

И. Н. ГИХМАН
А. В. СКОРОХОД
М. Н. ЯДРЕНКО

ТЕОРИЯ
ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ
СТАТИСТИКА

ББК22.17
517.8
Г51

УДК 519.2(075.8)

Теория вероятностей и математическая статистика. Гихман И. И., Скороход А. В., Ядренко М. И. Киев, Издательское объединение «Вища школа», Головное изд-во, 1979, 408 с.

В учебнике изложены основные сведения по теории вероятностей, теории случайных процессов, математической статистике. Приведены определения вероятностного пространства, случайной величины, математического ожидания, условной вероятности и условного математического ожидания, доказаны теоремы о законе больших чисел, центральная предельная теорема. Рассмотрены процессы восстановления, случайные блуждания, цепи Маркова со счетным множеством состояний, стационарные процессы (в том числе обобщенные). Определены основные задачи математической статистики, изложены методы проверки статистических гипотез и теория оценивания параметров вероятностных распределений. Рассматривается большое количество примеров и задач, иллюстрирующих основные понятия, а также поясняющих возможные практические применения доказанных теорем.

Рассчитан на студентов механико-математических факультетов университетов, факультетов прикладной математики и кибернетики технических вузов, физико-математических факультетов педагогических институтов.

Табл. 6. Ил. 18. Список лит.: 12 назв.

Редакция литературы по математике и физике
Зав. редакцией *Е. Л. Корженевич*

20203—085
Г М211(04)—79 117—79 1702060000

© Издательское объединение «Вища школа», 1979

СЛУЧАЙНЫЕ СОБЫТИЯ

§ 1. Стохастический эксперимент, случайные события

Стохастический эксперимент, пространство элементарных событий. Исходными понятиями теории вероятностей являются понятия стохастического эксперимента, случайного события и вероятности случайного события. *Стохастическими называются эксперименты, результаты которых нельзя предугадать заранее.*

В основе распространенного в настоящее время теоретико-множественного метода изложения теории вероятностей, который принят и в нашей книге, лежит предположение, что рассматриваемому эксперименту поставлено в соответствие некоторое множество Ω , точки которого изображают наиболее полную информацию о предполагаемых результатах в данном эксперименте. Множество Ω называют пространством элементарных событий, а его точки — элементарными событиями.

ПРИМЕРЫ

1. Производится эксперимент: один раз бросают монету. Пространство элементарных событий этого эксперимента имеет вид $\Omega = \{Г, Р\}$, где буква Г означает появление герба, буква Р — появление решки.

2. Монету бросают дважды. Пространством элементарных событий этого эксперимента является множество

$$\Omega = \{ГГ, ГР, РГ, РР\}.$$

Здесь ГР означает, например, что при первом бросании появился герб, а при втором — решка.

3. Бросают шестигранную игральную кость, на которой выбиты очки от 1 до 6. Нас интересует число выпавших очков. Пространством элементарных событий здесь может быть множество, состоящее из чисел 1, 2, 3, 4, 5, 6.

4. Игральную кость бросают m раз подряд. В качестве пространства элементарных событий можно рассматривать множество всех m -мерных векторов вида (i_1, i_2, \dots, i_m) , где каждая компонента i_k принимает значения 1, 2, 3, 4, 5, 6.

В рассмотренных выше примерах пространство элементарных событий было конечным множеством. Однако во многих задачах теории вероятностей приходится иметь дело с экспериментами, имеющими бесконечное число возможных исходов.

ПРИМЕРЫ

1. Предположим, что монету бросают до первого появления герба. Пространством элементарных событий такого эксперимента является множество

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots; \omega_\infty\},$$

где $\omega_n = \underbrace{P \dots P}_{(n-1) \text{ раз}} \Gamma$ означает, что герб впервые появится при n -м бросании

монеты, а ω_∞ соответствует той возможности, что герб никогда не появится (в этом случае наш эксперимент продолжается бесконечно долго).

2. Стрелок стреляет по круглой мишени, и нас интересует точка, в которую попала пуля. В качестве пространства элементарных событий можно принять множество, состоящее из рассматриваемого круга K и одной дополнительной точки θ , обозначающей непопадание стрелка в мишень.

3. Задача о встрече. Два лица A и B условились встретиться в интервале времени $[0, T]$. Если через x обозначим время прихода A , а через y время прихода B , то пространством элементарных событий будет множество

$$\Omega = \{(x, y) : 0 \leq x \leq T, 0 \leq y \leq T\}.$$

4. Установлен автоматически работающий прибор, вычерчивающий график скорости ветра в данной точке атмосферы в течение данного отрезка времени $[T_1, T_2]$. Пространством элементарных событий может быть множество всех неотрицательных функций, заданных на отрезке времени $[T_1, T_2]$ и имеющих, скажем, только конечное число разрывов первого рода.

5. Наблюдается частица, совершающая броуновское движение. Пространство элементарных событий — множество всех возможных траекторий частицы.

Случайные события — подмножества в пространстве элементарных событий. Говоря об эксперименте в теории вероятностей, мы не интересуемся его технической стороной, а только тем, какие события в этом эксперименте могут наблюдаться и что в результате проведенного эксперимента действительно наблюдалось. Таким образом, с каждым экспериментом связывают некоторое множество событий, о которых можно судить, осуществилось ли оно в данном эксперименте или нет. Такие события называют **наблюдаемыми** в данном эксперименте.

Если в эксперименте примера 4 нас интересует, превысила ли скорость ветра 10 м/сек, то это событие наблюдаемо. Если же по графику скорости ветра мы захотим судить, попал ли данный стрелок в мишень в примере 2, то сделать этого мы не сможем. Это событие ненаблюдаемо в эксперименте примера 4.

Пусть A — произвольное наблюдаемое в данном эксперименте событие. Поскольку каждое элементарное событие ω дает полную информацию о результате эксперимента, то зная, что результат эксперимента описывается точкой ω , всегда можно сказать, произошло ли A или нет. Таким образом, по отношению к событию A все пространство элементарных событий Ω можно разбивать на два дополнительных множества A' и A'' ($A' \in \Omega$, $A'' \in \Omega$, $A' \cup A'' = \Omega$) так, что, если результат эксперимента описывается точкой ω из множества A' , то событие A в этом эксперименте произошло, если же $\omega \in A''$, то событие A не произошло. Точки ω из множества A называют **элементарными событиями, благоприятствующими**

щими событием A . Говорят, что множество A' является отражением или интерпретацией события A во множестве Ω .

Следующий этап в построении теоретико-множественной модели теории вероятностей состоит в отождествлении события A и множества A' , $A \equiv A'$.

Итак, в дальнейшем событие A — это некоторое подмножество Ω , состоящее из всех тех точек ω — элементарных событий, — которые благоприятствуют событию A . Если результат эксперимента описывается точкой ω и ω входит в A , то в данном эксперименте событие A произошло, если же $\omega \notin A$, то событие A в этом эксперименте не произошло.

ПРИМЕРЫ

1. Пусть монету бросают дважды и A — событие, состоящее в том, что хотя бы один раз появится герб. Тогда

$$\Omega = \{ГГ, ГР, РГ, РР\},$$

$$A = \{ГГ, ГР, РГ\}.$$

2. Предположим, что один раз бросают игральный кубик и A — событие, состоящее в том, что число появившихся очков делится на 3. Тогда

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

$$A = \{3, 6\}.$$

3. Предположим, что монету бросают до первого появления герба (пример 1, с. 4). Пусть A — событие, состоящее в том, что будет сделано не больше трех бросаний. Тогда

$$A = \{Г, РГ, РРГ\}.$$

4. Рассмотрим задачу о встрече (пример 3, с. 4). Предположим, что каждое из лиц A и B ожидает другого время, не большее чем τ , $0 < \tau < T$. Пусть C — событие, состоящее в том, что встреча произойдет. Тогда

$$C = \{(x, y) : |x - y| \leq \tau, 0 \leq x \leq T, 0 \leq y \leq T\}$$

(рис. 1).

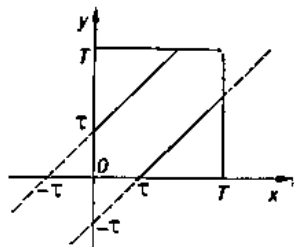


Рис. 1

Как известно, для множеств определено отношение порядка и над ними можно производить определенные алгебраические операции. Проанализируем содержательное значение этих понятий в том случае, когда подмножества некоторого множества Ω интерпретируются как события, наблюдаемые в некотором эксперименте.

1. Само множество Ω , рассматриваемое как событие, характеризуется тем, что в результате эксперимента оно необходимо происходит. Действительно, никакие другие результаты эксперимента, кроме тех, которые описываются точками $\omega \in \Omega$, по определению невозможны. Множество Ω называют достоверным событием.

2. Подмножеством любого множества Ω считается пустое множество \emptyset , не содержащее ни одной точки Ω . Если \emptyset отождествлять с событием, то это событие в эксперименте не происходит. Его называют невозможным событием.

3. Подмножества данного множества Ω частично упорядочены. Пишут $A \subset B$, если каждый элемент множества A содержится в B . Если A и B — события, то $A \subset B$ означает, что, если событие A происходит, то событие B также происходит. В этом случае говорят, что из события A следует событие B (событие A влечет за собой событие B). Очевидно, для любого A , $A \subset \Omega$. По определению принимают

$$\emptyset \subset A.$$

4. Суммой двух множеств $A \cup B$ называют множество, содержащее все элементы, входящие или во множество A , или во множество B и не содержащие никаких других элементов. Для событий это следует интерпретировать так: суммой (или объединением) двух событий A и B называют событие $A \cup B$, происходящее тогда и только тогда, когда происходит или событие A , или событие B .

Аналогичный смысл имеет сумма любого числа событий. Если I — произвольное множество значений некоторого индекса i , A_i ($i \in I$) — некоторое множество событий, то сумма (объединение) $\bigcup_{i \in I} A_i$ есть событие, происходящее тогда и только тогда, когда

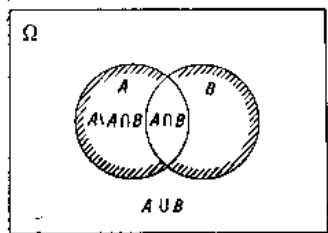


Рис. 2

происходит одно из событий A_i .

Заметим, что для любого A

$$A \cup \emptyset = A, \quad A \cup \Omega = \Omega.$$

Событие $A \cup B$ играет роль точной верхней грани событий A и B . Это надо понимать так: во-первых,

$$A \subset A \cup B, \quad B \subset A \cup B,$$

и если событие C таково, что $A \subset C$ и $B \subset C$, то

$$A \cup B \subset C.$$

5. Операция пересечения двух (или любого числа) множеств определяется следующим образом: пересечение $A \cap B$ (или $\bigcap_{i \in I} A_i$) есть множество, состоящее из тех и только тех элементов, которые принадлежат как множеству A , так и множеству B (всем множествам A_i , $i \in I$) (рис. 2).

На языке алгебры событий событие $A \cap B$ ($\bigcap_{i \in I} A_i$) происходит тогда и только тогда, когда происходит и событие A , и событие B (все события A_i , $i \in I$).

Пересечение событий называют также их совмещением.

Отметим очевидные соотношения:

$$A \cap \emptyset = \emptyset, \quad A \cap \Omega = A,$$

$$A \cap B \subset A, \quad A \cap B \subset B,$$

и если $C \subset A \cap B$, то $C \subset A$ и $C \subset B$.

Таким образом, пересечение $A \cap B$ можно рассматривать как точную нижнюю грань множеств A и B .

Отметим еще, что, как известно из теории множеств, в смысле введенного отношения порядка, операции \cup и \cap дистрибутивны:

$$A \cap (\cup_i B_i) = \cup_i (A \cap B_i),$$

$$A \cup (\cap_i B_i) = \cap_i (A \cup B_i).$$

Условимся два события A и B называть несовместимыми, или дизъюнктивными, если их совмещение есть событие невозможное, т. е.

$$A \cap B = \emptyset.$$

Несовместимые события изображаются множествами, не имеющими общих точек.

Последовательность событий A_1, A_2, \dots (конечная или бесконечная) называется дизъюнктивной, или последовательностью несовместимых событий, если каждая пара из них является несовместимой:

$$A_i \cap A_j = \emptyset \text{ для любых } i, j; i \neq j.$$

6. Каждому множеству A можно поставить в соответствие другое множество — его дополнение \bar{A} , состоящее из всех точек, которые не входят в A .

Таким образом,

$$A \cup \bar{A} = \Omega, \quad A \cap \bar{A} = \emptyset.$$

Если A — случайное событие, то \bar{A} — событие, происходящее тогда и только тогда, когда A не происходит. Событие \bar{A} называют событием, противоположным A .

7. Разность $A \setminus B$ двух множеств A и B есть множество, состоящее из элементов, входящих в A , но не входящих в B . Если A и B — события, то $A \setminus B$ есть событие, происходящее тогда и только тогда, когда происходит A , но не происходит B .

Очевидно,

$$\bar{A} = \Omega \setminus A, \quad A \setminus B = A \cap \bar{B}.$$

8. Пусть $\{A_n\}$ — бесконечная последовательность множеств, $A_n \subset \Omega$. Обозначим через A^* множество всех тех и только тех элементов, которые принадлежат бесконечному числу множеств A_n . Заметим, что

$$A^* = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

В самом деле, если $\omega \in A^*$, т. е. ω принадлежит бесконечному числу множеств A_m , то $\omega \in \bigcup_{m=n}^{\infty} A_m$ для каждого n , и, следовательно, $\omega \in \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$, то есть $A^* \subset \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$. Наоборот, если $\omega \in \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$, то $\omega \in A^*$.

Пусть A_* — множество тех и только тех ω , которые принадлежат всем A_n , за исключением конечного числа. Тогда

$$A_* = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m.$$

Множество A^* можно интерпретировать, как событие, состоящее в том, что произойдет бесконечно много событий из последовательности событий $\{A_n\}$. Множество A_* есть событие, состоящее в том, что произойдут все события A_n , начиная с некоторого номера, или, что то же самое, не произойдет только конечное число событий из последовательности A_n . Очевидно, $A_* \subset A^*$. Событие A^* называется **верхним пределом** последовательности событий $\{A_n\}$

$$A^* = \overline{\lim}_{n \rightarrow \infty} A_n,$$

а событие A_* называется **нижним пределом** последовательности событий $\{A_n\}$

$$A_* = \underline{\lim}_{n \rightarrow \infty} A_n.$$

Если

$$\overline{\lim}_{n \rightarrow \infty} A_n = \underline{\lim}_{n \rightarrow \infty} A_n,$$

то говорят, что последовательность событий $\{A_n\}$ имеет **предел**

$$\lim_{n \rightarrow \infty} A_n = \overline{\lim}_{n \rightarrow \infty} A_n = \underline{\lim}_{n \rightarrow \infty} A_n.$$

Пусть, например, A_n — «монотонно убывающая» последовательность событий, т. е. такая последовательность, что

$$A_1 \supset A_2 \supset \dots \supset A_n \supset A_{n+1} \supset \dots$$

Тогда

$$A^* = \overline{\lim}_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n,$$

$$A_* = \underline{\lim}_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n.$$

Следовательно, предел последовательности событий A_n существует и

$$\lim A_n = \bigcap_{n=1}^{\infty} A_n.$$

Если A_n — «монотонно возрастающая» последовательность событий

$$A_1 \subset A_2 \subset \dots \subset A_n \subset A_{n+1} \subset \dots,$$

то

$$A^* = \overline{\lim} A_n = \bigcup_{n=1}^{\infty} A_n,$$

$$A_* = \underline{\lim} A_n = \bigcap_{n=1}^{\infty} A_n,$$

и последовательность $\{A_n\}$ имеет предел, равный

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

Подытоживая все сказанное выше, приведем таблицу, показывающую, как некоторые понятия теории множеств интерпретируются в теории вероятностей.

Обозначения	Язык теории множеств	Язык теории вероятностей
Ω	Универсальное множество	Пространство элементарных событий (элементарных исходов эксперимента)
ω	Элемент Ω	Элементарное событие
A	Некоторое множество элементов ω	Событие A (если $\omega \in A$, то говорят, что наступило событие A)
Ω	Множество всех ω	Достоверное событие
\emptyset	Пустое множество	Невозможное событие
$A \subset B$	A подмножество B	Из наступления события A necessarily следует наступление B
$A \cup B$	Объединение множеств A и B ; множество точек, входящих или в A , или в B	Событие, состоящее в том, что произошло A или B
$A \cap B$	Пересечение множеств A и B ; множество точек, входящих и в A , и в B	Событие, состоящее в том, что произойдет и A , и B

Обозначения	Язык теории множеств	Язык теории вероятностей
$A \cap B = \emptyset$	A и B — непересекающиеся множества	A и B — несовместные события
$A \setminus B$	Разность множеств A и B	Событие, состоящее в том, что произойдет A , но не произойдет B
$\overline{\lim} A_n$	Множество всех тех ω , которые принадлежат бесконечному числу множеств из последовательности $\{A_n\}$	Событие, состоящее в том, что произойдет бесконечное число событий из последовательности событий $\{A_n\}$
$\lim A_n$	Множество всех тех ω , которые принадлежат всем A_n , за исключением конечного числа (множество тех ω , которые не принадлежат только конечному числу множеств A_n)	Событие, состоящее в том, что произойдут все события A_n , за исключением конечного числа (событие, состоящее в том, что не произойдет только конечное число из событий последовательности $\{A_n\}$)
$\lim A_n$	Если $\lim A_n = \overline{\lim} A_n$, то последовательность множеств $\{A_n\}$ имеет предел $\lim A_n = \overline{\lim} A_n = \lim A_n$	Если $\lim A_n = \overline{\lim} A_n$, то последовательность событий $\{A_n\}$ имеет предел $\lim A_n = \overline{\lim} A_n = \lim A_n$

§ 2. Основные понятия комбинаторики

Комбинаторика — это раздел математики, изучающий расположения объектов в соответствии со специальными правилами и методы подсчета числа всех возможных способов, которыми эти расположения могут быть сделаны.

Впервые теоретическое исследование проблем комбинаторики было проведено в XVII в. Паскалем, Ферма, Лейбницем и в XVIII в. Я. Бернулли, Эйлером. Тогда же сложилась и принята в комбинаторике терминология (сочетания, размещения, перестановки и т. п.). К началу XX в. комбинаторика считалась в основном завершенным разделом математики, лежащим вне основного русла развития математики и ее приложений. В XX в. комбинаторику стали рассматривать как раздел теории множеств, изучающий различные проблемы, возникающие при изучении конечных множеств. Такая точка зрения привела к более естественной и последовательной классификации основных понятий и задач комбинаторики. В последнее время роль комбинаторики возросла в связи с развитием теории вычислительных машин, а также теории информации, изучающей методы оптимального кодирования, декодирования и передачи информации. Ныне комбинаторика является одним из наиболее интенсивно развивающихся разделов математики.

Методы комбинаторики играют важную роль при вычислении вероятностей различных событий, связанных с экспериментами, имеющими конечное число исходов.

Изложим основные понятия комбинаторики.

Основной принцип комбинаторики (правило умножения). Рассмотрим следующую задачу:

Из города А в город В ведет m различных путей, а из города В в город С ведет n путей. Каким числом различных путей можно совершить путешествие из города А в город С через город В?

Решение. Выбрав один из m возможных путей из А в В, дальше можно продолжить путешествие n способами. Поэтому общее число путей равно $m \times n$.

Соображения, приведенные при рассмотрении предыдущего примера, доказывают справедливость следующего простого, но очень важного правила, которое называют основным принципом комбинаторики.

Если некоторый выбор А можно осуществить m различными способами, а для каждого из этих способов некоторый другой выбор В можно осуществить n способами, то выбор А и В (в указанном порядке) можно осуществить $m \times n$ способами.

Рассмотрим еще один пример, иллюстрирующий основной принцип комбинаторики.

В розыгрыше первенства страны по футболу принимают участие 16 команд. Сколькими способами могут быть распределены золотая и серебряная медали?

Решение. Золотую медаль может получить одна из 16 команд. После того как определен владелец золотой медали, серебряную медаль может иметь одна из 15 команд. Следовательно, общее число способов, которыми могут быть распределены золотая и серебряная медали, равно $16 \times 15 = 240$.

Сформулируем теперь основное правило комбинаторики (правило умножения) в общем виде.

Пусть требуется выполнить одно за другим k действий. Если первое действие можно выполнить n_1 способами, второе — n_2 способами, третье — n_3 способами и так до k -го действия, которое можно выполнить n_k способами, то все k действий вместе могут быть выполнены

$$n_1 \times n_2 \times n_3 \times \dots \times n_k$$

способами.

Доказательство этого утверждения может быть легко проведено методом полной математической индукции. Рекомендуем читателю выполнить это самостоятельно.

ЗАДАЧА

Сколько четырехзначных чисел можно составить из цифр 0, 1, 2, 3, 4, 5, 6, 7, 8, 9?

а) ни одна из цифр не повторяется более одного раза;

б) цифры могут повторяться;

в) числа должны быть нечетными (цифры могут повторяться)?

Решение. а) Первой цифрой числа может быть одна из пяти цифр 2, 3, 4, 5. Если первая цифра выбрана, то вторая может быть выбрана способами, третья — 4 способами, четвертая — 3 способами. Согласно правилу умножения, общее число способов равно $5 \times 5 \times 4 \times 3 = 300$.

б) Первой цифрой может быть одна из цифр 1, 2, 3, 4, 5 (5 случаев), для каждой из следующих цифр имеем 6 случаев (0, 1, 2, 3, 4, 5). Следовательно, число искомых чисел равно $5 \times 6 \times 6 \times 6 = 5 \times 6^3 = 1080$.

в) Первой цифрой может быть одна из цифр 1, 2, 3, 4, 5, а последней — одна из цифр 1, 3, 5 (числа должны быть нечетными). Следовательно, общее количество чисел равно $5 \times 6 \times 6 \times 3 = 540$.

Сочетания из n по k . Пусть Ω — множество из n элементов. Любое k -элементное подмножество множества из n элементов называется **сочетанием из n элементов по k** . Порядок элементов в подмножестве не существен. Число k -элементных подмножеств множества из n элементов обозначают C_n^k .

Например, пусть $\Omega = \{a, b, c\}$. Тогда $\{a\}, \{b\}, \{c\}$ — все возможные сочетания из 3 по 1 (следовательно, $C_3^1 = 3$), $\{a, b\}, \{a, c\}, \{b, c\}$ — все возможные сочетания из 3 по 2 (таким образом, $C_3^2 = 3$).

Пусть $n! = 1 \cdot 2 \cdot \dots \cdot n$; удобно считать по определению, что $1! = 1$.

Теорема 1. Число всех k -элементных подмножеств множества из n элементов равно

$$C_n^k = \frac{n(n-1) \dots (n-k+1)}{1 \cdot 2 \dots k} = \frac{n!}{k!(n-k)!}. \quad (1)$$

Доказательство. Чтобы построить k -элементное подмножество множества Ω , нужно к $(k-1)$ -элементному подмножеству присоединить один из $n-k+1$ элементов, который не входит в это подмножество. Поскольку $(k-1)$ -элементных подмножеств имеется C_n^{k-1} и каждое из них можно сделать k -элементным $n-k+1$ способами, то таким образом мы получим $(n-k+1)C_n^{k-1}$ подмножеств. Но не все они будут разными, так как каждое k -элементное множество можно построить k способами: присоединением любого из k его элементов. Поэтому найденное нами число в k раз больше, чем число C_n^k k -элементных подмножеств. Следовательно,

$$kC_n^k = (n-k+1)C_n^{k-1}.$$

отсюда

$$\begin{aligned} C_n^k &= \frac{n-k+1}{k} C_n^{k-1} = \frac{(n-k+1)(n-k+2)}{k(k-1)} C_n^{k-2} = \dots \\ &= \frac{(n-k+1) \dots (n-1)}{k(k-1) \dots 2} C_n^1. \end{aligned}$$

Поскольку число одноэлементных подмножеств множества A равно количеству элементов n , то, подставив вместо C_n^1 число n , получим равенство (1).

1. Сколькими способами читатель может выбрать 3 книги из 7?

Решение. Искомое число равно числу трехэлементных подмножеств во множестве из 7 элементов:

$$C_7^3 = \frac{7!}{3!4!} = 35.$$

2. В скольких точках пересекаются диагонали выпуклого n -угольника, если никакие три из них не пересекаются в одной точке?

Решение. Каждая точка пересечения двух диагоналей определяется четырьмя вершинами n -угольника (концами соответствующих диагоналей); каждому четверю вершинам n -угольника соответствует одна точка пересечения (точка пересечения диагоналей четырехугольника с вершинами в данных четырех точках). Поэтому число всех точек пересечения равно числу способов, которыми среди n вершин можно выбрать 4 вершины:

$$C_n^4 = \frac{n!}{4!(n-4)!} = \frac{n(n-1)(n-2)(n-3)}{24}.$$

3. Сколькими способами можно составить комиссию в составе трех человек, выбирая их из четырех супружеских пар, если:

- 1) в комиссию могут входить любые три из восьми человек;
- 2) в комиссию не могут входить члены одной семьи?

Решение. 1) Если в комиссию войдут любые 3 из 8 человек, то число всех возможных комиссий равно $C_8^3 = 56$.

2) Если в комиссию не входят члены одной семьи, то в ней будут представлены 3 из 4 семей. Эти семьи можно выбрать $C_4^3 = 4$ способами. После этого в каждой из них можно двумя способами выбрать представителя — мужа или жену. По правилу умножения число всех возможных комиссий равно:

$$C_4^3 \cdot 2 \cdot 2 \cdot 2 = 32.$$

В следующей задаче дана интересная и важная геометрическая интерпретация для чисел C_n^k .

4. Рассмотрим «шахматный город» (прямоугольную сетку квадратов размерами $m \times n$), состоящий из $m \times n$ прямоугольных кварталов, разделенных $n-1$ «горизонтальными» и $m-1$ «вертикальными» улицами (рис. 3). Каково число различных кратчайших путей на этой сетке, ведущих из левого нижнего угла (точки $(0; 0)$) в правый верхний угол (точку $(m; n)$)?

Решение. Каждый кратчайший путь из точки $(0; 0)$ в точку $(m; n)$ состоит из $m+n$ отрезков, причем среди них есть m горизонтальных и n вертикальных отрезков. Разные пути отличаются лишь порядком чередования горизонтальных и вертикальных отрезков. Поэтому общее число путей равно числу способов, которыми из $m+n$ отрезков можно выбрать n вертикальных отрезков, т. е. C_{m+n}^n .

Можно было бы рассматривать число способов выбора не n вертикальных, а m горизонтальных отрезков, и мы получили бы тогда ответ C_{m+n}^m . Итак, мы установили геометрически равенство $C_{m+n}^m = C_{m+n}^n$, в справедливости которого нетрудно убедиться, непосредственно используя соотношение (1).

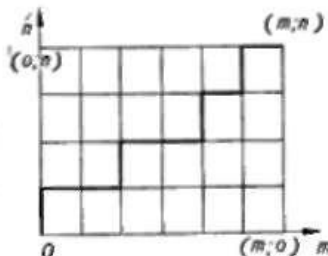


Рис. 3

Таким образом, число кратчайших путей из точки $(0; 0)$ в точку $(m; n)$ равно $C_{m+n}^m = C_{m+n}^n$.

Числа C_n^k называются биномиальными коэффициентами. Следующая теорема выражает важное свойство биномиальных коэффициентов.

Теорема 2. *Имеет место равенство*

$$C_n^k = C_{n-1}^k + C_{n-1}^{k-1}. \quad (2)$$

Легко убедиться в справедливости равенства (2), используя соотношение (1).

Приведем геометрическое доказательство равенства (2).

Число кратчайших путей из точки $O(0; 0)$ в точку $A(n; n-k)$ равно $C_{n+(n-k)}^k = C_{2n-k}^k$ (рис. 4). Все такие пути можно разделить на две группы: пути, проходящие через точку $A_1(k-1, n-k)$ (число их равно $C_{(k-1)+(n-k)}^{k-1} = C_{n-1}^{k-1}$), и пути, проходящие через точку $A_2(k; n-k-1)$

(число их равно $C_{k+(n-k-1)}^k = C_{n-1}^k$). Следовательно, имеет место равенство (2).

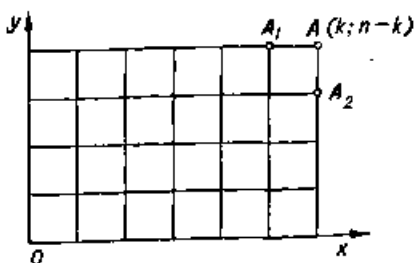


Рис. 4

Упорядоченные множества, перестановки. Множество называется упорядоченным, если каждому элементу этого множества поставлено в соответствие некоторое число (номер элемента) от 1 до n , где n — число элементов множества, так, что различным элементам соответствуют различные

числа. Всякое конечное множество можно сделать упорядоченным, если, например, записать все элементы множества в некоторый список (a, b, c, \dots) , а затем поставить в соответствие каждому элементу номер места, на котором он стоит в списке. Очевидно, каждое множество, содержащее более одного элемента, можно упорядочить не единственным способом. Упорядоченные множества считаются различными, если они отличаются либо своими элементами, либо их порядком. Различные упорядоченные множества, которые отличаются лишь порядком элементов (т. е. могут быть получены из того же самого множества), называются перестановками этого множества.

ПРИМЕР

Перестановки множества $\Omega = (a, b, c)$ из трех элементов имеют вид

$$(a, b, c), (a, c, b), (b, a, c), \\ (b, c, a), (c, a, b), (c, b, a).$$

Теорема 3. Пусть P_n — число перестановок множества, содержащего n элементов. Имеет место равенство

$$P_n = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n.$$

Доказательство. Будем последовательно выбирать элементы множества Ω и размещать их в определенном порядке на n местах. На первое место можно поставить любой из n элементов. После того как заполнено первое место, на второе место можно поставить любой из оставшихся $n-1$ элементов и т. д. По правилу умножения все n мест можно заполнить

$$n(n-1)(n-2)\dots 2\cdot 1 = n!$$

способами. Следовательно, множество Ω из n элементов можно упорядочить $n!$ способами.

задача

Сколькими способами можно упорядочить множество $\{1, 2, \dots, 2n\}$ так, чтобы каждое четное число имело четный номер?

Решение. Четные числа можно расставить на местах с четными номерами (таких мест n) $n!$ способами; каждому способу размещения четных чисел на местах с четными номерами соответствует $n!$ способов размещения нечетных чисел на местах с нечетными номерами. Поэтому общее число перестановок указанного типа по правилу умножения равно $n! \cdot n! = (n!)^2$.

Размещения из n по k . Рассмотрим теперь упорядоченные подмножества данного множества Ω . Само множество Ω считаем неупорядоченным, поэтому каждое его подмножество может быть упорядочено каким-либо возможным способом. Число всех k -элементных подмножеств множества Ω равно C_n^k . Каждое такое подмножество можно упорядочить $k!$ способами. Таким образом, получим все упорядоченные k -элементные подмножества множества Ω . Следовательно, их число будет $k!C_n^k$.

Теорема 4. Число упорядоченных k -элементных подмножеств множества, состоящего из n элементов, равно

$$A_n^k = k!C_n^k = \frac{n!}{(n-k)!} = n(n-1)\dots(n-k+1). \quad (3)$$

Упорядоченные k -элементные подмножества множества из n элементов называются размещениями из n элементов по k . Различные размещения из n по k отличаются количеством элементов либо их порядком.

задачи

1. Студенту необходимо сдать 4 экзамена на протяжении 8 дней. Сколькими способами это можно сделать?

Решение. Искомое число способов равно числу 4-элементных упорядоченных подмножеств (дни сдачи экзаменов) множества из 8 элементов, т. е. $A_8^4 = 8 \cdot 7 \cdot 6 \cdot 5 = 1680$ способов. Если известно, что последний экзамен будет сдаваться на восьмой день, то число способов равно:

$$4 \cdot A_7^3 = 7 \cdot 6 \cdot 5 \cdot 4 = 840.$$

2. Сколько существует телефонных номеров, состоящих из шести различных цифр?

Решение. Число таких номеров равно $A_{10}^6 = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 = 151200$ (число размещений на 6 местах шести из 10 цифр 0, 1, 2, 3, 4, 5, 6, 7, 8, 9).

Разбиения на группы; перестановки с повторениями. Пусть Ω — множество из n элементов. Поставим следующий вопрос: каким числом способов множество Ω можно представить в виде объединения m попарно непересекающихся множеств B_1, \dots, B_m , содержащих соответственно k_1, \dots, k_m элементов ($k_1 + k_2 + \dots + k_m = n$)?

Все разбиения множества Ω на m групп B_1, \dots, B_m можно получить так: возьмем произвольное k_1 -элементное подмножество B_1 множества Ω (это можно сделать $C_n^{k_1}$ способами), среди $n - k_1$ оставшихся элементов возьмем k_2 -элементное подмножество B_2 (это можно сделать $C_{n-k_1}^{k_2}$ способами) и т. д. Общее число способов выбора различных множеств B_1, \dots, B_m по правилу умножения равно:

$$\begin{aligned} & C_n^{k_1} \cdot C_{n-k_1}^{k_2} \cdot C_{n-k_1-k_2}^{k_3} \cdots C_{n-k_1-\dots-k_{m-1}}^{k_m} = \dots = k_{m-1} = \\ &= \frac{n!}{k_1! (n-k_1)!} \cdot \frac{(n-k_1)!}{k_2! (n-k_1-k_2)!} \cdot \frac{(n-k_1-k_2)!}{k_3! (n-k_1-k_2-k_3)!} \cdots \\ & \quad \cdots \frac{(n-k_1-\dots-k_{m-1})!}{k_m! (n-k_1-\dots-k_m)!} = \frac{n!}{k_1! k_2! \dots k_m!} \end{aligned}$$

(напомним, что $0! = 1$).

Итак, мы доказали следующую теорему:

Теорема 5. Пусть k_1, \dots, k_m — целые неотрицательные числа, причем $k_1 + \dots + k_m = n$. Число способов, которыми можно представить множество A из n элементов в виде суммы m множеств B_1, \dots, B_m , число элементов которых составляет соответственно k_1, k_2, \dots, k_m , равно:

$$C_n(k_1, \dots, k_m) = \frac{n!}{k_1! \dots k_m!}. \quad (4)$$

Числа $C_n(k_1, \dots, k_m)$ называются полиномиальными коэффициентами. Они имеют еще одну очень важную комбинаторную интерпретацию.

Пусть имеется n букв: k_1 — букв a_1 , k_2 — букв a_2 , ..., k_m — букв a_m , $k_1 + \dots + k_m = n$. Определим, сколько различных слов можно составить из этих букв. Перенумеруем места, на которых стоят буквы, числами $1, 2, \dots, n$. Каждое слово определяется множествами B_1 (номера мест, где стоит буква a_1), B_2 (номера мест, где стоит буква a_2) ..., B_m (номера мест, где стоит буква a_m). Следовательно, число различных слов равно числу способов, которыми можно представить множество $\Omega = \{1, 2, \dots, n\}$ в виде объединения множеств B_1, \dots, B_m , т. е. это число определяется соотношением (4).

Слова длины n , которые можно получить из k_1 букв a_1 , k_2 букв a_2 , ..., k_m букв a_m , называются еще перестановками

с повторениями. Проведенное выше рассуждение показывает, что справедлива следующая теорема:

Теорема 6. Число различных перестановок, которые можно составить из n элементов, среди которых имеется k_1 элементов первого типа, k_2 элементов второго типа, ..., k_m элементов m -го типа, равно:

$$C_n(k_1, \dots, k_m) = \frac{n!}{k_1! \dots k_m!}.$$

ЗАДАЧИ

1. Сколькими способами можно расселить 8 студентов по трем комнатам: одноместной, трехместной, четырехместной?

Решение. По теореме 5 искомое число равно:

$$C_8(1, 3, 4) = \frac{8!}{1!3!4!} = 280.$$

2. Сколько различных слов можно получить, переставляя буквы слова «математика»?

Решение. Искомое число равно:

$$C_{10}(3, 2, 2, 1, 1, 1) = \frac{10!}{3! \cdot 2! \cdot 2!} = 151200.$$

Полиномиальная формула. Рассмотрим вопрос о том, как раскрывать скобки при вычислении выражения вида

$$(a_1 + a_2 + \dots + a_k)^n.$$

Теорема 7. (Полиномиальная теорема). Выражение

$$(a_1 + a_2 + \dots + a_k)^n$$

равно сумме всех возможных слагаемых вида

$$\frac{n!}{r_1! r_2! \dots r_k!} a_1^{r_1} a_2^{r_2} \dots a_k^{r_k},$$

где

$$r_1 + r_2 + \dots + r_k = n,$$

т. е.

$$\begin{aligned} & (a_1 + a_2 + \dots + a_k)^n = \\ &= \sum_{\substack{r_1 > 0, \dots, r_k > 0 \\ r_1 + r_2 + \dots + r_k = n}} \frac{n!}{r_1! \dots r_k!} a_1^{r_1} \dots a_k^{r_k}. \end{aligned} \quad (5)$$

Доказательство. Перемножим $a_1 + \dots + a_k$ последовательно n раз. Тогда получим k^n слагаемых вида $d_1 d_2 \dots d_n$, где каждый множитель d_i равен или a_1 , или a_2 , ..., или a_k . Обозначим через $B(r_1, \dots, r_n)$ совокупность всех тех слагаемых, где a_1 встречается множителем r_1 раз, a_2 — r_2 раз, ..., a_k — r_k раз. Число таких слагаемых равно $C_n(r_1, \dots, r_k)$ — числу способов представления множества из n элементов $\{1, 2, \dots, n\}$ в виде суммы k

множеств B_1, B_2, \dots, B_k так, чтобы множество B_s имело r_s элементов ($r_s \geq 0, r_1 + \dots + r_k = n$; множество B_s — это множество тех i , для которых $d_i = a_s$). По теореме 5

$$C_n(r_1, \dots, r_k) = \frac{n!}{r_1! \dots r_k!}.$$

Следовательно,

$$\begin{aligned} & (a_1 + a_2 + \dots + a_k)^n = \\ &= \sum_{\substack{r_1 \geq 0, \dots, r_k \geq 0 \\ r_1 + \dots + r_k = n}} \frac{n!}{r_1! \dots r_k!} a_1^{r_1} \dots a_k^{r_k}. \end{aligned}$$

При $k = 2$ равенство (3) имеет вид

$$(a_1 + a_2)^n = \sum_{r=0}^n \frac{n!}{r! (n-r)!} a_1^{n-r} a_2^r = \sum_{r=0}^n C_n^r a_1^{n-r} a_2^r. \quad (6)$$

Равенство (6) называется биномиальной формулой.

Сочетания с повторениями. Сочетаниями из m элементов по n элементов с повторениями называются группы, содержащие n элементов, причем каждый элемент принадлежит одному из m типов.

Например, из трех элементов a, b, c можно составить такие сочетания по два с повторениями:

$$aa, ac, bc, ab, bb, cc.$$

Теорема 8. Число различных сочетаний из m элементов по n с повторениями равно:

$$f_m^n = C_{m+n-1}^{n-1} = C_{n+m-1}^n. \quad (7)$$

Доказательство. Каждое сочетание полностью определяется, если указать, сколько элементов каждого из m типов в него входит. Поставим в соответствие каждому сочетанию последовательность нулей и единиц, составленную по такому правилу: напишем подряд столько единиц, сколько элементов первого типа входит в сочетание, далее поставим нуль и после него напишем столько единиц, сколько элементов второго типа содержит это сочетание и т. д. Например, написанным выше сочетаниям из трех букв по две будут соответствовать такие последовательности: 1100, 1001, 0101, 1010, 0110, 0011.

Таким образом, каждому сочетанию из m по n соответствует последовательность из n единиц и $m-1$ нулей и, наоборот, по каждой такой последовательности однозначно восстанавливается такое сочетание. Поэтому число сочетаний из m по n с повторениями равно числу последовательностей из n единиц и $m-1$ нулей, т. е.

$$C_{n+m-1}^{m-1} = C_{n+m-1}^n.$$

ПРИМЕРЫ

1. Кости домино можно рассматривать как сочетания с повторениями по два из семи цифр 0, 1, 2, 3, 4, 5, 6. Число таких сочетаний равно:

$$f_7^2 = C_{7+2-1}^2 = C_8^2 = \frac{8 \cdot 7}{2} = 28.$$

2. Сколькими способами можно выбрать три из двенадцати букв А, А, А, Т, Т, Т, Г, Г, Г, Ц, Ц, Ц?

Решение. В рассматриваемом случае $m=4$ (имеем три сорта предметов А, Т, Г, Ц), а $n=3$. Поэтому искомое число равно:

$$f_4^3 = C_{3+4-1}^3 = C_6^3 = 20.$$

Задача имеет отношение к теории белкового кода, предложенной американским физиком Гамовым. Буквы А, Т, Г, Ц обозначают нуклеотиды Аденин, Тимин, Гуанин, Цитозин. Число троек нуклеотидов оказывается равным 20 — числу стандартных аминокислот, на которые разлагаются молекулы белка.

3. Сколько целых неотрицательных решений имеет уравнение

$$x_1 + x_2 + \dots + x_m = n? \quad (8)$$

Решение. Существует тесная связь между решениями указанного уравнения и сочетаниями из m элементов по n . Если имеем целые неотрицательные числа x_1, \dots, x_m такие, что $x_1 + x_2 + \dots + x_m = n$, то можем составить сочетание из m элементов, взяв x_1 элементов первого типа, x_2 — второго типа, x_m — m -го типа. Наоборот, имея сочетание из m элементов по n , получим некоторое решение уравнения (8) (x_1 — число элементов первого типа, ..., x_m — число элементов m -го типа) в целых неотрицательных числах. Следовательно, между множеством всех сочетаний из m элементов по n с повторениями и множеством всех целых неотрицательных решений уравнения (8) установлено взаимно однозначное соответствие. Поэтому число решений уравнения (8) равно $f_m^n = C_{n+m-1}^n$.

4. Сколько существует различных частных производных порядка n у бесконечно дифференцируемой функции m переменных?

Решение. Частные производные порядка n бесконечно дифференцируемой функции $f(x_1, \dots, x_m)$ m переменных не зависят от порядка дифференцирования, а зависят лишь от того, сколько раз мы дифференцируем по каждому переменному. Если по первому переменному дифференцируем k_1 раз, по второму — k_2 раз, ..., по m -му — k_m раз ($k_1 + \dots + k_m = n$), то получим частную производную $\frac{\partial^n f}{\partial x_1^{k_1} \dots \partial x_m^{k_m}}$. Поэтому число различных частных про-

изводных равно числу различных неотрицательных решений уравнения $k_1 + \dots + k_m = n$, т. е.

$$f_m^n = C_{n+m-1}^{m-1}.$$

Например, функция трех переменных имеет 15 различных производных четвертого порядка.

Теорема 9. Пусть $n > m$. Тогда число таких сочетаний из m по n с повторениями, в которых каждый предмет представлен хотя бы один раз, есть C_{n-1}^{m-1} .

Доказательство. Поставим в соответствие каждому сочетанию из m по n набор из $m-1$ нулей и n единиц по правилу, описанному в доказательстве теоремы 8. Тогда тем сочетаниям, которые каждый из m элементов содержат хотя бы один раз, будут соответствовать наборы единиц и нулей, в которых нет нулей.

стоящих рядом (в таком наборе каждый нуль стоит между двумя единицами). Каждый такой набор кончается единицей. Все такие наборы можно указать, отметив те из $n-1$ единиц, справа от которых стоит нуль. Это можно сделать C_{n-1}^{m-1} способами.

Размещения частиц по ячейкам (модель Максвелла—Больцмана и модель Бозе—Эйнштейна). В заключение рассмотрим два комбинаторных результата о распределении различных и неразличимых частиц по ячейкам, играющих важную роль в статистической физике.

Теорема 10. *Предположим, что n различных частиц распределяются по m ячейкам (областям пространства). Тогда число всех возможных размещений частиц равно m^n . Число таких размещений, при которых в первой ячейке находится k_1 частиц, во второй — k_2 частиц, ..., в m -й — k_m частиц, равно*

$$\frac{n!}{k_1! \dots k_m!}.$$

Доказательство. Пронумеруем ячейки номерами $1, 2, \dots, m$ и частицы — номерами $1, 2, \dots, n$. Расположение частиц задано, если каждой частице с номером k указан номер ячейки i_k , т. е. расположение частиц описывается набором (i_1, \dots, i_n) , где каждое i_k может принимать любое из m значений $1, 2, \dots, m$. Число таких наборов, согласно правилу умножения, равно $m \times m \times \dots \times m = m^n$. Следовательно, число всех возможных расположений частиц равно m^n .

Пусть B_1 — номера тех частиц, которые попали в первую ячейку, B_2 — номера тех частиц, которые попали во вторую ячейку, ..., B_m — номера частиц, которые попали в m -ю ячейку. Тогда

$$\{1, 2, \dots, n\} = \bigcup_{i=1}^m B_i.$$

По теореме 5 число тех расположений частиц, при которых в первой ячейке находится k_1 частиц, во второй — k_2 частиц, ..., в m -й — k_m частиц, равно $\frac{n!}{k_1! \dots k_m!}$. Теорема доказана. Рассмотренная в теореме 10 модель размещения частиц называется моделью Максвелла—Больцмана.

Теорема 11. *Пусть n неразличимых частиц распределяются по m ячейкам (областям пространства). Тогда число всех возможных расположений частиц равно $f_m^n = C_{n+m-1}^{m-1} = C_{n+m-1}^n$. Число таких размещений, при которых в каждой ячейке есть хотя бы одна частица, равно C_{n-1}^{m-1} ($n > m$).*

Доказательство. Так как все частицы неразличимы, расположение частиц задается указанием набора (x_1, x_2, \dots, x_m) , где x_i — число частиц в i -й ячейке. Два расположения различны, если соответствующие наборы не совпадают. Так как

$$x_1 + x_2 + \dots + x_m = n,$$

то утверждения теоремы следуют из результатов примера 3, с. 19, и теоремы 9. Рассмотренная модель размещения частиц называется моделью Бозе—Эйнштейна.

§ 3. Классическое определение вероятности

Частота случайного события. В дальнейшем будем предполагать, что рассматриваются только такие стохастические эксперименты, которые можно повторить любое число раз, и что при многократном повторении эксперимента события, происшедшие при предыдущих экспериментах, никак не влияют на события, которые произойдут при данном эксперименте.

Рассмотрим некоторый стохастический эксперимент и событие A , наблюдаемое в этом эксперименте. Повторим эксперимент n раз. Пусть $k_n(A)$ — число экспериментов, в которых произошло событие A .

Отношение

$$v_n(A) = \frac{k_n(A)}{n}$$

называется частотой события A в проведенной серии экспериментов.

Частота $v_n(A)$ обладает следующими свойствами:

1) $0 \leq v_n(A) \leq 1$;

2) $v_n(\Omega) = 1$;

3) если A и B — два наблюдаемых несовместных события, то

$$v_n(A \cup B) = v_n(A) + v_n(B).$$

Эти свойства следуют из того, что $k_n(A) \leq n$, $k_n(\Omega) = n$ (достоверное событие происходит при каждом осуществлении эксперимента), и, если A и B — несовместные события, то

$$k_n(A \cup B) = k_n(A) + k_n(B).$$

Частота может быть вычислена лишь после того, как проведена серия экспериментов, и, вообще говоря, частота изменяется, если мы проведем другую серию из n экспериментов, или если изменим n . Однако, как показывает опыт, при достаточно больших n для большинства таких серий экспериментов частота сохраняет почти постоянную величину, причем большие отклонения наблюдаются тем реже, чем больше n .

Так, например, пусть много раз бросают симметричную монету. Тогда, как показывают многочисленные эксперименты, частота появления герба в длинных сериях экспериментов мало отличается от $\frac{1}{2}$. В приведенной ниже таблице даны результаты трех экспериментов, состоящих в бросании симметричной монеты.

Экспериментатор	Число бросаний	Число выпадений герба	Частота
Ж. Бюффон	4040	2048	0,5080
К. Пирсон	12000	6019	0,5016
К. Пирсон	24000	12012	0,5005

Если при больших n частота $\nu_n(A)$ события A мало отличается от некоторого фиксированного значения p , то говорят что событие A стохастически устойчиво, а число p является вероятностью события A .

Приведенное эмпирическое определение статистической устойчивости и вероятности события характеризует естественно-научное содержание понятия вероятности, но не является его формальным определением. Чтобы прийти к формальному определению вероятности, следует аксиоматизировать свойства вероятности, вытекающие из того естественно-научного содержания этого понятия, которое мы желаем ему приписать. После того как аксиомы будут сформулированы, мы можем произвольную систему величин, удовлетворяющих сформулированной системе аксиом, называть вероятностью. Разумеется, при дальнейшем развитии формальной теории несущественно, какой реальный смысл вкладывается в понятие вероятности, но если мы захотим применить результаты теории к явлениям реального мира, то нужно дать интерпретацию отношения понятия вероятности к объективным явлениям природы. В дальнейшем мы будем придерживаться ранее высказанной точки зрения: вероятность события A есть число, близкое к частоте наступления события A в длинной серии тождественных экспериментов.

Таким образом, основными задачами теории вероятностей являются следующие:

1. Формализация понятия вероятности и изучение общих свойств математических объектов, в которых введено понятие вероятности.

2. Создание и изучение различных вероятностных моделей, отражающих те конкретные ситуации, возникающие в человеческой практике или при изучении природы, которые описываются с помощью теоретико-вероятностных понятий.

3. Разработка методов проверки соответствия предлагаемых вероятностных моделей тем реальным ситуациям или явлениям, которые они должны описывать. Исследования, посвященные последней проблеме, часто выделяют в самостоятельную дисциплину и называют ее математической статистикой. В нашем пособии мы такого разделения не придерживаемся.

Прежде чем сформулировать аксиомы теории вероятностей, мы сформулируем и изучим простейшие вероятностные модели.

Вероятностная модель эксперимента с конечным или счетным числом исходов. Рассмотрим стохастический эксперимент с конечным или счетным числом возможных исходов $\omega_1, \dots, \omega_i, \dots$,

$$\Omega = \{\omega_1, \dots, \omega_i, \dots\}.$$

Предположим, что каждому элементарному событию ω_i приписан некоторый «вес» p_i , называемый «вероятностью» элементарного события ω_i , и что веса p_i обладают следующими свойствами:

$$1) p_i \geq 0,$$

$$2) \sum_{i=1}^{\infty} p_i = 1.$$

Пусть A — произвольное случайное событие, наблюдаемое в данном эксперименте, т. е. A — произвольное подмножество Ω .

Определение. Вероятностью $P(A)$ события A называют сумму вероятностей элементарных событий, составляющих событие A

$$P(A) = \sum_{\omega_i \in A} p_i. \quad (1)$$

Введенная таким способом вероятность обладает следующими свойствами:

$$a) 0 \leq P(A) \leq 1,$$

$$б) P(\Omega) = 1,$$

в) если A и B — несовместные события ($A \subset \Omega, B \subset \Omega$), то

$$P(A \cup B) = P(A) + P(B).$$

Действительно,

$$0 \leq P(A) = \sum_{\omega_i \in A} p_i \leq \sum_{i=1}^{\infty} p_i = 1,$$

$$P(\Omega) = \sum_{\omega_i \in \Omega} p_i = \sum_{i=1}^{\infty} p_i = 1,$$

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

если $A \cap B = \emptyset$.

Определение. Говорят, что для данного эксперимента построена вероятностная модель (Ω, P) , если:

1) указано пространство элементарных событий

$$\Omega = \{\omega_1, \dots, \omega_i, \dots\};$$

2) каждому ω_i из Ω приписана «вероятность» p_i , причем

$$p_i \geq 0, \quad \sum_{i=1}^{\infty} p_i = 1.$$

Вероятность произвольного события A , наблюдаемого в данном эксперименте, определяется по формуле (1). Теория вероятностей не учит тому, как «правильно» определять вероятности p_i элемен-

тарных событий ω_i . При определении этих вероятностей принимаются во внимание интуитивные представления о p_i , как частоте элементарного события ω_i в серии из большого числа повторений эксперимента. Теория вероятностей не занимается также вопросом о том, правильно ли построена вероятностная модель (Ω, P) , она отвечает лишь на вопрос, как вычислить вероятности различных сложных событий, связанных с построенной моделью (Ω, P) .

ПРИМЕРЫ

1. Пусть бросают симметричную шестигранную игральную кость. Тогда в качестве Ω естественно рассматривать множество $\Omega = \{1, 2, 3, 4, 5, 6\}$. Если кость симметрична, то каждому элементарному событию $\omega_i = i$ припишем вероятность $\frac{1}{6}$. Тем самым будет построена вероятностная модель эксперимента, состоящего в бросании шестигранной симметричной игральной кости. Если A случайное событие, состоящее в том, что появившееся число очков кратно 3, т. е. $A = \{3, 6\}$, то $P(A) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Предположим, что игральная кость не является симметричной и массы в ней распределены так, что масса каждой грани пропорциональна ее номеру. Тогда припишем элементарному событию $\omega_i = i$ вероятность $p_i = \frac{i}{\sum_{k=1}^6 k} = \frac{i}{21}$.

Вероятность рассмотренного выше события $A = \{3, 6\}$ будет равна $P(A) = \frac{3}{21} + \frac{6}{21} = \frac{3}{7}$.

2. Построим вероятностную модель эксперимента, состоящего в двухкратном бросании симметричной монеты. В этом случае

$$\Omega = \{ГГ, ГР, РГ, РР\}.$$

Так как монета симметрична, то все элементарные события считаем одинаково возможными, и каждому из них припишем вероятность $\frac{1}{4}$. Пусть, например, A — событие, состоящее в том, что хотя бы один раз появится герб, т. е.

$$A = \{ГГ, ГР, РГ\}.$$

Тогда

$$P(A) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}.$$

3. Пусть симметричная монета бросается до тех пор, пока впервые не появится герб. Тогда

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots; \omega_\infty\},$$

где $\omega_n = \underbrace{РР \dots Р}_{n-1 \text{ раз}} Г$, ω_∞ — элементарное событие, означающее, что герб никогда не появится. Припишем ω_n вероятность $\frac{1}{2^n}$, а ω_∞ — вероятность 0.

Тогда

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

Таким образом, мы построили вероятностную модель эксперимента, состоящего в бросании монеты до первого появления герба. Подсчитаем теперь вероят-

ность события A , состоящего в том, что будет произведено не более трех бросаний монеты ($A = \{Г, РГ, РРГ\}$). Имеем

$$P(A) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}.$$

Предположим, что рассматривается эксперимент с конечным числом n одинаково возможных исходов $\{\omega_1, \omega_2, \dots, \omega_n\}$. Тогда естественно при построении вероятностной модели этого эксперимента приписать каждому ω_i вероятность $\frac{1}{n}$. Пусть A ($A \subset \Omega = \{\omega_1, \dots, \omega_n\}$) — случайное событие, состоящее из m элементарных событий (иногда говорят в этом случае, что событию A благоприятствует m элементарных исходов). Тогда

$$P(A) = \sum_{\omega_i \in A} p_i = \frac{m}{n}.$$

Определение. (Классическое определение вероятности). Рассмотрим стохастический эксперимент, имеющий n одинаково возможных исходов. Предположим, что событию A благоприятствует m из этих исходов. Тогда

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

ПРИМЕР

Пусть симметричную игральную кость бросают дважды. В этом случае пространство элементарных событий

$$\Omega = \{i, j\} : i = \overline{1, 6}; j = \overline{1, 6}\}$$

состоит из 36 элементарных событий. Будем считать их одинаково возможными. Пусть A — событие, состоящее в том, что сумма выпавших очков равна 5. Тогда

$$A = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$$

и

$$P(A) = \frac{4}{36} = \frac{1}{9}.$$

§ 4. Алгебры и σ -алгебры множеств; теорема о продолжении меры

При построении вероятностных моделей для экспериментов с бесконечным числом исходов возникают определенные трудности, но прежде чем рассматривать пути преодоления этих трудностей и изучать системы аксиом теории вероятностей, изложим некоторые сведения из теории множеств и теории меры, которые в дальнейшем будут играть существенную роль.

Будем рассматривать некоторое фиксированное множество Ω и подмножества этого множества.

Определение. Класс \mathfrak{M} подмножеств из Ω называется алгеброй, если:

- 1) $\Omega \in \mathfrak{A}$;
- 2) из того, что $A \in \mathfrak{A}$, следует, что $\bar{A} \in \mathfrak{A}$;
- 3) из того, что $A \in \mathfrak{A}$, $B \in \mathfrak{A}$, следует, что $A \cup B \in \mathfrak{A}$.

Замечание. Если \mathfrak{A} — алгебра, то из того, что $A \in \mathfrak{A}$, $B \in \mathfrak{A}$, следует $A \cap B$ принадлежит \mathfrak{A} .

Действительно, $A \cap B = \overline{\bar{A} \cup \bar{B}}$.

ПРИМЕРЫ

1. Пусть A — некоторое подмножество Ω . Тогда класс множеств

$$\mathfrak{A} = \{\Omega, \emptyset, A, \bar{A}\}$$

образует алгебру.

2. Пусть $\Omega = [0, 1]$ и \mathfrak{A} — система подмножеств из Ω , каждое из которых представляет собой конечную сумму непересекающихся интервалов вида $[a, b]$. Тогда \mathfrak{A} — алгебра.

3. Пусть $\Omega = [0, 1] \times [0, 1]$ и \mathfrak{A} — множество всех прямоугольников

$$[a_1, b_1] \times [a_2, b_2]$$

и конечных сумм таких непересекающихся прямоугольников. Тогда \mathfrak{A} — алгебра.

Определение. Класс \mathcal{F} множеств из Ω называется σ -алгеброй, если:

- 1) $\Omega \in \mathcal{F}$;
- 2) из того, что $A \in \mathcal{F}$, следует $\bar{A} \in \mathcal{F}$;
- 3) из того, что $A_i \in \mathcal{F}$ ($i = 1, 2, \dots$), следует, что

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

ПРИМЕР. Множество всех подмножеств Ω образует σ -алгебру.

Определение. Пусть K — некоторый класс подмножеств из Ω . σ -алгебра $\sigma(K)$ называется наименьшей σ -алгеброй, содержащей класс K , если

- 1) $K \subset \sigma(K)$;
- 2) для любой σ -алгебры \mathcal{F} такой, что $K \subset \mathcal{F}$, имеем $\sigma(K) \subset \mathcal{F}$.

Теорема 1. Для любого класса множеств K существует наименьшая σ -алгебра $\sigma(K)$, содержащая класс K .

Заметим сначала, что существует хотя бы одна σ -алгебра, содержащая класс K . Такой σ -алгеброй является, например, σ -алгебра всех подмножеств из Ω . Обозначим через $\sigma(K)$ пересечение всех σ -алгебр, содержащих класс K . Тогда $\sigma(K)$ есть σ -алгебра. Действительно, легко проверить, что пересечение любого числа σ -алгебр является σ -алгеброй. Если \mathcal{F} — любая σ -алгебра, содержащая K , то $\sigma(K) \subset \mathcal{F}$, т. е. $\sigma(K)$ есть наименьшая σ -алгебра, содержащая класс K .

Определение. Класс множеств M называется монотонным классом, если:

1) для любой последовательности множеств $\{A_n\}$ такой, что $A_1 \subset A_2 \subset \dots \subset A_n \subset A_{n+1} \subset \dots$, из того что $A_n \in M$, следует, что

$$\lim A_n = \bigcup_{n=1}^{\infty} A_n \in M;$$

2) для любой последовательности множеств $\{A_n\}$ такой, что $A_1 \supset A_2 \supset \dots \supset A_n \supset A_{n+1} \supset \dots$, из того что $A_n \in M$, следует, что

$$\lim A_n = \bigcap_{n=1}^{\infty} A_n \in M.$$

ПРИМЕР. Всякая σ -алгебра является монотонным классом.

Теорема 2. Для того чтобы алгебра \mathfrak{A} была σ -алгеброй, необходимо и достаточно, чтобы она была монотонным классом.

Необходимость. Если алгебра \mathfrak{A} является σ -алгеброй, то она является и монотонным классом.

Достаточность. Пусть алгебра \mathfrak{A} является монотонным классом. Докажем, что если $A_n \in \mathfrak{A}$, то и $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$. Заметим, что если $B_m = \bigcup_{n=1}^m A_n$, то $B_m \subset B_{m+1}$, $B_m \in \mathfrak{A}$, поскольку \mathfrak{A} — алгебра. Так как по предположению \mathfrak{A} — монотонный класс, то

$$\lim B_m = \bigcup_{m=1}^{\infty} B_m = \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}.$$

Определение. Монотонный класс $M(K)$ называется наименьшим монотонным классом, содержащим класс K , если:

- 1) $K \subset M(K)$;
- 2) для любого монотонного класса \mathfrak{M} , содержащего класс K , $M(K) \subset \mathfrak{M}$.

Теорема 3. Для любого класса множеств K существует наименьший монотонный класс, содержащий K .

Доказательство. Существует хотя бы один монотонный класс, содержащий класс K (это класс всех подмножеств множества Ω). Обозначим через $M(K)$ — пересечение всех монотонных классов, содержащих K . Тогда $M(K)$ — монотонный класс (пересечение любого числа монотонных классов — монотонный класс!), причем для любого монотонного класса \mathfrak{M} , содержащего K , $M(K) \subset \mathfrak{M}$. Итак, $M(K)$ является наименьшим монотонным классом, содержащим класс K .

Теорема 4. Пусть \mathfrak{A} — алгебра. Тогда

$$\sigma(\mathfrak{A}) = M(\mathfrak{A}),$$

т. е. наименьшая σ -алгебра, содержащая алгебру \mathfrak{A} , и наименьший монотонный класс, содержащий \mathfrak{A} , совпадают.

Доказательство. Так как \mathfrak{A} — алгебра, то по теореме 2 $\sigma(\mathfrak{A})$ является монотонным классом, и поэтому

$$M(\mathfrak{A}) \subseteq \sigma(\mathfrak{A}).$$

Если доказать, что $M(\mathfrak{A})$ есть σ -алгебра, то тогда

$$\sigma(\mathfrak{A}) \subseteq M(\mathfrak{A}),$$

что и будет доказывать требуемое равенство. Поскольку $M(\mathfrak{A})$ — монотонный класс, то, принимая во внимание теорему 2, достаточно доказать, что $M(\mathfrak{A})$ — алгебра. Обозначим $M(\mathfrak{A}) = M$.

Пусть $A \in M$; докажем, что $\bar{A} \in M$. Для этого рассмотрим класс множеств

$$\tilde{M} = \{B : B \in M, \bar{B} \in M\}.$$

Тогда

$$a \subset \tilde{M} \subset M.$$

Класс \tilde{M} — монотонный класс. Действительно, если $B_n \subset \tilde{M}$, $B_n \subset B_{n+1}$, то $B_n \in M$, $\bar{B}_n \in M$, $\bar{B}_n \supset \bar{B}_{n+1}$. Так как M — монотонный класс, то $\lim B_n = \bigcup B_n \in M$, $\lim \bar{B}_n = \bigcap \bar{B}_n = \overline{\bigcup B_n} = \overline{\lim B_n} \in M$. Поскольку M — наименьший монотонный класс, содержащий \mathfrak{A} , то $M \subset \tilde{M}$, т. е. $M = \tilde{M}$. Таким образом, если $B \in M$, то $\bar{B} \in M$.

Докажем теперь, что класс M замкнут относительно взятия конечных пересечений: если $A \in M$, $B \in M$, то $A \cap B \in M$.

Пусть $A \in M$. Рассмотрим класс множеств

$$M_A = \{B : B \in M, A \cap B \in M\}.$$

Так как для любой монотонной последовательности множеств $\{B_n\}$

$$\lim A \cap B_n = A \cap \lim B_n,$$

то M_A — монотонный класс. Далее, если $A \in \mathfrak{A}$, то для всякого $B \in \mathfrak{A}$, $A \cap B \in \mathfrak{A}$ (класс \mathfrak{A} — алгебра) и, следовательно,

$$\mathfrak{A} \subset M_A \subset M.$$

Поскольку M_A — монотонный класс, содержащий \mathfrak{A} , а M — наименьший монотонный класс, содержащий алгебру \mathfrak{A} , то $M_A = M$ для любого $A \in \mathfrak{A}$.

