = i\}$ может произойти двумя несовместными способами.

1. Может случиться, что в интервале $(0, t)$ процесс $\xi_k(t)$ ни разу не вышел из состояния k. Это событие имеет вероятность $1 - F_k(t)$; но тогда в момент t процесс $\xi_k(t)$ окажется в состоянии i только в том случае, если $i = k$.

2. Может случиться, что в некоторый момент τ, $0 < \tau < t$, процесс вышел из состояния k. [Вероятность того, что это произойдет в интервале $d\tau$, есть $dF_k(\tau)$.] При этом условии процесс $\xi_k(t)$ в момент τ перейдет в состояние j с вероятностью p_{kj}. Если же он перешел в состояние j, то его поведение будет совпадать с поведением процесса $\xi_j(t - \tau)$.

На основании формулы полной вероятности

$$p_{ki}(t) = [1 - F_k(t)]\,\delta_{ik} + \sum_j p_{kj} \int_0^t p_{ji}(t-\tau)\,dF_k(\tau), \quad (8.70)$$

где $\delta_{ik}=1$ при $i = k$, $\delta_{ik} = 0$ при $i \neq k$.

Систему уравнений (8.70) удобнее всего решать методом преобразования Лапласа. Предположим, что функции распределения $F_k(x)$ непрерывно-дискретны, т. е.

$$F_k(x) = \int_{\infty}^{x} f_k(t)\,dt$$

где $f_k(t)$—сумма обычной функции, и не более счетного множества δ-функций. Введем преобразования Лапласа (при $\operatorname{Re} s > 0$):

$$\varphi_k(s) = \int_0^{\infty} e^{-sx} f_k(x)\,dx^*, \quad (8.71)$$

$$\pi_{ki}(s) = \int_0^{\infty} e^{-sx} p_{ki}(x)\,dx \quad (8.72)$$

и заметим что

$$\int_0^{\infty} e^{-sx} [1 - F_k(x)]\,dx = \frac{1 - \varphi_k(s)}{s}. \quad (8.73)$$

Справедливость этого равенства доказывается интегрированием по частям выражения (8.71).

Умножив все части формулы (8.70) на e^{-st} и проинтегрировав от 0 до ∞, получим

$$\pi_{ki}(s) = \frac{1 - \varphi_k(s)}{s}\,\delta_{ik} + \sum_j p_{kj}\,\pi_{ji}(s)\,\varphi_k(s). \quad (8.74)$$

Сделаем следующие предположения:

1) число состояний полумарковского процесса конечно: $k, i = 1, 2, \ldots, m$;

2) $F_k(+0) = 0$, т. е. время пребывания полумарковского процесса в любом состоянии положительно.

---

* Все дальнейшие рассуждения справедливы и в самом общем случае, если $\varphi_k(s)$ понимать в смысле интеграла Стильтьеса:

$$\varphi_k(s) = \int_0^{\infty} e^{-sx}\,dF_k(x).$$

Заметим, что $\varphi_k(s)$ есть характеристическая функция распределения $F_k(x)$ при значении аргумента $t = is$.

Из второго предположения следует, что $\varphi_k(s) \to 0$ при $\operatorname{Re} s \to \infty$. Действительно, для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что $\int\limits_0^\delta f_k(x)\,dx < \varepsilon$. Так как при $\operatorname{Re} s > 0$, $x \geqslant 0$ $|e^{-sx}| \leqslant 1$, то

$$\left|\int\limits_0^\delta e^{-sx} f_k(x)\,dx\right| \leqslant \int\limits_0^\delta |e^{-sx}|\, f_k(x)\,dx \leqslant \int\limits_0^\delta f_k(x)\,dx < \varepsilon.$$

Далее,

$$\left|\int\limits_\delta^\infty e^{-sx} f_k(x)\,dx\right| \leqslant \int\limits_\delta^\infty |e^{-sx}|\, f_k(x)\,dx \leqslant e^{-\operatorname{Re} s\cdot\delta} \int\limits_\delta^\infty f_k(x)\,dx \leqslant e^{-s\delta}.$$

Окончательно имеем

$$|\varphi_k(s)| \leqslant \varepsilon + e^{-\operatorname{Re} s\cdot\delta}. \tag{8.75}$$

Поскольку $\varepsilon > 0$ — произвольное число, то из оценки (8.75) следует, что $\varphi_k(s) \to 0$ при $\operatorname{Re} s \to \infty$.

Определитель системы уравнений (8.74), которая рассматривается как система относительно $\pi_{ki}(s)$ при фиксированном i, имеет вид

$$\Delta = \begin{vmatrix} 1 - p_{11}\varphi_1(s) & -p_{12}\varphi_1(s) & \ldots & -p_{1m}\varphi_1(s) \\ -p_{21}\varphi_2(s) & 1 - p_{22}\varphi_2(s) & \ldots & -p_{2m}\varphi_2(s) \\ \cdots & \cdots & \cdots & \cdots \\ -p_{m1}\varphi_m(s) & -p_{m2}\varphi_m(s) & \ldots & 1 - p_{mm}\varphi_m(s) \end{vmatrix} \tag{8.76}$$

По доказанному, диагональные элементы этого определителя при $\operatorname{Re} s \to \infty$ стремятся к 1, а недиагональные — к 0. Отсюда следует, что при $\operatorname{Re} s > d$, где d — некоторое конечное число $\Delta \neq 0$, а стало быть, система (8.74) имеет единственное решение. При этом $\pi_{ki}(s)$ имеют вид дробно-рациональных функций $\{\varphi_k(s)\}$

Рассмотрим функции $p_i(t)$. Событие $\{\xi(t) = i\}$ может произойти двумя несовместными способами:

1) $\xi(0) = i$ и за время от 0 до t процесс $\xi(t)$ ни разу не вышел из состояния i. Вероятность этого события равна $F_1(\infty, i) - F_1(t, i)$.

2) $\xi(0) = j$, где j — любое состояние полумарковского процесса; в некоторый момент τ произошел выход процесса из состояния j. [Вероятность такого события в интервале $d\tau$ есть $dF_1(\tau, j)$.] После этого с вероятностью p_{jk} процесс $\xi(t)$ перешел в состояние k. Процесс $\xi_k(t)$ в момент $t - \tau$ находится в состоянии i.

По формуле полной вероятности

$$p_i(t) = F_1(\infty, i) - F_1(t, i) + \sum_j \int\limits_0^t \sum_k p_{jk} p_{ki}(t - \tau)\, dF_1(\tau, j). \tag{8.77}$$

Умножим обе части равенства (8.77) на e^{-st} ($\operatorname{Re} s > 0$) и проинтегрируем по t от 0 до ∞. Пусть

$$\pi_i(s) = \int_0^\infty e^{-st} p_i(t)\,dt \tag{8.78}$$

и

$$\int_0^\infty e^{-st}\,dF_1(t,\ j) = \psi_j(s); \tag{8.79}$$

тогда, учитывая равенства

$$\int_0^\infty e^{-st}[F_1(\infty,\ i) - F_1(t,i)]\,dt = \frac{F_1(\infty, i) - \psi_i(s)}{s}, \tag{8.80}$$

получим

$$\pi_i(s) = \frac{F_1(\infty,\ i) - \psi_i(s)}{s} + \sum_j \sum_k p_{jk}\pi_{ki}(s)\,\psi_j(s). \tag{8.81}$$

В правой части равенства (8.81) находятся заданные функции и функции, определенные выше, в левой — функции $\pi_i(s)$, представляющие собой преобразования Лапласа искомых функций $p_i(t)$. Воспользовавшись обратным преобразованием Лапласа можно найти и сами эти функции.

Отметим, что если распределения $F_i(t)$, $\dfrac{F_1(t,\ i)}{F_1(\infty,\ i)}$ являются гиперэрланговскими, то функции $\varphi_i(s)$ и $\psi_i(s)$ являются дробно-рациональными функциями s; тогда и функции $\pi_i(s)$ также будут дробно-рациональными, что значительно облегчает нахождение функций по их преобразованиям Лапласа.

§ 8. МАРКОВСКИЙ ПРОЦЕСС С КОНЕЧНЫМ МНОЖЕСТВОМ СОСТОЯНИЙ

Пусть X — конечное множество. Для упрощения обозначений будем считать, что $X = \{1, 2, \ldots, m\}$. Случайный процесс с множеством значений X, как было определено выше, является функцией двух переменных:

$$\xi(t) = \xi(t,\ \omega),\quad t \geqslant 0,$$

где $\omega \in \Omega$, $(\Omega,\ \mathcal{U},\ \mathsf{P})$ — вероятностное пространство. Ступенчатые случайные процессы были определены нами как процессы, для которых при некоторых последовательностях $\{a_n\}$ и $\{x_n\}$, зависящих от ω и таких, что $a_n \xrightarrow[n\to\infty]{} \infty$ с вероятностью 1, $\xi(t,\ \omega) = x_n$ при $a_{n-1} < t < a_n$, $n \geqslant 1$, где $a_0 = 0$. Согласно принятому условию события $\{x_n = i\}$ и $\{a_n < x\}$ имеют определенные вероятности при любых $i = 1, 2, \ldots, m$, $0 \leqslant x < \infty$.

Предположим, что для любого отрезка $\Delta = [a,\ b]$ и любого i, $1 \leqslant i \leqslant m$, $A_i(\Delta)$ есть событие, состоящее в том, что $\xi(t) = i$

при всех $t \in \Delta$. Покажем, что событие $A_i(\Delta)$ имеет определенную вероятность. В самом деле, отрезок Δ покрывается некоторым конечным набором полуинтервалов $[a_{n-1}, a_n)$, $n \in I$, для каждого из которых $a_{n-1} < a_n$ и каждый из которых имеет с Δ хотя бы одну общую точку. Тогда $A_i(\Delta)$ имеет место в том и только том случае, если $x_n = i$ для всех $n \in I$. Поскольку множество I можно выбрать не более чем счетным числом способов, и каждый из этих способов имеет определенную вероятность, так, например $P(I = \{n\}) = P(a_n \leqslant a,\ b < a_{n+1})$, а условная вероятность $A_i(\Delta)$ при фиксированном I есть $P(x_i = i,\ n \in I)$, то $A_i(\Delta)$ имеет определенную вероятность. В частности, положив $\Delta = \{t\}$, найдем, что вероятность события $\{\xi(t) = i\}$ всегда определена.

Пусть L_t — множество событий, каждое из которых есть результат конечного или счетного числа операций объединения, пересечения и взятия дополнения над событиями $A_j(\Delta)$), где $\Delta \subset (0, t)$, U_t — аналогичным образом определенное множество событий с условием $\Delta \subset (t, \infty)$ вместо $\Delta \subset (0, t)$.

Ступенчатый процесс $\xi(t)$ называется марковским, если для любых $t > 0$, $i = 1, 2, \ldots, m$, любых $A \in L_t$, $B \subset U_t$

$$P(AB/\xi(t) = i) = P(A/\xi(t) = i)\, P(B/\xi(t) = i), \tag{8.82}$$

если только $P(\xi(t) = i) > 0$.

Короче можно сказать, что при фиксированном «настоящем» ($\xi(t)$) будущее поведение процесса ($\{\xi(\tau),\ \tau > t\}$) не зависит от прошлого ($\{\xi(\tau),\ \tau < u\}$).

Формулу (8.82) можно представить иначе. Разделим обе ее части на $P(A/\xi(t) = i)$, предположив, что $P(A/\xi(t) = i) > 0$:

$$\frac{P(AB/\xi(t) = i)}{P(A/\xi(t) = i)} = P(B/\xi(t) = i),$$

или, что то же,

$$P(B/\xi(t) = i,\ A) = P(B/\xi(t) = i). \tag{8.83}$$

Итак, если фиксировано значение $\xi(t)$, то любая дополнительная информация о поведении процесса до момента t не может повлиять на вероятностные характеристики, связанные с его поведением после момента t.

Теорема 8.4. *Если ступенчатый случайный процесс $\xi(t)$ является марковским процессом, то последовательность (a_1, x_1), (a_2, x_2), ..., (a_n, x_n), ... является однородной цепью Маркова.*

Доказательство. Пусть известно, что $a_n = t$, $x_n = i$. Тогда событие $a_{n+1} < x$, $x_{n+1} = j$ при любых x и j зависит от $\xi(\tau)$, $\tau > t$, в то время как события $a_k \in \Delta_k$, $x_k = j'_k$, $k < n$, относятся к поведению процесса до момента t (ведь $a_k < a_n = t$ при $k < n$). Поэтому

$$P(a_{n+1} < x,\ x_{n+1} = j/a_n = t,\quad x_n = i;\quad a_k \in \Delta_k,$$

$$x_k = j'_k,\quad k < n) = P(a_{n+1} < x,\quad x_{n+1} = j/a_n = t,\ x_n = i), \tag{8.84}$$

Следовательно, последовательность (a_n, x_n) представляет собой цепь Маркова. Докажем, что эта цепь Маркова однородна.

Заметим, что событие $\{a_{n+1} < x,\ x_{n+1} = j\}$ при условии, что

$a_n = t$, входит в U_t, так что его вероятность не должна зависеть от поведения процесса до момента t. Поскольку указание n является информацией относительно поведения процесса до момента t, то вероятность рассматриваемого события при фиксированных $a_n = t$, $x_n = i$ от n не зависит, а это и есть свойство однородности цепи Маркова. Теорема доказана.

Следствие. Пусть $a_1 < a_2 < \ldots < a_n < \ldots$ — моменты смены состояний ступенчатого марковского процесса, x_n — значение процесса в полуинтервале $[a_{n-1}, a_n)$, $n \geqslant 1$, где $a_0 = 0$. Тогда совокупность всех конечномерных распределений a_n, x_n, $n \geqslant 1$, полностью задается функциями

$$F_{ij}(t, \tau) = \mathrm{P}(a_{n+1} < t + \tau, \quad x_{n+1} = j / a_n = t, \quad x_n = i). \quad (8.85)$$

Заметим, что эти функции не зависят от n, $n \geqslant 0$. Пользуясь формулой полной вероятности, можно записать

$$P(a_{n+1} < t + \tau, \quad x_{n+1} = j / a_n = t, \quad x_n = i) =$$

$$= \int_t^{t+\tau} \mathrm{P}(x_{n+1} = j / a_n = t, \quad x_n = i, \quad a_{n+1} = u)\, dG(u/t, i), \quad (8.86)$$

где

$$G(u/t, i) = P(a_{n+1} < u / a_n = t, x_n = i). \quad (8.87)$$

Исследуем функцию $G(u/t, i)$. Заметим, что функция $G(t + h/t, i)$, $1 < i < m$ является условной вероятностью выхода процесса $\xi(t)$ из состояния i в интервале $(t, t + h)$ при условии, что $\xi(t) = i$. Предположим, что

$$G(t + h/t, i) = \lambda_i(t) h + o(h) \quad (h \to 0), \quad (8.88)$$

где $\lambda_i(t)$ — интегрируемые в любом конечном интервале функции. Тогда

$$1 - G(u + h/t, i) = [1 - G(u/t, i)]\,[1 - G(u + h/u, i)]. \quad (8.89)$$

Действительно, в левой части этого равенства имеем вероятность невыхода процесса из состояния i в интервале $(t, u + h)$, а в правой — произведение вероятности невыхода за время от t до u на вероятность невыхода в интервале $(u, u + h)$. Так как $\xi(t)$ — марковский процесс, то при фиксированном $\xi(u)$ [именно при $\xi(u) = i$] второе событие не зависит от первого. Подставив равенство (8.88) в формулу (8.89), имеем

$$1 - G(u + h/t, i) = [1 - G(u/t, i)] \quad [1 - \lambda_i(u) h + o(h)]. \quad (8.90)$$

или (после приведения подобных членов, деления на h и устремления h к нулю)

$$[1 - G(u/t, i)]' = -\lambda_i(u)[1 - G(u/t, i)] \quad u \geqslant t. \quad (8.91)$$

Решение дифференциального уравнения (8.91) с учетом очевидного условия

$$1 - G(t/t, i) = 1 \quad (8.92)$$

имеет вид

$$1 - G(u/t,\ i) = \exp\left\{-\int_t^u \lambda_i(x)\,dx\right\},\quad u \geqslant t. \tag{8.93}$$

Итак, условная функция распределения $G(u/t,\ i)$ случайной величины a_{n+1} при условии, что $a_n = t, x_n = i$, задается формулой

$$G(u/t,\ i) = 1 - \exp\left\{-\int_t^u \lambda_i(x)\,dx\right\},\quad u \geqslant t. \tag{8.94}$$

Функция $\lambda_i(t)$ (если она существует) называется *интенсивностью выхода* процесса $\xi(t)$ из состояния i.

Рассмотрим функцию $\mathrm{P}(x_{n+1} = j/a_n = t,\ x_n = i,\ a_{n+1} = u)$, $t < u$. По определению ступенчатого процесса, из того, что $x_n = i$ следует, что $\xi(\tau) = i$ при всех τ, $t \leqslant \tau < u$. Поэтому при любом τ $t \leqslant \tau \leqslant u$,

$$\mathrm{P}(x_{n+1} = j/a_n = t,\quad x_n = i,\quad a_{n+1} = u) = \mathrm{P}(x_{n+1} =$$
$$= j/\xi(\tau) = i,\quad a_n = t,\quad x_n = i,\quad a_{n+1} = u). \tag{8.95}$$

Однако при условии, что $\xi(\tau) = i$, т. е. при фиксированном значении процесса в момент τ, событие $\{a_n = t,\ x_n = i\}$ относится к поведению процесса до момента τ, а значит, согласно определению марковского процесса может быть из числа условий исключено.

Итак,

$$\mathrm{P}(x_{n+1} = j/a_n = t,\quad x_n = i,\quad a_{n+1} = u) = \mathrm{P}(x_{n+1} = j/x_n = i,$$
$$a_{n+1} = u). \tag{8.96}$$

Пусть

$$\mathrm{P}_{ij}(u) = \mathrm{P}(x_{n+1} = j/x_n = i,\quad a_{n+1} = u). \tag{8.97}$$

Тогда, учитывая равенства (8.94) и (8.97), формулу (8.86) можно записать так:

$$\mathrm{P}(a_{n+1} < t + \tau,\quad x_{n+1} = j/a_n = t,\quad x_n = i) =$$
$$= \int_t^{t+\tau} \lambda_i(u) P_{ij}(u) \exp\left\{-\int_t^u \lambda_i(x)\,dx\right\} du,\quad \tau > 0. \tag{8.98}$$

Функция

$$\lambda_{ij}(t) = \lambda_i(t) P_{ij}(t) \tag{8.99}$$

называется *интенсивностью перехода* процесса $\xi(t)$ из состояния i в состояние j.

Пусть γ_t — первый после момента t момент изменения состояния процесса $\xi(t)$. Докажем следующую формулу:

$$\mathrm{P}(\gamma_t < t + \tau,\ \xi(\gamma_t) = j/\xi(t) = i) =$$

$$= \int_t^{t+\tau} \lambda_{ij}(u) \exp\left\{-\int_t^u \lambda_i(x)\,dx\right\} du, \quad \tau > 0. \tag{8.100}$$

Имеем

$$P(\gamma_t < t+\tau, \quad \xi(\gamma_t) = j / \xi(t) = i) = P(t < a_{n+1} < t+\tau,$$
$$x_{n+1} = j / a_n \leqslant t < a_{n+1}, \quad x_n = i) = P(a_{n+1} < t+\tau,$$
$$x_{n+1} = j / a_n = t, \quad x_n = i).$$

Подставив в полученное выражение формулу (8.98), придем к формуле (8.100).

Докажем следующее важное равенство:

$$P(\xi(t+h) = j / \xi(t) = i) = \lambda_{ij}(t)\,h + o(h) \ (h \to 0), \quad i \neq j, \tag{8.101}$$

если $\lambda_{ij}(t)$ непрерывна в данной точке t. Заметим, что при условии $\xi(t) = i$ событие $\{\xi(t+h) = j\}$ может произойти двумя способами:

1) $\gamma_t < t+h$, $\xi(\gamma_t) = j$; 2) за время от t до $t+h$ процесс изменит свое состояние по меньшей мере дважды.

Пусть $\psi_i(t, h)$ — вероятность события B, состоящего в том, что за время от t до $t+h$ произойдет хотя бы два изменения состояний процесса $\xi(t)$ при условии, что $\xi(t) = i$. Если произойдет событие B, то произойдут следующие события: $\gamma_t < t+h$, $\gamma_{t+\gamma_t} < t+h$. Из формулы (8.100) следует, что если

$$\lambda(x) = \max_{1 \leqslant i \leqslant m} \lambda_i(x),$$

то

$$P(\gamma_t < t+\tau) \leqslant \int_t^{t+\tau} \lambda(u)\,du$$

при любом фиксированном состоянии процесса в момент t. Отсюда

$$P(\gamma_t < t+h, \quad \gamma_{t+\gamma_t} < t+h) \leqslant \left[\int_t^{t+h} \lambda(u)\,du\right]^2 = o(h)$$

при $h \to 0$. Следовательно, $\psi_i(t+h) = o(h)$. Однако если $\gamma_t < t+h$, $\xi(\gamma_t) = j$, то $\xi(t+h) \neq j$ лишь в том случае, если произойдет событие B; с другой стороны, $P(\xi(t+h) = j, \xi(\gamma_t \neq j) / \xi(t) = i) \leqslant \psi_i(t, \ t+h)$. Окончательно имеем

$$P(\xi(t+h) = j / \xi(t) = i) =$$
$$= \int_t^{t+h} \lambda_{ij}(u) \exp\left\{-\int_t^u \lambda_i(x)\,dx\right\} du + o(h) = \lambda_{ij}(t)\,h + o(h),$$

что и требовалось доказать.

Пусть

$$p_{ij}(t_0, t) = P(\xi(t) = j \mid \xi(t_0) = i), \quad t_0 < t. \tag{8.102}$$

Функции вида (8.102) являются основными характеристиками марковского процесса с конечным множеством состояний. Эти функции определяются из уравнений, к выводу которых мы приступаем.

По формуле полной вероятности можно записать

$$p_{ij}(t_0\ \ t+h) = \sum_{k=1}^{m} p_{ik}\,(t_0, t)\, p_{kj}\,(t, t+h) =$$

$$= p_{ij}(t_0,\ t)\, p_{jj}\,(t,\ t+h) + \sum_{k\neq j} p_{kj}\,(t,\ t+h). \qquad (8.103)$$

Нами было выяснено, что $p_{kj}(t,\ \ t + h) = \lambda_{kj}(t)\, h + o\,(h)$ при $k \neq j$. Так как $\sum\limits_{l=1}^{m} p_{jl}(t,\ t + h) = 1$, то

$$p_{jj}(t,\ t+h) = 1 - \sum_{l\neq j} p_{jl}\,(t+h) = 1 - \sum_{l\neq j} \lambda_{jl}\,(t)\, h + o\,(h) =$$

$$= 1 - \lambda_j\,(t)\, h + o\,(h) \qquad (8.104)$$

Подставив равенства (8.101) и (8.104) в формулу (8.103), найдем

$$p_{ij}(t_0,\ t+h) = p_{ij}\,(t_0,\ h)\,[1 - \lambda_j\,(t)\, h] + \sum_{k\neq j} \lambda_{kj}(t)\, p_{ik}(t_0,\ t) h + o(h),$$

или

$$\frac{p_{ij}(t_0,\ t+h) - p_{ij}\,(t_0,\ t)}{h} + \lambda_j(t) p_{ij}(t_0,\ t) = \sum_{k\neq j} \lambda_{kj}(t) p_{ik}(t_0,\ t) + o\,(1).$$

Устремив h к нулю, придем к уравнениям

$$\frac{\partial\, p_{ij}\,(t_0,\ t)}{\partial\, t} + \lambda_j(t) p_{ij}(t_0,\ t) = \sum_{k\neq j} \lambda_{kj}(t) p_{ik}(t_0,\ t). \qquad (8.105)$$

Уравнения (8.105) называются *дифференциальными уравнениями Колмогорова — Феллера*. Система уравнений (8.105) в случае непрерывных $\lambda_{ij}(t)$ при $t > t_0$ обладает единственным решением при условиях

$$p_{ij}\,(t_0,\ t_0) = \begin{cases} 1 \text{ при } i = j, \\ 0 \text{ при } i \neq j. \end{cases} \qquad (8.106)$$

Марковский процесс с конечным множеством состояний называется *однородным*, если функции $\lambda_{ij}(t)$ существуют и не зависят от t: $\lambda_{ij}(t) = \lambda_{ij}$. Для однородных марковских процессов система уравнений (8.105) представляет собой систему обыкновенных дифференциальных уравнений (t_0 в них входит как параметр) с постоянными коэффициентами. Эта система решается методом преобразования Лапласа.

Как следует из формулы (8.98), время пребывания однородного марковского процесса в состоянии i есть показательно распределенная случайная величина с параметром λ_i; при выходе из состояния i процесс принимает значение j с вероятностью $p_{ij} = \frac{\lambda_{ij}}{\lambda_{i1} + \ldots + \lambda_{im}} = \frac{\lambda_{ij}}{\lambda_i}$. С другой стороны, как следует из формулы (8.100), если процесс в данный момент времени находится в состоянии i, то время до следующего изменения состояния имеет показательное распределение с параметром λ_i; при этом вероятность перехода из состояния i в состояние j также равна $p_{ij}= \frac{\lambda_{ij}}{\lambda_i}$.

Легко доказать также, что однородный марковский процесс с конечным множеством состояний является в то же время полумарковским процессом.