Так как A и B в определение класса M_A входят симметрично, то при $A \in M$, $B \in M$ имеем:

$$A \in M_B \Rightarrow B \in M_A,$$

$$B \in M_A \Rightarrow A \in M_B.$$

Пусть $A \in \mathfrak{A}$, $B \in M = M_A$. Тогда $A \in M_B$ для любого $B \in M$. В силу произвольности A получаем

$$\mathfrak{A} \subset M_B \subset M.$$

Таким образом, для всякого $B \in M$

$$M_B = \{C : C \in M, B \cap C \in M\} = M.$$

Итак, мы доказали, что из $B \in M$, $C \in M$ следует, что $B \cap C \in M$, т. е. M — алгебра. Теорема доказана.

определение. Пусть $\Omega = R^1 = (-\infty, +\infty)$, K — класс всех интервалов вида $[a, b)$. Наименьшая σ -алгебра \mathfrak{B} , содержащая класс K , называется σ -алгеброй борелевских множеств в R^1 .

Заметим, что одноточечное множество $\{a\}$ является борелевским, так как

$$\{a\} = \bigcap_{n=1}^{\infty} \left[a, a + \frac{1}{n} \right).$$

Интервал $(a, b) = [a, b) \setminus \{a\}$ есть борелевское множество. Каждое открытое множество является борелевским (любое открытое множество в R^1 есть сумма конечного или счетного числа интервалов). Каждое замкнутое множество является также борелевским (дополнение к замкнутому множеству — множество открытое).

Определение. Пусть $\Omega = R^m$, K — класс параллелепипедов вида

$$\prod_{i=1}^m [a_i, b_i) = \{x: x = (x_1, \dots, x_m): a_i \leq x_i < b_i, i = \overline{1, m}\}.$$

Наименьшая σ -алгебра \mathfrak{B}^m , содержащая класс K , называется σ -алгеброй борелевских множеств в R^m .

Определение. Пусть \mathfrak{A} — некоторая алгебра подмножеств из Ω . Функция множества $P(\cdot)$, определенная на \mathfrak{A} , называется конечно-аддитивной вероятностной мерой на \mathfrak{A} , если:

- 1) $P(A) \geq 0$ для каждого A из \mathfrak{A} ;
- 2) $P(\Omega) = 1$;
- 3) $P(A \cup B) = P(A) + P(B)$, если $A \in \mathfrak{A}$, $B \in \mathfrak{A}$, $A \cap B = \emptyset$.

Из свойства 3) методом математической индукции легко получить, что для любого конечного числа m множеств A_1, \dots, A_m из \mathfrak{A}

$$P\left(\bigcup_{i=1}^m A_i\right) = \sum_{i=1}^m P(A_i),$$

если $A_i \cap A_j = \emptyset$ ($i \neq j$; $i = \overline{1, m}$, $j = \overline{1, m}$).

Далее, так как $A \cup \bar{A} = \Omega$, то, в силу 3) и 2),

$$P(A) + P(\bar{A}) = 1,$$

и, следовательно,

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

Определение. Функция множества $P(\cdot)$, определенная на алгебре \mathfrak{A} , называется счетно-аддитивной вероятностной мерой на \mathfrak{A} , если:

- 1) $P(A) \geq 0$ для каждого множества $A \in \mathfrak{A}$;
- 2) $P(\Omega) = 1$;

$$3) P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i), \text{ если } A_i \in \mathfrak{A} \ (i = 1, 2, \dots),$$

$$A_i \cap A_j = \emptyset \ (i \neq j) \quad \bigcup_{i=1}^{\infty} A_i \in \mathfrak{A}.$$

Замечание. Из конечной аддитивности вероятностной меры, определенной на алгебре \mathfrak{A} , не следует счетная аддитивность. Приведем простой пример.

Пусть Ω — множество всех рациональных чисел r из отрезка $[0, 1]$. \mathfrak{A} — совокупность всех множеств вида $\{r: a \leq r < b\}$, $\{r: a \leq r \leq b\}$, $\{r: a < r \leq b\}$, $\{r: a < r < b\}$ (a и b — рациональные числа) и всех конечных объединений таких множеств. Тогда \mathfrak{A} образует алгебру множеств. Для каждого множества $A \subset \mathfrak{A}$ вида

$$\{r: a \leq r < b\}, \{r: a \leq r \leq b\}, \{r: a < r \leq b\}, \{r: a < r < b\}$$

положим $P(A) = b - a$, а если $B \in \mathfrak{A}$ и $B = \bigcup_{i=1}^m A_i$, где A_i попарно не пересекаются, то возьмем

$$P(B) = \sum_{i=1}^m P(A_i).$$

Тогда $P(\cdot)$ будет конечно-аддитивной вероятностной мерой на алгебре \mathfrak{A} . Одноточечные множества $\{r\}$ принадлежат \mathfrak{A} и $P(\{r\}) = 0$. Остается заметить, что свойством счетной аддитивности мера $P(\cdot)$ не обладает

$$P(\Omega) = 1 \neq \sum_{r \in \Omega} P(\{r\}) = 0.$$

Докажем следующую теорему, показывающую, какие дополнительные предположения надо наложить на конечно-аддитивную вероятностную меру, чтобы она была счетно-аддитивной.

Теорема 5. Пусть $P(\cdot)$ — конечно-аддитивная вероятностная мера на \mathfrak{A} . Тогда следующие четыре условия являются эквивалентными:

1) $P(\cdot)$ — счетно-аддитивна, т. е., если $A_i \in \mathfrak{A}$, $i = 1, 2, \dots$, $A_i \cap A_j = \emptyset$, $i \neq j$, $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{A}$, то

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i);$$

2) если A_i , $i = 1, 2, \dots$, — неубывающая последовательность множеств ($A_i \subset A_{i+1}$) и $\lim A_i = \bigcup_{i=1}^{\infty} A_i \in \mathfrak{A}$, то

$$P(\bigcup_{i=1}^{\infty} A_i) = \lim_{i \rightarrow \infty} P(A_i);$$

3) если $A_i, i = 1, 2, \dots$, — невозрастающая последовательность множеств ($A_i \supset A_{i+1}$) и $\lim A_i = \bigcap_{i=1}^{\infty} A_i \subset \mathfrak{A}$, то

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i \rightarrow \infty} P(A_i);$$

4) $P(\cdot)$ непрерывна в \emptyset , т. е. если $A_i \supset A_{i+1}$ и $\bigcap_{i=1}^{\infty} A_i = \emptyset$, то

$$\lim_{i \rightarrow \infty} P(A_i) = 0.$$

Доказательство. Докажем следующую цепочку заключений $1) \Rightarrow 2) \Rightarrow 3) \Rightarrow 4) \Rightarrow 1)$, из которой и будет следовать, что все условия 1), 2), 3), 4) эквивалентны.

$1) \Rightarrow 2)$. Пусть $A_i, i = 1, 2, \dots$, — неубывающая последовательность множеств из \mathfrak{A} . Положим $A_0 = \emptyset, B_i = A_{i+1} \setminus A_i \in \mathfrak{A}$. Тогда

$$A_n = \bigcup_{i=1}^{n-1} B_i,$$

и согласно свойству конечной аддитивности

$$P(A_n) = \sum_{i=1}^{n-1} P(B_i).$$

Заметим далее, что

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i,$$

причем множества B_i попарно не пересекаются. Поэтому из 1) следует, что

$$\begin{aligned} P(\lim A_n) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \\ &= \sum_{i=1}^{\infty} P(B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} P(B_i) = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

Таким образом, из 1) следует 2).

$2) \Rightarrow 3)$. Пусть A_i — невозрастающая последовательность множеств ($A_i \supset A_{i+1}$). Тогда \bar{A}_i есть неубывающая последовательность множеств и, согласно 2), имеем

$$\begin{aligned} P(\lim A_n) &= P\left(\bigcap_{i=1}^{\infty} A_i\right) = 1 - P\left(\overline{\bigcap_{i=1}^{\infty} A_i}\right) = \\ &= 1 - P\left(\bigcup_{i=1}^{\infty} \bar{A}_i\right) = 1 - \lim_{n \rightarrow \infty} P(\bar{A}_n) = \lim_{n \rightarrow \infty} [1 - P(\bar{A}_n)] = \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

$3) \Rightarrow 4)$. Так как $P(\emptyset) = P(\bar{\Omega}) = 1 - P(\Omega) = 0$, то для любой невозрастающей последовательности, такой, что $\bigcap_{i=1}^{\infty} A_i = \emptyset$, по свойству 3) имеем

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = 0 = \lim_{n \rightarrow \infty} P(A_n).$$

4) \Rightarrow 1). Пусть $A_i, i = 1, 2, \dots$ — последовательность непересекающихся множеств из \mathfrak{M} и

$$C_n = \bigcup_{i=n+1}^{\infty} A_i.$$

Тогда $C_{n+1} \subset C_n$ и $\bigcap_{n=1}^{\infty} C_n = \emptyset$.

В силу свойства конечной аддитивности

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^n A_i\right) + P(C_n) = \sum_{i=1}^n P(A_i) + P(C_n).$$

Поскольку $\lim_{n \rightarrow \infty} P(C_n) = 0$, то

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A_i) = \sum_{i=1}^{\infty} P(A_i).$$

Теорема доказана.

Заметим, что свойства 2) и 3) можно сформулировать следующим образом: если A_n — монотонная последовательность множеств, то

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

Свойство, выраженное этим равенством, называют свойством непрерывности счетно-аддитивной вероятностной меры.

В дальнейшем при изложении некоторых вопросов важную роль играет следующая теорема о продолжении меры.

Теорема 6. (Теорема Каратеодори о продолжении меры). Пусть P — счетно-аддитивная вероятностная мера на алгебре \mathfrak{M} . Тогда существует единственная счетно-аддитивная вероятностная мера $Q(\cdot)$, определенная на наименьшей σ -алгебре $\mathcal{F} = \sigma(\mathfrak{M})$, содержащей \mathfrak{M} , и являющаяся продолжением меры P , т. е. мера $Q(\cdot)$ такая, что $Q(A) = P(A)$ для каждого $A \in \mathfrak{M}$.

Мы не будем доказывать все это утверждение. Докажем лишь единственность продолжения. Пусть $Q_1(\cdot)$ и $Q_2(\cdot)$ — две меры, определенные на $\mathcal{F} = \sigma(\mathfrak{M})$ и такие, что

$$Q_1(A) = Q_2(A) = P(A),$$

если $A \in \mathfrak{M}$. Обозначим через M класс множеств

$$M = \{A : A \in \mathcal{F}, Q_1(A) = Q_2(A)\}.$$

По предположению $\mathfrak{M} \subset M$. Далее заметим, что M — монотонный класс: если A_n — монотонная последовательность множеств из M , то по свойству непрерывности счетно-аддитивной меры

$$Q_1(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} Q_1(A_n) = \lim_{n \rightarrow \infty} Q_2(A_n) = Q_2(\lim_{n \rightarrow \infty} A_n)$$

и, следовательно, $\lim_{n \rightarrow \infty} A_n \in M$. По теореме 4 $M = \sigma(\mathfrak{M}) = \mathcal{F}$. Таким образом, $Q_1(A) = Q_2(A)$ для каждого множества A из $\mathcal{F} = \sigma(\mathfrak{M})$.

§ 5. Построение вероятностных моделей для экспериментов с несчетным числом исходов; геометрические вероятности

Выше мы рассматривали примеры стохастических экспериментов, множество элементарных событий у которых является несчетным множеством. Естественно возникает вопрос: как построить вероятностную модель для каждого из таких экспериментов? Приведем несколько примеров.

Рассмотрим эксперимент, состоящий в бросании наугад точки на отрезок $[0, 1]$. Предполагается, что все мыслимые положения точки «одинаково возможны». Пространство элементарных событий для такого эксперимента — отрезок $\Omega = [0, 1]$. Однако в данном случае уже нельзя построить вероятностную модель эксперимента, приписав вероятности лишь отдельным элементарным событиям, как это мы делали в случае экспериментов со счетным числом элементарных событий; если по аналогии с дискретным случаем задавать для каждого ω вероятности $P(\omega)$, то поскольку все ω «одинаково возможны», получим $P(\omega) = 0$ для каждого ω . Оказывается, выход из создавшихся затруднений можно найти в том, чтобы приписывать вероятности лишь некоторым множествам элементарных событий, причем задание вероятностей должно происходить согласованным образом. Например, естественно считать, что вероятности попадания в интервалы $\left[0, \frac{1}{2}\right)$ и $\left[\frac{1}{2}, 1\right)$ должны быть одинаковыми (и, следовательно, каждая из них равна $\frac{1}{2}$); вероятность попадания в отрезок $\left[0, \frac{1}{2}\right)$ должна быть больше, чем вероятность попадания в отрезок $\left[0, \frac{1}{3}\right)$ и т. д.

Кроме того, желательно, чтобы класс тех множеств, которым приписаны вероятности, был замкнут относительно операций объединения, пересечения, перехода к дополнению, т. е. если множествам A и B приписаны вероятности, желательно, чтобы множествам $A \cup B$, $A \cap B$, \bar{A} , \bar{B} также были приписаны вероятности.

Будем считать, что вероятность попадания точки в отрезок $[a, b] \subset [0, 1]$ равна $b - a$. Объединению конечного числа непересекающихся отрезков $\bigcup_{i=1}^m [a_i, b_i] \subset [0, 1]$ припишем вероятность

$\sum_{i=1}^m (b_i - a_i)$. Рассмотрим множество \mathcal{A} всех интервалов вида $[a, b]$

и конечных сумм таких непересекающихся интервалов. Пусть $\mathfrak{B} = \sigma(\mathcal{A})$ — наименьшая σ -алгебра, содержащая класс \mathcal{A} (как известно, \mathfrak{B} это σ -алгебра борелевских множеств отрезка $[0, 1]$). По теореме о продолжении меры следует, что существует единственная мера $P(\cdot)$ на \mathfrak{B} такая, что $P([a, b]) = b - a$. Будем

рассматривать в качестве случайных событий только борелевские множества из $[0, 1]$ и вероятностью $A \subset \mathfrak{B}$ назовем $P(A)$. Итак, мы построили вероятностную модель $(\Omega, \mathfrak{B}, P)$ эксперимента, состоящего в бросании наугад точки на $[0, 1]$, указав пространство элементарных событий $\Omega[0, 1]$ и σ -алгебру \mathfrak{B} случайных событий, для которых определена вероятность $P(\cdot)$.

Заметим, что на множестве всех подмножеств из $[0, 1]$ нельзя определить меру $P(\cdot)$ так, чтобы $P([a, b]) = b - a$ для всех $[a, b] \subset [0, 1]$ (см. Натансон Г. И. Теория функций вещественного переменного. М., Наука, 1974, с. 89). Таким образом, существенно то, что в качестве случайных событий рассматривается лишь часть подмножеств из $\Omega = [0, 1]$, а именно σ -алгебра всех борелевских множеств.

Задача о встрече. Предположим, что два лица A и B условились встретиться в интервале времени $[0, T]$. Пришедший первым ждет другого в течение τ единиц времени, а затем уходит.

Пусть x — время прихода A , y — время прихода B . Тогда пространством элементарных исходов эксперимента является множество

$$\Omega = \{(x, y): 0 \leq x \leq T, 0 \leq y \leq T\} = [0, T] \times [0, T]$$

(рис. 1), представляющее собой квадрат со стороной T .

Будем считать, опираясь на интуитивное представление, что все элементарные события «одинаково возможны». Вычислим теперь вероятность того, что встреча A и B состоится.

В этой задаче также нельзя построить вероятностную модель, приписав вероятность каждому элементарному событию. Рассмотрим сначала в качестве случайных событий многоугольники, содержащиеся в $\Omega = [0, T] \times [0, T]$.

Припишем каждому многоугольнику K в качестве вероятности $P(K) = \frac{\text{пл } K}{T^2}$.

Пусть $\mathfrak{B} = \mathfrak{B}(K)$ наименьшая σ -алгебра, содержащая класс K многоугольников. Каждому множеству $C \in \mathfrak{B}$ припишем вероятность, применяя теорему о продолжении меры (получим $P(C) = \frac{\text{пл } C}{T^2}$; отметим, что все множества из \mathfrak{B} имеют площадь). Таким образом, мы построим вероятностную модель эксперимента $(\Omega, \mathfrak{B}, P)$.

Интересующее нас событие C (встреча произойдет) изображено на рис. 1:

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

Его вероятность

$$P(C) = \frac{T^2 - (T - \tau)^2}{T^2} = 1 - \left(1 - \frac{\tau}{T}\right)^2.$$

В частности, если $T = 1$ ч, $\tau = \frac{1}{3}$ ч, то $P(C) = 1 - \left(1 - \frac{1}{3}\right)^2 = \frac{5}{9}$.

Сделаем теперь следующее общее замечание. Предположим, что рассматривается эксперимент, пространство элементарных исходов

которого представляет собой область Ω в n -мерном евклидовом пространстве R^n (пример такого эксперимента: точку бросают наугад в область Ω). Предположим, что область Ω имеет меру Лебега $m(\Omega)$ (в случае $n=2$ — площадь, в случае $n=3$ — объем). Рассмотрим σ -алгебру \mathfrak{A} всех подмножеств из Ω , имеющих меру Лебега (как известно из теории меры, класс таких множеств образует σ -алгебру). Если все точки Ω «одинаково возможны», то положим для каждого $A \in \mathfrak{A}$

$$P(A) = \frac{m(A)}{m(\Omega)}. \quad (1)$$

Тогда $(\Omega, \mathfrak{A}, P)$ есть вероятностная модель рассматриваемого стохастического эксперимента. В этом случае говорят, что имеют дело с геометрическими вероятностями. Заметим, что вероятность $P(\cdot)$, определяемая соотношением (1), обладает следующими свойствами:

- 1) $0 \leq P(A) \leq 1$;
- 2) $P(\Omega) = 1$;
- 3) если $A_i \in \mathfrak{A}$, $A_i \cap A_k = \emptyset$ ($i \neq k$), то

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

Задача Бюффона. Плоскость разграфлена параллельными прямыми, отстоящими друг от друга на расстоянии $2a$. На плоскость наугад бросают иглу длиной $2l$ ($l < a$). Найти вероятность того, что игла пересечет какую-нибудь прямую.

Пусть x — расстояние от центра K иглы до ближайшей прямой (рис. 5), а φ — угол, составленный иглой с этой прямой. Величины x и φ полностью определяют положение иглы. Следовательно, пространство элементарных событий — это прямоугольник

$$\Omega = \{(x, \varphi) : 0 \leq x \leq l, 0 \leq \varphi \leq \pi\}.$$

Будем считать, что каждая точка $(x, \varphi) \in \Omega$ одинаково возможна. Именно такой смысл вкладывается в содержание фразы: «иглу бросают наугад». Игла пересекает прямую тогда и только тогда, когда выполнено условие $x \leq l \sin \varphi$.

Пусть C — следующее множество

$$C = \{(x, \varphi) : x \leq l \sin \varphi, (x, \varphi) \in \Omega\}$$

(на рис. 6 множество C заштриховано). Рассмотрим σ -алгебру \mathfrak{A} всех подмножеств из Ω , имеющих площадь (меру Лебега), и опре-

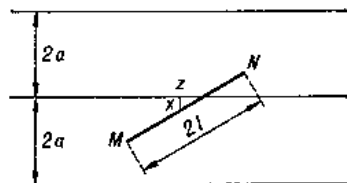


Рис. 5

делим вероятность для каждого множества $A \in \mathfrak{A}$ соотношением (1). Тогда будет построена вероятностная модель $(\Omega, \mathfrak{A}, P)$ эксперимента, состоящего в бросании иглы на разграфленную плоскость. В частности, вероятность интересующего нас события

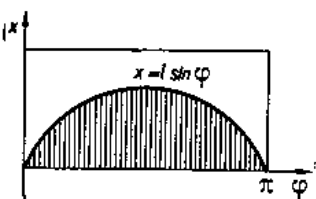


Рис. 6

$$P(C) = \frac{\int_0^\pi l \sin \varphi d\varphi}{a\pi} = \frac{2l}{a\pi}. \quad (2)$$

Формулу (2) можно использовать для экспериментального определения числа π . Предположим, что игла брошена на плоскость n раз, причем n довольно велико, и пусть m — число происшедших пересечений. Если построенная модель адекватно описывает эксперимент, то при большом n частота $\frac{m}{n}$ числа пересечений должна быть близка к вероятности $P(C)$, т. е., должно выполняться соотношение $\frac{2l}{a\pi} \sim \frac{m}{n}$, откуда получаем

$$\pi \sim \frac{2ln}{am}. \quad (3)$$

Таких экспериментов для определения π было проделано много и различными исследователями.

Экспериментатор, год	$\frac{l}{a}$	Число бросаний	Число пересечений	Экспериментальное значение
Вольф, 1850	0,8	5000	2532	3,1596
Смит, 1855	0,6	3204	1218	3,1553
Де Морган, 1860	1	600	383	3,137
Лазерини, 1901	0,83	3408	1808	3,145929
Гриджман, 1960	0,7857	2	1	3,143

Данные заимствованы из книги М. Кендалл, П. Моран. Геометрические вероятности. М., Наука, 1972.

§ 6. Аксиомы теории вероятностей

Сформулируем теперь аксиомы теории вероятностей. При этом мы исходим из свойств частот и свойств вероятности в изложенных выше вероятностных моделях экспериментов.

Рассмотрим некоторый стохастический эксперимент. Пусть Ω — пространство элементарных событий. Предположим, что в Ω выделена система \mathcal{F} подмножеств, являющаяся σ -алгеброй. Это означает, что:

$A_1) \Omega \in \mathcal{F}$;

$A_2) \text{ если } A \in \mathcal{F}, \text{ то } \bar{A} = \Omega \setminus A \in \mathcal{F}$;

A_3) из того, что $A_i \in \mathcal{F}$, $i = 1, 2, \dots$ следует, что $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Множества из \mathcal{F} называют случайными событиями. Предположим, что каждому случайному событию A (множеству из \mathcal{F}) поставлено в соответствие число $P(A)$ (назовем его вероятностью случайного события A), обладающее следующими свойствами:

P_1) $P(A) \geq 0$ для каждого $A \in \mathcal{F}$;

P_2) $P(\Omega) = 1$;

P_3) если $\{A_i\}$ — последовательность случайных событий такая, что $A_i \cap A_j = \emptyset$ ($i \neq j$), то

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

Утверждения A_1 , A_2 , A_3 , P_1 , P_2 , P_3 и составляют систему аксиом теории вероятностей. В таком виде аксиоматика теории вероятностей была сформулирована А. Н. Колмогоровым и оказалась исключительно плодотворной для развития теории вероятностей и целого ряда ее новых разделов, в первую очередь теории случайных процессов.

Заметим, что аксиомы P_1 и P_2 означают, что функция множества $P(A)$, определенная на \mathcal{F} , является мерой, удовлетворяющей дополнительному условию $P(\Omega) = 1$. Такая мера называется вероятностной мерой. Тройка (Ω, \mathcal{F}, P) , где \mathcal{F} есть σ -алгебра подмножеств из Ω , а $P(\cdot)$ — вероятностная мера на \mathcal{F} , называется вероятностным пространством. Говорят, что построена вероятностная модель эксперимента, если построено вероятностное пространство (Ω, \mathcal{F}, P) , т. е. указано пространство элементарных событий Ω , σ -алгебра \mathcal{F} случайных событий и определена вероятностная мера $P(\cdot)$ на \mathcal{F} .

ПРИМЕРЫ

1. Рассматривают стохастический эксперимент с конечным числом одинаково возможных элементарных событий $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$. В качестве \mathcal{F} возьмем σ -алгебру всех подмножеств из Ω . Положим $P(A) = \frac{m}{n}$, где m — число элементарных событий, входящих в A . Тогда все утверждения A_1 , A_2 , A_3 , P_1 , P_2 , P_3 выполнены. Таким образом, (Ω, \mathcal{F}, P) — вероятностная модель рассматриваемого эксперимента.

2. Задача о встрече. Предположим, что в квадрате $\Omega = [0, T] \times [0, T]$ наугад выбирается точка, причем «одинаково возможен» выбор любой точки квадрата Ω (другая интерпретация рассматриваемого эксперимента описана в примере 3, с. 5). Пусть \mathcal{F} есть σ -алгебра борелевских множеств из Ω . Положим для каждого $A \in \mathcal{F}$

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

где $m(A)$ — лебегова мера (площадь) множества A . Легко проверить справедливость аксиом A_1 , A_2 , A_3 , P_1 , P_2 , P_3 . Вероятностная модель эксперимента построена.

Примеры 1 и 2 показывают, что предложенная система аксиом непротиворечива: существуют реальные объекты, которые удовлетворяют этой системе аксиом.

Свойства вероятности. Используя аксиомы $A_1, A_2, A_3, P_1, P_2, P_3$, установим ряд важных свойств вероятности.

Теорема 1. Вероятность события, противоположного событию A , равна $1 - P(A)$:

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

Так как $A \cup \bar{A} = \Omega$, $A \cap \bar{A} = \emptyset$, то согласно аксиоме P_3

$$P(A) + P(\bar{A}) = P(\Omega).$$

Поэтому $P(\bar{A}) = 1 - P(A)$. Теорема доказана.

задача

Выбирается наугад четырехзначный номер, составленный из цифр 0, 1, ..., 9. Какова вероятность события A , состоящего в том, что этот номер содержит, по крайней мере, одну из трех данных цифр?

Решение. Вычислим вероятность противоположного события \bar{A} . Событие состоит в том, что выбранный номер не содержит ни одной из трех данных цифр. Так как число всех четырехзначных номеров по правилу умножения равно 10^4 , а число номеров, не содержащих ни одной из трех данных цифр, равно 7^4 , то $P(\bar{A}) = \left(\frac{7}{10}\right)^4$. Поэтому, согласно (1),

$$P(A) = 1 - P(\bar{A}) = 1 - (0,7)^4 \approx 0,76.$$

Следствие 1. Вероятность невозможного события равна нулю

$$P(\emptyset) = 0.$$

Положим в равенстве (1), которое справедливо для каждого случайного события A из \mathcal{F} , $A = \Omega$. Тогда

$$P(\emptyset) = P(\bar{\Omega}) = 1 - P(\Omega) = 0,$$

поскольку из P_2 имеем $P(\Omega) = 1$.

Замечание. Иногда ошибочно считают, что событие нулевой вероятности обязательно есть невозможное событие. Это не так. Например, пусть выбирается наугад число из отрезка $[0, 1]$. Положим $\Omega = [0, 1]$. Пусть \mathcal{F} есть σ -алгебра борелевских множеств отрезка $[0, 1]$ и $P(A) = m(A)$, где $m(A)$ — лебегова мера множества, а C — множество рациональных чисел отрезка $[0, 1]$. Тогда вероятность того, что событие C произойдет, равна нулю. Однако событие C может произойти, оно не является невозможным.

Теорема 2. Пусть A и B — случайные события, такие, что $A \subset B$. Тогда

$$P(B \setminus A) = P(B) - P(A). \quad (2)$$

Доказательство. Так как $A \subset B$, то

$$B = A \cup (B \setminus A),$$

причем $A \cap (B \setminus A) = \emptyset$. Поэтому по аксиоме P_3

$$P(B) = P(A) + P(B \setminus A),$$

а отсюда следует равенство (2).

Следствие 2. Если $A \subset B$, то

$$P(A) \leq P(B).$$

Действительно, по аксиоме P_1 имеем $P(B \setminus A) \geq 0$.

Тогда из равенства (2) получаем $P(B) \geq P(A)$.

Следствие 3. Для каждого случайного события A

$$P(A) \leq 1.$$

В самом деле, для каждого случайного события A $A \subset \Omega$. Поэтому

$$P(A) \leq P(\Omega) = 1.$$

Теорема 3. Пусть A и B — случайные события. Тогда

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (3)$$

Теорему 3 иногда называют теоремой сложения.

Для доказательства заметим, что множество $A \cup B$ можно представить в виде объединения трех попарно непересекающихся множеств

$$A \cup B = (A \setminus A \cap B) \cup (B \setminus A \cap B) \cup (A \cap B)$$

(рис. 2).. Поэтому по аксиоме P_3 и теореме 2

$$\begin{aligned} P(A \cup B) &= P(A \setminus A \cap B) + P(B \setminus A \cap B) + P(A \cap B) = \\ &= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) = \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

Из (3) легко вывести формулу сложения для трех событий

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) - \\ &- P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3). \end{aligned} \quad (4)$$

Действительно,

$$\begin{aligned} P((A_1 \cup A_2) \cup A_3) &= P(A_1 \cup A_2) + P(A_3) - P((A_1 \cup A_2) \cap A_3) = \\ &= P(A_1) + P(A_2) - P(A_1 \cap A_2) + P(A_3) - P(A_1 \cap A_3) - \\ &- P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3). \end{aligned}$$

Легко получить обобщение формул (3) и (4) для любого числа событий.

Теорема 4. Пусть A_1, \dots, A_n — случайные события. Тогда

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \\ &+ \sum_{1 \leq i_1 < i_2 < i_3 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n). \end{aligned} \quad (5)$$

Доказательство. Проведем доказательство методом математической индукции. При $n=2$ и $n=3$ теорема доказана.

Предположим, что формула 5 верна для любых $n-1$ событий. Это означает, что:

$$\begin{aligned} P\left(\bigcup_{i=2}^n A_i\right) &= \sum_{i=2}^n P(A_i) - \sum_{2 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) - \\ &- \sum_{2 \leq i_1 < i_3 \leq n} P(A_{i_1} \cap A_{i_3}) - \dots + (-1)^{n-2} P(A_2 \cap \dots \cap A_n), \\ P\left(\bigcup_{i=2}^n (A_1 \cap A_i)\right) &= \sum_{i=2}^n P(A_1 \cap A_i) - \sum_{2 \leq i_1 < i_2 \leq n} P(A_1 \cap A_{i_1} \cap A_{i_2}) + \\ &+ \sum_{2 \leq i_1 < i_2 < i_3 \leq n} P(A_1 \cap A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \dots + \\ &+ (-1)^{n-2} P(A_1 \cap A_2 \cap \dots \cap A_n). \end{aligned}$$

Далее имеем, применяя формулу (3),

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= P(A_1) + P\left(\bigcup_{i=2}^n A_i\right) - P\left(\bigcup_{i=2}^n (A_1 \cap A_i)\right) = \\ &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \\ &+ \sum_{1 \leq i_1 < i_2 < i_3 \leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \dots + \\ &+ (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n). \end{aligned}$$

Теорема доказана.

В качестве иллюстрации общей формулы сложения (5) рассмотрим следующий пример:

Задача о совпадениях. На отдельных карточках написаны числа $1, 2, \dots, n$. Карточки расположены в «абсолютно случайном» порядке. Какова вероятность того, что хотя бы одно из чисел окажется на месте с таким же номером?

Эта задача часто встречается в различных шуточных вариантах. Вот, например, один из них. Некто написал n писем, а затем написал адреса на конвертах в «абсолютно случайном» порядке. Какова вероятность того, что хотя бы одно письмо будет получено своим адресатом?

Под «абсолютно случайным» расположением карточек мы понимаем следующее: все $n!$ возможных перестановок карточек одинаково возможны. Пусть A_i — событие, состоящее в том, что карточ-

ка с номером i окажется на месте с номером i . Надо вычислить $P(\bigcup_{i=1}^n A_i)$. Имеем:

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n},$$

$$P(A_{i_1} \cap A_{i_2}) = \frac{(n-2)!}{n!},$$

$$\dots \dots \dots$$

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = \frac{(n-k)!}{n!}.$$

Используя формулу (5), получим

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= C_n^1 \frac{(n-1)!}{n!} - C_n^2 \frac{(n-2)!}{n!} + C_n^3 \frac{(n-3)!}{n!} - \dots + \\ &+ (-1)^{n-1} C_n^n \frac{1}{n!} = 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n-1} \frac{1}{n!}. \end{aligned}$$

Заметим, что искомая вероятность является частичной суммой ряда Тейлора функции $1 - e^x$ при $x = -1$. Поэтому для больших n имеем

$$P\left(\bigcup_{i=1}^n A_i\right) \approx 1 - e^{-1} \approx 0,63.$$

При доказательстве сформулированных выше свойств вероятности мы использовали аксиому P_3 лишь для случая конечного числа случайных событий. Поэтому теоремы 1—4 справедливы и тогда, когда относительно вероятности предполагается лишь конечная аддитивность. Следующая теорема играет важную роль, и для ее справедливости требуется выполнение счетной аддитивности вероятности.

Теорема 5. Если $\{A_n\}$ — монотонно неубывающая последовательность случайных событий, т. е.

$$A_1 \subset A_2 \subset \dots \subset A_n \subset A_{n+1} \subset \dots,$$

то

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n). \quad (6)$$

Если A_n — монотонно невозрастающая последовательность событий, т. е.

$$A_1 \supset A_2 \supset \dots \supset A_n \supset A_{n+1} \supset \dots,$$

то

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} P(A_n). \quad (7)$$

Заметим, что пользуясь понятием предела монотонной последовательности событий (см. § 1), равенства (6) и (7) можно записать в следующем виде:

$$P(\lim_{n \rightarrow \infty} A_n) = \lim_{n \rightarrow \infty} P(A_n), \quad (8)$$

где $\{A_n\}$ — монотонная последовательность. Поэтому теорему 5 часто называют свойством непрерывности вероятности. Теорема 5 не нуждается в отдельном доказательстве, так как в § 4 доказана более общая теорема, устанавливающая эквивалентность свойства счетной аддитивности конечной меры и свойства непрерывности.

Теорема 6. Для любого конечного или счетного числа случайных событий $\{A_n\}$ имеют место следующие неравенства:

$$P(\bigcup_n A_n) \leq \sum_n P(A_n), \quad (9)$$

$$P(\bigcap_n A_n) \geq 1 - \sum_n P(\bar{A}_n). \quad (10)$$

Для доказательства неравенства (9) заметим следующее. Легко показать, что

$$\bigcup_n A_n = \bigcup_n B_n,$$

где $B_1 = A_1$, $B_2 = \bar{A}_1 \cap A_2$, ..., $B_n = \bar{A}_1 \cap \dots \cap \bar{A}_{n-1} \cap A_n$, События B_n попарно несовместны и $B_n \subset A_n$. Поэтому

$$P(\bigcup_n A_n) = P(\bigcup_n B_n) = \sum_n P(B_n) \leq \sum_n P(A_n).$$

Неравенство (10) вытекает из следующих рассуждений:

$$P(\bigcap_n A_n) = 1 - P(\overline{\bigcap_n A_n}) = 1 - P(\bigcup_n \bar{A}_n) \geq 1 - \sum_n P(\bar{A}_n),$$

здесь использовано неравенство

$$P(\bigcup_n \bar{A}_n) \leq \sum_n P(\bar{A}_n).$$

§ 7. Условные вероятности

Условные вероятности. В ряде случаев приходится рассматривать вероятности случайных событий, если известно, что произошло некоторое случайное событие B , имеющее положительную вероятность. Прежде чем давать точное определение условной вероятности, рассмотрим несколько примеров.

ПРИМЕРЫ

1. Пусть дважды брошена симметричная игральная кость. B — событие, состоящее в том, что сумма появившихся очков меньше 4, а A — событие, состоящее в том, что первый раз появилась 1. Вычислим условную вероятность $P(A/B)$ события A , если известно, что произошло событие B .

Пространство элементарных событий состоит из 36 исходов:

$$\Omega = \{(i, j) : i = \overline{1, 6}, j = \overline{1, 6}\}.$$

События A и B — это следующие подмножества Ω :

$$A = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6)\}$$

$$B = \{(1, 1), (1, 2), (2, 1)\}.$$

Поэтому

$$P(A) = \frac{6}{36} = \frac{1}{6}, \quad P(B) = \frac{3}{36} = \frac{1}{12}.$$

$$P(A \cap B) = \frac{1}{36}.$$

Если известно, что событие B произошло, то событие A может появиться лишь тогда, когда произойдет элементарное событие $(1, 1)$. Поэтому естественно считать, что

$$P(A/B) = \frac{1}{3} = \frac{P(A \cap B)}{P(B)}.$$

2. Пусть пространство элементарных событий Ω состоит из n одинаково возможных элементарных событий, событию A благоприятствует m элементарных событий, B состоит из l исходов, событию $A \cap B$ благоприятствует r исходов. Найдем $P(A/B)$.

Заметим, что

$$P(B) = \frac{l}{n}, \quad P(A \cap B) = \frac{r}{n}.$$

Если произошло событие B , то произошло одно из l элементарных событий, событие A произойдет только тогда, когда произойдет одно из r элементарных событий, составляющих $A \cap B$. Поэтому

$$P(A/B) = \frac{r}{l} = \frac{\frac{r}{n}}{\frac{l}{n}} = \frac{P(A \cap B)}{P(B)}.$$

3. Пусть в квадрат $\Omega = [0, 1]^2$ наугад бросают точку. Вероятность попадания во множество A равна $m(A)$, где $m(A)$ — площадь A .

Если мы интересуемся вероятностью $P(A/B)$ того, что точка попадает во множество A , при дополнительном предположении, что точка попадает во множество B , то естественно считать, что вероятность $P(A/B)$ пропорциональна $P(A \cap B)$, т. е.

$$P(A/B) = \frac{P(A \cap B)}{P(B)}.$$

Определение. Пусть (Ω, \mathcal{F}, P) — вероятностное пространство, $P(B) > 0$, $B \in \mathcal{F}$. Условной вероятностью события A при условии, что произошло событие B , называют

$$P(A/B) = \frac{P(A \cap B)}{P(B)}. \quad (1)$$

Из этого определения и свойств вероятности непосредственно следует, что:

$$1) \quad P(A/B) \geq 0,$$

$$2) P(\Omega/B) = 1,$$

$$3) P(B/B) = 1,$$

4) если $\{A_i\}$ — последовательность попарно несовместных случайных событий ($A_i \cap A_j = \emptyset$), то

$$P\left(\bigcup_{i=1}^{\infty} A_i/B\right) = \sum_{i=1}^{\infty} P(A_i/B).$$

Докажем 2) и 4). Имеем:

$$\begin{aligned} P(\Omega/B) &= \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1, \\ P\left(\bigcup A_i/B\right) &= \frac{P\left(\left(\bigcup A_i\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup (A_i \cap B)\right)}{P(B)} = \\ &= \frac{\sum P(A_i \cap B)}{P(B)} = \sum \frac{P(A_i \cap B)}{P(B)} = \sum P(A_i/B). \end{aligned}$$

Заметим также, что $P(B/B) = 1$. Обозначим через \mathcal{F}_B δ -алгебру всех множеств вида $A \cap B$, где $A \in \mathcal{F}$. Из свойств 1)–4) следует, что $(B, \mathcal{F}_B, P(\cdot/B))$ — вероятностное пространство.

Соотношение (1) можно записать в виде $P(A \cap B) = P(B) P(A/B)$. Если $P(A) > 0$, то можно записать, что $P(A \cap B) = P(A) P(B/A)$. Итак, имеет место следующее утверждение.

Теорема 1. Если $P(B) > 0$, $P(A) > 0$, то имеют место равенства

$$P(A \cap B) = P(B) P(A/B) = P(A) P(B/A). \quad (2)$$

Соотношение (2) называют формулой умножения. Формула (2) допускает следующее обобщение.

Пусть A_1, \dots, A_n — случайные события такие, что $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$. Поскольку

$$\bigcap_{i=1}^{n-1} A_i \subset \bigcap_{i=1}^{n-2} A_i \subset \dots \subset A_2 \cap A_1 \subset A_1,$$

то из неравенства $P(A_1 \cap \dots \cap A_{n-1}) > 0$ следует, что

$$P\left(\bigcap_{i=1}^{n-l} A_i\right) > 0, \quad l = 1, 2, \dots, n-1.$$

Следовательно, все условные вероятности $P(A_n/\bigcap_{i=1}^{n-l} A_i)$ определены.

Легко проверить (можно воспользоваться методом математической индукции), что

$$\begin{aligned} P\left(\bigcap_{i=1}^n A_i\right) &= P(A_1) P(A_2/A_1) P(A_3/A_1 \cap A_2) \dots \\ &\dots P(A_n/A_1 \cap \dots \cap A_{n-1}). \end{aligned} \quad (3)$$

Формула полной вероятности.

Определение. Говорят, что набор случайных событий H_1, \dots, H_n образует полную группу событий, если:

$$1) H_1 \cup H_2 \cup \dots \cup H_n = \Omega; \quad (4)$$

$$2) H_i \cap H_j = \emptyset \quad (i \neq j). \quad (5)$$

Теорема 2. Если H_1, \dots, H_n — полная группа событий и $P(H_i) > 0, i = \overline{1, n}$, то для любого случайного события $A \in \mathcal{F}$ имеет место равенство

$$P(A) = \sum_{i=1}^n P(H_i) P(A/H_i). \quad (6)$$

Равенство (4) называют формулой полной вероятности. Для доказательства заметим, что в силу (4)

$$A = A \cap \Omega = \bigcup_{i=1}^n (H_i \cap A).$$

В силу (5) случайные события $H_i \cap A$ попарно несовместны. Поэтому

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

Формула (6) справедлива и для счетного набора событий, удовлетворяющего условиям (4) и (5).

Рассмотрим примеры применения формулы полной вероятности.

задачи

1. Пусть имеется n одинаковых урн. Известно, что урна с номером i содержит m_i белых шаров, всего же шаров в этой урне N_i . Наугад выбрана урна, а из нее шар. Какова вероятность того, что взят белый шар?

Пусть H_i — событие, состоящее в том, что выбрана урна с номером i , а A — событие, состоящее в том, что взят белый шар. Тогда

$$P(H_i) = \frac{1}{n}, \quad P(A/H_i) = \frac{m_i}{N_i}.$$

Следовательно, по формуле полной вероятности

$$P(A) = \sum_{i=1}^n P(H_i) P(A/H_i) = \frac{1}{n} \sum_{i=1}^n \frac{m_i}{N_i}.$$

2. Среди N экзаменационных билетов n «счастливых». Студенты подходят за билетами один за другим. У кого больше вероятность взять «счастливый» билет: у того, кто подошел первым или у того, кто подошел вторым?

Вероятность взять «счастливый» билет для первого студента равна, очевидно, $\frac{n}{N}$. Пусть A событие, состоящее в том, что второй студент взял «счастливый» билет. Можно сделать два предположения: H_1 — первый студент

взял «счастливый» билет, H_2 — первый не взял «счастливый» билет. По формуле полной вероятности

$$\begin{aligned} P(A) &= P(H_1) P(A/H_1) + P(H_2) P(A/H_2) = \\ &= \frac{n}{N} \cdot \frac{n-1}{N-1} + \frac{N-n}{N} \cdot \frac{n}{N-1} = \frac{n}{N}. \end{aligned}$$

Итак, вероятность взять «счастливый» билет для второго студента также равна $\frac{n}{N}$.

Формулы Бейеса. Докажем следующую теорему.

Теорема. Пусть набор событий H_i , $i = \overline{1, n}$ образует полную группу событий, причем $P(H_i) > 0$ для каждого $i = \overline{1, n}$. Тогда для любого случайного события B такого, что $P(B) > 0$, выполнены равенства

$$P(H_i/B) = \frac{P(H_i) P(B/H_i)}{\sum_{k=1}^n P(H_k) P(B/H_k)}. \quad (7)$$

Формулы (7) называются формулами Бейеса. Общая схема применения этих формул при решении практических задач следующая. Пусть событие B может происходить в различных условиях, о характере которых можно сделать n гипотез H_1, H_2, \dots, H_n . Из каких-то соображений известны вероятности этих гипотез $P(H_1), \dots, P(H_n)$ (априорные вероятности), известны также условные вероятности $P(B/H_1), \dots, P(B/H_n)$. Предположим, что произведен опыт, в результате которого наступило событие B . Это должно вызвать переоценку вероятностей гипотез H_i ; формулы Бейеса и дают выражение для условных вероятностей $P(H_i/B)$ (эти вероятности называются апостериорными вероятностями).

Для доказательства (7) используем формулу умножения и формулу полной вероятности. Имеем

$$P(H_i/B) = \frac{P(H_i \cap B)}{P(B)} = \frac{P(H_i) P(B/H_i)}{\sum_{k=1}^n P(H_k) P(B/H_k)}.$$

Рассмотрим примеры применения формул Бейеса.

ЗАДАЧИ

1. В урне находится n шаров. Возможна $(n+1)$ гипотеза о количестве белых шаров в урне: H_0, H_1, \dots, H_n (H_i — гипотеза, состоящая в том, что в урне i белых шаров). Предположим, что все эти гипотезы одинаково возможны

$$P(H_0) = P(H_1) = \dots = P(H_n) = \frac{1}{n+1}.$$

Из урны наугад взяли шар, который оказался белым. Пусть B — событие, состоящее в том, что наугад взятый шар из урны — белый. Вычислить $P(H_i/B)$.

Имеем $P(B/H_1) = \frac{1}{n}$ и по формуле (7)

$$P(H_1/B) = \frac{\frac{1}{n+1} \cdot \frac{1}{n}}{\sum_{k=0}^n \frac{1}{n+1} \cdot \frac{k}{n}} = \frac{2l}{n(n+1)}.$$

Таким образом, наиболее вероятной является гипотеза H_n .

2. Некоторая деталь производится на двух заводах. Известно, что объем продукции первого завода в n раз превышает объем продукции второго завода. Доля брака на первом заводе P_1 , на втором — P_2 . Наугад взятая деталь оказалась бракованной. Какова вероятность того, что эта деталь выпущена первым заводом?

Пусть H_1 — событие, состоящее в том, что взятая деталь изготовлена на первом заводе, H_2 — событие, состоящее в том, что эта деталь изготовлена на втором заводе. Заметим, что

$$P(H_1) = \frac{n}{n+1}, \quad P(H_2) = \frac{1}{n+1}.$$

Пусть B — событие, состоящее в том, что наугад взятая деталь оказалась бракованной. По условию задачи $P(B/H_1) = P_1$, $P(B/H_2) = P_2$. По формуле (7)

$$P(H_1/B) = \frac{\frac{n}{n+1} \cdot P_1}{\frac{n}{n+1} P_1 + \frac{1}{n+1} P_2} = \frac{nP_1}{nP_1 + P_2}.$$

§ 8. Независимые случайные события

В этом параграфе мы введем понятие независимости случайных событий, которое играет в теории вероятностей очень важную роль.

Определение. Пусть (Ω, \mathcal{F}, P) — вероятностное пространство. Случайные события A и B ($A \in \mathcal{F}$, $B \in \mathcal{F}$) называются независимыми, если

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

ПРИМЕРЫ

1. Дважды бросается игральная кость.

Тогда $\Omega = \{ГГ, ГР, РГ, РР\}$. Пусть A — событие, состоящее в том, что первый раз выпал герб, B — событие, состоящее в том, что второй раз выпал герб. Тогда

$$A = \{ГГ, ГР\}; \quad B = \{ГГ, РГ\}; \quad A \cap B = \{ГГ\}$$

и

$$P(A) = \frac{2}{4} = \frac{1}{2}, \quad P(B) = \frac{2}{4} = \frac{1}{2}, \quad P(A \cap B) = \frac{1}{4}.$$

Итак, $P(A \cap B) = P(A) P(B)$. Следовательно, случайные события A и B независимы.

2. Предположим, что в квадрат $\Omega = [0, 1]^2$ наугад бросают точку. Пусть \mathcal{F} — σ -алгебра всех борелевских множеств из Ω , и если $A \in \mathcal{F}$, то $P(A) = m(A)$, где $m(A)$ — площадь множества A . Мы имеем дело с «геометриче-

скими вероятностями». Пусть A — событие, состоящее в том, что абсцисса брошенной точки не меньше a , $0 < a < 1$, а B — событие, состоящее в том, что ордината брошенной точки не меньше b , $0 < b < 1$. Тогда (рис. 7)

$$P(A) = 1 - a, \quad P(B) = 1 - b, \\ P(A \cap B) = (1 - a)(1 - b),$$

и, следовательно, события A и B независимы.

Рассмотрим некоторые простые, но важные свойства независимых событий.

Теорема 1. Пусть $P(B) > 0$. Случайные события A и B независимы тогда и только тогда, когда $P(A/B) = P(A)$ (появление события B не изменяет вероятности события A).

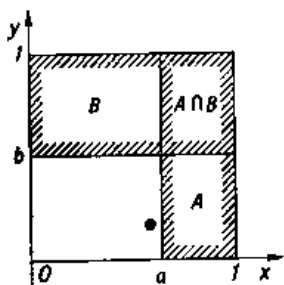


Рис. 7

Действительно, если A и B независимы и $P(B) > 0$, то $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$. Если же известно, что

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = P(A),$$

то $P(A \cap B) = P(A)P(B)$, а это означает, что A и B независимы.

Теорема 2. Если A и B независимы, то события A и \bar{B} , \bar{A} и \bar{B} также независимы.

Достаточно доказать, что A и \bar{B} независимы. Заметим, что $A \cap \bar{B} = A \setminus (A \cap B)$. Поэтому

$$P(A \cap \bar{B}) = P(A) - P(A \cap B).$$

Если A и B независимы, то

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

Следовательно, A и \bar{B} независимы.

Теорема 3. Если случайные события A и B_1 независимы, A и B_2 независимы, B_1 и B_2 несовместны ($B_1 \cap B_2 = \emptyset$), то A и $B_1 \cup B_2$ независимы.

Действительно,

$$P(A \cap (B_1 \cup B_2)) = P((A \cap B_1) \cup (A \cap B_2)).$$

Так как $A \cap B_1$ и $A \cap B_2$ несовместны, то

$$P((A \cap B_1) \cup (A \cap B_2)) = P(A \cap B_1) + P(A \cap B_2) = \\ = P(A)P(B_1) + P(A)P(B_2) = P(A)[P(B_1) + P(B_2)] = \\ = P(A)P(B_1 \cup B_2).$$

Следовательно,

$$P(A \cap (B_1 \cup B_2)) = P(A)P(B_1 \cup B_2),$$

т. е. A и $B_1 \cup B_2$ независимы.

Определение. Случайные события B_1, \dots, B_n независимы в совокупности, если для любого k , $1 \leq k \leq n$, и для любого набора индексов i_1, i_2, \dots, i_k , $1 \leq i_1 \leq i_2 < \dots < i_k \leq n$,

$$P\left(\bigcap_{r=1}^k B_{i_r}\right) = \prod_{r=1}^k P(B_{i_r}).$$

В частности, если B_1, \dots, B_n независимы в совокупности, то любые два события B_i и B_j , $i \neq j$, независимы. Однако из попарной независимости независимость в совокупности не следует. Это показывает следующий пример:

Пример С. Н. Бериштейна. На плоскость бросают тетраэдр, три грани которого окрашены соответственно в красный, зеленый, голубой цвета, а на четвертую грань нанесены все три цвета.

Пусть событие K состоит в том, что при бросании тетраэдра на плоскость выпала грань красного цвета и пусть аналогично определены события $З$ и $Г$. Так как каждый из трех цветов нанесен на две грани, то

$$P(K) = P(З) = P(Г) = \frac{2}{4} = \frac{1}{2}.$$

Далее

$$P(K \cap З) = P(K \cap Г) = P(З \cap Г) = \frac{1}{4},$$

и, следовательно, события K , $З$, $Г$ попарно независимы. Однако эти события не являются независимыми в совокупности, потому что

$$P(K \cap З \cap Г) = \frac{1}{4} \neq P(K)P(З)P(Г) = \frac{1}{8}.$$

Глава II

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

§ 1. Случайные величины

Случайные величины — функции на пространстве элементарных событий. Одним из основных понятий теории вероятностей является понятие случайной величины. Случайная величина — это величина, принимающая те или иные значения, в зависимости от случая. Примерами случайных величин могут быть число очков, выпадающих при одном бросании игральной кости, число бракованных изделий среди взятых наугад n изделий, число попа-

даний в цель при n выстрелах, время безотказной работы прибора, дальность полета баллистической ракеты и т. д. Случайная величина ξ есть число, которое ставится в соответствие каждому возможному исходу эксперимента. Поскольку исходы эксперимента описываются элементарными событиями, случайную величину можно рассматривать как функцию $\xi = \xi(\omega)$ на пространстве элементарных событий Ω .

Прежде чем давать определение случайной величины, рассмотрим несколько примеров.

ПРИМЕРЫ

1. Пусть дважды бросают монету.

Пространство элементарных событий имеет вид

$$\Omega = \{ГГ, ГР, РГ, РР\}.$$

Пусть ξ — число появлений герба. Величина ξ является функцией $\xi = \xi(\omega)$ элементарного события. Таблица значений функции $\xi(\omega)$ имеет следующий вид:

ω	ГГ	ГР	РГ	РР
$\xi(\omega)$	2	1	1	0

2. Бросают игральную кость.

В этом случае пространство элементарных событий имеет вид

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\},$$

где ω_i означает, что выпало i очков. Случайная величина ξ — число появившихся очков — является функцией элементарного события, причем $\xi(\omega) = i$, если $\omega = \omega_i$.

3. Монету бросают до тех пор, пока не появится герб. В этом случае

$$\begin{aligned} \Omega &= \{Г, РГ, РРГ, \dots, \underbrace{РР \dots РГ}_{n-1}, \dots\} = \\ &= \{\omega_1, \omega_2, \omega_3, \dots, \omega_n, \dots\}. \end{aligned}$$

Пусть ξ — число произведенных бросаний. Тогда величина ξ является функцией элементарного события, причем $\xi(\omega) = n$, если $\omega = \omega_n$ ($n = 1, 2, \dots$).

Определение случайной величины. Рассмотренные примеры подтверждают, что случайные величины можно интерпретировать как функции на пространстве элементарных событий стохастического эксперимента. Однако не любые функции, определенные на Ω , можно рассматривать в качестве случайных величин. При изучении величин нам часто придется отвечать на вопрос: какова вероятность того, что значения случайной величины $\xi(\omega)$ принадлежат тому или иному множеству. Следовательно, для достаточно широкого класса множеств $\{B\}$ на числовой прямой мы должны быть уверены, что множество

$$\{\omega : \xi(\omega) \in B\}$$

принадлежит σ -алгебре \mathcal{F} случайных событий, и поэтому можно рассматривать вероятность $P\{\omega : \xi(\omega) \in B\}$. Оказывается (в этом мы убедимся ниже), что достаточно предположить, что для каждого интервала $(-\infty, x)$ множество

$$\{\omega : \xi(\omega) \in (-\infty, x)\} = \{\omega : \xi(\omega) < x\}$$

принадлежит σ -алгебре \mathcal{F} случайных событий, и тогда для каждого борелевского множества B получим

$$\{\omega : \xi(\omega) \in B\} \in \mathcal{F}.$$

Определение. Пусть (Ω, \mathcal{F}, P) — вероятностное пространство. Всякая действительная функция $\xi = \xi(\omega)$ на Ω такая, что для каждого действительного x

$$\{\omega : \xi(\omega) < x\} \in \mathcal{F} \quad (1)$$

называется случайной величиной.

Заметим, что в теории функций функция $\xi = \xi(\omega)$ на Ω называется измеримой относительно σ -алгебры \mathcal{F} , если для каждого вещественного x выполнено условие (1). Таким образом, случайная величина на (Ω, \mathcal{F}, P) это функция, измеримая относительно σ -алгебры \mathcal{F} .

Определение. Функция

$$F_{\xi}(x) = P\{\omega : \xi(\omega) < x\} \quad (2)$$

называется функцией распределения случайной величины $\xi(\omega)$.

Рассмотрим еще несколько примеров случайных величин.

ПРИМЕРЫ

1. Пусть один раз бросают симметричную монету.

В этом случае $\Omega = \{\omega_1, \omega_2\}$, где $\omega_1 = \Gamma$, $\omega_2 = \Pi$; \mathcal{F} — совокупность всех подмножеств Ω , $P(\{\omega_1\}) = P(\{\omega_2\}) = \frac{1}{2}$. Предположим, что

$$\xi(\omega) = \begin{cases} 1, & \omega = \omega_1; \\ -1, & \omega = \omega_2. \end{cases}$$

Заметим

$$\{\omega : \xi(\omega) < x\} = \begin{cases} \emptyset, & x \leq -1; \\ \{\omega_2\}, & -1 < x \leq 1; \\ \Omega, & 1 < x. \end{cases}$$

Таким образом, $\xi(\omega)$ — случайная величина и

$$F_{\xi}(x) = \begin{cases} 0, & x \leq -1; \\ \frac{1}{2}, & -1 < x \leq 1; \\ 1, & x > 1. \end{cases}$$

График функции распределения изображен на рис. 8.

2. Равномерное распределение на отрезке $[a, b]$. На отрезок $[a, b]$ числовой прямой наугад бросают точку, причем считают, что все положения точки «одинаково возможны».

В рассматриваемом стохастическом эксперименте $\Omega = [a, b]$, \mathcal{F} есть σ -алгебра борелевских подмножеств из $[a, b]$, $P(\cdot)$ — такая мера на \mathcal{F} , что $P\{[\alpha, \beta]\} = \frac{\beta - \alpha}{b - a}$, если $[\alpha, \beta] \subset [a, b]$. Определим случайную величину ξ следующим образом: $\xi(\omega) = \omega$, если $\omega \in [a, b]$, т. е. ξ — координата полученной точки. Заметим, что

$$\{\omega : \xi(\omega) < x\} = \begin{cases} \emptyset, & x \leq a; \\ [a, x), & a < x \leq b; \\ \Omega = [a, b], & x > b. \end{cases}$$

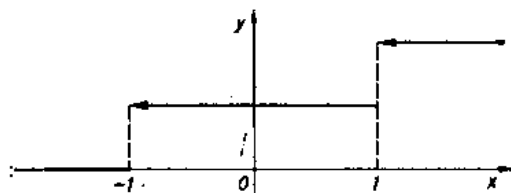


Рис. 8

Таким образом, при каждом x $\{\omega : \xi(\omega) < x\} \in \mathcal{F}$, т. е. $\xi(\omega)$ — случайная величина. Функция распределения случайной величины $\xi(\omega)$ имеет вид

$$F_{\xi}(x) = \begin{cases} 0, & x \leq a; \\ \frac{x - a}{b - a}, & a < x \leq b; \\ 1, & x > b. \end{cases}$$

Эта функция распределения (рис. 9) определяет так называемое равномерное распределение на $[a, b]$.

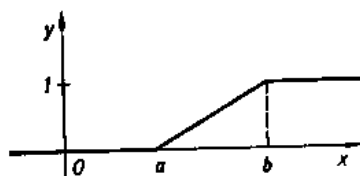


Рис. 9

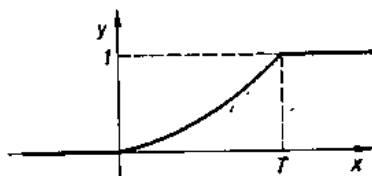


Рис. 10

3. Задача о встрече. Два лица A и B условились о встрече на интервале времени $[0, T]$.

Пусть u — время прихода A , v — время прихода B . Пространство элементарных событий имеет вид

$$\Omega = \{(u, v) : 0 \leq u \leq T, 0 \leq v \leq T\}$$

(см. пример 3, с. 4). В качестве σ -алгебры \mathcal{F} случайных событий рассмотрим σ -алгебру измеримых по Борелю подмножеств квадрата $[0, T] \times [0, T]$. Если $C \in \mathcal{F}$, то $P(C) = \frac{m(C)}{T^2}$, где $m(C)$ — площадь множества C . Пусть ξ — время ожидания. Тогда $\xi(\omega) = |u - v|$, если $\omega = (u, v)$. Заметим, что

$$\{\omega : \xi(\omega) < x\} = \begin{cases} \emptyset, & x \leq 0; \\ \{(u, v) : |u - v| < x, 0 \leq u \leq T, 0 \leq v \leq T\}, & 0 < x \leq T; \\ [0, T] \times [0, T], & x > T. \end{cases}$$

Следовательно, функция $\xi(\omega)$ измерима относительно σ -алгебры \mathcal{F} . Функция распределения $\xi(\omega)$ равна

$$P(\omega : \xi(\omega) < x) = \begin{cases} 0, & x \leq 0; \\ \frac{T^2 - (T-x)^2}{T^2}, & 0 < x \leq T; \\ 1, & x > T. \end{cases}$$

График функции распределения изображен на рис. 10.

4. Предположим, что в круг K_R радиуса R с центром в начале координат наугад бросают точку. Пусть ξ — расстояние от точки до начала координат. Пространство элементарных событий рассматриваемого эксперимента имеет вид

$$\Omega = \{(u, v) : u^2 + v^2 \leq R^2\} = K_R,$$

причем $\xi = \xi(\omega) = \sqrt{u^2 + v^2}$, если $\omega = (u, v)$. Пусть \mathcal{F} есть σ -алгебра борелевских множеств из K_R . Функция $\xi(\omega)$ измерима относительно \mathcal{F} , так как множество

$$\{\omega : \xi(\omega) < x\} = \begin{cases} \emptyset, & x \leq 0; \\ \{(u, v) : u^2 + v^2 < x^2\}, & 0 < x \leq R; \\ \Omega, & x > R \end{cases}$$

при каждом x принадлежит σ -алгебре \mathcal{F} . Если $A \in \mathcal{F}$, то

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

где $m(\cdot)$ — «площадь» множества A .

Функция распределения случайной величины $\xi(\omega)$ имеет вид

$$F_\xi(x) = \begin{cases} 0, & x \leq 0; \\ \frac{x^2}{R^2}, & 0 < x \leq R; \\ 1, & x \geq R. \end{cases}$$

5. Пусть (Ω, \mathcal{F}, P) — произвольное вероятностное пространство. Предположим, что A — некоторое случайное событие ($A \in \mathcal{F}$) и

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

Заметим, что

$$\{\omega : I_A(\omega) < x\} = \begin{cases} \emptyset, & x \leq 0; \\ \bar{A}, & 0 < x \leq 1; \\ \Omega, & x > 1 \end{cases}$$

и, следовательно, множество $\{\omega : \xi(\omega) < x\} \in \mathcal{F}$.

Итак, $\xi(\omega)$ — случайная величина, функция распределения которой равна

$$P(\omega : I_A(\omega) < x) = \begin{cases} 0, & x \leq 0; \\ P(\bar{A}), & 0 < x \leq 1; \\ 1, & x > 1. \end{cases}$$

Случайную величину $I_A(\omega)$ называют индикатором случайного события A .

Рассмотрим еще пример вероятностного пространства и функции на нем, не являющейся случайной величиной.

6. Пусть $\Omega = [0, 1]$, \mathcal{M} есть σ -алгебра измеримых по Лебегу множеств из $[0, 1]$. Положим $P(A) = m(A)$, если $A \in \mathcal{M}$, где $m(A)$ — мера Лебега множества A . Тогда (Ω, \mathcal{M}, P) — вероятностное пространство.

Предположим, что E — неизмеримое по Лебегу множество из отрезка $[0, 1]^1$. Рассмотрим функцию

$$\xi(\omega) = \begin{cases} 1, & \omega \in E; \\ -1, & \omega \in [0, 1] \setminus E = \bar{E}. \end{cases}$$

Тогда

$$\{\omega : \xi(\omega) < x\} = \begin{cases} \emptyset, & x \leq -1; \\ \bar{E}, & -1 < x \leq 1; \\ [0, 1], & x > 1. \end{cases}$$

Следовательно, при $-1 < x \leq 1$ $\{\omega : \xi(\omega) < x\} = \bar{E} \notin \mathcal{A}$, и функция $\xi(\omega)$ не является случайной величиной на вероятностном пространстве (Ω, \mathcal{A}, P) .

Случайные события, порожденные случайной величиной или системой случайных величин. Пусть $\xi(\omega)$ — случайная величина, т. е. функция $\xi(\omega) : \Omega \rightarrow R = (-\infty, +\infty)$, измеримая относительно σ -алгебры. Предположим, что B — некоторое множество на числовой прямой. Рассмотрим множество

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

Является ли это множество случайным событием? Иначе говоря, можем ли мы утверждать, что

$$\xi^{-1}(B) \in \mathcal{F}?$$

Если $B = (-\infty, x)$, то по соотношению (1) из определения случайной величины $\xi^{-1}((-\infty, x)) = \{\omega : \xi(\omega) < x\} \in \mathcal{F}$. Отсюда уже автоматически следует, что $\xi^{-1}(B) \in \mathcal{F}$ для довольно широкого класса множеств B .

Теорема 1. Пусть (Ω, \mathcal{F}, P) — вероятностное пространство и $\xi(\omega)$ — случайная величина на нем. Тогда каждое из множеств множества Ω

$$\begin{aligned} &\{\omega : \xi(\omega) \geq x\}, \quad \{\omega : \xi(\omega) \leq x\}, \\ &\{\omega : \xi(\omega) > x\}, \quad \{\omega : a \leq \xi(\omega) < b\}, \\ &\{\omega : a < \xi(\omega) \leq b\}, \quad \{\omega : \xi(\omega) = x\} \end{aligned}$$

является случайным событием, т. е. каждое из этих множеств принадлежит σ -алгебре \mathcal{F} .

Доказательство. Достаточно заметить, что

$$\begin{aligned} &\{\omega : \xi(\omega) \geq x\} = \Omega \setminus \{\omega : \xi(\omega) < x\}, \\ &\{\omega : \xi(\omega) \leq x\} = \bigcap_{n=1}^{\infty} \left\{ \omega : \xi(\omega) < x + \frac{1}{n} \right\}, \\ &\{\omega : \xi(\omega) > x\} = \Omega \setminus \{\omega : \xi(\omega) \leq x\}, \\ &\{\omega : a \leq \xi(\omega) < b\} = \{\omega : \xi(\omega) < b\} \setminus \{\omega : \xi(\omega) < a\}, \\ &\{\omega : a < \xi(\omega) \leq b\} = \{\omega : \xi(\omega) < b\} \setminus \{\omega : \xi(\omega) \leq a\}, \\ &\{\omega : \xi(\omega) = x\} = \{\omega : \xi(\omega) \leq x\} \setminus \{\omega : \xi(\omega) < x\}. \end{aligned}$$

Так как при каждом действительном x $\{\omega : \xi(\omega) < x\} \in \mathcal{F}$, то каждое из множеств в правых частях написанных выше равенств принадлежит \mathcal{F} .

¹ См. Г. И. Натансон. Теория функций вещественной переменной. М., Наука, 1974.

Обобщением теоремы 1 является следующее утверждение.

Теорема 2. Пусть (Ω, \mathcal{F}, P) — вероятностное пространство, $\xi(\omega)$ — случайная величина на нем, B — произвольное борелевское множество на числовой прямой. Тогда множество

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

является случайным событием, т. е. $\xi^{-1}(B) \in \mathcal{F}$.

Доказательство. Заметим, что операция «взятие прообраза» (переход от множества $B \subset (-\infty, +\infty)$ к множеству $\xi^{-1}(B) \subset \Omega$ сохраняет теоретико-множественные операции). Следовательно, имеют место равенства:

$$\xi^{-1}\left(\bigcup_{\alpha \in J} B_\alpha\right) = \bigcup_{\alpha \in J} \xi^{-1}(B_\alpha),$$

$$\xi^{-1}\left(\bigcap_{\alpha \in J} B_\alpha\right) = \bigcap_{\alpha \in J} \xi^{-1}(B_\alpha),$$

$$\xi^{-1}(\bar{B}) = \overline{\xi^{-1}(B)}.$$

Рассмотрим теперь класс K тех множеств B , для которых

$$K = \{B : B \subset (-\infty, +\infty), \xi^{-1}(B) \in \mathcal{F}\}.$$

Из сделанного выше замечания следует, что класс K образует σ -алгебру множеств. При доказательстве теоремы 1 мы уже убедились в том, что $[a, b] \in K$. Поэтому наименьшая σ -алгебра \mathcal{B} , содержащая все интервалы вида $[a, b]$, содержится в классе K . Итак, утверждение теоремы справедливо для любого борелевского множества B .

Теорема 3. Пусть (Ω, \mathcal{F}, P) — вероятностное пространство, а $\xi(\omega)$ и $\eta(\omega)$ — случайные величины на нем. Тогда

$$\{\omega : \xi(\omega) < \eta(\omega)\} \in \mathcal{F},$$

$$\{\omega : \xi(\omega) > \eta(\omega)\} \in \mathcal{F},$$

$$\{\omega : \xi(\omega) \leq \eta(\omega)\} \in \mathcal{F},$$

$$\{\omega : \xi(\omega) = \eta(\omega)\} \in \mathcal{F}.$$

Доказательство. Пусть $Z = \{r_k\}$ — множество всех рациональных чисел. Тогда

$$\{\omega : \xi(\omega) < \eta(\omega)\} = \bigcap_{r_k \in Z} \{\omega : \xi(\omega) < r_k < \eta(\omega)\}.$$

Поскольку

$$\{\omega : \xi(\omega) < r_k < \eta(\omega)\} = \{\omega : \xi(\omega) < r_k\} \cap \{\omega : \eta(\omega) > r_k\} \in \mathcal{F},$$

то $\{\omega : \xi(\omega) < \eta(\omega)\} \in \mathcal{F}$. Так как в предыдущем рассуждении $\xi(\omega)$ и $\eta(\omega)$ входят совершенно равноправным образом, то $\{\omega : \xi(\omega) > \eta(\omega)\} \in \mathcal{F}$. Далее заметим, что:

$$\{\omega : \xi(\omega) \leq \eta(\omega)\} = \Omega \setminus \{\omega : \xi(\omega) > \eta(\omega)\} \in \mathcal{F},$$

$$\{\omega : \xi(\omega) = \eta(\omega)\} = \{\omega : \xi(\omega) \leq \eta(\omega)\} \cap \{\omega : \xi(\omega) \geq \eta(\omega)\} \in \mathcal{F}.$$

В качестве следствия из теоремы 3 получаем следующее утверждение:

Теорема 4. Пусть (Ω, \mathcal{F}, P) — вероятностное пространство, а $\xi(\omega)$ и $\eta(\omega)$ — случайные величины на нем. Тогда

$$c \cdot \xi(\omega), \xi(\omega) + c, |\xi(\omega)|, \xi(\omega) \pm \eta(\omega),$$

$\xi(\omega) \cdot \eta(\omega), \frac{\xi(\omega)}{\eta(\omega)}$ — также случайные величины (в последнем случае предполагается, что $P(\omega : \eta(\omega) \neq 0) = 1$).

Для доказательства теоремы достаточно заметить, что

$$\{\omega : c \cdot \xi(\omega) < x\} = \begin{cases} \left\{ \omega : \xi(\omega) < \frac{x}{c} \right\}, & c > 0; \\ \left\{ \omega : \xi(\omega) > \frac{x}{c} \right\}, & c < 0. \end{cases}$$

ПРИМЕРЫ

1. Пусть (Ω, \mathcal{F}, P) — произвольное вероятностное пространство и $\xi(\omega) = c$ для каждого ω из Ω . Тогда $\xi(\omega)$ — случайная величина.

В самом деле,

$$\{\omega : \xi(\omega) < x\} = \begin{cases} \emptyset, & x \leq c; \\ \Omega, & x > c. \end{cases}$$

и при каждом действительном x $\{\omega : \xi(\omega) < x\} \in \mathcal{F}$.

2. Пусть Ω — некоторое пространство, а \mathcal{F} — система всех подмножеств Ω . Тогда каждая функция $\xi(\omega)$ на Ω является случайной величиной.

Если $\mathcal{F} = \{\emptyset, \Omega\}$ (это самая «тощая» σ -алгебра подмножеств в Ω), то случайными величинами могут быть только константы.

Предлагаем читателю убедиться в этом самостоятельно.

Докажем следующее утверждение:

Теорема 5. Пусть $\xi_1(\omega), \dots, \xi_n(\omega)$ — случайные величины на вероятностном пространстве (Ω, \mathcal{F}, P) , а B — борелевское множество в n -мерном пространстве. Тогда

$$\{\omega : [\xi_1(\omega), \dots, \xi_n(\omega)] \in B\} \in \mathcal{F}.$$

Доказательство. Рассмотрим сначала в качестве множества параллелепипед

$$B = \prod_{k=1}^n [a_k, b_k].$$

Тогда

$$\{\omega : [\xi_1(\omega), \dots, \xi_n(\omega)] \in B\} = \{\omega : a_1 \leq \xi_1(\omega) < b_1, \dots,$$

$$a_n \leq \xi_n(\omega) < b_n\} = \bigcap_{k=1}^n \{\omega : a_k \leq \xi_k(\omega) < b_k\} \in \mathcal{F},$$

так как по теореме 1 множества $\{\omega : a_k \leq \xi_k(\omega) < b_k\}$, $k = \overline{1, n}$, являются случайными событиями.

Пусть K — класс тех подмножеств B n -мерного пространства R^n , для которых

$$\{\omega : [\xi_1(\omega), \dots, \xi_n(\omega)] \in B\}.$$

Рассмотрим отображение

$$\xi : \Omega \rightarrow R^n, \xi(\omega) = [\xi_1(\omega), \dots, \xi_n(\omega)].$$

Тогда из соотношений между образами, приведенными в доказательстве теоремы 2, следует, что класс K есть σ -алгебра. Но мы уже убедились, что класс K содержит все параллелепипеды. Следовательно, классу K принадлежит и наименьшая σ -алгебра, содержащая все параллелепипеды, т. е. K содержит все борелевские множества. Теорема доказана.

В дальнейшем мы будем рассматривать и случайные события, связанные с бесконечными последовательностями случайных величин. Имеет место, например, следующее утверждение:

Теорема 6. Пусть $\xi_n(\omega)$ — последовательность случайных величин, заданных на вероятностном пространстве (Ω, \mathcal{F}, P) , и $\xi(\omega)$ — случайная величина на (Ω, \mathcal{F}, P) . Тогда $\{\omega: \lim_{n \rightarrow \infty} \xi_n(\omega) \text{ существует}\} \in \mathcal{F}$,

$$\{\omega: \lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)\} \in \mathcal{F}.$$

Доказательство. Пусть

$$\omega \in \{\omega: \lim_{n \rightarrow \infty} \xi_n(\omega) \text{ существует}\}.$$

Тогда для каждого натурального k существует такое $N(k)$, что для всех n и m , $n > N$, $m > N$, $|\xi_n(\omega) - \xi_m(\omega)| < \frac{1}{k}$. Следовательно,

$$\omega \in \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{\substack{m > N \\ n > N}} \left\{ \omega: |\xi_n(\omega) - \xi_m(\omega)| < \frac{1}{k} \right\}.$$

Наоборот, если

$$\omega \in \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{\substack{m > N \\ n > N}} \left\{ \omega: |\xi_n(\omega) - \xi_m(\omega)| < \frac{1}{k} \right\},$$

то $\lim_{n \rightarrow \infty} \xi_n(\omega)$ существует.

Таким образом,

$$\begin{aligned} & \{\omega: \lim_{n \rightarrow \infty} \xi_n(\omega) \text{ существует}\} = \\ & = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{\substack{m > N \\ n > N}} \left\{ \omega: |\xi_n(\omega) - \xi_m(\omega)| < \frac{1}{k} \right\}. \end{aligned}$$

Из теоремы 3 следует, что

$$\left\{ \omega: |\xi_n(\omega) - \xi_m(\omega)| < \frac{1}{k} \right\} \in \mathcal{F}.$$

Поэтому

$$\{\omega: \lim_{n \rightarrow \infty} \xi_n(\omega) \text{ существует}\} \in \mathcal{F}.$$

Это означает, что можно рассматривать вероятность

$$P \{ \omega: \lim_{n \rightarrow \infty} \xi_n(\omega) \text{ существует} \}.$$

Аналогично доказывается, что

$$\{\omega: \lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)\} = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n > N} \left\{ \omega: |\xi_n(\omega) - \xi(\omega)| < \frac{1}{k} \right\} \in \mathcal{F}.$$

В дальнейшем будем рассматривать без специальных разъяснений вероятности

$$P\{\lim \xi_n \text{ существует}\}, P\{\lim \xi_n = \xi\},$$

не указывая зависимость случайных величин от ω .

Теорема 7. Если $\{\xi_n(\omega)\}$ — последовательность случайных величин на вероятностном пространстве (Ω, \mathcal{F}, P) , то функции

$$\sup_n \xi_n(\omega), \quad \inf_n \xi_n(\omega), \\ \overline{\lim} \xi_n(\omega), \quad \underline{\lim} \xi_n(\omega)$$

являются случайными величинами.

Доказательства вытекают из следующих соотношений:

$$\{\omega : \sup_n \xi_n(\omega) < x\} = \Omega \setminus \{\omega : \sup_n \xi_n(\omega) \geq x\} =$$

$$= \Omega \setminus \bigcup_{n=1}^{\infty} \{\omega : \xi_n(\omega) \geq x\},$$

$$\{\omega : \inf_n \xi_n(\omega) < x\} = \bigcup_{n=1}^{\infty} \{\omega : \xi_n(\omega) < x\},$$

$$\{\omega : \overline{\lim} \xi_n(\omega) < x\} = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcap_{j=n}^{\infty} \left\{ \omega : \xi_j(\omega) < x - \frac{1}{k} \right\},$$

$$\{\omega : \underline{\lim} \xi_n(\omega) > x\} = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} \left\{ \omega : \xi_j(\omega) > x + \frac{1}{k} \right\}.$$

Борелевские функции. Функции от случайных величин.

Пусть $f(x)$ — функция на m -мерном пространстве R^m со значениями в R^1 , \mathcal{B} есть σ -алгебра борелевских множеств в R^m , т. е. наименьшая σ -алгебра, содержащая все «полузамкнутые» параллелепипеды $\prod_{k=1}^m [a_k, b_k]$.

Определение 3. Функция $f(x) : R^m \rightarrow R^1$, измеримая относительно σ -алгебры \mathcal{B} , называется борелевской функцией.

Класс таких функций достаточно богат, он содержит, в частности, все непрерывные функции. В самом деле, если $f(x)$ — непрерывная функция, то множество $\{x : f(x) < c\}$ является открытым, следовательно, принадлежит σ -алгебре \mathcal{B} .

Если $f_n(x)$ — непрерывные функции на R^m и $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ для каждого $x \in R^m$, то $f(x)$ — борелевская функция (см. теорему 7).

Часто приходится рассматривать различные функции от случайных величин и при этом, конечно, возникает вопрос, являются ли эти функции случайными величинами. Имеет место следующее утверждение.

Теорема 8. Пусть $f(x)$ — борелевская функция на R^m , а $\xi_1(\omega), \dots, \xi_m(\omega)$ — случайные величины на вероятностном пространстве (Ω, \mathcal{F}, P) . Тогда $f(\xi_1(\omega), \dots, \xi_m(\omega))$ — случайная величина.

Доказательство. Поскольку функция $f(x): R^m \rightarrow R^1$ борелевская, то для каждого $c \in R^1$

$$\{\omega: f(x) < c\} = f^{-1}((-\infty, c)) \in \mathfrak{B}.$$

Следовательно, по теореме 2

$$\begin{aligned} \{\omega: f(\xi_1(\omega), \dots, \xi_m(\omega)) < c\} = \\ = \{\omega: [\xi_1(\omega), \dots, \xi_m(\omega)] \in f^{-1}((-\infty, c))\} \in \mathcal{F}. \end{aligned}$$

Заметим, что функции

$$f(x) = |x| \quad (x \in R^1);$$

$$f(x) = \sin x \quad (x \in R^1);$$

$$f(x) = \begin{cases} x, & x > 0; \\ 0, & x \leq 0; \end{cases}$$

$$f(x) = \begin{cases} 0, & x > 0; \\ -x, & x \leq 0; \end{cases}$$

$$f(x) = x_1 + \dots + x_m \quad (x \in R^m);$$

$$f(x) = |x| = \sqrt{x_1^2 + \dots + x_m^2} \quad (x \in R^m);$$

$$f(x) = x_1 \dots x_m \quad (x \in R^m)$$

являются борелевскими функциями (все эти функции непрерывны). Поэтому из теоремы 8 следует, что если $\xi(\omega), \xi_1(\omega), \dots, \xi_m(\omega)$ — случайные величины на вероятностном пространстве (Ω, \mathcal{F}, P) , то

$$\eta(\omega) = |\xi(\omega)|,$$

$$\eta(\omega) = \sin \xi(\omega),$$

$$\eta(\omega) = \begin{cases} \xi(\omega), & \xi(\omega) > 0 \\ 0, & \xi(\omega) \leq 0; \end{cases}$$

$$\eta(\omega) = \begin{cases} 0, & \xi(\omega) > 0, \\ -\xi(\omega), & \xi(\omega) \leq 0; \end{cases}$$

$$\eta(\omega) = \xi_1(\omega) + \dots + \xi_m(\omega);$$

$$\eta(\omega) = \sqrt{\xi_1^2(\omega) + \dots + \xi_m^2(\omega)};$$

$$\eta(\omega) = \xi_1(\omega) \dots \xi_m(\omega)$$

также являются случайными величинами на этом вероятностном пространстве.

§ 2. Распределения случайных величин

Дискретные случайные величины. Пусть (Ω, \mathcal{F}, P) — вероятностное пространство. Случайная величина $\xi(\omega)$ называется дискретной случайной величиной, если она принимает конечное или счетное число значений.

Теорема 1. Пусть $\xi(\omega): \Omega \rightarrow R$ — функция, принимающая конечное или счетное число значений x_1, \dots, x_n, \dots . Функция $\xi(\omega)$ измерима относительно \mathcal{F} тогда и только тогда, когда для каждого n

$$\{\omega: \xi(\omega) = x_n\} \in \mathcal{F}. \quad (1)$$

Действительно, если для функции $\xi(\omega)$ имеет место соотношение (1), то эта функция измерима относительно \mathcal{F} , так как для каждого вещественного x

$$\{\omega : \xi(\omega) < x\} = \bigcup_{ix_i < x} \{\omega : \xi(\omega) = x_i\} \in \mathcal{F}.$$

Кроме того, если $\xi(\omega)$ измерима относительно σ -алгебры \mathcal{F} , то по теореме 1 предыдущего параграфа для каждого вещественного x $\{\omega : \xi(\omega) = x\} \in \mathcal{F}$.

Таким образом, если $\xi(\omega)$ — дискретная случайная величина на вероятностном пространстве (Ω, \mathcal{F}, P) , принимающая значения x_1, \dots, x_n, \dots , то для каждого n определена вероятность

$$p_n = P\{\omega : \xi(\omega) = x_n\}. \quad (2)$$

Набор вероятностей (2) называется распределением случайной величины $\xi(\omega)$.

Распределение дискретной случайной величины удобно характеризовать с помощью следующей таблицы:

Значения $\xi(\omega)$	x_1	x_2	\dots	x_n	\dots
Вероятности	p_1	p_2	\dots	p_n	\dots

Заметим, что $p_n \geq 0$, $\sum_{n=1}^{\infty} p_n = 1$.

Рассмотрим некоторые наиболее важные примеры дискретных случайных величин.

Биномиальное распределение. Предположим, что производятся независимые испытания и при каждом испытании может быть два исхода — успех с вероятностью p или неудача с вероятностью q ($p + q = 1$). Например, стрельба по цели (при каждом выстреле два исхода — попадание или непопадание); проверка наугад выбранного изделия, которое может оказаться качественным или бракованным; подбрасывание симметричной монеты (при каждом подбрасывании появится герб или решка). Термины «успех» и «неудача» мы употребляем для удобства, важно только, чтобы при каждом испытании было два исхода.

Предположим, что произведено n испытаний, пусть ξ — число успехов при n испытаниях.

Построим вероятностную модель эксперимента, состоящего в последовательном осуществлении n независимых испытаний с двумя исходами.

Пусть, например, $n = 2$. Тогда пространством элементарных событий является множество $\Omega = \{УУ; УН; НУ; НН\}$. Поскольку испытания независимы, то вероятности элементарных исходов следует определить так: $УУ - p^2$; $УН - pq$; $НУ - pq$; $НН - q^2$.

В общем случае пространство элементарных событий состоит из 2^n всевозможных наборов ω длины n , состоящих из букв У и Н. Принимая во внимание независимость испытаний, вероятность одного отдельного набора ω положим равной $p^k q^{n-k}$, если в набор ω входит k букв У.

Случайная величина $\xi = \xi(\omega)$ — число успехов при n испытаниях — равна k , если набор ω состоит из k букв У. Так как число наборов ω длины n , содержащих ровно k букв У, равно C_n^k , то

$$P_n(k) = P\{\omega : \xi(\omega) = k\} = C_n^k p^k q^{n-k} \quad (k = 0, 1, \dots, n). \quad (3)$$

Набор вероятностей (3) называется **биномиальным распределением**. Найдем k , при котором вероятность $P_n(k)$ максимальна, то есть вычислим наиболее вероятное число успехов. Заметим, что

$$\frac{P_n(k)}{P_n(k-1)} = \frac{n-k+1}{k} \cdot \frac{p}{q} = 1 + \frac{(n-k+1)p-kq}{kq} = 1 + \frac{(n+1)p-k}{kq}.$$

Поэтому, если $k < (n+1)p$, то $P_n(k) > P_n(k-1)$ (с возрастанием k вероятности $P_n(k)$ возрастают); если же $k > (n+1)p$, то $P_n(k) < P_n(k-1)$ (с возрастанием k вероятности $P_n(k)$ убывают). Пусть $m = [(n+1)p]$ — наибольшее целое число, не превосходящее $(n+1)p$. Тогда $P_n(k)$ принимает наибольшее значение при $k = m$. Если же $(n+1)p$ — целое число, то наиболее вероятных значений k два: $m = (n+1)p$ и $m-1$.

Сформулируем полученный результат.

Теорема 2. Пусть $(n+1)p$ — не целое число. Тогда с изменением k от 0 до n вероятность $P_n(k)$ сначала монотонно возрастает, а затем монотонно убывает, достигая наибольшего значения при $k = m = [(n+1)p]$. Если $(n+1)p = m$ — целое число, то $P_n(m-1) = P_n(m)$ и при $k < m-1$ вероятность $P_n(k)$ монотонно возрастает, а при $k > m$ монотонно убывает.

Геометрическое распределение. Предположим, что производятся независимые испытания, причем при каждом испытании может быть два исхода — успех с вероятностью p или неудача с вероятностью q . Пусть испытания проводятся до первого появления успеха. Обозначим через ξ — число испытаний до первого появления успеха. Найдем распределение случайной величины ξ .

В качестве пространства элементарных событий можно рассматривать множество

$$\Omega = \{U; NU; ННУ; НННУ; \dots; \underbrace{НН \dots НУ}_{n-1}; \dots\}$$

Согласно предположению независимости испытаний $P\{НН \dots НУ\} = q^{n-1}p$. Таким образом,

$$P\{\xi(\omega) = n\} = q^n p \quad (n = 0, 1, \dots). \quad (4)$$

Набор вероятностей (4) называется геометрическим распределением.

Пусть ξ имеет геометрическое распределение. Тогда

$$P\{\xi = n + m/\xi \geq n\} = P\{\xi = m\}. \quad (5)$$

$(m \geq 1)$

Действительно,

$$\begin{aligned} P\{\xi = n + m/\xi \geq n\} &= \frac{P\{\xi = n + m, \xi \geq n\}}{P\{\xi \geq n\}} = \\ &= \frac{P\{\xi = n + m\}}{P\{\xi \geq n\}} = \frac{q^{n+m}p}{\sum_{k=n}^{\infty} q^k p} = pq^m. \end{aligned}$$

Равенство (5) можно интерпретировать следующим образом. Рассмотрим телефонные разговоры. Предположим, что их длина измеряется целым числом минут и в начале каждой минуты с вероятностью p принимается решение закончить разговор и с вероятностью $q = 1 - p$ продолжить разговор. Тогда длительность телефонного разговора будет случайной величиной, имеющей геометрическое распределение. Равенство (5) означает, что условная вероятность того, что телефонный разговор будет продолжаться ровно $n + m$ минут, если известно, что разговор не закончился за $n - 1$ минуту, совпадает с безусловной вероятностью того, что разговор будет продолжаться ровно m минут. Свойство, выражаемое равенством (5), называется отсутствием последствия.

Интересно отметить, что среди всех дискретных распределений свойством отсутствия последствия обладает только геометрическое распределение.

Действительно, пусть $\{p_m\}$ — распределение случайной величины, принимающей значения $0, 1, \dots$. Предположим, что это распределение обладает свойством (5), т. е. для всех $n \geq 1$ и для всех $m > 0$

$$\frac{p_{n+m}}{\sum_{k=n}^{\infty} p_k} = p_m. \quad (6)$$

Полагая в равенстве (6) $n = 1$, имеем

$$p_{m+1} = p_m \sum_{k=1}^{\infty} p_k = p_m (1 - p_0).$$

Отсюда следует, что $p_m = (1 - p_0)^m p_0$, следовательно, распределение $\{p_m\}$ является геометрическим.

Гипергеометрическое распределение. Предположим, что в урне N шаров, среди них n белых и $N - n$ черных. Наугад из урны взято k шаров. Пусть ξ — число белых шаров среди k взятых. Найдем распределение случайной величины ξ . Число всевозможных

партий, содержащих ровно k шаров, равно C_n^k ; среди них имеется $C_n^r C_{N-n}^{k-r}$ партий, содержащих ровно r белых шаров. Поэтому

$$P\{\xi = r\} = \frac{C_n^r C_{N-n}^{k-r}}{C_N^k} \quad (7)$$

$$(0 \leq r \leq \min(n, k)).$$

Набор вероятностей (7) называется гипергеометрическим распределением. Это распределение часто используется при решении некоторых задач.

ЗАДАЧИ

1. Из партии, содержащей N изделий, среди которых n бракованных, взято наугад k изделий. Какова вероятность того, что среди них не больше s бракованных ($s \leq \min(n, k)$)?

Решение. Искомая вероятность, согласно формуле (7), равна

$$\frac{\sum_{r=0}^s C_n^r C_{N-n}^{k-r}}{C_N^k}.$$

2. В лотерее типа «спортлото» имеется N видов спорта. Участник лотереи называет k видов спорта; если при этом окажется, что не менее s видов спорта совпадает с k видами, отмеченными ранее, но неизвестными участнику, то участник лотереи выигрывает. Какова вероятность выигрыша?

Решение. По формуле (7) вероятность выигрыша равна

$$\frac{\sum_{r=s}^k C_n^r C_{N-n}^{k-r}}{C_N^k}.$$

Заметим, что в обычной лотерее «спортлото» $N = 49$, $k = 6$, $s = 3$.

Распределение Пуассона. Случайная величина ξ имеет распределение Пуассона с параметром λ ($\lambda > 0$), если она принимает значения $0, 1, 2, 3, \dots$, причем

$$P\{\xi = n\} = \frac{\lambda^n e^{-\lambda}}{n!}. \quad (8)$$

Распределение Пуассона играет важную роль в теории случайных процессов, теории массового обслуживания, теории надежности (см. § 7).

Функции распределения случайных величин и их свойства. Перейдем теперь к изучению произвольных случайных величин. Пусть (Ω, \mathcal{F}, P) — вероятностное пространство, а $\xi(\omega)$ — случайная величина на этом пространстве. В § 1 мы назвали функцией распределения случайной величины $\xi(\omega)$ вероятность случайного события $\{\omega : \xi(\omega) < x\}$:

$$F(x) = P\{\omega : \xi(\omega) < x\}.$$

Отметим следующие важные свойства функций распределения:
 а) Если $a < b$, то

$$P\{\omega : a \leq \xi(\omega) < b\} = F(b) - F(a). \quad (9)$$

Действительно, если $a < b$, то $\{\omega : a \leq \xi(\omega) < b\} = \{\omega : \xi(\omega) < b\} \setminus \{\omega : \xi(\omega) < a\}$, причем $\{\omega : \xi(\omega) < a\} \subset \{\omega : \xi(\omega) < b\}$. Отсюда и следует (9).

б) Если $a < b$, то $F(a) \leq F(b)$, т. е. $F(x)$ — неубывающая функция.

Действительно, если $a < b$, то

$$F(b) - F(a) = P\{\omega : a \leq \xi(\omega) < b\} \geq 0.$$

$$\text{в) } \lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

Достаточно доказать, что для любых последовательностей $\{x_n\}$ и $\{y_n\}$ таких, что

$$x_{n+1} < x_n, \quad \lim_{n \rightarrow +\infty} x_n = -\infty, \\ y_{n+1} > y_n, \quad \lim_{n \rightarrow +\infty} y_n = +\infty,$$

имеют место равенства:

$$\lim_{n \rightarrow \infty} F(x_n) = 0, \quad \lim_{n \rightarrow \infty} F(y_n) = 1.$$

Заметим, что

$$\bigcap_{n=1}^{\infty} \{\omega : \xi(\omega) < x_n\} = \emptyset$$

и

$$\{\omega : \xi(\omega) < x_{n+1}\} \subset \{\omega : \xi(\omega) < x_n\}.$$

Поэтому, согласно свойству непрерывности вероятности,

$$0 = P\left(\bigcap_{n=1}^{\infty} \{\omega : \xi(\omega) < x_n\}\right) = \lim_{n \rightarrow \infty} P\{\omega : \xi(\omega) < x_n\} = \lim_{n \rightarrow \infty} F(x_n).$$

Далее отметим также, что

$$\bigcup_{n=1}^{\infty} \{\omega : \xi(\omega) < y_n\} = \Omega$$

и

$$\{\omega : \xi(\omega) < y_n\} \subset \{\omega : \xi(\omega) < y_{n+1}\}.$$

Поэтому по свойству непрерывности вероятности

$$1 = P(\Omega) = P\left(\bigcup_{n=1}^{\infty} \{\omega : \xi(\omega) < y_n\}\right) = \\ = \lim_{n \rightarrow \infty} P\{\omega : \xi(\omega) < y_n\} = \lim_{n \rightarrow \infty} F(y_n).$$

2) $F(x)$ непрерывна слева.

Пусть $\{x_n\}$ — произвольная последовательность такая, что

$$x_n < x_{n+1}, \quad x_n \rightarrow x, \quad x_n < x.$$

Докажем, что

$$\lim_{n \rightarrow \infty} F(x_n) = F(x).$$

Имеет место равенство

$$\{\omega : \xi(\omega) < x\} = \bigcup_{n=1}^{\infty} \{\omega : \xi(\omega) < x_n\}.$$

Так как

$$\{\omega : \xi(\omega) < x_n\} \subset \{\omega : \xi(\omega) < x_{n+1}\},$$

то по свойству непрерывности вероятности

$$\begin{aligned} F(x) &= P\{\omega : \xi(\omega) < x\} = P\left\{\bigcup_{n=1}^{\infty} \omega : \xi(\omega) < x_n\right\} = \\ &= \lim_{n \rightarrow \infty} P\{\omega : \xi(\omega) < x_n\} = \lim_{n \rightarrow \infty} F(x_n). \end{aligned}$$

$$д) P\{\omega : \xi(\omega) \leq x\} = F(x+0).$$

По определению $F(x+0) = \lim_{n \rightarrow \infty} F(x_n)$, где $\{x_n\}$ — последовательность такая, что $x_{n+1} < x_n$, $x_n \rightarrow x$, $x_n > x$. Поскольку

$$\{\omega : \xi(\omega) \leq x\} = \bigcap_{k=1}^{\infty} \{\omega : \xi(\omega) < x_k\}$$

и $\{\omega : \xi(\omega) < x_n\}$ — монотонно убывающая последовательность событий, то, согласно свойству непрерывности вероятности,

$$\begin{aligned} P\{\omega : \xi(\omega) \leq x\} &= P\left\{\bigcap_{n=1}^{\infty} \{\omega : \xi(\omega) < x_n\}\right\} = \\ &= \lim_{n \rightarrow \infty} P\{\omega : \xi(\omega) < x_n\} = \lim_{n \rightarrow \infty} F(x_n) = F(x+0). \end{aligned}$$

$$е) P\{\omega : \xi(\omega) = x\} = F(x+0) - F(x).$$

Так как

$$\{\omega : \xi(\omega) = x\} = \{\omega : \xi(\omega) \leq x\} \setminus \{\omega : \xi(\omega) < x\}$$

и

$$\{\omega : \xi(\omega) < x\} \subset \{\omega : \xi(\omega) \leq x\},$$

то

$$\begin{aligned} P\{\omega : \xi(\omega) = x\} &= P\{\omega : \xi(\omega) \leq x\} - \\ &- P\{\omega : \xi(\omega) < x\} = F(x+0) - F(x). \end{aligned}$$

Докажем следующую теорему:

Теорема 3. Пусть $F(x)$ обладает следующими свойствами:

- 1) $F(x)$ неубывающая функция на $(-\infty, +\infty)$;
- 2) $F(x)$ непрерывна слева;
- 3) $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$.

Тогда существует вероятностное пространство (Ω, \mathcal{F}, P) и случайная величина $\xi(\omega)$ на нем такая, что функция распределения $\xi(\omega)$ равна $F(x)$.

Доказательство. Построим требуемое вероятностное пространство. Пусть $\Omega = (-\infty, +\infty)$, $\mathcal{F} = \mathcal{B}$ — σ -алгебра борелевских множеств на числовой прямой, \mathcal{A} — класс всех интервалов вида $[a, b]$ и конечных объединений таких интервалов (допускается, в частности, $a = -\infty$, $b = +\infty$). Класс \mathcal{A} — алгебра множеств.

Определим на классе \mathcal{A} функцию множества $P(\cdot)$ следующим образом: если $A \in \mathcal{A}$, то

$$A = \bigcup_{i=1}^n [a_i, b_i],$$

где интервалы $[a_i, b_i]$ между собой попарно не пересекаются; положим

$$P(A) = \sum_{i=1}^n [F(b_i) - F(a_i)].$$

Из предположений 1) и 2) теоремы следует, что $P(A) \geq 0$ для каждого множества. Кроме того, по определению $P(A)$ — конечноаддитивная функция множеств на \mathcal{A} и $P(\Omega) = 1$.

Из предположения 3) вытекает, что $P(\cdot)$ — счетно-аддитивная функция множеств на \mathcal{A} . Согласно теореме 5 (§ 3, гл. I), достаточно проверить, что функция множества $P(A)$ на алгебре \mathcal{A} обладает свойством непрерывности: для любой последовательности множеств такой, что $A_n \in \mathcal{A}$, $A_{n+1} \subset A_n$,

$$\bigcap_{n=1}^{\infty} A_n = A \in \mathcal{A},$$

выполнено равенство

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

Не ограничивая общности, можно считать, что $A = [a, b]$. Так как $A \subset A_n$ для всех n , то A_n содержит интервал $[a_n, b_n]$ такой, что $a_n < a$, $b_n > b$. Пусть это будет максимальный интервал в A_n , содержащий A . Нетрудно видеть, что, начиная с некоторого n ,

$$A_n = [a_n, b_n] = [a, b].$$

Тогда

$$\lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} [F(b) - F(a_n)] = F(b) - F(a) = P(A).$$

Следовательно, $P(A)$ — счетно-аддитивная мера на \mathcal{A} . По теореме 6 о продолжении меры (§ 3, гл. I) меру $P(\cdot)$ можно продолжить на наименьшую σ -алгебру $\sigma(\mathcal{A})$, содержащую класс \mathcal{A} , а это есть σ -алгебра \mathcal{B} борелевских множеств. Итак, мы построили вероятностное пространство (Ω, \mathcal{F}, P) ($\Omega = R^1$, $\mathcal{F} = \mathcal{B}$, P). Рассмотрим функцию $\xi(\omega) = \omega$. Тогда

$$P\{\omega : \xi \leq x\} = P\{(-\infty, x]\} = F(x).$$

Следовательно, $\xi(\omega) = \omega$ — случайная величина с функцией распределения $F(x)$.

Плотность распределения случайной величины. Пусть ξ — случайная величина с функцией распределения $F(x)$. Говорят, что случайная величина имеет плотность распределения, если существ-

вует интегрируемая борелевская функция $p(u)$ такая, что для всех x выполнено равенство

$$F(x) = \int_{-\infty}^x p(u) du. \quad (10)$$

Функция $p(x)$ называется плотностью распределения случайной величины ξ .

Если функция $p(u)$ является непрерывной функцией, то из курса математического анализа известно, что функция $F(x)$, определяемая равенством (10), дифференцируема и $F'(x) = p(x)$. Заметим также, что $F'(x) \geq 0$, так как $F(x)$ — неубывающая функция.

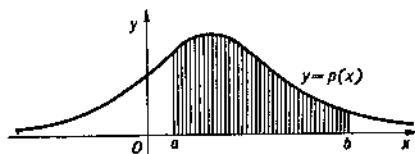


Рис. 11

Следовательно, $p(x)$ — неотрицательная функция. В общем же случае из (10) следует, что равенство $F'(x) = p(x)$ выполнено почти всюду относительно меры Лебега.

Далее, если случайная величина ξ имеет плотность распределения $p(x)$, то

$$P\{a \leq \xi < b\} = \int_a^b p(u) du. \quad (11)$$

Действительно, как было показано выше,

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

Отсюда и из равенства (10) получаем равенство (11), которое показывает, что вероятность попадания случайной величины ξ в отрезок $[a, b]$ равна площади, заштрихованной на рис. 11. В случае непрерывной плотности распределения из равенства (11) имеем

$$P\{x \leq \xi < x + \Delta x\} = p(x) \Delta x + o(\Delta x). \quad (12)$$

Поскольку $F(+\infty) = 1$, то учитывая равенство (10), получаем

$$\int_{-\infty}^{+\infty} p(u) du = 1. \quad (13)$$

Подытоживая все сказанное выше, отметим, что плотность распределения — неотрицательная функция, удовлетворяющая равенству (13), и, наоборот, каждая неотрицательная функция $p(x)$, обладающая свойством (13), является плотностью распределения

некоторой случайной величины. Для доказательства достаточно заметить, что если $p(x)$ — такая функция, то функция $F(x)$, определенная равенством (10), удовлетворяет всем условиям теоремы 3.

Теорема 4. Если случайная величина ξ имеет плотность распределения $p(x)$, то для любого борелевского множества B имеет место равенство

$$P\{\omega : \xi(\omega) \in B\} = \int_B p(x) dx. \quad (14)$$

Доказательство. Рассмотрим на σ -алгебре \mathfrak{B} борелевских множеств две вероятностные меры

$$\mu_1(B) = P\{\omega : \xi(\omega) \in B\}$$

и

$$\mu_2(B) = \int_B p(x) dx.$$

Согласно равенству (11), $\mu_1(B) = \mu_2(B)$, если $B = [a, b]$. Эти меры совпадают и на алгебре множеств \mathcal{A} , введенной при доказательстве теоремы 3. Следовательно, они совпадают и на σ -алгебре, являющейся наименьшей σ -алгеброй, содержащей алгебру \mathcal{A} .

Три типа функций распределения. Теорема Лебега о разложении.

Точку x называют точкой роста функции распределения $F(x)$, если для каждого $\varepsilon > 0$ выполняется неравенство

$$F(x + \varepsilon) - F(x - \varepsilon) > 0.$$

Например, функция распределения, рассмотренная в примере 1, с. 51, имеет две точки роста $x = -1$ и $x = 1$, а в следующем примере множество точек роста функции распределения — отрезок $[a, b]$.

Сформулируем теорему, принадлежащую А. Лебегу.

Теорема 5. Каждая функция распределения $F(x)$ единственным образом может быть представлена в виде суммы

$$F(x) = F_d(x) + F_{an}(x) + F_c(x)$$

трех неубывающих функций $F_d(x)$, $F_{an}(x)$, $F_c(x)$, где:

1) $F_d(x)$ — дискретная компонента, представимая в виде

$$F_d(x) = \sum_{x_i \leq x} p(x_i), \quad p(x_i) \geq 0, \quad \sum_i p(x_i) \leq 1;$$

2) $F_{an}(x)$ — абсолютно непрерывная компонента, имеющая вид

$$F_{an}(x) = \int_{-\infty}^x p(u) du,$$

$$p(u) \geq 0, \quad \int_{-\infty}^{\infty} p(u) du \leq 1;$$

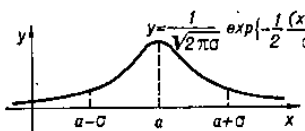
3) $F_c(x)$ — сингулярная компонента, представляющая собой непрерывную функцию, множество точек роста которой имеет лебегову меру нуль.

К сингулярным распределениям принадлежит, например, известная кривая Кантора¹.

Рассмотрим некоторые наиболее важные распределения случайных величин.

Равномерное распределение. Равномерным распределением на отрезке $[a, b]$ называется распределение с плотностью распределения

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b); \\ 0 & x \notin (a, b) \end{cases} \quad (15)$$



(см. пример 2, с. 51). График функции распределения изображен на рис. 10.

Рис. 12

Нормальное распределение. Исключительно важную роль в теории вероятностей играет нормальное распределение.

Нормальным распределением $N(a, \sigma^2)$ с параметрами a и σ^2 называется распределение с плотностью распределения

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-a)^2}{\sigma^2}}. \quad (16)$$

График функции $p(x)$ изображен на рис. 12. Заметим, что $p(x)$ действительно является плотностью распределения, так как

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \frac{(x-a)^2}{\sigma^2}} dx = 1.$$

Чтобы убедиться в этом, достаточно положить $z = \frac{x-a}{\sigma}$ и воспользоваться известным из курса математического анализа интегралом Пуассона

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

Нормальное распределение часто называют еще законом Гаусса.

Важная роль этого распределения объясняется тем, что при весьма широких предположениях распределение суммы большого

¹ Процесс построения такой кривой описан в книге: Б. В. Гнеденко. Курс теории вероятностей. М., Наука, 1965, с. 137.

числа случайных величин оказывается близким к нормальному распределению.

Показательное распределение. Пусть рассматривается некоторый прибор, начавший работать в момент времени $t = 0$. Предположим, что условная вероятность выхода прибора из строя в интервале времени $(t, t + \Delta t)$ при условии, что прибор не выйдет из строя до момента времени t , равна $\lambda \Delta t + o(\Delta t)$. Пусть ξ — продолжительность безотказной работы прибора. Найдем функцию распределения случайной величины ξ .

Положим $Q(t) = P\{\xi \geq t\}$.

Тогда

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

Отсюда

$$\frac{Q(t + \Delta t) - Q(t)}{\Delta t} = -\lambda Q(t) + Q(t) \frac{o(\Delta t)}{\Delta t}.$$

Переходя в этом равенстве к пределу при $\Delta t \rightarrow 0$, получим

$$\frac{dQ}{dt} = -\lambda Q(t).$$

Интегрируя это дифференциальное уравнение и принимая во внимание условие $Q(0) = 1$, получим, что $Q(t) = e^{-\lambda t}$. Следовательно, функция распределения времени ξ безотказной работы прибора равна

$$F(t) = P\{\xi < t\} = 1 - Q(t) = \begin{cases} 1 - e^{-\lambda t}, & t \geq 0; \\ 0, & t \leq 0. \end{cases}$$

Определение. Случайная величина ξ имеет показательное распределение с параметром λ , если функция распределения ξ равна

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0; \\ 0, & x \leq 0. \end{cases} \quad (17)$$

Плотность распределения такой случайной величины равна

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0; \\ 0, & x \leq 0. \end{cases}$$

Заметим, что показательное распределение обладает свойством отсутствия последствия. Это свойство выражается следующим равенством:

$$P\{t < \xi < s/\xi > t\} = P\{\xi < s - t\}, \quad (18) \\ t < s.$$

Действительно,

$$P\{t < \xi < s/\xi > t\} = \frac{F(s) - F(t)}{1 - F(t)} = 1 - e^{-\lambda(s-t)} = F(s-t).$$

Интересно отметить, что свойство отсутствия последствия является характеристическим для показательного распределения.

Среди всех распределений с непрерывной функцией распределения свойством (18), отсутствия последствия, обладает только показательное распределение. Положим $1 - F(t) = Q(t)$. Тогда из (18) получаем

$$\frac{Q(t) - Q(s)}{Q(t)} = 1 - Q(s - t)$$

при $s > t$. Отсюда:

$$\begin{aligned} Q(t) - Q(s) &= Q(t) [1 - Q(s - t)], \\ Q(s) &= Q(t) Q(s - t). \end{aligned}$$

Пусть $s - t = u$. Тогда

$$Q(t + u) = Q(t) Q(u). \quad (19)$$

Функция $Q(t)$ является непрерывной и ограниченной функцией. Легко убедиться, что все непрерывные ограниченные решения функционального уравнения (19) имеют вид $Q(t) = e^{-\lambda t}$, где $\lambda > 0$.

Распределение Коши. Рассмотрим следующую задачу. Предположим, что построена окружность единичного радиуса с центром в начале координат. На этой окружности наугад выбирают точку. В выбранной точке проводят касательную к окружности. Найдем функцию распределения длины этой касательной до ее пересечения с осью Ox .

Пусть ξ — длина касательной. Построим точки M, M_1, M_2, M_3 на окружности, длина касательных из которых равна x (рис. 13). Тогда событие $\xi < x$ произойдет тогда и только тогда, когда точка будет выбрана на одной из двух дуг M_1M_2 или M_3M . Поэтому

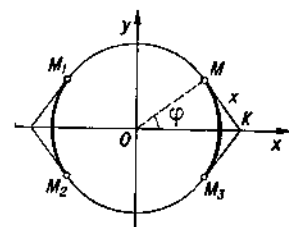


Рис. 13

$$P\{\xi < x\} = \begin{cases} 0, & x \leq 0; \\ \frac{2}{\pi} \operatorname{arctg} x, & x > 0. \end{cases} \quad (20)$$

Распределение с функцией распределения (20) называют распределением Коши. Плотность этого распределения равна

$$p(x) = \begin{cases} 0, & x < 0; \\ \frac{1}{\pi} \frac{1}{1 + x^2}, & x > 0. \end{cases}$$

Часто распределением Коши называют также распределение с плотностью

$$p(x) = \frac{1}{\pi} \frac{1}{1 + x^2}, \quad -\infty < x < +\infty.$$

Некоторые задачи, связанные с нахождением распределений функций от случайных величин. Мы уже отмечали, что если $g(x)$ — борелевская функция, а $\xi(\omega)$ — случайная величина, то $g(\xi(\omega))$ —

также случайная величина. Рассмотрим несколько задач, в которых требуется найти функцию распределения функции от случайной величины.

ЗАДАЧИ

1. Пусть функция распределения случайной величины ξ равна $F(x)$. Найти функцию распределения случайной величины $\eta = \xi^2$.

Решение. Имеем

$$F_{\eta}(x) = P\{\eta < x\} = P\{\xi^2 < x\} = \begin{cases} 0, & x \leq 0; \\ P\{-\sqrt{x} < \xi < \sqrt{x}\}, & x > 0. \end{cases}$$

Следовательно,

$$F_{\eta}(x) = \begin{cases} 0, & x \leq 0; \\ F(\sqrt{x}) - F(-\sqrt{x}), & x > 0. \end{cases}$$

Если случайная величина ξ имеет плотность распределения $p(x)$, то плотность распределения η равна:

$$p_{\eta}(x) = \begin{cases} 0, & x < 0; \\ \frac{1}{2\sqrt{x}} [p(\sqrt{x}) + p(-\sqrt{x})], & x > 0. \end{cases}$$

2. Пусть ξ — случайная величина с непрерывной функцией распределения $F(x)$. Найти функцию распределения случайной величины $\eta = F(\xi)$.

Решение. Так как при всех x $0 \leq F(x) \leq 1$, то $F_{\eta}(x) = 0$ при $x < 0$, $F_{\eta}(x) = 1$ при $x > 1$. Пусть $0 < x < 1$. Обозначим через z точку $z = F^{-1}(x)$ такую, что $F(z) = x$. Событие

$$\{\eta = F(\xi) < x\}$$

произойдет тогда и только тогда, когда произойдет событие $\{\xi < z\}$ (рис. 14). Следовательно,

$$F_{\eta}(x) = \begin{cases} 0, & x \leq 0; \\ F_{\xi}(z) = F(F^{-1}(x)) = x, & 0 < x < 1; \\ 1, & x > 1. \end{cases}$$

Рис. 14

Таким образом, случайная величина η имеет равномерное распределение на отрезке $[0, 1]$.

3. Пусть ξ равномерно распределена на отрезке $[0, 1]$. Найти функцию распределения случайной величины $\eta = \ln \frac{1}{\xi}$.

Решение. Так как $\ln \frac{1}{\xi} > 0$, если $\xi \in (0, 1)$, то $P\{\eta < x\} = 0$ при $x < 0$. Пусть $x > 0$. Тогда

$$P\left\{\ln \frac{1}{\xi} < x\right\} = P\left\{\frac{1}{\xi} < e^x\right\} = P\{\xi > e^{-x}\} = 1 - e^{-x}.$$

Следовательно, η имеет показательное распределение с параметром $\lambda = 1$.

§ 3. Математическое ожидание случайной величины

Математическое ожидание дискретной случайной величины.

Пусть случайная величина ξ принимает конечное число различных значений a_1, a_2, \dots, a_r ; $p_k = P\{\xi = a_k\}$. Будем повторять n раз эксперимент, в котором наблюдается величина ξ . Обозначим

через x_1, x_2, \dots, x_n наблюдаемые значения величины ξ . Тогда среднее значение наблюдаемых величин представимо следующим образом:

$$\frac{1}{n} \sum_{k=1}^n x_k = \sum_{i=1}^r a_i \frac{k_i}{n},$$

где k_i — число появлений в нашей серии экспериментов события $\{\xi = a_i\}$. Таким образом,

$$\frac{1}{n} \sum_{k=1}^n x_k = \sum_{i=1}^r a_i v_n(\{\xi = a_i\}), \quad (1)$$

где $v_n(A)$ — частота события A . Заменяя в формуле (1) частоты на вероятности, получим величину

$$\sum_{i=1}^r a_i p_i = M\xi, \quad (2)$$

которая называется математическим ожиданием (или вероятностным средним) случайной величины ξ . По аналогии с формулой (2) определяется математическое ожидание произвольной дискретной случайной величины.

Определение. Если ξ — дискретная случайная величина, принимающая значения x_1, x_2, \dots с вероятностями p_1, p_2, \dots , то ее математическим ожиданием называется величина

$$M\xi = \sum_k p_k x_k \quad (3)$$

при условии, что ряд справа абсолютно сходится.

Замечания. 1. При определении дискретной величины для нас не важен порядок нумерации ее возможных значений, поэтому естественно предполагать, что сумма ряда в (3) не зависит от порядка слагаемых, а это возможно при абсолютной сходимости ряда.

2. Если величина ξ неотрицательна, то ряд справа в (3) либо сходится абсолютно, либо расходится к $+\infty$. В последнем случае считают, что $M\xi = +\infty$.

Таким образом, среди дискретных случайных величин выделяется класс величин с конечным математическим ожиданием. Для таких величин

$$M|\xi| = \sum_k p_k |x_k| < \infty. \quad (4)$$

Рассмотрим несколько примеров вычисления математического ожидания дискретной случайной величины.

ПРИМЕРЫ

1. Предположим, что случайная величина ξ принимает значения x_1, \dots, x_n с одинаковыми вероятностями $p \{ \xi = x_i \} = \frac{1}{n}$. Тогда

$$M\xi = \frac{1}{n} x_1 + \frac{1}{n} x_2 + \dots + \frac{1}{n} x_n = \frac{x_1 + \dots + x_n}{n}.$$

2. Математическое ожидание биномиального распределения. Предположим, что производятся независимые испытания, и при каждом испытании может быть два исхода: успех, с вероятностью p , или неудача, с вероятностью $q = 1 - p$. Пусть произведено n испытаний, обозначим через Y_n число появлений успеха. Как известно, случайная величина Y_n имеет биномиальное распределение

$$p \{ Y_n = k \} = C_n^k p^k q^{n-k} \\ (k = 0, 1, \dots, n).$$

Математическое ожидание случайной величины Y_n равно

$$MY_n = \sum_{k=0}^n k C_n^k p^k q^{n-k} = np \sum_{k=1}^n \frac{k}{n} C_n^k p^{k-1} q^{n-k} = \\ = np \sum_{k=1}^n C_{n-1}^{k-1} p^{k-1} q^{n-k} = np \sum_{r=0}^{n-1} C_{n-1}^r p^r q^{n-1-r} = np.$$

3. Математическое ожидание геометрического распределения. Предположим, что производятся независимые испытания и при каждом испытании может быть два исхода: успех, с вероятностью p , или неудача, с вероятностью q . Испытания производятся до первого появления успеха. Пусть ξ — число испытаний до первого появления успеха. Случайная величина ξ имеет геометрическое распределение

$$p \{ \xi = k \} = q^k p \quad (k = 0, 1, \dots).$$

Математическое ожидание случайной величины ξ равно

$$M\xi = \sum_{k=0}^{\infty} k q^k p = qp (1 + 2q + 3q^2 + \dots + nq^{n-1} + \dots) = \frac{qp}{q^2} = \frac{p}{q}.$$

4. Математическое ожидание распределения Пуассона. Пусть случайная величина ξ имеет распределение Пуассона с параметром λ . Тогда

$$P \{ \xi = k \} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots,$$

и

$$M\xi = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda} \lambda \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} = e^{-\lambda} \lambda e^{\lambda} = \lambda.$$

Итак, математическое ожидание случайной величины ξ , распределенной по закону Пуассона с параметром λ , равно λ .

Определение математического ожидания в общем случае. Пусть ξ — дискретная случайная величина. Заметим, что при выполнении условия (4) для всех $h > 0$

$$\sum_{k=1}^{\infty} k h P\{kh \leq |\xi| < (k+1)h\} < \infty, \quad (5)$$

так как

$$k h P\{kh \leq |\xi| < (k+1)h\} = k \sum_{kh \leq |x_i| < (k+1)h} < \frac{1}{h} \sum_{kh \leq |x_i| < (k+1)h} |x_i| p_i.$$

Далее, если хотя бы при одном $h > 0$ ряд (5) сходится, то сходится и ряд (4), так как

$$\begin{aligned} \sum_i |x_i| p_i &= \sum_{k=0}^{\infty} \sum_{kh \leq |x_i| < (k+1)h} |x_i| p_i < \sum_{k=0}^{\infty} (k+1) h P\{kh \leq |\xi| < \\ &< (k+1)h\} \leq h + 2h \sum_{k=1}^{\infty} k h P\{kh \leq \xi < (k+1)h\}. \end{aligned}$$

Итак, для существования математического ожидания случайной величины ξ необходимо и достаточно, чтобы при некотором $h > 0$ выполнялось условие (5).

Пусть теперь ξ имеет произвольное распределение.

Определение. Случайная величина называется *интегрируемой*, если для нее при некотором $h > 0$ выполнено условие (5).

Выше мы установили, что дискретная случайная величина имеет математическое ожидание тогда и только тогда, когда она интегрируема. Для произвольной интегрируемой случайной величины справедливо следующее утверждение.

Теорема 1. Для всякой интегрируемой величины ξ существует предел

$$\lim_{h \downarrow 0} \sum_{k=-\infty}^{\infty} k h P\{kh \leq \xi < (k+1)h\}.$$

Прежде чем доказывать эту теорему, установим несколько свойств для математического ожидания дискретной случайной величины.

Лемма 1. Если ξ и η — интегрируемые дискретные случайные величины, а, b — вещественные числа, то величина $a\xi + b\eta$ интегрируема и справедлива формула

$$M(a\xi + b\eta) = aM\xi + bM\eta. \quad (6)$$

Действительно, пусть ξ принимает значения x_1, x_2, \dots ; η — значения y_1, y_2, \dots . Тогда $a\xi + b\eta$ принимает значения $ax_i + by_j$ и

$$M|a\xi + b\eta| = \sum_{i,j} |ax_i + by_j| P(\{\xi = x_i\} \cap \{\eta = y_j\}) \leq$$

$$\leq |a| \sum_i |x_i| P(\{\xi = x_i\} \cap \{\eta = y_i\}) + |b| \sum_j |y_j| P(\{\xi = x_i\} \cap \{\eta = y_j\}) = |a| \sum_i |x_i| P(\xi = x_i) + b \sum_j |y_j| P(\eta = y_j),$$

так как

$$\begin{aligned} \sum_j P(\{\xi = x_i\} \cap \{\eta = y_j\}) &= P(\xi = x_i), \\ \sum_i P(\{\xi = x_i\} \cap \{\eta = y_j\}) &= P(\eta = y_j). \end{aligned} \quad (7)$$

Это доказывает интегрируемость величины $a\xi + b\eta$. Далее

$$\begin{aligned} M(a\xi + b\eta) &= \sum_{i,j} (ax_i + by_j) P(\{\xi = x_i\} \cup \{\eta = y_j\}) = \\ &= a \sum_{i,j} x_i P(\{\xi = x_i\} \cap \{\eta = y_j\}) + b \sum_{i,j} y_j P(\{\xi = x_i\} \cap \{\eta = y_j\}) = \\ &= a \sum_i x_i P(\xi = x_i) + b \sum_j y_j P(\eta = y_j) = aM\xi + bM\eta, \end{aligned}$$

мы воспользовались равенствами (7).

Лемма 2. Если ξ — неотрицательная интегрируемая случайная величина, а $|\eta| \leq \xi$, то случайная величина η также интегрируема, причем $M|\eta| \leq M\xi$.

Действительно,

$$\begin{aligned} \sum_{k=1}^{\infty} kP\{kh \leq |\eta| < (k+1)h\} &= \sum_{k=0}^{\infty} P\{|\eta| \geq kh\} \leq \\ &\leq \sum_{k=0}^{\infty} P\{\xi \geq kh\} = \sum_{k=1}^{\infty} kP\{kh \leq \xi < (k+1)h\} < +\infty. \end{aligned}$$

Здесь мы воспользовались тем, что событие $\{|\eta| > c\}$ влечет событие $\{\xi > c\}$ и, следовательно,

$$P\{\xi > c\} \geq P\{|\eta| > c\}.$$

Было использовано также равенство

$$\sum_{k=1}^{\infty} P\{|\eta| \geq kh\} = \sum_{k=0}^{\infty} kP\{kh \leq |\eta| < (k+1)h\}.$$

Для его доказательства заметим, что

$$\sum_{k=1}^{\infty} P\{|\eta| \geq kh\} = \sum_{k=1}^{\infty} \sum_{r=k}^{\infty} P\{rh \leq |\eta| < (r+1)h\}.$$

Меняя порядок суммирования в двойном ряде, стоящем справа, получим

$$\begin{aligned} \sum_{k=1}^{\infty} P\{|\eta| \geq kh\} &= \sum_{r=1}^{\infty} \sum_{k=1}^r P\{rh \leq |\eta| < (r+1)h\} = \\ &= \sum_{r=1}^{\infty} rP\{rh \leq |\eta| < (r+1)h\} = \sum_{k=0}^{\infty} kP\{kh \leq |\eta| < (k+1)h\}. \end{aligned}$$

Лемма 3. Если ξ — неотрицательная случайная величина, то $M\xi \geq 0$.

Лемма 4. Если $P\{\xi = c\} = 1$, то $M\xi = c$.

Утверждения лемм 3 и 4 следуют из формулы 3.

Доказательство теоремы 1. Обозначим через ξ_h дискретную случайную величину, для которой $\xi_h = kh$ при $kh \leq \xi < (k+1)h$. Тогда при любом $h > 0$

$$0 \leq \xi - \xi_h < h.$$

Заметим также, что при любых $h > 0$ и $h_1 > 0$

$$|\xi_h - \xi_{h_1}| \leq |\xi - \xi_h| + |\xi - \xi_{h_1}| \leq h + h_1. \quad (8)$$

Пусть h таково, что

$$\sum_{k=1}^{\infty} kP\{kh \leq |\xi| < (k+1)h\} < +\infty.$$

Такое h существует согласно предположению об интегрируемости случайной величины ξ . Тогда

$$\begin{aligned} \sum_{k=1}^{\infty} kP\{kh \leq |\xi_h| < (k+1)h\} &= \sum_{k=1}^{\infty} kP\{|\xi_h| = kh\} = \\ &= \sum_{k=1}^{\infty} kP\{kh \leq \xi < (k+1)h\} + \sum_{k=1}^{\infty} kP\{-kh \leq \xi < (-k+1)h\} \leq \\ &\leq \sum_{k=1}^{\infty} kP\{kh \leq \xi < (k+1)h\} + \sum_{k=1}^{\infty} kP\{(-k-1)h \leq \xi < (-k+1)h\} = \\ &= \sum_{k=1}^{\infty} kP\{kh \leq |\xi| < (k+1)h\} + \sum_{k=0}^{\infty} (k+1)P\{-(k+1)h < \\ &< \xi \leq -kh\} = 2 \sum_{k=1}^{\infty} kP\{kh \leq |\xi| < (k+1)h\} + \\ &+ \sum_{k=0}^{\infty} P\{-(k+1)h < \xi \leq -kh\} < +\infty. \end{aligned}$$

Следовательно, величина ξ_h интегрируема. Величина ξ_{h_1} также интегрируема для всех $h_1 > 0$ вследствие леммы 2, так как на основании неравенств (8)

$$|\xi_{h_1}| \leq |\xi_h| + h + h_1,$$

и величина в правой части неравенства интегрируема. Заметим, что

$$M\xi_{h_1} = \sum kh_1P\{kh_1 \leq \xi < (k+1)h_1\}.$$

Согласно лемм 1, 2, 3

$$\begin{aligned} \left| \sum_{k=-\infty}^{\infty} kh_2P\{kh_2 \leq \xi < (k+1)h_2\} - \sum_{k=-\infty}^{\infty} kh_1P\{kh_1 \leq \xi < (k+1)h_1\} \right| = \\ = |M\xi_{h_2} - M\xi_{h_1}| \leq M|\xi_{h_2} - \xi_{h_1}| \leq h_2 + h_1. \end{aligned}$$

Следовательно,

$$\lim_{\substack{h_1 \rightarrow 0 \\ h_2 \rightarrow 0}} |M\xi_{h_1} - M\xi_{h_2}| = 0$$

и предел

$$\lim_{h \rightarrow 0} M\xi_h = \lim_{h \rightarrow 0} \sum_{k=-\infty}^{\infty} khP\{kh \leq \xi < (k+1)h\}.$$

Определение. Математическим ожиданием интегрируемой случайной величины ξ называется величина

$$M\xi = \lim_{h \rightarrow 0} \sum_{k=-\infty}^{\infty} khP\{kh \leq \xi < (k+1)h\}.$$

Замечания. 1. Это определение не противоречит данному выше определению математического ожидания дискретной случайной величины, так как

$$0 \leq M\xi - M\xi_h = M(\xi - \xi_h) \leq h$$

согласно неравенству $0 \leq \xi - \xi_h \leq h$ и лемме 2.

2. При доказательстве теоремы 1 установлено, что для всякой интегрируемой случайной величины ξ

$$M\xi = \lim_{h \rightarrow 0} M\xi_h = \lim_{h \rightarrow 0} \sum_{k=-\infty}^{\infty} khP\{kh \leq \xi < (k+1)h\}. \quad (9)$$

Свойства математического ожидания.

1) Если $P\{\xi = c\} = 1$, то $M\xi = c$. Это вытекает из того, что величина ξ дискретная, принимающая одно значение, и из формулы (2).

2) Если ξ — интегрируемая неотрицательная случайная величина, то $M\xi \geq 0$.

Действительно, $\xi_h \geq 0$, для всех $h > 0$, $M\xi_h \geq 0$ по лемме 2, поэтому и $M\xi \geq 0$ на основании формулы (9).

3) Если $|\xi| \leq C$, то и $|M\xi| \leq C$.