ГЛАВА IX

КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

Если две случайные величины ξ и η не являются независимыми, то, как известно, они называются *зависимыми случайными величинами*. При этом зависимость между величинами ξ и η не является, вообще говоря, функциональной и носит *вероятностный* (*стохастический*) характер. Такая зависимость изучается методами теории вероятностей.

Изучению стохастической зависимости случайных величин и посвящена настоящая глава.

§ 1. КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Наиболее важными и простыми характеристиками степени зависимости двух случайных величин являются корреляционный момент (или ковариация этих величин) и их коэффициент корреляции, уже рассмотренные в гл. V.

В настоящем параграфе эти характеристики будут рассмотрены более подробно.

Согласно определению, корреляционный момент $\mu_{\xi,\eta}$ случайных величин ξ и η равен

$$\mu_{\xi,\eta} = M(\xi - M\xi)(\eta - M\eta).$$

Корреляционный момент $\mu_{\xi,\eta}$ величин ξ и η в случае, если величина (ξ, η) дискретна, и в случае существования плотности $p_{\xi,\eta}(x, y)$ распределения вероятностей величины (ξ, η) вычисляется соответственно по формулам:

$$\mu_{\xi,\eta} = \sum_{i,j} (x_i - M\xi)(y_j - M\eta)\, p_{ij},$$

$$\mu_{\xi,\eta} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x - M\xi)(y - M\eta)\, p_{\xi,\eta}(x, y)\, dx dy,$$

где (x_i, y_j) — возможные значения величины (ξ, η); p_{ij} — вероятности этих значений $(p_{ij} = \mathrm{P}(\xi = x_i, \eta =$

$= y_j$). Корреляционный момент используют как меру зависимости случайных величин, что оправдывается следующей теоремой.

Теорема 9.1. *Корреляционный момент* $\mu_{\xi,\eta}$ *двух независимых случайных величин* ξ *и* η *равен нулю, т. е.* $\mu_{\xi,\eta} = 0$, если величины ξ и η независимы.

Доказательство. По условию, величины ξ и η независимы, поэтому являются независимыми величины $\xi_1 = \xi - M\xi$ и $\eta_1 = \eta - M\eta$. Заметим, что $M\xi_1 = M(\xi - M\xi) = M\xi - M\xi = 0$ и $M\eta_1 = 0$. По определению, $\mu_{\xi,\eta} = M(\xi_1, \eta_1)$; согласно теореме о математическом ожидании произведения независимых величин имеем

$$M_{\xi,\eta} = M(\xi_1 \eta_1) = M\xi_1 \cdot M\eta_1 = 0,$$

что и требовалось доказать.

Утверждение, обратное утверждению теоремы 9.1, вообще говоря, неверно, т. е., если корреляционный момент равен нулю, то величины не обязательно независимы. Приведем пример таких величин. Пусть

$$g(x) = \begin{cases} e^{-x} & \text{при } x \geqslant 0, \\ 0 & \text{при } x < 0. \end{cases}$$

Функция $g(x)$ является плотностью распределения вероятностей некоторой случайной величины, имеющей конечную дисперсию. Положим

$$p(x, y) = \frac{g\left(\sqrt{x^2 + y^2}\right)}{2\pi\sqrt{x^2 + y^2}}.$$

Функция $p(x, y)$ является плотностью распределения некоторой двумерной случайной величины (ξ, η), так как $p(x, y) \geqslant 0$ и

$$\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} p(x, y)\, dy = \int_0^{\infty} d\rho \int_0^{2\pi} d\varphi \, \frac{g(\rho)}{2\pi\rho} \cdot \rho = 1.$$

Так определенная плотность постоянна на любой окружности с центром в начале координат, причем, по симметрии соответствующей двумерной плотности относительно начала координат, $M\xi = M\eta = 0$ (что можно проверить и аналитически). В силу симметрии распределения, корреляционный момент величин ξ и η равен нулю: $\mu_{\xi,\eta}=0$.

Из существования плотности распределения вероятностей двумерной случайной величины вытекает существование плотностей распределения компонент. В рассматриваемом случае по соображениям симметрии можно заключить, что эти плотности одинаковы:

$$p_\xi(x)=p_\eta(x).$$

Как известно (см. гл. V), для независимости величин ξ и η необходимо и достаточно, чтобы выполнялось следующее условие:

$$p_{\xi,\eta}(x, y)=p_\xi(x)\cdot p_\eta(y).$$

Предположим, что это условие выполнено, и положим $x=y$, тогда имеем

$$p_{\xi,\eta}(x, x)=e^{-|x|\sqrt{2}}\cdot\frac{1}{2\pi|x|\sqrt{2}}=p_\xi(x)\cdot p_\eta(x)=p_\xi^2(x),$$

откуда

$$p_\xi(x)=\frac{1}{\sqrt{2\pi}\sqrt{|x|}\sqrt[4]{2}}\cdot e^{-\frac{|x|}{\sqrt{2}}}$$

и

$$p_{\xi,\eta}(x, y)=\frac{e^{-\sqrt{x^2+y^2}}}{2\pi\sqrt{x^2+y^2}}\neq\frac{1}{2\pi\sqrt{|x|}\sqrt{|y|}\sqrt{2}}\times$$

$$\times e^{-\frac{|x|+|y|}{\sqrt{2}}}=p_\xi(x)\cdot p_\eta(y).$$

Таким образом, мы пришли к противоречию. Следовательно, рассматриваемые величины ξ и η, для которых $\mu_{\xi,\eta}=0$, зависимы.

Величина корреляционного момента $\mu_{\xi,\eta}$ зависит от выбора единиц измерения величин ξ и η, поэтому использовать эту характеристику не всегда удобно. От этого недостатка свободна характеристика, называемая *коэффициентом корреляции*. Как известно (см. гл. V), коэффициентом корреляции $\rho_{\xi,\eta}$ случайных величин ξ и η называется отношение их корреляционного момента к произведению их средних квадратичных отклонений:

$$\rho_{\xi,\eta}=\frac{\mu_{\xi,\eta}}{\sigma_\xi\cdot\sigma_\eta}.$$

Величина коэффициента корреляции не зависит от выбора единиц измерения величин ξ и η, т. е. $\rho_{\xi,\eta}=\rho_{c\xi,d\eta}$ ($c>0$, $d>0$ — некоторые постоянные). В самом деле,

$$\rho_{c\xi,d\eta}=\frac{\mu_{c\xi,d\eta}}{\sigma_{c\xi}\cdot\sigma_{d\eta}}=\frac{cd\mu_{\xi,\eta}}{c\sigma_\xi\cdot d\sigma_\eta}=\rho_{\xi,\eta}.$$

Из теоремы 9.1 вытекает важное следствие.

Следствие. Коэффициент корреляции независимых случайных величин ξ и η равен нулю, т. е. $\rho_{\xi,\eta}=0$, если эти величины независимы.

Случайные величины ξ и η называются *некоррелированными*, если их коэффициент корреляции $\rho_{\xi,\eta}=0$. Таким образом, из независимости следует некоррелированность, а обратное же утверждение, как показывает приведенный пример, неверно.

В частном случае при $\xi=\eta$ имеем

$$\mu_{\xi,\xi}=D\xi \quad \text{и} \quad \rho_{\xi,\xi}=1.$$

Таким образом, коэффициент корреляции величины ξ с самой собой равен единице. Следует ожидать, что среди всех величин η величина $\eta=\xi$ имеет наибольший коэффициент корреляции с величиной ξ, т. е. $\rho_{\xi,\eta}\leqslant\rho_{\xi,\xi}=1$ так как среди всех величин η величина $\eta=\xi$ «наибольшим образом» зависит от величины ξ. Имеет место следующая теорема.

Теорема 9.2. *Коэффициент корреляции $\rho_{\xi,\eta}$ величин ξ и η не превосходит по абсолютной величине единицы, т. е.*

$$|\rho_{\xi,\eta}|\leqslant 1.$$

Доказательство. Рассмотрим величину

$$M[(\xi-M\xi)+t(\eta-M\eta)]^2=D\xi+2t\mu_{\xi,\eta}+t^2D\eta.$$

Левая часть равенства неотрицательна при всех t, поэтому квадратный трехчлен, находящийся в правой части равенства, также неотрицателен при всех t. Следовательно, его дискриминант

$$(\mu_{\xi,\eta})^2-D\xi\cdot D\eta\leqslant 0,$$

или

$$|\mu_{\xi,\eta}|\leqslant\sqrt{D\xi\cdot D\eta},$$

Разделив обе части неравенства на $\sqrt{D\xi \cdot D\eta}$, получим

$$|\rho_{\xi,\eta}| \leqslant 1.$$

Теорема доказана.

Рассмотрим отдельно экстремальные случаи ($\rho_{\xi,\eta} = \pm 1$).

Теорема 9.3. *Коэффициент корреляции величин ξ и η $\rho_{\xi,\eta} = \pm 1$ тогда и только тогда, когда с вероятностью единица между величинами ξ и η существует точная функциональная зависимость, являющаяся линейной.*

Доказательство. Очевидно, что $\rho_{\xi,\eta} = \pm 1$ тогда и только тогда, когда квадратный трехчлен, рассмотренный при доказательстве теоремы 9.2, имеет равные вещественные корни $t_1 = t_2 = t_0$. При t_0 имеем

$$M\,[(\xi - M\xi) + t_0(\eta - M\eta)]^2 = 0.$$

Следовательно, с вероятностью 1

$$\xi - M\xi + t_0(\eta - M\eta) = 0,$$

или

$$\xi = t_0\eta + M\xi - t_0 M\eta. \quad (\ast)$$

Линейная функциональная зависимость всегда может быть задана в виде ($\ast$) при некотором t_0, что означает существование равных вещественных корней $t_1 = t_2 = t_0$ квадратного трехчлена и, следовательно, справедливость равенства $|\rho_{\xi,\eta}| = 1$. Теорема доказана.

В силу теоремы 9.3 коэффициенты корреляции $\rho_{\xi,\eta}$ можно рассматривать как *меру прямолинейности* распределения двумерной величины (ξ, η). В самом деле, мера, отвечающая двумерному распределению величины (ξ, η) при $\rho_{\xi,\eta} = \pm 1$ (и только в этом случае), сосредоточена на одной прямой, уравнение которой имеет вид

$$x = t_0 y + M\xi - t_0 M\eta.$$

Введенные понятия и полученные результаты допускают геометрическую интерпретацию в терминах бесконечномерного евклидова пространства*.

---

* Все используемые в настоящей главе сведения из курса линейной алгебры, в частности из теории евклидовых пространств, можно найти в кн.: А. Н. Рублев Линейная алгебра. М., «Высшая школа», 1968.

Рассмотрим совокупность всех случайных величин $\xi(\omega)$, определенных на одном и том же пространстве элементарных событий Ω, имеющих нулевое математическое ожидание $M\xi = 0$ и конечные дисперсии $D\xi < \infty$. Нетрудно проверить, что эта совокупность является линейным пространством с операциями сложения и умножения на число, понимаемыми в обычном смысле.

Элементами линейного пространства являются случайные величины, которые, следуя принятой терминологии, будем называть векторами. Введем в этом пространстве скалярное произведение

$$(\xi, \eta) = M(\xi\eta) = \mu_{\xi, \eta}.$$

В силу доказанного нами неравенства $|\rho_{\xi, \eta}| \leqslant 1$ имеем

$$|(\xi, \eta)| = |\mu_{\xi, \eta}| \leqslant \sigma_\xi \cdot \sigma_\eta,$$

поэтому (ξ, η) существует для любой пары случайных величин, принадлежащих нашему пространству.

Линейное пространство с определенным подобным образом скалярным произведением становится евклидовым, так как при этом все аксиомы евклидова пространства выполняются.

Корреляционный момент $\mu_{\xi, \eta}$, по определению, равен скалярному произведению (ξ, η) векторов ξ и η. Дисперсия

$$D\xi = \|\xi\|^2,$$

где через $\|\xi\|$ обозначена длина вектора ξ, а среднее квадратичное отклонение σ_ξ совпадает с длиной вектора ξ:

$$\sigma_\xi = \|\xi\|.$$

Коэффициент корреляции $\rho_{\xi, \eta}$, равный отношению скалярного произведения векторов ξ и η к произведению их длин:

$$\rho_{\xi, \eta} = \frac{\mu_{\xi, \eta}}{\sigma_\xi \cdot \sigma_\eta} = \frac{(\xi, \eta)}{\|\xi\| \cdot \|\eta\|},$$

оказывается равным косинусу угла между векторами ξ и η, т. е.

$$\rho_{\xi, \eta} = \frac{(\xi, \eta)}{\|\xi\|\,\|\eta\|} = \cos\varphi,$$

Некоррелированность случайных величин при этом означает ортогональность векторов ξ и η и из теоремы 9.1 следует, что из независимости случайных величин ξ и η вытекает их ортогональность как векторов евклидова пространства. Доказанное нами неравенство $|\rho_{\xi,\eta}| \leqslant 1$ переходит в неравенство

$$|\cos\varphi| \leqslant 1 \quad \text{или} \quad |(\xi, \eta)| \leqslant \|\xi\| \cdot \|\eta\|.$$

Последнее неравенство является неравенством Коши—Буняковского для векторов ξ и η евклидова пространства. Заметим, что по существу доказательство теоремы 9.2 и есть вывод неравенства Коши—Буняковского для произвольного евклидова пространства. Это неравенство в различных евклидовых пространствах имеет различные конкретные формы (сравните с известными из анализа неравенствами

$$\left|\int_a^b f(x)\,g(x)\,dx\right| \leqslant \sqrt{\int_a^b f^2(x)\,dx \cdot \int_a^b g^2(x)\,dx},$$

или

$$\left|\sum_{k=1}^{n} a_k b_k\right| \leqslant \sqrt{\sum_{k=1}^{n} a_k^2 \cdot \sum_{k=1}^{n} b_k^2}\Biggr).$$

В теории вероятностей это неравенство принимает следующий вид: $|\rho_{\xi,\eta}| \leqslant 1$. При этом условие $\rho_{\xi,\eta} = \pm 1$ означает, что угол φ между векторами ξ и η равен 0 или π, т. е. векторы коллинеарны и направлены в одну сторону, если $\rho_{\xi,\eta} = 1$, и в противоположные стороны, если $\rho_{\xi,\eta} = -1$. Как известно из элементарной геометрии, коллинеарные векторы отличаются друг от друга числовым множителем. Поэтому естественно, что в этом случае $\xi = t_0\eta$ (напомним, что $M\xi = M\eta = 0$). Существование такой зависимости между ξ и η и устанавливает теорема 9.3.

§ 2. ЛИНЕЙНАЯ РЕГРЕССИЯ

Пусть (ξ, η) — двумерная случайная величина, например ξ и η — сигналы на входе и выходе некоторого устройства. Во многих случаях бывает важно уметь представить одну из величин, например η, как функцию $g(\xi)$ другой величины. Точное представление $\eta = g(\xi)$, как правило, невозможно, потому возникает задача о возможности приближенного представления: $\eta \approx g(\xi)$. Итак, требуется найти среди всех функций $g(\xi)$ величины ξ такую, которая являлась бы *наилучшим приближением величины* η. Термину «наилучшее приближение» следует придать точный смысл, что можно сделать различными способами. Самым удобным и общепринятым является так называемое приближение *по методу наименьших квадратов*.

Определение. *Величина* $g(\xi)$ *называется наилучшим приближением величины* η *в смысле метода наименьших квадратов, если* $M[\eta - g(\xi)]^2$ *принимает наименьшее возможное значение; при этом величина* $g(\xi)$ *называется средней квадратической регрессией величины* η *на величину* ξ.

Ограничимся нахождением линейной средней квадратической регрессии величины η на величину ξ, т. е. линейной функции $g(\xi) = \alpha + \beta\xi$, которая наилучшим образом среди всех линейных функций величины ξ приближает величину η.

Без доказательства отметим, что в случае нормального закона распределения двумерной случайной величины (ξ, η) средняя квадратическая регрессия η на ξ (так же как и ξ на η) является линейной функцией от ξ (соответственно от η). Таким образом, минимальное значение величины $M[\eta - g(\xi)]^2$ достигается в этом случае для некоторой линейной функции $g(\xi)$, которая, очевидно, является линейной средней квадратической регрессией η на ξ.

Пусть $m_1 = M\xi$, $m_2 = M\eta$; $\sigma_1^2 = D\xi$, $\sigma_2^2 = D\eta$; $\mu = \mu_{\xi, \eta}$ — корреляционный момент величин ξ и η; $\rho = \rho_{\xi, \eta}$ — коэффициент корреляции величин ξ и η. Имеет место следующая теорема.

Теорема 9.4. *Линейная средняя квадратическая регрессия* $g(\xi)$ *величины* η *на величину* ξ *имеет вид*:

$$g(\xi) = m_2 + \rho\frac{\sigma_2}{\sigma_1}(\xi - m_1),$$

или

$$\frac{g(\xi) - m_2}{\sigma_2} = \rho\frac{\xi - m_1}{\sigma_1}.$$

Доказательство. Для доказательства найдем такие α и β, при которых величина $\Phi(\alpha, \beta) = M(\eta - \alpha - \beta\xi)^2$ принимает наименьшее из возможных значений. Имеем

$$\Phi(\alpha, \beta) = M(\eta - \alpha - \beta\xi)^2 = M[(\eta - m_2) - \beta(\xi - m_1) + m_2 - \alpha - \beta m_1]^2 = \sigma_2^2 + \beta^2\sigma_1^2 - 2\mu\beta + (m_2 - \alpha - \beta m_1)^2.$$

Исследуем функцию $\Phi(\alpha, \beta)$ на экстремум:

$$\frac{\partial \Phi}{\partial \alpha} = -2(m_2 - \alpha - \beta m_1) = 0, \quad \frac{\partial \Phi}{\partial \beta} = 2\beta\sigma_1^2 - 2\mu = 0,$$

откуда

$$\beta = \frac{\mu}{\sigma_1^2} = \rho\frac{\sigma_2}{\sigma_1}, \quad \alpha = m_2 - \beta m_1 = m_2 - \rho\frac{\sigma_2}{\sigma_1}m_1.$$

Нетрудно убедиться, что при найденных величинах α и β функция $\Phi(\alpha, \beta)$ действительно принимает наименьшее значение. Таким образом, величина

$$g(\xi) = \alpha + \beta\xi = m_2 - \rho\frac{\sigma_2}{\sigma_1}m_1 + \rho\frac{\sigma_2}{\sigma_1}\xi = m_2 + \rho\frac{\sigma_2}{\sigma_1}(\xi - m_1)$$

является линейной регрессией величины η на величину ξ, что и требовалось доказать.

Определение. *Коэффициент* $\beta = \rho\frac{\sigma_2}{\sigma_1}$ *линейной регрессии* $\alpha + \beta\xi$ *величины* η *на величину* ξ *называется коэффициентом регрессии величины* η *на величину* ξ, *а прямая, определяемая уравнением*

$$\frac{y - m_2}{\sigma_2} = \rho\frac{x - m_1}{\sigma_1},$$

называется прямой регрессии (или прямой средней квадратичной регрессии) величины η *на величину* ξ.

Наименьшее значение величины $M(\eta - \alpha - \beta\xi)^2$, которое достигается при найденных нами значениях α и β, как нетрудно подсчитать, таково:

$$\min_{\alpha, \beta} M(\eta - \alpha - \beta\xi)^2 = \sigma_2^2(1 - \rho^2).$$

Эта величина называется *остаточной дисперсией величины* η *относительно* ξ. Остаточная дисперсия измеряет (в соответствии с принципом наименьших квадратов) величину ошибки, допускаемой при использовании приближенного равенства $\eta \approx g(\xi) = \alpha + \beta\xi$.

Остаточная дисперсия показывает, насколько можно уменьшить дисперсию величины η, вычитая из нее линейную функцию величины ξ.

Разность η_1 между величиной η и линейной регрессией η на ξ называется *остатком величины* η *относительно* ξ:

$$\eta_1 = \eta - \alpha - \beta\xi = \eta - m_2 - \rho \frac{\sigma_2}{\sigma_1} (\xi - m_1).$$

Подсчитаем корреляционный момент величин η_1 и ξ:

$$\mu_{\xi, \eta_1} = M(\xi - m_1)\left(\eta - m_2 - \rho \cdot \frac{\sigma_2}{\sigma_1} (\xi - m_1\right) =$$

$$= M(\xi - m_1)(\eta - m_2) - \rho \frac{\sigma_2}{\sigma_1} M(\xi - m_1)^2 = \mu_{\xi, \eta} - \rho \sigma_1 \sigma_2 = 0.$$

Таким образом, остаток величины η относительно ξ оказывается некоррелированным с величиной ξ. Остаточная дисперсия является математическим ожиданием квадрата остатка $M\eta_1^2$.

Мы рассмотрели линейную регрессию величины η на величину ξ. Аналогично можно найти линейную регрессию величины ξ на величину η:

$$h(\eta) = m_1 + \rho \frac{\sigma_1}{\sigma_2} (\eta - m_2),$$

при этом коэффициент регрессии ξ на η равен $\rho \frac{\sigma_1}{\sigma_2}$, а прямая линейной регрессии ξ на η определяется уравнением

$$\frac{y - m_2}{\sigma_2} = \frac{1}{\rho} \frac{x - m_1}{\sigma_1}.$$

Остаточная дисперсия величины ξ при этом равна

$$\min_{\alpha, \beta} M(\xi - \alpha - \beta\eta)^2 = \sigma_1^2 (1 - \rho^2),$$

Обе прямые регрессии проходят через точку (m_1, m_2) и никогда не совпадают, за исключением случаев $\rho = \pm 1$. В этих частных случаях, как известно, существует линейная зависимость между величинами ξ и η. Уравнение, задающее эту зависимость, является одновременно уравнением линейной регрессии η на ξ и ξ на η. Обе остаточные дисперсии в этом случае равны нулю. В случае, если величины ξ и η некоррелированы, т. е. $\rho = 0$, регрессия η на ξ (соответст-

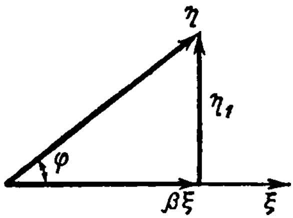

Рис. 9.1

венно ξ на η) принимает вид $g(\xi) = m_2$ [соответственно $h(\eta) = m_1$].

Введенные понятия и полученные результаты можно интерпретировать геометрически. Ограничимся рассмотрением случайных величин ξ таких, что $M\xi = 0$, $D\xi < \infty$. Совокупность таких величин (заданных на одном и том же пространстве элементарных исходов) является, как было отмечено в § 1 настоящей главы, евклидовым пространством.

Пусть ξ и η — два вектора этого пространства, т. е. две такие случайные величины, что

$$M\xi = M\eta = 0,$$

$$D\xi = \|\xi\|^2 < \infty, \quad D\eta = \|\eta\|^2 < \infty.$$

Рассмотрим задачу о проектировании вектора η на направление вектора ξ (рис. 9.1), т. е. задачу о представлении вектора η в виде

$$\eta = \beta\xi + \eta_1,$$

где $\eta_1 \perp \xi$. Умножая обе части последнего равенства скалярно на ξ, имеем

$$(\xi, \eta) = \beta(\xi, \xi) + (\xi, \eta_1) = \beta(\xi, \xi),$$

откуда находим

$$\beta = \frac{(\xi, \eta)}{(\xi, \xi)} = \frac{\mu_{\xi, \eta}}{\sigma_\xi^2} = \frac{\rho\sigma_1\sigma_2}{\sigma_1^2} = \rho\frac{\sigma_2}{\sigma_1}.$$

Величина β оказалась равной коэффициенту регрессии η на ξ. Вектор $\beta\xi$ при найденном значении β называется *проекцией вектора* η *на направление вектора* ξ.

Вектор $\eta_1 = \eta - \rho\frac{\sigma_2}{\sigma_1}\xi$ ортогонален к вектору ξ и является перпендикуляром, опущенным из конца η на ξ. Как известно из курса алгебры, в произвольном евклидовом пространстве перпендикуляр короче любой наклонной, поэтому среди всех векторов вида $\eta - \beta\xi$ вектор $\eta_1 = \eta - \rho\frac{\sigma_2}{\sigma_1}\xi$ имеет минимальную длину. Итак, величина $\|\eta_1\|^2 = M(\eta_1)^2 = M(\eta - \beta\xi)^2$ минимальна при $\beta = \rho\cdot\frac{\sigma_2}{\sigma_1}$.

Учитывая сделанное предположение о том, что $M\xi = m_1 = 0$, $M\eta = m_2 = 0$, приходим к следующему выводу:

линейная регрессия η *на* ξ *является проекцией вектора* η *на направление вектора* ξ. Перпендикуляр $\eta_1 = \eta - \beta\xi$, являющийся разностью между вектором η и его проекцией на направление ξ, ортогонален к вектору ξ, что соответствует установленной нами ранее некоррелированности величин η_1 и ξ. Перпендикуляр η_1 оказывается при этом остатком величины η относительно величины ξ. По теореме Пифагора, справедливой, как известно, в произвольном евклидовом пространстве, имеем

$$\|\eta\|^2 = \|\eta_1\|^2 + \|\beta\xi\|^2,$$

откуда

$$\|\eta_1\|^2 = \|\eta\|^2 - \beta^2\|\xi\|^2 = \sigma_2^2 - \rho^2 \frac{\sigma_2^2}{\sigma_1^2}\sigma_1^2 = \sigma_2^2(1-\rho^2).$$

Мы получили остаточную дисперсию, которая, таким образом, оказывается равной квадрату длины перпендикуляра, опущенного из конца η на направление ξ. Величина ошибки, допускаемой при использовании приближенного равенства $\eta \approx \rho \frac{\sigma_2}{\sigma_1}\xi$, измеряется длиной этого перпендикуляра.

§ 3. МНОЖЕСТВЕННАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

В настоящем параграфе мы рассмотрим обобщение на n-мерный случай результатов, полученных в § 2 настоящей главы.

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ — произвольные случайные величины, имеющие конечные математические ожидания $m_i = M\xi_i$ и дисперсии $\sigma_i^2 = D\xi_i$ $(i = 1, 2 \ldots n)$. Вычитая из случайной величины ее математическое ожидание, получим новую случайную величину $\xi_i - m_i$, для которой $M(\xi_i - m_i) = 0$, а дисперсия по-прежнему равна σ_i^2. Поэтому в дальнейшем для упрощения изложения будем предполагать, что $m_i = 0$ $(i = 1, 2, \ldots, n)$. Все формулы в общем случае могут быть получены из формул данного параграфа заменой ξ_i на $\xi_i - m_i$. Поставим перед собой задачу найти такую линейную функцию

$$g(\xi_2, \xi_3, \ldots, \xi_n) = \beta_{12.34\ldots n}\xi_2 + \beta_{13.24\ldots n}\xi_3 + \ldots + \beta_{1n.23\ldots(n-1)}\xi_n,$$

которая наилучшим образом в смысле метода наименьших квадратов давала бы приближение величины ξ_1. Таким образом, следует так подобрать коэффициенты $\beta_{12.34\ldots n}$, $\beta_{13.24\ldots n}, \ldots, \beta_{1n.23\ldots(n-1)}$, чтобы величина

$$M\left(\xi_1-\beta_{12.34\ldots n}\xi_2-\beta_{13.24\ldots n}\xi_3-\ldots-\beta_{1n.23\ldots(n-1)}\xi_n\right)^2 \quad (9.1)$$

была минимальной. При таком выборе коэффициентов функция $g(\xi_2, \xi_3, \ldots, \xi_n)$ дает наилучшую линейную оценку величины ξ_1 и называется *линейной* (*множественной*) *регрессией величины* ξ_1 *на величины* $\xi_2, \xi_3, \ldots, \xi_n$. Коэффициенты $\beta_{12.34\ldots n}$, $\beta_{13.24\ldots n}$, $\beta_{1n.23\ldots(n-1)}$ называются *коэффициентами множественной регрессии*, *или* (*частными*) *коэффициентами регрессии*. Каждый из этих коэффициентов имеет n индексов, причем первые два отделяются от остальных точкой; первый индекс указывает номер величины, для которой рассматривается регрессия, второй индекс— номер величины, при которой стоит этот коэффициент, а остальные индексы указывают на номера величин, входящих в уравнение регрессии. В некоторых случаях мы будем опускать индексы, стоящие после точки: так, вместо $\beta_{12.34\ldots n}$ будем писать β_{12} и т. д.

Для нахождения коэффициентов регрессии приравняем к нулю частные производные величины (9.1) по каждому из $(n-1)$ коэффициентов, в результате получаем систему вида

$$\begin{aligned}
&\mu_{22}\beta_{12}+\mu_{23}\beta_{13}+\ldots+\mu_{2n}\beta_{1n}=\mu_{21},\\
&\mu_{32}\beta_{12}+\mu_{33}\beta_{13}+\ldots+\mu_{3n}\beta_{1n}=\mu_{31},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\mu_{n2}\beta_{12}+\mu_{n3}\beta_{13}+\ldots+\mu_{nn}\beta_{1n}=\mu_{n1}.
\end{aligned} \quad (9.2)$$

Здесь через μ_{ij} обозначен корреляционный момент величин ξ_i и ξ_j, т. е. $\mu_{ij}=M\xi_i\xi_j$ (напомним, что в нашем случае $M\xi_i=M\xi_j=0$). Если определитель

$$D_{11}=\begin{vmatrix}\mu_{22} & \mu_{23} & \cdots & \mu_{2n}\\ \mu_{32} & \mu_{33} & \cdots & \mu_{3n}\\ \cdot & \cdot & \cdot & \cdot\\ \mu_{n2} & \mu_{n3} & \cdots & \mu_{nn}\end{vmatrix}\neq 0,$$

то, как известно из алгебры, система имеет, и притом единственное, решение, которое можно получить или обычным методом последовательного исключения неизвестных, или по правилу Крамера. Без доказательства отметим, что определитель $D_{11} \neq 0$, если распределение n-мерной случайной величины $(\xi_1, \xi_2, \ldots, \xi_n)$ является *собственным*, т. е. если в n-мерном пространстве не существует подпространства меньшей размерности, содержащего всю распределенную меру. Можно доказать, что распределение n-мерной случайной величины $(\xi_1, \xi_2, \ldots \xi_n)$ является собственным тогда и только тогда, когда

$$D = \begin{vmatrix} \mu_{11} \; \mu_{12} \cdots \mu_{1n} \\ \mu_{21} \; \mu_{22} \cdots \mu_{2n} \\ \cdot \;\; \cdot \;\; \cdot \;\; \cdot \;\; \cdot \;\; \cdot \\ \mu_{n1} \; \mu_{n2} \cdots \mu_{nn} \end{vmatrix} \neq 0.$$

При $n = 2$ двумерное распределение сосредоточено на прямой в том случае, если коэффициент корреляции $\rho = \pm 1$ (в этих случаях двумерное распределение оказывается несобственным). В дальнейшем мы будем предполагать, что распределение n-мерной случайной величины $(\xi_1, \xi_2, \ldots, \xi_n)$ — собственное. Тогда $D_{11} \neq 0$ и система (9.2) имеет единственное решение, которое находится по формулам

$$\beta_{1k} = -\frac{D_{1k}}{D_{11}} \; (k = 2, 3, \ldots n), \qquad (9.3)$$

где через D_{ij} обозначено алгебраическое дополнение элемента μ_{ij} определителя D.

При $n = 2$ имеем

$$\beta_{12} = -\frac{-\sigma_1 \sigma_2 \rho}{\sigma_2^2} = \frac{\sigma_1}{\sigma_2} \rho,$$

что согласуется с результатами, полученными в § 2.

Очевидно, что во всех рассуждениях величину ξ_1 можно заменить любой из величин $\xi_2, \xi_3, \ldots, \xi_n$.

Таким образом, в случае собственного распределения n-мерной случайной величины $(\xi_1, \xi_2, \ldots, \xi_n)$ задача нахождения линейной регрессии любой из величин $\xi_1, \xi_2, \ldots, \xi_n$ на остальные имеет, и притом единственное, решение. Если величины $\xi_1, \xi_2, \ldots, \xi_n$ некоррелированы (т. е. $\mu_{ij} = 0$ при $i \neq j$), то, как следует из системы (9.2), все коэффициенты регрессии равны нулю.

Разность $\eta_{1.23\ldots n}$ между величиной ξ_1 и ее линейной регрессией на величины $\xi_2, \xi_3, \ldots, \xi_n$

$$\eta_{1.23\ldots n}=\xi_1-\beta_{12}\xi_2-\beta_{13}\xi_3-\ldots-\beta_{1n}\xi_n$$

называется *остатком величины* ξ_1 *относительно величин* $\xi_2, \xi_3, \ldots, \xi_n$. Покажем, что остаток некоррелирован ни с одной из величин $\xi_2, \xi_3, \ldots, \xi_n$. В самом деле, из первого уравнения системы (9.2) следует, что корреляционный момент

$$M(\eta_{1.23\ldots n}\ \xi_2)=M(\xi_1-\beta_{12}\xi_2-\beta_{13}\xi_3-\ldots-\beta_{1n}\xi_n)\xi_2=$$
$$=\mu_{12}-\beta_{12}\mu_{22}-\mu_{13}\mu_{23}-\ldots-\beta_{1n}\mu_{2n}=0.$$

Аналогично устанавливается некоррелированность остатка с величинами $\xi_3, \xi_4, \ldots, \xi_n$. Система уравнений (9.2) эквивалентна условию некоррелированности разности $\xi_1-\beta_{12}\xi_2-\beta_{13}\xi_3-\ldots-\beta_{1n}\xi_n$ со всеми величинами ξ_2, $\xi_3, \ldots, \xi_n$. Таким образом, требование обращения в минимум дисперсии этой разности и требование ее некоррелированности со всеми величинами $\xi_2, \xi_3, \ldots, \xi_n$ эквивалентны.

Математическое ожидание квадрата остатка $\eta_{1.23\ldots n}$ величины ξ_1 относительно величин $\xi_2, \xi_3, \ldots, \xi_n$

$$M\eta^2_{1.23\ldots n}=\sigma^2_{1.23\ldots n}$$

называется *остаточной дисперсией величины* ξ_1 *относительно величин* $\xi_2, \xi_3, \ldots, \xi_n$. В силу некоррелированности остатка со всеми величинами $\xi_2, \xi_3, \ldots, \xi_n$ для остаточной дисперсии имеем

$$\sigma^2_{1.23\ldots n}=M\eta^2_{1.23\ldots n}=M(\xi_1-\beta_{12}\xi_2-$$
$$-\ldots-\beta_{1n}\xi_n)\eta_{1.23\ldots n}=M\xi_1\cdot\eta_{1.23\ldots n}=\mu_{11}-\beta_{12}\mu_{12}-$$
$$-\beta_{13}\mu_{13}-\ldots-\beta_{1n}\mu_{1n}. \tag{9.4}$$

Заменяя по формулам (9.3) коэффициенты регрессии их выражениями и используя теорему о разложении определителя по элементам строки, имеем

$$\sigma^2_{1.23\ldots n}=\mu_{11}+\frac{D_{12}}{D_{11}}\mu_{12}+\frac{D_{13}}{D_{11}}\mu_{13}+\ldots+\frac{D_{1n}\mu_{1n}}{D_{11}}=$$
$$=\frac{D}{D_{11}}. \tag{9.5}$$

При $n=2$ остаточная дисперсия

$$\sigma_{1.2}^2 = \frac{\begin{vmatrix} \sigma_1^2 & \rho\,\sigma_1\,\sigma_2 \\ \rho\,\sigma_1\,\sigma_2 & \sigma_2^2 \end{vmatrix}}{\sigma_2^2} = \sigma_1^2\,(1-\rho^2),$$

что согласуется с результатами, полученными в § 2.

Рассмотрим величину ξ_1 и ее линейную регрессию на величины $\xi_2, \xi_3, \ldots, \xi_n$.

$$\xi_1^* = \beta_{12}\,\xi_2 + \beta_{13}\,\xi_3 + \ldots + \beta_{1n}\,\xi_n.$$

Коэффициент корреляции величин ξ_1 и ξ_1^* можно рассматривать как *меру зависимости между величиной* ξ_1 *и совокупностью величин* $\xi_2, \xi_3, \ldots, \xi_n$. Коэффициент корреляции $\rho_{\xi_1,\,\xi_1^*}$ называется *сводным коэффициентом корреляции между величиной* ξ_1 *и совокупностью величин* $\xi_2, \xi_3, \ldots, \xi_n$ и обозначается так:

$$\rho_{1.\,(23\,\ldots\,n)} = \frac{M\,(\xi_1\,\xi_1^*)}{\sqrt{M\xi_1^2 \cdot M\,(\xi_1^*)^2}}.$$

Опуская индексы 2, 3...n у остатка $\eta_{1.23\,\ldots\,n}$, имеем $\xi_1 = \xi_1^* + \eta_1$, откуда в силу формул (9.4) и (9.5) получаем

$$M\,(\xi_1\,\xi_1^*) = M\,\xi_1\,(\xi_1 - \eta_1) = \mu_{11} - \frac{D}{D_{11}}.$$

Аналогично находим

$$M\,(\xi_1^*)^2 = M\,(\xi_1^2 - 2\xi_1\,\eta_1 + \eta_1^2) = \mu_{11} - \frac{D}{D_{11}},$$

и сводный коэффициент корреляции

$$\rho_{1\,(23\,\ldots\,n)} = \sqrt{1 - \frac{D}{\mu_{11}\,D_{11}}}.$$

Для сводного коэффициента корреляции, являющегося коэффициентом корреляции двух величин ξ_1 и ξ_1^*, справедливы все известные свойства коэффициента корреляции, в частности $-1 \leqslant \rho_{1\,(23\,\ldots\,n)} \leqslant 1$.

Можно показать, что сводный коэффициент корреляции обладает одним экстремальным свойством, а именно: среди всех коэффициентов корреляции величины ξ_1 с линейными функциями величин $\xi_2, \xi_3, \ldots, \xi_n$ наибольшим оказывается коэффициент корреляции ξ_1 и ее линейной регрессии ξ_1^* на $\xi_2, \xi_3, \ldots, \xi_n$, т. е. сводный коэффициент корреляции $\rho_{1\,(23\ldots n)}$.

Рассмотрим линейную регрессию величин ξ_1 и ξ_2 на величины $\xi_3, \xi_4, \ldots, \xi_n$. Учитывая принятые обозначения, имеем:

$$\xi_1 = \beta_{13.45\ldots n}\,\xi_3 + \beta_{14.35\ldots n}\,\xi_4 + \ldots + \beta_{1n.34\ldots(n-1)}\,\xi_n + \\ + \eta_{1.34\ldots n},$$

$$\xi_2 = \beta_{23.45\ldots n}\,\xi_3 + \beta_{24.35\ldots n}\,\xi_4 + \ldots + \beta_{2n.34\ldots(n-1)}\,\xi_n + \\ + \eta_{2.34\ldots n}.$$

Коэффициент корреляции между остатками $\eta_{1.34\ldots n}$ и $\eta_{2.34\ldots n}$ можно рассматривать как характеристику корреляции величин ξ_1 и ξ_2 после вычитания из них наилучших линейных оценок величинами $\xi_3, \xi_4, \ldots \xi_n$. Этот коэффициент корреляции называется *частным коэффициентом корреляции величин* ξ_1 и ξ_2 *относительно величин* $\xi_3, \xi_4, \ldots \xi_n$ и обозначается так:

$$\rho_{12.34\ldots n} = \rho_{\eta_{1.34\ldots n},\ \eta_{2.34\ldots n}}.$$

Частный коэффициент корреляции, являющийся обычным коэффициентом корреляции двух величин, удовлетворяет неравенству $-1 \leqslant \rho_{12.34\ldots n} \leqslant 1$. Для получения частного коэффициента корреляции воспользуемся формулами (9.5) применительно к системам $(n-1)$ величин $\xi_1, \xi_3, \ldots \xi_n$ и $\xi_2, \xi_3, \ldots \xi_n$. Имеем:

$$M\left(\eta_{1.34\ldots n}^2\right) = M\left(\xi_1 \cdot \eta_{1.34\ldots n}\right) = \frac{D_{22}}{D_{22.11}},$$

$$M\left(\eta_{2.34\ldots n}^2\right) = M\left(\xi_2 \cdot \eta_{2.34\ldots n}\right) = \frac{D_{11}}{D_{11.22}}. \qquad (9.6)$$

Здесь, как и ранее, через D_{11} и D_{22} обозначены алгебраические дополнения элементов μ_{11} и μ_{22} определителя D, а через $D_{11.22}$ или $D_{22.11}$ — алгебраическое дополнение элемента μ_{22} в определителе D_{11}. В силу некоррелированности остатка $\eta_{2.34\ldots n}$ со всеми величинами $\xi_3, \xi_4, \ldots, \xi_n$ по формулам (9.3) применительно к $(n-1)$ величинам $\xi_2, \xi_3, \ldots,$

ξ_n имеем

$$M(\eta_{1.34\ldots n}\cdot\eta_{2.34\ldots n}) \doteq M(\xi_1\cdot\eta_{2.34\ldots n}) = M\,\xi_1(\xi_2 -$$

$$-\beta_{23.4\ldots n}\xi_3 - \beta_{24.3\ldots n}\xi_4 - \ldots \beta_{2n.3\ldots(n-1)}\xi_n) = \mu_{12}\frac{D_{11.22}}{D_{11.22}} +$$

$$+\mu_{13}\frac{D_{11.23}}{D_{11.22}} + \mu_{14}\frac{D_{11.24}}{D_{11.22}} + \ldots + \mu_{1n}\frac{D_{11.2n}}{D_{11.22}} = -\frac{D_{21}}{D_{11.22}} =$$

$$= -\frac{D_{12}}{D_{11.22}}, \tag{9.7}$$

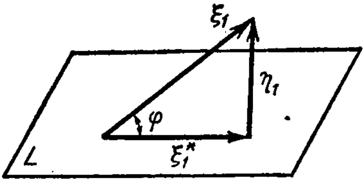

Рис. 9.2

где через $D_{11.2k}$ обозначено алгебраическое дополнение элемента μ_{2k} в определителе D_{11}. Мы воспользовались теоремой о разложении определителя D_{21} по элементам первой строки; $D_{12} = D_{21}$, так как $\mu_{ij} = \mu_{ji}$, т. е. матрица вторых моментов симметрична.

Используя равенства (9.6) и (9.7), получаем выражение для частного коэффициента корреляции:

$$\rho_{12.34\ldots n} = -\frac{D_{12}}{\sqrt{D_{11}D_{22}}}. \tag{9.8}$$

Если величины $\xi_1, \xi_2, \ldots, \xi_n$ некоррелированы, то из последнего равенства следует, что все частные коэффициенты корреляции равны нулю. Если же величины $\xi_1, \xi_2, \ldots, \xi_n$ коррелированы, то коэффициент корреляции ρ_{ξ_1,ξ_2}, вообще говоря, отличен от частного коэффициента корреляции этих величин $\rho_{12.34\ldots n}$.

Введенные в этом параграфе понятия и полученные результаты можно интерпретировать геометрически, аналогично тому, как это было сделано в § 2.

Рассмотрим вектор ξ_1 и совокупность всевозможных линейных комбинаций векторов $\xi_2, \xi_3, \ldots, \xi_n$, которая называется *линейной оболочкой этих векторов*. Поставим перед собой задачу спроектировать вектор ξ_1 на линейную оболочку L векторов $\xi_2, \xi_3, \ldots, \xi_n$, т. е. задачу о возможности представления вектора ξ_1 в виде

$$\xi_1 = \beta_{12}\xi_2 + \beta_{13}\xi_3 + \ldots + \beta_{1n}\xi_n + \eta_1, \qquad (9.9)$$

где $\xi_1^* = \beta_{12}\xi_2 + \beta_{13}\xi_3 + \ldots \beta_{1n}\xi_n \in L$, а $\eta_1 \perp L$, т. е. вектор η_1 ортогонален ко всем элементам линейной оболочки L (рис. 9.2).

Для ортогональности вектора η_1 к L достаточно, чтобы вектор η_1 был ортогонален к векторам $\xi_2, \xi_3, \ldots, \xi_n$. Вектор ξ_1^* называется *проекцией вектора* ξ_1 *на линейную оболочку* L, а вектор η_1 — *перпендикуляром, опущенным из конца* ξ_1 *на* L. Умножим скалярно обе части равенства (9.9) последовательно на $\xi_2, \xi_3, \ldots, \xi_{11}$. В результате получим систему для определения коэффициентов $\beta_{12}, \beta_{13}, \ldots, \beta_{1n}$.

$$\begin{aligned}
\mu_{12} &= \beta_{12}\mu_{22} + \beta_{13}\mu_{32} + \ldots + \beta_{1n}\mu_{2n},\\
\mu_{13} &= \beta_{12}\mu_{23} + \beta_{13}\mu_{33} + \ldots + \beta_{1n}\mu_{3n},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
\mu_{1n} &= \beta_{12}\mu_{2n} + \beta_{13}\mu_{3n} + \ldots + \beta_{1n}\mu_{nn}.
\end{aligned}$$

Мы снова получили систему (9.2). Таким образом, коэффициенты искомой линейной комбинации, являющейся *проекцией* вектора ξ_1 на линейную оболочку L векторов $\xi_2, \xi_3, \ldots, \xi_n$, оказываются коэффициентами регрессии величины ξ_1 на величины $\xi_2, \xi_3, \ldots, \xi_n$. Величина ξ_1^*, являющаяся *регрессией величины* ξ_1 *на величины* $\xi_2, \xi_3, \ldots, \xi_n$, *оказывается, таким образом, проекцией вектора* ξ_1 *на линейную оболочку векторов* $\xi_2, \xi_3, \ldots, \xi_n$. Остаток η_1 является перпендикуляром, опущенным из конца вектора ξ_1 на оболочку L. Перпендикуляр короче любой наклонной, поэтому величина $M(\eta_1)^2 = M(\xi_1 - \beta_{12}\xi_2 - \ldots - \beta_{1n}\xi_n)^2$ минимальна в случае, если коэффициенты $\beta_{12}, \beta_{13}, \ldots, \beta_{1n}$ являются коэффициентами регрессии вектора ξ_1 на векторы $\xi_2, \xi_3, \ldots \xi_n$. Таким образом, остаточная дисперсия оказывается равной квадрату длины перпендикуляра η_1. По теореме Пифагора,

$$\|\xi_1\|^2 = \|\xi_1^*\|^2 + \|\eta_1\|^2,$$

откуда

$$\|\eta_1\|^2 = \sigma^2_{1.23\ldots n} = \mu_{11} - \|\xi_1^*\|^2.$$

Сводный коэффициент корреляции, равный коэффициенту корреляции между величинами ξ_1 и ξ_1^*, равен косинусу угла φ между векторами ξ_1 и ξ_1^*. Как известно из алгебры, среди всех линейных комбинаций векторов $\xi_2, \xi_3, \ldots, \xi_n$ наименьший угол φ с вектором ξ_1 образует проекция ξ_1^*

вектора ξ_1 на линейную оболочку L этих векторов. При этом $\cos\varphi$ будет наибольшим (что и было отмечено выше как экстремальное свойство сводного коэффициента корреляции).

Частный коэффициент корреляции $\rho_{12.34\ldots n}$ величин ξ_1 и ξ_2 относительно $\xi_3, \xi_4, \ldots, \xi_n$, равный коэффициенту корреляции остатков $\eta_{1.34\ldots n}$ и $\eta_{2.34\ldots n}$, равен косинусу угла между перпендикулярами, опущенными из концов векторов ξ_1 и ξ_2 на линейную оболочку векторов $\xi_3, \xi_4, \ldots, \xi_n$.

Упражнения к главе IX

1. Пусть ξ и η — случайные величины, имеющие конечные дисперсии. Доказать, что $D(\xi+\eta) = D\xi + D\eta$ тогда и только тогда, когда величины ξ и η некоррелированы.

2. Случайные величины ξ и η независимы и нормально распределены с параметрами a и σ. Найти коэффициент корреляции величин $\xi_1 = \alpha\xi + \beta\eta$ и $\xi_2 = \alpha\xi - \beta\eta$.

3. Случайные величины ξ и η связаны соотношением $\alpha\xi + \beta\eta = \gamma$, где α, β и γ неслучайны, $\alpha \neq 0$, $\beta \neq 0$. Найти коэффициент корреляции $\rho_{\xi,\eta}$, а также коэффициенты регрессии η на ξ и ξ на η.

4. Пусть совместное распределение величин ξ и η нормально, причем $M\xi = M\eta = 0$, а коэффициент корреляции этих величин равен ρ. Показать, что коэффициент корреляции величин ξ^2 и η^2 равен ρ^2.

5. Случайные величины $\xi_1, \xi_2, \ldots, \xi_{n+m}$ $(n > m)$ независимы, одинаково распределены и имеют конечную дисперсию. Найти коэффициент корреляции между суммами $\zeta_1 = \xi_1 + \xi_2 + \ldots + \xi_n$ и $\zeta_2 = \xi_{m+1} + \xi_{m+2} + \ldots + \xi_{m+n}$.

6. Случайный вектор (ξ, η) равномерно распределен в квадрате со стороной, равной единице, и диагоналями, совпадающими с осями координат. Найти коэффициент корреляции величин ξ и η.

7. Показать, что остаток $\eta_{1.23\ldots n}$ величины ξ_1 относительно величин $\xi_2, \xi_3, \ldots, \xi_n$ равен остатку величины $\eta_{1.23\ldots(n-1)}$ относительно единственной величины $\eta_{n.23\ldots(n-1)}$.

ГЛАВА X

СТАТИСТИЧЕСКАЯ ОЦЕНКА НЕИЗВЕСТНЫХ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

§ 1. ПОСТАНОВКА ЗАДАЧИ

На практике во многих случаях функция распределения рассматриваемой случайной величины ξ неизвестна; ее определяют по результатам наблюдений, или, как говорят, по *выборке*.

Выборкой объема n для данной случайной величины ξ называется последовательность $X_1, X_2, \ldots, X_n$ n независимых наблюдений этой величины. Таким образом, $X_1, X_2, \ldots, X_n$ — это совокупность значений, принятых n независимыми случайными величинами $\xi_1, \xi_2, \ldots, \xi_n$, имеющими тот же закон распределения $F_\xi(x)$, что и рассматриваемая величина ξ, В этом случае говорят, что выборка $X_1, X_2, \ldots, X_n$ взята из *генеральной совокупности* величины ξ, а под законом распределения генеральной совокупности понимают закон распределения случайной величины ξ. Значения $X_1, X_2, \ldots, X_n$ называются *выборочными значениями*.