Это вытекает из неравенства

$$-C - h \leq \xi_h \leq C,$$

откуда, согласно свойству 2) и формулы 9, имеем

$$-C - h \leq M\xi_h \leq C, \quad -C \leq M\xi \leq C.$$

4) Если ξ — интегрируемая случайная величина, то $a\xi$ также интегрируема и

$$Ma\xi = aM\xi. \quad (10)$$

Пусть сначала $a > 0$. Тогда, если $(a\xi)_h = kh$ при $kh \leq a\xi < (k+1)h$, то

$$(a\xi)_h = a\xi_h$$

и $M(a\xi)_h = aM\xi_h$ на основании леммы 1. Переходя к пределу

при $h \rightarrow 0$, получаем формулу (10).

Пусть $a = -1$. Тогда, полагая $(-\xi)_h = kh$ при $kh \leq -\xi < (k+1)h$, получаем $\xi_h + (-\xi)_h = -h$, если $\frac{\xi}{h}$ — не целое число, и $\xi_h + (-\xi)_h = 0$, если $\frac{\xi}{h}$ — целое. Поэтому

$$-h \leq M\xi_h + M(-\xi)_h \leq 0$$

на основании леммы 2. Переходя к пределу, получаем $M\xi + M(-\xi) = 0$. Значит и для $a < 0$

$$Ma\xi = M(-|a|\xi) = -|a|M\xi = aM\xi.$$

5) Для интегрируемости величины ξ необходимо и достаточно, чтобы при некотором $h > 0$

$$\sum_{k=1}^{\infty} P\{|\xi| > kh\} < \infty.$$

Это утверждение вытекает из равенства

$$\sum_{k=1}^{\infty} kP\{kh \leq |\xi| < (k+1)h\} = \sum_{k=1}^{\infty} P\{|\xi| \geq kh\},$$

которое уже было использовано выше при доказательстве леммы 2.

6) Если случайная величина η неотрицательна и интегрируема, а $|\xi| \leq \eta$, то ξ также интегрируема.

Действительно, для всех x $P\{\eta \geq x\} \geq P\{|\xi| \geq x\}$. Поэтому

$$\sum_{k=1}^{\infty} P\{|\xi| \geq kh\} \leq \sum_{k=1}^{\infty} P\{\eta \geq kh\} < \infty.$$

7) Если ξ и η интегрируемы, то сумма $\xi + \eta$ также интегрируема и

$$M(\xi + \eta) = M\xi + M\eta. \quad (11)$$

Действительно, каково бы ни было $h > 0$,

$$\begin{aligned} \xi_h &\leq \xi < \xi_h + h, \\ \eta_h &\leq \eta < \eta_h + h, \\ \xi_h + \eta_h &\leq \xi + \eta < \xi_h + \eta_h + 2h. \end{aligned} \quad (12)$$

По лемме 1 $|\xi_h| + |\eta_h|$ — интегрируемая величина, поэтому из неравенства

$$|\xi + \eta| \leq |\xi_h| + |\eta_h| + 2h$$

и свойства 6) вытекает интегрируемость суммы $\xi + \eta$. Если $(\xi + \eta)_h = kh$ при $kh \leq \xi + \eta < (k+1)h$, то из двух неравенств (12) имеем

$$\xi_h + \eta_h \leq (\xi + \eta)_h \leq \xi_h + \eta_h + h.$$

Поскольку величины в этой цепочке неравенств дискретны, то, на основании леммы 2, получаем неравенства для математических ожиданий

$$M\xi_h + M\eta_h \leq M(\xi + \eta)_h \leq M\xi_h + M\eta_h + h.$$

Переходя к пределу при $h \rightarrow 0$, получим равенство (11).

8) Если ξ и η — интегрируемые случайные величины, то при любых постоянных a и b случайная величина $a\xi + b\eta$ также интегрируема, причем

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

Это свойство следует из свойств 4) и 7).

9) Если ξ и η — интегрируемые случайные величины и $\xi \leq \eta$, то $M\xi \leq M\eta$.

Действительно, $\eta - \xi$ — неотрицательная интегрируемая случайная величина. Поэтому, согласно 2), $M(\eta - \xi) \geq 0$. Используя свойство 8, получим $M\eta \geq M\xi$.

10) Случайные величины ξ и $|\xi|$ одновременно интегрируемы и

$$|M\xi| \leq M|\xi|.$$

Это утверждение вытекает из неравенств

$$-|\xi| \leq \xi \leq |\xi|$$

и свойства 9).

11) Пусть ξ_n — последовательность неотрицательных интегрируемых случайных величин $\xi_n \leq \xi_{n+1}$ и $\xi_n \rightarrow \xi$. Если $\sup_n M\xi_n < \infty$, то ξ также интегрируема и $M\xi = \lim_{n \rightarrow \infty} M\xi_n$ ¹.

Докажем интегрируемость ξ . Так как

$$\{\xi \geq kh\} \subset \bigcup_n \{\xi_n \geq (k-1)h\},$$

то

$$P\{\xi \geq kh\} \leq \lim_{n \rightarrow \infty} P\{\xi_n \geq (k-1)h\}. \quad (13)$$

Пусть $(\xi_n)_n = kh$, если $kh \leq \xi_n < (k+1)h$. Тогда

$$M(\xi_n)_h = \sum_{k=1}^{\infty} kh P\{kh \leq \xi_n < (k+1)h\} = h \sum_{k=1}^{\infty} P\{\xi_n \geq kh\},$$

и, следовательно,

$$\sum_{k=1}^{\infty} P\{\xi_k \geq (k-1)h\} = 1 + \frac{1}{h} M(\xi_n)_h \leq 1 + \frac{1}{h} M\xi_n, \quad (14)$$

¹ Свойство 11) часто называют теоремой о монотонной сходимости.

так как $(\xi_n)_h \leq \xi_n$. Имеем для всякого N

$$\begin{aligned} \sum_{k=1}^N P\{\xi \geq kh\} &= \lim_{n \rightarrow \infty} \sum_{k=1}^N P\{\xi_n \geq (k-1)h\} \leq \\ &\leq \lim_{n \rightarrow \infty} \left(1 + \frac{1}{h} M\xi_n\right) \leq 1 + \frac{1}{h} \sup_n M\xi_n. \end{aligned}$$

Переходя к пределу при $N \rightarrow \infty$, убеждаемся в интегрируемости ξ . Так как $\xi_n \leq \xi$, то согласно свойствам 2), 4) и 7)

$$0 \leq M(\xi - \xi_n) = M\xi - M\xi_n,$$

т. е.

$$M\xi \geq \lim_{n \rightarrow \infty} M\xi_n$$

(последний предел существует, так как $\xi_n \leq \xi_{n+1}$, $M\xi_n \leq M\xi_{n+1}$ и, значит, последовательность $M\xi_n$ — возрастающая и ограниченная). Воспользовавшись равенством

$$M(\xi)_h = h \sum_{k=1}^{\infty} P\{\xi \geq kh\}$$

и соотношением (13) и (14), запишем

$$\begin{aligned} M\xi &\leq M(\xi_n) + h = h + h \sum_{k=1}^{\infty} P\{\xi \geq kh\} \leq \\ &\leq h + h \left(\sum_{k=1}^{\infty} \lim_{n \rightarrow \infty} P\{\xi_n \geq (k-1)h\} \right) \leq \\ &\leq h + h \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} P\{\xi_n \geq (k-1)h\} \leq 2h + \lim_{n \rightarrow \infty} M\xi_n. \end{aligned}$$

Мы воспользовались тем, что для неотрицательных $a_k(n)$, для которых $\sum_k a_k(n) < \infty$ и существует предел

$$\begin{aligned} \lim_{n \rightarrow \infty} a_k(n) &= a_k, \\ \sum_{k=1}^N a_k &= \lim_{n \rightarrow \infty} \sum_{k=1}^N a_k(n) \leq \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} a_k(n), \end{aligned}$$

следовательно, и

$$\sum_{k=1}^{\infty} a_k \leq \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} a_k(n).$$

Из соотношения $M\xi \leq 2h + \lim_{n \rightarrow \infty} M\xi_n$ и произвольности $h > 0$ вытекает $M\xi \leq \lim_{n \rightarrow \infty} M\xi_n$. Противоположное неравенство было установлено выше.

Таким образом, $M\xi = \lim_{n \rightarrow \infty} M\xi_n$.

12) Если ξ — неотрицательная величина и $a > 0$, то $P\{\xi \geq a\} \leq \frac{1}{a} M\xi$. Это неравенство называется неравенством Чебышева.

Для доказательства рассмотрим величину η , такую, что $\eta = 0$ при $\xi < a$ и $\eta = a$ при $\xi \geq a$. Имеем $\xi \geq \eta$ и, следовательно, $M\xi \geq M\eta$. Величина η принимает два значения 0 и a , поэтому

$$M\eta = aP\{\eta = a\} = aP\{\xi \geq a\} \leq M\xi.$$

Вычисление математического ожидания. По формуле (3) можно вычислять математическое ожидание для дискретной величины. Для произвольной случайной величины ξ на основании формулы (9), если $F(x)$ — функция распределения величины, имеем

$$\begin{aligned} M\xi &= \lim_{h \rightarrow \infty} \sum_{k=-\infty}^{\infty} khP\{kh \leq \xi < (k+1)h\} = \\ &= \lim_{h \rightarrow \infty} \sum_{k=-\infty}^{\infty} kh[F((k+1)h) - F(kh)] = \int x dF(x), \end{aligned}$$

где $\int x dF(x)$ — интеграл Стильеса, он определяется как предыдущий предел. Для интегрируемости величины ξ необходимо и достаточно, чтобы при некотором $h > 0$ сходилась ряд

$$\sum_{k=-\infty}^{\infty} |k| [F((k+1)h) - F(kh)],$$

при этом существует

$$\lim_{h \rightarrow 0} \sum_{k=-\infty}^{\infty} h |k| [F((k+1)h) - F(kh)] = \int |x| dF(x).$$

Предположим теперь, что величина ξ имеет плотность распределения $f(x)$. Тогда для интегрируемости величины необходимо и достаточно, чтобы

$$\int |x| f(x) dx < \infty, \quad (16)$$

при этом

$$M\xi = \int xf(x) dx, \quad (17)$$

т. е. формула (17) справедлива при условии, что интеграл справа сходится абсолютно.

Действительно,

$$\begin{aligned} khP\{kh \leq \xi < (k+1)h\} &= kh \int_{-(k+1)h}^{-kh} f(x) dx + kh \int_{kh}^{(k+1)h} f(x) dx = \\ &= \left(\int_{-(k+1)h}^{-kh} |x| f(x) dx + \int_{kh}^{(k+1)h} |x| f(x) dx \right) (1 + O(h)). \end{aligned}$$

Поэтому ряд $\sum_{k=-\infty}^{\infty} kP\{kh \leq |\xi| < (k+1)h\}$ сходится одновременно с интегралом (16). Для доказательства формулы (17) нужно лишь заметить, что

$$M\xi_{\xi,h} = \sum_{k=-\infty}^{\infty} kh \int_{kh}^{(k+1)h} f(x) dx = \sum_{k=-\infty}^{\infty} \int_{kh}^{(k+1)h} xf(x) dx + \\ + \sum_{k=-\infty}^{\infty} \int_{kh}^{(k+1)h} (kh - x) f(x) dx = \int xf(x) dx + O(h).$$

Рассмотрим некоторые примеры.

ПРИМЕРЫ

1. Пусть ξ — случайная величина, имеющая равномерное распределение на отрезке $[a, b]$.

Тогда плотность распределения ξ равна

$$f_{\xi}(x) = \begin{cases} \frac{1}{b-a}, & a < x < b, \\ 0, & x \notin [a, b], \end{cases}$$

и по формуле (17)

$$M\xi = \int xf_{\xi}(x) dx = \int_a^b \frac{1}{b-a} x dx = \frac{a+b}{2}.$$

2. Предположим, что случайная величина ξ имеет показательное распределение с параметром λ . Тогда

$$M\xi = \int_0^{\infty} x\lambda e^{-\lambda x} dx = - \int_0^{\infty} x d(e^{-\lambda x}) = -xe^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda}.$$

Таким образом, если продолжительность безотказной работы некоторого прибора имеет показательное распределение с параметром λ , то математическое ожидание (среднее значение) длительности безотказной работы прибора равно $\frac{1}{\lambda}$.

3. Математическое ожидание нормального распределения. Предположим, что случайная величина ξ имеет нормальное распределение $N(a, \sigma^2)$. Это означает, что

$$f_{\xi}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Тогда

$$M\xi = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} xe^{-\frac{(x-a)^2}{2\sigma^2}} dx = \\ = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-a) e^{-\frac{(x-a)^2}{2\sigma^2}} dx + a \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = a,$$

так как

$$\int_{-\infty}^{\infty} (x-a) e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} ze^{-\frac{z^2}{2}} dz = 0,$$

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = 1.$$

Таким образом, параметр a нормального распределения $N(a, \sigma^2)$ равен его математическому ожиданию.

Заметим, что не для всех случайных величин существует математическое ожидание.

ПРИМЕРЫ

1. Пусть случайная величина ξ имеет распределение Коши с плотностью

$$f_{\xi}(x) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

В этом случае

$$\int_{-\infty}^{\infty} |x| f_{\xi}(x) dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|x|}{1+x^2} dx = +\infty.$$

Математическое ожидание случайной величины ξ не существует.

2. Предположим, что производятся независимые испытания. При каждом испытании может быть два исхода — успех или неудача. Вероятность появления успеха при испытании с номером i равна p_i . Обозначим через Y_n число успехов при n испытаниях. Вычислим MY_n .

Пусть

$$\xi_i = \begin{cases} 1, & \text{если в } i\text{-м испытании произошёл успех,} \\ 0, & \text{если в } i\text{-м испытании произошла неудача.} \end{cases}$$

Тогда

$$Y_n = \xi_1 + \dots + \xi_n$$

и

$$MY_n = \sum_{i=1}^n M\xi_i = \sum_{i=1}^n p_i.$$

В частности, если $p_1 = \dots = p_n = p$, то $MY_n = np$.

Математическое ожидание как интеграл Лебега. Для читателей, знакомых с понятием интеграла Лебега по произвольной мере (абстрактного интеграла Лебега)¹, заметим следующее. Пусть (Ω, \mathcal{F}, P) — вероятностное пространство, $\xi(\omega)$ — дискретная случайная величина, принимающая значения $x_1, x_2, \dots, x_n, \dots$ с ве

¹ Понятие интеграла Лебега излагается в учебнике: А. Н. Колмогоров, С. В. Фомин. Введение в теорию функций и функциональный анализ. М., Наука, 1974.

роятностями $p_1, p_2, \dots, p_n, \dots$, причем $\sum |x_n| p_n < +\infty$. Тогда

$$M\xi(\omega) = \sum_n x_n p_n = \int_Q \xi(\omega) dP,$$

где справа стоит интеграл Лебега от функции $\xi(\omega)$ по мере $P(\cdot)$. Если $\xi(\omega)$ — произвольная случайная величина, то по определению считают, что случайная величина $\xi(\omega)$ имеет математическое ожидание, если существует интеграл Лебега

$$\int_Q \xi(\omega) dP.$$

Этот интеграл и называют математическим ожиданием случайной величины $\xi(\omega)$. Из известных свойств интеграла Лебега непосредственно вытекают все сформулированные выше свойства математического ожидания случайной величины.

§ 4. Математическое ожидание функции от случайной величины.

Моменты. Дисперсия

Математическое ожидание функции от случайной величины. Пусть ξ — некоторая случайная величина с функцией распределения $F(x)$, а $g(x)$ — некоторая борелевская функция, заданная на R^1 . Тогда $g(\xi)$ также является случайной величиной. Оказывается, что для вычисления математического ожидания величины $g(\xi)$ не обязательно предварительно находить функцию распределения этой величины, а можно выразить математическое ожидание $g(\xi)$ через функцию $g(x)$ и функцию распределения $F(x)$. Для вывода формулы нам потребуется определение интеграла Стильтьеса

$$\int g(x) dF(x). \quad (1)$$

Мы убедимся, что для всякой непрерывной ограниченной функции $g(x)$ существует предел

$$\lim_{h \rightarrow 0} \sum_{k=-\infty}^{\infty} g(kh) [F((k+1)h) - F(kh)], \quad (2)$$

который и называют в данном случае интегралом (1). При этом

$$Mg(\xi) = \int g(x) dF(x). \quad (3)$$

Действительно, пусть $\xi_h = kh$ при $kh \leq \xi < (k+1)h$. Тогда $g(\xi_h)$ — дискретная величина и

$$Mg(\xi_h) = \sum_{k=-\infty}^{\infty} g(kh) [F((k+1)h) - F(kh)]. \quad (4)$$

Пусть δ таково, что при $|x - y| \leq \delta$, $|x| \leq c$ $|g(x) - g(y)| < \varepsilon$. Тогда при $h < \delta$ $|g(\xi) - g(\xi_h)| \leq \varepsilon$, если $|\xi| \leq c$, и $|g(\xi) - g(\xi_h)| \leq 2m$, если $|\xi| > c$, где $m = \sup_x |g(x)|$. Пусть η — случайная величина такая, что

$$\eta = \begin{cases} \varepsilon, & \text{если } |\xi| \leq c, \\ 2m, & \text{если } |\xi| > c. \end{cases}$$

Тогда $|g(\xi) - g(\xi_h)| \leq \eta$ и, следовательно,

$$M|g(\xi) - g(\xi_h)| \leq M\eta = \varepsilon P\{|\xi| \leq c\} + 2mP\{|\xi| > c\}.$$

Выражение в правой части последнего неравенства можно сделать сколь угодно малым путем выбора достаточно большого c и достаточно малого ε . Поэтому, принимая во внимание неравенство

$$|Mg(\xi) - Mg(\xi_h)| \leq M|g(\xi) - g(\xi_h)|,$$

имеем

$$Mg(\xi) = \lim_{h \rightarrow 0} Mg(\xi_h).$$

На основании формулы (4) можно сделать вывод о существовании предела (2) и получить формулу (3) для непрерывной ограниченной функции $g(x)$.

Определим далее интеграл (1) для неотрицательных непрерывных функций $g(x)$ с помощью равенства

$$\int g(x) dF(x) = \lim_{m \rightarrow \infty} \int [g(x) \wedge m] dF(x), \quad (5)$$

$$g \wedge m = \min[g, m].$$

Если последовательность

$$\int [g(x) \wedge m] dF(x) = M[g(\xi) \wedge m]$$

ограничена, то поскольку $[g(\xi) \wedge m] \uparrow g(\xi)$ при $m \rightarrow \infty$, на основании свойства 11), § 3,

$$Mg(\xi) = \lim_{m \rightarrow \infty} M[g(\xi) \wedge m].$$

Отсюда вытекает существование (конечного или бесконечного) предела в формуле (5) и формула (3) для произвольных непрерывных неотрицательных функций $g(x)$.

Если теперь $g(x)$ — знакопеременная непрерывная функция, для которой

$$\int |g(x)| dF(x) < \infty,$$

то тогда

$$M|g(\xi)| = \int |g(x)| dF(x).$$

Функция $|g(x)| - g(x)$ также неотрицательна, $||g(x)| - g(x)| \leq 2|g(x)|$ и, значит, $M(|g(\xi)| - g(\xi)) < \infty$, поэтому

$$M(|g(\xi)| - g(\xi)) = \int (|g(x)| - g(x)) dF(x).$$

Определим тогда

$$\int g(x) dF(x) = \int |g(x)| dF(x) - \int (|g(x)| - g(x)) dF(x).$$

При таком определении интеграла (1) формула (3) по-прежнему остается в силе.

Рассмотрим теперь случай, когда функция имеет конечное число точек разрыва первого рода. Пусть сначала $g(x) = a_1$ при $x < x_0$, $g(x) = a_2$, $x > x_0$, $g(x_0)$ также определено. Тогда $g(\xi)$ принимает значение a_1 , если $\xi < x_0$, значение $g(x_0)$, если $\xi = x_0$, и значение a_2 , если $\xi > x_0$. Следовательно,

$$Mg(\xi) = a_1 F(a_1) + g(x_0) [F(x_0 + 0) - F(x_0)] + a_2 [1 - F(x_0 + 0)].$$

Правую часть этого равенства обозначим через

$$\int g(x) dF(x).$$

Поскольку функция, имеющая конечное число точек разрыва первого рода x_1, x_2, \dots, x_r , представима в виде

$$g(x) = g_0(x) + \sum_{i=1}^r g_i(x),$$

где $g_0(x)$ непрерывна, а $g_i(x)$ постоянна на интервалах $(-\infty, x_i)$, (x_i, ∞) , то мы можем определить интеграл

$$\int g(x) dF(x) = \sum_{i=0}^r \int g_i(x) dF(x),$$

если существует интеграл от $g_0(x)$ (от $g_i(x)$ интегралы всегда существуют). Так как для каждого слагаемого выполняется формула (3), то эта формула будет справедлива и для функции $g(x)$.

Сформулируем окончательно полученный результат.

Теорема 1. Пусть ξ — случайная величина с функцией распределения $F(x)$, $g(x)$ — непрерывная функция, или $g(x)$ имеет только конечное число точек разрыва первого рода. Пусть

$\int |g(x)| dF(x) < +\infty$. Тогда имеет место формула (3).

Рассмотрим два частных случая формулы (3).

1) Если величина ξ имеет дискретное распределение $P\{\xi = x_k\} = p_k$, $k = 1, 2, \dots$, $\sum p_k = 1$, то

$$Mg(\xi) = \sum g(x_k) p_k, \quad (6)$$

если ряд справа сходится абсолютно. Эта формула вытекает из того, что $g(\xi)$ также дискретная величина и на основании определения математического ожидания дискретной величины.

2) Если величина ξ имеет плотность распределения $f(x)$, то

$$Mg(\xi) = \int g(x) f(x) dx, \quad (7)$$

если $g(x)$ такая борелевская функция, для которой интеграл справа существует. Действительно, если $g(x)$ ограничена и непрерывна, то

$$\begin{aligned} Mg(\xi) &= \lim_{h \rightarrow 0} \sum_{kh}^{(k+1)h} g(kh) \int_{kh}^{(k+1)h} f(x) dx = \\ &= \int g(x) f(x) dx + \lim_{h \rightarrow 0} \int [g_h(x) - g(x)] f(x) dx, \end{aligned} \quad (8)$$

где $g_h(x) = g(kh)$ при $kh \leq x < (k+1)h$. Для всякого c $g_h(x) - g(x) \rightarrow 0$ равномерно на $[-c, c]$, поэтому

$$\lim_{h \rightarrow 0} \int_{-c}^c [g_h(x) - g(x)] f(x) dx = 0.$$

Если $|g(x)| \leq m$, то

$$\left| \int_{|x| > c} [g_h(x) - g(x)] f(x) dx \right| \leq 2m \int_{|x| > c} f(x) dx.$$

Следовательно,

$$\lim_{h \rightarrow 0} \left| \int [g_h(x) - g(x)] f(x) dx \right| \leq 2m \int_{|x| > c} f(x) dx.$$

Так как правая часть стремится к 0 при $c \rightarrow \infty$, то предел в правой части равенств (8) равен 0, следовательно, в этом случае равенство (7) справедливо. Для произвольных борелевских функций $g(x)$ формула (7) получается с помощью предельного перехода.

Приведем несколько примеров использования рассмотренных формул.

ПРИМЕРЫ

1. Пусть ξ — случайная величина, имеющая распределение Пуассона с параметром λ . Вычислить $M \frac{1}{\xi + 1}$.

Решение. Используя формулу (6), получаем

$$\begin{aligned} M \frac{1}{1 + \xi} &= \sum_{k=0}^{\infty} \frac{1}{1+k} \frac{\lambda^k}{k!} e^{-\lambda} = \frac{e^{-\lambda}}{\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{(k+1)!} = \\ &= \frac{e^{-\lambda}}{\lambda} \sum_{r=1}^{\infty} \frac{\lambda^r}{r!} = \frac{e^{-\lambda}}{\lambda} (e^{\lambda} - 1) = \frac{1 - e^{-\lambda}}{\lambda} \end{aligned}$$

(использовано разложение $e^{\lambda} = \sum_{r=0}^{\infty} \frac{\lambda^r}{r!}$).

2. Пусть v_n — случайная величина, имеющая биномиальное распределение с параметрами n и $p = x$, $0 < x < 1$, а $f(x)$ — непрерывная функция. Вычислить $Mf\left(\frac{v_n}{n}\right)$.

Решение. По формуле (6)

$$B_n(x) = Mf\left(\frac{v_n}{n}\right) = \sum_{k=0}^n f\left(\frac{k}{n}\right) C_n^k x^k (1-x)^{n-k}.$$

Многочлен $B_n(x)$ называется многочленом Бернштейна для функции $f(x)$.

3. Пусть ξ — случайная величина, равномерно распределенная на $[0, 1]$. Тогда по формуле (7)

$$M \sin^2 \pi \xi = \int_0^1 \sin^2 \pi x \, dx = \frac{1}{2} \int_0^1 (1 - \cos 2\pi x) \, dx = \frac{1}{2}.$$

4. Пусть ξ — случайная величина с плотностью распределения

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

распределение Коши). Вычислить $M \min(|\xi|, 1)$.

Решение. По формуле (7) имеем

$$\begin{aligned} M \min(|\xi|, 1) &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \min(|x|, 1) \frac{1}{1+x^2} \, dx = \\ &= \frac{2}{\pi} \left[\int_0^1 \frac{x}{1+x^2} \, dx + \int_1^{+\infty} \frac{dx}{1+x^2} \right] = \frac{1}{2} + \frac{1}{\pi} \ln 2. \end{aligned}$$

Обобщенное неравенство Чебышева. Установим в качестве применения равенства (3) одно важное неравенство.

Теорема 2. Пусть $g(x)$ — неотрицательная неубывающая на множестве значений случайной величины ξ функция. Предположим, что существует $Mg(\xi)$. Тогда для каждого $\varepsilon > 0$

$$P\{\xi > \varepsilon\} \leq \frac{Mg(\xi)}{g(\varepsilon)}.$$

Пусть $F(x)$ — функция распределения ξ . Согласно формуле (3) и свойствам функции $g(x)$ имеем

$$\begin{aligned} Mg(\xi) &= \int_{-\infty}^{+\infty} g(x) \, dF(x) \geq \int_{\varepsilon}^{+\infty} g(x) \, dF(x) \geq \\ &\geq g(\varepsilon) \int_{\varepsilon}^{+\infty} dF(x) = g(\varepsilon) P\{\xi > \varepsilon\}, \end{aligned}$$

что и доказывает требуемое неравенство. В частности, если ξ — неотрицательная случайная величина, то взяв $g(x) = x$, получим неравенство Чебышева (см. § 3)

$$P\{\xi > \varepsilon\} \leq \frac{M\xi}{\varepsilon}. \quad (9)$$

Моменты. Дисперсия. Для целых неотрицательных m величина $M\xi^m$, если она определена, называется моментом m -го порядка. В этом случае существует также величина $M|\xi|^m$, которая называется абсолютным моментом m -го порядка. Моменты величины $(\xi - M\xi)$ называются центральными моментами. Наиболее употребительной числовой характеристикой случайной величины ξ является ее второй центральный момент, называемый дисперсией величины ξ :

$$D\xi = M(\xi - M\xi)^2. \quad (10)$$

Так как $(\xi - M\xi)^2$ — неотрицательная величина, то $D\xi$ всегда определена, она может принимать значение $+\infty$. Величина

$$\sigma_\xi = \sqrt{D\xi}$$

называется среднеквадратическим отклонением, или стандартным отклонением величины ξ . Величины $D\xi$, σ_ξ характеризуют, как тесно группируются значения ξ вокруг среднего значения. Установим некоторые свойства дисперсии и моментов.

1) Если $D\xi$ конечна, то для всякого $\varepsilon > 0$

$$P\{|\xi - M\xi| > \varepsilon\} \leq \frac{D\xi}{\varepsilon^2}. \quad (11)$$

Действительно, на основании неравенства Чебышева

$$P\{|\xi - M\xi| > \varepsilon\} = P\{(\xi - M\xi)^2 > \varepsilon^2\} \leq \frac{M(\xi - M\xi)^2}{\varepsilon^2}.$$

Неравенство (11) также называют неравенством Чебышева.

2) Если $D\xi = 0$, то $P\{\xi = M\xi\} = 1$. Действительно, из свойства 1) вытекает, что $P\{|\xi - M\xi| > \varepsilon\} = 0$, каково бы ни было $\varepsilon > 0$. Следовательно,

$$P\{|\xi - M\xi| = 0\} = 1.$$

$$\begin{aligned} 3) D\xi = M(\xi - M\xi)^2 &= M(\xi^2 - 2\xi M\xi + (M\xi)^2) = \\ &= M\xi^2 - (M\xi)^2. \end{aligned} \quad (12)$$

Из этого соотношения вытекают следующие важные равенства и неравенства:

$$\begin{aligned} M\xi^2 &= D\xi + (M\xi)^2, \\ |M\xi| &\leq \sqrt{M\xi^2}, \\ D\xi &\leq M\xi^2. \end{aligned}$$

4) Если для величин ξ и η величины $M\xi^2$ и $M\eta^2$ конечны, то справедливо неравенство Коши — Буняковского

$$(M\xi\eta)^2 \leq M\xi^2 M\eta^2. \quad (13)$$

Действительно, квадратный трехчлен

$$M(\xi + i\eta)^2 = M\xi^2 + 2iM\xi\eta + i^2M\eta^2$$

неотрицателен, поэтому его дискриминант не положителен. Существование $M\xi\eta$ вытекает из неравенства

$$|\xi\eta| \leq \frac{1}{2}(\xi^2 + \eta^2).$$

5) Если $M\xi^2 < \infty$ и $M\eta^2 < \infty$, то справедливо неравенство Шварца

$$\sqrt{M(\xi + \eta)^2} \leq \sqrt{M\xi^2} + \sqrt{M\eta^2}. \quad (14)$$

Действительно, из неравенства (13) имеем

$$M(\xi + \eta)^2 = M\xi^2 + 2M\xi\eta + M\eta^2 \leq M\xi^2 + 2\sqrt{M\xi^2 M\eta^2} + M\eta^2.$$

6) Если $D\xi$ существует, то для всех a и b

$$D(a\xi + b) = a^2 D\xi. \quad (15)$$

Действительно,

$$\begin{aligned} M(a\xi + b - M(a\xi + b))^2 &= M(a\xi + b - aM\xi - b)^2 = \\ &= a^2 M(\xi - M\xi)^2. \end{aligned}$$

7) Если $D\xi$ существует, то для всех c имеет место неравенство

$$M(\xi - c)^2 \geq D\xi = M(\xi - M\xi)^2,$$

и, следовательно, $M(\xi - c)^2$ принимает наименьшее значение при $c = M\xi$.

В самом деле, используя свойства математического ожидания, получаем

$$\begin{aligned} M(\xi - c)^2 &= M[(\xi - M\xi) + (M\xi - c)]^2 = \\ &= M(\xi - M\xi)^2 + 2(M\xi - c)M(\xi - M\xi) + (c - M\xi)^2 = \\ &= D\xi + (c - M\xi)^2 \geq D\xi. \end{aligned}$$

Рассмотрим, наконец, формулы для вычисления моментов и дисперсии. В общем случае,

$$M\xi^m = \int_{-\infty}^{+\infty} x^m dF(x), \quad (16)$$

$$M|\xi|^m = \int_{-\infty}^{+\infty} |x|^m dF(x), \quad M(\xi - a)^m = \int_{-\infty}^{+\infty} (x - a)^m dF(x).$$

Все эти моменты определены одновременно. В частности,

$$D\xi = \int_{-\infty}^{+\infty} (x - M\xi)^2 dF(x) = \int x^2 dF(x) - \left(\int x dF(x) \right)^2. \quad (17)$$

Для дискретной величины ξ , принимающей значения x_k с вероятностью p_k ,

$$D\xi = \sum_k (x_k - M\xi)^2 p_k = \sum_k x_k^2 p_k - \left(\sum_k x_k p_k\right)^2. \quad (18)$$

Если же величина ξ имеет плотность распределения $f(x)$, то

$$D\xi = \int (x - M\xi)^2 f(x) dx = \int x^2 f(x) dx - \left(\int x f(x) dx\right)^2. \quad (19)$$

Рассмотрим несколько примеров.

ПРИМЕРЫ

1. Пусть ξ — случайная величина, имеющая распределение Пуассона. Вычислить дисперсию и среднеквадратическое отклонение случайной величины ξ .
Решение. Найдем сначала $M\xi^2$. Имеем

$$\begin{aligned} M\xi^2 &= \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} = \\ &= \lambda e^{-\lambda} \sum_{r=0}^{\infty} (r+1) \frac{\lambda^r}{r!} = \lambda e^{-\lambda} \sum_{r=0}^{\infty} r \frac{\lambda^r}{r!} + \lambda e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} = \\ &= e^{-\lambda} \left(\lambda^2 \sum_{r=1}^{\infty} \frac{\lambda^{r-1}}{(r-1)!} + \lambda \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} \right) = \lambda^2 + \lambda. \end{aligned}$$

Выше мы установили, что $M\xi = \lambda$. Следовательно,

$$D\xi = M\xi^2 - (M\xi)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Итак, дисперсия случайной величины ξ , распределенной по закону Пуассона с параметром λ , равна λ ; среднеквадратическое отклонение ξ равно $\sqrt{\lambda}$.

2. Дисперсия равномерного распределения. Пусть ξ — случайная величина, имеющая равномерное распределение на отрезке $[a, b]$. Тогда $M\xi = \frac{a+b}{2}$. Вычислить $M\xi^2$.

Решение. Имеем

$$M\xi^2 = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{3} \frac{b^3 - a^3}{b-a} = \frac{1}{3} (b^2 + ab + a^2).$$

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{1}{3} (b^2 + ab + a^2) - \frac{1}{4} (a+b)^2 = \frac{(b-a)^2}{12}.$$

Таким образом, дисперсия равномерно распределенной величины зависит только от длины интервала и является возрастающей функцией от длины интервала.

3. Центральные моменты нормального распределения. Предположим, что случайная величина имеет нормальное распределение с параметрами a и σ^2 . Плотность распределения в этом случае равна

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(x-a)^2}{\sigma^2}}.$$

Как известно, $a = M\xi$. Вычислять центральные моменты случайной величины ξ .

Решение. Имеем

$$M(\xi - a)^n = \frac{1}{\sqrt{2\pi}\sigma} \int (x - a)^n \exp \left\{ -\frac{1}{2} \frac{(x - a)^2}{\sigma^2} \right\} dx.$$

В результате подстановки $\frac{x - a}{\sigma} = z$ получим

$$M(\xi - a)^n = \frac{\sigma^n}{\sqrt{2\pi}} \int z^n e^{-\frac{1}{2} z^2} dz.$$

Если n — нечетное, то $M(\xi - a)^n = 0$; если n — четное ($n = 2k$), то

$$M(\xi - a)^{2k} = \frac{2\sigma^{2k}}{\sqrt{2\pi}} \int_0^{+\infty} z^{2k} e^{-\frac{1}{2} z^2} dz.$$

Выполняя замену переменных $\frac{1}{2} z^2 = t$, получим

$$\begin{aligned} M(\xi - a)^{2k} &= \frac{2^k \sigma^{2k}}{\sqrt{\pi}} \int_0^{+\infty} t^{k-\frac{1}{2}} e^{-t} dt = \frac{2^k \sigma^{2k}}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right) = \\ &= 1 \cdot 3 \dots (2k-1) \sigma^{2k} = (2k-1)!! \sigma^{2k}. \end{aligned}$$

В частности,

$$D_\xi^2 = M(\xi - a)^2 = \sigma^2.$$

Таким образом, мы выяснили вероятностный смысл параметра σ^2 нормального закона. Нормальный закон распределения $N(a, \sigma^2)$ полностью определяется заданием двух параметров: параметра $a = M_\xi$ и параметра $\sigma^2 = D_\xi^2$.

4. Моменты показательного распределения. Пусть случайная величина ξ имеет показательное распределение с параметром λ . Вычислить моменты случайной величины ξ .

Решение. Имеем

$$\begin{aligned} M_\xi^k &= \int_0^{+\infty} x^k \lambda e^{-\lambda x} dx = \lambda \int_0^{+\infty} x^k e^{-\lambda x} dx = \\ &= \frac{1}{\lambda^k} \int_0^{+\infty} t^k e^{-t} dt = \frac{\Gamma(k+1)}{\lambda^k} = \frac{k!}{\lambda^k}. \end{aligned}$$

В частности, $M_\xi^1 = \frac{1}{\lambda}$, $M_\xi^2 = \frac{2}{\lambda^2}$, и поэтому

$$D_\xi^2 = M_\xi^2 - (M_\xi^1)^2 = \frac{1}{\lambda^2}.$$

5. Некоторые применения неравенства Чебышева. Неравенство Чебышева дает оценку вероятностей «значительных» отклонений случайной величины от своего математического ожидания. Запишем неравенство Чебышева в следующей форме:

$$P(|\xi - M_\xi| > k \sqrt{D_\xi^2}) < \frac{D_\xi^2}{k^2 D_\xi^2} = \frac{1}{k^2}, \quad (20)$$

где k — произвольное положительное число. Таким образом,

$$P(|\xi - M_\xi| < k \sqrt{D_\xi^2}) = 1 - P(|\xi - M_\xi| > k \sqrt{D_\xi^2}) > 1 - \frac{1}{k^2}. \quad (21)$$

Например, если взять $k=3$, то с вероятностью большей, чем $1 - \frac{1}{9} = \frac{8}{9}$, можно утверждать, что значение случайной величины ξ принадлежит интервалу $(M\xi - 3\sqrt{D\xi}, M\xi + 3\sqrt{D\xi})$, т. е.

$$P\{|\xi - M\xi| < 3\sqrt{D\xi}\} > 1 - \frac{1}{9} = \frac{8}{9}. \quad (22)$$

Предположим, что случайная величина ξ нормально распределена с параметрами a и σ^2 . Сравним точное значение вероятности события $P\{|\xi - a| < 3\sigma\}$ с оценкой (20). Так как случайная величина $\frac{\xi - a}{\sigma}$ имеет нормальное распределение с параметрами 0 и 1, то

$$P\{|\xi - a| < 3\sigma\} = P\left\{\left|\frac{\xi - a}{\sigma}\right| < 3\right\} = 2\Phi(3) \approx 0,997,$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

Неравенство Гельдера. Пусть $p > 1$, $q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$ и случайные величины ξ и η таковы, что $M|\xi|^p < +\infty$, $M|\eta|^q < +\infty$. Тогда

$$M|\xi\eta| \leq (M|\xi|^p)^{\frac{1}{p}} (M|\eta|^q)^{\frac{1}{q}}. \quad (23)$$

Это неравенство называется неравенством Гельдера. Частным случаем неравенства Гельдера является установленное выше неравенство Коши—Буняковского (оно получается из неравенства Гельдера при $p = q = 2$).

Рассмотрим кривую OBA , уравнение которой

$$y = x^{p-1},$$

или

$$x = y^{q-1}.$$

Из рис. 15, соответствующего случаю $b < a^{p-1}$, вытекает, что сумма площадей

$$S = \int_0^a x^{p-1} dx = \frac{1}{p} a^p \text{ и } T = \int_0^b y^{q-1} dy = \frac{1}{q} b^q$$

не меньше, чем ab :

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Положим в этом неравенстве

$$a = \frac{|\xi|}{(M|\xi|^p)^{\frac{1}{p}}}, \quad b = \frac{|\eta|}{(M|\eta|^q)^{\frac{1}{q}}}$$

и возьмем математическое ожидание от обеих частей полученного неравенства. Тогда, используя свойство 2) математического ожидания (см. § 3), получим

$$M \frac{|\xi|}{(M|\xi|^p)^{\frac{1}{p}}} = \frac{|\eta|}{(M|\eta|^q)^{\frac{1}{q}}} \leq \frac{1}{p} + \frac{1}{q} = 1,$$

откуда и следует (23).

Неравенство Иенсена. Пусть $g(x)$ — функция, определенная на открытом интервале \mathcal{J} и $C = (x_0, g(x_0))$ — точка, лежащая на графике этой функции. Проходящая через точку C прямая l называется **опорной прямой** функции $g(x)$ в точке x_0 , если график $g(x)$ целиком лежит над l или на l . Это означает, что существует $k(x_0)$ такое, что при всех $x \in \mathcal{J}$ выполняется неравенство

$$g(x) \geq g(x_0) + k(x_0)(x - x_0), \quad (24)$$

где $k(x_0)$ — угловой коэффициент прямой l . Из курса математического анализа известно, что дважды дифференцируемая функция $g(x)$ выпукла на \mathcal{J} тогда и только тогда, когда $g''(x) \geq 0$ ($x \in \mathcal{J}$).

Пусть ξ — случайная величина с областью значений \mathcal{J} , а $g(x)$ — выпуклая функция на \mathcal{J} . Положим в (23) $x = \xi$, $x_0 = M\xi$ и возьмем математическое ожидание от обеих частей полученного неравенства. Тогда, используя свойство 9) математического ожидания (§ 3), получим

$$Mg(\xi) \geq g(M\xi). \quad (25)$$

Неравенство (25) называется **неравенством Иенсена**. В частности, если $g(x) = x^2$, то $(M\xi)^2 \leq M\xi^2$; если $g(x) = |x|^r$ ($x \geq 1$), то $|M\xi|^r \leq M|\xi|^r$.

Неравенство Ляпунова. Если $0 < s < t$, то

$$(M|\xi|^s)^{\frac{1}{s}} \leq (M|\xi|^t)^{\frac{1}{t}}. \quad (26)$$

Для доказательства положим $r = \frac{t}{s}$, $g(x) = |x|^r$, $r \geq 1$, $\eta = |\xi|^s$. Тогда, применяя неравенство Иенсена, получим

$$(M|\xi|^s)^{\frac{t}{s}} \leq M(|\xi|^s)^{\frac{t}{s}} = M|\xi|^t.$$

Из этого неравенства и следует неравенство (26), которое называется **неравенством Ляпунова**.

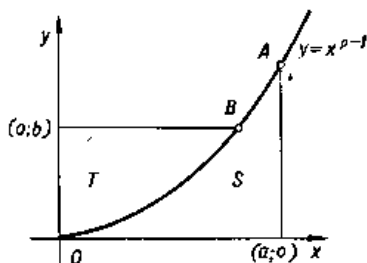


Рис. 15

§ 5. Независимые случайные величины

Определение. Величины $\xi_1, \xi_2, \dots, \xi_n$ называются *независимыми*, если каковы бы ни были x_1, x_2, \dots, x_n события $\{\xi_1 < x_1\}, \{\xi_2 < x_2\}, \dots, \{\xi_n < x_n\}$ независимые, т. е. для всех x_1, x_2, \dots, x_n

$$P\left(\bigcap_{k=1}^n \{\xi_k < x_k\}\right) = \prod_{k=1}^n P\{\xi_k < x_k\}. \quad (1)$$

Замечание. Поскольку некоторые x_k можно положить равными $+\infty$, то для всякой подпоследовательности k_1, \dots, k_m из соотношения (1) вытекает соотношение

$$P\left(\bigcap_{i=1}^m \{\xi_{k_i} < x_{k_i}\}\right) = \prod_{i=1}^m P\{\xi_{k_i} < x_{k_i}\}, \quad (2)$$

которое и означает независимость событий $\{\xi_1 < x_1\}, \{\xi_2 < x_2\}, \dots, \{\xi_n < x_n\}$.

Рассмотрим основные свойства независимых случайных величин.

Теорема 1. Если $\xi_1, \xi_2, \dots, \xi_n$ — независимые величины, A_1, A_2, \dots, A_n — произвольные борелевские множества, то

$$P\left(\bigcap_{k=1}^n \{\xi_k \in A_k\}\right) = \prod_{k=1}^n P\{\xi_k \in A_k\}. \quad (3)$$

Доказательство. Утверждение (3) справедливо, если A_k являются полуинтервалами вида $(-\infty, x_k)$. Совокупность \mathcal{A} множеств A , для которых выполнено условие (3), удовлетворяет следующим условиям: а) вместе с A'_1 и A''_1 совокупность \mathcal{A} содержит $A'_1 \setminus A''_1$, если $A'_1 \supset A''_1$. Действительно,

$$\begin{aligned} P\left(\bigcap_{k=2}^n \{\xi_k \in A_k\} \cap \{\xi_1 \in A'_1 \setminus A''_1\}\right) &= P(\{\xi_1 \in A'_1\} \cap \bigcap_{k=2}^n \{\xi_k \in A_k\}) - \\ &- P(\{\xi_1 \in A''_1\} \cap \bigcap_{k=2}^n \{\xi_k \in A_k\}) = P\{\xi_1 \in A'_1\} \prod_{k=2}^n P\{\xi_k \in A_k\} - \\ &- P\{\xi_1 \in A''_1\} \prod_{k=2}^n P\{\xi_k \in A_k\} = P\{\xi_1 \in A'_1 \setminus A''_1\} \prod_{k=2}^n P\{\xi_k \in A_k\}. \end{aligned}$$

б) вместе с A'_1 и A''_1 она содержит и $A'_1 \cup A''_1$, если $A'_1 \cap A''_1 = \emptyset$ (доказательство аналогично предыдущему, только знаки \setminus и $-$ нужно заменить на \cup и $+$).

в) если $A_1^{(n)} \subset A_1^{(n+1)}$ ($A_1^{(n)} \supset A_1^{(n+1)}$), $A_1^{(n)} \in \mathcal{A}$ для всех n , то и $\bigcup_n A_1^{(n)}$ ($\bigcap_n A_1^{(n)}$) принадлежит \mathcal{A} (для этого достаточно подставить в равенство (3)

вместо A — $A_1^{(n)}$ и перейти к пределу при $n \rightarrow \infty$). Из условий а) и б) вытекает, что \mathcal{A} содержит алгебру, порожденную открытыми справа полуинтервалами, а из в) вытекает, что \mathcal{A} — монотонный класс. Следовательно, \mathcal{A} содержит σ -алгебру борелевских множеств (теорема 4, § 5). Итак, в равенстве (3) A_k может быть произвольным борелевским множеством. Аналогичные рассуждения применимы к каждому из множеств A_k .

Следствие 1. Если $\xi_1, \xi_2, \dots, \xi_n$ — независимые дискретные величины, то каковы бы ни были x_1, x_2, \dots, x_n ,

$$P\left(\bigcap_{k=1}^n \{\xi_k = x_k\}\right) = \prod_{k=1}^n P\{\xi_k = x_k\}. \quad (4)$$

Наоборот, если для дискретных величин $\xi_1, \xi_2, \dots, \xi_n$ для всех x_1, x_2, \dots, x_n выполнено (4), то эти величины независимы.

Соотношение (4) получается из (3), если взять в качестве A_k одноточечное множество $\{x_k\}$. Пусть далее ξ_k с положительной вероятностью принимает значения $y_k^{(m)}$, $m = 1, 2, \dots$. Тогда

$$\begin{aligned} P\left(\bigcap_{k=1}^n \{\xi_k < x_k\}\right) &= \sum_{y_1^{(m)} < x_1, \dots, y_n^{(m)} < x_n} P\left(\bigcap_{k=1}^n \{\xi_k = y_k^{(m_k)}\}\right) = \\ &= \sum_{y_1^{(m_1)} < x_1, \dots, y_n^{(m_n)} < x_n} \prod_{k=1}^n P\{\xi_k = y_k^{(m_k)}\} = \\ &= \prod_{k=1}^n \sum_{y_k^{(m_k)} < x_k} P\{\xi_k = y_k^{(m_k)}\} = \prod_{k=1}^n P\{\xi_k < x_k\}. \end{aligned}$$

Следствие 2. Если $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины, $g_1(x), g_2(x), \dots, g_n(x)$ — борелевские функции, то и величины $g_1(\xi_1), g_2(\xi_2), \dots, g_n(\xi_n)$ также независимы.

Действительно, пусть $A_k = \{x : g_k(x) < y_k\}$. Это борелевское множество. Поэтому

$$\begin{aligned} P\left(\bigcap_{k=1}^n \{g_k(\xi_k) < y_k\}\right) &= P\left(\bigcap_{k=1}^n \{\xi_k \in A_k\}\right) = \\ &= \prod_{k=1}^n P\{\xi_k \in A_k\} = \prod_{k=1}^n P\{g_k(\xi_k) < y_k\}. \end{aligned}$$

Теорема 2. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — интегрируемые независимые случайные величины. Тогда $\xi_1 \xi_2 \dots \xi_n$ — также интегрируемая случайная величина и

$$M\xi_1 \xi_2 \dots \xi_n = M\xi_1 M\xi_2 \dots M\xi_n. \quad (5)$$

Доказательство. Предположим сначала, что ξ_k — дискретные величины со значениями $y_k^{(m)}$. Тогда формулу (5) получаем из равенства

$$\begin{aligned} M\xi_1 \xi_2 \dots \xi_n &= \sum_{m_1, \dots, m_n} y_1^{(m_1)} \dots y_n^{(m_n)} P\left(\bigcap_{k=1}^n \{\xi_k = y_k^{(m_k)}\}\right) = \\ &= \sum_{m_1, \dots, m_n} y_1^{(m_1)} \dots y_n^{(m_n)} \prod_{k=1}^n P\{\xi_k = y_k^{(m_k)}\} = \prod_{k=1}^n y_k^{(m_k)} P\{\xi_k = y_k^{(m_k)}\}, \end{aligned}$$

абсолютная сходимость рядов гарантируется сходимостью рядов

$$\sum_{m_k} |y_k^{(m_k)}| \mathbf{P} \{ \xi_k = y_k^{(m_k)} \}, \quad k = 1, \dots, n.$$

Введем теперь величины $\xi_k^{(h)} = \{h$ при $lh \leq \xi_k < (l+1)h$. Для всякого $h > 0$ эти величины интегрируемы и независимы, на основании следствия 2. Таким образом,

$$\mathbf{M} \prod_{k=1}^n \xi_k^{(h)} = \prod_{k=1}^n \mathbf{M} \xi_k^{(h)}.$$

Из неравенств $\xi_k^{(h)} \leq \xi_k < \xi_k^{(h)} + h$ вытекает, что $|\xi_k| \leq |\xi_k^{(h)}| + h$. Тогда

$$\prod_{k=1}^n |\xi_k| \leq \prod_{k=1}^n (|\xi_k^{(h)}| + h)$$

и величина $\prod_{k=1}^n \xi_k$ интегрируема. Далее

$$\prod_{k=1}^n \xi_k - \prod_{k=1}^n \xi_k^{(h)} = \sum_{k=1}^n \left(\prod_{i=1}^{k-1} \xi_i \right) (\xi_k - \xi_k^{(h)}) \prod_{j=k+1}^n \xi_j^{(h)}.$$

Следовательно,

$$\left| \prod_{k=1}^n \xi_k - \prod_{k=1}^n \xi_k^{(h)} \right| \leq h \sum_{k=1}^n \sum_{i=1}^{k-1} (|\xi_i^{(h)}| + h) \prod_{j=k+1}^n (|\xi_j^{(h)}|).$$

Согласно следствию 2, в произведении под знаком суммы стоят независимые величины, кроме того, они интегрируемы, поэтому

$$\mathbf{M} \left| \prod_{k=1}^n \xi_k - \prod_{k=1}^n \xi_k^{(h)} \right| \leq h \sum_{k=1}^n \prod_{i=1}^{k-1} (\mathbf{M} |\xi_i^{(h)}| + h) \prod_{j=k+1}^n \mathbf{M} |\xi_j^{(h)}|.$$

Таким образом,

$$\left| \mathbf{M} \prod_{k=1}^n \xi_k - \prod_{k=1}^n \mathbf{M} \xi_k^{(h)} \right| = O(h).$$

Переходя к пределу при $h \downarrow 0$ и учитывая, что $\mathbf{M} \xi_k^{(h)} \rightarrow \mathbf{M} \xi_k$, получаем доказательство теоремы.

Следствие. Если $\xi_1, \xi_2, \dots, \xi_n$ независимы, а g_1, g_2, \dots, g_n — такие борелевские функции, что случайные величины $g_k(\xi_k)$ интегрируемы при $k = 1, 2, \dots, n$, то

$$\mathbf{M} \prod_{k=1}^n g_k(\xi_k) = \prod_{k=1}^n \mathbf{M} g_k(\xi_k) \quad (6)$$

(мы воспользовались следствием 2 из теоремы 1).

Замечание. Если для всяких ограниченных непрерывных функций $g_1(x), \dots, g_n(x)$ выполнено равенство (6), то величины ξ_1, \dots, ξ_n независимы.

Действительно, пусть $g_k^{(n)}(x) = 1$ при $x \leq x_k - \frac{1}{n}$, $g_k^{(n)}(x) = 0$ при $x \geq x_k$, $g_k^{(n)}(x) = n(x_k - x)$ при $x_k - \frac{1}{n} \leq x \leq x_k$. Тогда

$$M \prod_{k=1}^n g_k^{(n)}(\xi_k) = \prod_{k=1}^n M g_k^{(n)}(\xi_k),$$

$$g_k^{(n)}(\xi_k) \uparrow \begin{cases} 1, & \xi_k < x_k, \\ 0, & \xi_k \geq x_k. \end{cases}$$

Переходя к пределу при $n \rightarrow \infty$ и учитывая свойство II, § 4, получим

$$P\left(\bigcap_{k=1}^n \{\xi_k < x_k\}\right) = \prod_{k=1}^n P\{\xi_k < x_k\}.$$

В дальнейшем нам часто придется использовать независимость величин, выражающихся через различные функции независимых величин. Установим нужный нам факт.

Теорема 3. Пусть величины $\xi_1, \xi_2, \dots, \xi_n, \eta_1, \dots, \eta_m$ независимы, а $g(x_1, \dots, x_n), h(y_1, \dots, y_m)$ — две борелевские функции. Тогда случайные величины $g(\xi_1, \dots, \xi_n)$ и $h(\eta_1, \dots, \eta_m)$ независимы.

Доказательство. Для доказательства, на основании предыдущего замечания, достаточно показать, что каковы бы ни были ограниченные непрерывные функции $u(y)$ и $v(y)$, выполняется соотношение

$$Mu(g(\xi_1, \dots, \xi_n))v(h(\eta_1, \dots, \eta_m)) =$$

$$= Mu(g(\xi_1, \dots, \xi_n))Mv(h(\eta_1, \dots, \eta_m)).$$

Для этого достаточно, чтобы для любых ограниченных борелевских функций g и h имело место равенство

$$Mg(\xi_1, \dots, \xi_n)h(\eta_1, \dots, \eta_m) = Mg(\xi_1, \dots, \xi_n)Mh(\eta_1, \dots, \eta_m). \quad (7)$$

Множество \mathfrak{L} функций g , для которых при фиксированном h выполнено равенство (7), удовлетворяет следующим условиям:

- если $g_1, g_2 \in \mathfrak{L}$, то $ag_1 + bg_2 \in \mathfrak{L}$ для всех вещественных a, b ;
- если $g_n \in \mathfrak{L}$ и g_n ограничены одной и той же постоянной, то $\lim g_n \in \mathfrak{L}$, если только этот предел существует;

в) \mathfrak{L} содержит функции g вида $g(x_1, \dots, x_n) = \prod_{k=1}^n \varphi_k(x_k)$, где φ_k — ограниченные борелевские функции. Но множество функций, удовлетворяющее условиям а)–в), содержит все ограниченные борелевские функции. Аналогичные рассуждения применимы и к h .

Следствие 1. Пусть $\xi_1^{(1)}, \dots, \xi_{n_1}^{(1)}, \xi_1^{(2)}, \dots, \xi_{n_2}^{(2)}, \dots, \xi_1^{(r)}, \dots, \xi_{n_r}^{(r)}$ — независимые случайные величины, а $g_1(x_1, \dots, x_{n_1}), \dots, g_r(x_1, \dots, x_{n_r})$ — произвольные борелевские функции. Тогда случайные величины

$$\eta_k = g_k(\xi_1^{(k)}, \dots, \xi_{n_k}^{(k)}), \quad k = 1, \dots, r$$

также независимы.

Действительно, если φ_k — ограниченные непрерывные функции, то, согласно (7),

$$M \prod_{k=1}^r \varphi_k(\eta_k) = M \left(\prod_{k=1}^{r-1} \varphi_k(\eta_k) \varphi_r(\eta_r) \right) = M \left(\prod_{k=1}^{r-1} \varphi_k(\eta_k) \right) M \varphi_r(\eta_r) = \prod_{k=1}^r M \varphi_k(\eta_k)$$

Остается воспользоваться замечанием.

Следствие 2. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые случайные величины с конечными дисперсиями. Тогда

$$D \sum_{k=1}^n \xi_k = \sum_{k=1}^n D \xi_k. \quad (8)$$

Действительно, для двух величин ξ и η

$$D(\xi + \eta) = M(\xi + \eta) - (M\xi + M\eta)^2 = M\xi^2 + 2M\xi\eta + M\eta^2 - (M\xi)^2 - 2M\xi M\eta - (M\eta)^2 = M\xi^2 - (M\xi)^2 + M\eta^2 - (M\eta)^2,$$

так как $M\xi\eta = M\xi M\eta$. Далее формула (8) проверяется по индукции числа слагаемых, при этом используется независимость величин

$$\sum_{k=1}^{n-1} \xi_k \text{ и } \xi_n.$$

ПРИМЕР

В качестве иллюстрации следствия 2 вычислим дисперсию случайной величины, имеющей биномиальное распределение.

Найдем дисперсию случайной величины γ_n , введенной при рассмотрении примера 2, с. 92. Согласно следствию 2,

$$D\gamma_n = \sum_{k=1}^n D\xi_k.$$

Так как $M\xi_k^2 = p_k$, то $D\xi_k = p_k - p_k^2 = p_k q_k$. Следовательно,

$$D\gamma_n = \sum_{k=1}^n p_k q_k.$$

В частности, если $p_1 = \dots = p_n = p$, то

$$D\gamma_n = npq.$$

§ 6. Предельные теоремы для биномиального распределения

Предположим, что произведено n независимых испытаний, в каждом из которых может быть два исхода: успех с вероятностью p и неудача с вероятностью q ($p + q = 1$). Пусть γ_n — число успехов при n испытаниях. Как известно (см. § 2, 3, 5),

$$B_p(n, m) = P\{\gamma_n = m\} = C_n^m p^m q^{n-m}, \quad (1)$$

$$m = 0, 1, \dots, n,$$

$$M\gamma_n = np, \quad D\gamma_n = npq. \quad (2)$$

При больших значениях n и m вычисление вероятностей $B_p(n, m)$ по формуле 1 представляет значительные трудности. Возникает необходимость в асимптотических формулах, позволяющих с достаточной степенью точности определять эти вероятности.

Изучим асимптотическое поведение вероятностей $B_p(n, m)$ при $n \rightarrow \infty$. Асимптотическое поведение этих вероятностей зависит от характера изменения Dy_n ; мы будем предполагать, что вероятности p также зависят от n .

Теорема 1. Локальная предельная теорема. Обозначим $a_n = np$, $b_n = npq$, $x_{n,m} = (m - a_n) b_n^{-\frac{1}{2}}$. Тогда, если при $n \rightarrow \infty$ $b_n \rightarrow \infty$, а $|x_{n,m}| \leq c$, где c — некоторая постоянная, то

$$\lim_{n \rightarrow \infty} \frac{B_p(n, m)}{\frac{1}{\sqrt{2\pi b_n}} \exp\left\{-\frac{1}{2} x_{n,m}^2\right\}} = 1. \quad (3)$$

Доказательство. Заметим, что по условию теоремы $m \rightarrow \infty$ и $n - m \rightarrow \infty$, поскольку

$$m = a_n + \sqrt{b_n} x_{n,m} = \sqrt{b_n} \left(\frac{\sqrt{b_n}}{q} + x_{n,m} \right) \geq \sqrt{b_n} \left(\frac{\sqrt{b_n}}{q} - c \right),$$

$$n - m = nq - \sqrt{b_n} x_{n,m} = \sqrt{b_n} \left(\frac{\sqrt{b_n}}{q} - c \right).$$

В дальнейшем символ $a_n \sim b_n$, где $\{a_n\}$ и $\{b_n\}$ — две числовые последовательности, будет означать, что

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1.$$

Воспользуемся формулой Стирлинга для факториалов

$$n! \sim \sqrt{2\pi n} n^n e^{-n}.$$

Получим

$$\begin{aligned} \sqrt{2\pi b_n} B_p(n, m) &\sim \sqrt{2\pi p q n} \frac{\sqrt{2\pi n n^n e^{-n}}}{2\pi \sqrt{m(n-m)} m^m e^{-m} e^{-(n-m)}} \times \\ &\times p^m q^{n-m} = \sqrt{\frac{p q n^2}{m(n-m)}} \left(\frac{np}{m} \right)^m \left(\frac{nq}{n-m} \right)^{n-m}. \end{aligned}$$

При сделанных предположениях

$$\frac{m}{n} = p + \frac{\sqrt{b_n}}{n} x_{n,m} \sim p, \quad \frac{n-m}{n} = 1 - \frac{m}{n} \sim 1 - p = q.$$

Поэтому

$$\sqrt{2\pi b_n} B_p(n, m) \sim \left(\frac{np}{m} \right)^m \left(\frac{nq}{n-m} \right)^{n-m}.$$

Пусть $x_{n,m} = x$. Поскольку $m = np + x\sqrt{b_n}$, $n - m = nq - x\sqrt{b_n}$, то

$$\begin{aligned} \left(\frac{np}{m}\right)^m \left(\frac{nq}{n-m}\right)^{n-m} &= \left(\frac{np}{np + x\sqrt{b_n}}\right)^{np+x\sqrt{b_n}} \left(\frac{nq}{nq - x\sqrt{b_n}}\right)^{nq-x\sqrt{b_n}} = \\ &= \left(1 + x\frac{\sqrt{b_n}}{np}\right)^{-np-x\sqrt{b_n}} \left(1 - x\frac{\sqrt{b_n}}{nq}\right)^{-nq+x\sqrt{b_n}} = \theta_n(x). \end{aligned}$$

Для доказательства теоремы достаточно показать, что $\theta_n(x) \rightarrow e^{-\frac{1}{2}x^2}$, или $\ln \theta_n(x) \rightarrow -\frac{1}{2}x^2$ равномерно на каждом конечном промежутке при $n \rightarrow \infty$. Имеем

$$\begin{aligned} \ln \theta_n(x) &= -(np + x\sqrt{b_n}) \ln\left(1 + \frac{qx}{\sqrt{b_n}}\right) - \\ &- (nq - x\sqrt{b_n}) \ln\left(1 - \frac{px}{\sqrt{b_n}}\right) = -(np + x\sqrt{b_n}) \left(\frac{qx}{\sqrt{b_n}} - \frac{1}{2} \frac{q^2 x^2}{b_n} + \right. \\ &\quad \left. + O\left(\frac{q^3}{b_n^{3/2}}\right)\right) - (nq - x\sqrt{b_n}) \left(-\frac{px}{\sqrt{b_n}} - \frac{1}{2} \frac{p^2 x^2}{b_n} + O\left(\frac{p^3}{b_n^{3/2}}\right)\right) = \\ &= \frac{npq}{\sqrt{b_n}} x - x^2 q + \frac{1}{2} q x^2 + O\left(\frac{1}{\sqrt{b_n}}\right) + \frac{npq}{\sqrt{b_n}} x - p x^2 + \frac{1}{2} p x^2 = \\ &= -\frac{x^2}{2} + O\left(\frac{1}{\sqrt{b_n}}\right). \end{aligned}$$

Из этого соотношения и вытекает доказательство теоремы.