Для нахождения неизвестной функции распределения в такой общей постановке следует оценить значение теоретической функции распределения $F_\xi(x)$, являющееся вероятностью события $\{\xi < x\}$, с помощью частоты этого события в полученной выборке.

Пусть $X_1, X_2, \ldots, X_n$ — выборка, x — произвольное число. Обозначим через ν_x количество выборочных значений, меньших x. Тогда $\frac{\nu_x}{n}$ является частотой попадания выборочных значений левее точки x в данной выборке, т. е. частотой события $\{\xi < x\}$. Эта частота является функцией от x и называется *эмпирической функцией распределения* случайной величины ξ, полученной по данной выборке. Обозначая эту функцию через $F_n^*(x)$, имеем

$$F_n^*(x) = \frac{\nu_x}{n}$$

Эмпирическая функция распределения обладает всеми свойствами функции распределения. Она равна нулю для x, меньших меньшего из выборочных значений, и равна единице для x, больших большего из выборочных значений. При этом она растет скачками величины $\frac{1}{n}$ в каждой из точек $X_1, X_2, \ldots, X_n$. Такая функция распределения называется ступенчатой. Она соответствует дискретному распределению, сосредоточенному в точках $X_1, X_2, \ldots, X_n$ с вероятностями $\frac{1}{n}$ в каждой точке.

Как известно (см. гл. VI), частота события в n независимых опытах является оценкой для вероятности этого события, поэтому значение эмпирической функции распределения в точке x является оценкой для вероятности события $\{\xi < x\}$, т. е. для теоретической функции распределения $F_\xi(x)$

$$F_\xi(x) \approx F_n^*(x) = \frac{\nu_x}{n}.$$

Более того, согласно закону больших чисел Бернулли при каждом фиксированном x имеем

$$F_n^*(x) \xrightarrow{\text{по вер.}} F_\xi(x) \quad \text{при} \quad n \to \infty.$$

Таким образом, чем больше объем выборки, тем более точное представление дает эмпирическая функция распределения о теоретической функции распределения.

Однако приведенный способ нахождения неизвестной функции распределения на практике мало приемлем и в большинстве случаев не является необходимым.

Во многих случаях бывает заранее известно, что функция распределения $F_\xi(x)$ принадлежит к определенному классу функций, зависящему от одного или нескольких параметров: $F_\xi(x) = F(x, \alpha_1, \alpha_2 \ldots \alpha_k)$. В этом случае определение неизвестной функции распределения сводится к оценке неизвестных параметров по выборке.

В случае, если рассматриваемая случайная величина ξ — результат прямого измерения некоторой физической постоянной a (систематическая ошибка отсутствует), то распределение вероятностей величины ξ, как известно, нормальное с параметрами $a = M\xi$ и $\sigma = \sigma_\xi$. В этом случае вид функции распределения известен заранее, а оценка неизвестных параметров по выборке заключается в определении по результатам измерений истинного значения a измеряемой величины и средней квадратичной ошибки σ измерения.

Следует обратить внимание на тот факт, что ни при каком n, вообще говоря, нельзя определить по выборке точное значение неизвестного параметра, а можно лишь найти его приближенное значение, которое называется *оценкой* по выборке неизвестного параметра.

Всякая оценка неизвестного параметра α по выборке является функцией α^* $(X_1, X_2, \ldots, X_n)$ от выборочных значений $X_1, X_2, \ldots, X_n$. Значения X_i являются конкретными значениями случайных величин ξ_i, полученными по выборке, поэтому они меняются от выборки к выборке.

Таким образом, величины X_i можно рассматривать как числа (если эксперимент проведен и выборка состоялась) и как случайные величины ξ_i (до проведения эксперимента). В дальнейшем мы преимущественно будем пользоваться последней точкой зрения, сохраняя при этом те же обозначения X_i. Функция $\alpha^*(X_1, X_2, \ldots, X_n)$ также является случайной величиной, так как аргументы случайны.

Возникает вопрос о выборе вида функции α^* $(X_1, X_2, \ldots, X_n)$ для оценки неизвестного параметра α функции распределения $F_\xi(x) = F(x, \alpha)$. (Здесь и в дальнейшем рассматривается случай одного неизвестного параметра.) Если параметр α является средним значением, дисперсией или какой-либо иной числовой характеристикой распределения, естественно находить функцию α^* $(X_1, X_2, \ldots, X_n)$ следующим образом: сначала по выборке построить эмпирическую функцию распределения, являющуюся аналогом теоретической функции распределения $F_\xi(x) = F(x, \alpha)$, затем вычислить соответствующую числовую характеристику эмпирического распределения и принять ее за оценку неизвестного параметра α. Числовые характеристики эмпирического распределения называются *эмпирическими*, или *выборочными*, *характеристиками* распределения. Так, например, *выборочное среднее*

$$\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$$

является средним арифметическим выборочных значений;

выборочная дисперсия

$$S^2=\frac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2 \quad \text{и т. д.}$$

В соответствии с вышесказанным за оценку математического ожидания $M\xi$ и дисперсии $D\xi$ можно принять соответствующие выборочные характеристики, т. е.

$$M\xi\approx\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k, \tag{10.1}$$

$$D\xi\approx S^2=\frac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2. \tag{10.2}$$

Аналогично поступают с другими числовыми характеристиками. Известно, что $M\overline{X}=M\xi$, и по закону больших чисел имеем

$$\overline{X}\xrightarrow{\text{по вер.}}M\xi \quad \text{при } n\to\infty.$$

Приведенный способ оценки параметров пригоден лишь для параметров, являющихся известными числовыми характеристиками распределения. Кроме того, он не всегда приводит к наилучшим оценкам. В связи с этим возникает ряд вопросов и прежде всего вопрос о том, какую оценку, т. е. какую функцию $\alpha^*(X_1, X_2, \ldots, X_n)$, считать наилучшей или по крайней мере «хорошей», т. е. какие требования следует предъявлять к оценкам.

Затем следует решить вопрос о способе получения оценок. Получив ту или иную приближенную формулу, необходимо оценить точность полученных приближенных равенств.

§ 2. КЛАССИФИКАЦИЯ ОЦЕНОК

Как уже отмечалось, оценка $\alpha^*(X_1, X_2, \ldots, X_n)$ является случайной величиной как функция от n независимых случайных величин $X_1, X_2, \ldots, X_n$ или от одного n-мерного случайного аргумента $\overline{\mathbf{X}}=\{X_1, X_2, \ldots, X_n\}$. Величины $X_1, X_2, \ldots, X_n$ независимы и имеют одну и ту же функцию распределения $F_\xi(x)$, поэтому функция распределения $F_{\mathrm{X}}(t_1, t_2, \dots t_n) = \mathrm{P}(X_1 < t_1, \quad X_2 < t_2, \quad \dots, X_n < t_n)$ однозначно определена и имеет следующий вид:

$$F_{\overline{\mathrm{X}}}(t_1, t_2 \dots, t_n) = \mathrm{P}(X_1 < t_1, \ X_2 < t_2 \ \dots, X_n < t_n) =$$

$$= \mathrm{P}(X_1 < t_1) \cdot \mathrm{P}(X_2 < t_1) \dots \mathrm{P}(X_n < t_n) = F_\xi(t_1) \times$$

$$\times F_\xi(t_2) \dots F_\xi(t_n).$$

Закон распределения оценки $\alpha^*(X_1, X_2, \dots, X_n)$ также однозначно определяется неизвестной функцией распределения $F_\xi(x)$ (см. гл. IV), а точнее, неизвестным параметром α. Функцию $\alpha^*(X_1, X_2, \dots, X_n)$ надо подобрать так, чтобы случайная величина $\alpha^*(X_1, X_2, \dots, X_n)$ по возможности более точно аппроксимировала неслучайное неизвестное число α. При этом оценка $\alpha^*(X_1, X_2, \dots, X_n)$ не должна давать систематического завышения или занижения результата, т. е. необходимо, чтобы $M\alpha^*(X_1, X_2, \dots, X_n) = \alpha$.

Определение. *Оценка* $\alpha^*(X_1, X_2, \dots, X_n)$, *обладающая следующим свойством*:

$$M\alpha^*(X_1, X_2, \dots, X_n) = \alpha,$$

называется несмещенной оценкой. В противном случае оценка называется смещенной.

Если оценка является смещенной, то, вычислив ее математическое ожидание и введя поправочный коэффициент, иногда можно получить несмещенную оценку.

Для того чтобы случайное число $\alpha^*(X_1, X_2, \dots, X_n)$ было возможно ближе к неслучайному числу α, необходимо, чтобы разброс величин $\alpha^*(X_1, X_2, \dots, X_n)$ вокруг α был возможно меньшим. Выбор меры разброса в значительной мере произволен; наиболее удобной и распространенной мерой разброса служит величина второго момента $M(\alpha^* - \alpha)^2$. Если α^* — оценка несмещенная, то эта величина совпадает с дисперсией $D\alpha^*$.

Из двух оценок $\alpha_1^*(X_1, X_2, \dots, X_n)$ и $\alpha_2^*(X_1, X_2, \dots, X_n)$ считается лучшей и называется более эффективной та, для которой величина второго момента наименьшая.

Определение. *Оценка* $\alpha_1^*(X_1, X_2, \dots, X_n)$ *называется более эффективной, чем оценка* $\alpha_2^*(X_1, X_2, \dots, X_n)$, *если*

$$M(\alpha_1^* - \alpha)^2 < M(\alpha_2^* - \alpha)^2.$$

Рассмотрим нижнюю грань $\inf\limits_{\alpha^*} M(\alpha^* - \alpha)^2$ множества величин $M(\alpha^* - \alpha)^2$ по всем возможным функциям $\alpha^*(X_1, X_2, \ldots, X_n)$. Как известно, нижняя грань может и не принадлежать множеству, однако для широкого класса распределений эта нижняя грань достигается. Оценка $\alpha^*(X_1, X_2, \ldots, X_n)$, для которой достигается нижняя грань, называется *эффективной оценкой*. Таким образом, эффективная оценка не всегда существует. Оценка $\alpha^*(X_1, X_2, \ldots, X_n)$, для которой мера разброса $M(\alpha^* - \alpha)^2$ при $n \to \infty$ ведет себя так же, как и для эффективной оценки, т. е. для которой

$$\lim_{n \to \infty} \frac{M(\alpha^*(X_1, X_2, \ldots, X_n) - \alpha)^2}{\inf\limits_{\alpha^*} M(\alpha^*(X_1, X_2, \ldots, X_n) - \alpha)^2} = 1$$

называется *асимптотически эффективной оценкой*.

Рассматривая оценку $\alpha^*(X_1, X_2, \ldots, X_n)$ параметра α, естественно потребовать, чтобы при $n \to \infty$ эта оценка стремилась к параметру α в каком-нибудь смысле (например, по вероятности, в среднем квадратичном или с вероятностью 1).

Определение. *Оценка* $\alpha^*(X_1, X_2, \ldots, X_n)$ *параметра* α *называется состоятельной, если она при* $n \to \infty$ *сходится по вероятности к* α, *т. е. если*

$$\alpha^*(X_1 X_2, \ldots, X_n) \xrightarrow{\text{по вер.}} \alpha \quad \text{при} \quad n \to \infty.$$

Согласно закону больших чисел $\overline{X} \xrightarrow{\text{по вер.}} M\xi$ при $n \to \infty$, т. е. оценка $\overline{X}$ является состоятельной оценкой для $M\xi$. Таким образом, оценка (10.1) является состоятельной и несмещенной, так как $M\overline{X} = M\xi$. Можно показать, что в случае нормального распределения эта оценка будет также и эффективной.

Рассмотрим оценку (10.2) для дисперсии, несколько преобразовав ее:

$$S^2 = \frac{1}{n} \sum_{k=1}^{n} (X_k - \overline{X})^2 = \frac{1}{n} \sum_{k=1}^{n} [X_k - m - (\overline{X} - m)]^2 =$$

$$= \frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2 - \frac{2}{n}(\overline{X} - m)\sum_{k=1}^{n}(X_k - m) +$$

$$+ \frac{n}{n}(\overline{X} - m)^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2 - \frac{2}{n}(\overline{X} - m)(\overline{X} -$$

$$- m)\, n + (\overline{X} - m)^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2 - (\overline{X} - m)^2, \quad (10.3)$$

где $m = M\xi$.

Найдем математическое ожидание S^2:

$$MS^2 = \frac{1}{n} M \sum_{k=1}^{n}(X_k - m)^2 - M(\overline{X} - m)^2 = \frac{1}{n}\cdot n\sigma^2 -$$

$$- \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma^2 \neq \sigma^2, \quad \text{где} \quad \sigma^2 = D\xi.$$

Здесь использован тот факт, что $M(\overline{X} - m)^2 = D\overline{X} = \frac{1}{n}\sigma^2$ (см. гл. III). Таким образом, оценка (10.2) является смещенной; несмещенную оценку можно получить, если считать

$$\sigma^2 \approx \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2 = (S^*)^2. \quad (10.4)$$

Покажем, что оценка (10.2) является состоятельной, для этого воспользуемся равенством (10.3). Первый член в правой части равенства (10.3) $\left(\text{величина } \frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2\right)$ является средним арифметическим n независимых, одинаково распределенных величин $(X_k - m)^2$. Согласно закону больших чисел имеем

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2 \xrightarrow{\text{по вер.}} M(X_i - m)^2 = \sigma^2.$$

Второй член равенства (10.3) [величина $(\overline{X}-m)^2$] является квадратом величины, стремящейся по вероятности к нулю (по закону больших чисел), и поэтому сама величина также стремится к нулю*. Величина S^2 [по формуле (10.3)] является суммой двух слагаемых, одно из которых стремится по вероятности к σ^2, другое — к нулю, поэтому $S^2 \xrightarrow{\text{по вер.}} \sigma^2$ при $n \to \infty$**, что и означает состоятельность оценки (10.2).

Оценка (10.4) также является состоятельной, так как отличается от состоятельной оценки (10.2) множителем $\frac{n}{n-1}$, стремящимся к единице при $n \to \infty$. В самом деле,

$$\frac{n}{n-1}S^2-\sigma^2=S^2-\sigma^2+\frac{S^2-\sigma^2}{n-1}+\frac{\sigma^2}{n-1}$$

Первое из слагаемых этого равенства стремится к нулю по вероятности по доказанному, второе и третье из слагаемых также стремятся к нулю по вероятности при $n \to \infty$, поэтому и левая часть равенства стремится к нулю по вероятности при $n \to \infty$ как сумма величин, стремящихся к нулю по вероятности.

---

* В самом деле, пусть $\xi_n \xrightarrow{\text{по вер}} 0$ при $n \to \infty$, т. е., каково бы ни было $\varepsilon > 0$, имеем $P(|\xi_n| > \varepsilon) \to 0$ при $n \to \infty$ Зададим $\varepsilon > 0$, тогда, по условию,

$$P\left(\xi_n^2 > \varepsilon\right) = P\left(|\xi_n| > \sqrt{\varepsilon}\right) \to 0 \quad \text{при} \quad n \to \infty.$$

** Достаточно показать, что если $\xi_n^{(1)} \xrightarrow{\text{по вер.}} 0$ и $\xi_n^{(2)} \xrightarrow{\text{по вер.}} 0$ при $n \to \infty$, то $\xi_n^{(1)} + \xi_n^{(2)} \xrightarrow{\text{по вер.}} 0$, при $n \to \infty$. Пусть $\varepsilon > 0$, тогда

$$P\left(\left|\xi_n^{(1)}+\xi_n^{(2)}\right| > \varepsilon\right) \leqslant P\left(\left|\xi_n^{(1)}\right| > \frac{\varepsilon}{2} \text{ или } \left|\xi_n^{(2)}\right| > \frac{\varepsilon}{2}\right) \leqslant P\left(\left|\xi_n^{(1)}\right| > \frac{\varepsilon}{2}\right) + P\left(\left|\xi_n^{(2)}\right| > \frac{\varepsilon}{2}\right).$$

Но каждое из слагаемых, находящихся в правой части последнего неравенства, стремится к нулю при $n \to \infty$ по условию, поэтому и левая часть стремится к нулю при $n \to \infty$, что и требовалось доказать.

§ 3. МЕТОДЫ ПОЛУЧЕНИЯ ОЦЕНОК

Один из методов получения оценок неизвестных параметров распределения по выборке для частного случая, когда оцениваемый параметр α является математическим ожиданием или дисперсией распределения (т. е. первым или вторым моментом распределения), изложен в § 2 настоящей главы. Пусть теперь α — совершенно произвольный параметр функции распределения $F_\xi(x) = F(x, \alpha)$.

Тогда $M\xi = \int\limits_{-\infty}^{+\infty} x dF(x, \alpha) = m_1(\alpha)$ является функцией от α.

Эмпирическое среднее $\overline{X} = \frac{1}{n} \sum\limits_{k=1}^{n} X_k = \int\limits_{-\infty}^{+\infty} x dF_n^*(x)$ является функцией выборочных значений. Приравнивая первый теоретический и первый эмпирический моменты, получаем уравнение для определения неизвестного параметра α:

$$m_1(\alpha) = \overline{X}.$$

Если по выборке надо оценить не один, а несколько (пусть k неизвестных параметров $\alpha_1, \alpha_2, \ldots, \alpha_k$), следует найти первый, второй, k-й моменты теоретического распределения $F(x, \alpha_1, \alpha_2, \ldots, \alpha_k)$:

$$m_j(\alpha_1, \alpha_2, \ldots, \alpha_k) = \int\limits_{-\infty}^{+\infty} x^j \, dF(x, \alpha_1, \alpha_2, \ldots, \alpha_k), \quad j = 1, 2, \ldots, k,$$

затем соответствующие эмпирические моменты:

$$\mu_j(X_1, X_2, \ldots X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i^j = \int\limits_{-\infty}^{+\infty} x^j \, dF_n^*(x)$$

и приравнять их. Получаем систему k уравнений с k неизвестными:

$$m_j(\alpha_1, \alpha_2, \ldots, \alpha_k) = \mu_j(X_1, X_2, \ldots, X_n), \; j = 1, 2, \ldots, k. \quad (10.5)$$

Решая систему (10.5), получаем оценки неизвестных параметров. Изложенный метод получения оценок, состоящий в приравнивании теоретических моментов распределения к эмпирическим, носит название *метода моментов*. Преимуществом метода моментов является сравнительная простота его применения. Однако такой способ получения оценок содержит в себе элемент произвола, так как кроме моментов можно рассматривать и другие числовые характеристики распределений, например моду, медиану, коэффициент асимметрии, эксцесс и т. п. Приравнивая друг к другу значения соответствующих теоретических и эмпирических числовых характеристик, мы также получаем систему уравнений для определения неизвестных параметров. Таким образом, нет оснований предполагать, что с помощью метода моментов можно получить наилучшие в каком-либо смысле оценки. Можно показать, что оценки, полученные по методу моментов, вообще говоря, имеют не наименьший возможный разброс и не являются асимптотически эффективными. Однако в силу простоты метод моментов иногда используют на практике.

Для получения оценки неизвестного параметра α естественно попытаться найти такое значение α^*, при котором, грубо говоря, вероятность реализации полученной выборки $X_1, X_2, \ldots, X_n$ была бы максимальной. При таком способе получения оценки можно ожидать от нее некоторых оптимальных свойств. Пусть величина ξ имеет дискретное распределение с возможными значениями $a_1, a_2, \ldots, a_k$, причем $p_1(\alpha), p_2(\alpha), \ldots, p_k(\alpha)$ — вероятности этих значений, т. е.

$$P(\xi = a_l) = p_l(\alpha),$$

где α — неизвестный параметр. Пусть в выборке $X_1, X_2, \ldots, X_n$ значение a_j встречалось n_j раз ($j = 1, 2, \ldots, k$). Тогда по правилу умножения вероятностей вероятность при n независимых наблюдениях величины ξ получить выборку $X_1, X_2, \ldots, X_n$ равна

$$L(X_1, X_2, \ldots, X_n \; \alpha) = p_1^{n_1}(\alpha) \cdot p_2^{n_2}(\alpha), \ldots p_k^{n_k}(\alpha). \quad (10.6)$$

Величина (10.6) называется *функцией правдоподобия*, а величина α^*, являющаяся точкой максимума этой функции, называется *оценкой*, полученной по *методу максимума правдоподобия*, или по *методу наибольшего правдоподобия*.

Пусть величина ξ имеет плотность распределения вероятностей $p(x, \alpha)$. Тогда вероятность получить при n независимых наблюдениях величины ξ выборку $X_1, X_2, \ldots, X_n$ будет всегда равна нулю, так как многомерная случайная величина $\{X_1, X_2, \ldots, X_n\}$ имеет плотность распределения

$$L(X_1, X_2, \ldots, X_n, \alpha) = p(x_1, \alpha)\, p(x_2, \alpha) \ldots p(x_n, \alpha). \quad (10.7)$$

Поэтому будем искать параметр α из условия обращения в максимум вероятности попадания выборки в n-мерный параллелепипед с центром в $(X_1, X_2, \ldots, X_n)$ и ребрами $dx_1, dx_2, \ldots, dx_n$ (при этом выборка рассматривается как точка n-мерного пространства). Эта вероятность равна

$$p(X_1, \alpha)\, p(X_2, \alpha) \ldots p(X_n, \alpha)\, dx_1\, dx_2 \ldots dx_n =$$
$$= L(X_1, X_2, \ldots, X_n\, \alpha)\, dx_1\, dx_2 \ldots dx_n. \quad (10.8)$$

Величина α^* — точка максимума вероятности (10.8); является, очевидно, точкой максимума функции $L(X_1, X_2, \ldots, X_n\, \alpha)$. Эта функция также называется функцией правдоподобия, а величина α^*, являющаяся точкой максимума этой функции,— оценкой, полученной по *методу максимума правдоподобия*, или *оценкой максимального правдоподобия*.

При $a = \alpha^*$ функция правдоподобия принимает наибольшее значение, и событие, состоящее в том, что новая выборка совпадает с наблюдаемой в дискретном случае или окажется близкой к наблюдаемой в непрерывном случае, имеет максимальную вероятность, т. е. максимально правдоподобно.

Как известно, для нахождения точек максимума функции $L(X_1, X_2, \ldots, X_n, \alpha)$ надо решить

$$\frac{\partial L(X_1, X_2, \ldots, X_n, \alpha)}{\partial \alpha} = 0 \quad (10.9)$$

Однако точки максимума функций $L(X_1, X_2, \ldots, X_n\, \alpha)$ и $\ln L(X_1, X_2, \ldots, X_n, \alpha)$ совпадают, поэтому вместо уравнения (10.9) можно решать уравнение

$$\frac{\partial \ln L(X_1, X_2, \ldots, X_n, \alpha)}{\partial \alpha} =$$
$$= \frac{1}{L} \frac{\partial L(X_1, X_2, \ldots, X_n, \alpha)}{\partial \alpha} = 0. \quad (10.10)$$

Если надо оценить не один, а несколько (пусть k неизвестных параметров $\alpha_1, \alpha_2, \ldots, \alpha_k$), то оценки наибольшего правдоподобия для этих параметров находятся из системы уравнений

$$\frac{\partial L(X_1, X_2, \ldots, X_n \alpha_1, \ldots, \alpha_k)}{\partial \alpha_j} = 0 \quad (j = 1, 2, \ldots, k).$$

Ценность оценок максимального правдоподобия обусловлена следующими их свойствами, которые выполняются при весьма общих условиях и приводятся здесь без доказательства* (при этом мы ограничиваемся случаем одного параметра):

1) уравнение правдоподобия имеет решение, которое является состоятельной оценкой для α;

2) решение является асимптотической эффективной оценкой для α;

3) решение имеет асимптотически нормальное предельное распределение, т. е. при соответствующей нормировке предельное распределение при $n \to \infty$ полученной оценки оказывается нормальным;

4) если для параметра α существует эффективная оценка, то уравнение правдоподобия имеет единственное решение, совпадающее с этой оценкой.

Оценка максимального правдоподобия может оказаться, однако, смещенной.

На практике метод максимума правдоподобия приводит иногда к довольно сложным системам уравнений.

Пример 1. Пусть производится последовательность n испытаний Бернулли с неизвестной вероятностью p успеха. Пусть в этой последовательности событие «успех» осуществилось m раз, «неудача» — $(n - m)$ раз. Величина ξ имеет дискретное распределение; возможные значения — количество успехов в одном испытании — 0 и 1 с вероятностями $q = 1 - p$ и p соответственно. Неизвестный параметр — вероятность p; выборка $X_1, X_2, \ldots, X_n$ — последовательность из чисел нуль и единица, причем единица встречается m раз, нуль — $(n - m)$ раз.

Составим функцию правдоподобия. По формуле (10. 6) имеем

$$L(X_1, X_2, \ldots X_n, p) = p^m (1 - p)^{n-m},$$

и уравнение (10.10) принимает вид

$$\frac{\partial \ln L}{\partial p} = \frac{m}{p} - \frac{n - m}{1 - p} = 0.$$

Отсюда находим единственное решение:

$$p^* (X_1, X_2, \ldots X_n) = \frac{m}{n}$$

* Доказательство можно найти в кн.: Г. Крамер. Математические методы статистики М., ИЛ. 1948

Таким образом, оценкой наибольшего правдоподобия вероятности события служит его частота в данной последовательности испытаний. Частота является несмещенной оценкой вероятности $\left(\text{так как } M\dfrac{m}{n} = p\right)$ и состоятельной оценкой $\left(\text{так как } \dfrac{m}{n} \xrightarrow{\text{по вер.}} p\right.$ при $n \to \infty$ согласно закону больших чисел Бернулли). Кроме того, по теореме Муавра — Лапласа частота является асимптотически нормальной оценкой.

Пример 2. Рассмотрим оценки неизвестных параметров a и σ нормального распределения по методу максимума правдоподобия. В этом случае имеем:

$$L(X_1, X_2, \dots, X, a, \sigma) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n} \cdot e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(X_i - a)^2},$$

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - a)^2.$$

Дифференцируем $\ln L$ по a и σ^2:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - a) = 0,$$

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i - a)^2 = 0,$$

откуда находим:

$$a^*(X_1, X_2, \dots, X_n) = \overline{X},$$

$$(\sigma^*)^2(X_1, X_2, \dots, X_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = S^2.$$

Эти оценки совпадают с оценками, полученными по методу моментов. Они являются, как было показано в § 2 настоящей главы, состоятельными, причем первая из оценок оказывается несмещенной, а вторая — смещенной.

Возможны и другие подходы к вопросу получения оценок. Целую систему различных методов нахождения оценок можно получить из следующих соображений. Зададим каким-либо образом меру D отклонения эмпирической функции распределения $F_n^*(x)$ от теоретической функции распределения $F(x, \alpha)$. Эта мера зависит от α и от выборочных значений $X_1, X_2, \dots, X_n$:

$$D = D(X_1, X_2, \dots, X_n, \alpha).$$

При этом за оценку параметра α естественно принять то значение, которое минимизирует меру D отклонения. Образуя величину D различными способами, будем получать различные способы нахождения оценок. Наиболее важной и употребительной является мера D отклонения, предложенная К. Пирсоном и носящая название величины χ^2. Эта мера отклонения определяется следующим образом: пусть множество значений случайной величины разбито на r непересекающихся множеств (интервалов или полуинтервалов, конечных или бесконечных); пусть ν_i ($i = 1, 2, \ldots, r$) — количество выборочных значений, попавших в i-е множество; $p_i(\alpha)$ — вероятность попадания выборочного значения в i-е множество; при этом

$$\sum_{i=1}^{n} \nu_i = n, \qquad \sum_{i=1}^{r} p_i(\alpha) = 1,$$

где $\frac{\nu_i}{n}$ — наблюдаемая частота; $p_i(\alpha)$ — ожидаемая частота попадания выборочных значений в i-е множество.

За меру отклонения D можно принять величину

$$D = \sum_{i=1}^{r} c_i \left(\frac{\nu_i}{n} - p_i(\alpha)\right)^2,$$

где c_i — некоторые коэффициенты. Если считать $c_i = \frac{n}{p_i(\alpha)}$, то полученная мера отклонения обладает рядом простых свойств и носит название величины χ^2. Таким образом,

$$\chi^2 = \sum_{i=1}^{r} \frac{n}{p_i(\alpha)} \left(\frac{\nu_i}{n} - p_i(\alpha)\right)^2 = \sum_{i=1}^{r} \frac{(\nu_i - np_i(\alpha))^2}{np_i(\alpha)}.$$

Оценка, получаемая из условия обращения величины χ^2 в минимум, называется оценкой по *методу минимума* χ^2. Эта оценка при весьма общих условиях является состоятельной, асимптотически эффективной и асимптотически нормальной, так же как и оценка по методу максимума правдоподобия.

Рассмотрим еще один способ получения оценок. Пусть D равна так называемой величине ω^2 (впервые введена

Р. Мизесом):

$$D(X_1, X_2, \ldots, X_n, \alpha) = \omega^2 = \int_{-\infty}^{+\infty} \left[F_n^*(x) - F(x, \alpha)\right]^2 dF(x, \alpha).$$

Оценка, полученная из условия обращения в минимум величины ω^2, также в весьма общих условиях оказывается асимптотически нормальной, но уже не будет, вообще говоря, асимптотически эффективной.

§ 4. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ ДЛЯ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ И ДИСПЕРСИИ

В предыдущих параграфах были рассмотрены различные оценки неизвестных параметров распределения; при этом рассматривалось приближенное равенство

$$\alpha \approx \alpha^* \ (X_1, X_2, \ldots, X_n). \tag{10.11}$$

В левой части приближенного равенства находится неслучайное число α, а в правой — случайная величина, являющаяся функцией выборочных значений. Возникает вопрос об оценке точности приближенного равенства (10.11).

Рассмотрим выборку $X_1, X_2, \ldots, X_n$ объема n из генеральной совокупности, закон распределения которой задается плотностью $p(x, \alpha)$, где значение α неизвестно. Пусть $\alpha^* = \alpha^*(X_1, X_2, \ldots, X_n)$ — некоторая оценка для α, являющаяся, очевидно, случайной величиной, и пусть $g(x/\alpha)$ — плотность распределения вероятностей этой оценки, зависящая от неизвестного значения α. Зададимся целью, вычислив по выборке значение α^*, указать для параметра α такие два предела $\underline{\alpha}$ и $\overline{\alpha}$, чтобы истинное значение α с некоторой определенной вероятностью $\mathcal{P}$ лежало в заданных пределах, т. е. чтобы

$$\mathrm{P}\left(\underline{\alpha} < \alpha < \overline{\alpha}\right) = \mathcal{P}.$$

Существуют два метода решения поставленной задачи: так называемый *классический* (*или байесовский метод*) и *метод доверительных интервалов*, предложенный Нейманом. Рассмотрим оба метода.

1. Классический метод применим в тех случаях, когда неизвестный параметр α является случайной величиной, имеющей некоторое распределение вероятностей, называемое *априорным распределением*. Предположим, что априорное распределение параметра α имеет известную плотность $\varphi(x)$; плотность $g(x/\alpha)$ является условной плотностью распределения оценки α^*—при заданном значении α. Условная плотность распределения вероятности величины α при заданном значении α^* по формуле Байеса равна

$$h(x/\alpha^*) = \frac{\varphi(x)\, g(\alpha^*/x)}{\int\limits_{-\infty}^{+\infty} \varphi(x)\, g(\alpha^*/x)\, dx}.$$

Поэтому условная вероятность события $\underline{\alpha} < \alpha < \overline{\alpha}$ при заданном значении α^*

$$P\left(\underline{\alpha} < \alpha < \overline{\alpha}/\alpha^*\right) = \int\limits_{\underline{\alpha}}^{\overline{\alpha}} h(x/\alpha^*)\, dx.$$

Эта вероятность называется *апостериорной вероятностью* события $\left(\underline{\alpha} < \alpha < \overline{\alpha}\right)$ в отличие от *априорной вероятности*, равной

$$P\left(\underline{\alpha} < \alpha < \overline{\alpha}\right) = \int\limits_{\underline{\alpha}}^{\overline{\alpha}} \varphi(x)\, dx.$$

Итак, задавая вероятность $\mathcal{P}$ и определяя числа $\underline{\alpha}$ и $\overline{\alpha}$ так, чтобы

$$\int\limits_{\underline{\alpha}}^{\overline{\alpha}} h(x/\alpha^*)\, dx = \mathcal{P},$$

получаем

$$P\left(\underline{\alpha} < \alpha < \overline{\alpha}/\alpha^*\right) = \mathcal{P}.$$

Таким образом, при условии, что наша оценка приняла значение α^*, истинное значение параметра α лежит между $\underline{\alpha}$ и $\overline{\alpha}$ с вероятностью $\mathcal{P}$. Выбирая $\mathcal{P}$ достаточно близким к единице, можно считать событие $\left\{\underline{\alpha} < \alpha < \overline{\alpha}\right\}$ практически достоверным при условии, что оценка приняла значение α^*. На практике изложенный метод часто неприменим, так как параметр α является просто неизвестной постоянной, а не случайной величиной. Кроме того, если даже можно считать α случайной величиной, то очень часто ее априорное распределение бывает неизвестным. От всех этих недостатков свободен метод доверительных интервалов.

2. **Метод доверительных интервалов** является более общим методом, поскольку он применим и в случае, если α является неизвестным фиксированным числом и если α — случайная величина.

Зададим число $0 < \mathscr{P} < 1$, которое на практике выбирают достаточно близким к единице так, чтобы событие с вероятностью $\mathscr{P}$ можно было считать практически достоверным. При каждом фиксированном значении параметра α плотность распределения $g(x/\alpha)$ задает распределение оценки α^*, которое будем рассматривать как распределение единичной массы на вертикальной прямой, проходящей через точку $(\alpha, 0)$ в плоскости (α, α^*) (рис. 10.1). Определим для каждого значения α числа $\gamma_1(\alpha, \mathscr{P})$ и $\gamma_2(\alpha, \mathscr{P})$ так, чтобы количество массы, попавшей на отрезок $[\gamma_1, \gamma_2]$ рассматриваемой вертикальной прямой, было равно $\mathscr{P}$, т. е. чтобы

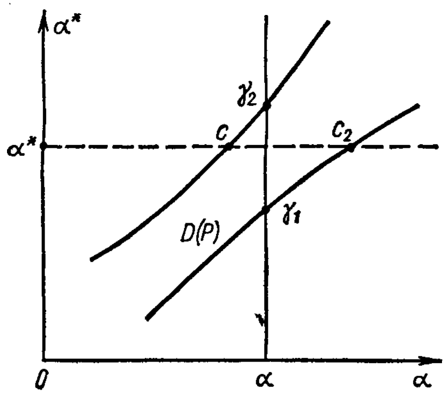

Рис. 10.1

$$P(\gamma_1(\alpha, \mathscr{P}) < \alpha^* < \gamma_2(\alpha, \mathscr{P})) = \int_{\gamma_1}^{\gamma_2} g(x/\alpha)\, dx = \mathscr{P}.$$

Такие числа можно выбрать, очевидно, бесчисленным множеством способов. Числа γ_1 и γ_2 зависят от α; с изменением α точки (α, γ_1) и (α, γ_2) описывают в плоскости (α, α^*) две кривые (см. рис. 10.1). Предположим, что всякая прямая, параллельная оси α, пересекает каждую из этих кривых лишь в одной точке. Обозначим через $c_1(\alpha^*, \mathscr{P})$ и $c_2(\alpha^*, \mathscr{P})$ точки пересечения этих кривых с прямой, проходящей через точку $(0, \alpha^*)$ параллельно оси α. Пусть $D(\mathscr{P})$ — область, заключенная между двумя кривыми. Очевидно, что утверждение

$$\gamma_1(\alpha, \mathscr{P}) < \alpha^* < \gamma_2(\alpha, \mathscr{P}) \tag{10.12}$$

эквивалентно утверждению $(\alpha, \alpha^*) \in D(\mathscr{P})$.

Аналогично, утверждение

$$c_1(\alpha^*, \mathscr{P}) < \alpha < c_2(\alpha^*, \mathscr{P}) \qquad (10.13)$$

также эквивалентно утверждению $(\alpha, \alpha^*) \in D(\mathscr{P})$. Таким образом, если для какой-нибудь выборки $X_1, X_2, \ldots, X_n$ справедливо неравенство (10.12), то справедливо и неравенство (10.13), и наоборот. Поэтому при любом α

$$\mathrm{P}\{c_1(\alpha^*, \mathscr{P}) < \alpha < c_2(\alpha^* \mathscr{P})\} = \mathrm{P}(\gamma_1(\alpha, \mathscr{P}) < \alpha^* < \gamma_2(\alpha, \mathscr{P})) = \mathscr{P}.$$

Соотношение (10.12) означает, что случайная величина α^* заключена между пределами γ_1 и γ_2, а неравенство (10.13) означает, что величина α (которая может быть как неслучайной, так и случайной) заключена между случайными пределами c_1 и c_2.

Таким образом, случайный интервал $(c_1(\alpha^*, \mathscr{P}), c_2(\alpha^*, \mathscr{P}))$ с вероятностью $\mathscr{P}$ содержит внутри себя неизвестное значение α.

Случайный интервал (c_1, c_2) называется *доверительным интервалом* для параметра α, соответствующим *коэффициенту доверия* $\mathscr{P}$ (или *доверительному уровню* $\mathscr{P}$), а числа c_1 и c_2— *доверительными пределами*. Заметим, что доверительные интервалы, отвечающие коэффициенту доверия $\mathscr{P}$, можно строить различными способами, подобно тому, как можно разными способами оценивать неизвестные параметры распределения по выборке. Способ построения доверительных интервалов следует выбирать так, чтобы при данном коэффициенте доверия $\mathscr{P}$ они оказывались возможно более короткими, т. е. полоса $D(\mathscr{P})$ между нашими кривыми была бы по возможности у́же.

Построим доверительные интервалы для математического ожидания и дисперсии в случае нормально распределенной генеральной совокупности.

1. Построение доверительного интервала для математического ожидания a при известной дисперсии σ^2 нормально распределенной генеральной совокупности. Пусть выборка $X_1, X_2, \ldots, X_n$ состоит из независимых нормально распределенных с параметрами a и σ случайных величин, причем σ известно, а величину a оцениваем по выборке:

$$a \approx \overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

Поставим перед собой задачу оценить точность этого приближенного равенства, т. е. указать границы (доверительные пределы), в которых практически достоверно лежит неизвестное число a. Предварительно покажем, что сумма $\zeta_n = \sum_{k=1}^{n} \xi_k$ независимых нормально распределенных с параметрами a и σ случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ распределена также нормально с параметрами a и $\sqrt{n}\sigma$. В самом деле, переходя к характеристическим функциям (см. пример 1 § 2 гл. VII), имеем

$$\varphi_{\xi_k}(t) = e^{iat - \frac{\sigma^2 t^2}{2}}.$$

Согласно свойству 2) характеристических функций, можно записать

$$\varphi_{\zeta_n}(t) = \prod_{k=1}^{n} \varphi_{\xi_k}(t) = e^{inat - \frac{n\sigma^2 t^2}{2}},$$

что в силу теоремы единственности для характеристических функций означает нормальность суммы ζ_n. Таким же образом можно показать, что любая линейная комбинация $c_1X_1 + c_2X_2 + \ldots + c_nX_n$ независимых нормально распределенных случайных величин $X_1, X_2, \ldots, X_n$ распределена нормально.

Итак, величина $\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$ распределена в нашем случае нормально с параметрами a и $\frac{\sigma}{\sqrt{n}}$. Поэтому вероятность того, что $(\overline{X} - a)$ не превзойдет по абсолютной величине некоторого наперед заданного числа $\varepsilon > 0$, равна

$$P\left(|\overline{X} - a| < \varepsilon\right) = F\left(\frac{\varepsilon}{\sigma_{\overline{X}}}\right) - F\left(-\frac{\varepsilon}{\sigma_{\overline{X}}}\right) =$$

$$= F\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - F\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right),$$

или

$$P\left(|\overline{X} - a| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right),$$

где $F(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ — нормальная функция распределения, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{x} e^{-\frac{t^2}{2}} dt$ — функция Лапласа.

Зададим коэффициент доверия $\mathcal{P}$ таким, чтобы событие с вероятностью $\mathcal{P}$ можно было считать практически достоверным, и пусть $t_{\mathcal{P}}$ — корень уравнения $2\Phi(t_{\mathcal{P}}) = \mathcal{P}$, который можно найти по таблицам функции Лапласа или нормальной функции распределения. Например, при $\mathcal{P} = 0{,}999$ имеем $t_{\mathcal{P}} = 3{,}29$. Определим из условия $\frac{\varepsilon\sqrt{n}}{\sigma} = t_{\mathcal{P}}$ число ε:

$$\varepsilon = \frac{t_{\mathcal{P}}\sigma}{\sqrt{n}},$$

Для данного ε

$$P\left(|\overline{X} - a| < \frac{t_{\mathcal{P}}\sigma}{\sqrt{n}}\right) = 2\Phi\left(t_{\mathcal{P}}\right) = \mathcal{P}.$$

Таким образом, практически достоверно (точнее, с вероятностью $\mathcal{P}$), что

$$|\overline{X} - a| < \frac{t_{\mathcal{P}}\sigma}{\sqrt{n}}, \quad \text{где} \quad 2\Phi\left(t_{\mathcal{P}}\right) = \mathcal{P}.$$

Последнее неравенство запишем в виде

$$\overline{X} - t_{\mathcal{P}}\frac{\sigma}{\sqrt{n}} < a < \overline{X} + t_{\mathcal{P}}\frac{\sigma}{\sqrt{n}} \tag{10.14}$$

Мы получили так называемую *классическую оценку*.

Таким образом, интервал со случайными концами $\overline{X} - t_{\mathcal{P}}\frac{\sigma}{\sqrt{n}}$ и $\overline{X} + t_{\mathcal{P}}\frac{\sigma}{\sqrt{n}}$ с вероятностью $\mathcal{P}$ покрывает неизвестное значение $a = MX_k$. Этот интервал является *доверительным интервалом для a*, соответствующим *коэф-*

фициенту доверия $\mathcal{P}$. Доверительные пределы в этом случае таковы: $\overline{X}-t_{\mathcal{P}}\frac{\sigma}{\sqrt{n}}$ и $\overline{X}+t_{\mathcal{P}}\frac{\sigma}{\sqrt{n}}$.

Оценка (10.14) предполагает известным среднее квадратичное отклонение σ, которое на практике чаще всего бывает неизвестно. Если величину σ в неравенстве (10.14) заменить ее приближенным значением

$$\sigma\approx\sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\overline{X})^2},$$

то коэффициент доверия оценки (10.14) уменьшится. Поэтому, если величина σ неизвестна, пользуются другим способом построения доверительного интервала для математического ожидания.

2. Построение доверительного интервала для математического ожидания a при неизвестной дисперсии σ^2 нормально распределенной генеральной совокупности. Для построения доверительного интервала докажем следующую лемму.

Лемма. *В выборке $X_1, X_2, \ldots, X_n$ из нормально распределенной генеральной совокупности выборочное среднее $\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$ и выборочная дисперсия $S^2=\frac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2$ взаимно независимы. Величина $\overline{X}$ распределена нормально с параметрами a и $\frac{\sigma}{\sqrt{n}}$, а величина $\frac{nS^2}{\sigma^2}$ имеет распределение χ^2 с $(n-1)$ степенями свободы.*

Доказательство. Заметим, что выборочная дисперсия $S^2=\frac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2$ не изменится, если каждую из величин X_k заменить на $X_k'=X_k-a$. Поэтому, рассматривая величину S^2, можно без ограничения общности считать, что $a=0$, т. е. что величины X_k распределены нормально с параметрами 0 и σ. Перейдем от величин X_k к величинам Y_k, используя формулы

$$Y_k = \sum_{l=1}^{n} c_{kl} X_l \quad (k = 1, 2, \dots, n), \tag{10.15}$$

где числа c_{kl} образуют квадратную матрицу C порядка n. Пусть все элементы последней строки этой матрицы равны $\frac{1}{\sqrt{n}}$, т. е. $c_{nl} = \frac{1}{\sqrt{n}}$, $l = 1, 2, \dots, n$, тогда

$$Y_n = \frac{1}{\sqrt{n}} \sum_{l=1}^{n} X_l = \sqrt{n}\, \overline{X} \quad \text{и} \quad \sum_{l=1}^{n} c_{nl}^2 = 1.$$

Как известно*, остальные строки матрицы можно при этом подобрать так, что условия ортогональности выполняются: сумма произведений соответствующих элементов любых двух различных строк равна нулю, а сумма квадратов элементов любой строки равна единице, т. е.

$$\sum_{l=1}^{n} c_{kl}\, c_{jl} = \begin{cases} 1 & \text{при} \quad k = j \\ 0 & \text{при} \quad k \neq j \end{cases} \qquad k, j = 1, 2, \dots, n.$$

Если для матрицы C выполнены условия ортогональности (такая матрица называется *ортогональной*), то преобразование (10.15) называется *ортогональным преобразованием.*

При ортогональном преобразовании, означающем вращение n-мерного пространства около начала координат, длина векторов и углы между ними не меняются. Поэтому длины векторов $(X_1, X_2, \dots, X_n)$ и $(Y_1, Y_2, \dots, Y_n)$ равны

$$\sum_{k=1}^{n} X_k^2 = \sum_{k=1}^{n} Y_k^2. \tag{10.16}$$

Согласно п. 1 настоящего параграфа, распределение каждой из величин Y_k, $k = 1, 2, \dots, n$, нормально, причем

---

* См. в кн.: А. Н. Рублев. Линейная алгебра. М., «Высшая школа», 1968

$$MY_k = MX_k = 0, \quad DY_k = MY_k^2 = \sum_{l=1}^{n} c_{kl}^2 MX_l^2 =$$
$$= \sigma^2 \sum_{l=1}^{n} c_{kl}^2 = \sigma^2.$$

В силу условия ортогональности имеем при $k \neq j$

$$MY_k Y_j = \sum_{l=1}^{n} c_{kl}\, c_{jl}\, MX_l^2 = \sigma^2 \sum_{l=1}^{n} c_{kl} c_{jl} = 0.$$

Таким образом, нормально распределенные с параметрами 0 и σ случайные величины $Y_1, Y_2, \ldots, Y_n$ оказываются некоррелированными и, следовательно, независимыми. Для величины S^2 в силу равенства (10.16) и учитывая, что $Y_n = \sqrt{n}\, \overline{X}$, имеем

$$nS^2 = \sum_{k=1}^{n} (X_k - \overline{X})^2 = \sum_{k=1}^{n} X_k^2 - n\overline{X}^2 = \sum_{k=1}^{n} Y_k^2 - Y_n^2 =$$
$$= \sum_{k=1}^{n-1} Y_k^2.$$

Разделив на σ^2 обе части полученного равенства, получим

$$\frac{n}{\sigma^2} S^2 = \sum_{k=1}^{n-1} \left(\frac{Y_k}{\sigma}\right)^2, \tag{10.17}$$

где $\frac{Y_k}{\sigma}$, $(k = 1, 2, \ldots, n-1)$ — независимые нормально распределенные с параметрами 0 и 1 величины. Как известно (см. гл. IV), сумма квадратов таких величин имеет распределение χ^2 с $(n-1)$ степенями свободы. Таким образом, распределение для величины $\frac{n}{\sigma^2} S^2$ получено; распределение для величины $\overline{X}$ уже известно (см. п. 1 настоящего параграфа), независимость величин $\overline{X}$ и S^2 следует из того факта, что $\overline{X} = \frac{Y_n}{\sqrt{n}}$, а S^2 выражается через величины $Y_1, Y_2, \ldots, Y_{n-1}$, не зависящие от Y_n. Лемма доказана.

Перейдем к построению доверительного интервала для a. Рассмотрим две величины

$$Z = \sqrt{n}\,\frac{\overline{X} - a}{\sigma} \quad \text{и} \quad V = \frac{nS^2}{\sigma^2},$$

которые согласно лемме независимы, причем величина Z распределена нормально с параметрами 0 и 1, а величина V распределена по закону χ^2 с $(n - 1)$ степенями свободы. Как известно (см. гл. IV) в этом случае величина

$$\zeta = \frac{Z}{\sqrt{V}}\sqrt{n-1} = \frac{\overline{X} - a}{S}\sqrt{n-1}$$

имеет распределение Стьюдента с $(n - 1)$ степенями свободы. Зададим коэффициент доверия $\mathcal{P}$ и предположим, что $t_{\mathcal{P}}$ — корень уравнения

$$\int_{-t}^{t} S_{n-1}(x)\,dx = \mathcal{P},$$

где $S_{n-1}(x)$ — плотность распределения вероятностей закона Стьюдента с $(n-1)$ степенями свободы. Для значения $t_{\mathcal{P}}$, которое находится из таблиц, имеем

$$\mathrm{P}\left(|\zeta| < t_{\mathcal{P}}\right) = \int_{-t_{\mathcal{P}}}^{t_{\mathcal{P}}} S_{n-1}(x)\,dx = \mathcal{P}.$$

Таким образом, с коэффициентом доверия $\mathcal{P}$ выполняется неравенство

$$|\zeta| < t_{\mathcal{P}} \quad \text{или} \quad \left|\frac{\overline{X} - a}{S}\sqrt{n-1}\right| < t_{\mathcal{P}}.$$

Преобразуя последнее неравенство, получаем

$$\overline{X} < t_{\mathcal{P}}\frac{S}{\sqrt{n-1}} < a < \overline{X} + t_{\mathcal{P}}\frac{S}{\sqrt{n-1}}. \qquad (10.18)$$

Итак, случайный интервал с концами в точках $\overline{X} - t_{\mathcal{P}}\frac{S}{\sqrt{n-1}}$ и $\overline{X} + t_{\mathcal{P}}\frac{S}{\sqrt{n-1}}$ с вероятностью $\mathcal{P}$ содержит внутри себя неизвестное значение a. Таким образом, построен доверительный интервал для величины a, соответствующий коэффициенту доверия $\mathcal{P}$.