Следствие. Интегральная предельная теорема. В условиях теоремы 1 при $-\infty \leq a < b \leq \infty$

$$\lim_{n \rightarrow \infty} P\left\{a < \frac{y_n - np}{\sqrt{npq}} < b\right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}x^2} dx. \quad (4)$$

Действительно, пусть $-\infty < a < b < \infty$. Тогда

$$P\left\{a < \frac{y_n - np}{\sqrt{npq}} < b\right\} = \sum_{a < x_{n,m} < b} B_p(n, m).$$

При доказательстве теоремы 1 установлено, что при $|x_{n,m}| < c$

$$B_p(n, m) = \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{1}{2}x_{n,m}^2\right\} \left(1 + O\left(\frac{1}{\sqrt{npq}}\right)\right),$$

каково бы ни было $c > 0$. Поэтому

$$\begin{aligned} P\left\{a \leq \frac{y_n - np}{\sqrt{npq}} < b\right\} &= \left(\sum_{a < x_{n,m} < b} \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{1}{2}x_{n,m}^2\right\}\right) \times \\ &\times \left[1 + O\left(\frac{1}{\sqrt{npq}}\right)\right]. \end{aligned}$$

Поскольку $x_{n, m+1} - x_{n, m} = \frac{1}{\sqrt{npq}}$, то

$$\sum_{a < x_{n, m} < b} \frac{1}{\sqrt{npq}} \exp \left\{ -\frac{1}{2} x_{n, m}^2 \right\}$$

является интегральной суммой для функции $e^{-\frac{1}{2}x^2}$ на отрезке $[a, b]$. Поэтому

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left\{ \frac{y_n - np}{\sqrt{npq}} < b \right\} &= 1 - \lim_{n \rightarrow \infty} P \left\{ \frac{y_n - np}{\sqrt{npq}} \geq b \right\} = \\ &= 1 - \frac{1}{\sqrt{2\pi}} \int_b^{+\infty} \exp \left\{ -\frac{1}{2} x^2 \right\} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{1}{2}x^2} dx \end{aligned}$$

и стремится к $\int_a^b e^{-\frac{1}{2}x^2} dx$ при $n \rightarrow \infty$. Тем самым соотношение (4) доказано при $-\infty < a, b < \infty$. При $c < b$

$$\lim_{n \rightarrow \infty} P \left\{ \frac{y_n - np}{\sqrt{npq}} < b \right\} \geq \lim_{n \rightarrow \infty} P \left\{ c < \frac{y_n - np}{\sqrt{npq}} < b \right\} = \frac{1}{\sqrt{2\pi}} \int_c^b e^{-\frac{1}{2}x^2} dx.$$

Переходя к пределу $c \rightarrow -\infty$, найдем

$$\lim_{n \rightarrow \infty} P \left\{ \frac{y_n - np}{\sqrt{npq}} < b \right\} \geq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{1}{2}x^2} dx. \quad (5)$$

Аналогично

$$\lim_{n \rightarrow \infty} P \left\{ \frac{y_n - np}{\sqrt{npq}} \geq b \right\} \geq \frac{1}{\sqrt{2\pi}} \int_b^{\infty} e^{-\frac{1}{2}x^2} dx. \quad (6)$$

Так как

$$P \left\{ \frac{y_n - np}{\sqrt{npq}} < b \right\} + P \left\{ \frac{y_n - np}{\sqrt{npq}} \geq b \right\} = 1$$

и

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{1}{2}x^2} dx + \frac{1}{\sqrt{2\pi}} \int_b^{\infty} e^{-\frac{1}{2}x^2} dx = 1,$$

то в обеих формулах (5) и (6) возможен только знак равенства. Отсюда вытекает соотношение (4) для всех значений a и b . Формула (4) показывает, что распределение величины $\frac{y_n - np}{\sqrt{npq}}$ стремится к нормальному распределению со средним значением 0 и дисперсией 1.

Рассмотрим теперь асимптотическое поведение вероятностей $B_p(n, m)$ при условии, что $npq = D\gamma_n$ ограничено при $n \rightarrow \infty$. Тогда либо $p \rightarrow 1$, либо $q \rightarrow 1$. Так как

$$B_p(n, m) = B_q(n, n - m),$$

то достаточно рассмотреть лишь один из этих случаев. Если $q \rightarrow 1$, то $npq \sim np$.

Теорема 2. Если $a_n = np \leq c$, где c — произвольная постоянная, то для всех m

$$B_p(n, m) \sim \frac{a_n^m}{m!} e^{-a_n}. \quad (7)$$

Доказательство.

$$\begin{aligned} B_p(n, m) &= \frac{n(n-1) \dots (n-m+1)}{m!} p^{n-m} q^{n-m} = \\ &= \frac{n(n-1) \dots (n-m+1)}{n^m m!} a_n^m \left(1 - \frac{a_n}{n}\right)^{n-m} \sim \frac{a_n^m}{m!} e^{-a_n}. \end{aligned}$$

Формула (7) называется формулой Пуассона. Из нее следует, что при большом числе n испытаний число появлений события, имеющих вероятность порядка $\frac{1}{n}$, имеет приближенно распределение Пуассона. Поэтому теорема 2 называется еще законом редких событий. Обычно этому закону подчиняется число появлений некоторого события, зависящего от большого числа независимых факторов, например, число частиц, зарегистрированных счетчиком космических частиц, число катастроф, число вызовов, поступивших на телефонную станцию, и т. п.

Рассмотрим несколько примеров на применение асимптотических формул.

ЗАДАЧИ

1. Вероятность некоторого изделия быть бракованным равна $P = 0,005$. Нему равна вероятность того, что среди 10 000 наугад взятых изделий число бракованных равно 40?

Решение. Здесь имеем распределение Бернулли, причем $n = 10\,000$, $p = 0,005$. Тогда искомая вероятность равна

$$B_{0,005}(10\,000, 40) = C_{10\,000}^{40} (0,005)^{40} (0,995)^{9960}.$$

Для вычисления этой вероятности используем теорему 1. В данном случае

$b_n = \sqrt{npq} = \sqrt{10\,000 \cdot 0,005 \cdot 0,995} \approx 7,05$, $(m - np) b^{-\frac{1}{2}} = -1,42$. Следовательно, по теореме 1

$$B_{0,005}(10\,000, 40) \approx \frac{1}{\sqrt{2\pi} \cdot 7,05} e^{-\frac{1}{2}(1,42)^2} = 0,0206.$$

Заметим, что точные подсчеты дают ответ 0,0197.

2. Известно, что вероятность рождения мальчика равна приблизительно 0,515. Какова вероятность того, что среди 10 000 новорожденных число мальчиков будет не больше, чем число девочек?

Решение. Пусть Y — число мальчиков среди 10 000 новорожденных. Случайная величина Y имеет распределение Бернулли, причем $n = 10\,000$, $p = 0,515$. Надо вычислить вероятность

$$P\{Y \leq 10\,000 - Y\} = P\{Y \leq 5000\} = P\left\{\frac{Y - np}{\sqrt{npq}} \leq \frac{5000 - np}{\sqrt{npq}}\right\}.$$

В данном случае

$$\sqrt{npq} \approx 49,98, \quad \frac{5000 - 5150}{49,98} \approx -3,0012.$$

Поэтому по интегральной предельной теореме

$$P\left\{\frac{Y - np}{\sqrt{npq}} < -3,0012\right\} = \Phi(-3,0012),$$

где

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

Функция $\Phi(x)$ табулирована. Эти таблицы имеются в задачниках по теории вероятностей и сборниках статистических таблиц¹. С помощью таблицы значений функции $\Phi(x)$ находим, что $\Phi(-3,0012) = 0,0010$. Искомая вероятность равна 0,0010.

3. Вероятность попадания в цель при одном выстреле равна 0,001. Для поражения цели необходимо не менее двух попаданий. Найти вероятность поражения цели, если число выстрелов равно 5 000.

Решение. Пусть Y — число попаданий при 5 000 выстрелах. Необходимо вычислить вероятность

$$P\{Y \geq 2\} = 1 - P\{Y < 2\} = 1 - B_p(n, 0) - B_p(n, 1),$$

где $p = 0,001$, $n = 5\,000$. В рассматриваемом случае $a_n = np = 5$, и используя формулу (7), получим

$$P\{Y \geq 2\} \approx 1 - e^{-5} - 5e^{-5} = 1 - 6e^{-5} \approx 0,9596.$$

§ 7. Процесс Пуассона

Рассмотрим события, которые могут происходить в каждый момент непрерывно меняющегося времени. К таким событиям можно отнести регистрацию частицы счетчиком, поступление телефонного вызова на телефонную станцию, испускание ядром частицы при радиоактивном распаде и т. п. Если принять, что за конечное время происходит лишь конечное число событий, то при весьма общих предположениях число появлений события за любое время t имеет распределение Пуассона.

Пусть $\xi(t)$ — число появлений события на промежутке $[0, t]$. Будем считать, что $\xi(t)$ определено для всех t и является неотри-

¹ См., например, Б о л ь ш е в Л. Н., С м и р н о в С. В. Таблицы математической статистики. М., Наука, 1968; Я . Я н к о. Математико-статистические таблицы. М., Госстатиздат, 1961; М е ш а л к и н Л. Д. Сборник задач по теории вероятностей. М., Изд-во МГУ, 1963.

пательной целочисленной случайной величиной. Тогда для $t_1 < t_2$ число появлений событий на $(t_1, t_2]$ равно $\xi(t_2) - \xi(t_1)$.

Основное предположение относительно совокупности рассматриваемых событий обобщает предположение об их независимости:

1) числа событий, появившихся на непересекающихся промежутках, являются независимыми случайными величинами. Это означает, что при $0 < t_1 < t_2 < \dots < t_n$ случайные величины $\xi(t_1)$,

$$\xi(t_2) - \xi(t_1), \dots, \xi(t_n) - \xi(t_{n-1})$$

независимы между собой.

Остальные два условия обеспечивают однородность процесса появления событий во времени (это условие аналогично предположению о том, что в дискретной последовательности события имеют одинаковую вероятность) и невозможность появления одновременно нескольких событий (это свойство называется ординарностью потока событий). Получим эти условия в математической форме.

Однородность процесса означает, что число появлений событий на некотором промежутке времени зависит лишь от длины этого промежутка, т. е. распределение величины $\xi(t+h) - \xi(t)$ зависит от h , но не зависит от t . Найдем условие ординарности потока. Пусть B_t — событие, заключающееся в том, что на отрезке $[0, t]$ произошло одновременно, по крайней мере, два события. Тогда

$$B_t = \bigcap_{n=1}^{\infty} \bigcup_{k=1}^n \left\{ \xi\left(\frac{kt}{n}\right) - \xi\left(\frac{k-1}{n}t\right) > 1 \right\}. \quad (1)$$

Замечание. Если происходит событие B_t , то для всех n происходит, по крайней мере, одно из событий: на отрезке $\left[\frac{k-1}{n}t, \frac{k}{n}t\right]$ произошло не менее двух событий ($k = 1, 2, \dots, n$), т. е. $\xi\left(\frac{k}{n}t\right) - \xi\left(\frac{k-1}{n}t\right) > 1$; наоборот, если для всех n найдется такое k , что $\xi\left(\frac{k}{n}t\right) - \xi\left(\frac{k-1}{n}t\right) > 1$, т. е. на отрезке $\left[\frac{k-1}{n}t, \frac{k}{n}t\right]$ происходит два события, то найдется такая точка s , что при всех $\varepsilon > 0$ $\xi(s+\varepsilon) - \xi(s-\varepsilon) > 1$ и, следовательно, в момент s произошло не менее двух событий.

Из соотношения (1) вытекает, что

$$P(B_t) = \lim_{n \rightarrow \infty} P\left\{ \bigcup_{k=1}^n \left\{ \xi\left(\frac{kt}{n}\right) - \xi\left(\frac{k-1}{n}t\right) > 1 \right\} \right\}.$$

Сумму событий, стоящую под знаком P в правой части, можно записать как сумму несовместных событий

$$\bigcup_{k=1}^n \left[\bigcap_{i=1}^{k-1} \left\{ \xi\left(\frac{kt}{n}\right) - \xi\left(\frac{i-1}{n}t\right) \leq 1 \right\} \right] \cap \left\{ \xi\left(\frac{k}{n}t\right) - \xi\left(\frac{k-1}{n}t\right) > 1 \right\}.$$

Поскольку события под знаком $\bigcap_{i=1}^k$ независимы, то

$$\begin{aligned} P(B_t) &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \left(\prod_{i=1}^{k-1} P \left\{ \xi \left(\frac{i}{n} \right) - \xi \left(\frac{i-1}{n} \right) \leq 1 \right\} \times \right. \\ &\quad \left. \times P \left\{ \xi \left(\frac{i}{n} t \right) - \xi \left(\frac{i-1}{n} t \right) > 1 \right\} \right). \end{aligned}$$

Обозначим $\alpha(h) = P \{ \xi(t+h) - \xi(t) > 1 \}$ (эта вероятность не зависит от t). Тогда

$$\begin{aligned} P(B_t) &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \left(1 - \alpha \left(\frac{t}{n} \right) \right)^{k-1} \alpha \left(\frac{t}{n} \right) = \\ &= \lim_{n \rightarrow \infty} \left(1 - \left(1 - \alpha \left(\frac{t}{n} \right) \right)^n \right). \end{aligned}$$

Так как $P(B_t) = 0$, то $\lim_{n \rightarrow \infty} \left(1 - \alpha \left(\frac{t}{n} \right) \right)^n = 1$, но

$$\lim_{n \rightarrow \infty} \left(1 - \alpha \left(\frac{t}{n} \right) \right)^n = e^{-\lim_{n \rightarrow \infty} n \alpha \left(\frac{t}{n} \right)}.$$

Следовательно, $n \left(\alpha \left(\frac{t}{n} \right) \right) \rightarrow 0$. Таким образом, доказано, что для ординарности однородного потока необходимо выполнение условия:

$$2) \lim_{h \rightarrow 0} \frac{1}{h} P \{ \xi(t+h) - \xi(t) > 1 \} = 0.$$

Покажем теперь, что для однородного ординарного потока событий существует такое $\lambda > 0$, что

$$3) \lim_{h \rightarrow 0} \frac{1}{h} P \{ \xi(t+h) - \xi(t) > 0 \} = \lambda.$$

Действительно, обозначим $\varphi(t) = P \{ \xi(t) = 0 \}$. Тогда согласно условию 1)

$$\begin{aligned} \varphi(t+h) &= P \{ \xi(t+h) = 0 \} = P \{ \xi(t) = 0 \} \cap \{ \xi(t+h) - \xi(t) = 0 \} = \\ &= P \{ \xi(t) = 0 \} P \{ \xi(t+h) - \xi(t) = 0 \} = \\ &= P \{ \xi(t) = 0 \} P \{ \xi(h) = 0 \} = \varphi(t) \varphi(h). \end{aligned}$$

Так как $\xi(0) = 0$ и для $t_n \downarrow 0$

$$1 = \varphi(0) = P \{ \xi(0) = 0 \} = P \left\{ \bigcap_{n=1}^{\infty} \{ \xi(t_n) = 0 \} \right\} = \lim_{h \rightarrow 0} \varphi(t_n),$$

то функция $\varphi(t)$ непрерывна в точке 0 справа. Поскольку $\varphi(t+h) = \varphi(t) \varphi(h) \rightarrow \varphi(t)$ при $h \downarrow 0$, то $\varphi(t)$ непрерывна справа. Далее

$$\varphi(1) = \varphi \left(\frac{1}{n} \right)^n = \lim_{n \rightarrow \infty} \exp \left\{ n \left[\varphi \left(\frac{1}{n} \right) - 1 \right] \right\},$$

и, следовательно, существует предел

$$\lim_{n \rightarrow \infty} n \left[1 - \varphi \left(\frac{1}{n} \right) \right] = \lambda.$$

Тогда $\varphi(1) = e^{-\lambda}$, и для всех рациональных t имеем $\varphi(t) = e^{-\lambda t}$. Используя непрерывность $\varphi(t)$, убеждаемся, что эта формула справедлива для всех $t > 0$. Из соотношения

$$P\{\xi(t+h) - \xi(t) > 0\} = P\{\xi(h) > 0\} = 1 - e^{-\lambda h}$$

вытекает 3).

Теорема. Пусть для семейства целочисленных неотрицательных случайных величин $\xi(t)$ выполнено условие 1), распределение величины $\xi(t+h) - \xi(t)$ не зависит от t и выполнены условия 2) и 3). Тогда

$$P\{\xi(t) = k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \dots \quad (2)$$

Доказательство. Справедливость этой формулы для $k=0$ установлена выше.

Пусть $P_k(t) = P\{\xi(t) = k\}$. Тогда

$$\begin{aligned} P_{k+1}(t+h) &= P\{(\xi(t) < k) \cap (\xi(t+h) - \xi(t) > 1)\} + \\ &+ P\{(\xi(t) = k) \cap (\xi(t+h) - \xi(t) = 1)\} + \\ &+ P\{(\xi(t) = k+1) \cap (\xi(t+h) - \xi(t) = 0)\}. \end{aligned}$$

Так как

$$\begin{aligned} P\{\xi(t+h) - \xi(t) > 1\} &= o(h), \\ P\{\xi(t+h) - \xi(t) = 0\} &= e^{-\lambda h} = 1 - \lambda h + o(h), \\ P\{\xi(t+h) - \xi(t) = 1\} &= 1 - P\{\xi(t+h) - \xi(t) > 1\} - \\ &- P\{\xi(t+h) - \xi(t) = 0\} = \lambda h + o(h), \end{aligned}$$

то

$$P_{k+1}(t+h) = P_k(t) \lambda h + P_{k+1}(t) [1 - \lambda h] + o(h).$$

Из этого соотношения, учитывая, что $\frac{o(h)}{h} \rightarrow 0$ при $h \rightarrow 0$ равномерно относительно t , находим

$$\frac{d}{dt} P_{k+1}(t) = -\lambda P_{k+1}(t) + \lambda P_k(t). \quad (3)$$

Кроме того, если $k \geq 0$, $P_{k+1}(0) = 1$. Выразив из уравнения (3) $P_{k+1}(t)$ через $P_k(t)$, получим

$$P_{k+1}(t) = \lambda \int_0^t e^{\lambda(s-t)} P_k(s) ds.$$

Из этого соотношения по индукции, учитывая значение $P_0(t)$, убеждаемся в справедливости формулы (2).

При изучении потока событий мы ввели понятие семейства случайных величин $\xi(t)$, определенных для каждого $t \in [0, \infty)$. Другими словами, мы рассматривали функцию, определенную на некотором множестве вещественных чисел, значениями которой являются случайные величины. Такого рода функции называются случайными функциями. Как и в рассматриваемом нами случае, аргумент функции часто интерпретируется как время. Тогда такая случайная функция называется случайным процессом.

Случайный процесс, рассмотренный в этом параграфе, называется процессом Пуассона. Условие 1), которому он удовлетворяет, это условие независимости приращений процесса $\xi(t)$. Всякий процесс, для которого свойство 1) выполнено, называется процессом с независимыми приращениями. Теорема, доказанная в этом параграфе, дает условия, при которых процесс с независимыми приращениями есть процесс Пуассона.

§ 8. Суммы независимых случайных величин

Понятие независимых случайных величин было введено выше (см. § 5). Рассмотрим теперь суммы независимых случайных величин. Изучение поведения сумм большого числа независимых случайных величин — одна из важнейших задач теории вероятностей. Найдем сначала распределение суммы случайных величин через распределения отдельных слагаемых.

Теорема 1. Пусть ξ_1 и ξ_2 — независимые случайные величины с функциями распределения $F_1(x)$ и $F_2(x)$ соответственно. Тогда функция распределения $F(x)$ величины $\xi_1 + \xi_2$ задается равенством:

$$F(x) = \int_{-\infty}^{+\infty} F_1(x-y) dF_2(y) = \int_{-\infty}^{+\infty} F_2(x-y) dF_1(y). \quad (1)$$

Доказательство. Поскольку величина суммы не зависит от порядка слагаемых, то достаточно доказать первое из равенств (1). Рассмотрим два частных случая.

а) Пусть $F_2(y)$ — непрерывная функция. Тогда для всякого $h > 0$ имеет место соотношение

$$\{\xi_1 + \xi_2 < x\} = \bigcup_{k=-\infty}^{\infty} \{\xi_1 + \xi_2 < x\} \cap \{kh \leq \xi_2 < (k+1)h\},$$

но для всех k

$$\begin{aligned} \{\xi_1 < x - (k+1)h\} \cap \{kh \leq \xi_2 < (k+1)h\} &\subset \\ &\subset \{\xi_1 + \xi_2 < x\} \cap \{kh \leq \xi_2 < (k+1)h\} \subset \\ &\subset \{\xi_1 < x - kh\} \cap \{kh \leq \xi_2 < (k+1)h\}. \end{aligned}$$

Так как события в этой цепочке включений при разных k несовместны, то, записав неравенство для вероятностей и суммируя по k , получим

$$\sum_k \mathbf{P}(\{\xi_1 < x - (k+1)h\} \cap \{kh \leq \xi_2 < (k+1)h\}) \leq \\ \leq \mathbf{P}(\xi_1 + \xi_2 < x) \leq \sum_k \mathbf{P}(\{\xi_1 < x - kh\} \cap \{kh \leq \xi_2 < (k+1)h\}). \quad (2)$$

Воспользуемся независимостью величин ξ_1 и ξ_2

$$\mathbf{P}(\{\xi_1 < t\} \cap \{kh \leq \xi_2 < (k+1)h\}) = \\ = \mathbf{P}(\xi_1 < t) \mathbf{P}(kh \leq \xi_2 < (k+1)h) = F_1(t) [F_2((k+1)h) - F_2(kh)].$$

Поэтому

$$\sum_k F_1(x - (k+1)h) [F_2((k+1)h) - F_2(kh)] \leq \\ \leq \mathbf{P}(\xi_1 + \xi_2 < x) \leq \sum_k F_1(x - kh) [F_2((k+1)h) - F_2(kh)].$$

Перейдем к пределу при $h \rightarrow 0$. Тогда левая и правая суммы сходятся к интегралу $\int F_1(x - y) dF_2(y)$. Таким образом, равенство (1) выполнено.

б) Пусть ξ_2 имеет дискретное распределение $\mathbf{P}(\xi_2 = x_k) = p_k = F_2(x_k + 0) - F_2(x_k)$, $\sum p_k = 1$. Тогда

$$\mathbf{P}(\xi_1 + \xi_2 < x) = \sum_k \mathbf{P}(\{\xi_1 + \xi_2 < x\} \cap \{\xi_2 = x_k\}) = \\ = \sum_k \mathbf{P}(\xi_1 + x_k < x) \mathbf{P}(\xi_2 = x_k) = \\ = \sum_k F_1(x - x_k) [F_2(x_k + 0) - F_2(x_k)] = \int F_1(x - y) dF_2(y).$$

В общем случае $F_2(x) = p\Phi_1(x) + q\Phi_2(x)$, где $p + q = 1$, $\Phi_1(x)$ — непрерывная функция распределения, а $\Phi_2(x)$ — дискретная функция распределения. Пусть $\{x_k, k = 1, \dots\} = N$ — множество всех точек разрыва функции Φ_2 . Тогда

$$\mathbf{P}(\xi_1 + \xi_2 < x) = \mathbf{P}(\xi_1 + \xi_2 < x, \xi_2 \in N) + \mathbf{P}(\xi_1 + \xi_2 < x, \xi_2 \notin N) = \\ = \mathbf{P}(\xi_1 + \xi_2 < x | \xi_2 \in N) q + \mathbf{P}(\xi_1 + \xi_2 < x | \xi_2 \notin N) p.$$

Заметим, что условное распределение величины ξ_2 при условии $\{\xi_2 \in N\}$ будет $\Phi_2(x)$:

$$\mathbf{P}(\xi_2 < x | \xi_2 \in N) = \frac{\sum_{x_k < x} \mathbf{P}(\xi_2 = x_k)}{\mathbf{P}(\xi_2 \in N)} = \\ = \frac{1}{q} \sum_{x_k < x} q [\Phi_2(x_k + 0) - \Phi_2(x_k)] = \Phi_2(x).$$

Аналогично, условное распределение ξ_2 при условии $\{\xi_1 \in N\}$ будет $\Phi_1(x)$. Поэтому, как было установлено в случаях а) и б),

$$P\{\xi_1 + \xi_2 < x/\xi_1 \in N\} = \int F_1(x-y) d\Phi_1(y),$$

$$P\{\xi_1 + \xi_2 < x/\xi_2 \in N\} = \int F_2(x-y) d\Phi_2(y).$$

Умножая эти равенства на p и q и складывая их, получаем доказательство теоремы.

Замечание. Пусть ξ_1 имеет интегрируемую и ограниченную плотность распределения $f_1(x)$. Тогда $\xi_1 + \xi_2$ также имеет плотность распределения $f(x)$, причем

$$f(x) = \int f_1(x-y) dF_2(y). \quad (3)$$

Действительно, правая часть этого соотношения определена и интегрируема. Поэтому, меняя порядок интегрирования, что возможно вследствие неотрицательности f , получим

$$\int_{x_1}^{x_2} dx \int f_1(x-y) dF_2(y) = \int [F_1(x_2-y) - F_1(x_1-y)] dF_2(y) = F(x_2) - F(x_1).$$

Это и доказывает формулу (3).

Следствие 1. Если ξ_1 и ξ_2 имеют плотность распределения $f_1(x)$ и $f_2(x)$ соответственно, то

$$f(x) = \int f_1(x-y) f_2(y) dy. \quad (4)$$

Следствие 2. Если ξ_1 и ξ_2 независимы, имеют плотности распределения $f_1(x)$ и $f_2(x)$ соответственно, то плотность распределения $\xi_1 - \xi_2$ равна

$$f(x) = \int f_1(x+z) f_2(z) dz. \quad (5)$$

Для доказательства следует заметить, что $\xi_1 - \xi_2 = \xi_1 + (-\xi_2)$, причем величины ξ_1 и $-\xi_2$ независимы, и плотность распределения величины $-\xi_2$ равна $f_2(-x)$. Поэтому, согласно (4),

$$f(x) = \int f_1(x-y) f_2(-y) dy = \int f_1(x+z) f_2(z) dz.$$

Определение. Если $F_1(x)$ и $F_2(x)$ — две функции распределения, то функция $F(x)$, определяемая соотношением (1), называется сверткой функций распределения F_1 и F_2 , она также является функцией распределения и обозначается:

$$F = F_1 \times F_2.$$

Если f_1 и f_2 — две плотности распределения, то плотность f , определяемая равенством (4), называется сверткой плотностей f_1 и f_2 и обозначается

$$f = f_1 \times f_2.$$

Операция свертки функций распределения, по доказанному, коммутативна. Кроме того, она ассоциативна

$$F_1 \times [F_2 \times F_3] = [F_1 \times F_2] \times F_3,$$

так как обе функции распределения являются функциями распределения величины $\xi_1 + \xi_2 + \xi_3$, где ξ_1, ξ_2, ξ_3 независимы, $F_k(x) = P\{\xi_k < x\}$, поскольку величина ξ_1 не зависит от $\xi_2 + \xi_3$, а величина $\xi_1 + \xi_2$ — от ξ_3 .

$$\begin{aligned} P\{\xi_1 < x, \xi_2 + \xi_3 < y\} &= \\ &= \lim_{k \rightarrow \infty} P\left\{\left(\xi_1 < x\right) \cap \left[\bigcup_{n=-\infty}^{\infty} \left\{\frac{n}{k} \leq \xi_2 < \frac{n+1}{k}\right\} \cap \left\{\xi_3 < y - \frac{n+1}{k}\right\}\right]\right\} = \\ &= P\{\xi_1 < x\} \lim_{k \rightarrow \infty} \sum_{n=-\infty}^{\infty} P\left\{\frac{n}{k} \leq \xi_2 < \frac{n+1}{k}\right\} P\left\{\xi_3 < y - \frac{n+1}{k}\right\} = \\ &= P\{\xi_1 < x\} P\{\xi_2 + \xi_3 < y\}. \end{aligned}$$

Следовательно, выражение $F_1 \times F_2 \times \dots \times F_k$ имеет смысл для любого набора функций распределения $F_1(x), \dots, F_k(x)$, не зависит от порядка применения операции свертки и является функцией распределения суммы независимых случайных величин $\xi_1 + \dots + \xi_k$, если $F_i(x) = P\{\xi_i < x\}$. Из замечания к теореме 1 вытекает, что сумма независимых случайных величин имеет плотность распределения, когда хотя бы одно слагаемое имеет плотность. Пусть, например, плотность распределения ξ_1 равна $f_1(x)$. Тогда плотность распределения $f(x)$ суммы $\xi_1 + \dots + \xi_k$ задается формулой

$$f(x) = \int f_1(x-y) dF_2 \times \dots \times F_k(y). \quad (6)$$

Отметим особо случай целочисленных случайных величин. Пусть ξ_1, \dots, ξ_k — независимые величины и $P_i(n) = P\{\xi_i = n\}$, $\sum_n P_i(n) = 1$. Тогда $\xi_1 + \dots + \xi_k$ — также целочисленная случайная величина, при этом

$$P\{\xi_1 + \dots + \xi_k = n\} = \sum_{n_1 + \dots + n_k = n} P_1(n_1) \dots P_k(n_k). \quad (7)$$

В частности, для двух величин имеем

$$P\{\xi_1 + \xi_2 = n\} = \sum_{m=-\infty}^{\infty} P_1(m) P_2(n-m). \quad (8)$$

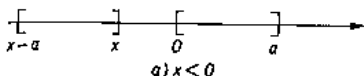
Рассмотрим несколько примеров применения полученных выше формул:

1. Предположим, что ξ_1 и ξ_2 — независимые случайные величины, имеющие распределение Пуассона с параметрами λ_1 и λ_2 соответственно. По формуле (8)

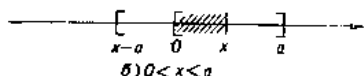
$$P\{\xi_1 + \xi_2 = n\} = \sum_{m=0}^n \frac{\lambda_1^m e^{-\lambda_1}}{m!} \frac{\lambda_2^{n-m}}{(n-m)!} e^{-\lambda_2} =$$

$$= e^{-(\lambda_1 + \lambda_2)} \frac{1}{n!} \sum_{m=0}^n \frac{n!}{m!(n-m)!} \lambda_1^m \lambda_2^{n-m} = \frac{(\lambda_1 + \lambda_2)^n}{n!} e^{-(\lambda_1 + \lambda_2)}.$$

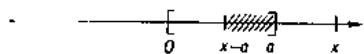
Таким образом, сумма двух независимых случайных величин, имеющих распределение Пуассона с параметрами λ_1 и λ_2 соответственно, имеет распределение Пуассона с параметром $\lambda_1 + \lambda_2$.



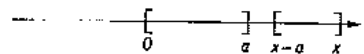
a) $x < 0$



б) $0 < x < a$



в) $a < x < 2a$



г) $x > 2a$

Рис. 16

2. Распределение суммы независимых равномерно распределенных случайных величин. Пусть случайные ве-

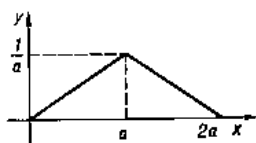


Рис. 17

личины ξ_1 и ξ_2 независимы и равномерно распределены на отрезке $[0, a]$, $a > 0$. Тогда

$$f_1(x) = f_2(x) = \begin{cases} \frac{1}{a}, & x \in (0, a), \\ 0, & x \notin (0, a). \end{cases}$$

Обозначим через $u_2(x)$ плотность распределения суммы $\xi_1 + \xi_2$. Получим

$$u_2(x) = \int f_1(x-y) f_2(y) dy = \frac{1}{a} \int_0^a f_1(x-y) dy = \frac{1}{a} \int_{x-a}^x f_1(z) dz.$$

Интеграл $\int_{x-a}^x f_1(z) dz$ представляет собой длину общей части отрезков $[0, a]$ и $[x-a, x]$, разделенную на a . Поэтому (рис. 16)

$$u_2(x) = \begin{cases} 0, & x < 0; \\ \frac{x}{a^2}, & 0 < x \leq a; \\ \frac{2a-x}{a^2}, & a < x \leq 2a; \\ 0, & x > 2a. \end{cases}$$

График плотности $u_2(x)$ представляет собой треугольник (рис. 17), и поэтому распределение с плотностью $u_2(x)$ называется треугольным распределением. Введем следующее обозначение:

$$x_+ = \frac{x + |x|}{2} = \begin{cases} x, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Тогда выражение для $u_2(x)$ можно записать в виде

$$u_2(x) = \frac{1}{a^2} [x_+ - 2(x-a)_+]. \quad (9)$$

Пусть ξ_1, \dots, ξ_n — независимые случайные величины, имеющие равномерное распределение на $[0, a]$. Положим $\eta_n = \xi_1 + \dots + \xi_n$ и найдем плотность распределения $u_n(x)$ случайной величины η_n .

Теорема 2. При $n \geq 2$ имеет место соотношение

$$u_n(x) = \frac{1}{a^n (n-1)!} \sum_{k=0}^n (-1)^k C_n^k (x - ka)_+^{n-1}. \quad (10)$$

Функция распределения $U_n(x)$ случайной величины η_n равна

$$U_n(x) = \frac{1}{a^n n!} \sum_{k=0}^n (-1)^k C_n^k (x - ka)_+^n. \quad (11)$$

Доказательство. Докажем формулу (10), используя метод математической индукции. Предположим, что (10) имеет место. Тогда функция распределения η_n вычисляется при $x > 0$ по формуле

$$U_n(x) = \int_0^x u_n(z) dz.$$

Учитывая, что

$$\int_0^x (z - ka)_+^{n-1} dz = \frac{1}{n} (x - ka)_+^n,$$

получим формулу (11). Формула (10) при $n = 2$ справедлива (в этом случае она принимает вид (9)). Далее,

$$\begin{aligned} u_{n+1}(x) &= \int u_n(x-y) f_1(y) dy = \frac{1}{a} \int_0^a u_n(x-y) dy = \\ &= \frac{1}{a} [U_n(x) - U(x-a)]. \end{aligned}$$

Используя (11), получим для $u_{n+1}(x)$ следующее выражение:

$$u_{n+1}(x) = \frac{1}{a^{n+1}n!} \left\{ \sum_{k=0}^n (-1)^k C_n^k (x - ka)_+^n + \right. \\ \left. + \sum_{k=0}^n (-1)^{k+1} C_n^k [x - (k+1)a]_+^n \right\}.$$

Во второй сумме произведем замену $k+1 = r$. Тогда

$$u_{n+1}(x) = \frac{1}{a^{n+1}n!} \left\{ \sum_{k=0}^n (-1)^k C_n^k (x - ka)_+^n + \right. \\ \left. + \sum_{r=1}^{n+1} (-1)^r C_n^{r-1} (x - ra)_+^n \right\} = \\ = \frac{1}{a^{n+1}n!} \left\{ \sum_{k=0}^n (-1)^k [C_n^k + C_n^{k-1}] (x - ka)_+^n + \right. \\ \left. + (-1)^{n+1} C_n^n (x - (n+1)a)_+^n \right\}.$$

Используя соотношение

$$C_n^k + C_n^{k-1} = C_{n+1}^k,$$

получим для $u_{n+1}(x)$ выражение, совпадающее с правой частью (10), где n заменено на $n+1$. Теорема доказана.

С помощью этой теоремы можно получить распределение суммы независимых равномерно распределенных на $[-b, b]$ случайных величин. Действительно, если случайная величина ξ_k равномерно распределена на $[0, 2b]$, то случайная величина $\xi_k - b$ равномерно распределена на $[-b, b]$, и поэтому распределение суммы

$$\gamma_k = \sum_{k=1}^n (\xi_k - b) = \sum_{k=1}^n \xi_k - nb$$

совпадает с распределением суммы n независимых случайных величин, имеющих равномерное распределение на $[-b, b]$. Так как функция распределения $\sum \xi_k$ равна $U_n(x)$ при $a = 2b$, то функция распределения γ_n равна $U_n(x + nb)$, а плотность распределения γ_n^* равна

$$u_n(x + nb) = \frac{1}{(2b)^n (n-1)!} \sum_{k=0}^n (-1)^k C_n^k (x + (n-2k)b)_+^{n-1}. \quad (12)$$

3. Гамма-распределение. Введем следующее определение:

Определение. Случайная величина ξ имеет гамма-распределение с параметрами α и β , если плотность распределения ξ равна

$$f(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x > 0; \\ 0, & x < 0. \end{cases}$$

Частным случаем гамма-распределения является показательное распределение с параметром λ (в этом случае $\alpha = 1$, $\beta = \lambda$). Докажем следующее утверждение.

Теорема 3. Если ξ_1 и ξ_2 — независимые случайные величины, имеющие гамма-распределение с параметрами (α_1, β) и (α_2, β) соответственно, то случайная величина $\eta = \xi_1 + \xi_2$ имеет гамма-распределение с параметрами $(\alpha_1 + \alpha_2, \beta)$.

Для доказательства заметим, что при $x > 0$ вследствие (4),

$$\begin{aligned} f_\eta(x) &= \int_{-\infty}^{\infty} f_1(x-y) f_2(y) dy = \\ &= \int_0^x \frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)} (x-y)^{\alpha_1-1} e^{-\beta(x-y)} \frac{\beta^{\alpha_2}}{\Gamma(\alpha_2)} y^{\alpha_2-1} e^{-\beta y} dy = \\ &= \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} e^{-\beta x} \int_0^x (x-y)^{\alpha_1-1} y^{\alpha_2-1} dy. \end{aligned}$$

Сделаем в интеграле замену переменной $\frac{y}{x} = t$. Тогда

$$f_\eta(x) = \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} x^{\alpha_1+\alpha_2-1} e^{-\beta x} \int_0^1 t^{\alpha_2-1} (1-t)^{\alpha_1-1} dt.$$

Так как

$$\int_0^1 t^{\alpha_2-1} (1-t)^{\alpha_1-1} dt = B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1) \Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)},$$

то при $x > 0$

$$f_\eta(x) = \frac{\beta^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1 + \alpha_2)} x^{\alpha_1+\alpha_2-1} e^{-\beta x},$$

а это есть плотность гамма-распределения с параметрами $\alpha_1 + \alpha_2$ и β .

Гамма-распределение играет важную роль в теории массового обслуживания и математической статистике.

4. Распределение Эрланга. Распределением Эрланга с параметрами n и λ называется распределение суммы n независимых

случайных величин, каждая из которых имеет показательное распределение с параметром λ .

По теореме 3 распределение Эрланга представляет собой гамма-распределение с параметрами $\alpha = n$, $\beta = \lambda$.

5. χ^2 -распределение с n степенями свободы. Распределением χ^2 с n степенями свободы называется распределение случайной величины

$$\chi^2 = \xi_1^2 + \dots + \xi_n^2,$$

где ξ_k — независимые одинаково распределенные случайные величины, имеющие нормальное $N(0, 1)$ распределение.

Для вычисления плотности распределения случайной величины χ^2 заметим, что случайная величина $\eta = \xi^2$, где ξ — нормально $N(0, 1)$ распределенная случайная величина, имеет гамма-распределение с параметрами $\alpha = \frac{1}{2}$ и $\beta = \frac{1}{2}$. В самом деле, при $x > 0$

$$P\{\xi^2 < x\} = P\{|\xi| < \sqrt{x}\} = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-\frac{u^2}{2}} du,$$

и, следовательно, плотность распределения ξ^2 равна

$$\frac{1}{\sqrt{2\pi}} x^{\frac{1}{2}-1} e^{-\frac{x}{2}},$$

что представляет собой плотность гамма-распределения с параметрами $\alpha = \frac{1}{2}$, $\beta = \frac{1}{2}$. По теореме 3 распределение случайной величины χ^2 есть гамма-распределение с параметрами $\alpha = \frac{n}{2}$, $\beta = \frac{1}{2}$. Следовательно, плотность распределения случайной величины χ^2 равна

$$f_{\chi^2}(x) = \begin{cases} \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}, & x > 0; \\ 0, & x < 0. \end{cases}$$

В статистической механике величина

$$\chi_3^2 = \xi_1^2 + \xi_2^2 + \xi_3^2$$

появляется как квадрат скорости частицы в предположении, что проекции скорости на оси имеют нормальное распределение. Соответствующее распределение называется распределением Максвелла.

6. Распределение Лапласа. Пусть ξ_1 и ξ_2 — независимые одинаково распределенные случайные величины, имеющие показательное распределение с параметром λ . Тогда, используя соотношение

(5), получим, что плотность распределения случайной величины $\eta = \xi_1 - \xi_2$ равна

$$f_{\eta}(x) = \frac{\lambda^2}{2} e^{-\lambda(x)}.$$

Распределение с такой плотностью называется распределением Лапласа.

7. Распределение произведения n независимых равномерно распределенных случайных величин.

Пусть ξ_1, \dots, ξ_n — независимые случайные величины, имеющие равномерное распределение на $[0, 1]$. Найдем плотность распределения случайной величины

$$\gamma_n = \xi_1 \xi_2 \dots \xi_n.$$

Рассмотрим случайную величину

$$\eta_n = -\ln \gamma_n = \sum_{k=1}^n (-\ln \xi_k) = \sum_{k=1}^n \ln \frac{1}{\xi_k}.$$

Если ξ_k имеет равномерное распределение на $[0, 1]$, то случайная величина $-\ln \xi_k = \ln \frac{1}{\xi_k}$ распределена по показательному закону с параметром $\lambda = 1$ (см. пример 3, с. 72). Следовательно, случайная величина η_n имеет распределение Эрланга с параметрами n и $\lambda = 1$. Это означает, что плотность распределения η_n при $x > 0$ равна

$$f_{\eta_n}(x) = \frac{1}{\Gamma(n)} x^{n-1} e^{-x} = \frac{1}{(n-1)!} x^{n-1} e^{-x}.$$

Далее заметим, что при $0 < x < 1$

$$\begin{aligned} F_{\gamma_n}(x) &= P\{\gamma_n < x\} = P\{e^{-\eta_n} < x\} = P\{\eta_n > -\ln x\} = \\ &= 1 - P\left\{\eta_n < \ln \frac{1}{x}\right\}; \end{aligned}$$

следовательно, при $0 < x < 1$

$$f_{\gamma_n}(x) = f_{\eta_n}(x) \frac{1}{x} = \frac{1}{(n-1)!} \left(\ln \frac{1}{x}\right)^{n-1};$$

при $x < 0$ или $x > 1$

$$f_{\gamma_n}(x) = 0.$$

§ 9. Характеристические и производящие функции

Характеристические параметры. При изучении сумм независимых случайных величин используют характеристические функции случайных величин.

Определение. Характеристической функцией $\varphi(z)$ случайной величины ξ называется комплексно-значная функция, определенная при $z \in R$ соотношением

$$\varphi(z) = M e^{iz\xi} = M(\cos z\xi + i \sin z\xi). \quad (1)$$

Если $F(x)$ — функция распределения величины ξ , то

$$\varphi(z) = \int_{-\infty}^{+\infty} e^{izx} dF(x). \quad (2)$$

Существование интеграла (и математического ожидания), определяющего характеристическую функцию, вытекает из непрерывности функции e^{izx} и ее ограниченности: $|e^{izx}| = 1$.

Рассмотрим несколько примеров вычисления характеристических функций.

ПРИМЕРЫ

1. Пусть

$$\xi = \begin{cases} 1 & \text{с вероятностью } \frac{1}{2}, \\ -1 & \text{с вероятностью } \frac{1}{2}. \end{cases}$$

Тогда

$$\varphi(z) = Me^{iz\xi} = \frac{1}{2} (e^{iz} + e^{-iz}) = \cos z.$$

2. Пусть ξ — случайная величина, имеющая распределение Пуассона с параметром λ . Тогда

$$\varphi(z) = Me^{iz\xi} = \sum_{k=0}^{\infty} e^{izk} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{iz})^k}{k!} = e^{\lambda(e^{iz}-1)}.$$

3. Пусть случайная величина ξ равномерно распределена на $[a, b]$. Тогда

$$\varphi(z) = \frac{1}{b-a} \int_a^b e^{izx} dx = \frac{1}{b-a} \frac{e^{izb} - e^{iza}}{iz}.$$

В частности, если ξ равномерно распределена на $[-a, a]$, то

$$\varphi(z) = \frac{\sin az}{az}.$$

4. Характеристическая функция нормального распределения. Пусть ξ — случайная величина, имеющая нормальное распределение $N(0, 1)$. Тогда

$$\psi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{izx - \frac{x^2}{2}} dx.$$

Заметим, что

$$\varphi'(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ixe^{izx - \frac{x^2}{2}} dx$$

(дифференцирование под знаком интеграла по z законно, так как интеграл

$$\int ixe^{izx - \frac{x^2}{2}} dx$$

сходится равномерно относительно $z \in (-\infty, \infty)$). Интегрируя по частям, получим

$$\begin{aligned}\varphi'(z) &= -\frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{izx} d\left(e^{-\frac{x^2}{2}}\right) = \\ &= -\frac{i}{\sqrt{2\pi}} e^{izx} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} - \frac{z}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{izx} e^{-\frac{x^2}{2}} dx = -z\varphi(z).\end{aligned}$$

Следовательно, $\varphi(z)$ удовлетворяет дифференциальному уравнению

$$\varphi'(z) = -z\varphi(z),$$

решая которое, находим

$$\varphi(z) = Ce^{-\frac{z^2}{2}},$$

где C — некоторая константа. Поскольку $\varphi(0) = 1$, то

$$\varphi(z) = e^{-\frac{z^2}{2}}.$$

Пусть ξ имеет нормальное распределение $N(a, \sigma)$. Рассмотрим случайную величину

$$\eta = \frac{\xi - a}{\sigma}.$$

Тогда η имеет нормальное распределение $N(0, 1)$ и $\xi = \sigma\eta + a$. Характеристическая функция ξ равна

$$\varphi_{\xi}(z) = Me^{iz\xi} = e^{iza} Me^{iza\eta} = e^{iza - \frac{\sigma^2 z^2}{2}}.$$

Итак, мы установили следующий важный факт, которым в дальнейшем будем неоднократно пользоваться: *характеристическая функция случайной величины с нормальным распределением $N(a, \sigma^2)$ равна*

$$\varphi(z) = e^{iza - \frac{\sigma^2 z^2}{2}}.$$

5. Пусть ξ имеет показательное распределение с параметром λ . Тогда

$$\varphi(z) = \int_0^{+\infty} e^{izx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - iz}.$$

6. Характеристическая функция распределения Коши. Пусть ξ — случайная величина с плотностью распределения

$$p_{\xi}(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2}.$$

Вычислим характеристическую функцию

$$\varphi(z) = \frac{a}{\pi} \int_{-\infty}^{\infty} e^{izx} \frac{1}{a^2 + x^2} dx = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{\cos zx}{a^2 + x^2} dx.$$

Заметим, что $\varphi(z)$ — четная функция, и найдем $\varphi(z)$ при $z > 0$. Дифференцируя под знаком интеграла по z , получим

$$\varphi'(z) = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{-x \sin zx}{a^2 + x^2} dx \quad (3)$$

Воспользуемся известным из курса математического анализа интегралом

$$\pi = \int_{-\infty}^{\infty} \frac{\sin zx}{x} dx \quad (z > 0) \quad (4)$$

(интеграл Дирихле).

Запишем равенство (4) в виде

$$a = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{\sin zx}{x} dx \quad (5)$$

и сложим почленно (3) и (5). Тогда

$$\varphi'(z) + a = \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{a^2}{a^2 + x^2} \frac{\sin zx}{x} dx.$$

Дифференцируя по z под знаком интеграла, получим дифференциальное уравнение для $\varphi(z)$

$$\varphi''(z) = a^2 \varphi(z).$$

Следовательно,

$$\varphi(z) = C_1 e^{az} + C_2 e^{-az}.$$

Так как $\varphi(z)$ — ограниченная функция на всей числовой оси, то $C_1 = 0$. Из условия $\varphi(0) = 1$ получаем $C_2 = 1$. Итак, при $z > 0$ $\varphi(z) = e^{-az}$. Так как $\varphi(z)$ — четная функция, то

$$\varphi(z) = e^{-a|z|}.$$

Предоставляем читателю возможность самостоятельно убедиться в том, что, если плотность распределения ξ равна

$$p_{\xi}(x) = \frac{a}{\pi} \frac{1}{a^2 + (x - \theta)^2},$$

то характеристическая функция ξ равна

$$\varphi(z) = e^{i\theta z - a|z|}.$$

Отметим следующие свойства характеристических функций, вытекающие из равенств (1) и (2).

1) $\varphi(0) = 1$, $|\varphi(z)| \leq 1$ для всех вещественных z .

2) $\varphi(-z) = \overline{\varphi(z)}$.

Доказательство следует из равенств

$$\varphi(-z) = M e^{-iz\xi} = \overline{M e^{iz\xi}} = \overline{\varphi(z)}.$$

3) $\varphi(z)$ равномерно непрерывная функция на всей числовой оси. Для доказательства заметим, что

$$\begin{aligned} |\varphi(z_1) - \varphi(z_2)| &\leq \int_{-\infty}^{\infty} |e^{iz_1x} - e^{iz_2x}| dF(x) = \\ &= \int_{-\infty}^{\infty} |e^{ix(z_1 - z_2)} - 1| dF(x) = \\ &= \int_{|x| < A} |e^{ix(z_1 - z_2)} - 1| dF(x) + 2 \int_{|x| > A} dF(x) = \\ &= \int_{|x| < A} |e^{ix(z_1 - z_2)} - 1| dF(x) + 2P\{|\xi| > A\} = I_1 + I_2. \end{aligned}$$

Для заданного $\varepsilon > 0$ выберем сначала A столь большим, чтобы $I_2 < \frac{\varepsilon}{2}$. Далее зафиксируем A и выберем $\delta(\varepsilon)$ такое, что $|e^{ix(z_1 - z_2)} - 1| < \frac{\varepsilon}{2}$ при $|z_1 - z_2| < \delta(\varepsilon)$; тогда $I_1 < \frac{\varepsilon}{2}$ и $|\varphi(z_1) - \varphi(z_2)| < \varepsilon$ при $|z_1 - z_2| < \delta(\varepsilon)$.

4) Для любого n , любых вещественных z_1, \dots, z_n и любых комплексных чисел c_1, \dots, c_n

$$\sum_{k=1}^n \sum_{r=1}^n \varphi(z_k - z_r) c_k \bar{c}_r \geq 0.$$

Действительно,

$$\begin{aligned} \sum_{k=1}^n \sum_{r=1}^n M e^{i(z_k - z_r)\xi} c_k \bar{c}_r &= M \left(\sum_{k=1}^n c_k e^{iz_k \xi} \right) \left(\sum_{r=1}^n \bar{c}_r e^{-iz_r \xi} \right) = \\ &= M \left| \sum_{k=1}^n c_k e^{iz_k \xi} \right|^2 \geq 0. \end{aligned}$$

5) Если существует момент порядка n , а именно $M|\xi|^n$, то функция $\varphi(z)$ имеет n непрерывных производных и

$$\varphi^{(n)}(0) = i^n M \xi^n. \quad (6)$$

Предположим, что существует $M|\xi|$. Рассмотрим интеграл.

$$\int_{-\infty}^{\infty} i x e^{izx} dF(x), \quad (7)$$

полученный формальным дифференцированием в (2) под знаком интеграла. Так как

$$\int_{-\infty}^{\infty} |x| dF(x) < +\infty,$$

то интеграл (7) сходится равномерно относительно z . Поэтому возможно дифференцирование под знаком интеграла по z и

$$\varphi'(z) = i \int_{-\infty}^{\infty} x e^{izx} dF(x), \quad \varphi'(0) = iM\xi.$$

Далее можно воспользоваться методом математической индукции.

б) *Характеристическая функция однозначно определяет распределение случайной величины.*

Это свойство будет доказано ниже.

Одно из **важнейших** свойств характеристических функций сформулируем в виде теоремы.

Теорема 1. Если ξ_1 и ξ_2 — независимые случайные величины, а $\varphi_1(z)$ и $\varphi_2(z)$ — их характеристические функции, то характеристическая функция суммы $\xi_1 + \xi_2$ равна произведению $\varphi_1(z) \times \varphi_2(z)$.

Доказательство. Характеристическую функцию $\varphi(z)$ суммы можно записать в виде

$$\begin{aligned} \varphi(z) &= M[\cos z(\xi_1 + \xi_2) + i \sin z(\xi_1 + \xi_2)] = \\ &= M[\cos z\xi_1 \cos z\xi_2 - \sin z\xi_1 \sin z\xi_2 + i \sin z\xi_1 \cos z\xi_2 + \\ &\quad + i \cos z\xi_1 \sin z\xi_2]. \end{aligned}$$

Как было доказано, для независимых величин ξ_1, ξ_2 и ограниченных непрерывных функций $g_1(x), g_2(x)$ справедливо равенство

$$Mg(\xi_1)g(\xi_2) = Mg(\xi_1)Mg(\xi_2).$$

Поэтому

$$\begin{aligned} \varphi(z) &= M \cos z\xi_1 M \cos z\xi_2 - M \sin z\xi_1 M \sin z\xi_2 + \\ &\quad + i M \sin z\xi_1 M \cos z\xi_2 + i M \cos z\xi_1 M \sin z\xi_2 = \\ &= (M \cos z\xi_1 + i M \sin z\xi_1) (M \cos z\xi_2 + i M \sin z\xi_2) = \varphi_1(z) \varphi_2(z). \end{aligned}$$

Следствие. Если $\xi_1, \xi_2, \dots, \xi_k$ — независимые случайные величины, то характеристическая функция суммы $\xi_1 + \dots + \xi_k$ равна произведению характеристических функций слагаемых.

В качестве применения теоремы 1 докажем следующее утверждение:

Теорема 2. Пусть ξ_1 и ξ_2 — независимые нормально $N(a_1, \sigma_1^2)$ и $N(a_2, \sigma_2^2)$ распределенные случайные величины. Тогда $\xi_1 + \xi_2$ имеет нормальное $N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$ распределение.

По теореме 1 характеристическая функция $\xi_1 + \xi_2$ равна

$$e^{ia_1z - \frac{\sigma_1^2 z^2}{2}} e^{ia_2z - \frac{\sigma_2^2 z^2}{2}} = e^{i(a_1+a_2)z - \frac{1}{2}z^2(\sigma_1^2 + \sigma_2^2)}.$$

Это есть характеристическая функция нормального $N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$ распределения. Принимая во внимание, что каждой харак-

теристической функции соответствует только одно распределение вероятностей, получаем утверждение теоремы 2.

Дальше в этом параграфе мы будем рассматривать целочисленные величины. Пусть ξ — такая величина и $p_k = P\{\xi = k\}$. Тогда характеристическая функция величины ξ определяется равенством

$$\varphi(z) = \sum_{k=-\infty}^{+\infty} p_k e^{ikz}.$$

Умножая это равенство на e^{-imz} и интегрируя ряд почленно (он сходится равномерно, так как $|p_k e^{ikz}| = p_k$, $\sum_k p_k = 1$) на $[-\pi, \pi]$, получим

$$p_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(z) e^{-imz} dz \quad (8)$$

(мы воспользовались тем, что $\int_{-\pi}^{\pi} e^{ilz} dz = 0$ при $l \neq 0$).

Соотношения (8) показывают, что распределение целочисленной случайной величины полностью определяется заданием характеристической функции на интервале $[-\pi, \pi]$.

Производящие функции. Для неотрицательных целочисленных величин удобно пользоваться производящими функциями.

Определение. Пусть ξ — целочисленная неотрицательная величина, для которой $P\{\xi = k\} = p_k$, $k \geq 0$. Производящей функцией $\psi(\lambda)$ величины ξ называется функция

$$\psi(\lambda) = M\lambda^\xi = \sum_{k=0}^{\infty} p_k \lambda^k, \quad (9)$$

определенная при комплексных λ , для которых $|\lambda| \leq 1$.

Поскольку $|e^{iz}| \leq 1$, то определена величина $\psi(e^{iz})$. Легко видеть, что $\psi(e^{iz}) = \varphi(z)$, где $\varphi(z)$ — характеристическая функция величины ξ . Поэтому производящая функция суммы независимых целочисленных случайных величин равна произведению производящих функций отдельных слагаемых.

Производящие функции можно определить для любой последовательности $\{a_n, n \geq 0\}$, полагая

$$\psi(\lambda) = \sum_{n=0}^{\infty} a_n \lambda^n \quad (10)$$

для тех комплексных λ , для которых ряд справа сходится. Как известно, степенной ряд сходится в круге комплексной плоскости с центром в точке O , радиус которого определяется формулой Коши

$$r = (\overline{\lim} \sqrt[n]{|a_n|})^{-1}. \quad (11)$$

С помощью производящей функции можно определять члены последовательности $\{a_n\}$

$$a_n = \frac{1}{2\pi r^n} \int_{-\pi}^{\pi} \psi(\rho e^{iz}) e^{inz} dz, \quad (12)$$

где $\rho < r$. Использование производящих функций часто основывается на следующей лемме.

Лемма. Пусть $\{a_n\}$ и $\{b_n\}$ — две последовательности, производящие функции которых $\psi_1(\lambda)$ и $\psi_2(\lambda)$ определены в некотором круге ненулевого радиуса. Тогда $\psi_1(\lambda)\psi_2(\lambda)$ является производящей функцией последовательности

$$c_n = a_0 b_n + a_1 b_{n-1} + \dots + a_n b_0, \quad (13)$$

называемой *сверткой* последовательностей $\{a_n\}$ и $\{b_n\}$.

Доказательство леммы вытекает из правила умножения степенных рядов.

Зная производящую функцию случайной величины ξ , можно вычислить все моменты случайной величины. Докажем следующее утверждение:

Теорема 3. Пусть ξ имеет математическое ожидание. Тогда производящая функция $\psi(\lambda)$ величины ξ дифференцируема на отрезке $[-1, 1]$ и $M\xi = \psi'(1)$.

Действительно, $M\xi = \sum_{k=0}^{\infty} k p_k < +\infty$. Поэтому ряд

$$\sum k \lambda^{k-1} p_k,$$

полученный почленным дифференцированием (9), равномерно сходится. Следовательно, почленное дифференцирование законно и

$$\psi'(1) = M\xi.$$

Рассмотрим несколько примеров вычисления производящих функций:

ПРИМЕРЫ

1. Пусть γ — число успехов при n испытаниях в схеме Бернулли с вероятностью успеха p . Тогда

$$\psi(\lambda) = M\lambda^\gamma = \sum_{k=0}^n \lambda^k C_n^k p^k q^{n-k} = (q + p\lambda)^n.$$

2. Предположим, что ξ имеет геометрическое распределение с параметром p . Тогда $P(\xi = n) = q^n p$, $n = 0, 1, \dots$ и

$$\psi(\lambda) = M\lambda^\xi = \sum_{n=0}^{\infty} \lambda^n q^n p = \frac{p}{1 - q\lambda}.$$

3. Однородное случайное блуждание на прямой. Рассмотрим случайное блуждание частицы по целочисленным точкам числовой

прямой, описываемое следующей схемой: в начальный момент времени $t=0$ частица находится в точке $x=0$, а затем в моменты времени $t=1, 2, \dots, n, \dots$ совершает скачки на единицу вправо с вероятностью p , или на единицу влево с вероятностью q .

Рассмотрим такие вероятности:

f_n — вероятность того, что частица, выйдя из точки $x=0$, впервые на n -м шаге вернется в исходную точку;

u_n — вероятность того, что частица на n -м шаге вернется в исходную точку;

f_n^+ — вероятность того, что частица, выйдя из точки $x=0$, впервые на n -м шаге попадет в точку $x=1$;

f_n^- — вероятность того, что частица, выйдя из точки $x=0$, впервые на n -м шаге попадет в точку $x=-1$.

Найдем производящие функции последовательностей $\{f_n^+\}$, $\{f_n^-\}$, $\{f_n\}$, $\{u_n\}$. Заметим, что

$$f_1^+ = p \quad (14)$$

и

$$f_n^+ = q \sum_{k=0}^{n-1} f_k^+ f_{n-1-k}^+, \quad n > 1, \quad (15)$$

так как при $n > 1$ первое попадание в точку $x=1$ может произойти следующим образом: первый шаг происходит влево (вероятность этого q), затем на каком-то k -м шаге частица впервые попадает в точку $x=0$ и на $(n-1-k)$ -м шаге впервые попадает в точку $x=1$ (вероятность этого равна $f_k^+ f_{n-1-k}^+$). Пусть

$$F^+(\lambda) = \sum_{k=1}^{\infty} f_k^+ \lambda^k.$$

Тогда, умножая равенство (14) на λ , а равенство (15) — на λ^n и суммируя по всем $n=1, 2, \dots$, получим, принимая во внимание утверждение леммы, следующее равенство:

$$F^+(\lambda) = p\lambda + q\lambda [F^+(\lambda)]^2. \quad (16)$$

Таким образом,

$$F^+(\lambda) = \frac{1 - \sqrt{1 - 4pq\lambda^2}}{2q\lambda}. \quad (17)$$

Второй корень квадратного уравнения не ограничен при $\lambda \rightarrow 0$ и поэтому не является производящей функцией. Производящую функцию $F^-(\lambda)$ можно получить, меняя местами p и q . Поэтому

$$F^-(\lambda) = \frac{1 - \sqrt{1 - 4pq\lambda^2}}{2p\lambda}. \quad (18)$$

Заметим, что по формуле полной вероятности

$$f_n = pf_{n-1}^- + qf_{n-1}^+. \quad (19)$$

Умножая равенство (19) на λ^n и суммируя по всем $n=1, 2, \dots$, получим

$$F(\lambda) = p\lambda F^-(\lambda) + q\lambda F^+(\lambda) = 1 - \sqrt{1 - 4pq\lambda^2}. \quad (20)$$

Чтобы найти

$$U(\lambda) = \sum_{n=0}^{\infty} u_n \lambda^n,$$

нужно учесть, что

$$u_n = f_n + \sum_{k=0}^{n-1} f_k u_{n-k} = \sum_{k=0}^n f_k u_{n-k}.$$

Умножив это равенство на z^n и суммируя по всем n , получим справа $U(\lambda) F(\lambda)$ (по лемме о производящей функции свертки), а слева $U(\lambda) - 1$. Поэтому

$$U(\lambda) = \frac{1}{1 - F(\lambda)} = (1 - 4pq\lambda^2)^{-\frac{1}{2}}.$$

Заметим, что

$$F(1) = \sum_{n=1}^{\infty} f_n = 1 - \sqrt{1 - 4pq} = 1 - \sqrt{(p+q)^2 - 4pq} = 1 - |p - q|$$

представляет собой вероятность того, что частица когда-либо вернется в исходную точку. Следовательно, при $p = q = \frac{1}{2}$ вероятность возвращения в исходную точку равна 1, а при $p \neq q$ существует положительная вероятность того, что частица не вернется в исходное положение. Вероятность f^+ того, что частица когда-либо вернется в соседнюю правую точку, равна

$$\sum_{n=1}^{\infty} f_n^+ = F^+(1) = \frac{1 - |p - q|}{2q} = \begin{cases} 1, & p \geq q, \\ \frac{p}{q}, & p < q. \end{cases}$$

Математическое ожидание числа шагов до первого возвращения частицы в исходное положение равно

$$F'(1) = \frac{4pq}{|p - q|}.$$

Следовательно, в симметричном ($p = q = \frac{1}{2}$) случайном блуждании среднее число шагов до первого возвращения бесконечно.

4. Суммы случайного числа случайных величин. Пусть $\{\eta_i\}$ — последовательность взаимно независимых одинаково распределенных случайных величин с производящей функцией $P(\lambda)$. Предположим, что N — случайная величина, принимающая значения $0, 1, \dots, n, \dots$ и не зависящая от случайных величин η_i . Пусть

$$q_k = P\{N = k\}, \quad q(\lambda) = \sum_{k=0}^{\infty} q_k \lambda^k.$$

Рассмотрим случайную величину

$$\xi = \eta_1 + \dots + \eta_N.$$

Теорема 4. Производящая функция случайной величины ξ равна $q(P(\lambda))$.