3. Построение доверительного интервала для математического ожидания a в случае ненормально распределенной генеральной совокупности. Каков бы ни был закон распределения независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, имеющих конечную дисперсию, их сумма $\zeta_n = \sum_{k=1}^{n} \xi_k$ распределена *приближенно нормально при достаточно больших n* (согласно центральной предельной теореме). Поэтому все выводы п. 1 будут приближенно справедливы при достаточно больших n и для ненормально распределенной генеральной совокупности. Оценка (10.14) имеет место вероятностью, близкой к $\mathcal{P}$ при достаточно больших n, и в случае, когда закон распределения генеральной совокупности не является нормальным, т. е.

$$P\left(\overline{X} - t_{\mathcal{P}} \frac{\sigma}{\sqrt{n}} < a < \overline{X} + t_{\mathcal{P}} \frac{\sigma}{\sqrt{n}}\right) \approx \mathcal{P}. \quad (10.19)$$

Здесь предполагается известным значение σ. Если же σ неизвестно, то можно использовать оценку величины σ по выборке:

$$\sigma \approx \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2} = \sigma^*$$

и заменить в равенстве (10.19) неизвестную величину σ величиной σ^*. При больших значениях n такая замена мало влияет на коэффициент доверия, и мы имеем

$$P\left(\overline{X} - t_{\mathcal{P}} \frac{\sigma^*}{\sqrt{n}} < a < \overline{X} + t_{\mathcal{P}} \frac{\sigma^*}{n}\right) \approx \mathcal{P},$$

т. е. интервал $\overline{X} - t_{\mathcal{P}} \frac{\sigma^*}{\sqrt{n}}, \ \overline{X} + t_{\mathcal{P}} \frac{\sigma^*}{\sqrt{n}}$ является *доверительным интервалом для a с коэффициентом доверия, близким к $\mathcal{P}$.*

4. Построение доверительного интервала для среднего квадратического отклонения σ и дисперсии σ^2 нормально

распределенной генеральной совокупности. Пусть X_1, X_2, ..., X_n — выборка из нормальной генеральной совокупности. Согласно лемме, доказанной в п. 2 настоящего параграфа, величина

$$\frac{nS^2}{\sigma^2}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

распределена по закону χ^2 с $(n-1)$ степенями свободы. Зададим коэффициент доверия $\mathscr{P}$ и определим числа χ_1^2 и χ_2^2 из условия

$$\int_{\chi_1^2}^{\chi_2^2} k_{n-1}(x)\,dx=\mathscr{P},$$

где $k_{n-1}(x)$ — плотность распределения вероятности закона χ^2 с $(n-1)$ степенями свободы. Очевидно, числа χ_1^2 и χ_2^2, удовлетворяющие данному условию, можно выбрать бесчисленным множеством способов. Потребуем дополнительно, чтобы

$$\int_{0}^{\chi_1^2} k_{n-1}(x)\,dx=\frac{1-\mathscr{P}}{2},$$

тогда

$$\int_{\chi_2^2}^{\infty} k_{n-1}(x)\,dx=\frac{1-\mathscr{P}}{2},$$

и числа χ_1^2 и χ_2^2 однозначно определены (их значения находятся из таблиц распределения χ^2 с $(n-1)$ степенями). Для величины $\frac{nS^2}{\sigma^2}$ имеем

$$P\left(\chi_1^2<\frac{nS^2}{\sigma^2}<\chi_2^2\right)=\int_{\chi_1^2}^{\chi_2^2} k_{n-1}(x)\,dx=\mathscr{P}.$$

Итак, с вероятностью $\mathscr{P}$ выполнены неравенства

$$\chi_1^2<\frac{nS^2}{\sigma^2}<\chi_2^2,$$

откуда

$$\frac{nS^2}{\chi_2^2} < \sigma^2 < \frac{nS^2}{\chi_1^2}, \tag{10.20}$$

или, что то же самое,

$$\frac{\sqrt{n}\,S}{\chi_2} < \sigma < \frac{\sqrt{n}\,S}{\chi_1}. \tag{10.21}$$

Таким образом, интервалы $\left(\frac{nS^2}{\chi_2^2}, \frac{nS^2}{\chi_1^2}\right)$ и $\left(\frac{\sqrt{n}\,S}{\chi_2}, \frac{\sqrt{n}\,S}{\chi_1}\right)$ являются доверительными интервалами для дисперсии σ^2 и среднего квадратичного отклонения σ, соответствующими коэффициенту доверия $\mathscr{P}$ в случае нормально распределенной генеральной совокупности.

Пример. Пусть известны результаты $X_1, X_2, \ldots, X_{22}$ двадцати двух измерений некоторого угла в градусах: 3,1; 3,3; 2,9; 3,0; 3,1; 3,2; 2,8; 2,7; 3,1; 3,2; 2,9; 3,0; 2,9; 3,1; 2,8; 2,9; 3,2; 3,3; 2,9; 3,1; 3,2; 3,0.

Предположим, что результаты измерений распределены нормально. Тогда имеем

$$a \approx \overline{X} = \frac{1}{22}\sum_{k=1}^{22} X_k = 3{,}0318,$$

$$\sigma^2 \approx \frac{22}{21} S^2 = \frac{1}{21}\sum_{k=1}^{22}(X_k - \overline{X})^2 = 0{,}0303$$

— оценки для математического ожидания a и дисперсии σ^2.

Для того чтобы построить доверительный интервал для математического ожидания a, выберем коэффициент доверия $\mathscr{P} = 0{,}98$ и по таблицам распределения Стьюдента с $n - 1 = 21$ степенями свободы найдем $t_{\mathscr{P}} = 2{,}518$. Доверительный интервал для a с коэффициентом доверия $\mathscr{P} = 0{,}98$ имеет концы в точках

$$\overline{X} - t_{\mathscr{P}} \frac{S}{\sqrt{21}} = 3{,}0318 - 0{,}0934 = 2{,}9384$$

$$\overline{X} + t_{\mathscr{P}} \frac{S}{\sqrt{21}} = 3{,}0318 + 0{,}0934 = 3{,}1252.$$

Чтобы определить доверительный интервал для дисперсии σ^2 с тем же коэффициентом доверия $\mathscr{P} = 0{,}98$, найдем по таблицам распределения χ^2 с $(n - 1) = 21$ степенями свободы числа $\chi_1^2 =$

$= 8{,}897$ и $\chi_2^2 = 38{,}932$. Для σ^2 доверительный интервал имеет концы

$$\frac{nS^2}{\chi_2^2} = \frac{22 \cdot 0{,}0289}{38{,}932} = 0{,}0163 \quad \text{и} \quad \frac{nS^2}{\chi_1^2} = \frac{22 \cdot 0{,}0289}{8{,}897} = 0{,}0715.$$

§ 5. ОБРАБОТКА РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ

1. Оценка «истинного значения» измеряемой величины и среднеквадратичной ошибки измерения. Как правило, для получения «истинного значения» a измеряемой величины (а также для оценки средней квадратичной ошибки σ измерения) производят некоторое число n независимых измерений этой величины. Обозначим результаты измерений через $X_1, X_2, \ldots, X_n$. Как известно (см. гл. VII § 4), результат измерения есть случайная величина, распределенная нормально. Предположим, что $MX_i = a$ — условие отсутствия систематической ошибки, и положим $DX_i = \sigma^2$. Таким образом, величины $X_1, X_2, \ldots, X_n$ оказываются независимыми нормально распределенными с параметрами a и σ случайными величинами. Параметры a и σ неизвестны и подлежат определению по результатам измерений, т. е. по выборке. «Истинное значение» a измеряемой величины и средняя квадратичная ошибка σ измерения определяются по формулам

$$a \approx \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

$$\sigma \approx \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}.$$

Для оценки точности данных приближенных равенств можно построить доверительные интервалы (см. § 4 настоящей главы).

2. Сглаживание экспериментальных зависимостей. Пусть величины X и Y связаны функциональной зависимостью вида $Y = \varphi(X)$, причем функция φ нам неизвестна и ее требуется определить по результатам наблюдений.

Предположим, что имеется возможность на опыте измерять значения величины Y в различных точках x_i. Обозначая результат i-го измерения через y_i, имеем,

$$y_i = \varphi(x_i) + \delta_i,$$

где δ_i — случайная ошибка измерения. Таким образом, величина y_i как всякий результат измерения является случайной величиной. Если нанести на график точки (x_i, y_i) и соединить их кривой, вид этой кривой будет отличаться от кривой $Y = \varphi(X)$ из-за наличия случайных ошибок при определении ее ординат. Возникает вопрос: как обработать опытные данные, чтобы наилучшим образом определить зависимость Y от X?

Подобная задача носит название задачи *о сглаживании экспериментальных зависимостей*. Рассмотрим частный, но наиболее важный для приложений случай, когда заранее известно, что функция $\varphi(X)$ принадлежит к некоторому классу функций, зависящему от одного или нескольких параметров $\varphi(X) = \varphi(X, \alpha_1, \alpha_2, \dots, \alpha_k)$. В этом случае задача отыскания наилучшей функции $\varphi(X)$ сводится к задаче наилучшего определения параметров $\alpha_1, \alpha_2, \dots, \alpha_k$ по опытным данным. Словам «наилучшим образом» следует придать точный смысл, что можно сделать по-разному. В соответствии с этим возможны разные способы решения задачи о сглаживании. Слова «наилучшим образом» мы будем понимать в дальнейшем в смысле так называемого метода наименьших квадратов, так как такое понимание является общепринятым и на практике приводит обычно к несложным вычислениям.

Будем говорить, что неизвестные параметры $\alpha_1, \alpha_2, \dots, \alpha_k$ функции $\varphi(X, \alpha_1, \alpha_2, \dots, \alpha_k)$, задающей зависимость $Y = \varphi(X, \alpha_1, \alpha_2, \dots, \alpha_k)$, определены *наилучшим образом в смысле метода наименьших квадратов*, если сумма квадратов отклонений экспериментальных точек y_i от ординат сглаживающей кривой $\varphi(x_i, \alpha_1, \alpha_2, \dots, \alpha_k)$ минимальна, т. е. минимальна величина

$$\delta^2 = \sum_{i=1}^{n} [y_i - \varphi(x_i, \alpha_1, \alpha_2, \dots, \alpha_k)]^2,$$

Для нахождения точки минимума величины δ^2 надо приравнять нулю ее частные производные по $\alpha_1, \alpha_2, \dots, \alpha_k$:

$$\sum_{i=1}^{n} (y_i - \varphi(x_i, \alpha_1, \alpha_2 \ \dots, \alpha_k))\varphi'_{\alpha_1} = 0,$$

$$\sum_{i=1}^{n} (y_i - \varphi(x_i \ \alpha_1, \alpha_2, \dots, \alpha_k))\varphi'_{\alpha_2} = 0, \qquad (10.22)$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\sum_{i=1}^{n} (y_i - \varphi(x_i \ \alpha_1, \alpha_2, \dots, \alpha_k))\varphi'_{\alpha_k} = 0.$$

Таким образом, имеем систему k уравнений с k неизвестными, из которой определяем искомые значения $\alpha_1, \alpha_2, \dots, \alpha_k$. Заметим, что система содержит случайные величины $y_1, y_2, \dots, y_n$, поэтому и решение системы $\overset{*}{\alpha}_1, \overset{*}{\alpha}_2, \dots, \overset{*}{\alpha}_k$ также будет случайно. Величины $\overset{*}{\alpha}_1, \overset{*}{\alpha}_2, \dots, \overset{*}{\alpha}_k$ являются оценками неизвестных параметров $\alpha_1, \alpha_2, \dots, \alpha_k$ по результатам наблюдений. Рассмотренная задача отличается от задачи оценки неизвестных параметров распределения, изученной в § 1—3 настоящей главы, так как величины $y_1, y_2, \dots, y_n$ хотя и предполагаются независимыми, но имеют, вообще говоря, различные распределения.

Рассмотрим оценку по методу наименьших квадратов параметров линейной функции $Y = kX + b$. Пусть из опыта известна совокупность значений (x_i, y_i). Рассмотрим величину

$$\sigma^2 = \sum_{i=1}^{n} (y_i - kx_i - b)^2.$$

Дифференцируя, получим следующую систему:

$$\sum_{i=1}^{n} (y_i - kx_i - b)x_i = 0, \qquad \sum_{i=1}^{n} (y_i - kx_i - b) = 0.$$

Из второго уравнения находим

$$b = \bar{y} - k\bar{x}, \quad \text{где} \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Подставляя найденное значение в первое уравнение и преобразуя его, имеем

$$\sum_{i=1}^{n} y_i x_i - k \sum_{i=1}^{n} x_i^2 - n\bar{x}(\bar{y} - k\bar{x}) = 0,$$

откуда находим

$$k = \frac{\sum_{i=1}^{n} y_i x_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2},$$

или

$$k = \frac{m^*_{x,\,y}}{S_x^2},$$

где

$$m^*_{x,\,y} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}), \quad S_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

(сравните эмпирический смешанный момент величин X и Y и эмпирическую дисперсию величины X).

Таким образом, задача решена, и линейная функция

$$Y = \frac{m^*_{xy}}{S_x^2} X + \bar{y} - \frac{m^*_{xy}}{S_x^2} \bar{x} \tag{10.23}$$

наилучшим образом среди всех линейных функций выражает зависимость Y от X. Уравнение (10.23) похоже на уравнение линейной регрессии случайной величины Y на случайную величину X.

Последнее уравнение можно получить, если в (10.23) вместо $\bar{x}$ и $\bar{y}$ подставить MX и MY, вместо S_x^2 подставить DX, а вместо m^*_{xy} — корреляционный момент μ_{xy}. Следует, однако, помнить, что в равенстве (10.23) величины $x_1, x_2, \ldots, x_n$ не являются случайными, в то время как соответствующие ординаты $y_1, y_2, \ldots, y_n$ случайны.

Заметим, что мы получили бы тот же результат, если бы решали чисто формально задачу нахождения линейной регрессии величины Y на величину X в предположении, что двумерное распределение величины (X, Y) сосредоточено в точках $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, причем в каждой точке сосредоточена масса $\frac{1}{n}$ (см. § 6 настоящей главы).

Пример*. В таблице приведены результаты опытов, в которых исследовалась зависимость глубины h проникновения снаряда в преграду от удельной энергии ε (т. е. энергии, приходящейся на 1 $см^2$ площади соударения).

i	ε_i	h_i
1	41	4
2	50	8
3	81	10
4	104	14
5	120	16
6	139	20
7	154	19
8	180	23
9	208	26
10	241	30
11	250	31
12	269	36
13	301	37

Найдем по методу наименьших квадратов параметры k и b линейной функции $h = k\varepsilon + b$, выражающей зависимость h от ε. Имеем:

$$\overline{\varepsilon} = \frac{1}{13} \sum_{i=1}^{13} \varepsilon_i \approx 164{,}4;$$

$$\overline{h} = \frac{1}{13} \sum_{i=1}^{13} h_i \approx 21{,}1.$$

$$S_\varepsilon^2 = \frac{1}{13} \sum_{i=1}^{13} (\varepsilon_i - \overline{\varepsilon})^2 \approx 6662;$$

$$\overset{*}{m}_{\varepsilon h} = \frac{1}{13} \sum_{i=1}^{13} (h_i - \overline{h})(\varepsilon_i - \overline{\varepsilon}) \approx 826;$$

$$k = \frac{\overset{*}{m}_{\varepsilon h}}{S_\varepsilon^2} \approx 0{,}124.$$

Подставляя полученные значения в равенство (10.23), имеем

$$h - 21{,}1 = 0{,}124\,(\varepsilon - 164{,}4).$$

* Пример заимствован из книги Е. С. Вентцель. Теория вероятностей. М., Физматгиз, 1962.

Мы получили уравнение прямой, наилучшим образом в смысле метода наименьших квадратов изображающей зависимость h от ε.

§ 6. ОЦЕНКА КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ И РЕГРЕССИИ ПО ВЫБОРКЕ

Пусть $\{X_1, Y_1\}$, $\{X_2, Y_2\}$, ..., $\{X_n, Y_n\}$ — выборка из генеральной совокупности двумерной случайной величины (ξ, η). Пусть $M\xi = m_1$, $M\eta = m_2$, $D\xi = \sigma_1^2$, $D\eta = \sigma_2^2$, $\mu_{\xi,\eta} = \mu$, $\rho_{\xi,\eta} = \rho$, $\beta_{1,2} = \frac{\rho\sigma_1}{\sigma_2}$ (β — коэффициент регрессии величины ξ на величину η). *Эмпирическим распределением двумерной выборки* называется двумерное дискретное распределение, сосредоточенное в n выборочных точках (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n), причем в каждой точке сосредоточена масса величины $\frac{1}{n}$. По эмпирическому распределению, так же как по любому распределению, можно подсчитать все числовые характеристики, в том числе коэффициенты корреляции и регрессии. Характеристики, полученные по эмпирическому распределению, называются *эмпирическими, или выборочными, характеристиками*. Основные выборочные характеристики двумерной выборки следующие:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$ — выборочные средние

$$S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$S_2^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$ — выборочные дисперсии,

$$m_{12} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})$$ — эмпирический корреляционный момент.

По этим характеристикам можно вычислить эмпирический коэффициент корреляции

$$r = \frac{m_{12}}{S_1 S_2}$$

и два эмпирических коэффициента регрессии:

$$b_{12} = \frac{rS_1}{S_2} = \frac{m_{12}}{S_2^2} \quad \text{и} \quad b_{12} = \frac{rS_2}{S_1} = \frac{m_{12}}{S_1^2}.$$

Выборочные характеристики используются в качестве оценок для соответствующих характеристик теоретического распределения:

$$m_1 \approx \overline{X}, \quad m_2 \approx \overline{Y}, \quad \sigma_1^2 \approx S_1^2, \quad \sigma_2^2 \approx S_2^2,$$

$$\mu \approx m_{12}, \quad \rho \approx r, \quad \beta_{12} \approx b_{12}, \quad \beta_{21} \approx b_{21}.$$

Все приведенные оценки являются состоятельными, т. е. при $n \to \infty$ сходятся по вероятности к соответствующим теоретическим характеристикам (для первых двух оценок это следует из закона больших чисел, для следующих двух доказано в § 2 настоящей главы). В случае, когда генеральная совокупность имеет нормальное распределение, можно оценить точность приведенных приближенных равенств и построить доверительные интервалы для ρ, β_{12} и β_{21}. В этом случае при больших n можно считать, что выборочный коэффициент корреляции r распределен приближенно нормально с параметрами ρ и $\sigma_r \approx \frac{1-r^2}{\sqrt{n}}$, а выборочный коэффициент регрессии b_{12} распределен приближенно нормально с параметрами β_{12} и $\sigma_{B_{12}} \approx$ $\approx \frac{S_1}{S_2} \cdot \frac{1-r^2}{\sqrt{n}}$.

Для построения доверительных интервалов для ρ и β_{12} следует задать коэффициент доверия $\mathscr{P}$ и найти по таблицам нормального распределения число $t_{\mathscr{P}}$ такое, что

$$\frac{1}{\sqrt{2\pi}} \int_{-t_{\mathscr{P}}}^{t_{\mathscr{P}}} e^{-\frac{x^2}{2}} dx = \mathscr{P}.$$

Тогда имеем

$$P\left(-t_P < \frac{\rho - r}{\frac{1-r^2}{\sqrt{n}}} < t_P\right) \approx \mathscr{P}$$

и

$$P\left(-t_{\mathscr{P}} < \frac{\beta_{12} - b_{12}}{\frac{S_1}{S_2}\,\frac{1-r^2}{\sqrt{n}}} < t_{\mathscr{P}}\right) \approx \mathscr{P},$$

откуда получаем доверительные интервалы для ρ и β_{12} с коэффициентом доверия $\mathscr{P}$:

$$r - t_{\mathscr{P}}\frac{1-r^2}{\sqrt{n}} < \rho < r + t_{\mathscr{P}}\frac{1-r^2}{\sqrt{n}};$$

$$b_{12} - t_{\mathscr{P}}\frac{S_1}{S_2}\,\frac{1-r^2}{\sqrt{n}} < \beta_{12} < b_{12} + t_{\mathscr{P}}\frac{S_1}{S_2}\,\frac{1-r^2}{\sqrt{n}}.$$

Приведенными доверительными интервалами можно пользоваться лишь при больших n. При небольших n и нормально распределенной генеральной совокупности используют преобразование, введенное Фишером:

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r}, \quad \zeta = \frac{1}{2}\ln\frac{1+\rho}{1-\rho}.$$

Фишер показал, что величина Z уже при небольших значениях n распределена приблизительно нормально, причем

$$MZ \approx \zeta + \frac{\rho}{2(n-1)}, \quad DZ \approx \frac{1}{n-3}.$$

Для коэффициента регрессии β_{12} при небольших n вычислим величину

$$t = \frac{S_2\sqrt{n-2}}{S_1\sqrt{1-r^2}}\,(b_{12} - \beta_{12}).$$

Можно показать, что в случае нормально распределенной генеральной совокупности величина t имеет распределение Стьюдента с $(n-2)$ степенями свободы. Доверительный интервал для β_{12} строится обычным образом. С коэффициентом доверия $\mathscr{P}$ имеем

$$b_{12} - t_{\mathscr{P}}\frac{S_1\sqrt{1-r^2}}{S_2\sqrt{n-2}} < \beta_{12} < b_{12} + t_{\mathscr{P}}\frac{S_1\sqrt{1-r^2}}{S_2\sqrt{n-2}},$$

где $t_{\mathscr{P}}$ — доверительный предел, который находится по таблицам распределения Стьюдента с $(n-2)$ степенями свободы.

ГЛАВА XI
ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

§ 1. ПОСТАНОВКА ЗАДАЧИ

Часто функция распределения случайной величины бывает заранее неизвестна, и возникает необходимость ее определения по эмпирическим данным. Однако решение задачи в такой общей постановке вызывает значительные трудности и в большинстве случаев не является необходимым.

Во многих случаях из некоторых дополнительных соображений могут быть сделаны предположения о виде функции распределения $F_\xi(x)$. Простейшим и в то же время наиболее сильным предположением такого рода является предположение, что функция распределения $F_\xi(x)$ есть вполне определенная функция $F_\xi(x) = F(x)$. В тех случаях, когда нет оснований сделать такое предположение, часто оказывается возможным предположить, что функция распределения $F_\xi(x)$ принадлежит к некоторому классу функций, зависящих от одного или нескольких параметров $\alpha_1, \alpha_2, \ldots, \alpha_k$ $(k \geqslant 1)$: $F_\xi(x) = F(x_1, \alpha_1, \alpha_2, \ldots, \alpha_k)$. При этом параметры $\alpha_1, \alpha_2, \ldots, \alpha_k$ неизвестны, и их значения следует получить из опытных данных, т. е. оценить по выборке. Возможны и другие предположения о виде функции распределения $F_\xi(x)$, например $F_\xi(x)$ задает симметричное распределение с медианой в данной точке и т. д. Все эти гипотезы необходимо проверить по эмпирическим данным, т. е. по выборке

Таким образом, необходимы критерии, которые позволяли бы судить, согласуются ли наблюдаемые значения $X_1, X_2, \ldots, X_n$ величины ξ с гипотезой относительно ее функции распределения. Разработка таких критериев, называемых *критериями согласия*, составляет одну из важных задач математической статистики.

Рассмотрим случай, когда гипотетическое распределение полностью задано, т. е. задана функция $F_\xi(x) = F(x)$.

Пусть X_1, X_2, ..., X_n — случайная выборка, т. е. наблюдаемые значения случайной величины ξ, и пусть $F_n^*(x)$ — эмпирическая функция распределения выборки.

Определим некоторую неотрицательную меру D отклонения эмпирической функции распределения $F_n^*(x)$ от предполагаемой (теоретической) функции распределения $F(x)$: $D = D\{F_n^*, F\}$. Величину D можно определить многими способами, в соответствии с которыми получаются различные критерии для проверки интересующей нас гипотезы. Например, можно положить

$$D\{F_n^*, F\} = \sup_x |F_n^*(x) - F(x)|,$$

где $\sup\limits_x$ — верхняя грань по x, или

$$D\{F_n^*, F\} = \int_{-\infty}^{+\infty} \left[F_n^*(x) - F(x)\right]^{2k} g(x)\, dx,$$

где $g(x) \geqslant 0$, $\int\limits_{-\infty}^{+\infty} g(x)\, dx < \infty$ и т. п. В первом случае для проверки нашей гипотезы получим критерий Колмогорова, во втором случае (при $k = 1$) — критерий ω^2 Мизеса.

Величины X_1, X_2, ..., X_n, образующие выборку, в случае справедливости выдвинутой гипотезы можно рассматривать как независимые одинаково распределенные случайные величины с функцией распределения $F(x)$. Но тогда величина D, каким бы образом она ни была определена, является функцией от случайных величин и поэтому сама есть величина случайная.

Предположим, что выдвинутая гипотеза верна, т. е. $F_\xi(x) = F(x)$. Тогда распределение случайной величины D может быть найдено. Зададим число $\varepsilon > 0$ столь малое, что можно считать практически невозможным осуществление события с вероятностью ε в единичном опыте. Считая известным распределение случайной величины D, можно найти такое число D_0, что $P\{D > D_0\} = \varepsilon$*. Пусть имеются фактически наблюдаемые значения X_1, X_2, ..., X_n.

---

* Здесь для простоты предполагается что распределение величины D непрерывного типа

По этим значениям строим функцию $F_n^*(x)$ и вычисляем величину $D\{F_n^*, F\}$. Если полученная при этом величина D окажется больше D_0, то это означает, что событие с вероятностью ε произошло (т. е. произошло событие, которое считаем практически невозможным).

Таким образом, если $D > D_0$, то предположение о справедливости выдвинутой гипотезы привело к выводу, что произошло практически невозможное событие, т. е. гипотеза опровергнута опытом. Если же вычисленная величина $D\{F_n^*, F\}$ окажется меньше ε, то считают, что гипотеза не противоречит опытным данным и, возможно, может быть принята.

Следует отметить, что опровержение гипотезы при $D > D_0$ ни в коем случае не означает логического опровержения, равно как и подтверждение гипотезы в случае $D < D_0$ не означает логического доказательства справедливости гипотезы. В самом деле, событие $D > D_0$ может произойти и в случае справедливости гипотезы, но если ε достаточно мало, то на практике этой возможностью можно пренебречь. Событие $D < D_0$ может осуществиться и в случае, если наша гипотеза неверна, поэтому ее необходимо проверить с помощью (по возможности большего числа) различных критериев, прежде чем считать ее подтвержденной опытными данными.

Число ε, выбор которого зависит от характера задачи, называют *уровнем значимости критерия*, а величину D_0, определяемую из условия $\mathrm{P}\{D > D_0\} = \varepsilon$, — *пределом значимости*.

Распределение величины $D\{F_n^*, F\}$ зависит от n, и вычисление его при конечных значениях n оказывается трудным и нецелесообразным. Вместо этого вычисляют предельное (при $n \to \infty$) распределение величины D и пользуются им в качестве приближения для распределения величины D при достаточно больших значениях n.

В случае, когда гипотетическая функция распределения $F(x, \alpha_1, \alpha_2, \ldots, \alpha_k)$ содержит неизвестные параметры $\alpha_1, \alpha_2, \ldots, \alpha_k$, подлежащие оценке по выборке, также рассматривается некоторая мера $D\{F_n^*, F\}$ отклонения эмпирической функции распределения выборки от теоретической функции распределения $F(x, \alpha_1, \alpha_2, \ldots, \alpha_n)$. Последняя в этом случае сама является величиной случайной, так как $\alpha_1, \alpha_2, \ldots, \alpha_k$ являются функциями наблюдаемых значений, и, следовательно, случайными величинами.

§ 2. КРИТЕРИЙ χ^2 В СЛУЧАЕ ПОЛНОСТЬЮ ОПРЕДЕЛЕННОГО ГИПОТЕТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ

При получении критерия для проверки гипотезы, состоящей в том, что функция распределения $F_\xi(x)$ случайной величины ξ есть вполне определенная функция $F(x)$, мы условились образовывать меру D отклонения эмпирической функции распределения $F_n^*(x)$ выборки $X_1, X_2, \ldots, X_n$ от предполагаемой (теоретической) функции распределения $F(x)$. Наиболее употребительной является мера, введенная Пирсоном, приводящая к так называемому критерию χ^2 Пирсона. Рассмотрим эту меру. Разобьем множество значений величины ξ на r множеств $S_1, S_2, \ldots, S_r$, без общих точек.

Практически такое разбиение обычно осуществляется при помощи $(r-1)$ чисел $a_1 < a_2 < \ldots < a_{r-1}$. При этом правый конец каждого интервала исключается из соответствующего множества, а левый — включается (рис. 11.1).

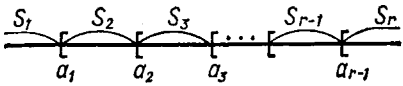

Рис. 11.1

Пусть p_i, $i = 1, 2, \ldots, r$,— вероятность того, что величина ξ принадлежит множеству S_i $\left(\text{очевидно, } \sum\limits_{i=1}^{r} p_i = 1\right)$. Пусть ν_i, $i = 1, 2, \ldots, r$ — количество величин из числа наблюдаемых величин $X_1, X_2, \ldots, X_n$, принадлежащих множеству S_i.

Тогда $\frac{\nu_i}{n}$ — частота попадания величины ξ в множество S_i при n наблюдениях. Очевидно, что

$$\sum_{i=1}^{r} \nu_i = n, \quad \sum_{i=1}^{r} \frac{\nu_i}{n} = 1.$$

Для разбиения, приведенного на рис. 11.1, p_i есть приращение гипотетической функции распределения на множестве S_i, а $\frac{\nu_i}{n}$ — приращение эмпирической функции распределения $F_n^*(x)$ выборки на том же множестве. За меру D отклонения эмпирической функции распределения от теоретической принимается величина, обозначаемая обычно χ^2 и равная

$$\chi^2 = \sum_{i=1}^{r} \frac{n}{p_i}\left(\frac{\nu_i}{n} - p_i\right)^2 = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i}.$$

χ^2— величина случайная и нас интересует ее распределение, вычисленное в предположении, что наша гипотеза верна, т. е. $F_\xi(x) = F(x)$.

Если распределение величины χ^2 известно, то по заданному уровню значимости можно найти предел значимости для проверки принятой гипотезы. Мы не будем вычислять распределение величины χ^2 при каждом значении n, а лишь укажем предельное (при $n \to \infty$) распределение этой величины. Ответ на вопрос о предельном распределении величины χ^2 дает теорема Пирсона.

Теорема Пирсона*. *Какова бы ни была функция распределения $F(x)$ случайной величины ξ, при $n \to \infty$ распределение величины χ^2 стремится к χ^2-распределению с $(r-1)$ степенями свободы, т. е. при $n \to \infty$*

$$P(\chi^2 < x) \to \int_{\infty}^{x} k_{r-1}(t)\,dt$$

в каждой точке x, где $k_{r-1}(x)$ — плотность распределения с χ^2 $(r-1)$ степенями свободы.

С помощью теоремы Пирсона введем критерий для проверки гипотезы. Зададим число $\varepsilon > 0$ такое, что событие с вероятностью ε (ε — уровень значимости) можно считать практически невозможным. По таблице для распределения χ^2 с $(r-1)$ степенями свободы найдем такое число χ_ε^2 (предел значимости), что

$$\int_{\chi_\varepsilon^2}^{\infty} k_{r-1}(x)\,dx = \varepsilon.$$

* Доказательство теоремы см. в кн.: Г. Крамер Математические методы статистики. М., ИЛ, 1948.

Предположим, что n достаточно велико, тогда, по теореме Пирсона, вероятность $\mathcal{P}(\chi^2 > \chi^2_\varepsilon)$ будет приблизительно составлять ε, т. е. событие $\chi^2 > \chi^2_\varepsilon$ можно считать практически невозможным. Таким образом, если гипотеза верна, т. е. $F_\xi(x) = F(x)$, то значения χ^2, превышающие предел значимости χ^2_ε, практически невозможны. Если для данной выборки окажется, что $\chi^2 > \chi^2_\varepsilon$, то гипотеза считается опровергнутой опытными данными; если же $\chi^2 \leqslant \chi^2_\varepsilon$, то опытные данные можно считать совместимыми с принятой гипотезой, однако одного этого еще недостаточно для установления истинности гипотезы.

Применение теоремы Пирсона на практике дает достаточно хорошие результаты во всех случаях, когда величины

$$np_i \geqslant 10, \quad i = 1, 2, \ldots, r.$$

Пример. Случайная величина ξ — показание часов, выставленных в витрине часовщика; проверяемая гипотеза — величина имеет равномерное распределение на интервале (0, 12); интервалы группировки (0, 1), (1, 2), ... , (11, 12), так что $p_i = \frac{1}{12}$ ($i=1, 2, \ldots, 12$). Взяты две выборки $n_1 = n_2 = 500$.

Выборка	Часы											
	0	1	2	3	4	5	6	7	8	9	10	11
1	41	34	54	39	49	45	41	33	37	41	47	39
2	36	47	41	47	49	45	32	37	40	41	37	48

Здесь 0 означает промежуток времени от 0 до 1 ч, 1 — от 1 ч до 2 и т. д.

Для первой выборки имеем $(\chi^2)_1 = 10{,}000$, для второй — $(\chi^2)_2 = 8{,}032$, для составной выборки объема 1000 — $\chi^2 = 9{,}464$. В каждом случае имеем $12 - 1 = 11$ степеней свободы и по таблицам убеждаемся, что согласие хорошее. Можно рассмотреть величину $(\chi^2)_1 + (\chi^2)_2 = 18{,}032$, имеющую 22 степени свободы; согласие получается снова хорошим*.

* Г. Крамер. Математические методы статистики М., ИЛ, 1948.

§ 3. КРИТЕРИЙ χ^2 В СЛУЧАЕ, КОГДА ПО ВЫБОРКЕ ОЦЕНИВАЮТСЯ ПАРАМЕТРЫ

Полностью определенное гипотетическое распределение встречается на практике довольно редко, гораздо чаще распределение $F(x, \alpha_1, \alpha_2, \ldots, \alpha_k)$ содержит некоторые неизвестные параметры $\alpha_1, \alpha_2, \ldots, \alpha_k$, значения которых приходится оценивать по выборке. Гипотеза, подлежащая проверке, состоит в том, что функция распределения $F_\xi(x)$ наблюдаемой величины ξ равна $F(x, \alpha_1, \alpha_2, \ldots, \alpha_k)$ при *некоторых значениях параметров* $\alpha_1, \alpha_2, \ldots, \alpha_k$. Разобьем множество значений величины ξ на r множеств $S_1, S_2, \ldots, S_r$ без общих точек. Обозначим частоты попадания величины ξ в наблюдаемой выборке $X_1, X_2, \ldots, X_n$ в множества $S_1, S_2, \ldots, S_r$ через $\frac{\nu_1}{n}, \frac{\nu_2}{n}, \ldots, \frac{\nu_r}{n}$, а вероятности попадания величины ξ в эти множества при справедливости гипотезы — через $p_1(\alpha_1, \alpha_2, \ldots, \alpha_k), \ldots, p_r(\alpha_1, \alpha_2, \ldots, \alpha_k)$. Составим величину

$$\chi^2 = \sum_{i=1}^{r} \frac{[\nu_i - np_i(\alpha_1, \alpha_2, \ldots, \alpha_k)]^2}{np_i(\alpha_1, \alpha_2, \ldots, \alpha_k)}. \qquad (11.1)$$

Однако воспользоваться теоремой Пирсона нельзя, так как значения $\alpha_1, \alpha_2, \ldots, \alpha_k$ неизвестны. Если же в приведенном выражении для χ^2 заменить величины $\alpha_1, \alpha_2, \ldots, \alpha_k$ их оценками по выборке, т. е. определенными функциями от случайных величин $X_1, X_2, \ldots, X_n$, то величины $p_i(\alpha_1, \alpha_2, \ldots, \alpha_k)$ уже не будут постоянными, а сами станут случайными величинами, поэтому и в этом случае теорему Пирсона применять нельзя.

Поэтому возникает необходимость нахождения предельного при $n \to \infty$ распределения величины χ^2. Предельное распределение величины χ^2 зависит от принятого метода оценки параметров. Задача нахождения предельного при $n \to \infty$ распределения величины χ^2 при наличии оцениваемых по выборке параметров была впервые рассмотрена Фишером, который показал при весьма общих условиях, что предельным при $n \to \infty$ распределением величины χ^2, если неизвестные параметры $\alpha_1, \alpha_2, \ldots, \alpha_k$ оцениваются по выборке по методу минимума χ^2 (или максимума правдоподобия), является распределение χ^2 с $(r - 1 - k)$ степенями свободы. Таким образом, наличие оцениваемых по выборке параметров (если оценка производится одним из перечисленных методов) не меняет характера предельного распределения величины χ^2, а *лишь уменьшает число степеней свободы этого предельного распределения на столько единиц, каково число параметров, оцениваемых по выборке**. Отметим, что если неизвестные параметры оцениваются иными методами, то предельное при $n \to \infty$ распределение величины χ^2 будет, вообще говоря, отличным от χ^2-распределения. Дальнейшие рассуждения при проверке гипотезы аналогичны рассуждениям, проведенным в § 2 настоящей главы.

Пример.** Требуется проверить гипотезу о нормальности закона распределения размера изделий, сделанных на некотором станке. Пусть имеется выборка объема 200 изделий. Эмпирическое среднее $\overline{X}$ и эмпирическое среднее квадратичное отклонение S, полученные по этой выборке, соответственно равны: $X = 4{,}3$ *мк*, $S = 9{,}7$ *мк*. Разобьем множество возможных значений размеров изделий на $r = 10$ интервалов равной длины. В нижеследующей таблице приведены границы соответствующих интервалов и количества ν_i выборочных значений, попавших в эти интервалы (рассматриваются не сами наблюдаемые значения, а их отклонения от номинала).

№ интервала	Границы интервалов, *мк*	Числа ν_i попаданий в интервалы	Оценки вероятностей p_i попадания в интервалы
1	—20, —15	7	0,0233
2	—15, —10	11	0,0475
3	—10, — 5	15	0,0977
4	— 5, 0	24	0,1615
5	0, 5	49	0,1979
6	5, 10	41	0,1945
7	10, 15	26	0,1419
8	15, 20	17	0,0831
9	20, 25	7 } 10	0,0526
10	25, 30	3	

* Точную формулировку теоремы и ее доказательство см. в кн.: Г. Крамер. Математические методы статистики. М., ИЛ, 1948.

** Пример заимствован из кн.: Н. В. Смирнов и И. В. Дунин-Барковский. Краткий курс математической статистики для технических приложений. М., Физматгиз, 1959.

Зная границы интервалов и оценки для параметров a и σ нормального закона, можно с помощью таблиц нормального закона найти оценки для вероятностей p_i попадания выборочных значений в соответствующие интервалы (полученные таким образом оценки приведены в последнем столбце таблицы).

Наконец, по формуле (11.1) вычисляем значение χ^2: $\chi^2 = 7{,}19$. Число k оцениваемых по выборке параметров равно двум, поэтому число степеней свободы распределения χ^2 в нашем случае равно $r' - 1 - 2 = 9 - 1 - 2 = 6$ (число r' интервалов после объединения двух последних интервалов стало равным 9). По таблицам распределения χ^2 с шестью степенями свободы находим $\chi^2_{0,30} = 7{,}2$.

Таким образом, полученное значение оказалось меньше 30% уровня значимости, т. е. в 30% случаев значение χ^2 будет больше наблюдаемого, если гипотеза справедлива. Таким образом, наша гипотеза о нормальности закона распределения не противоречит опытным данным.

В заключение рассмотрим две задачи, которые решаются на основании результатов, полученных в настоящем параграфе.

1) *Задача проверки независимости признаков.* Пусть результаты эксперимента классифицируются по двум признакам, т. е. одновременно наблюдаются две случайные величины (или двумерная случайная величина). Требуется проверить гипотезу независимости признаков, т. е. независимости случайных величин. Пусть различаются r значений первого признака и s значений второго признака, т. е. множество значений первой наблюдаемой величины разбито на r множеств, а второй — на s множеств, не содержащих общих точек. Обозначим через p_{ij} вероятность того, что первая величина принадлежит множеству с номером i, а вторая — с номером j. Тогда гипотеза независимости эквивалентна гипотезе о существовании $r + s$ постоянных p_i и q_j таких, что

$$p_{ij} = p_i q_j; \quad \sum_{i=1}^{r} p_i = \sum_{j=1}^{s} q_j = 1. \qquad (*)$$

В соответствии с гипотезой совместное распределение двух величин содержит $r + s - 2$ неизвестных параметров, так как с помощью соотношений (*) два из $r + s$ параметров могут быть выражены через остальные $r + s - 2$ параметра.

Обозначим через ν_{ij} общее количество таких случаев когда первая величина принадлежит множеству с номером i, а вторая — с номером j. Очевидно,

$$\sum_{i=1}^{r} \sum_{j=1}^{s} \nu_{ij} = n.$$

Рассмотрим величину

$$\chi^2 = \sum_{i, j=1,1}^{r, s} \frac{(\nu_{ij} - np_i q_j)^2}{np_i q_j},$$

которая, согласно теореме Фишера, имеет в пределе при $n \to \infty$ распределение χ^2 с $rs - 1 - (r + s - 2) = (r - 1)(s - 1)$ степенями свободы (в случае справедливости гипотезы и если оценки для величин p_i и q_j получены методом максимума правдоподобия). Соответствующая система уравнений для определения величин p_i и q_j имеет вид

$$\sum_{j=1}^{s} \left(\frac{\nu_{ij}}{p_i} - \frac{\nu_{rj}}{p_r} \right) = 0, \quad i = 1, 2, \ldots, r - 1,$$

$$\sum_{i=1}^{r} \left(\frac{\nu_{ij}}{q_j} - \frac{\nu_{is}}{q_s} \right) = 0, \quad j = 1, 2, \ldots, s - 1.$$

Решением полученной системы уравнений являются величины

$$p_i = \frac{1}{n} \sum_{j=1}^{s} \nu_{ij} = \frac{\nu_i}{n}, \qquad q_j = \frac{1}{n} \sum_{i=1}^{r} \nu_{ij} = \frac{\varkappa_j}{n}.$$

Подставляя найденные значения p_i и q_j в выражение для χ^2, окончательно получим величину

$$\chi^2 = n \sum_{i, j=1,1}^{r, s} \frac{\left(\nu_{ij} - \frac{\nu_i \varkappa_j}{n} \right)^2}{\nu_i \varkappa_j} = \sum_{i, j=1,1}^{r, s} \left(\frac{\nu_{ij}^2}{\nu_i \varkappa_j} - 1 \right), \quad (11.2)$$

имеющую в пределе распределение χ^2 с $(r - 1)(s - 1)$ степенями свободы. Далее критерий для проверки гипотезы независимости строится обычным образом. Если результаты эксперимента расположены в таблице, где на пересечении i-й строки и j-го столбца стоят числа ν_{ij}, то такую таблицу называют *таблицей сопряженности признаков.*

2) *Задача проверки однородности данных.* Пусть имеется s выборок, состоящих каждая из n_1, n_2, ..., n_s наблюдений, т. е. s независимых случайных величин ξ_1, ξ_2, ... , ξ_s, и пусть мы располагаем n_i наблюдаемыми значениями величины ξ_i.

Требуется проверить гипотезу об однородности данных, т. е. гипотезу о том, что случайные величины ξ_1, ξ_2, ... , ξ_n распределены одинаково. Пусть множество возможных значений рассматриваемых величин разбито на r множеств без общих точек. Обозначим через ν_{ij} количество наблюдаемых значений, принадлежащих множеству с номером i в выборке с номером j. Помещая числа ν_{ij} в таблицу (в i-ю строку и j-й столбец), получим таблицу формально такого же вида, что и таблица сопряженности признаков. Для проверки интересующей нас гипотезы вычислим величину χ^2 по формуле (11.2) и воспользуемся тем обстоятельством, что величина χ^2 имеет распределение χ^2 с $(r-1)(s-1)$ степенями свободы (если гипотеза справедлива).

Строго говоря, последнее утверждение не следует непосредственно из теоремы Фишера и требует специального доказательства*.

§ 4. КРИТЕРИЙ СОГЛАСИЯ КОЛМОГОРОВА

Вернемся к случаю, когда гипотетическая функция распределения полностью определена, т. е. задана функция $F(x) = F_\xi(x)$, которую предположим непрерывной. Мера D отклонения эмпирической функции распределения $F_n^*(x)$ выборки X_1, X_2, ... , X_n от гипотетической функции распределения $F(x)$, предложенная А. Н. Колмогоровым, определяется следующим образом:

$$D_n = D\{F_n^*, F\} = \sup_x \left| F_n^*(x) - F(x) \right|,$$

где $\sup_x$ — верхняя грань множества по всевозможным значениям x. Очевидно, D_n — величина случайная, и нас, как и в ранее разобранных случаях, интересует ее предельное при $n \to \infty$ распределение, вычисленное в предположении, что гипотеза справедлива. Ответ на постав-

* См. кн.: Г. Крамер. Математические методы статистики. М., ИЛ, 1948.

ленный вопрос дает следующая теорема, которую приводим без доказательства.

Теорема Колмогорова. *Если функция распределения $F(x)$ непрерывна, то при $n \to \infty$*

$$P\left(\sqrt{n}\ D_n < x\right) \to K(x) = \begin{cases} \sum\limits_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2x^2} & \text{при } x > 0, \\ 0 & \text{при } x \leqslant 0. \end{cases}$$

Функция распределения $K(x)$ табулирована ввиду ее важности для практики*. Схема применения теоремы Колмогорова к построению критерия согласия находится в полном соответствии с изложенным в § 1 настоящей главы.

§ 5. КРИТЕРИЙ ω^2 МИЗЕСА

Предположим, что гипотетическая функция распределения полностью задана: $F_\xi(x) = F(x)$. Р. Мизес предложил в качестве меры D отклонения эмпирической функции распределения $F_n^*(x)$ выборки от (теоретической) функции $F(x)$ величину

$$\omega^2 = \int\limits_{-\infty}^{+\infty} \left[F_n^*(x) - F(x)\right]^2 dF(x).$$

Предельное при $n \to \infty$ распределение величины ω^2 было получено Н. В. Смирновым, это распределение оказалось не зависящим от вида непрерывной функции распределения $F(x)$. На основании теоремы Н. В. Смирнова соответствующий критерий согласия, называемый ω^2-критерием (Мизеса), строится обычным образом. В случае, если гипотетическое распределение содержит неизвестные параметры, оцениваемые по выборке, предельное при $n \to \infty$ распределение соответствующей величины ω^2 может быть также найдено и может быть построен критерий согласия.

---

* Соответствующие таблицы можно найти в кн.: Б. В. Гнеденко. Курс теории вероятностей. М., «Наука», 1969.

§ 6. ПРЕДЕЛЬНЫЕ РАСПРЕДЕЛЕНИЯ ВЕЛИЧИН, ВСТРЕЧАЮЩИХСЯ В МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Пусть $X_1, X_2, \ldots, X_n$ — выборка, т. е. независимые, одинаково распределенные случайные величины. В математической статистике обычно рассматривают симметричные функции выборочных значений, т. е. функции, которые не меняются при любой перестановке аргументов. Всякую симметричную функцию от величин $X_1, X_2, \ldots, X_n$ можно рассматривать как число T, поставленное в соответствие эмпирической функции распределения $F_n^*(x)$ выборки $X_1, X_2, \ldots, X_n$, или, как говорят, *функционал T от эмпирической функции распределения* $F_n^*(x)$: $T = T[F_n^*]$.

Предположим, что в область определения функционала T, являющуюся множеством функций, входит как функция распределения $F(x)$ генеральной совокупности, называемая в дальнейшем теоретической функцией распределения, так и все функции вида:

$$F_n^t(x) = F(x) + t\left(F_n^*(x) - F(x)\right) \quad (0 \leqslant t \leqslant 1).$$

При $t = 0$ имеем $F_n^{(0)}(x) = F(x)$;

при $t = 1$ имеем $F_n^{(1)}(x) = F_n^*(x)$.

Рассмотрим функцию $T[F_n^{(t)}] = f(t)$ в интервале $0 \leqslant t \leqslant 1$. Предполагая, что функция $f(t)$ имеет производные всех порядков до $(k+1)$ включительно, можно по формуле Тейлора разложить функцию $f(t) = T\left[F_n^{(t)}\right]$:

$$f(t) - f(0) = \frac{f'(0)}{1!} t + \frac{f''(0)}{2!} t^2 + \cdots + \frac{f^{(k)}_{(0)}}{k!} t^k + \frac{f^{(k+1)}(\theta_t)}{(k+1)!} t^{k+1},$$

где $0 \leqslant \theta_t \leqslant t$.

Подставляя $t = 1$, имеем

$$f(1) - f(0) = T\left[F_n^*\right] - T[F] = \frac{f'(0)}{1!} + \frac{f''(0)}{2!} + \cdots + \frac{f^{(k)}(0)}{k!} + \frac{f^{(k+1)}(\theta_1)}{(k+1)!}, \qquad (11.3)$$

где $0 \leqslant \theta_1 \leqslant 1$.

Левая часть полученной формулы представляет собой разность между интересующим нас значением функционала от эмпирической функции распределения $T\left[F_n^*\right]$ и неслучайной величиной $T[F]$, а в правой, вообще говоря, находятся случайные величины. Большинство функционалов, встречающихся в математической статистике, таковы, что функция $f_1(t)$ дифференцируема необходимое число раз и более того, ее производные в нуле можно представить в виде

$$f^{(m)}(0) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \psi(x_1, x_2, \dots, x_m) \prod_{i=1}^{m} d\left[F_n^*(x) - F(x)\right],$$

$$m = 1, 2, \dots, k.$$

При $m = 1$ имеем

$$f'(0) = \int_{-\infty}^{+\infty} \psi(x)\, d\left[F_n^*(x) - F(x)\right] = \int_{-\infty}^{+\infty} \psi(x)\, dF_n^*(x) -$$

$$- \int_{-\infty}^{+\infty} \psi(x)\, dF(x) = \frac{1}{n} \sum_{i=1}^{n} \psi(X_i) - M\psi(X_i).$$

Таким образом, первый член разложения Тейлора функции $f_1(t)$ при $t = 0$ является с точностью до числового множителя и постоянного слагаемого суммой независимых, одинаково распределенных случайных величин $\psi(X_1), \psi(X_2), \dots, \psi(X_n)$. Закон распределения такой суммы при соответствующей нормировке стремится к нормальному закону, когда $n \to \infty$ (согласно центральной предельной теореме) при условии существования дисперсии

$$D\psi(X_i) = \int_{-\infty}^{+\infty} \left[\psi(x) - \int_{-\infty}^{+\infty} \psi(y)\, dF(y)\right]^2 dF(x),$$

а именно величина

$$\frac{\sum_{i=1}^{n} \psi(X_i) - M\psi(X_i)}{\sqrt{n}\,\sqrt{D\psi(X_i)}}$$

имеет в пределе при $n \to \infty$ нормальное распределение с параметрами нуль и единица.

Для широкого класса функционалов можно показать, что предельное при $n \to \infty$ распределение разности $T\left[F_n^*\right] - T[F]$ совпадает с предельным распределением первого отличного от нуля члена в разложении Тейлора. Если в разложении (11.3) первый член отличен от нуля (именно этот случай надо считать основным), то согласно изложенному выше предельное распределение величины $T\left[F_n^*\right] - T[F]$ при $n \to \infty$ оказывается нормальным [если при этом дисперсия $D\psi(X_i)$ конечна].