Действительно, по формуле полной вероятности

$$\begin{aligned} P\{\xi = j\} &= \sum_{n=0}^{\infty} P\{N = n\} P\{\eta_1 + \dots + \eta_n = j \mid N = n\} = \\ &= \sum_{n=0}^{\infty} P\{N = n\} P\{\eta_1 + \dots + \eta_n = j\}. \end{aligned}$$

Умножим эти равенства на λ^j и просуммируем по всем j

$$\sum_{j=0}^{\infty} \lambda^j P(\xi = j) = \sum_{n=0}^{\infty} g_n \left(\sum_{j=0}^{\infty} \lambda^j P(\eta_1 + \dots + \eta_n = j) \right),$$

но

$$\sum_{j=0}^{\infty} \lambda^j P(\eta_1 + \dots + \eta_n = j)$$

есть производящая функция суммы n независимых величин и равна, следовательно, $P^n(\lambda)$. Таким образом,

$$\sum_{j=0}^{\infty} \lambda^j P(\xi = j) = \sum_{n=0}^{\infty} g_n P^n(\lambda) = g(P(\lambda)).$$

5. Ветвящийся процесс Гальтона—Ватсона. Рассмотрим частицы, способные производить новые такие же частицы. Некоторая частица образует нулевое поколение; порожденные ею частицы образуют первое поколение, потомки частиц первого поколения образуют второе поколение и т. д. Предположим, что каждая частица с вероятностью p_k , $k=0, 1, \dots$, порождает k новых частиц, и что частицы каждого поколения размножаются независимо друг от друга.

Пусть $P(\lambda) = \sum_{k=0}^{\infty} p_k \lambda^k$ — производящая функция распределения $\{p_k\}$, ξ_n — число частиц в n -м поколении, $P_n(\lambda)$ — производящая функция ξ_n . Тогда

$$\xi_n = \eta_1 + \dots + \eta_{\xi_{n-1}},$$

где η_k — число потомков k -й частицы из $n-1$ поколения. По теореме 5

$$P_n(\lambda) = P_{n-1}(P(\lambda)).$$

Производящую функцию ξ_n можно также получить, используя следующее представление

$$\xi_n = \gamma_1 + \dots + \gamma_{\xi_1},$$

где γ_k — число тех частиц n -го поколения, которые являются потомками k -й частицы первого поколения. Заметим, что производящая функция γ_k равна $P_{n-1}(\lambda)$. Следовательно, согласно теореме 5, $P_n(\lambda) = P(P_{n-1}(\lambda))$.

Пусть A_n — событие, состоящее в том, что в n -м поколении или ранее уже нет частиц; пусть также x_n — вероятность события A_n . Поскольку $A_n \subset A_{n+1}$, то $x_n \leq x_{n+1}$ и

$$P(\cup A_n) = \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} x_n = x$$

есть вероятность вырождения процесса. Заметим, что

$$x_1 = P(0) = p_0, \quad x_n = P_n(0) = P(P_{n-1}(0)) = P(x_{n-1}).$$

Таким образом, вероятность вырождения x является корнем уравнения

$$x = P(x). \quad (21)$$

Легко убедиться, что x — наименьший корень этого уравнения. Действительно, $x_1 = p_0 = P(0) < P(u) = u$, где u — некоторый корень уравнения (21). Если предположить, что $x_n < u$, то

$$x_{n+1} = P(x_n) < P(u) = u,$$

и для всех n получим

$$x_n \leq u.$$

Поэтому $x = \lim_{n \rightarrow \infty} x_n \leq u$.

График функции $y = P(x)$ — выпуклая книзу кривая ($y'' = P''(x) \geq 0$). Поэтому прямая $y = x$ пересекает эту кривую не более, чем в двух точках. Одной такой точкой является точка $(1, 1)$, так как $P(1) = 1$, и поэтому уравнение (21) имеет не более одного корня на интервале $(0, 1)$.

Если уравнение (21) имеет корень x , $0 < x < 1$, то по теореме Лагранжа найдется точка c , $x < c < 1$, такая, что

$$\frac{1 - P(x)}{1 - x} = P'(c) = 1 < P'(1).$$

Следовательно, уравнение (21) имеет корень $x < 1$ только тогда, когда $P'(1) > 1$. Если $P'(1) \leq 1$, то

$$\frac{1 - P(s)}{1 - s} = P'(c) < P'(1) \leq 1, \\ 1 - P(s) < 1 - s, \quad P(s) > s,$$

при всех $0 < s < 1$. Это означает, что на отрезке $[0, 1]$ график функции $y = P(x)$ лежит выше прямой $y = x$, и поэтому минимальный корень уравнения (21) равен 1. Заметим, что $P'(1) = \sum k p_k$ есть математическое ожидание потомков одной частицы.

Итак, доказано следующее утверждение:

Теорема 5. Пусть $\mu = \sum_{k=0}^{\infty} k p_k$ — математическое ожидание числа непосредственных потомков одной частицы. Если $\mu \leq 1$, то вероятность x того, что процесс оборвется, равна 1. Если $\mu > 1$, то уравнение (21) имеет единственный корень x на интервале $(0, 1)$, и x есть вероятность того, что процесс оборвется после конечного числа поколений.

Найдем математическое ожидание числа частиц в n -м поколении. Имеем

$$M_{E_n} = P_n'(1) = P'(1) P_{n-1}'(1) = \mu M_{E_{n-1}},$$

где $\mu = M_{E_1}$. Следовательно, $M_{E_n} = \mu^n$.

Изложенная задача впервые была рассмотрена Гальтоном в 1889 году под названием задачи о выживании фамилии. В ней учитываются лишь потомки мужского рода, играющие роль частиц; p_k есть вероятность того, что новорожденный мальчик произведет в течение своей жизни k потомков мужского рода.

Решетчатые распределения. Сделаем несколько замечаний о решетчатых распределениях. Будем говорить, что дискретная случайная величина ξ имеет решетчатое распределение, если существуют числа a и h ($h > 0$) такие, что все возможные значения ξ могут быть представлены в виде $a + kh$, где k принимает любые целые значения. Число h называют шагом распределения.

Условие решетчатости распределения можно выразить в терминах характеристической функции.

Теорема 6. Для того чтобы случайная величина ξ имела решетчатое распределение, необходимо и достаточно, чтобы при некотором $z \neq 0$ модуль ее характеристической функции был равен единице.

Действительно, если случайная величина ξ имеет решетчатое распределение, причем $p_k = P(\xi = a + kh)$, то характеристическая функция φ равна

$$\varphi(z) = \sum_{k=-\infty}^{\infty} p_k e^{iz(a+kh)} = e^{iaz} \sum_{k=-\infty}^{\infty} p_k e^{ikhz}.$$

Поэтому

$$\varphi\left(\frac{2\pi}{h}\right) = e^{ia \frac{2\pi}{h}} \sum_{k=-\infty}^{\infty} p_k e^{i2\pi k} = e^{ia \frac{2\pi}{h}}.$$

Следовательно,

$$\left| \varphi\left(\frac{2\pi}{h}\right) \right| = 1.$$

Предположим, что при некотором $z_1 \neq 0$ $|\varphi(z_1)| = 1$. Убедимся, что тогда ξ обязательно имеет решетчатое распределение. Так как $\varphi(z_1) = e^{iz_1 a}$, где a — некоторое вещественное число, то

$$\int_{-\infty}^{\infty} e^{iz_1 x} dF(x) = e^{iz_1 a},$$

откуда вытекает, что

$$\begin{aligned} \int_{-\infty}^{\infty} \cos(z_1 x - \alpha) dF(x) &= 1, \\ \int_{-\infty}^{\infty} [1 - \cos(z_1 x - \alpha)] dF(x) &= 0, \\ \int_{-\infty}^{\infty} \sin^2 \frac{z_1 x - \alpha}{2} dF(x) &= 0. \end{aligned}$$

Последнее равенство возможно лишь тогда, когда функция $F(x)$ возрастает только при тех значениях x , при которых

$$\sin \frac{z_1 x - \alpha}{2} = 0.$$

Это означает, что возможные значения ξ имеют вид $\frac{\alpha}{z_1} + k \frac{2\pi}{z_1}$, т. е. случайная величина ξ имеет решетчатое распределение.

Шаг распределения h называется максимальным, если ни при каких b и $h_1 > h$ нельзя представить все возможные значения ξ в виде $b + kh_1$. Например, если ξ — целочисленная случайная величина, принимающая только нечетные значения, то все ее значения можно представить в виде $a + kh$, где $a = 0$, $h = 1$. Однако шаг $h = 1$ не является максимальным, так как все возможные значения ξ можно также записать в виде $b + kh_1$, где $b = 1$, $h_1 = 2$.

Условие максимальной шага можно сформулировать следующим образом:

Теорема 7. Шаг распределения h будет максимальным тогда и только тогда, когда модуль характеристической функции при $0 < |z| < \frac{2\pi}{h}$ меньше единицы и $\varphi\left(\frac{2\pi}{h}\right) = 1$.

В самом деле, если при $0 < z_1 < \frac{2\pi}{h} |\varphi(z_1)| = 1$, то из доказанного выше вытекает, что $\frac{2\pi}{z_1}$ — шаг распределения, а так как $h < \frac{2\pi}{z_1}$, то шаг h не может быть максимальным.

Глава III

ПОСЛЕДОВАТЕЛЬНОСТИ СЛУЧАЙНЫХ ВЕЛИЧИН

§ 1. Неравенство Колмогорова

При изучении поведения сумм независимых случайных величин часто приходится оценивать вероятности отклонения максимальной суммы. Наиболее простая, и в то же время, общая оценка такой вероятности получена А. Н. Колмогоровым.

Теорема. Пусть ξ_1, \dots, ξ_n — независимые случайные величины, для которых $M\xi_k = 0$, $k = 1, \dots, n$, $D\xi_k < \infty$, $k = 1, \dots, n$. Положим $\zeta_k = \xi_1 + \dots + \xi_k$. Тогда для $a > 0$

$$P\left\{\sup_{1 \leq k \leq n} |\zeta_k| \geq a\right\} \leq \frac{1}{a^2} \sum_{k=1}^n D\xi_k. \quad (1)$$

Неравенство (1) называется неравенством Колмогорова.

Доказательство. Введем случайные величины χ_k , $k = 1, \dots, n$, определяемые следующим образом: $\chi_k = 1$, если $|\xi_1| < a, \dots, |\xi_{k-1}| < a$; $|\xi_k| \geq a$; $\chi_k = 0$ — в остальных случаях. Величина $\chi_k = 1$, если k является первым номером, для которого $|\xi_k| \geq a$. В остальных случаях $\chi_k = 0$. Поэтому среди величин χ_k лишь одна величина может быть равной 1. $\sum_{k=1}^n \chi_k = 1$ тогда и только тогда, когда $\sup_{1 \leq k \leq n} |\xi_k| \geq a$; если же $\sup_{1 \leq k \leq n} |\xi_k| < a$, то

$\sum_{k=1}^n \chi_k = 0$. Поэтому

$$M \sum_{k=1}^n \chi_k = P\left\{\sup_{1 \leq k \leq n} |\xi_k| \geq a\right\}. \quad (2)$$

Величина χ_k определяется величинами ξ_1, \dots, ξ_k и, следовательно, не зависит от величин ξ_{k+1}, \dots, ξ_n (см. § 5, гл. II). Поэтому

$$\begin{aligned} M\xi_n^2 &\geq M \sum_1^n \chi_k \xi_k^2 = M \sum_1^n \chi_k [\xi_k^2 + 2(\xi_n - \xi_k) \xi_k + \\ &+ (\xi_n - \xi_k)^2] \geq M \sum_1^n \chi_k [\xi_k^2 + 2(\xi_n - \xi_k) \xi_k]. \end{aligned} \quad (3)$$

Заметим далее, что

$$\chi_k \xi_k^2 \geq a^2 \chi_k,$$

а

$$M\chi_k \xi_k (\xi_n - \xi_k) = M\chi_k \xi_k \sum_{i=k+1}^n \xi_i = \sum_{i=k+1}^n M\chi_k \xi_k M\xi_i = 0,$$

так как ξ_i , $i < k$, не зависит от $\chi_k \xi_k$. Следовательно,

$$M\xi_n^2 \geq a^2 \sum_1^n M\chi_k.$$

Чтобы получить отсюда требуемое неравенство, нужно воспользоваться соотношением (2) и равенством $M\xi_n^2 = \sum_1^n D\xi_k$. Теорема доказана.

Следствие. Пусть ξ_k , $k = 1, 2, \dots$, — бесконечная последовательность независимых случайных величин, $M\xi_k = 0$, $D\xi_k < \infty$ и $\sum_{k=1}^{\infty} D\xi_k < \infty$. Положим $\xi_k = \xi_1 + \dots + \xi_k$. Тогда

$$P \left\{ \sup_k |\xi_k| > a \right\} \leq \frac{1}{a^2} \sum_{k=1}^{\infty} D\xi_k.$$

Действительно, для всякого n

$$P \left\{ \sup_{k \leq n} |\xi_k| > a \right\} \leq \frac{1}{a^2} \sum_{k=1}^n D\xi_k \leq \frac{1}{a^2} \sum_{k=1}^{\infty} D\xi_k. \quad (4)$$

Поскольку $\{\sup_k |\xi_k| > a\} = \bigcup_n \{\sup_{k \leq n} |\xi_k| > a\}$, то переходя к пределу в соотношении (1) при $n \rightarrow \infty$, получаем требуемое.

§ 2. Сходимость по вероятности

Определение 1. Говорят, что последовательность случайных величин ξ_n по вероятности сходится к случайной величине ξ , если для всякого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P \{ |\xi_n - \xi| > \varepsilon \} = 0.$$

Обозначим сходимость по вероятности последовательности ξ_n

к ξ так: $\xi = p - \lim \xi_n$, или $\xi_n \xrightarrow{p} \xi$.

Отметим важнейшие свойства сходящихся по вероятности последовательностей.

1) Пусть $\xi_n \xrightarrow{p} \xi$ и $f(x)$ — непрерывная функция, определенная на R^1 . Тогда

$$f(\xi_n) \xrightarrow{p} f(\xi).$$

Действительно, пусть $\varepsilon > 0$, $c > 0$ и $\delta > 0$ выбрано так по c и ε , что при $|x_1 - x_2| \leq \delta$, $|x_1| \leq c$ $|f(x_1) - f(x_2)| \leq \varepsilon$ (такое δ можно выбрать согласно равномерной непрерывности функции f на каждом ограниченном множестве). Тогда

$$P\{|f(\xi_n) - f(\xi)| > \varepsilon\} \leq P\{|\xi| > c\} + P\{|\xi_n - \xi| > \delta\}.$$

Следовательно,

$$\overline{\lim}_{n \rightarrow \infty} P\{|f(\xi_n) - f(\xi)| > \varepsilon\} \leq P\{|\xi| > c\}.$$

Поскольку величина в правой части выбором $c > 0$ может быть сделана сколь угодно малой, то

$$\overline{\lim}_{n \rightarrow \infty} P\{|f(\xi_n) - f(\xi)| > \varepsilon\} = 0.$$

Отсюда и вытекает доказываемое утверждение.

2) Пусть заданы m последовательностей случайных величин $\xi_n^{(k)}$, $k = 1, \dots, m$, $\xi_n^{(k)} \xrightarrow{p} \xi^{(k)}$, а $\Phi(x_1, \dots, x_m)$ — непрерывная функция, определенная на R^m . Тогда

$$\Phi(\xi_n^{(1)}, \dots, \xi_n^{(m)}) \xrightarrow{p} \Phi(\xi^{(1)}, \dots, \xi^{(m)}).$$

Доказательство, как и в 1), вытекает из неравенства

$$\begin{aligned} P\{|\Phi(\xi_n^{(1)}, \dots, \xi_n^{(m)}) - \Phi(\xi^{(1)}, \dots, \xi^{(m)})| > \varepsilon\} &\leq \\ &\leq \sum_{k=1}^m (P\{|\xi^{(k)}| > c\} + P\{|\xi_n^{(k)} - \xi^{(k)}| > \delta\}), \end{aligned}$$

где δ таково, что при $|x_1| \leq c, \dots, |x_m| \leq c$

$$|\Phi(x_1, \dots, x_m) - \Phi(y_1, \dots, y_m)| \leq \varepsilon,$$

если только $|x_1 - y_1| \leq \delta, \dots, |x_m - y_m| \leq \delta$.

3) Если $\xi_n \xrightarrow{p} \xi$ и $P\{|\xi_n| \leq c\} = 1$ при некотором $c > 0$, то

$$\lim_{n \rightarrow \infty} M\xi_n = M\xi.$$

Заметим, что и $P\{|\xi| \leq c\} = 1$. Действительно, пусть $f(x)$ непрерывна, $f(x) = 0$ при $|x| \leq c$ и $f(x) > 0$ при $|x| > c$. Тогда $P\{f(\xi_n) = 0\} = 1$ и $f(\xi_n) \xrightarrow{p} f(\xi)$. Следовательно, $P\{f(\xi) = 0\} = 1$ и $P\{|\xi| \leq c\} = 1$. Пусть $\chi_n(\delta) = 1$, если $|\xi_n - \xi| > \delta$, $\chi_n(\delta) = 0$, если $|\xi_n - \xi| \leq \delta$. Тогда

$$|\xi_n - \xi| \leq \delta + 2c\chi_n(\delta).$$

Таким образом,

$$|M\xi_n - M\xi| \leq M|\xi_n - \xi| \leq \delta + 2cM\chi_n(\delta) \leq \delta + 2cP\{|\xi_n - \xi| > \delta\}.$$

Поэтому

$$\overline{\lim}_{n \rightarrow \infty} |M\xi_n - M\xi| \leq \delta,$$

каково бы ни было $\delta > 0$.

4) Если $P\{\xi_n \geq 0\} = 1$ и $\xi_n \xrightarrow{p} \xi$, то

$$M\xi \leq \lim_{n \rightarrow \infty} M\xi_n.$$

Пусть $f_N(x) = x$ при $|x| \leq N$ и $f_N(x) = N \operatorname{sign} x$ при $|x| > N$. Тогда на основании свойства 3)

$$\lim_{n \rightarrow \infty} Mf_N(\xi_n) = Mf_N(\xi).$$

Следовательно,

$$Mf_N(\xi) \leq \lim_{n \rightarrow \infty} M\xi_n. \quad (1)$$

Используя свойство 11) (гл. II, § 3) математических ожиданий, имеем $M\xi = \lim_{N \rightarrow \infty} Mf_N(\xi)$. Переходя к пределу при $N \rightarrow \infty$, в (1) получаем требуемое.

5) Если $\xi_n \xrightarrow{p} \xi$ и при некотором $\alpha > 0 \sup_n M|\xi_n|^{1+\alpha} < \infty$, то

$$\lim_{n \rightarrow \infty} M\xi_n = M\xi.$$

Пусть $f_N(x)$ такое, как в 4). Тогда

$$|f_N(\xi) - \xi| \leq \frac{|\xi|^{1+\alpha}}{N^\alpha}.$$

Поэтому

$$\begin{aligned} |M\xi_n - M\xi| &\leq |Mf_N(\xi_n) - Mf_N(\xi)| + M|f_N(\xi) - \xi| + \\ &+ M|f_N(\xi_n) - \xi_n| \leq |Mf_N(\xi_n) - Mf_N(\xi)| + \\ &+ \frac{1}{N^\alpha} (M|\xi_n|^{1+\alpha} + M|\xi|^{1+\alpha}). \end{aligned}$$

На основании свойства 3) $|M/N(\xi_n) - M/N(\xi)| \rightarrow 0$, а согласно 4)

$$M|\xi|^{1+\alpha} \leq \sup_n M|\xi_n|^{1+\alpha}.$$

Следовательно,

$$\lim_{n \rightarrow \infty} |M\xi_n - M\xi| \leq \frac{2}{N^\alpha} \sup_n M|\xi_n|^{1+\alpha}.$$

Переходя к пределу при $N \rightarrow \infty$, убеждаемся в справедливости доказываемого.

Пусть $\xi_n \xrightarrow{p} \xi$. Тогда $|\xi_n - \xi| \xrightarrow{p} 0$. Если при некотором $c > 0$ $|\xi_n - \xi| \leq c$, то согласно 3), $\lim_{n \rightarrow \infty} M|\xi_n - \xi| = 0$.

Определение. Последовательность случайных величин ξ_n сходится к случайной величине ξ в среднем, если для всех n существуют $M\xi_n$, $M\xi$ и $\lim_{n \rightarrow \infty} M|\xi_n - \xi| = 0$. Очевидно, что в этом случае $\lim_{n \rightarrow \infty} M\xi_n = M\xi$.

6) Если $\xi_n \rightarrow \xi$ в среднем, то $\xi_n \xrightarrow{p} \xi$. Действительно, в силу свойства (2) (гл. II, § 3) для всех $\varepsilon > 0$

$$P\{|\xi_n - \xi| > \varepsilon\} \leq \frac{1}{\varepsilon} M|\xi_n - \xi|.$$

Определение. Последовательность случайных величин ξ_n сходится к случайной величине ξ в среднеквадратическом, если $M\xi_n^2 < \infty$, $M\xi^2 < \infty$ и

$$\lim_{n \rightarrow \infty} M|\xi_n - \xi|^2 = 0.$$

7) Если $\xi_n \rightarrow \xi$ в среднеквадратическом, то $\xi_n \rightarrow \xi$ в среднем, а следовательно, и $\xi_n \xrightarrow{p} \xi$. Это вытекает из неравенства (см. гл. II, § 4)

$$M|\xi_n - \xi| \leq (M|\xi_n - \xi|^2)^{\frac{1}{2}}.$$

8) Если $\xi_n \rightarrow \xi$ в среднеквадратическом, то

$$M\xi_n \rightarrow M\xi \text{ и } M\xi_n^2 \rightarrow M\xi^2.$$

Первое утверждение вытекает из 7).

Для доказательства второго заметим, что

$$\xi_n^2 = [\xi + (\xi_n - \xi)]^2 \leq 2[\xi^2 + (\xi_n - \xi)^2],$$

здесь использовано неравенство

$$(a + b)^2 \leq 2[a^2 + b^2].$$

Следовательно,

$$M\xi_n^2 \leq 2[M\xi^2 + M(\xi_n - \xi)^2].$$

Поэтому

$$\begin{aligned} |M\xi_n^2 - M\xi^2| &= |M(\xi_n - \xi)(\xi_n + \xi)| \leq \\ &\leq [M(\xi_n - \xi)^2]^{\frac{1}{2}} [M(\xi_n + \xi)^2]^{\frac{1}{2}} \leq \\ &\leq [M(\xi_n - \xi)^2]^{\frac{1}{2}} [2M(\xi_n^2 + \xi^2)]^{\frac{1}{2}} \leq \\ &\leq [M(\xi_n - \xi)^2]^{\frac{1}{2}} [5M\xi^2 + 4M(\xi_n - \xi)^2]^{\frac{1}{2}}, \end{aligned}$$

а правая часть стремится к нулю.

9) Для того, чтобы последовательность ξ_n сходилась в среднеквадратическом к некоторой постоянной c , необходимо и достаточно, чтобы $M\xi_n \rightarrow c$, $D\xi_n \rightarrow 0$. Действительно, если $\xi_n \rightarrow c$ в среднеквадратическом, то $M\xi_n \rightarrow Mc = c$, $M\xi_n^2 \rightarrow c^2$, поэтому $D\xi_n = M\xi_n^2 - (M\xi_n)^2 \rightarrow 0$. Наоборот, если $M\xi_n \rightarrow c$, $D\xi_n \rightarrow 0$, то

$$M(\xi_n - c)^2 = M(\xi_n - M\xi_n)^2 + (M\xi_n - c)^2 \rightarrow 0.$$

§ 3. Закон больших чисел

Закон больших чисел. Пусть ξ_n — последовательность случайных величин, для которых существуют $M\xi_n$. Законом больших чисел называются теоремы, утверждающие, что разность

$$\frac{1}{n} \sum_{k=1}^n \xi_k - \frac{1}{n} \sum_{k=1}^n M\xi_k$$

сходится к нулю по вероятности.

Теорема Чебышева. Пусть ξ_n — последовательность независимых случайных величин, $M\xi_n = a$, $D\xi_n \leq c$. Тогда

$$\frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow{p} a.$$

Доказательство. Докажем даже больше, что $\frac{1}{n} \sum_{k=1}^n \xi_k \rightarrow a$

в среднеквадратическом. Так как $M \frac{1}{n} \sum_{k=1}^n \xi_k = a$, то на основании 9), § 2, для доказательства теоремы достаточно показать, что $D\left(\frac{1}{n} \sum_{k=1}^n \xi_k\right) \rightarrow 0$. Вследствие независимости величин ξ_k

$$D\left(\frac{1}{n} \sum_{k=1}^n \xi_k\right) = \frac{1}{n^2} \sum_{k=1}^n D\xi_k \leq \frac{1}{n^2} cn = \frac{c}{n} \rightarrow 0.$$

Замечания. 1. Для справедливости теоремы достаточно, чтобы

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{k=1}^n D\xi_k = 0.$$

Это будет выполнено, если $\lim_{n \rightarrow \infty} \frac{1}{n} D\xi_n = 0$, так как на основании теоремы Штольца¹

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{k=1}^n D\xi_k = \lim_{n \rightarrow \infty} \frac{D\xi_n}{n^2 - (n-1)^2} = \lim_{n \rightarrow \infty} \frac{D\xi_n}{2n-1} = 0.$$

2. Можно оценить вероятность

$$\begin{aligned} P\left\{\left|\frac{1}{n} \sum_{k=1}^n \xi_k - a\right| > \varepsilon\right\} &\leq \frac{1}{\varepsilon^2} M\left(\frac{1}{n} \sum_{k=1}^n \xi_k - a\right)^2 = \\ &= \frac{\sum_{k=1}^n D\xi_k}{\varepsilon^2 n^2} \leq \frac{c}{\varepsilon^2 n}. \end{aligned}$$

Мы воспользовались неравенством Чебышева (неравенство (11), § 4).

Следствие. Пусть $\xi_1, \dots, \xi_n, \dots$ — последовательность независимых случайных величин такая, что

$$M\xi_n = a, D\xi_n \leq c, n = 1, 2, \dots$$

Тогда для каждого $\varepsilon > 0$

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

Этот частный случай теоремы Чебышева дает обоснование правилу среднего арифметического в теории обработки результатов измерений. Предположим, что нужно измерить некоторую физическую величину a . Повторив измерения n раз в одинаковых условиях, наблюдатель получает результаты измерений ξ_1, \dots, ξ_n .

¹ Теоремой Штольца называется следующее утверждение из теории пределов:

Пусть $\{x_n\}$ и $\{y_n\}$ — две числовые последовательности такие, что:

а) $y_{n+1} > y_n$,

б) $\lim y_n = +\infty$,

в) существует предел $\lim \frac{x_n - x_{n-1}}{y_n - y_{n-1}}$.

Тогда последовательность $\frac{x_n}{y_n}$ имеет предел, причем

$$\lim_{n \rightarrow \infty} \frac{x_n}{y_n} = \lim_{n \rightarrow \infty} \frac{x_n - x_{n-1}}{y_n - y_{n-1}}.$$

В качестве приближенного значения a принимается среднее арифметическое результатов измерений

$$\frac{\xi_1 + \dots + \xi_n}{n}.$$

Если наблюдения лишены систематической ошибки, т. е. $M\xi_n = a$, то согласно сформулированному выше следствию,

$$\frac{\xi_1 + \dots + \xi_n}{n} - a \xrightarrow{p} 0.$$

Теорема Хинчина. Если $\{\xi_n\}$ — последовательность независимых одинаково распределенных случайных величин, то закон больших чисел к такой последовательности применим и без предположения о существовании дисперсий. Именно, имеет место следующее утверждение.

Теорема Хинчина. Пусть $\{\xi_n\}$ — последовательность независимых одинаково распределенных величин, имеющих конечное математическое ожидание $M\xi_n = a$. Тогда для каждого $\varepsilon > 0$

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

Мы не будем приводить доказательство этой теоремы, а отметим лишь, что она следует из приведенной в § 5 теоремы Колмогоры об усиленном законе больших чисел.

Теорема Бернулли. Рассмотрим еще один частный случай теоремы Чебышева. Пусть имеем последовательность испытаний, в каждом из которых может быть два исхода — успех Y (с вероятностью p) или неудача N (с вероятностью $q = 1 - p$) независимо от исходов других испытаний. Образует последовательность случайных величин следующим образом. Пусть $\xi_k = 1$, если в k -м испытании произошел успех, и $\xi_k = 0$, если в k -м испытании произошла неудача. Тогда $\{\xi_k\}$ есть последовательность независимых одинаково распределенных случайных величин, $M\xi_k = p$, $D\xi_k = M\xi_k^2 - (M\xi_k)^2 = pq$. Случайная величина

$$v_n = \frac{1}{n} \sum_{k=1}^n \xi_k$$

представляет собой частоту появления успеха в первых n испытаниях. Так как для последовательности $\{\xi_k\}$ выполнены условия теоремы Чебышева, то мы из теоремы Чебышева получаем следующее утверждение.

Теорема Бернулли. Для любого $\varepsilon > 0$ $P\{|v_n - p| > \varepsilon\} \rightarrow 0$ при $n \rightarrow \infty$.

Смысл этого утверждения состоит в том, что введенное нами определение вероятности соответствует интуитивному пониманию вероятности как предела частоты.

Многочлены Бернштейна. Закон больших чисел можно использовать для доказательства известной из курса математического анализа теоремы Вейерштрасса о равномерном приближении непрерывной функции многочленами.

Предположим, что производятся независимые испытания, в каждом из которых может произойти либо событие A (успех) с вероятностью x , либо противоположное событие \bar{A} (неудача) с вероятностью $1 - x$ ($0 < x < 1$). Пусть γ_n — число появлений A при n испытаниях, а $f(x)$ — непрерывная функция на $[0, 1]$. Как известно,

$$P(\gamma_n = k) = C_n^k x^k (1 - x)^{n-k}, \quad k = 0, 1, \dots, n.$$

Поэтому

$$B_n(x) = Mf\left(\frac{\gamma_n}{n}\right) = \sum_{k=0}^n f\left(\frac{k}{n}\right) C_n^k x^k (1 - x)^{n-k}. \quad (1)$$

Многочлен $B_n(x)$ называется многочленом Бернштейна для функции $f(x)$.

Выше мы отметили, что $\frac{\gamma_n}{n} \xrightarrow{P} x$. Естественно ожидать, что $Mf\left(\frac{\gamma_n}{n}\right) \rightarrow f(x)$ при $n \rightarrow \infty$. Именно, мы докажем следующее утверждение.

Теорема Бернштейна. Последовательность многочленов $B_n(x)$, определенных равенством (1), сходится к функции $f(x)$ равномерно относительно $x \in [0, 1]$.

Так как $f(x)$ равномерно непрерывна на $[0, 1]$, то для каждого $\varepsilon > 0$ найдется $\delta(\varepsilon)$ такое, что $|f(x_1) - f(x_2)| < \frac{\varepsilon}{2}$, как только $|x_1 - x_2| < \frac{\delta}{2}$. Функция $f(x)$ ограничена на $[0, 1]$. Поэтому существует такая постоянная C , что $|f(x)| \leq C$ для всех $x \in [0, 1]$. Заметим также, что

$$\sum_{k=0}^n C_n^k x^k (1 - x)^{n-k} = 1.$$

Поэтому

$$B_n(x) - f(x) = \sum_{k=0}^n \left\{ f\left(\frac{k}{n}\right) - f(x) \right\} C_n^k x^k (1 - x)^{n-k},$$

и имеем далее

$$|B_n(x) - f(x)| \leq \sum_{k=0}^n \left| f\left(\frac{k}{n}\right) - f(x) \right| C_n^k x^k (1 - x)^{n-k} =$$

$$\begin{aligned}
&= \sum_{k: \left| \frac{k}{n} - x \right| < \delta} \left| f\left(\frac{k}{n}\right) - f(x) \right| C_n^k x^k (1-x)^{n-k} + \\
&+ \sum_{k: \left| \frac{k}{n} - x \right| \geq \delta} \left| f\left(\frac{k}{n}\right) - f(x) \right| C_n^k x^k (1-x)^{n-k} \leq \\
&\leq \frac{\varepsilon}{2} + 2C \sum_{\left| \frac{k}{n} - x \right| \geq \delta} C_n^k x^k (1-x)^{n-k} = \frac{\varepsilon}{2} + 2C \mathbf{P} \left\{ \left| \frac{\gamma_n}{n} - x \right| \geq \delta \right\}.
\end{aligned}$$

Вследствие неравенства Чебышева,

$$\mathbf{P} \left\{ \left| \frac{\gamma_n}{n} - x \right| > \delta \right\} \leq \frac{D \left(\frac{\gamma_n}{n} \right)}{\delta^2} = \frac{x(1-x)}{n\delta^2} \leq \frac{1}{4n\delta^2}$$

(использовано неравенство $x(1-x) \leq \frac{1}{4}$ при $0 \leq x \leq 1$). Пусть $N(\delta)$ такое, что $\frac{1}{4n\delta^2} < \frac{\varepsilon}{2}$. Тогда при $n \geq N$

$$|B_n(x) - f(x)| < \varepsilon$$

при всех $x \in [0, 1]$. Теорема доказана.

Метод статистических испытаний вычисления интегралов (метод Монте-Карло).

Предположим, что надо вычислить интеграл $\int_0^1 g(x) dx$, где $g(x)$ — некоторая непрерывная функция. Пусть $\xi_1, \dots, \xi_n, \dots$ — последовательность независимых случайных величин, имеющих равномерное распределение на $[0, 1]$. Заметим, что согласно соотношению (7), § 4, гл. II,

$$Mg(\xi_n) = \int_{-\infty}^{\infty} g(x) p_{\xi_n}(x) dx = \int_0^1 g(x) dx.$$

По теореме Хинчина

$$\frac{g(\xi_1) + \dots + g(\xi_n)}{n} \xrightarrow[n \rightarrow \infty]{P} \int_0^1 g(x) dx.$$

Таким образом, теоремы о законе больших чисел дают теоретическое обоснование следующего приближенного алгоритма вычисления интеграла $\int_0^1 g(x) dx$:

1. Моделируем последовательность $\xi_1, \dots, \xi_n, \dots$ независимых случайных величин, имеющих равномерное распределение на $0, 1^1$.
2. Полагаем

$$\int_0^1 g(x) dx \approx \frac{g(\xi_1) + \dots + g(\xi_n)}{n}.$$

Описанный метод называется методом статистических испытаний вычисления интегралов (методом Монте-Карло)².

Метод Монте-Карло особенно эффективен при вычислении интегралов большой кратности; для вычисления таких интегралов другие методы приближенного анализа непригодны.

§ 4. Сходимость с вероятностью 1

Пусть ξ_n — некоторая последовательность случайных величин. Рассмотрим событие A : « ξ_n сходится к ξ »:

$$A = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ |\xi_n - \xi| \leq \frac{1}{k} \right\}.$$

По теореме 6, § 1, гл. II, событие A является случайным событием, если ξ_n и ξ случайные величины. Здесь напомним лишь,

что $A_{N,k} = \bigcap_{n=N}^{\infty} \left\{ |\xi_n - \xi| \leq \frac{1}{k} \right\}$ — событие, заключающееся в том,

что для $n \geq N$ выполнено неравенство $|\xi_n - \xi| \leq \frac{1}{k}$, $B_k =$

$= \bigcup_{N=1}^{\infty} A_{N,k}$ — событие, заключающееся в том, что существует

такое N , что указанное неравенство выполняется при $n \geq N$,

наконец, $\bigcap_{k=1}^{\infty} B_k$ — событие, что для всех k существует такое N ,

что при $n \geq N$ выполнено неравенство $|\xi_n - \xi| \leq \frac{1}{k}$, т. е. событие, состоящее в том, что $\lim \xi_n = \xi$.

Определение 1. Если

$$P(\lim \xi_n = \xi) = P\left\{ \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ |\xi_n - \xi| \leq \frac{1}{k} \right\} \right\} = 1,$$

¹ В книгах, посвященных методу статистического моделирования, описаны различные методы моделирования случайных последовательностей с заданным распределением; см., например, С. М. Ермаков, Г. А. Михайлов. Курс статистического моделирования. М., Наука, 1976.

² Название «Монте-Карло» происходит от города Монте-Карло — столицы княжества Монако, знаменитого своими игорными домами. (Заметим, что одним из простейших приборов для моделирования случайных величин является рулетка).

то говорят, что последовательность случайных величин $\{\xi_n\}$ сходится к случайной величине ξ с вероятностью 1, и записывают $P\{\xi_n \rightarrow \xi\} = 1$ или $P\{\lim \xi_n = \xi\} = 1$.

Если $P\{\bigcap_{k=1}^{\infty} B_k\} = 1$, то и $P\{B_k\} = 1$ для всех k , и, наоборот,

из того что $P\{B_k\} = 1$ для всех k , вытекает, что $P\{\bigcap_{k=1}^{\infty} B_k\} = 1$.

Таким образом, для того чтобы с вероятностью 1 $\xi_n \rightarrow \xi$ необходимо и достаточно, чтобы для всех k

$$P\left\{\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{|\xi_n - \xi| \leq \frac{1}{k}\right\}\right\} = 1. \quad (1)$$

Поскольку $\bigcap_{n=N}^{\infty} \left\{|\xi_n - \xi| \leq \frac{1}{k}\right\}$ — возрастающая последовательность событий, то (1) выполняется тогда и только тогда, когда для всех k

$$\lim_{N \rightarrow \infty} P\left\{\bigcap_{n=N}^{\infty} \left\{|\xi_n - \xi| \leq \frac{1}{k}\right\}\right\} = 1.$$

Заметим, что число $\frac{1}{k}$ в последнем соотношении можно заменить произвольным $\varepsilon > 0$. Итак, справедливо утверждение:

Лемма 1. Для того чтобы последовательность ξ_n сходилась с вероятностью 1 к величине ξ , необходимо и достаточно, чтобы для всех $\varepsilon > 0$

$$\lim_{N \rightarrow \infty} P\left\{\bigcap_{n=N}^{\infty} \{|\xi_n - \xi| \leq \varepsilon\}\right\} = 1.$$

Замечание. Поскольку

$$P\{|\xi_n - \xi| \leq \varepsilon\} \geq P\left\{\bigcap_{k=n}^{\infty} \{|\xi_k - \xi| \leq \varepsilon\}\right\},$$

то из сходимости ξ_n к ξ с вероятностью 1 вытекает, что $\xi_n \xrightarrow{P} \xi$, так как в этом случае

$$P\{|\xi_n - \xi| < \varepsilon\} \rightarrow 1.$$

Следствие. Условие леммы эквивалентно следующему: для всякого $\varepsilon > 0$

$$\lim_{N \rightarrow \infty} P\left\{\bigcup_{n=N}^{\infty} \{|\xi_n - \xi| > \varepsilon\}\right\} = 0.$$

Поскольку

$$P\left\{\bigcup_{n=N}^{\infty} \{|\xi_n - \xi| > \varepsilon\}\right\} \leq \sum_{n=N}^{\infty} P\{|\xi_n - \xi| > \varepsilon\},$$

то для того чтобы $P\{\xi_n \rightarrow \xi\} = 1$, достаточно выполнить неравенство

$$\sum_{n=1}^{\infty} P(|\xi_n - \xi| > \varepsilon) < \infty$$

для всякого $\varepsilon > 0$.

При изучении сходимости величин с вероятностью 1 важную роль играет следующая теорема.

Теорема 1. (Бореля — Кантелли). Пусть A_n — последовательность событий. Для того чтобы в этой последовательности с вероятностью 1 наступило конечное число событий, достаточно, а в случае независимых событий A_k и необходимо, чтобы выполнялось условие

$$\sum_{k=1}^{\infty} P(A_k) < \infty. \quad (2)$$

Доказательство. 1) Пусть C — событие, заключающееся в том, что среди событий A_n происходит бесконечное число. Тогда

$$C = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Для всякого n

$$C \subset \bigcup_{k=n}^{\infty} A_k, \quad P(C) \leq \sum_{k=n}^{\infty} P(A_k) \rightarrow 0 \text{ при } n \rightarrow \infty.$$

2) Пусть A_k — независимы и $P(C) = 0$. Тогда

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} A_k\right) = 0$$

и, следовательно, при некотором n имеем $P\left(\bigcup_{k=n}^{\infty} A_k\right) < 1$. Но для

противоположного события $\bigcup_{k=n}^{\infty} A_k = \bigcap_{k=n}^{\infty} \bar{A}_k$

$$\begin{aligned} 0 < P\left(\bigcap_{k=n}^{\infty} \bar{A}_k\right) &= \lim_{m \rightarrow \infty} \prod_{k=n}^m P(\bar{A}_k) = \lim_{m \rightarrow \infty} \prod_{k=n}^m (1 - P(A_k)) \leq \\ &\leq \lim_{m \rightarrow \infty} \exp\left(-\sum_{k=n}^m P(A_k)\right) = \exp\left(-\sum_{k=n}^{\infty} P(A_k)\right). \end{aligned}$$

Из этого неравенства и вытекает условие (2).

Следствие. Если $\xi_n \xrightarrow{P} \xi$, то можно указать такую подпоследовательность ξ_{n_k} , что $\xi_{n_k} \rightarrow \xi$ с вероятностью 1. Действительно, для всякого k можно указать такое n_k , что

$$P\left\{|\xi_{n_k} - \xi| > \frac{1}{k}\right\} \leq \frac{1}{k^2}.$$

$$\sum \mathbf{P} \left\{ |\xi_{n_k} - \xi| > \frac{1}{k} \right\} < \infty,$$

то на основании теоремы Бореля—Кантелли происходит лишь конечное число событий $\left\{ |\xi_{n_k} - \xi| > \frac{1}{k} \right\}$. Таким образом, начиная с некоторого номера k , выполнено неравенство $|\xi_{n_k} - \xi| < \frac{1}{k}$, т. е. $\xi_{n_k} \rightarrow \xi$ (все это происходит с вероятностью 1).

Рассмотрим теперь такое событие: «последовательность случайных величин ξ_n имеет предел». Это действительно событие, потому что «последовательность ξ_n имеет предел» тогда и только тогда, когда происходит событие

$$A = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{\substack{n \geq N \\ m \geq N}} \left\{ |\xi_n - \xi_m| \leq \frac{1}{k} \right\}. \quad (3)$$

Возникает вопрос, если последовательность случайных величин ξ_n сходится с вероятностью 1 к некоторому пределу, т. е. если событие A , определяемое равенством (3), имеет вероятность 1, то существует ли случайная величина, к которой ξ_n сходится с вероятностью 1?

Пусть случайные величины ξ_n заданы на вероятностном пространстве $\{\Omega, \mathcal{F}, \mathbf{P}\}$, $\xi_n = \xi_n(\omega)$. Существует $\lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)$ почти для всех ω по мере \mathbf{P} . Тогда

$$\{\xi(\omega) < x\} = \bigcup_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ \xi_n(\omega) < x - \frac{1}{k} \right\}.$$

Таким образом, $\{\xi(\omega) < x\} \in \mathcal{F}$, поэтому $\xi(\omega)$ — случайная величина.

Используя это обстоятельство и следствие из теоремы Бореля—Кантелли, можно сформулировать критерий Коши для сходимости последовательности случайных величин по вероятности и в среднем квадратическом.

Теорема 2. Для того чтобы последовательность ξ_n сходилась по вероятности к некоторому пределу ξ , необходимо и достаточно, чтобы для всех $\varepsilon > 0$

$$\lim_{n, m \rightarrow \infty} \mathbf{P} \{ |\xi_n - \xi_m| > \varepsilon \} = 0.$$

Доказательство. Достаточность. Выберем так последовательность n_k , чтобы $n_k > n_{k-1}$,

$$\sup_{m > n_k} \mathbf{P} \left\{ |\xi_{n_k} - \xi_m| > \frac{1}{k^2} \right\} \leq \frac{1}{k^2}.$$

Тогда

$$P\left\{|\xi_{n_k} - \xi_{n_{k+1}}| > \frac{1}{k^2}\right\} < \frac{1}{k^2}.$$

Следовательно, по теореме Бореля — Кантелли с вероятностью 1 можно указать такой номер (он, вообще говоря, будет случайным), что для $k \geq N$

$$|\xi_{n_k} - \xi_{n_{k+1}}| < \frac{1}{k^2}.$$

Поэтому ряд $\xi_{n_1} + \sum_{k=2}^{\infty} (\xi_{n_k} - \xi_{n_{k-1}})$, члены которого, начиная с определенного номера, мажорируются членами сходящегося ряда $\frac{1}{k^2}$, будет с вероятностью 1 сходиться, т. е. существует предел $\xi = \lim_{k \rightarrow \infty} \xi_{n_k}$, являющийся случайной величиной. На основании замечания к лемме 1, ξ_{n_k} сходится по вероятности к ξ . Поэтому

$$\lim_{n \rightarrow \infty} P\{|\xi - \xi_n| > \varepsilon\} \leq \overline{\lim}_{\substack{n \rightarrow \infty \\ k \rightarrow \infty}} \left(P\left\{|\xi - \xi_{n_k}| > \frac{\varepsilon}{2}\right\} + P\left\{|\xi_{n_k} - \xi_n| > \frac{\varepsilon}{2}\right\} \right) = 0.$$

Необходимость. Пусть $\xi_n \xrightarrow{P} \xi$. Тогда

$$P\{|\xi_n - \xi_m| > \varepsilon\} \leq P\left\{|\xi_n - \xi| > \frac{\varepsilon}{2}\right\} + P\left\{|\xi_m - \xi| > \frac{\varepsilon}{2}\right\}$$

и правая часть этого неравенства стремится к нулю при $n \rightarrow \infty$, $m \rightarrow \infty$. Теорема доказана.

Теорема 3. Для того, чтобы последовательность ξ_n , для которой $M\xi_n^2 < \infty$, сходилась в среднеквадратическом к некоторому пределу, необходимо и достаточно выполнения условия

$$\lim_{n, m \rightarrow \infty} M|\xi_n - \xi_m|^2 = 0. \quad (4)$$

Доказательство. Достаточность. Согласно неравенству Чебышева,

$$P\{|\xi_n - \xi_m| > \varepsilon\} \leq \frac{1}{\varepsilon^2} M|\xi_n - \xi_m|^2.$$

Поскольку правая часть полученного неравенства стремится к нулю при $n, m \rightarrow \infty$, то, на основании теоремы 2 ξ_n сходится по вероятности к некоторому пределу ξ . Поэтому и $(\xi_n - \xi_m)^2$ при $n \rightarrow \infty$ сходится по вероятности к $(\xi - \xi_m)^2$. В силу свойства 4), § 2,

$$M(\xi - \xi_m)^2 \leq \lim_{n \rightarrow \infty} M(\xi_n - \xi_m)^2.$$

Переходя в этом соотношении к пределу при $m \rightarrow \infty$ и воспользовавшись условием (4), убеждаемся, что ξ_m сходится к ξ в среднем квадратическом. Для доказательства необходимости следует воспользоваться неравенством

$$M|\xi_n - \xi_m|^2 \leq 2M|\xi_n - \xi|^2 + 2M|\xi_m - \xi|^2,$$

в котором ξ является среднеквадратическим пределом последовательности ξ_n . Так как правая часть стремится к нулю при $n \rightarrow \infty$, $m \rightarrow \infty$, получаем (4).

§ 5. Усиленный закон больших чисел

Рассмотрим последовательность случайных величин ξ_k с конечными математическими ожиданиями. Теоремы, утверждающие, что разность

$$\frac{1}{n} \sum_{k=1}^n \xi_k - \frac{1}{n} \sum_{k=1}^n M\xi_k$$

сходится с вероятностью 1 к нулю, называются усиленным законом больших чисел (сравним с законом больших чисел, утверждающим, что указанная разность сходится к нулю лишь по вероятности). Ниже приведены две теоремы об усиленном законе больших чисел, обе они доказаны А. Н. Колмогоровым.

Теорема 1. Пусть ξ_n — последовательность независимых случайных величин, для которой $M\xi_n$ и $D\xi_n$ определены. Если

$$\sum_{k=1}^{\infty} \frac{1}{k^2} D\xi_k < \infty,$$

то

$$P\left\{\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n (\xi_k - M\xi_k) = 0\right\} = 1.$$

Доказательство. Не ограничивая общности, можно считать, что $M\xi_k = 0$. Обозначим

$$\eta_n = \sup_{m < 2^n} \left| \sum_{k=1}^m \xi_k \right|.$$

Так как при

$$2^{n-1} \leq m \leq 2^n,$$

$$\left| \frac{1}{m} \sum_{k=1}^m \xi_k \right| \leq 2^{-n+1} \sup_{m < 2^n} \left| \sum_{k=1}^m \xi_k \right| = 2^{-n+1} \eta_n,$$

$$P\{\lim_{n \rightarrow \infty} 2^{-n} \eta_n = 0\} = 1.$$

Чтобы выполнялось это соотношение, достаточно, чтобы для всякого $\varepsilon > 0$

$$\sum_{n=1}^{\infty} P\{2^{-n} \eta_n > \varepsilon\} < \infty$$

(см. следствие из леммы 1, § 4). На основании неравенства Колмогорова

$$P\{2^{-n} \eta_n > \varepsilon\} = P\{\eta_n > 2^n \varepsilon\} \leq \frac{1}{\varepsilon^2 2^{2n}} \sum_{k=1}^{2^n} D\xi_{kn}.$$

Поэтому

$$\begin{aligned} \sum_{n=1}^{\infty} P\{2^{-n} \eta_n > \varepsilon\} &\leq \varepsilon^{-2} \sum_{n=1}^{\infty} 2^{-2n} \sum_{k=1}^{2^n} D\xi_{kn} \leq \\ &\leq \varepsilon^{-2} \sum_{k=1}^{\infty} D\xi_k \sum_{2^n > k} 4^{-n} \leq \varepsilon^2 \sum_{k=1}^{\infty} D\xi_k \frac{k^{-1}}{1 - \frac{1}{4}} < \infty. \end{aligned}$$

Теорема доказана.

Следствие (Теорема Бореля). *Предположим, что рассматривается последовательность независимых испытаний, в каждом из которых появляется успех У с вероятностью p или неудача Н с вероятностью $q = 1 - p$. Пусть μ_n — число успехов при n испытаниях. Тогда*

$$P\left\{\frac{\mu_n}{n} \rightarrow p\right\} = 1. \quad (2)$$

Это вытекает из того, что

$$\frac{\mu_n}{n} = \frac{1}{n} \sum_{k=1}^n \xi_k,$$

где ξ_k — последовательность независимых случайных величин, введенных в § 2 при доказательстве теоремы Бернулли.

Условия теоремы 1 выполнены, так как $D\xi_k = p(1 - p)$, равенство (2) вытекает из (1), потому что $M\xi_k = p$. Сформулированное выше утверждение оправдывает статистический подход к определению вероятности как предела частоты.

Замечание. Следует, однако, иметь в виду, что утверждается сходимость частоты к вероятности лишь с вероятностью 1, т. е. строго говоря, частоту можно использовать для вычисления вероятности, но не для логического определения.

Теорема 2. Пусть ξ_n последовательность независимых одинаково распределенных величин с конечным математическим ожиданием $M\xi_n = a$. Тогда

$$P\left\{\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \xi_k = a\right\} = 1.$$

Доказательство. Введем величины ξ'_n , а именно $\xi'_n = \xi_n$ при $|\xi_n| \leq n$, $\xi'_n = 0$ при $|\xi_n| > n$. Пусть $\xi''_n = \xi_n - \xi'_n$. Заметим, что среди величин ξ''_n только конечное число отличных от нуля (на основании теоремы Бореля—Кантелли), так как

$$\begin{aligned} \sum_{n=1}^{\infty} P\{\xi''_n \neq 0\} &= \sum_{n=1}^{\infty} P\{|\xi_n| > n\} = \sum_{n=1}^{\infty} P\{|\xi_1| > n\} = \\ &= \sum_{n=1}^{\infty} \sum_{m=n}^{\infty} P\{m < |\xi_1| \leq m+1\} = \\ &= \sum_{m=1}^{\infty} mP\{m < |\xi_1| \leq m+1\} < M|\xi_1| < \infty. \end{aligned}$$

Следовательно,

$$P\left\{\lim_{k \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \xi''_k = 0\right\} = 1. \quad (3)$$

Рассмотрим далее величины ξ'_n . Пусть $F(x)$ — функция распределения величины ξ_k . Тогда

$$\begin{aligned} D\xi'_k &\leq M(\xi'_k)^2 = \int_{-k}^k x^2 dF(x), \\ \sum_{n=1}^{\infty} \frac{1}{n^2} \int_{-n}^n x^2 dF(x) &= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^n \int_{k-1 < |x| < k} x^2 dF(x) = \\ &= \sum_{k=1}^{\infty} \int_{k-1 < |x| < k} x^2 dF(x) \sum_{n=k}^{\infty} \frac{1}{n^2} \leq \\ &\leq \sum_{k=1}^{\infty} \int_{k-1 < |x| < k} |x| dF(x) \left(k \sum_{n=k}^{\infty} \frac{1}{n^2}\right). \end{aligned}$$

Поскольку $k \sum_{n=k}^{\infty} \frac{1}{n^2}$ величина ограниченная, а

$$\sum_{k=1}^{\infty} \int_{k-1 < |x| < k} |x| dF(x) = M|\xi_1| < \infty,$$

$$\sum_{n=1}^{\infty} \frac{1}{n^2} D\xi'_n < \infty.$$

Следовательно, на основании теоремы 1,

$$P \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n (\xi'_k - M\xi'_k) = 0 \right\},$$

по теореме Штольца

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n M\xi'_k = \lim_{n \rightarrow \infty} M\xi'_n = \lim_{n \rightarrow \infty} \int_{-\infty}^n x dF(x) = a.$$

Таким образом,

$$P \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \xi'_k = a \right\} = 1.$$

Это соотношение совместно с (3) дает доказательство теоремы.

§ 6. Случайные блуждания

Рассмотрим последовательность независимых одинаково распределенных случайных величин $\{\xi_k, k=1, 2, \dots\}$. Последовательность случайных величин $\{\zeta_n, n=0, 1, \dots\}$, где $\zeta_0=0$, $\zeta_n = \zeta_{n-1} + \xi_n$ ($n \geq 1$), называется случайным блужданием, ζ_n называется положением блуждания в момент n , ξ_n — n -м шагом блуждания. Случайное блуждание обычно интерпретируется как движение некоторой частицы на прямой, ζ_n — положение частицы в момент n , ξ_{n+1} — смещение частицы на отрезке $[n, n+1]$. Предполагается, что смещение частицы на отрезке времени $[n, n+1]$ не зависит от характера движения частицы на $[0, n]$.

Мы рассмотрим только тот случай, когда ξ_n — целочисленные случайные величины. Так как $\zeta_n = \xi_1 + \dots + \xi_n$, то ζ_n — также целочисленные случайные величины. Такое блуждание называется блужданием по целочисленной решетке.

Определение. Целое положительное число d называется шагом блуждания, если для всех n

$$\sum_k P\{\zeta_n = kd\} = 1$$

и d — наибольшее из чисел, для которых это выполнено. Если $d=1$, то блуждание называется *нерешетчатым*.

Очевидно, d есть наибольшее из чисел, для которых $\frac{\xi_1}{d}$ — целочисленная величина.

Лемма 1. Пусть $\varphi(z)$ — характеристическая функция величины ξ_1 , d — шаг блуждания. Тогда $\varphi(2\pi/d) = 1$ и $\varphi(z) \neq 1$ при $z \in (0, 2\pi/d)$.

Доказательство. Если d — шаг блуждания, то $\eta = \xi/d$ — целочисленная величина и

$$\varphi(2\pi/d) = M e^{i2\pi \xi_1/d} = M e^{i2\pi \eta} = \sum P\{\eta = m\} e^{2\pi m i} = 1.$$

Предположим, что при некотором $s \in (0, 2\pi/d)$ $\varphi(s) = 1$. Тогда

$$1 = \operatorname{Re} \varphi(s) = \sum p_k \cos ks.$$

Следовательно,

$$0 = \sum_k p_k (1 - \cos ks) = \sum_k p_k 2 \sin^2 \frac{ks}{2}$$

и для всех целых k

$$p_k \sin^2 \frac{ks}{2} = 0.$$

Пусть N — множество тех k , для которых $p_k > 0$. Для таких k выполняется равенство $\sin^2 \frac{ks}{2} = 0$ и, следовательно, $\frac{ks}{2} = l\pi$. Обозначим через l_0 наибольший общий делитель чисел $l = ks/2\pi$, $k \in N$. Тогда $l_0\pi = k_0 s/2$, где k_0 — целое число, поскольку d_1 представимо в виде $\sum (m_i k_i s/2\pi)$, $k_i \in N$. Поэтому $s = 2l_0\pi/k_0$. Таким образом, $s = 2\pi \cdot m/n$, где m/n — несократимая дробь. Поскольку для всех $k \in N$ число $2\pi \cdot km/n$ целое, то n является общим делителем чисел из \bar{N} . При этом $2\pi m/n < 2\pi/d$ и тогда $n > d$. Но ξ_1/n — целочисленная величина, так как $P\{\xi_1 \in N\} = 1$, а все числа из N делятся на n , но это противоречит определению d . Лемма доказана.

Заметим, что рассматривая вместо блуждания $\{\xi_n, n = 0, 1, \dots\}$ блуждание $\{\xi'_n = \xi_n/d, n = 0, 1, \dots\}$, можно свести изучение блуждания с шагом d к нерешетчатому блужданию.

Рассмотрим событие $\bigcup_{n=1}^{\infty} \{\xi_n = 0\}$. Блуждание называется возвратным, если

$$P\left\{\bigcup_{n=1}^{\infty} \{\xi_n = 0\}\right\} = 1, \quad (1)$$

и невозвратным — в противном случае. Заметим, что случайное событие

$$\bigcup_{n=1}^{\infty} \{\xi_n = 0\}$$

можно интерпретировать как событие, состоящее в том, что частица, блуждающая по целочисленной решетке прямой, хотя бы один раз вернется в исходное положение. Поэтому блуждание возвратно, если с вероятностью 1 частица вернется в исходное состояние хотя бы один раз (а, следовательно, и бесконечное число раз), так как после каждого возвращения весь процесс блуждания начинается сначала. Если блуждание невозвратно, то с положительной вероятностью частица может не вернуться в исходное состояние.

Предположим, что $P\{\bigcup_{n=1}^{\infty} \{\xi_n = 0\}\} > 0$. Введем обобщенную случайную величину ν , называемую временем первого возвращения блуждания в начальное состояние: $\nu = k$, если $\xi_n \neq 0$ при $0 < n < k-1$, $\xi_k = 0$. Если $\xi_n \neq 0$ для всех $n > 0$, то полагаем $\nu = +\infty$ (поскольку ν может принимать значение $+\infty$, ее и называют обобщенной случайной величиной).

Заметим, что

$$\bigcup_{n=1}^{\infty} \{\xi_n = 0\} = \bigcup_{i=1}^{\infty} \{\nu = i\},$$

причем при разных i события $\{\nu = i\}$ несовместны. Следовательно,

$$P\{\bigcup_{n=1}^{\infty} \{\xi_n = 0\}\} = \sum_{i=1}^{\infty} P\{\nu = i\} = \sum_{i=1}^{\infty} f_i.$$

Таким образом, блуждание возвратно тогда и только тогда, когда

$$\sum_{i=1}^{\infty} f_i = 1.$$

ПРИМЕР

Пусть $\{\xi_k\}$ — последовательность независимых случайных величин такая, что

$$\xi_k = \begin{cases} 1 & \text{с вероятностью } p, \\ -1 & \text{с вероятностью } q = 1 - p. \end{cases}$$

В этом случае блуждание $\{\xi_n, n = 1, 2, \dots\}$ называется простейшим случайным блужданием (никогда употребляется термин бернуллиевское случайное блуждание). Это блуждание было рассмотрено в примере 3, с. 125. Там найдена производящая функция случайной величины ν ; напомним, что если

$$f_n = P\{\nu = n\},$$

то

$$F(s) = \sum_{n=1}^{\infty} f_n s^n = 1 - (1 - 4pq s^2)^{1/2}.$$

Заметим теперь, что

$$F(1) = \sum_{i=1}^{\infty} f_i = 1 - (1 - 4pq)^{1/2} = 1 - |p - q|.$$

Поэтому при $p \neq q$ блуждание не возвратно, при $p = q = 1/2$ — блуждание возвратно (в этом случае блуждание называется простейшим симметричным блужданием).

Найдем условия, при которых блуждание возвратно.

Теорема 1. Для того чтобы блуждание было возвратным, необходимо и достаточно, чтобы

$$\sum_{n=1}^{\infty} P\{\zeta_n = 0\} = +\infty.$$

Доказательство. Отметим, что возвратность блуждания эквивалентна соотношению $P\{v = +\infty\} = 0$. Событие $\{\zeta_n = 0\}$ происходит лишь тогда, когда происходит одно из несовместных между собой событий $\{v = k\} \cap \{\zeta_n - \zeta_k = 0\}$, $k = 1, \dots, n$. Событие $\{v = k\}$ определяется величинами ξ_1, \dots, ξ_k , которые не зависят от ξ_{k+1}, \dots, ξ_n , которые, в свою очередь, определяют $\zeta_n - \zeta_k$. Поэтому события $\{v = k\}$, $\{\zeta_n - \zeta_k = 0\}$ независимы. Таким образом,

$$P\{\zeta_n = 0\} = \sum_{k=1}^n P\{v = k\} P\{\zeta_n - \zeta_k = 0\}.$$

Так как величины ξ_i одинаково распределены и независимы, то распределение $\sum_{i=l+1}^{l+N} \xi_i$ зависит лишь от числа слагаемых. Поэтому

$$P\{\zeta_n - \zeta_k = 0\} = P\left\{\sum_{i=k}^n \xi_i = 0\right\} = P\left\{\sum_{i=1}^{n-k} \xi_i = 0\right\} = P\{\zeta_{n-k} = 0\}.$$

Следовательно,

$$P\{\zeta_n = 0\} = \sum_{k=1}^n P\{v = k\} P\{\zeta_n - \zeta_k = 0\} \quad (n > 0). \quad (2)$$

Переходя к производящим функциям последовательностей $\{P\{v = k\}\}$ и $\{P\{\zeta_n = 0\}\}$, $(\sum_{n=0}^{\infty} \lambda^n P\{\zeta_n = 0\}, \sum_{k=1}^{\infty} \lambda^k P\{v = k\})$ соответственно), на основании леммы § 9, гл. II, из соотношения (2) получаем:

$$\sum_{n=1}^{\infty} \lambda^n P\{\zeta_n = 0\} = \left(\sum_{k=1}^{\infty} \lambda^k P\{v = k\}\right) \left(\sum_{k=0}^{\infty} \lambda^k P\{\zeta_k = 0\}\right).$$

Так как $P\{\xi_0 = 0\} = 1$, то при $|\lambda| < 1$

$$\sum_{k=1}^{\infty} \lambda^k P\{v = k\} = \frac{\sum_{n=1}^{\infty} \lambda^n P\{\xi_n = 0\}}{1 + \sum_{n=1}^{\infty} \lambda^n P\{\xi_n = 0\}}.$$

Пусть $\lambda \uparrow 1$. Тогда слева имеем

$$\sum_{k=1}^{\infty} P\{v = k\} = P\{v = +\infty\},$$

а справа

$$\frac{1}{1 + \lim_{\lambda \uparrow 1} \frac{1}{\sum_{n=1}^{\infty} \lambda^n P\{\xi_n = 0\}}}.$$

Это выражение равно 1 тогда и только тогда, когда

$$\lim_{\lambda \uparrow 1} \sum_{n=1}^{\infty} \lambda^n P\{\xi_n = 0\} = \sum_{n=1}^{\infty} P\{\xi_n = 0\} = +\infty.$$

Теорема доказана.

ПРИМЕР

Для простейшего случайного блуждания

$$P\{\xi_n = 0\} = \begin{cases} 0, & \text{если } n = 2k + 1; \\ C_{2k}^k p^k q^k, & \text{если } n = 2k. \end{cases}$$

Отсюда, используя формулу Стирлинга, находим, что при $n = 2k$

$$P\{\xi_n = 0\} \sim \frac{(4pq)^k}{(\pi k)^{1/2}}.$$

При $p \neq q$ $4pq < 1$ и, следовательно, $\sum P\{\xi_n = 0\} < +\infty$, в этом случае блуждание, согласно теореме 1, не возвратно. Если $p = q = \frac{1}{2}$, то $4pq = 1$ и

$$\sum P\{\xi_n = 0\} = +\infty.$$

В этом случае на основании теоремы 1 блуждание возвратно.

Целое натуральное число p называется периодом случайного блуждания, если для всех n существует такое $m < p$, что

$$P\{\xi_n = m \pmod{p}\} = 1 \quad (3)$$

и p — наибольшее из чисел, удовлетворяющих этому условию. Блуждание называется непериодическим, если $p = d$, где d — шаг блуждания.

Лемма 2. Если ξ — целочисленная случайная величина, для которой существует такое m , что $P\{\xi = m \pmod{p}\} = 1$, а $\varphi(z)$ — характеристическая функция величины ξ , то $\left|\varphi\left(\frac{2\pi}{p}\right)\right| = 1$. Если $\left|\varphi\left(\frac{2\pi}{p}\right)\right| = 1$, то при некотором m выполнено (3).

Доказательство. Если выполнено (1), то

$$\begin{aligned}\varphi(z) &= \sum_{k=-\infty}^{\infty} e^{iz(kp+m)} P\{\xi = kp + m\} = \\ &= e^{izm} \sum_{k=-\infty}^{\infty} P\{\xi = kp + m\} e^{izkp}.\end{aligned}$$

Поэтому

$$\varphi\left(\frac{2\pi}{p}\right) = e^{\frac{iz\pi m}{p}} \sum_{k=-\infty}^{\infty} P\{\xi = kp + m\} = e^{i\frac{2\pi m}{p}}.$$

Пусть $\left|\varphi\left(\frac{2\pi}{p}\right)\right| = 1$. Тогда

$$\left|\sum_k P\{\xi = k\} e^{i\frac{2\pi k}{p}}\right| = \sum_k P\{\xi = k\}.$$

Это возможно лишь в том случае, когда для всех k , для которых $p_k > 0$, аргументы комплексных чисел $e^{i\frac{2\pi k}{p}}$ (считаем, что аргумент принимает значения из $[0, 2\pi]$) совпадают, т. е. величина остатка от деления k на p одинакова для всех k , для которых $p_k > 0$. Это означает, что эти $k = lp + m$, где m — постоянно. Лемма доказана.

Теорема 2. Если p — период случайного блуждания, а $\varphi(z)$ — характеристическая функция величины ξ_1 , то $|\varphi(z)| = 1$ при $z \in \left(0, \frac{2\pi}{p}\right)$, $|\varphi(2\pi/p)| = 1$.

Доказательство. Из леммы 2 вытекает второе утверждение теоремы. Пусть p — период блуждания и $s \in (0, 2\pi/p)$ такая точка, что $|\varphi(z)| = 1$. Тогда $\varphi(s) = e^{i\alpha}$ при некотором $\alpha \in [0, 2\pi]$ и, следовательно,

$$\begin{aligned}e^{i\alpha} &= \sum_k p_k e^{iks}, \quad 1 = \sum_k p_k e^{i(ks-\alpha)}, \quad s = \sum_k p_k \cos(ks - \alpha) \\ &\quad (p_k = P\{\xi = k\}).\end{aligned}$$

Поэтому

$$0 = \sum_k p_k (1 - \cos(ks - \alpha)) = 2 \sum_k p_k \sin^2 \frac{ks - \alpha}{2}.$$

Обозначим множество тех k , для которых $p_k > 0$ через N . Тогда для всех $k \in N$

$$ks - \alpha = 2\pi l.$$

Поскольку все $k \in N$ имеют вид $k = rp + m$ (p — период блуждания, m — некоторое натуральное число между 0 и p , p — целое), то

$$rp \frac{s}{2\pi} + m \frac{s}{2\pi} - \frac{\alpha}{2\pi} = l.$$

Из того что p — максимальное число, для которого $k \in N$ представимы в виде $k = rp + m$, вытекает, что p — наибольший общий делитель разностей $k_1 - k_2$, $k_1, k_2 \in N$ и, следовательно, наибольший общий делитель разностей $r_1 - r_2$, где $r_i = [ki/p]$, $k_i \in N$, равен 1. Поскольку

$$(r_1 - r_2) \frac{ps}{2\pi} = r_1 p \frac{s}{2\pi} - r_2 p \frac{s}{2\pi} = l_1 - l_2$$

является целым числом при любых $r_1 - r_2$ и можно выбрать такие целые q_1, \dots, q_i и $r_1^{(i)}, r_2^{(i)}$, что

$$\sum q_i (r_1^{(i)} - r_2^{(i)}) = 1,$$

то $ps/2\pi$ целое число, $ps/2\pi > 1$, что противоречит предположению $s < p/2\pi$. Заметим, наконец, что характеристическая функция величины ξ_n равна произведению характеристических функций слагаемых, т. е. $\varphi^n(z)$. Если $|\varphi(s)| = 1$, то $|\varphi^n(s)| = 1$ для всех n , если $|\varphi^n(s)| = 1$ при некотором n , то $|\varphi(s)| = 1$. Теорема доказана.

Следствие. Пусть s — минимальный положительный корень уравнения $|\varphi(s)| = 1$. Тогда $s = 2\pi/p$, где p — период блуждания.

§ 7. Процесс восстановления

Процессом восстановления называется случайное блуждание с положительными шагами, т. е. это последовательность $\{\xi_n, n = 0, 1, \dots\}$, где $\xi_0 = 0$, $\xi_n = \xi_1 + \dots + \xi_n$, а $\{\xi_k, k = 1, 2, \dots\}$ — последовательность независимых одинаково распределенных неотрицательных случайных величин. Будем предполагать, что $P\{\xi_1 = 0\} < 1$, так как в противном случае $P\{\xi_n = 0\} = 1$ для всех n , поэтому изучение такого блуждания не представляет интереса.

Название «процесс восстановления» связано с такой интерпретацией. Имеется прибор, выходящий из строя через некоторое время ξ_1 , после чего он заменяется идентичным прибором, который работает время ξ_2 , а потом заменяется третьим прибором и т. д. Идентичность приборов означает, что времена их работы одинаково распределены. Моменты замены приборов $\{\xi_0, \xi_1, \dots, \xi_n, \dots\}$ называются моментами восстановления.

Обозначим через $v(t)$ число использованных приборов до момента t (включая работающий в этот момент). Тогда $v(t) = n$, если $\xi_{n-1} < t \leq \xi_n$.

$$N(t) = Mv(t)$$

называется функцией восстановления. Обозначим $\Phi_n(t) = P\{\xi_n < t\}$. Тогда для $n > 0$

$$P\{v(t) = n\} = P\{\xi_{n-1} < t \leq \xi_n\} = P\{\xi_{n-1} < t\} - P\{\xi_n < t\} = \Phi_{n-1}(t) - \Phi_n(t).$$

Поскольку $P\{v(t) = 1\} = P\{\xi_1 > t\} = 1 - \Phi_1(t)$, то естественно считать, что $\Phi_0(t) = 1$.

Следовательно,

$$N(t) = \sum_{n=1}^{\infty} n [\Phi_{n-1}(t) - \Phi_n(t)] = \sum_{n=0}^{\infty} \Phi_n(t).$$

ПРИМЕР

Вычислим $N(t)$, предполагая, что ξ_k имеет показательное распределение с параметром λ . Как известно (§ 8, гл. II), случайная величина ξ_n имеет распределение Эрланга. Следовательно, при $n \geq 1$

$$\Phi_n(t) = \int_0^t \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} dx.$$

Поэтому

$$\begin{aligned} N(t) &= \sum_{n=0}^{\infty} \Phi_n(t) = 1 + \sum_{n=1}^{\infty} \lambda \int_0^t \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x} dx = \\ &= 1 + \lambda \int_0^t \left(\sum_{n=1}^{\infty} \frac{(\lambda x)^{n-1}}{(n-1)!} \right) e^{-\lambda x} dx = 1 + \lambda t. \end{aligned}$$

Покажем, что функция восстановления $N(t)$ конечна при всех $t > 0$.

По предположению $P\{\xi_k < \delta\} < 1$ при достаточно малых δ . Пусть $\hat{\xi}_k = 0$, если $\xi_k < \delta$; $\hat{\xi}_k = 1$, если $\xi_k \geq \delta$. Обозначим через $[\alpha]$ целую часть α . Поскольку

$$\{\xi_n < t\} \subset \left\{ \sum_{k=1}^n \hat{\xi}_k \leq \left[\frac{t}{\delta} \right] \right\},$$

то

$$\begin{aligned} \Phi_n(t) &\leq \sum_{l < [t/\delta]} P\left\{ \sum_{k=1}^n \hat{\xi}_k = l \right\} = \sum_{l < [t/\delta]} C_n^l (1-\rho)^l \rho^{n-l} \leq \\ &\leq \sum_{l < [t/\delta]} \left(\frac{1-\rho}{\rho} \right)^l n^l \rho^n \leq k [t/\delta] n^{[t/\delta]} \rho^n, \end{aligned}$$

где k — некоторая постоянная. Таким образом,

$$N(t) \leq k[t/\delta] \sum_{n=1}^{\infty} n^{[t/\delta]} p^n < \infty,$$

так как ряд $\sum_{n=1}^{\infty} n^m p^n$ сходится при $0 < p < 1$ для всех m .

Далее будем рассматривать лишь случай целочисленных величин ξ_k . Введем обозначение:

$$q_k = \sum_{n=0}^{\infty} P\{\xi_n = k\}. \quad (1)$$

Тогда

$$N(t) = \sum_{n=0}^{\infty} \sum_{k \leq t} P\{\xi_n = k\} = \sum_{k \leq t} q_k.$$

Докажем теорему об асимптотике функции восстановления при $t \rightarrow \infty$. Эта теорема называется «теоремой восстановления».

Теорема. Если распределение величин нерешетчато, то $\lim_{k \rightarrow \infty} q_k = 1/M\xi_1$ (если $M\xi_1 = +\infty$, считаем $1/M\xi_1 = 0$).

Для доказательства нам понадобятся некоторые вспомогательные предложения.

Лемма 1. Пусть $\varphi(z)$ — характеристическая функция величины ξ_1 . Тогда

$$q_k = \lim_{\lambda \uparrow 1} (1/\pi i) \int_{-\pi}^{\pi} [1/(1 - \lambda\varphi(z))] \sin kz \, dz. \quad (2)$$

Доказательство. Так как $\xi_n = \xi_1 + \dots + \xi_n$ и ξ_k независимы и одинаково распределены, то характеристическая функция величины ξ_n будет $\varphi^n(z)$. Следовательно, на основании формулы обращения для целочисленных величин ($k > 0$) имеем

$$P\{\xi_n = k\} = (1/2\pi) \int_{-\pi}^{\pi} \varphi^n(z) e^{-ikz} \, dz.$$

Поэтому

$$\begin{aligned} q_k &= \lim_{\lambda \uparrow 1} \sum_{n=0}^{\infty} \lambda^n P\{\xi_n = k\} = \lim_{\lambda \uparrow 1} \sum_{n=0}^{\infty} \int_{-\pi}^{\pi} \lambda^n \varphi^n(z) e^{-ikz} \, dz = \\ &= \lim_{\lambda \uparrow 1} \int_{-\pi}^{\pi} \sum_{n=0}^{\infty} \lambda^n \varphi^n(z) e^{-ikz} \, dz. \end{aligned}$$

Почленное интегрирование ряда возможно, так как при $0 < \lambda < 1$ он сходится равномерно. Суммируя в последнем выражении геометрическую прогрессию под знаком интеграла, получим

$$q_k = \lim_{\lambda \uparrow 1} (1/2\pi) \int_{-\pi}^{\pi} [1/1 - \lambda\varphi(z)] e^{-ikz} \, dz. \quad (3)$$

Аналогично

$$0 = \lim_{\lambda \uparrow 1} \sum_{n=0}^{\infty} \lambda^n \mathbf{P} \{ \xi_n = -k \} = \lim_{\lambda \uparrow 1} (1/2\pi) \int_{-\pi}^{\pi} \{ e^{ikz} / 1 - \lambda \varphi(z) \} dz. \quad (4)$$

Вычитая из равенства (3) равенство (4), получим правую часть формулы (2).

Лемма 2. Если ξ_k имеют нерешетчатое распределение, то

$$\int_{-\pi}^{\pi} [z^2 / |1 - \varphi(z)|] dz < \infty.$$

Доказательство. Для всякого $\delta > 0$ $\inf_{\delta < |z| < \pi} |1 - \varphi(z)| > 0$.

Поэтому достаточно доказать, что при достаточно малых $\delta > 0$ функция $z^2 / |1 - \varphi(z)|$ ограничена при $|z| \leq \delta$. Пусть $\mathbf{P} \{ \xi_1 = l \} = \alpha > 0$. Тогда

$$\begin{aligned} |1 - \varphi(z)| &\geq \operatorname{Re}(1 - \varphi(z)) = \sum_k \mathbf{P} \{ \xi_1 = k \} (1 - \cos kz) \geq \\ &\geq \alpha (1 - \cos lz) = 2\alpha \sin^2(lz/2). \end{aligned}$$

Следовательно, $z^2 / |1 - \varphi(zs)|$ ограничено при $|lz| \leq \pi$.

Лемма 3.

$$\lim_{k \rightarrow \infty} (q_{k-1} + q_{k+1} - 2q_k) = 0.$$

Доказательство. На основании леммы 1

$$\begin{aligned} q_{k-1} + q_{k+1} - 2q_k &= (1/\pi i) \lim_{\lambda \uparrow 1} \int_{-\pi}^{\pi} [1/1 - \lambda \varphi(z)] [\sin(k-1)z + \\ &+ \sin(k+1)z - 2 \sin kz] dz = \\ &= (1/\pi i) \lim_{\lambda \uparrow 1} \int_{-\pi}^{\pi} [1/1 - \lambda \varphi(z)] 2(\cos z - 1) \sin kz dz. \end{aligned}$$

Заметим, что поскольку $\operatorname{Re}(1 - \varphi(z)) \geq 0$, $1 - \lambda > 0$, то

$$\begin{aligned} 1/|1 - \lambda \varphi(z)| &= 1/|(1 - \lambda) + \lambda(1 - \varphi(z))| \leq \\ &\leq 1/|\lambda(1 - \varphi(z))| = 1/\lambda |1 - \varphi(z)|. \end{aligned}$$

Из леммы 2 вытекает интегрируемость

$$|1 - \cos z| / |1 - \varphi(z)|.$$

Поэтому на основании теоремы о предельном переходе под знаком интеграла

$$\begin{aligned} \lim_{\lambda \uparrow 1} \int_{-\pi}^{\pi} [1/1 - \lambda \varphi(z)] (\cos z - 1) \sin kz dz &= \\ &= \int_{-\pi}^{\pi} [(\cos z - 1) / |1 - \varphi(z)|] \sin kz dz. \end{aligned}$$

Таким образом,

$$q_{k-1} + q_{k+1} - 2q_k = (1/\pi i) \int_{-\pi}^{\pi} \frac{\cos z - 1}{1 - \varphi(z)} \sin kz \, dz.$$

Правая часть этого равенства стремится к нулю при $k \rightarrow \infty$ (на основании теоремы Римана о коэффициентах Фурье интегрируемой функции).

Лемма 4. Пусть $r_k = \sum_{l \geq k} P(\xi_1 = l)$. Тогда для всех n

$$\sum r_k q_{n-k} = 1. \quad (5)$$

Доказательство. Рассмотрим производящую функцию последовательности q_k

$$Q(\lambda) = \sum_{k=0}^{\infty} q_k \lambda^k = \sum_{k=0}^{\infty} \lambda^k \sum_{n=0}^{\infty} P(\xi_n = k) = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \lambda^k P(\xi_n = k).$$

Обозначим через $\psi(\lambda)$ производящую функцию величины ξ_1

$$\psi(\lambda) = \sum_{k=0}^{\infty} P(\xi_1 = k) \lambda^k.$$

Тогда

$$\sum_{k=0}^{\infty} \lambda^k P(\xi_n = k) = \psi^n(\lambda)$$

(как производящая функция суммы одинаково распределенных независимых случайных величин). Следовательно, при $|\lambda| < 1$

$$Q(\lambda) = \sum_{n=0}^{\infty} \psi^n(\lambda) = 1/[1 - \psi(\lambda)],$$

так как $|\psi(\lambda)| < 1$. Запишем это соотношение так:

$$Q(\lambda) \frac{1 - \psi(\lambda)}{1 - \lambda} = \frac{1}{1 - \lambda}. \quad (6)$$

В правой части равенства имеем производящую функцию последовательности $\{a_n = 1, n = 0, 1, \dots\}$. Далее

$$\begin{aligned} \frac{1 - \psi(\lambda)}{1 - \lambda} &= \sum_{k=0}^{\infty} P(\xi_1 = k) \frac{1 - \lambda^k}{1 - \lambda} = \sum_{k=1}^{\infty} P(\xi_1 = k) (1 + \lambda + \dots + \lambda^{k-1}) = \\ &= \sum_{k=0}^{\infty} r_k \lambda^k. \end{aligned}$$

Воспользуемся теперь леммой, § 9, гл. II, по которой производящая функция свертки двух последовательностей есть произведение

производящих функций этих последовательностей, и получим из (6) равенство (5).

Доказательство теоремы. Из равенства (5) вытекает, что $r_0 q_n \leq 1$, и так как $r_0 > 0$, то $q_n \leq 1/r_0$, т. е. последовательность q_n ограничена. Пусть

$$\overline{\lim} q_n = s$$

и n_m — такая последовательность, что $q_{n_m} \rightarrow s$. Тогда на основании леммы 3

$$\lim_{m \rightarrow \infty} (q_{n_m-1} + q_{n_m+1} - 2s) = 0.$$

Таким образом,

$$\lim (q_{n_m-1} + q_{n_m+1}) = 2s,$$

но

$$\underline{\lim} q_{n_m-1} + \overline{\lim} q_{n_m+1} \geq \lim (q_{n_m-1} + q_{n_m+1}) = 2s,$$

так что

$$\underline{\lim} q_{n_m-1} \geq s, \quad \overline{\lim} q_{n_m-1} \leq s, \quad \lim q_{n_m-1} = s.$$

Аналогично

$$\lim q_{n_m} = s.$$

Следовательно,

$$\lim q_{n_m \pm 2} = s, \quad \lim q_{n_m \pm k} = s,$$

каково бы ни было k .

Теперь имеем

$$1 \geq \sum_{k=0}^N q_{n_m-k} r_k$$

каково бы ни было N ($n_m > N$).

Переходя к пределу при $m \rightarrow \infty$, получим

$$1 \geq s \sum_{k=0}^N r_k.$$

Если $M\xi_1 = \sum_{k=0}^{\infty} r_k = +\infty$, то получаем

$$s \leq \left(\sum_{k=0}^N r_k \right)^{-1} \rightarrow 0 \quad (N \rightarrow \infty).$$

Это доказывает теорему для случая бесконечного математического ожидания.

Для конечного $M\xi_1$ имеем

$$\overline{\lim}_{n \rightarrow \infty} q_n \leq 1/M\xi_1. \quad (7)$$

Пусть $t = \underline{\lim} q_k$, $q_{n_m} \rightarrow t$. Исходя из соотношения

$$q_{n_m-1} + q_{n_m+1} - 2t \rightarrow 0,$$

так, как и выше, находим

$$\lim q_{n_m+1} = t, \quad \lim q_{n_m \pm k} = t.$$

Поэтому

$$1 = \sum_{k=0}^{n_m} q_{n_m-k} r_k \leq \sum_{k=0}^N q_{n_m-k} r_k + c \sum_{k>N} r_k,$$

где $c = \sup_k q_k$ (сходимость ряда $\sum r_k = M\xi_1$ предполагается). Переходя к пределу при $m \rightarrow \infty$, получим

$$1 \leq t \sum_{k=0}^N r_k + c \sum_{k>N} r_k.$$

Пусть $N_k \rightarrow \infty$, тогда

$$1 \leq t \sum_{k=0}^{\infty} r_k = t M\xi_k.$$

Отсюда

$$\underline{\lim} q_k \geq \frac{1}{M\xi_1}. \quad (8)$$

Совместно с неравенством (7) это дает доказательство теоремы для конечного $M\xi_1$.

Глава IV

ЦЕПИ МАРКОВА

§ 1. Определение цепи Маркова.

Простейшие свойства

Пусть некоторая система меняет свое состояние в моменты времени $t = 1, 2, \dots$, множество возможных состояний системы X (фазовое пространство системы) предполагается конечным либо счетным. Среди реальных систем важный класс образуют системы, у которых переходы из состояния в состояние происходят случайным образом, при этом вероятности перехода из данного состояния в данный момент не зависят от того, как вела себя система в предыдущие моменты времени. Такие системы называются марковскими (или цепями Маркова). Дадим определение цепи Маркова.

Рассмотрим случайные величины со значениями в фазовом пространстве системы X , пусть $\{\xi_n, n = 0, 1, 2, \dots\}$ обозначают состояния системы в последовательные моменты времени $t = 0, 1, 2, \dots$ (Так как X не более чем счетно, то его можно отождествить с подмножеством натуральных чисел, и тогда ξ_n — целочисленные величины).

Последовательность $\{\xi_n, n = 0, 1, 2, \dots\}$ называется *цепью Маркова*, когда для всех $n \geq 1, x_0, \dots, x_{n+1} \in X$ выполнено равенство

$$P\{\xi_{n+1} = x_{n+1} / \xi_0 = x_0, \dots, \xi_n = x_n\} = P\{\xi_{n+1} = x_{n+1} / \xi_n = x_n\}, \quad (1)$$

если только условная вероятность слева в (1) определена, т. е. если

$$P\{\xi_0 = x_0, \dots, \xi_n = x_n\} > 0.$$

Функция $P_n(x, y), n = 0, 1, \dots, x, y \in X$ называется вероятностью перехода цепи Маркова, если выполнены следующие два условия:

1) для каждого $n \geq 1$ и $x \in X$

$$\sum_{y \in X} P_n(x, y) = 1, \quad P_n(x, y) \geq 0;$$

2) для всех x , для которых $P\{\xi_{n-1} = x\} > 0$,

$$P_n(x, y) = P\{\xi_n = y / \xi_{n-1} = x\}.$$

Переходная вероятность всегда существует. Если $P\{\xi_{n-1} = x\} > 0$, то положим

$$P_n(x, y) = P\{\xi_n = y / \xi_{n-1} = x\}$$

и для таких x условие 1) выполнено. Если же $P\{\xi_{n-1} = x\} = 0$, то $P_n(x, y)$ можем определить произвольно, лишь бы выполнялось 1).

Теорема 1. Если $P_n(x, y)$ — вероятность перехода цепи Маркова $\{\xi_n, n \geq 0\}$, то для всех $n > 0$ и $x_0, \dots, x_n \in X$ выполнено соотношение

$$P\{\xi_0 = x_0, \dots, \xi_n = x_n\} = P\{\xi_0 = x_0\} P_1(x_0, x_1) \dots P_n(x_{n-1}, x_n). \quad (2)$$

Доказательство. Соотношение (2) установим по индукции. При $n = 1$

$$\begin{aligned} P\{\xi_0 = x_0, \xi_1 = x_1\} &= P\{\xi_0 = x_0\} P\{\xi_1 = x_1 / \xi_0 = x_0\} = \\ &= P\{\xi_0 = x_0\} P_1(x_0, x_1), \end{aligned}$$

если только $P\{\xi_0 = x_0\} > 0$. В противном случае

$$P\{\xi_0 = x_0, \xi_1 = x_1\} = 0 = P\{\xi_0 = x_0\} P_1(x_0, x_1).$$

Пусть равенство (2) справедливо для $n = k$. Покажем, что тогда оно справедливо и для $n = k + 1$. Если $P\{\xi_0 = x_0, \dots, \xi_k = x_k\} > 0$, то на основании (1)

$$P\{\xi_0 = x_0, \dots, \xi_k = x_k, \xi_{k+1} = x_{k+1}\} = \\ = P\{\xi_0 = x_0, \dots, \xi_k = x_k\} P_{k+1}(x_k, x_{k+1}).$$

Выразив $P\{\xi_0 = x_0, \dots, \xi_k = x_k\}$ по формуле (2) (это возможно согласно предположению индукции), получим формулу (2) для $n = k + 1$. Если же $P\{\xi_0 = x_0, \dots, \xi_k = x_k\} = 0$, то

$$P\{\xi_0 = x_0, \dots, \xi_{k+1} = x_{k+1}\} = 0 = P\{\xi_0 = x_0, \dots, \xi_k = x_k\} \times \\ \times P_{k+1}(x_k, x_{k+1}) = P\{\xi_0 = x_0\} P_1(x_0, x_1) \dots P_{k+1}(x_k, x_{k+1})$$

(мы опять воспользовались формулой (2) для $n = k$). Таким образом, из справедливости (2) для $n = k$ вытекает справедливость (2) для $n = k + 1$.

Замечание. Если для последовательности $\{\xi_n, n \geq 0\}$ при всех $n > 0$ и $x_0, \dots, x_n \in X$ выполнено соотношение (2), то эта последовательность есть цепь Маркова с вероятностью перехода $P_n(x, y)$. Действительно, если $P\{\xi_0 = x_0, \dots, \xi_n = x_n\} > 0$, то на основании (2)

$$P\{\xi_{n+1} = x_{n+1} / \xi_0 = x_0, \dots, \xi_n = x_n\} = \\ = \frac{P\{\xi_0 = x_0, \dots, \xi_{n+1} = x_{n+1}\}}{P\{\xi_0 = x_0, \dots, \xi_n = x_n\}} = P_{n+1}(x_n, x_{n+1}). \quad (3)$$

Кроме того,

$$P\{\xi_n = x_n, \xi_{n+1} = x_{n+1}\} = \sum_{x_0, \dots, x_{n-1}} P\{\xi_0 = x_0, \dots, \xi_{n-1} = x_{n-1}, \\ \xi_n = x_n, \xi_{n+1} = x_{n+1}\} = \\ = \sum_{x_0, \dots, x_{n-1}} P\{\xi_0 = x_0, \dots, \xi_{n-1} = x_{n-1}, \xi_n = x_n\} P_{n+1}(x_n, x_{n+1}) = \\ = P\{\xi_n = x_n\} P_{n+1}(x_n, x_{n+1}).$$

Поскольку $P\{\xi_n = x_n\} > 0$, то

$$P_{n+1}(x_n, x_{n+1}) = P\{\xi_n = x_n, \xi_{n+1} = x_{n+1}\} / P\{\xi_n = x_n\} = \\ = P\{\xi_{n+1} = x_{n+1} / \xi_n = x_n\}. \quad (4)$$

Сравнивая (3) и (4), убеждаемся, что для нашей последовательности выполнено равенство (1), т. е. она является цепью Маркова, а для $P_n(x, y)$ выполнено (2), следовательно, $P_n(x, y)$ — вероятность перехода цепи Маркова $\{\xi_n, n \geq 0\}$.

Следствие 1. Рассмотрим последовательность $\hat{\xi}_n = \xi_{n+m}$, $n \geq 0$. Она является цепью Маркова с вероятностью перехода

$$\hat{P}_n(x, y) = P_{m+n}(x, y).$$

Действительно, подставив вместо n в соотношение (2) $n + m$ и просуммировав по x_0, \dots, x_{m-1} , получим

$$P\{\xi_m = x_m, \dots, \xi_{n+m} = x_{n+m}\} = \\ = P\{\xi_m = x_m\} P_{m+1}(x_m, x_{m+1}) \dots P_{m+n}(x_{m+n-1}, x_{m+n}). \quad (5)$$

Следовательно, для последовательности $\hat{\xi}_n$ выполнено соотношение (2), если $P_n(x, y)$ заменить на $\hat{P}_n(x, y)$. Остается воспользоваться замечанием к теореме.

Следствие 2. При $m < n$ случайные величины ξ_0, \dots, ξ_{m-1} и ξ_{m+1}, \dots, ξ_n условно независимы при заданном ξ_m , это означает, что с вероятностью 1

$$\begin{aligned} P\{\xi_0 = x_0, \dots, \xi_{m-1} = x_{m-1}, \xi_{m+1} = x_{m+1}, \dots, \xi_n = x_n / \xi_m\} = \\ = P\{\xi_0 = x_0, \dots, \xi_{m-1} = x_{m-1} / \xi_m\} P\{\xi_{m+1} = x_{m+1}, \dots, \xi_n = \\ = x_n / \xi_m\}. \end{aligned} \quad (6)$$

Действительно, пусть x_m таково, что $P\{\xi_m = x_m\} > 0$. Тогда

$$\begin{aligned} P\{\xi_0 = x_0, \dots, \xi_{m-1} = x_{m-1}, \xi_m = x_m, \dots, \xi_n = x_n\} = \\ = P\{\xi_0 = x_0\} P_1(x_0, x_1) \dots P_n(x_{n-1}, x_n) = \\ = [P\{\xi_0 = x_0\} P_1(x_0, x_1) \dots P_{m-1}(x_{m-1}, x_m)] [P_{m+1}(x_m, x_{m+1}) \dots \\ \dots P_n(x_{n-1}, x_n)] = \frac{P\{\xi_0 = x_0, \dots, \xi_m = x_m\}}{P\{\xi_m = x_m\}} [P\{\xi_m = x_m\} P_{m+1}(x_m, \\ x_{m+1}) \dots P_n(x_{n-1}, x_n)] = \frac{P\{\xi_0 = x_0, \dots, \xi_m = x_m\}}{P\{\xi_m = x_m\}} P\{\xi_m = x_m, \xi_{m+1} = \\ = x_{m+1}, \dots, \xi_n = x_n\} \end{aligned}$$

(мы воспользовались равенством (5) при $n = n - m$). Разделив обе части равенства на $P\{\xi_m = x_m\}$, найдем

$$\begin{aligned} \frac{P\{\xi_0 = x_0, \dots, \xi_{m-1} = x_{m-1}, \xi_m = x_m, \xi_{m+1} = x_{m+1}, \dots, \xi_n = x_n\}}{P\{\xi_m = x_m\}} = \\ = \frac{P\{\xi_0 = x_0, \dots, \xi_m = x_m\}}{P\{\xi_m = x_m\}} \frac{P\{\xi_m = x_m, \xi_{m+1} = x_{m+1}, \dots, \xi_n = x_n\}}{P\{\xi_m = x_m\}}, \end{aligned}$$

откуда и вытекает (6).

$P_n(x, y)$ при фиксированном n называется вероятностью перехода на n -м шаге. Введем вероятности перехода за несколько шагов: при $n < m$

$$P_{n, m}(x, y) = \sum_{x_{n+1}, \dots, x_{m-1}} P_{n+1}(x, x_{n+1}) \dots P_m(x_{m-1}, y) \quad (7)$$

(при $m = n + 1$ $P_{n, n+1}(x, y) = P_{n+1}(x, y)$). Покажем, что

$$P_{n, m}(x, y) = P\{\xi_m = y / \xi_n = x\}, \quad (8)$$

если только $P\{\xi_n = x\} > 0$. Имеем

$$\begin{aligned} P\{\xi_n = x, \xi_{n+1} = x_{n+1}, \dots, \xi_{m-1} = x_{m-1}, \xi_m = y\} = \\ = P\{\xi_n = x\} P_{n+1}(x, x_{n+1}) \dots P_{m-1}(x_{m-2}, x_{m-1}) P_m(x_{m-1}, y). \end{aligned}$$

Просуммировав по x_{n+1}, \dots, x_{m-1} , получим

$$P\{\xi_n = x, \xi_m = y\} = P\{\xi_n = x\} P_{n, m}(x, y),$$

отсюда и вытекает (8). Пусть x_0, \dots, x_n такие, что $P\{\xi_0 = x_0, \dots, \xi_n = x_n\} > 0$. Тогда

$$\begin{aligned} & P\{\xi_0 = x_0, \dots, \xi_n = x_n, \xi_m = y\} = \\ &= \sum_{x_{n+1}, \dots, x_{m-1}} P\{\xi_0 = x_0, \dots, \xi_n = x_n\} P_{n+1}(x_n, \\ & x_{n+1}) \dots P_m(x_{m-1}, y) = P\{\xi_0 = x_0, \dots, \xi_n = x_n\} P_{n,m}(x_n, y). \end{aligned}$$

Следовательно,

$$P_{n,m}(x_n, y) = P\{\xi_m = y / \xi_0 = x_0, \dots, \xi_n = x_n\}.$$

Значит, при $n < m$

$$P\{\xi_m = y / \xi_0 = x_0, \dots, \xi_n = x_n\} = P\{\xi_m = y / \xi_n = x_n\}. \quad (9)$$

Из формулы (7) вытекает следующее соотношение для вероятностей перехода за 1 шаг: при $k < n < m$

$$P_{k,m}(x, y) = \sum_{z \in X} P_{k,n}(x, z) P_{n,m}(z, y) \quad (10)$$

(равенство (10) называется уравнением Колмогорова — Чепмена). Действительно, на основании формулы (7),

$$\begin{aligned} & \sum_{z \in X} P_{k,n}(x, z) P_{n,m}(z, y) = \\ &= \sum_z \sum_{x_{k+1}, \dots, x_{n-1}} P_{n,k+1}(x, x_{k+1}) \dots P_{n-1,n}(x_{n-1}, z) \times \\ & \times \sum_{x_{n+1}, \dots, x_{m-1}} P_{n,n+1}(z, x_{n+1}) P_{m-1,m}(x_{m-1}, y) = \\ &= \sum_{x_{k+1}, \dots, x_{n-1}, x_n, x_{n+1}, \dots, x_{m-1}} P_{k,k+1}(x, x_{k+1}) \dots \\ & \dots P_{n-1,n}(x_{n-1}, x_n) P_{n,n+1}(x_n, x_{n+1}) \dots P_{m-1,m}(x_{m-1}, y) = P_{k,m}(x, y) \end{aligned}$$

(мы заменили переменную суммирования z через x_n , обе они пробегают X).

Теорема 2. Если $\{\xi_n, n \geq 0\}$ — цепь Маркова и $0 < n_0 < n_1 < \dots < n_k < \dots$, то последовательность $\hat{\xi}_k = \xi_{n_k}$ — также цепь Маркова с вероятностью перехода

$$\hat{P}_k(x, y) = P_{n_{k-1}, n_k}(x, y).$$

Доказательство. Рассмотрим вероятности

$$P\{\hat{\xi}_0 = y_0, \dots, \hat{\xi}_n = y_n\} = P\{\xi_{n_0} = y_0, \dots, \xi_{n_k} = y_n\}.$$

На основании формулы (9)

$$\begin{aligned} & P\{\xi_0 = x_0, \dots, \xi_{n-1} = x_{n_{k-1}}, \xi_{n_k} = x_{n_k}\} = \\ &= P\{\xi_0 = x_0, \dots, \xi_{n_{k-1}} = x_{n_{k-1}}\} \hat{P}_k(x_{n_{k-1}}, x_{n_k}). \end{aligned}$$

Просуммировав это равенство по x_i , $i \neq n_0, n_1, \dots, n_{k-1}$, получим

$$\begin{aligned} P\{\xi_{n_0} = x_{n_0}, \dots, \xi_{n_k} = x_{n_k}\} &= P\{\xi_{n_0} = x_{n_0}, \dots, \xi_{n_{k-1}} = \\ &= x_{n_{k-1}}\} \widehat{P}_k(x_{n_{k-1}}, x_{n_k}). \end{aligned}$$

Следовательно,

$$\begin{aligned} P\{\xi_{n_0} = x_{n_0}, \dots, \xi_{n_k} = x_{n_k}\} &= P\{\xi_{n_0} = x_{n_0}\} \widehat{P}_1(x_{n_0}, x_{n_1}) \dots \\ &\dots \widehat{P}_k(x_{n_{k-1}}, x_{n_k}). \end{aligned}$$

Полагая $x_{n_0} = y_0, x_{n_1} = y_1, \dots, x_{n_k} = y_k$, получим

$$P\{\widehat{\xi}_0 = y_0, \dots, \widehat{\xi}_n = y_n\} = P\{\widehat{\xi}_0 = y_0\} \widehat{P}_1(y_0, y_1) \dots \widehat{P}_k(y_{k-1}, y_k).$$

Остается воспользоваться замечанием к теореме 1.

Рассмотрим примеры цепей Маркова.

1. Пусть $\xi_0, \xi_1, \dots, \xi_n$ — последовательность независимых дискретных случайных величин, $X = \{x_1, x_2, \dots\}$ — множество значений величин ξ_n (каждая из величин может принимать значения из «своего» множества X_n , тогда $X = \bigcup_k X_k$, это множество счетно, если все X_k не более чем счетные).

Поскольку на основании независимости

$$P\{\xi_n = x_n / \xi_0, \dots, \xi_{n-1}\} = P\{\xi_n = x_n\} = P\{\xi_n = x_n / \xi_{n-1}\},$$

то $\{\xi_n\}$ — цепь Маркова, вероятность перехода которой

$$P_n(x, y) = P_n(y) = P\{\xi_n = y\}.$$

2. Пусть $\xi_0, \eta_1, \eta_2, \dots$ — независимые целочисленные случайные величины, $q_k(y) = P\{\eta_k = y\}$. Положим $\xi_n = \xi_0 + \eta_1 + \dots + \eta_n$. Тогда

$$\begin{aligned} P\{\xi_n = x_n / \xi_0 = x_0, \dots, \xi_{n-1} = x_{n-1}\} &= \\ &= P\{\xi_n = x_n / \xi_0 = x_0, \eta_1 = x_1 - x_0, \dots, \eta_{n-1} = x_{n-1} - x_{n-2}\} = \\ &= P\{x_{n-1} + \eta_n = x_n / \xi_0 = x_0, \eta_1 = x_1 - x_0, \dots, \eta_{n-1} = x_{n-1} - x_{n-2}\} = \\ &= P\{\eta_n = x_n - x_{n-1}\}. \end{aligned}$$

Мы воспользовались независимостью величин. Кроме того,

$$\begin{aligned} P\{\xi_n = x_n / \xi_{n-1} = x_{n-1}\} &= P\{\eta_n = x_n - x_{n-1} / \xi_{n-1}\} = \\ &= P\{\eta_n = x_n - x_{n-1}\}, \end{aligned}$$

поскольку η_n не зависит от ξ_{n-1} . Следовательно, $\{\xi_n\}$ — цепь Маркова с вероятностью перехода

$$P_n(x, y) = q_n(y - x).$$

§ 2. Однородные цепи Маркова

Цепь Маркова $\{\xi_n, n = 0, 1, 2, \dots\}$ называется однородной, если вероятность перехода $P_n(x, y)$ не зависит от n : $P_n(x, y) = P(x, y)$. В этом случае вероятность перехода за несколько шагов $P_{n, m}(x, y)$ ($m > n$) зависит лишь от разности $m - n$. Обозначим

$$P_{n, m}(x, y) = P(m - n, x, y).$$

Формулу для вероятности перехода за несколько шагов удобно записать, используя матричные обозначения. Пронумеруем состояния системы: $X = \{x_1, x_2, \dots\}$. Обозначим через Π бесконечную матрицу с элементами $P(x_i, y_j)$ (i — номер строки, j — столбца), а через $\Pi(n)$ — бесконечную матрицу с элементами $P(n, x_i, y_j)$. Тогда

$$\Pi(n) = \Pi^n \quad (1)$$

(справа стоит степень матрицы Π). Доказательство формулы (1) проведем по индукции. При $n = 1$ она справедлива на основании равенства $P(x_i, y_j) = P(1, x_i, y_j)$. Предположим, что она справедлива при $n = k$. Так как по формуле (10), § 1,

$$P(k+1, x_i, y_j) = \sum_l P(k, x_i, x_l) P(1, x_l, y_j),$$

то

$$\Pi(k+1) = \Pi(k) \Pi = \Pi^k \Pi = \Pi^{k+1}.$$

При изучении цепей Маркова удобно не фиксировать начальное состояние системы, а вместо одной случайной последовательности рассматривать семейство случайных последовательностей $\{\xi_n^x, n \geq 0\}$, $x \in X$, имеющих одинаковые вероятности перехода, при этом $\xi_0^x = x$, и цепь Маркова ξ_n^x трактуется как эволюция системы, находящейся в начальный момент в состоянии x . В дальнейшем вероятности событий, связанных с цепью Маркова $\{\xi_n^x\}$, будем обозначать через P_x , а математическое ожидание по этой вероятности — через M_x , опуская при этом индекс x у последовательности. Например,

$$P\{\xi_n^x = y\} = P\{\xi_n = y\},$$

а для функции φ на X

$$M\varphi(\xi_n^x) = M_x \varphi(\xi_n).$$

Введем случайную величину τ — время возвращения в начальное состояние. Полагаем $\tau = n$ при $\xi_1^x \neq x, \dots, \xi_{n-1}^x \neq x, \xi_n^x = x$. Если для всех $n \geq 1$ $\xi_n^x \neq x$, то считаем $\tau = +\infty$. Установим условие, при котором $P_x\{\tau < \infty\} = 1$. Такое состояние x называется возвратным.

Теорема 1. Для того чтобы состояние x было возвратным, необходимо и достаточно, чтобы

$$\sum_n P_n(x, x) = +\infty. \quad (2)$$

Доказательство. Обозначим $P_x\{\tau = k\} = q_k$. Тогда состояние x возвратно, если $\sum_k q_k = 1$, и невозвратно при $\sum_k q_k < 1$. Если $\{\xi_n^x = x\}$, то $\tau \leq n$ и, следовательно,

$$\{\xi_n^x = x\} = \bigcup_{k=1}^n \{\xi_n^x = x\} \cap \{\tau = k\}.$$

Так как события в правой части равенства несовместны, то

$$P\{\xi_n^x = x\} = \sum_{k=1}^n P(\{\xi_n^x = x\} \cap \{\tau = k\}).$$

Поскольку

$$\begin{aligned} P(\{\xi_n^x = x\} \cap \{\tau = k\}) &= P_x\{\xi_1^x \neq x, \dots, \xi_{k-1}^x \neq x, \xi_k^x = x, \dots, \\ \xi_n^x = x\} &= \sum_{x_1 \neq x, \dots, x_{k-1} \neq x} P(x, x_1) \dots P(x_{k-1}, x) P(n-k, x, x) = \\ &= P_x\{\tau = k\} P(n-k, x, x) = q_k P(n-k, x, x), \end{aligned}$$

то

$$P(n, x, x) = \sum_{k=1}^{n-1} q_k P(n-k, x, x) + q_n. \quad (3)$$

Возьмем $0 < \lambda < 1$, умножим это равенство на λ^n и просуммируем по n от 1 до ∞ . Получим

$$\begin{aligned} \sum_{n=1}^{\infty} \lambda^n P(n, x, x) &= \sum_{n=2}^{\infty} \lambda^n \sum_{k=1}^{n-1} q_k P(n-k, x, x) + \sum_{n=1}^{\infty} q_n \lambda^n = \\ &= \sum_{n=1}^{\infty} q_n \lambda^n + \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \lambda^{k+m} q_k P(m, x, x) = \\ &= \sum_{n=1}^{\infty} q_n \lambda^n \left(1 + \sum_{m=1}^{\infty} P(m, x, x) \lambda^m\right). \end{aligned}$$

Следовательно,

$$\sum_{n=1}^{\infty} q_n \lambda^n = \frac{\sum_{n=1}^{\infty} P(n, x, x) \lambda^n}{1 + \sum_{n=1}^{\infty} P(n, x, x) \lambda^n}. \quad (4)$$

Тогда

$$\sum_{n=1}^{\infty} q_n = \left(1 + \lim_{\lambda \uparrow 1} \frac{1}{\sum_{n=1}^{\infty} P(n, x, x) \lambda^n}\right)^{-1}.$$

Если выполнено (2), то $\sum_{n=1}^{\infty} q_n = 1$, так как предел справа равен 0, если же

$$\lim_{\lambda \uparrow 1} \sum_{n=1}^{\infty} P(n, x, x) \lambda^n = \sum_{n=1}^{\infty} P(n, x, x) < \infty,$$

то

$$\sum_{n=1}^{\infty} q_n = \sum P(n, x, x) / (1 + \sum P(n, x, x)) < 1.$$

Пусть y — некоторое состояние. Обозначим через τ^y — момент первого попадания в состояние y . Имеем $\tau^y = n$, если $\xi_1 \neq y, \dots, \xi_{n-1} \neq y, \xi_n = y$, и $\tau^y = \infty$, если для всех $n \geq 1$ $\xi_n \neq y$. Пусть

$$q_n(x, y) = P_n\{\tau^y = n\}.$$

Аналогично соотношению (3) находим, что

$$P(n, x, y) = q_n(x, y) + \sum_{k=1}^{n-1} q_k(x, y) P(n-k, y, y). \quad (5)$$

Из равенства (4) вытекает следующее утверждение.

Теорема 2. Если состояние y невозвратно, то для всех $x \in X$

$$\sum_{n=1}^{\infty} P(n, x, y) < \infty. \quad (6)$$

Доказательство. По формуле (5)

$$\begin{aligned} \sum_{n=1}^{\infty} P(n, x, y) &= \sum_{n=1}^{\infty} q_n(x, y) + \sum_{n=2}^{\infty} \sum_{k=1}^{n-1} q_k(x, y) P(n-k, y, y) = \\ &= \sum_{n=1}^{\infty} q_n(x, y) + \sum_{m=1}^{\infty} \sum_{k=1}^{\infty} q_k(x, y) P(m, y, y) = \\ &= \sum_{n=1}^{\infty} q_n(x, y) \left(1 + \sum_{m=1}^{\infty} P(m, y, y)\right), \end{aligned}$$

но $\sum_{n=1}^{\infty} q_n(x, y) = P_x\{\tau^y < \infty\} \leq 1$, а $\sum_{m=1}^{\infty} P(m, y, y) < \infty$ на основании теоремы 1, так как по предположению состояние y невозвратно.

Следствие 1. Если y невозвратно, то из событий $\{\xi_n = y\}$ с вероятностью $P_x = 1$ происходит лишь конечное число.

Это вытекает из неравенства (6) и теоремы Бореля—Кантелли. Таким образом, для всякого невозвратного y конечна величина ζ^y — момент последнего попадания в y , определяемая как наименьшее $n \geq 0$ такое, что $\xi_k \neq y$ при $k > n$. Аналогично, если $G \subset X$ —

некоторое конечное множество невозвратных состояний, то конечна величина $\xi(0) = \max_{y \in G} \xi^y$ — момент последнего попадания во множество G .

Следствие 2. Если y невозвратно, то для всех $x \in X$

$$\lim_{n \rightarrow \infty} P(n, x, y) = 0.$$

Это соотношение вытекает из неравенства (6).

Рассмотрим множество возвратных состояний: $R \subset X$. Говорят, что состояние y достижимо из x , если $P_x\{\tau^y < \infty\} > 0$.

Теорема 3. Если состояние x возвратно и $P_x\{\tau^y < \infty\} > 0$, то $P_x\{\tau^y < \infty\} = 1$, $P_y\{\tau^x < \infty\} = 1$ и состояние y также возвратно.

Доказательство. Обозначим через $r(x, y)$ вероятность того, что система, находясь в начальный момент в состоянии x , выходит из него и попадает в состояние y раньше, чем в состояние x . Тогда $1 - r(x, y)$ есть вероятность того, что система возвратится в исходное состояние, не побывав в состоянии y .

$$r(x, y) = P_x\{\tau^y < \tau^x\}, \quad 1 - r(x, y) = P_x\{\tau^x < \tau^y\}$$

(считаем, что $y \neq x$, тогда $\tau^y \neq \tau^x$). Имеем

$$P_x\{\tau^y < \infty\} = P_x\{\tau^y < \tau^x\} + P_x\{\tau^x < \tau^y < \infty\}.$$

Вторая вероятность справа представима в виде

$$\begin{aligned} P_x\{\tau^x < \tau^y < \infty\} &= \sum_{n=1}^{\infty} P_x\{\tau^x = n, n < \tau^y < \infty\} = \\ &= \sum_{n=1}^{\infty} P_x\{\xi_1 \in (x, y), \dots, \xi_{n-1} \in (x, y), \xi_n = x, \bigcup_{k=n}^{\infty} \{\xi_{k+1} = y\}\} = \\ &= \sum_{n=1}^{\infty} P_x\{\xi_1 \in (x, y), \dots, \xi_{n-1} \in (x, y), \xi_n = x\} \times \\ &\quad \times P_x\{\bigcup_{k=n}^{\infty} \{\xi_{k+1} = y\} / \xi_n = x\}. \end{aligned}$$

Воспользовавшись тем, что цепь является однородной, можем записать

$$P_x\{\bigcup_{k=n}^{\infty} \{\xi_{k+1} = y\} / \xi_n = x\} = P_x\{\bigcup_{k=0}^{\infty} \{\xi_{k+1} = y\} / \xi_0 = x\} = P_x\{\tau^y < \infty\}.$$

Кроме того,

$$\sum_{n=1}^{\infty} P_x\{\xi_1 \in (x, y), \dots, \xi_{n-1} \in (x, y), \xi_n = x\} = P_x\{\tau^x < \tau^y\}.$$

Следовательно,

$$P_x\{\tau^y < \infty\} = r(x, y) + (1 - r(x, y)) P_x\{\tau^x < \infty\}.$$

Отсюда вытекает, что $P_x\{\tau^y < \infty\} = 1$, если только $r(x, y) > 0$. Покажем, что последнее неравенство выполняется. Имеем

$$\begin{aligned} P_x\{\tau^y \leq m\} &\leq r(x, y) + P_x\{\tau^x < \tau^y \leq m\} = \\ &= r(x, y) + \sum_{k=1}^{m-1} P_x\{\xi_1 \in (x, y), \dots, \xi_{k-1} \in (x, y), \xi_k = x\} \times \\ &\quad \times P\left\{\bigcup_{n=k}^{m-1} \{\xi_{n+1} = y\} / \xi_k = x\right\} = \\ &= r(x, y) + \sum_{k=1}^{m-1} P_x\{\xi_1 \in (x, y), \dots, \xi_{k-1} \in (x, y), \xi_k = x\} \times \\ &\quad \times P_x\left\{\bigcup_{n=0}^{m-k-1} \{\xi_{n+1} = y\} / \xi_k = x\right\} \leq r(x, y) + P_x\{\tau^y \leq m-1\} \end{aligned}$$

и

$$P_x\{\tau^y \leq 1\} \leq r(x, y).$$

Таким образом,

$$P_x\{\tau^y \leq m\} \leq mr(x, y).$$

Если $r(x, y) = 0$, то $P_x\{\tau^y \leq m\} = 0$ для всех m , тогда $P_x\{\tau^y < \infty\} = 0$. Поэтому при $P_x\{\tau^y < \infty\} > 0$ и $r(x, y) > 0$

$$P_x\{\tau^y < \infty\} = 1.$$

Далее

$$\begin{aligned} r(x, y) &= P_x\{\tau^y < \tau^x < \infty\} = \\ &= \sum_{n=1}^{\infty} P_x\{\xi_1 \in (x, y), \dots, \xi_{n-1} \in (x, y), \xi_n = y\} P\left\{\bigcup_{k=n}^{\infty} \{\xi_{k+1} = \right. \\ &\quad \left. = x\} / \xi_n = y\right\} = P\left\{\bigcup_{k=0}^{\infty} \{\xi_{k+1} = x\} / \xi_0 = y\right\} P_x\{\tau^x < \infty\} = \\ &= r(x, y) P_y\{\tau^x < \infty\}. \end{aligned}$$

Так как по доказанному $r(x, y) > 0$, то $P_y\{\tau^x < \infty\} = 1$. Из неравенства

$$\begin{aligned} 0 < r(x, y) &= \sum_{n=1}^{\infty} P_x\{\xi_1 \in (x, y), \dots, \xi_{n-1} \in (x, y), \xi_n = y\} \leq \\ &\leq \sum_{n=1}^{\infty} P_x\{\xi_n = y\} = \sum_{n=1}^{\infty} P(n, x, y) \end{aligned}$$

вытекает, что при некотором n_1 имеем $P(n_1, x, y) > 0$. На основании соотношения

$$1 = P_y\{\tau^x < \infty\} = P_y\left\{\bigcup_{n=1}^{\infty} \{\xi_n = x\}\right\} \leq \sum_{n=1}^{\infty} P(n, y, x)$$

найдется такое n_2 , что $P(n_2, y, x) > 0$. Из уравнения Колмогорова — Чепмена получаем неравенство

$$P(n_1 + n_2 + m, y, y) \geq P(n_2, y, x) P(m, x, x) P(n_1, x, y).$$

Таким образом,

$$\begin{aligned} \sum_{n=1}^{\infty} P(n, y, y) &\geq \sum_{m=1}^{\infty} P(n_1 + n_2 + m, y, y) \geq \\ &\geq P(n_2, y, x) P(n_1, x, y) \sum_{m=1}^{\infty} P(m, x, x). \end{aligned}$$

Из возвратности x вытекает расходимость ряда справа, значит, и ряд слева расходится и y также возвратно (мы дважды используем теорему 1).

Множество состояний $G \subset X$ называется замкнутым, если для всех $x \in G$

$$\sum_{y \in G} P(x, y) = 1.$$

Если множество G замкнуто, то для всех n

$$\sum_{y \in G} P(n, x, y) = 1.$$

Это соотношение доказывается по индукции с помощью уравнения Чепмена — Колмогорова

$$\begin{aligned} \sum_{y \in G} P(n, x, y) &= \sum_{y \in G} \sum_z P(x, z) P(n-1, z, y) = \\ &= \sum_{z \in G} \sum_{y \in G} P(x, z) P(n-1, z, y) \end{aligned}$$

(мы воспользовались тем, что $P(x, z) = 0$ при $z \notin G$).

Если система в начальный момент находится в данном замкнутом множестве G , то тогда и при всех n она будет находиться в G : раз попав в замкнутое множество, система никогда его не покинет. Из теоремы 3 вытекает, что множество всех возвратных состояний замкнуто.

Разобьем все возвратные состояния на классы, отнеся два состояния x и y к одному классу, если $P_x\{\tau^y < \infty\} = 1$ (тогда и $P_y\{\tau^x < \infty\} = 1$). Класс, содержащий состояние x , это множество тех y , для которых $P_x\{\tau^y < \infty\} > 0$ (а, значит, и $P_x\{\tau^y < \infty\} = 1$). Действительно, если и $P_x\{\tau^z < \infty\} > 0$, то $P_z\{\tau^x < \infty\} > 0$, поэтому $P(n, z, x) > 0$ при некотором n . А так как $P(m, x, y) > 0$ при некотором m , то и

$$P(n+m, z, y) \geq P(n, z, x) P(m, x, y) > 0,$$

т. е. $P_z\{\tau^y < \infty\} = 1$. Тогда и $P_z\{\tau^y < \infty\} = 1$, $P_y\{\tau^z < \infty\} = 1$. Следовательно, y и z принадлежат одному классу. Каждый из классов является замкнутым множеством.

Пусть K — некоторый класс возвратных состояний. Говорят, что состояние x имеет период d , если

$$\sum_n P_x\{\tau^x = nd\} = 1,$$

и d — наибольшее число, для которого это выполняется. Обозначим через N — множество тех натуральных чисел, для которых

$$P(n, x, x) > 0.$$

Тогда d есть наибольший общий делитель чисел из N . Действительно, разрешив (4) относительно $\sum_{n=1}^{\infty} P(n, x, x) \lambda^n$, имеем

$$\sum_{n=1}^{\infty} P(n, x, x) \lambda^n = \sum_{n=1}^{\infty} q_n \lambda^n / (1 - \sum_{n=1}^{\infty} q_n \lambda^n) = \sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} q_n \lambda^n \right)^m.$$

Возводя ряды в степень и приравнявая коэффициенты при λ^n , найдем

$$P(n, x, x) = q_n + \sum_r q_{n-r} q_r + \sum_{k_1+k_2+k_3=n} q_{k_1} q_{k_2} q_{k_3} + \dots$$

Обозначим через N_1 множество тех n , для которых $q_n > 0$, $N_2 = N_1 + N_1$ (алгебраическая сумма, $r \in N_2$, если $r = r_1 + r_2$, где $r_1, r_2 \in N_1$), $N_3 = N_2 + N_1$ и т. д. Тогда $N = \bigcup N_i$. Если N_1 имеет наибольший общий делитель d , то и N имеет тот же общий делитель, так как N_i при $i > 1$ все делятся на d . Из этого рассуждения вытекает, что каково бы ни было $r \in N$, d совпадает с наибольшим делителем чисел $n - r$, где $n \in N$. Действительно, все эти числа делятся на d и, кроме того, содержат все числа из N_1 , для которого d является наибольшим общим делителем.

Покажем, что для состояний x и y из одного класса их периоды d_x и d_y совпадают. Пусть $P(s, x, y) > 0$ и $P(t, x, y) > 0$, N_x — множество тех n , для которых $P(n, x, x) > 0$. Аналогично определено N_y . Если $m \in N_x$, то

$$P(s + t + m, y, y) \geq P(t, y, x) P(s, x, y) P(m, x, x) > 0,$$

и, следовательно, $s + t + m \in N_y$. Если $m_0 \in N_x$, то $(s + t + m) - (s + t + m_0) = m - m_0$ имеют наибольший общий делитель d_x и, кроме того, все делятся на d_y . Значит $d_y \leq d_x$. Меняя x и y местами, убеждаемся, что $d_x = d_y$. Общее значение периодов всех состояний класса называется периодом класса. Если $d = 1$, класс называется непериодическим. Два состояния x и y относятся к одному подклассу, если

$$\sum_n P(nd, x, y) > 0. \quad (7)$$

Заметим, что в данном случае $P(md + r, y, x) > 0$ для всех $0 < r < d$, $m \geq 0$, так как в противном случае было бы $P(nd +$

$+r, x, x) > 0$ для некоторого n , что невозможно. Таким образом, и $\sum_n P(nd, y, x) = \sum_n P(n, y, x) = +\infty > 0$. Как было замечено, $P(md + r, x, y) = 0$ для всех m .

Подкласс, содержащий x , состоит из всех y , для которых выполнено неравенство (7). Если t принадлежит указанному множеству, то из z в y и из y в z можно перейти за число шагов, кратное d (через состояние x). Отсюда y и z действительно принадлежат одному подклассу. Обозначим построенный подкласс через $K^{(0)}$. Пусть $K^{(i)}$ — множество тех состояний, в которые можно перейти из $K^{(0)}$ за i шагов. Очевидно, $K^{(d)} = K^{(0)}$, а при $i = 1, \dots, d-1$ $K^{(i)}$ попарно не пересекаются. Если $z \in K^{(i)} \cap K^{(j)}$, $0 < i < j < d$, то из z в состояние x можно попасть за $nd - i$ шагов (при некотором n), а из x в z — за $md + j$ шагов, т. е. из x в x можно вернуться за $(n+m)d + j - i$ шагов, что невозможно при $0 < j < i < d$. Если y и z принадлежат $K^{(i)}$, то из y в z можно перейти за число шагов, кратное d ; за $d - i$ шагов попадем в $K^{(0)}$, а затем за $nd + i$ шагов — из $K^{(0)}$ в z . Таким образом, периодический класс с периодом d разбивается на d подклассов.

Проиллюстрируем введенные понятия на цепи Маркова, порождаемой случайным блужданием (см. § 6, гл. III). Множество состояний — все целые числа. Вероятность перехода за 1 шаг $P(x, y) = q(y - x) = P(\eta_1 = y - x)$, η_1 — шаг блуждания.

Условие возвратности случайного блуждания исследовалось в § 6, гл. III. Пусть m — шаг блуждания, т. е. m — максимальное число, для которого ξ_1/m — целочисленная величина. Тогда состояния разлагаются на m классов состояний

$$K_0 = \{nm, n = 0, \pm 1, \dots\}, \dots, K_i = \{nm + i, n = 0, \pm 1, \dots\}, \\ i = 0, \dots, m-1.$$

Пусть теперь x таково, что $q(x) > 0$. Обозначим через m_1 — наибольший общий делитель чисел вида $y - x$, где y таково, что $q(y) > 0$. Очевидно, m_1 делится на m , $m_1 = dm$. Число d будет периодом каждого из классов. Поскольку в нашем случае $P(n, x, y)$ зависит лишь от $y - x$, то достаточно рассмотреть класс K_0 . Можно считать, что блуждание имеет один класс, если рассматривать блуждание с шагом ξ_1/m .

Покажем, что блуждающая частица может вернуться в исходное состояние лишь за число шагов, кратное d . Все y , для которых $q(y) > 0$, имеют вид $y = x + nd$, $n = 0, \pm 1, \dots$, x и d должны быть взаимно просты, так как по предположению наибольший общий делитель таких y есть 1. За i шагов система переходит в состояние вида

$$ix + nd.$$

Поскольку x и d взаимно просты, то это выражение может обращаться в нуль лишь при i , кратном d .

§ 3. Эргодическая теорема для однородных цепей Маркова

Исследуем поведение вероятности перехода $P(n, x, y)$ при $n \rightarrow \infty$. В случае невозвратности состояния y уже установлено, что $P(n, x, y) \rightarrow 0$. Поэтому интерес представляет лишь случай возвратных состояний. Предположим, что состояния образуют один класс возвратных состояний.

Рассмотрим последовательные времена возвращения в данное состояние x : ξ_1, ξ_2, \dots . Здесь ξ_1 — момент первого возвращения (он определен выше), $\xi_2 = n$, если $\xi_1 < n$, $\xi_k \neq x$ при $\xi_1 < k < n$, $\xi_n = x$. Если ξ_{m-1} уже определено, то положим $\xi_m = n$, если $\xi_{m-1} < n$, $\xi_k \neq x$ при $\xi_{m-1} < k < n$, $\xi_n = x$. Если при некотором m $\xi_m = +\infty$, т. е. $\xi_k \neq x$ для всех $k > \xi_{m-1}$, то будем считать, что и $\xi_n = +\infty$ при $n > m$.

Теорема 1. Если состояние x возвратно, то $P\{\xi_m < \infty\} = 1$, для всех m и случайные величины $\tau_1 = \xi_1$, $\tau_2 = \xi_2 - \xi_1$, \dots , $\tau_m = \xi_m - \xi_{m-1}$ независимы и одинаково распределены.

Доказательство. Найдем распределение τ_{m+1} при условии, что ξ_m — конечная величина. Заданы τ_1, \dots, ξ_m .

$$\begin{aligned} P_x\{\tau_{m+1} = k / \xi_1 = n_1, \xi_2 = n_2, \dots, \xi_m = n_m\} &= \\ = P_x\{\xi_1 \neq x, \dots, \xi_{n_1-1} \neq x, \xi_{n_1} = x, \dots, \xi_{n_{m-1}+1} \neq x, \dots \\ \dots, \xi_{n_m} = x, \xi_{n_m+1} \neq x, \dots, \xi_{n_m+k-1} \neq x, \xi_{n_m+k} = x\} &= \\ = P_x\{\xi_{n_m+1} \neq x, \dots, \xi_{n_m+k-1} \neq x, \xi_{n_m+k} = x / \xi_{n_m} = x\} \times \\ \times P_x\{\xi_1 \neq x, \dots, \xi_{n_1-1} \neq x, \xi_{n_1} = x, \dots, \xi_{n_{m-1}+1} \neq x, \dots, \times \\ \times \xi_{n_{m-1}} \neq x, \xi_{n_m} = x\} &= P_x\{\xi_1 = n_1, \dots, \xi_m = n_m\} P_x\{\xi_1 \neq x, \dots \\ \dots, \xi_{k-1} \neq x, \xi_k = x\} &= P_x\{\xi_1 = n_1, \dots, \xi_m = n_m\} P\{\tau_1 = k\}. \end{aligned}$$

Следовательно,

$$P\{\tau_{m+1} < \infty / \xi_m < \infty\} = P\{\tau_1 < \infty\} P\{\xi_m < \infty\} / P\{\xi_m < \infty\} = 1.$$

Таким образом, ξ_{m+1} конечно с вероятностью 1, если только ξ_m конечно с вероятностью 1. Поскольку $P\{\xi_1 < \infty\} = 1$, то по индукции заключаем, что $P_x\{\xi_m < \infty\} = 1$ для всех m . Из соотношения

$$\begin{aligned} P_x &= \{\tau_{m+1} = k, \xi_1 = n_1, \dots, \xi_m = n_m\} = \\ &= P\{\tau_1 = k\} \cdot P\{\xi_1 = n_1, \dots, \xi_m = n_m\} \end{aligned}$$

вытекает, что τ_{m+1} не зависит от ξ_1, \dots, ξ_m и распределение τ_{m+1} совпадает с распределением τ_1 .