Равенство нулю первого члена в разложении Тейлора (11.3) можно ожидать лишь при отдельных, исключительных для данного функционала теоретических функциях распределения $F(x)$. Типичным же случаем является необращение в нуль первого члена. Таким образом, вообще говоря, предельное при $n \to \infty$ распределение функционала $T\left[F_n^*\right] - T[F]$ оказывается нормальным при соответствующей нормировке. При этом нормирующий множитель совпадает с нормирующим множителем первого члена формулы Тейлора и равен $\frac{\sqrt{n}}{\sqrt{D\psi(x_i)}}$.

Главный член математического ожидания $M\left(T\left[F_n^*\right] - T[F]\right)^2$ совпадает с дисперсией первого отличного от нуля члена в разложении Тейлора, в нашем случае с $D\left[\frac{1}{n}\sum_{i=1}^{n}\psi(X_i) - M\psi(X_i)\right] =$ $= \frac{1}{n} D\psi(X_i)$.

В случае функционалов, являющихся оценками по выборке неизвестных параметров распределения, следует ожидать нормальное распределение в пределе, так как способ оценки параметра не предполагает известной теоретическую функцию распределения и поэтому нельзя ожидать обращения в нуль производной $\frac{d}{dt} T\left[F_n^{(t)}\right]\Big|_{t=0}$.

Как известно, оценки по методу максимума правдоподобия, по методу минимума χ^2, минимума ω^2, а также оценки по методу моментов распределены в пределе при $n \to \infty$ нормально.

В случае так называемых критериев согласия функционал $T\left[F_n^*\right]$ является некоторой мерой уклонения эмпирической функции распределения от теоретической. Теоретическая функция распределения играет особую роль, поэтому неудивительно, что в этом случае первый член разложения Тейлора может оказаться равным нулю. Для функционалов, связанных с критериями χ^2, ω^2, первым отличным от нуля членом оказывается второй член:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \psi(x_1, x_2) \prod_{i=1}^{2} d\left[F_n^*(x_i) - F(x_i)\right].$$

Предельное при $n \to \infty$ распределение последней величины зависит от вида функции $\psi(x_1, x_2)$ и может быть найдено в различных конкретных случаях. Это распределение оказывается отличным от нормального (таковы распределения χ^2, ω^2 и др).

К функционалу $T\left[F_n^*\right] = \sup_x \left|F_n^*(x) - F(x)\right|$, связанному с критерием согласия Колмогорова, изложенное выше неприменимо, так как соответствующая функция $f(t) = T\left[F_n^{(t)}\right]$ оказывается недифференцируемой. Рассмотрим несколько примеров. При их решении будем предполагать выполненными все необходимые условия регулярности, равно как и законность всех производимых чисто формально операций (например, дифференцирование под знаком интеграла и т. д.).

Пример 1. Рассмотрим функцию Φ от эмпирического момента порядка p:

$$T\left[F_n^*\right] = \Phi\left(\int_{-\infty}^{+\infty} x^p dF_n^*(x)\right).$$

Дифференцируя функцию $f(t) = T\left[F_n^{(t)}\right] = \Phi\left(\int_{-\infty}^{+\infty} x^p dF_n^{(t)}(x)\right)$ формально по t, при $t = 0$ имеем

$$f'(0) = \frac{d}{dt} T\left[F_n^{(t)}\right]\Bigg|_{t=0} = \Phi'\left(\int_{-\infty}^{+\infty} x^p\, dF(x)\right)\int_{-\infty}^{+\infty} x^p d\left[F_n^*(x) - F(x)\right].$$

Если $\Phi'\left(\int_{-\infty}^{+\infty} x^p dF(x)\right) \neq 0$, то первый член разложения Тейлора отличен от нуля, сводится к сумме независимых слагаемых и, следовательно, в пределе при $n \to \infty$ распределение функции от эмпирического момента оказывается нормальным. Нормирующий множитель находится, как было указано выше.

Пример 2. Рассмотрим эмпирический центральный момент k-го порядка

$$T\left[F_n^*\right] = a_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k = \int_{-\infty}^{+\infty}\left[x - \int_{-\infty}^{+\infty} y dF_n^*(y)\right]^k dF_n^*(x).$$

Имеем

$$f(t) = T\left[F_n^{(t)}\right] = \int_{-\infty}^{+\infty}\left[x - \int_{-\infty}^{+\infty} y dF_n^{(t)}(y)\right]^k dF_n^{(t)}(x).$$

Дифференцируя по t формальным образом, при $t = 0$ имеем

$$f'(0) = \frac{d}{dt} T\left[F_n^{(t)}\right]\Big|_{t=0} = \int_{-\infty}^{+\infty}\left[x - \int_{-\infty}^{+\infty} y dF(y)\right]^k d\left[F_n^*(x) - F(x)\right] + k\int_{-\infty}^{+\infty}\left[x - \int_{-\infty}^{+\infty} y dF(y)\right]^{k-1} dF(x)\cdot\left(-\int_{-\infty}^{+\infty} y d\left[F_n^*(y) - F(y)\right]\right) = \int_{-\infty}^{+\infty}\left\{\left[x - \int_{-\infty}^{+\infty} y dF(y)\right]^k + Cx\right\} d\left[F_n^*(x) - F(x)\right],$$

где $\quad C = -k \int \left(x - \int\limits_{-\infty}^{+\infty} y dF(y) \right)^{k-1} dF(x).$

Таким образом, первый член разложения Тейлора отличен от нуля, сводится к сумме независимых слагаемых и предельное при $n \to \infty$ распределение функционала $T\left[F_n^*\right] = a_k$ асимптотически нормально.

Пример 3. Рассмотрим оценку по методу моментов неизвестного параметра α теоретической функции распределения $F(x, \alpha)$. Обозначим через $p(x, \alpha) = F'_x(x, \alpha)$ соответствующую плотность распределения, тогда для определения оценки параметра α имеем следующее уравнение:

$$\int\limits_{-\infty}^{+\infty} y^k p(y, \alpha)\, dy = \int x^k dF_n^*(x),$$

или

$$\int \left[x^k - \int\limits_{-\infty}^{+\infty} y^k p(y, \alpha)\, dy \right] dF_n^*(x) = 0.$$

Положим

$$H(t, \alpha) = \int \left[x^k - \int\limits_{-\infty}^{+\infty} y^k p(y, \alpha)\, dy \right] dF_n^{(t)}(x)$$

и рассмотрим уравнение $H(t, \alpha) = 0$, определяющее при определенных условиях величину α как функцию от t. Нас интересует значение α, соответствующее $t = 1$, так как уравнение для определения α может быть записано в виде $H(1, \alpha) = 0$. Оценка α является функционалом от эмпирической функции распределения $\alpha\left[F_n^*\right]$, и уравнение $H(t, \alpha) = 0$ определяет неявным образом значение этого функционала для $F_n^{(t)}(x)$, т. е. $\alpha\left[F_n^{(t)}\right] = f(t)$. По известному правилу дифференцирования неявных функций имеем

$$f'(0) = \frac{d}{dt} \alpha\left[F_n^{(t)}\right]\Big|_{t=0} = -\left.\frac{\dfrac{\partial H}{\partial t}}{\dfrac{\partial H}{\partial \alpha}}\right|_{t=0} = C^{-1} \int\limits_{-\infty}^{+\infty} \left[x^k - \right.$$

$$\left. - \int\limits_{-\infty}^{+\infty} y^k p(y, \alpha)\, dy \right] d\left[F_n^*(x) - F(x)\right],$$

где

$$C = \int \left[x^k - \int\limits_{-\infty}^{\infty} \left[y^k p'_\alpha(y, \alpha)\, dy \right] dF(x) \right].$$

Таким образом, первый член разложения Тейлора отличен от нуля и сводится к сумме независимых слагаемых, поэтому предельное при $n \to \infty$ распределение рассматриваемого функционала оказывается нормальным.

Пример 4. Рассмотрим оценку по моменту максимума правдоподобия неизвестного параметра α. Для определения параметра воспользуемся следующим уравнением:

$$\frac{\partial}{\partial\alpha} \ln (p (X_1, \alpha) \cdot p (X_2\alpha) \ldots p (X_3, \alpha)) = \frac{p'_\alpha (X_1, \alpha)}{p (X_1, \alpha)} +$$

$$+ \frac{p'_\alpha (X_2, \alpha)}{p (X_2, \alpha)} + \cdots + \frac{p'_\alpha (X_n, \alpha)}{p (X_n, \alpha)} = 0$$

или

$$\int\limits_{-\infty}^{+\infty} \frac{p'_\alpha (x, \alpha)}{p (x, \alpha)} \, dF_n^* (x) = 0.$$

Рассмотрим уравнение

$$H (t, \alpha) = \int\limits_{-\infty}^{+\infty} \frac{p'_\alpha (x, \alpha)}{p (x, \alpha)} \, dF_n^{(t)} (x) = 0,$$

которое при $t = 1$ переходит в уравнение для определения α. Предположим, что уравнение $H (t, \alpha) = 0$ определяет неявным образом значение функционала $\alpha \left[F_n^{(t)} \right]$ для всех $t \in [0, 1]$. Дифференцируя функцию $f (t) = \alpha \left[F_n^{(t)} \right]$ по t как неявную функцию, имеем

$$f' (0) = \frac{d}{dt} \, \alpha \left[F_n^{(t)} \right] \Big|_{t=0} = - \frac{\dfrac{\partial H}{\partial t}}{\dfrac{\partial H}{\partial \alpha}} =$$

$$= C^{-1} \int\limits_{-\infty}^{+\infty} \frac{p'_\alpha (x, \alpha)}{p (x, \alpha)} \, d \left[F_n^* (x) - F (x) \right]$$

где

$$C = \int\limits_{-\infty}^{+\infty} \frac{\left[p'_\alpha (x, \alpha) \right]^2}{p (x, \alpha)} \, dx.$$

Итак, оценка по методу максимума правдоподобия распределена асимптотически нормально. Главный член математического ожидания $M \left\{ \alpha \left[F_n^* \right] - \alpha [F] \right\}^2$ совпадает с дисперсией величины

$\left.\frac{d\alpha}{dt}\right|_{t=0}$ и равен $M\left(\left.\frac{d\alpha}{dt}\right|_{t=0}\right)^2 = (Cn)^{-1}$. Зная, что оценка максимального правдоподобия является асимптотически эффективной, можно сравнивать с ней другие оценки, вычисляя дисперсии первых членов. Таким образом, убеждаемся, что оценка по методу моментов не является асимптотически эффективной.

Пример 5. Рассмотрим величину

$$\omega^2\left[F_n^*\right] = \int_{-\infty}^{+\infty}\left[F_n^*(x) - F(x)\right]^2 dF(x).$$

Для $\omega^2\left[F_n^{(t)}\right]$ имеем

$$f(t) = \omega^2\left[F_n^{(t)}\right] = \int_{-\infty}^{+\infty}\left[F_n^{(t)}(x) - F(x)\right]^2 dF(x) =$$

$$= t^2\int_{-\infty}^{+\infty}\left[F_n^*(x) - F(x)\right]^2 dF(x) = t^2\int_{-\infty}^{+\infty}\int_{-\infty}^{x} d\left[F_n^*(s) -\right.$$

$$\left.- F(s)\right]\cdot\int_{-\infty}^{x} d\left[F_n^*(t) - F(t)\right] dF(x) =$$

$$= t^2\int_{-\infty}^{+\infty} dF(x)\int_{-\infty}^{+\infty}\chi_{(-\infty,\, x)}(s)\, dF_n^*(s) -$$

$$\left.- F(s)\right]\int_{-\infty}^{+\infty}\chi_{(-\infty,\, x)}(t)\, d\left[F_n^*(t) - F(t)\right] =$$

$$= t^2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}\chi_{(-\infty,\, x)}(s)\cdot\chi_{(-\infty,\, x)}(t)\, dF(x)\right)\times$$

$$\times d\left[F_n^*(t) - F(t)\right]\times d\left[F_n^*(s) - F(s)\right] =$$

$$= t^2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F(\max(s,t))\, d\left[F_n^*(t) - F(t)\right]\cdot d\left[F_n^*(s) - F(s)\right],$$

где $\chi_{(-\infty,\, x)}(s) = \begin{cases} 1 & \text{при } s \leqslant x; \\ 0 & \text{при } s > x \end{cases}$ — так называемая характеристическая функция множества $(-\infty, x)$.

Очевидно, что

$$f(0) = \omega^2[F] = 0,\ f'(0) = 0,$$

$$\left.\frac{d^2\omega^2\left[F_n^{(t)}\right]}{dt^2}\right|_{t=0} = 2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} F[\max(s,t)]\ d\left[F_n^*(t) - F(t)\right]\cdot d\left[F_n^*(s) - F(s)\right].$$

Все производные выше второго порядка равны нулю, так что в нашем случае функционал совпадает с величиной $\frac{f''(0)}{2}$, которая приведена к виду

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \psi(x_1, x_2)\prod_{i=1}^{2} d\left[F_n^*(x_i) - F(x_i)\right],$$

причем $\psi(s, t) = F[\max(s, t)]$. Предельное при $n \to \infty$ распределение этой величины отлично от нормального и называется ω^2-распределением.

Аналогично можно рассмотреть величину

$$\omega^2 = \int_{-\infty}^{+\infty}\left[F_n^*(x) - F(x, a)\right]^2 dF(x, a)$$

в случае, когда неизвестный параметр α оценивается по выборке, например, методом максимума правдоподобия. При этом дифференцируется по t величина

$$f(t) = \omega^2\left[F_n^{(t)}\right] = \int_{-\infty}^{+\infty}\left\{F_n^*(x) - F\left(x, a\left[F_n^{(t)}\right]\right)\right\}^2 p\left(x, a\left[F_n^{(t)}\right]\right)dx,$$

и первый член разложения Тейлора также является нулевым, вид второго члена зависит от способа оценки параметра α, от этого способа будет зависеть и предельное при $n \to \infty$ распределение рассматриваемого функционала.

Таким же методом можно исследовать величину χ^2, рассмотренную в § 2, 3 настоящей главы, оценку по методу минимума χ^2 и многие другие величины, используемые в математической статистике при оценке неизвестных параметров и при проверке статистических гипотез.

В заключение отметим, что изложенный подход к исследованию предельных свойств, в том числе и предельных распределений функционалов от эмпирических функций распределения, был предложен Мизесом и может быть строго обоснован.

§ 7. ВЫБОР КРИТЕРИЯ ДЛЯ ПРОВЕРКИ СТАТИСТИЧЕСКОЙ ГИПОТЕЗЫ. НАИБОЛЕЕ МОЩНЫЕ КРИТЕРИИ

Для проверки статистической гипотезы мы в соответствии со сказанным в § 1 настоящей главы конструируем некоторую неотрицательную меру D отклонения эмпирической функции распределения от предполагаемой теоретической функции распределения. Величина D как функция от выборочных значений $X_1, X_2, \ldots, X_n$ является случайной величиной; распределение вероятностей величины D определяется однозначно теоретической функцией распределения. Задавая $\varepsilon > 0$, находим такое число D_0, что

$$P(D > D_0) = \varepsilon.$$

Принятая гипотеза отвергается, если D, вычисленное по выборке, больше чем D_0, и принимается в противном случае. Образуя величину D различными способами, получают различные критерии для проверки гипотезы (см. § 2—5 настоящей главы). Пусть выбрана некоторая вполне определенная величина D, измеряющая расхождение между теоретической и эмпирической функциями распределения. Можно, очевидно, бесчисленным множеством способов выбрать множество S такое, что

$$P(D \in S) = \varepsilon$$

(здесь, как и ранее, для простоты предполагается, что распределение величины D принадлежит к непрерывному типу). Если гипотеза отвергается при $D \in S$ и принимается в противном случае, то вероятность отвергнуть принятую гипотезу, когда она в действительности верна, для всех таких множеств S будет равна ε. Таким образом, с помощью одной и той же величины D можно строить различные критерии для проверки гипотезы, по-разному выбирая множество S. Интуитивно представляется наиболее естественным выбрать в качестве S бесконечный интервал $(D_0, +\infty)$, т. е. поступить в соответствии с изложенным в § 1 настоящей главы.

Для доказательства правильности сделанного выбора необходимо сравнить между собой различные критерии, отвечающие одному и тому же значению ε, что можно сделать, применяя теорию Неймана—Пирсона, в основе которой лежит следующий общий принцип.

При решении на основании данных выборки вопроса о том, отвергнуть или принять некоторую гипотезу H, можно ошибиться в том и в другом случае. Если гипотеза H отвергается, т. е. если $D \in S$, то с вероятностью ε можно отвергнуть гипотезу, которая в действительности верна. Такая ошибка называется *ошибкой первого рода* и ее вероятность равна ε. Если же гипотеза принимается, то также можно ошибиться, приняв ложную гипотезу. Это будет в том случае, если в действительности справедлива некоторая другая гипотеза H_1. Такая ошибка называется *ошибкой второго рода*. Вероятность ошибки второго рода равна вероятности того, что $D \overline{\in} S$, вычисленной в предположении, что верна гипотеза H_1. При построении критерия следует стремиться уменьшить вероятности ошибок как первого, так и второго рода. Из двух критериев, соответствующих одной и той же вероятности ε ошибки первого рода, следует предпочесть тот, который дает меньшую вероятность ошибки второго рода.

Обозначим через $\mathrm{P}(A \mid H)$, соответственно $\mathrm{P}(A \mid H_1)$, вероятность события A, вычисленную в предположении, что верна гипотеза H (соответственно H_1). Тогда:

вероятность ошибки первого рода $\varepsilon = \mathrm{P}(D \in S|H)$,
вероятность ошибки второго рода $\mathrm{P}(D \overline{\in} S|H_1)$.

Мощностью некоторого критерия относительно гипотезы H_1 называется вероятность отвергнуть гипотезу H, если верна гипотеза H_1, т. е.

$$\mathrm{P}(D \in S|H_1).$$

Очевидно,

$$\mathrm{P}(D \overline{\in} S|H_1) + \mathrm{P}(D \in S|H_1) = 1,$$

т. е. сумма вероятности ошибки второго рода и мощности критерия относительно H_1 равна единице. Поэтому задача отыскания критерия, обладающего возможно меньшей вероятностью ошибки второго рода при заданной вероятности ошибки первого рода, есть задача отыскания критерия, обладающего возможно большей мощностью, т. е. наиболее мощного критерия.

Рассмотрим применение изложенного общего принципа на примере. Пусть имеется выборка $X_1, X_2, \ldots, X_n$ из генеральной совокупности, плотность распределения вероятностей которой $p(x, \alpha)$ (α может быть как числом, так и k-мерным вектором). Критерий для проверки некоторой гипотезы H может быть сформулирован следующим образом. Пусть $V \subset R^n$ — некоторое множество в n-мерном пространстве; гипотеза H отвергается, если выборочная точка $(X_1, X_2, \ldots, X_n) \in V$, и принимается в противном случае. В самом деле, если строить критерий с помощью величины $D = D(X_1, X_2, \ldots, X_n)$, то условие $D \in S$, где S — множество на прямой, эквивалентно условию $(X_1, X_2, \ldots, X_n) \in V$, где V — прообраз множества S при отображении $D = D(X_1, X_2, \ldots, X_n)$. Множество V называется *критическим множеством* данного критерия.

Вероятность отвергнуть гипотезу H совпадает с вероятностью попадания выборочной точки в критическое множество V. Пусть гипотеза H состоит в том, что точка α совпадает с данной точкой α_0. В соответствии с изложенным выше мы должны так подобрать множество V, чтобы вероятность попадания выборочной точки в V была мала, если гипотеза H верна, и велика в противном случае. Зададим вероятность ε ошибки первого рода, т. е. $\varepsilon = \mathrm{P}(V \mid \alpha_0)$.

Следует выбрать такое множество V, чтобы мощность $\mathrm{P}(V|\alpha_1)$ критерия была возможно большей для любого допустимого значения $\alpha_1 \neq \alpha_0$. Фиксируем некоторое допустимое значение $\alpha_1 \neq \alpha_0$. Без ограничения общности можно предположить, что $p(x, \alpha_0)$ и $p(x, \alpha_1)$ конечны и определены для всех x. Рассмотрим множество X тех точек $(X_1, X_2, \ldots, X_n)$, для которых

$$p(X_1, \alpha_1)\, p(X_2, \alpha_1), \ldots, p(X_n, \alpha_1) \geqslant$$

$$\geqslant Cp(X_1, \alpha_0)\, p(X_2, \alpha_0), \ldots, p(X_n, \alpha_0),$$

где $C \geqslant 0$ — произвольное неотрицательное число.

Плотность распределения вероятностей

$$p(x_1, \alpha_0)\, p(x_2, \alpha_0) \ldots p(x_n, \alpha_0)$$

[соответственно $p(x_1, \alpha_1)\, p(x_2, \alpha_1) \ldots p(x_n, \alpha_1)$] называется функцией правдоподобия гипотезы H (соответственно H_1).

Отношение

$$\frac{p(x_1, \alpha_1)\, p(x_2, \alpha_1) \ldots p(x_n, \alpha_1)}{p(x_1, \alpha_0)\, p(x_2, \alpha_0) \ldots p(x_n, \alpha_0)}$$

называют *отношением правдоподобия*, а критерий, основанный на критическом множестве X, *называется критерием отношения правдоподобия*.

Множество X зависит от C, поэтому $\mathrm{P}(X|\alpha_0) = g(C)$ является функцией от C. Функция $g(C)$ невозрастающая, так как множества X_1 и X_2, соответствующие двум различным значениям C_1 и C_2, связаны соотношением

$$X_{C_1} \supset X_{C_2} \text{ при } C_1 < C_2.$$

Очевидно, $g(C) \geqslant 0$, $g(0) = 1$. Покажем, что $g(C) \leqslant \frac{1}{C}$, и таким образом, $g(C) \to 0$ при $C \to \infty$. В самом деле, имеем

$$Cg(C) = C\mathrm{P}(X \mid \alpha_0) = C \int\limits_X p(x_1, \alpha_0)\, p(x_2, \alpha_0) \ldots$$

$$\ldots p(x_n, \alpha_0)\, dx_1 dx_2 \ldots dx_n \leqslant \int\limits_X p(x_1, \alpha_1)\, p(x_2, \alpha_1) \ldots$$

$$\ldots p(x_n, \alpha_1)\, dx_1 dx_2 \ldots dx_n \leqslant 1,$$

откуда и следует справедливость нашего утверждения.

Для упрощения вывода предположим, что существует такое значение C, что $g(C) = \varepsilon$ (это предположение не является необходимым). Для множества X, отвечающего выбранному таким образом значению C, имеем

$$\mathrm{P}(X|\alpha_0) = g(C) = \varepsilon.$$

Обозначим через V критическое множество какого-нибудь другого критерия, для которого ошибка первого рода также равна ε, т. е. $\mathrm{P}(V|\alpha_0) = \varepsilon$. Покажем, что $\mathrm{P}(X|\alpha_1) \geqslant \mathrm{P}(V|\alpha_1)$, т. е. мощность критерия, основанного на критическом множестве X, оказывается большей, чем мощность любого другого критерия относительно гипотезы H_1, где гипотеза H_1 состоит в том, что $\alpha = \alpha_1$. Имеем

$$\mathrm{P}(X - VX \mid \alpha_0) = \varepsilon - \mathrm{P}(VX \mid \alpha_0) = \mathrm{P}(V - VX \mid \alpha_0).$$

Согласно определению множества X для любой точки $(X_1, X_2, \ldots, X_n) \in X$ выполняется неравенство

$$Cp(X_1, \alpha_0)\, p(X_2, \alpha_0) \ldots p(X_n, \alpha_0) >$$

$$> p(X_1, \alpha_1)\, p(X_2, \alpha_1) \ldots p(X_n, \alpha_1).$$

Поэтому имеем

$$\mathrm{P}(X - VX \mid \alpha_1) \geqslant C\mathrm{P}(X - VX \mid \alpha_0) =$$

$$= C\mathrm{P}(V - VX \mid \alpha_0) \geqslant \mathrm{P}(V - VX \mid \alpha_1).$$

Прибавляя к обеим частям полученного неравенства величину $\mathrm{P}(VX \mid \alpha_1)$, получим

$$\mathrm{P}(X \mid \alpha_1) \geqslant \mathrm{P}(V \mid \alpha_1),$$

что и требовалось доказать.

ОГЛАВЛЕНИЕ

Игорь Николаевич Коваленко
Анна Алексеевна Филиппова

ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Редактор Ж. И. Яковлева
Художник А. В. Алексеев
Художественный редактор В. И. Пономаренко
Технический редактор А. К. Нестерова
Корректор Г. И. Кострикова

Сдано в набор 10/V—73 г. Подп. к печати 2/XI—73 г. Формат $84\times108^1/_{32}$. Объем. 11,5 печ. л. Усл. п. л. 19,32. Уч.-изд. л. 16,32. Изд. № ФМ-475. Тираж 52 000 экз. Бум. тип. № 2. Цена 73 коп. Зак. 400.
БЗ—10—2 от 5/II—73 г.

Москва, К-51, Неглинная ул., д. 29/14,
Издательство «Высшая школа»

Ярославский полиграфкомбинат «Союзполиграфпрома» при Государственном комитете Совета Министров СССР по делам издательств, полиграфии и книжной торговли. Ярославль, ул. Свободы, 97.