Теорема 2. Пусть $q_x = 1/M_x \tau_1$ (при $M_x \tau_1 = \infty$ считаем $q_x = 0$). Тогда, если x имеет период d , то

$$\lim_{n \rightarrow \infty} P(nd, x, x) = dq_x.$$

Доказательство. Событие $\xi_n^x = x$ происходит тогда и только тогда, когда при некотором m

$$\sum_{k=1}^m \tau_k = n.$$

Таким образом,

$$\begin{aligned} P(nd, x, x) &= \sum_{m=1}^{\infty} P_x \left\{ \sum_{k=1}^m \tau_k = nd \right\} = \\ &= \sum_{m=1}^{\infty} P_x \left\{ \sum_{k=1}^m \frac{\tau_k}{d} = n \right\}. \end{aligned}$$

Так как распределение величин τ_1/d будет уже нерешетчатым, то на основании теоремы восстановления

$$\lim_{n \rightarrow \infty} \sum_{m=1}^{\infty} P_x \left\{ \sum_{k=1}^m \frac{\tau_k}{d} = n \right\} = \frac{1}{M_x \left(\frac{\tau_1}{d} \right)} = \frac{d}{M_x \tau_1}.$$

Теорема 3. Пусть $0 \leq r < d$ таково, что $P(md + r, y, x) > 0$ при некотором m . Тогда

$$\lim_{n \rightarrow \infty} P(nd + r, y, x) = dq_x.$$

Доказательство. При сделанных предположениях относительно y и x из y в x можно попасть лишь за число шагов вида $nd + r$. Если τ_x — момент первого попадания в x , то

$$\sum_{n=0}^{\infty} P_y \{ \tau^x = nd + r \} = 1.$$

Далее

$$P(nd + r, x, y) = \sum_{i=0}^{n-1} P_y \{ \tau^x = id + r \} P((n-i)d, x, x).$$

Так как $\sum P_y \{ \tau^x = id + r \}$ сходится, то ряд справа сходится равномерно относительно n . Переходя к пределу при $n \rightarrow \infty$ и используя теорему 2, получим требуемое.

Следствие 1.

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P(k, y, x) = q_x. \quad (1)$$

Действительно,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P(k, y, x) = \frac{1}{n} \sum_{m \leq \frac{n-r}{d}} P(md + r, x, y).$$

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P(k, y, x) &= \lim_{n \rightarrow \infty} \frac{\left[\frac{n}{d} \right]}{n} \cdot \frac{1}{\left[\frac{n}{d} \right]} \sum_{m=1}^{\left[\frac{n}{d} \right]} P(md + r, y, x) = \\ &= \frac{1}{d} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P(md + r, y, x) = \\ &= \frac{1}{d} \lim_{n \rightarrow \infty} P(md + r, y, x) = \frac{1}{d} dq_x = q_x.\end{aligned}$$

Следствие 2. $\sum q_x \leq 1$. Имеем для любого конечного множества $X_1 \subset X$:

$$\sum_{x \in X_1} q_x = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \sum_{x \in X_1} P(k, y, x) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \sum_{x \in X} P(k, y, x) = 1.$$

Состояние x называется положительным, если $M_x \tau_1 < \infty$. В противном случае оно называется нулевым.

Теорема 4. Если для некоторого состояния x $M_x \tau_1 < \infty$, то тогда и для всякого y из того же класса $M_y \tau < \infty$.

Доказательство. Заметим сначала, что в этом случае $M_x \tau^y < \infty$. Действительно,

$$\begin{aligned}M_x \tau^y \chi \{ \tau^y \leq m \} &\leq M_x \tau^y \chi \{ \tau^x > \tau^y \} + M_x \tau^y \chi \{ \tau^x < \tau^y \leq m \} \leq \\ &\leq M_x \tau^x + \sum_{n=1}^m n P \{ \tau^x < \tau^y = n \} \leq \\ &\leq M_x \tau^x + \sum_{k=1}^{\infty} \sum_{i=1}^m (k+i) P_x \{ \tau^x = k, \tau^y = k+i \} = M_x \tau^x + \\ &+ \sum_{k=1}^{\infty} \sum_{i=1}^m (k+i) P_x \{ \xi_1 \in (x; y), \dots, \xi_{k-1} \in (x; y), \xi_k = x, \xi_{k+1} \neq \\ &\neq y, \dots, \xi_{k+i-1} \neq y, \xi_{k+i} = y \} = M_x \tau^x + \sum_{k=1}^{\infty} \sum_{i=1}^m (k+i) P_x \{ \xi_1 \in (x; y), \dots \\ &\dots, \xi_{k-1} \in (x; y), \xi_k = x \} P_x \{ \tau^y = i \} \leq \\ &\leq 2M_x \tau^x + M_x \tau^y \chi \{ \tau^y \leq m \} P_x \{ \tau^y > \tau^x \}.\end{aligned}$$

Таким образом,

$$M_x \tau^y \chi \{ \tau^y \leq m \} \leq 2M_x \tau^x / (1 - P_x \{ \tau^y > \tau^x \})$$

(как было показано в теореме 3, § 2, знаменатель не обращается в 0). Переходя к пределу при $m \rightarrow \infty$, убеждаемся, что $M_x \tau^y < \infty$.

Покажем теперь, что $M_y \tau^x < \infty$. Имеем

$$\begin{aligned} \infty > M_x \tau^x &\geq M_x \tau^x \chi \{ \tau^x > \tau^y \} = \\ &= \sum_{k, i=1}^{\infty} P_x \{ \tau^y = k, \tau^x = k + i \} (k + i) = \\ &= \sum_{i=1}^{\infty} (k + i) P_x \{ \xi_1 \in (x; y), \dots, \xi_{k+1} \in (x; y), \xi_k = \\ &= y, \xi_{k+1} \neq x, \dots, \xi_{k+i-1} \neq x, \xi_{k+i} = x \} = \\ &= \sum_{k, i=1}^{\infty} (k + i) P_x \{ \xi_1 \in (x; y), \dots, \xi_{k-1} \in (x; y), \xi_k = y \} \times \\ &\times P \{ \xi_{k+1} \neq x, \dots, \xi_{k+i-1} \neq x, \xi_{k+i} = x / \xi_k = y \} \geq \\ &\geq \sum_{k=1}^{\infty} P_x \{ \xi_1 \in (x; y), \dots, \xi_{k-1} \in (x; y), \xi_k = y \} \times \\ &\times \sum_{i=1}^{\infty} i P_y \{ \xi_1 \neq x, \dots, \xi_{i-1} \neq x, \xi_i = x \} = P_x \{ \tau^y > \tau^x \} M_y \tau^x. \end{aligned}$$

Поскольку $P_x \{ \tau^y > \tau^x \} > 0$, то $M_y \tau^x < \infty$. Наконец,

$$\begin{aligned} M_y \tau^y \chi \{ \tau^y \leq m \} &\leq M_y \tau^y \chi \{ \tau^x > \tau^y \} + M_y \tau^y \chi \{ \tau^x < \tau^y \leq m \} \leq \\ &\leq M_y \tau^x + \sum_{i=1}^m \sum_{k=1}^{\infty} P_y \{ \tau^x = k, \tau^y = k + i \} (k + i) \leq \\ &\leq M_y \tau^x + \sum_{k=1}^{\infty} k P_y \{ \tau^x = k \} + P \{ \tau^x < \tau^y \} M_x \tau^y \leq \\ &\leq 2M_y \tau^x + M_x \tau^y \end{aligned}$$

(мы использовали те же выражения для $P_y \{ \tau^x = k, \tau^y = k + i \}$, что и при выводе предыдущих неравенств).

Если все состояния класса положительны, то и класс называется положительным, если все нулевые, то класс называется нулевым.

Теорема 5. Если все состояния системы образуют один положительный класс, то числа q_x удовлетворяют условиям: а) $q_x > 0$, б) $\sum_{x \in X} q_x = 1$, в) q_x — единственное решение системы уравнений

$$\sum_{x \in X} P(x, y) q_x = q_y, \quad (2)$$

удовлетворяющее условиям а) и б).

Доказательство. Условие а) очевидно. Из (1) вытекает неравенство

$$\sum_{x \in X} q_x \leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \sum_{x \in X} P(k, y, x) = 1. \quad (3)$$

На основании уравнения Колмогорова—Чепмена

$$P(k+1, z, y) = \sum_{x \in X} P(k, z, x) P(x, y)$$

и тогда

$$\frac{1}{n} \sum_{k=1}^n P(k+1, z, y) = \sum_{x \in X} P(x, y) \frac{1}{n} \sum_{k=1}^n P(k, z, x). \quad (4)$$

Так как на основании равенства (1)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P(k+1, z, y) = \lim_{n \rightarrow \infty} \left(\frac{n}{n+1} \frac{1}{n+1} \sum_{k=1}^{n+1} P(k, z, y) - \frac{P(z, y)}{n} \right) = q_y,$$

то, переходя в (4) к пределу при $n \rightarrow \infty$, получим

$$q_y \geq \sum_{x \in X} P(x, y) \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P(k, z, x) = \sum_{x \in X} P(x, y) q_x. \quad (5)$$

Поскольку

$$\sum_{y \in X} (q_y - \sum_{x \in X} P(x, y) q_x) = \sum_{y \in X} q_y - \sum_{y \in X} P(x, y) \sum_{x \in X} q_x = 0,$$

то в формуле (5) должно быть равенство. Значит, выполнено равенство (2). Умножая (2) на $P(y, z)$ и суммируя по y , находим

$$\sum_{y \in X} \sum_{x \in X} P(x, y) P(y, z) q_x = \sum_{y \in X} q_y P(y, z) = q_z.$$

Если слева просуммировать сначала по y , то получим

$$\sum_{x \in X} P(z, x, z) q_x = q_z.$$

Таким же образом убеждаемся, что для всех k

$$\sum_{x \in X} P(k, x, z) q_x = q_z. \quad (6)$$

Просуммировав равенства (6) по k от 1 до n и разделив на n , получим

$$q_z = \sum_{x \in X} q_x \cdot \frac{1}{n} \sum_{k=1}^n P(k, x, z). \quad (7)$$

Переходя к пределу при $n \rightarrow \infty$, найдем

$$q_z = \sum_{x \in X} q_x q_z = q_z \sum_{x \in X} q_x \quad (8)$$

(почленный переход к пределу возможен, так как $\sum q_x \leq 1$). Поскольку $q_z > 0$, то из равенства (8) вытекает б). Докажем, наконец, единственность решения (2). Пусть \bar{q}_x — некоторое решение (2). Аналогично тому, как было получено (7), находим

$$\bar{q}_z = \sum_{x \in X} \bar{q}_x \cdot \frac{1}{n} \sum_{k=1}^n P(k, x, z).$$

Переходя к пределу, получим

$$\bar{q}_z = q_z \sum_{x \in X} \bar{q}_x = q_z.$$

Глава V

МАРКОВСКИЕ ПРОЦЕССЫ СО СЧЕТНЫМ МНОЖЕСТВОМ СОСТОЯНИЙ

§ 1. Определение марковского процесса с непрерывным временем

Рассмотрим систему, которая может находиться в одном из состояний множества X (фазового пространства системы). Будем считать, что система переходит из одного состояния в другое случайным образом. Пусть $x(t)$ — состояние системы в момент t , считаем, что $x(t)$ — случайные величины для всех $t \geq 0$ со значениями во множестве X , определенные на одном и том же вероятностном пространстве. Совокупность случайных величин $x(t)$ называется случайным процессом в X . В дальнейшем мы ограничимся случаем, когда X не более чем счетно. Пусть $X = \{x_1, x_2, \dots\}$. Набор вероятностей

$$P\{x(t_2) = x_{i_2}, \dots, x(t_n) = x_{i_n}\}$$

образуют совместные распределения случайного процесса $x(t)$. Этот процесс называется марковским, если какова бы ни была возрастающая последовательность моментов времени $0 \leq t_1 < t_2 < \dots < t_n < \dots$, последовательность случайных величин $x(t_1), \dots, x(t_n), \dots$ есть цепь Маркова. Обозначим

$$P(t, x, s, y) = P\{x(s) = y | x(t) = x\} \quad (1)$$

вероятность перехода процесса. Из определения цепи Маркова и формулы (2), § 1, гл. IV, вытекает, что совместные распределения марковского процесса выражаются через вероятность перехода следующим образом: при $t_1 < t_2 < \dots < t_n$

$$\begin{aligned} & P\{x(t_1) = x_{i_1}, \dots, x(t_n) = x_{i_n}\} = \\ & = P\{x(t_1) = x_{i_1}\} P(t_1, x_{i_1}, t_2, x_{i_2}) \dots P(t_{n-1}, x_{i_{n-1}}, t_n, x_{i_n}). \end{aligned} \quad (2)$$

Если соотношение (2) выполняется для всех n , $0 \leq t_1 < \dots < t_n$ и $x_{t_n} \in X$, то процесс $x(t)$ будет марковским и $P(t, x, s, y)$ является его вероятностью перехода. Будем считать, что вероятность перехода определена при $0 \leq t \leq s$, $x, y \in X$ и удовлетворяет условиям: б) $P(t, x, s, y) \geq 0$ при $0 \leq t < s$, $x \in X$ является распределением вероятностей по y , т. е. а) $P(t, x, s, y) \geq 0$ и б) $\sum_y P(t, x, s, y) = 1$; 2) $P(t, x, s, y)$ удовлетворяет уравнению Колмогорова—Чепмена: при $t < \tau < s$, $x, y \in X$

$$P(t, x, s, y) = \sum_z P(t, x, \tau, z) P(\tau, z, s, y) \quad (3)$$

(эти условия связаны со свойствами вероятности перехода цепи Маркова, см. § 1, гл. IV). 3) $P(t, x, t, y) = 0$ при $x \neq y$, $P(t, x, t, x) = 1$. В дальнейшем рассматриваются лишь однородные марковские процессы, так называются процессы, у которых вероятность перехода $P(t, x, s, y)$ зависит лишь от $s - t$. Вероятность перехода для однородного процесса обозначим $P(t, x, y)$:

$$P(t, x, s, y) = P(s - t, x, y).$$

Уравнение Колмогорова—Чепмена для однородного процесса принимает вид

$$P(t + s, x, y) = \sum_z P(t, x, z) P(s, z, y). \quad (4)$$

Удобно считать, что множество состояний X совпадает с множеством всех натуральных чисел (если число состояний счетно) или числами $1, \dots, n$ (если число состояний конечно и равно n). Вместо $P(t, i, j)$ принято писать $P_{ij}(t)$ — это вероятность перехода из состояния i в состояние j за время t . Формулу (4) можно записать в матричной форме, если ввести матрицу вероятностей перехода

$$\Pi(t) = \| P_{ij}(t) \|$$

(i — номер строки, j — номер столбца, где стоит элемент $P_{ij}(t)$, порядок матрицы совпадает с числом состояний системы). Из равенства (4) получаем

$$\Pi(s + t) = \Pi(t) \Pi(s). \quad (5)$$

Марковский процесс называется стохастически непрерывным, если

$$\lim_{t \rightarrow 0} P_{ij}(t) = \delta_{ij} \quad (6)$$

($\delta_{ij} = 1$ при $i = j$, $\delta_{ij} = 0$ при $i \neq j$). Для стохастически непрерывного процесса вероятность перехода равномерно непрерывна по t . Действительно,

$$\begin{aligned} P_{ij}(t + h) - P_{ij}(t) &= \sum_k P_{ik}(h) P_{kj}(t) - P_{ij}(t) = \\ &= P_{ij}(t) (P_{ii}(h) - 1) + \sum_{k \neq i} P_{ik}(h) P_{jk}(t). \end{aligned}$$

Следовательно,

$$|P_{ij}(t+h) - P_{ij}(t)| \leq 1 - P_{ii}(h) + \sum_{k \neq i} P_{ik}(h) = 2(1 - P_{ii}(h))$$

и правая часть стремится к нулю при $h \rightarrow 0$ равномерно относительно t .

Процесс называется локально регулярным, если он удовлетворяет двум условиям:

1) в каждом состоянии он проводит некоторое время, прежде чем выйдет из него;

2) из каждого состояния процесс непосредственно переходит в некоторое другое.

Условие 2) требует пояснения. Если τ — момент выхода из начального состояния, т. е. $x(t) = x(0)$ при $t < \tau$ и существуют сколь угодно малые $\delta > 0$ такие, что $x(\tau + \delta) \neq x(0)$, то можно указать такое $\varepsilon > 0$, что на интервале $(\tau, \tau + \varepsilon)$ $x(t)$ постоянно (это и есть то состояние, в которое $x(t)$ перешло из начального); другими словами, существует $x(\tau + 0)$.

Условия 1) и 2) накладывают определенные ограничения на вероятности перехода.

Теорема 1. Для локально регулярного процесса существуют пределы

$$\lim_{h \downarrow 0} \frac{P_{ij}(h) - \delta_{ij}}{h} = a_{ij}, \quad (7)$$

при этом $\sum_i a_{ij} = 0$.

Доказательство. Пусть τ — момент выхода из начального состояния i . Тогда

$$P\{\tau > t\} = \lim_{h \rightarrow 0} P\left\{x\left(\frac{kh}{n}\right) = i, kh < nt\right\} = \lim_{h \rightarrow 0} [P_{ii}(h)]^{[t/n]}$$

(здесь $[t/n]$ — целая часть t/n). Этот предел существует и отличен от 0 при достаточно малых t . Так как $P_{ii}(h) \rightarrow 1$, то

$$\lim_{h \rightarrow 0} (P_{ii}(h))^{[t/h]} = \exp \left\{ \lim_{h \rightarrow 0} \frac{t}{h} (P_{ii}(h) - 1) \right\}$$

и предел под знаком экспоненты существует. Следовательно, существует предел (7) при $t = j$ и

$$P_i\{\tau > t\} = \exp\{ta_{ii}\} \quad (8)$$

(P_i — вероятность при условии, что в начальный момент процесс находился в состоянии i). Обозначим через A_k событие, заключающееся в том, что после выхода из начального состояния i про-

цесс непосредственно попадает в состояние k . Если происходит событие A_k , то для всех достаточно малых h происходит событие

$$\bigcup_{n=0}^{\infty} \{x(lh) = i, l \leq n\} \cap \{x((n+1)h) = k\}.$$

Поэтому

$$\begin{aligned} P_i(A_k) &= \lim_{h \rightarrow 0} \sum_{n=0}^{\infty} P_i\{\{x(lh) = i, l \leq n\} \cap \{x((n+1)h) = k\}\} = \\ &= \lim_{h \rightarrow 0} P_{ik}(h) \sum_{n=0}^{\infty} P_{ii}(h)^n. \end{aligned}$$

Если $a_{ii} \neq 0$, то $P_{ii}(h) < 1$ при достаточно малых h и

$$\lim_{h \rightarrow 0} \frac{P_{ik}(h)}{h} = \lim_{h \rightarrow 0} \frac{1 - P_{ii}(h)}{h} P_i(A_n) = -a_{ii} P_i(A_n).$$

Доказано существование предела (7) при $a_{ii} \neq 0$. Если же $a_{ii} = 0$, то

$$\overline{\lim}_{h \rightarrow 0} \frac{P_{ik}(h)}{h} \leq \lim_{h \rightarrow 0} \frac{1 - P_{ii}(h)}{h} = 0.$$

В обоих случаях справедлива формула

$$a_{ik} = -a_{ii} P_i\{A_k\}. \quad (9)$$

Если все $a_{ik} = 0$, то и $\sum a_{ij} = 0$. Если же $a_{ii} \neq 0$, то $\sum_{k \neq i} P_i(A_n) = 1$ и по формуле (9) получаем

$$\sum_{k \neq i} a_{ik} = -a_{ii} \sum_{k \neq i} P_i(A_k) = -a_{ii}.$$

Замечание. Из равенства (8) вытекает, что время пребывания в состоянии i имеет показательное распределение с параметром $-a_{ii}$. Если $a_{ii} = 0$, то состояние i является поглощающим: раз попав в это состояние, процесс уже больше никогда не покинет. При $a_{ii} \neq 0$ величина $\pi_{ik} = -a_{ik}/a_{ii}$ имеет следующий смысл: это вероятность того, что после выхода из состояния i процесс непосредственно попадает в состояние k .

§ 2. Уравнения Колмогорова

В этом параграфе для локально регулярных процессов будут получены две системы дифференциальных уравнений для вероятностей перехода.

Теорема 1. Если однородный марковский процесс с переходными вероятностями $P_{ij}(t)$ локально регулярен, то тогда выполнена первая система уравнений Колмогорова

$$\frac{d}{dt} P_{ii}(t) = \sum_k a_{ik} P_{kj}(t). \quad (1)$$

Доказательство. Имеем

$$\begin{aligned} & \frac{P_{ij}(t+h) - P_{ij}(t)}{h} - \sum_k a_{ik} P_{kj}(t) = \\ &= \frac{\sum_k P_{ik}(h) P_{kj}(t) - P_{ij}(t)}{h} - \sum_k a_{ik} P_{kj}(t) = \\ &= \sum_k \left(\frac{P_{kj}(h) - \delta_{ik}}{h} - a_{ik} \right) P_{kj}(t). \end{aligned}$$

Поэтому для $n > i$

$$\begin{aligned} & \left| \frac{P_{ij}(t+h) - P_{ij}(t)}{h} - \sum_n a_{ik} P_{kj}(t) \right| \leq \\ & \leq \sum_{k < n} \left| \frac{P_{ik}(h) - \delta_{ik}}{h} - a_{ik} \right| + \sum_{k > n} \frac{P_{ik}(h)}{h} + \sum_{k > n} a_{ik} = \\ &= \sum_{k < n} \left| \frac{P_{ik}(h) - \delta_{ik}}{h} - a_{ik} \right| + \sum_{k < n} \left(\frac{P_{ik}(h) - \delta_{ik}}{h} - a_{ik} \right) + \\ &+ 2 \sum_{k > n} a_{ik} \leq 2 \sum_{k < n} \left| \frac{P_{ik}(h) - \delta_{ik}}{h} - a_{ik} \right| + 2 \sum_{k > n} a_{ik}. \end{aligned}$$

Значит, равномерно относительно

$$\lim_{h \rightarrow 0} \left| \frac{P_{ij}(t+h) - P_{ij}(t)}{h} - \sum_k a_{ik} P_{kj}(t) \right| \leq 2 \sum_{k > n} a_{ik}.$$

Так как левая часть от n не зависит, то, переходя к пределу при $n \rightarrow \infty$, получим уравнения (1).

Уравнения (1) можно рассматривать при фиксированном j . Это система обыкновенных дифференциальных уравнений с постоянными коэффициентами, она решается при начальном условии $P_{ij}(0) = \delta_{ij}$.

Укажем одно простое достаточное условие существования и единственности решения системы уравнений (1).

Теорема 2. Пусть выполнено условие

$$\sup_i |a_{ii}| < \infty. \quad (2)$$

Тогда система уравнений (1) имеет единственное ограниченное решение (при указанных начальных условиях).

Доказательство. Если обозначить через $\Pi(t)$ матрицу, составленную из функций $P_{ij}(t)$, а через A — матрицу из элементов a_{ij} , то

$$\Pi(t) = \exp\{tA\} = \sum_{n=0}^{\infty} \frac{t^n A^n}{n!} \quad (3)$$

($A^0 = I$ — единичная матрица). Покажем, что ряд справа сходится. Действительно, если $a_{kj}^{(n)}$ — элементы матрицы A^n , то

$$\begin{aligned} |a_{kj}^{(n)}| &= \left| \sum_i a_{ki} a_{ij}^{(n-1)} \right| \leq \sup_{i,j} |a_{ij}^{(n-1)}| \sum_i a_{ki} \leq \\ &\leq 2c \sup_{i,j} |a_{ij}^{(n-1)}|, \end{aligned}$$

где $c = \sup_i |a_{ii}|$. Следовательно,

$$\sup_{k,j} |a_{kj}^{(n)}| < 2c \sup_{i,j} |a_{ij}^{(n-1)}| \leq (2c)^n.$$

Ряд $\sum_{n=0}^{\infty} \frac{t^n (2c)^n}{n!} = e^{2ct}$ сходится для всех $t > 0$ и $c > 0$. Легко убедиться, что

$$\Pi'(t) = \sum_{n=1}^{\infty} \frac{n t^{n-1} A^n}{n!} = A \sum_{n=0}^{\infty} \frac{t^n A^n}{n!} = A \Pi(t).$$

Расписывая это соотношение поэлементно, получим уравнения (1). Для доказательства единственности достаточно показать, что при нулевых начальных данных решение будет также нулевым. Записав (1) в проинтегрированном виде, получим

$$P_{ij}(t) = \int_0^t \sum_k a_{ik} P_{kj}(s) ds.$$

Отсюда

$$\sup_i |P_{ij}(t)| \leq \int_0^t 2c \sup_k |P_{kj}(s)| ds. \quad (4)$$

Если функция $\alpha(t)$ ограничена, неотрицательна и удовлетворяет неравенству

$$\alpha(t) \leq 2c \int_0^t \alpha(s) ds,$$

то она равна нулю. Действительно, из неравенства (5) вытекает

$$\begin{aligned} \alpha(t) &\leq 2c \int_0^t 2c \int_0^s \alpha(u) du \leq 4c^2 \int_0^t (t-u) \alpha(u) du \leq \\ &\leq 4c^2 \int_0^t (t-u) \int_0^u 2c \alpha(s) ds \leq 8c^3 \int_0^t \frac{(t-u)^2}{2} \alpha(s) ds \leq \\ &\leq (2c)^n \int_0^t \frac{(t-u)^{n-1}}{(n-1)!} \alpha(u) du. \end{aligned}$$

Переходя к пределу при $n \rightarrow \infty$, получим требуемое. Следовательно, из (4) вытекает, что $P_{ij}(t) = 0$.

Замечания 1. Построенное решение, определяемое матрицей $\Pi(t)$ и задаваемой формулой (3), определяет вероятности перехода некоторого процесса. Действительно, из соотношения (3) получаем, что

$$\Pi(t+s) = \Pi(t) \Pi(s),$$

откуда вытекает уравнение Колмогорова—Чепмена (см. § 1, (4), (5)). Далее

$$\sum_j P_{ij}(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \sum_j a_{ij}^{(n)}.$$

При $n=0$ $\sum_j \delta_{ij}^{(n)} = 1$. При $n=1$ $\sum_j a_{ij}^{(1)} = 0$. По индукции устанавливаем, что

$$\sum_j a_{ij}^{(n)} = \sum_k a_{ik}^{(n-1)} \sum_j a_{kj}^{(n-1)} = 0.$$

Поэтому $\sum_j P_{ij}(t) = 1$ для всех i и $t \geq 0$. Для доказательства того, что $P_{ij}(t)$ неотрицательны, воспользуемся другим представлением $\Pi(t)$:

$$\Pi(t) = \lim_{n \rightarrow \infty} \left(I + \frac{t}{n} A \right)^n. \quad (6)$$

То, что предел в (6) существует и совпадает с правой частью (3), доказывается так, как в математическом анализе доказывается формула

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n} \right)^n = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Если только $n > t \sup_i |a_{ii}|$, то матрица $I + \frac{t}{n} A = \|\delta_{ij} + \frac{t}{n} a_{ij}\|$ имеет неотрицательные элементы, поэтому и ее n -я степень и предел этой степени имеют неотрицательные элементы.

2. В условиях теоремы 2 матрицы A и $\Pi(t)$ коммутируют. Действительно, для матрицы A^n справедливо соотношение

$$\sum_j |a_{ij}^{(n)}| \leq \sum_k |a_{ik}| \sum_j |a_{kj}^{(n-1)}| \leq 2c \sup_k \sum_k |a_{ki}^{(n-1)}|.$$

Поэтому

$$\sup_i \sum_j |a_{ij}^{(n)}| \leq 2c \sup_k \sum_l |a_{kl}^{(n-1)}| \leq (2c)^n,$$

так, что

$$a_{ij}^{(n+1)} = \sum_k a_{ik}^{(n)} a_{kj}$$

(ряд справа абсолютно сходится). Отсюда вытекает

$$\left(\sum_{n=0}^{\infty} \frac{t^n A^n}{n!} \right) A = \sum_{n=0}^{\infty} \frac{t^n A^n A}{n!} = \sum_{n=0}^{\infty} \frac{t^n A A^n}{n!} = A \sum_{n=0}^{\infty} \frac{t^n A^n}{n!}.$$

Таким образом, в условиях теоремы 2

$$\Pi'(t) = \Pi(t) A.$$

Расписывая это равенство поэлементно, получаем следующую систему уравнений для вероятностей перехода:

$$P'_{ij}(t) = \sum_k P_{ik}(t) a_{kj}, \quad (7)$$

которая называется второй (или прямой) системой уравнений А. Н. Колмогорова. Эта система удобна тем, что дает возможность вычислять безусловное распределение процесса.

Пусть

$$p_i = P\{x(0) = i\} -$$

начальное распределение процесса. Умножая равенство (7) на p_i и суммируя, получим систему уравнений

$$\frac{d}{dt} P_j(t) = \sum_k P_k(t) a_{kj}, \quad (8)$$

которая называется так же. Система (8) решается при начальных условиях: $P_j(0) = p_j$. Если вероятности p_i таковы, что $\frac{d}{dt} P_i(t) = 0$, т. е. они не меняются со временем, то они называются стационарными. Из равенства (8) вытекает такая система для определения стационарных вероятностей:

$$\sum_k p_k a_{kj} = 0. \quad (9)$$

§ 3. Применение теории марковских процессов к задачам массового обслуживания

Рассмотрим систему обслуживания, содержащую m одинаковых обслуживающих приборов. В систему поступают объекты, нуждающиеся в обслуживании (в принятой терминологии «требования»). Если имеется свободный прибор, он начинает обслуживать требование, если же свободных приборов нет, то требование теряется, или они выстраиваются в очередь в ожидании обслуживания. Охарактеризуем состояние системы числом наличных требований (тех, которые обслуживаются или ожидают обслуживания). Предположим, что поступления требований не зависят от состояния системы, если τ_1, τ_2, \dots — времена поступления требований, то

$\tau_1, \tau_2, \dots, \tau_1, \tau_2, \dots$ — последовательности независимых одинаково распределенных величин, имеющих показательное распределение с параметром μ

$$P\{\tau_n - \tau_{n-1} > t\} = e^{-\mu t}.$$

Времена обслуживания требований не зависят ни от состояния системы, ни от времени поступления, ни от прибора. Если ξ — время обслуживания некоторого требования, то оно имеет показательное распределение с параметром λ .

$$P\{\xi > t\} = e^{-\lambda t}.$$

Рассмотрим три возможных случая поведения требований в очереди:
а) требования ждут обслуживания независимо от длины очереди,
б) можно указать такое $n > m$, что если в системе имеется n требований, то новое требование теряется (не поступает в систему),
в) это же происходит при $n = m$.

Покажем, что в каждом из случаев процесс, описывающий состояние системы в момент t , является однородным марковским. Пусть в момент t в системе было k требований. Если из них обслуживаются i требований, то время до окончания обслуживания каждого из этих требований не зависит от времени их поступления и имеет показательное распределение, такое же, как и величина ξ . Далее, время до поступления очередного требования также не зависит от того, как давно поступило предыдущее требование. Оба эти свойства вытекают из равенства

$$P\{\xi > t + h \mid \xi > t\} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h}$$

(для показательного распределенной величины ξ остаток $\xi - t$ при условии $\xi > t$ имеет то же распределение, что и ξ). Таким образом, независимо от того, как поступали требования в систему до момента t и как они обслуживались (это определяет состояние системы до момента t), ее дальнейшая эволюция такова, как если бы в начальный момент было k требований, и i из них в этот момент начали обслуживаться. Согласно сделанным предположениям, $i = k$ при $k \leq m$ и $i = m$ при $k > m$. Следовательно, дальнейшая эволюция системы зависит только от состояния системы k . Это и означает, что система описывается однородным марковским процессом.

Получим для марковского процесса, описывающего систему, уравнения Колмогорова для вероятностей перехода. Удобно объединить все три случая а), б), в), полагая в первом $n = \infty$, а в третьем $n = m$. Пусть система находится в состоянии k . Для нахождения a_{kk} воспользуемся тем, что момент выхода из состояния k имеет показательное распределение с параметром $-\alpha_{kk}$. Если i — число занятых приборов, ξ_1, \dots, ξ_i — длительности обслуживания требований на этих приборах, а τ — время до поступления

очередного требования, то момент выхода η из этого состояния будет

$$\min [\tau, \xi_1, \dots, \xi_i].$$

Поэтому

$$\begin{aligned} e^{i a_{kk}} &= P \{ \min [\tau, \xi_1, \dots, \xi_i] > t \} = P \{ \tau > t, \xi_1 > t, \dots, \xi_i > t \} = \\ &= P \{ \tau > t \} P^i \{ \xi_1 > t \} = e^{-\mu t} e^{-i \lambda t} = \exp \{ -(\mu + i \lambda) t \}. \end{aligned}$$

Таким образом,

$$a_{kk} = -(\mu + i \lambda) = \begin{cases} -(\mu + k \lambda), & k \leq m \\ -(\mu + m \lambda), & k > m. \end{cases}$$

Из состояния k система может перейти непосредственно лишь в два соседних состояния: если один из занятых приборов окончил обслуживание, то система переходит в состояние $k-1$, если же поступило очередное требование, система переходит в состояние $k+1$. Используем введенные обозначения. Переход из k в $k-1$ происходит тогда, когда

$$\min [\xi_1, \dots, \xi_i] < \tau,$$

в противном случае происходит переход $k \rightarrow k+1$. Сумма вероятностей этих событий равна 1. Вычислим вероятность второго события. Так как ξ_1, \dots, ξ_i независимы и не зависят от τ , то

$$\begin{aligned} P \{ \min [\xi_1, \dots, \xi_i] > \tau \} &= \int_0^\infty P \{ \tau \in ds \} P \{ \min [\xi_1, \dots, \xi_i] > \\ &> \tau / \tau = s \} = \int_0^\infty P \{ \min [\xi_1, \dots, \xi_i] > s \} \mu e^{-\mu s} ds = \\ &= \mu \int_0^\infty e^{-\mu s - i \lambda s} ds = \frac{\mu}{\mu + i \lambda}. \end{aligned}$$

Эта вероятность равна $\pi_{k, k+1} = -\frac{a_{k, k+1}}{a_{kk}}$. Значит,

$$a_{k, k+1} = \mu, \quad a_{k, k-1} = -a_{kk} - \mu = \begin{cases} k \lambda; & k \leq m \\ m \lambda; & k > m \end{cases}$$

при $|k-i| > 1$, $a_{ki} = 0$. Кроме того, $-a_{00} = a_{01} = \mu$, переход в -1 невозможен. Наконец, при $k = n$ невозможны переходы в $n+1$, и тогда

$$a_{nn} = -a_{n, n-1} = \mu + m \lambda.$$

Выпишем вторую систему уравнений Колмогорова для безусловных вероятностей

$$\frac{dP_0(t)}{dt} = -\mu P_0(t) + \lambda P_1(t),$$

$$\frac{dP_j(t)}{dt} = \begin{cases} -(\mu + j\lambda) P_j(t) + (j+1)\lambda P_{j+1}(t) + \mu P_{j-1}(t), & 0 < j < m, \\ -(\mu + m\lambda) P_j(t) + m\lambda P_{j+1}(t) + \mu P_{j-1}(t), & m \leq j < n, \end{cases}$$

$$\frac{dP_n(t)}{dt} = -(\mu + m\lambda) P_n(t) + \mu P_{n-1}(t).$$

Найдем стационарные вероятности. Рассмотрим отдельно каждый из случаев. В случае а) система уравнений принимает вид

$$-\mu p_0 + \lambda p_1 = 0$$

$$-(\mu + j\lambda) p_j + (j+1)\lambda p_{j+1} + \mu p_{j-1} = 0, \quad 0 < j < m,$$

$$-(\mu + m\lambda) p_j + m\lambda p_{j+1} + \mu p_{j-1} = 0, \quad m \leq j.$$

При $j \geq m$ p_j удовлетворяют разностному уравнению второго порядка с характеристическим многочленом

$$m\lambda v^2 - (\mu + m\lambda) v + \mu = 0,$$

$$(m\lambda v - \mu)(v - 1) = 0.$$

Общее решение этого уравнения

$$p_j = \alpha + \beta \left(\frac{\mu}{m\lambda} \right)^j. \quad (1)$$

Поскольку $\sum p_j < \infty$, то стационарное распределение существует лишь при $\mu / m\lambda < 1$ и тогда $p_j = \beta \left(\frac{\mu}{m\lambda} \right)^j$ при $j \geq m$. Из первого уравнения находим

$$p_1 = \frac{\mu}{\lambda} p_0.$$

Далее

$$2\lambda p_2 = -(\mu + \lambda) p_1 - \mu p_0 = \frac{\mu + \lambda}{\lambda} \mu p_0 -$$

$$- \mu p_0 = \frac{\mu^2}{\lambda} p_0, \quad p_2 = \frac{\mu^2}{2\lambda^2} p_0.$$

По индукции устанавливаем, что при $j \leq m$

$$p_j = \frac{1}{j!} \left(\frac{\mu}{\lambda} \right)^j p_0.$$

При $m = j$ находим $\frac{\beta}{m!} = \frac{1}{m!}$. Итак, окончательно,

$$p_j = \frac{1}{j!} \left(\frac{\mu}{\lambda} \right)^j p_0, \quad 0 \leq j \leq m,$$

$$p_j = \frac{m^m}{m!} \left(\frac{\mu}{m\lambda} \right)^j p_0, \quad j > m.$$

Вероятность p_0 определяется из равенства

$$p_0 \left(\sum_{j=1}^m \frac{1}{j!} \left(\frac{\mu}{\lambda} \right)^j + \frac{m^m}{m!} \sum_{j=m+1}^{\infty} \left(\frac{\mu}{m\lambda} \right)^j \right) = 1.$$

В случае б) система имеет вид

$$\begin{aligned} -\mu p_0 + \lambda p_1 &= 0, \\ -(\mu + j\lambda) p_j + (j+1)\lambda p_{j+1} + \mu p_{j-1} &= 0, \quad 0 < j < m, \\ -(\mu + m\lambda) p_j + m\lambda p_{j+1} + \mu p_{j-1} &= 0, \quad j \leq m < n, \\ -(\mu + m\lambda) p_n + \mu p_{n-1} &= 0. \end{aligned} \quad (2)$$

При $j \leq m$ аналогично предыдущему,

$$p_j = \frac{1}{j!} \left(\frac{\mu}{\lambda} \right)^j p_0, \quad (3)$$

а при $m \leq j \leq n$ справедлива формула (1). Последнее из уравнений (2) позволяет исключить постоянную α :

$$\begin{aligned} -(\mu + m\lambda) \left[\alpha + \beta \left(\frac{\mu}{m\lambda} \right)^n \right] + \mu \left[\alpha + \beta \left(\frac{\mu}{m\lambda} \right)^{n+1} \right] &= 0, \\ \alpha = \frac{\beta}{m\lambda} \left(\frac{\mu}{m\lambda} \right)^{n+1} \left(\mu - \frac{(\mu - m\lambda)\mu}{m\lambda} \right) &= -\beta \left(\frac{\mu}{m\lambda} \right)^{n+1}. \end{aligned}$$

Таким образом, при $j \geq m$

$$p_j = \beta \left[\left(\frac{\mu}{m\lambda} \right)^j - \left(\frac{\mu}{m\lambda} \right)^{n+1} \right].$$

Поскольку

$$p_m = \frac{1}{m!} \left(\frac{\mu}{\lambda} \right)^m p_0 = \beta \left(\frac{\mu}{m\lambda} \right)^m \left[1 - \left(\frac{\mu}{m\lambda} \right)^{n+1-m} \right],$$

то

$$\beta = p_0 \frac{m^m}{m! \left(1 - \frac{\mu}{m\lambda} \right)^{n+1-m}}.$$

Окончательная формула для p_j такова:

$$p_j = \begin{cases} \frac{1}{j!} \left(\frac{\mu}{\lambda} \right)^j p_0, & j \leq m, \\ \frac{m^m \left[\left(\frac{\mu}{m\lambda} \right)^j - \left(\frac{\mu}{m\lambda} \right)^{n+1} \right]}{m! \left[1 - \left(\frac{\mu}{m\lambda} \right)^{n+1-m} \right]} p_0, & m < j \leq n. \end{cases} \quad (4)$$

p_0 определяется из условия $\sum_{j=0}^n p_j = 1$. Формула (4) справедлива при $\mu \neq m\lambda$. Однако с помощью предельного перехода можно

определить p_j и в этом случае. Первая строка формулы не изменится, а вторая примет вид

$$p_j = \frac{m^m}{m!} \frac{n+1-j}{n+1-m} p_0.$$

В случае $n = m$ имеем

$$p_j = \frac{1}{j!} \left(\frac{\mu}{\lambda} \right)^j p_0, \text{ где } p_0 = \sum_{k=0}^m \frac{1}{k!} \left(\frac{\mu}{\lambda} \right)^k. \quad (5)$$

Глава VI

СЛУЧАЙНЫЕ ВЕКТОРЫ

§ 1. Распределение случайного вектора

Пусть R^d — d -мерное линейное пространство, в котором выбран определенный базис. Точки этого пространства будем обозначать буквами x, y, a, b и отождествлять с последовательностью координат в выбранном базисе. Например, будем писать $x = (x_1, x_2, \dots, x_d)$, $a = (a_1, a_2, \dots, a_d)$ и т. д.

При фиксированном базисе пространство R^d можно рассматривать как полуупорядоченное. Полагаем $a \leq b$ тогда и только тогда, когда $a_k \leq b_k$ при $k = 1, 2, \dots, d$ и $a < b$, если $a_k < b_k$ при $k = 1, 2, \dots, d$.

Множество $[a, b]$ точек в R^d

$$\begin{aligned} [a, b] &= \{x : a \leq x < b\} = \\ &= \{(x_1, \dots, x_d) : a_1 \leq x_1 < b_1, \dots, a_d \leq x_d < b_d\} \end{aligned}$$

называют d -интервалом. Аналогично определяются d -интервалы вида $[a, b]$ ($[a, b]$ — замыкание $[a, b)$) и (a, b) ((a, b) — множество внутренних точек $[a, b)$).

Пусть задано некоторое вероятностное пространство (Ω, \mathcal{F}, P) . Случайным вектором ξ со значениями в R^d называют последовательность $(\xi_1, \xi_2, \dots, \xi_d)$ действительных случайных величин ξ_k , $k = 1, \dots, d$, определенных на данном вероятностном пространстве. Величину ξ_k называют k -й координатой вектора ξ . Так как $\xi_k = f_k(\omega)$, где $f_k(\omega)$ — \mathcal{F} -измеримая действительная функция, то $\xi = f(\omega)$, где $f(\omega) = (f_1(\omega), \dots, f_d(\omega))$ — функция со значениями в R^d с \mathcal{F} -измеримыми координатами. Такую функцию $f(\omega)$ называют \mathcal{F} -измеримой. Она отображает пространство Ω на R^d . Для каждого множества B из R^d положим

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

Множество $\xi^{-1}(B)$ называют прообразом (в некоторых случаях полным прообразом) множества B . Таким образом, ξ^{-1} является некоторым отображением, определенным на всех подмножествах R^d .

В § 1, гл. II, было показано, что для каждого борелевского множества B из R^d $\xi^{-1}(B) \in \mathcal{F}$ (см. теорему 2). Поэтому для любого борелевского множества B определена вероятность события — «случайный вектор ξ попадает во множество B ». Положим

$$F(B) = \mathbf{P}(\xi \in B), \quad B \in \mathfrak{B}^d.$$

Функция множества $F(B)$ является мерой:

если $B_k, k = 1, \dots, n, \dots$ последовательность непересекающихся борелевских множеств, то

$$F\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} F(B_n),$$

причем

$$F(R^d) = 1.$$

Действительно, $\xi^{-1}(B_n)$ также является последовательностью непересекающихся множеств, $\xi^{-1}(B_n) \in \mathcal{F}$ и

$$\begin{aligned} F\left(\bigcup_{n=1}^{\infty} B_n\right) &= \mathbf{P}(\xi^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right)) = \mathbf{P}\left(\bigcup_{n=1}^{\infty} \xi^{-1}(B_n)\right) = \\ &= \sum_{n=1}^{\infty} \mathbf{P}(\xi^{-1}(B_n)) = \sum_{n=1}^{\infty} F(B_n), \\ F(R^d) &= \mathbf{P}(\xi^{-1}(R^d)) = \mathbf{P}(\Omega) = 1. \end{aligned}$$

Меру $F(B), B \in \mathfrak{B}^d$ называют распределением случайного вектора ξ . Она несет в себе полную информацию о случайном векторе ξ самом по себе (т. е. если не затрагиваются связи между вектором ξ и другими случайными векторами или событиями, не выражающимися однозначно через ξ).

Как же можно задать распределение случайного вектора?

В общем случае это можно сделать с помощью функции распределения случайного вектора.

Определение. Функцией распределения $F(x)$ случайного вектора ξ называют функцию

$$F(x) = \underset{\text{Def}}{\mathbf{P}}(\xi_1 < x_1, \xi_2 < x_2, \dots, \xi_d < x_d).$$

Приведем некоторые свойства функции распределения $F(x)$, доказательства которых очевидны или такие же, как в одномерном случае:

- 1) $0 \leq F(x) \leq 1$;
- 2) если $x_k \rightarrow -\infty$ при некотором k , то

$$\lim F(x) = 0;$$

3) если $x_1 \rightarrow +\infty, x_2 \rightarrow +\infty, \dots, x_d \rightarrow +\infty$, то

$$\lim F(x) = 1;$$

4) функция $F(x)$ непрерывна слева.

Действительную функцию $F(x)$, $x \in R^d$, называют непрерывной слева, если для произвольной последовательности точек x_n такой, что $x_n \leq x$, $n = 1, 2, \dots$, и $\lim x_n = x$, имеем $\lim F(x_n) = F(x)$.

Покажем, что, зная функцию распределения случайного вектора ξ , можно найти вероятность события $\{\xi \in [a, b]\}$. Введем конечно-разностный оператор Δ_h^i , действующий на функцию $g(x) = g(x_1, \dots, x_d)$ по формуле

$$\Delta_h^i g(x) = g(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_d) - g(x_1, \dots, x_i, \dots, x_d) \quad i = 1, \dots, d.$$

Если оператор Δ_h^1 действует на функцию одного аргумента, то вместо Δ_h^1 пишут Δ_h .

Заметим, что если A — некоторое событие, η — случайная величина, $f(x) = P(A \cap \{\eta < x\})$, то

$$P(A \cap \{x \leq \eta < x + h\}) = \Delta_h f(x). \quad (1)$$

Отсюда легко получить по индукции формулу

$$P(\{\xi \in [a, a + h]\}) = \Delta_{h_1}^1 \Delta_{h_2}^2 \dots \Delta_{h_d}^d F(a).$$

Например, при $d = 2$

$$P(\{\xi \in [a, b]\}) = \Delta_{b_1 - a_1}^1 [F(a_1, b_2) - F(a_1, a_2)] = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2).$$

Формула (1) показывает, что функция распределения $F(x)$ однозначно определяет меру $F(B)$ на d -интервалах. В теории меры доказывается, что для любого борелевского множества B

$$F(B) = \inf (\sum F[a_k, b_k]), \quad (2)$$

где \inf означает точную нижнюю грань по всевозможным последовательностям $[a_k, b_k]$, $k = 1, 2, \dots$, образующим покрытие множества B , ($B \subset \bigcup_k [a_k, b_k]$). Отсюда вытекает, что функция

распределения $F(x)$ однозначно определяет распределение $F(B)$ случайного вектора. Более того, имеет место следующая теорема:

Теорема 1. Пусть $F(x)$, $x \in R^d$ — произвольная функция, удовлетворяющая условиям

$$a) 0 \leq F(x) \leq 1, \quad \lim_{x_k \rightarrow -\infty} F(x) = 0, \quad k = 1, \dots, d,$$

$$\lim_{\min(x_1, \dots, x_d) \rightarrow \infty} F(x) = 1;$$

б) функция $F(x)$ непрерывна слева;

в) для любых $x, h \in R^d$

$$\Delta_{h_1}^1 \Delta_{h_2}^2 \dots \Delta_{h_d}^d F(x) \geq 0.$$

Тогда на \mathfrak{B}^d можно и притом единственным образом определить меру $F(B)$ такую, что

$$F[a, b) = \Delta_{h_1}^1 \Delta_{h_2}^2 \dots \Delta_{h_d}^d F(a), \quad h = b - a.$$

При этом $F(B)$ определяется формулой (2).

Произвольную функцию $F(x)$, удовлетворяющую условиям а), б), в) теоремы 1, называют функцией распределения, а меру $F(B)$ — распределением, соответствующим функции распределения $F(x)$.

При решении конкретных задач пользоваться функцией распределения случайных векторов неудобно. В этих задачах, как правило, встречаются распределения двух типов.

1. Дискретные распределения. Случайный вектор ξ принимает конечное или счетное множество значений $x^1, x^2, \dots, x^n, \dots$. Распределение этого вектора можно задать с помощью набора чисел $p_n = P\{\xi = x^n\}$. Для любого множества B в R^d

$$F(B) = P\{\xi \in B\} = \sum_{x^n \in B} p_n.$$

2. Абсолютно непрерывные распределения. В этом случае существует такая борелевская функция $f(x)$, что для любого $B \in \mathfrak{B}^d$

$$F(B) = \int_B f(x) dx.$$

Интеграл в правой части равенства обозначает интеграл Лебега (или Римана, если функция $f(x)$ интегрируема по Риману) функции d -независимых переменных. Более пространная форма записи

$$\int_B f(x) dx = \int_B \dots \int f(x_1, \dots, x_d) dx_1 \dots dx_d.$$

Функцию $f(x)$ называют плотностью распределения $F(\cdot)$, или плотностью распределения случайного вектора ξ . В необходимых случаях вместо $f(x)$ пишут $f_\xi(x)$, чтобы подчеркнуть, что $f_\xi(x)$ есть плотность распределения случайной величины ξ .

Пусть $\xi_s = (\xi_1, \dots, \xi_d)$ — случайная величина с функцией распределения $F(x) = F(x_1, \dots, x_d)$. Рассмотрим «укороченный» вектор $\xi_s = (\xi_1, \dots, \xi_s)$, $s < d$. Его функция распределения выражается через функцию распределения вектора ξ следующим образом:

$$\begin{aligned} F_{\eta}(x_1, \dots, x_s) &= P\{\xi_1 < x_1, \dots, \xi_s < x_s\} = \\ &= P\{\xi_1 < x_1, \dots, \xi_s < x_s, \xi_{s+1} < +\infty, \dots, \xi_d < +\infty\} = \\ &= F_\xi(x_1, \dots, x_s, +\infty, \dots, +\infty). \end{aligned}$$

Пусть B_1 — борелевское множество в R^s , а B — множество в R^d , состоящее из всех точек $x = (x_1, \dots, x_d)$, для которых точка $(x_1, \dots, x_s) \in B_1$. Множество B называют цилиндрическим множеством с основанием B_1 в R^s . Очевидно, события $\{\eta \in B_1\}$ и $\{\xi \in B\}$ совпадают. Если распределение вектора ξ обладает плотностью $f_\xi(x_1, \dots, x_s)$, то

$$F_\eta(B_1) = P\{\eta \in B_1\} = P\{\xi \in B\}.$$

Поэтому

$$F_\eta(B_1) = \int \dots \int_{B_1} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_s, x_{s+1}, \dots, \dots, x_d) dx_1 \dots dx_s dx_{s+1} \dots dx_d.$$

Положив

$$f_\eta(x_1, \dots, x_s) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_s, x_{s+1}, \dots, \dots, x_d) dx_{s+1} \dots dx_d, \quad (3)$$

получим

$$F_\eta(B_1) = \int_{B_1} \dots \int f_\eta(x_1, \dots, x_s) dx_1 \dots dx_s.$$

Таким образом, если вектор ξ обладает плотностью распределения, то укороченный вектор η также обладает плотностью распределения $f_\eta(x_1, \dots, x_s)$.

Существование плотности распределения случайного вектора позволяет вычислять математическое ожидание случайных величин вида $\eta = g(\xi) = g(\xi_1, \xi_2, \dots, \xi_d)$ с помощью обычных интегралов, а именно, имеет место формула

$$Mg(\xi) = \int_{R^d} g(x) f(x) dx \quad (4)$$

для любой борелевской функции $g(x)$, $x \in R^d$, для которой одна из частей равенства (4) имеет смысл.

Доказательство. Допустим, что функция $g(x)$ неотрицательна. Обозначим $B_{nk} = \left\{x: \frac{k}{2^n} \leq g(x) < \frac{k+1}{2^n}\right\}$ — борелевские множества. Имеем

$$\begin{aligned} Mg(\xi) &= \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} \frac{k}{2^n} P\left\{\frac{k}{2^n} \leq g(\xi) < \frac{k+1}{2^n}\right\} = \\ &= \lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} \frac{k}{2^n} F(B_{nk}) = \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} \int_{B_{nk}} \frac{k}{2^n} f(x) dx = \lim_{n \rightarrow \infty} \int_{R^d} g^{(n)}(x) f(x) dx, \end{aligned}$$

где $g^{(n)}(x) = \sum_{k=0}^{\infty} \frac{k}{2^n} I_{B_{nk}}$. Последовательность простых функций $g^{(n)}(x)$ монотонно не убывает и равномерно сходится к $g(x)$, $0 \leq g(x) - g^{(n)}(x) < \frac{1}{2^n}$. По теореме о монотонной сходимости

$$\lim_{n \rightarrow \infty} \int_{R^d} g^{(n)}(x) f(x) dx = \int_{R^d} g(x) f(x) dx.$$

Для неотрицательных функций формула (4) доказана. Представив $g(x)$ в виде $g(x) = g^+(x) - g^-(x)$, докажем формулу (1) в общем случае ($g^+(x) = g(x)$ при $g(x) \geq 0$ и $g^+(x) = 0$ при $g(x) \leq 0$, $g^-(x) = -g(x)$ при $g(x) \leq 0$ и $g^-(x) = 0$ при $g(x) \geq 0$).

§ 2. Независимые случайные векторы

Совместное распределение случайных векторов. Пусть ξ, η — два случайных вектора со значениями в R^d и R^s соответственно $\xi = (\xi_1, \dots, \xi_d)$, $\eta = (\eta_1, \dots, \eta_s)$. Функцию распределения $F(x, y)$ составного вектора $(\xi, \eta) = (\xi_1, \dots, \xi_d, \eta_1, \dots, \eta_s)$, $F(x, y) = F(x_1, \dots, x_d, y_1, \dots, y_s) = P(\xi < x, \eta < y)$, $x = (x_1, \dots, x_d)$, $y = (y_1, \dots, y_s)$, называют совместной функцией распределения случайных векторов ξ и η .

Случайные векторы ξ и η называют независимыми, если для произвольных борелевских множеств B' и B'' , $B' \subset R^d$, $B'' \subset R^s$, события $\{\xi \in B'\}$ и $\{\eta \in B''\}$ независимы, то есть

$$P(\{\xi \in B'\} \cap \{\eta \in B''\}) = P\{\xi \in B'\} P\{\eta \in B''\}.$$

В частности, положив $B' = \{a : a < x\}$, $B'' = \{b : b < y\}$ ($a \in R^d$, $b \in R^s$), из условия независимости (5) получим следующее равенство

$$F(x, y) = F_\xi(x) F_\eta(y), \quad (1)$$

где $F_\xi(x)$ ($F_\eta(y)$) — функция распределения случайного вектора ξ (η). Итак, если случайные векторы ξ и η независимы, их совместная функция распределения равна произведению функций распределения векторов ξ и η . Имеет место и обратное утверждение:

если совместная функция распределения случайных векторов ξ и η удовлетворяет равенству (1), то векторы ξ и η независимы.

Действительно, пусть $F_\eta(y) > 0$. Тогда распределение в R^d $\frac{P\{\xi \in B, \eta < y\}}{F_\eta(y)}$ и $F_\xi(B)$ имеют одинаковые функции распределения. Поэтому по теореме 1, § 1, они совпадают на \mathfrak{B}^d , то есть $P\{\xi \in B, \eta < y\} = F_\xi(B) F_\eta(y)$, $\forall B \in \mathfrak{B}^d$. Это равенство, очевидно, сохраняется и в том случае, когда $F_\eta(y) = 0$. Пусть теперь $F_\xi(B) > 0$. Тогда распределения в R^s $\frac{P\{\xi \in B, \eta \in A\}}{F_\xi(B)}$ и $F_\eta(A)$

имеют одинаковую функцию распределения на \mathfrak{B}^s . Следовательно, согласно упомянутой теореме, они совпадают на \mathfrak{B}^s . Итак, при $F_\xi(B) > 0$

$$P\{\xi \in B, \eta \in A\} = F_\xi(B) F_\eta(A).$$

Очевидно, это равенство сохраняется для всех $A \in \mathfrak{B}^s$ и $B \in \mathfrak{B}^d$.

Предположим, что случайные векторы ξ и η независимы и обладают плотностями распределения $f(x)$ и $g(y)$, $x \in R^d$, $y \in R^s$ соответственно. Тогда

$$P\{\xi \in A, \eta \in B\} = P\{\xi \in A\} P\{\eta \in B\} = \int_A f(x) dx \int_B g(y) dy.$$

Произведение интегралов, стоящих в правой части равенства, можно рассматривать как один интеграл в пространстве $R^d \times R^s$ (теорема Фубини), так что

$$P\{\xi \in A, \eta \in B\} = \iint_{A \times B} f(x) g(y) dx dy.$$

Из того что два распределения в R^{d+s} , а именно $F_1(C) = P\{(\xi, \eta) \in C\}$ и $F_2(C) = \iint_{\{z \in C, z=(x, y)\}} f(x) g(y) dx dy$, совпадают на множествах вида $C = A \times B$, $A \in R^d$, $B \in R^s$, вытекает, что они совпадают на всех борелевских множествах в R^{d+s} . Итак,

$$P\{(\xi, \eta) \in C\} = \iint_{\{z \in C, z=(x, y)\}} f(x) g(y) dx dy, \quad \forall C \in \mathfrak{B}^{d+s}.$$

Мы получим следующий результат:

Теорема 1. Если независимые случайные векторы ξ и η обладают плотностями распределения, то их совместное распределение также обладает плотностью распределения $f_{(\xi, \eta)}(x, y)$, $x \in R^d$, $y \in R^s$ и $f_{(\xi, \eta)}(x, y) = f(x) g(y)$.

Пусть $h(x)$, $x \in R^d$ — произвольная действительная борелевская функция. Положим $\zeta = h(\xi)$. Тогда ζ — случайная величина. Действительно, пусть $B = \{x: h(x) < a\}$. Множество B — борелевское. Так как $\{\zeta < a\} = \{h(\xi) < a\} = \{\xi \in B\}$, то событие $\{\zeta < a\}$ — \mathcal{F} -измеримо при любом действительном a , то есть ζ — случайная величина.

Пусть теперь $f(x) g(y)$ — борелевская функция, $x \in R^d$ ($y \in R^s$), со значениями в R^d (R^s). Положим $\xi' = f(\xi)$ ($\eta' = g(\eta)$). Из предыдущего следует, что ξ' (η') — случайный вектор.

Теорема 2. Если векторы ξ и η независимы, то $f(\xi)$ и $g(\eta)$ также независимы.

Доказательство. Пусть B_1 (A_1) — произвольное борелевское множество в R^d (R^s). Тогда

$$\begin{aligned} \{\xi' \in B_1, \eta' \in A_1\} &= \{f(\xi) \in B_1, g(\eta) \in A_1\} = \\ &= \{\xi \in f^{-1}(B_1), \eta \in g^{-1}(A_1)\}, \end{aligned}$$

где $f^{-1}(B_1)$, $g^{-1}(A_1)$ — борелевские множества. Поэтому

$$\begin{aligned} \mathbf{P}\{\xi' \in B_1, \eta' \in A_1\} &= \mathbf{P}\{\xi \in f^{-1}(B_1), \eta \in g^{-1}(A_1)\} = \\ &= \mathbf{P}\{\xi \in f^{-1}(B_1)\} \mathbf{P}\{\eta \in g^{-1}(A_1)\} = \mathbf{P}\{\xi' \in B_1\} \mathbf{P}\{\eta' \in A_1\}. \end{aligned}$$

Пусть ξ и η — независимые векторы с распределениями $F_\xi(\cdot)$ и $F_\eta(\cdot)$ соответственно, $f(x, y)$ ($x \in R^d$, $y \in R^s$) — действительная борелевская функция в R^{d+s} . Выражение для $\mathbf{M}f(\xi, \eta)$ можно записать в виде

$$\mathbf{M}f(\xi, \eta) = \iint_{R^d \times R^s} f(x, y) F_\xi(dx) F_\eta(dy), \quad (2)$$

причем, согласно теореме Фубини, интеграл по произведению пространств $R^d \times R^s$ может быть сведен к повторному интегрированию

$$\mathbf{M}f(\xi, \eta) = \int_{R^d} \left(\int_{R^s} f(x, y) F_\eta(dy) \right) F_\xi(dx).$$

Правую часть полученной формулы следует понимать так: сначала вычисляется математическое ожидание случайной величины $f(x, y)$ (x — фиксировано), затем полученный результат рассматривают как функцию от x , подставляют ξ вместо x и снова вычисляют математическое ожидание. Таким образом,

$$\mathbf{M}h(\xi, \eta) = \mathbf{M}[(\mathbf{M}h(x, \eta))|_{x=\xi}]. \quad (3)$$

Функция распределения суммы двух независимых случайных векторов. Пусть ξ и η — два независимых вектора размерности d с функциями распределения $F_1(x)$ и $F_2(x)$ и совместной функцией распределения $F(x, y) = F_1(x) F_2(y)$. Найдем функцию распределения $F(z)$ суммы $\xi + \eta$. Имеем

$$F(z) = \mathbf{P}\{\xi + \eta < z\} = \mathbf{P}\{(\xi, \eta) \in B_z\},$$

где B_z — множество точек $z = (x_1, \dots, x_d, y_1, \dots, y_d)$ пространства R^{2d} , для которых $x_1 + y_1 < z_1, x_2 + y_2 < z_2, \dots, x_d + y_d < z_d$. Пусть $I_z(u, v)$, $v, u \in R_d$, v — индикатор множества B_z . Имеем, воспользовавшись формулой (3),

$$F(z) = \mathbf{M}I_z(\xi, \eta) = \mathbf{M}[(\mathbf{M}I_z(u, \eta))|_{u=\xi}] = \mathbf{M}F_2(z - u)|_{u=\xi},$$

или

$$F(z) = \mathbf{M}F_2(z - \xi) = \int_{R^d} F_2(z - x) F_1(dx). \quad (4)$$

Меняя местами случайные векторы ξ и η , получим

$$F(z) = \mathbf{M}F_1(z - \eta) = \int_{R^d} F_1(z - y) F_2(dy). \quad (5)$$

Операцию, ставящую в соответствие двум функциям распределения третью по формуле (4), называют сверткой функций распределения F_1 и F_2 и пишут

$$F = F_1 * F_2.$$

Формулы (4) и (5) показывают, что свертка распределений является коммутативной операцией

$$F_1 * F_2 = F_2 * F_1.$$

Если одна из функций $F_i(x)$ имеет плотность распределения, то $F(z)$ также имеет плотность. Действительно, пусть $F_2(x)$ обладает плотностью. Тогда

$$\begin{aligned} F(z) &= \int_{\mathbb{R}^d} \left(\int_{\{y: y < z-x\}} f_2(y) dy \right) F_1(dx) = \\ &= \int_{\mathbb{R}^d} \left(\int_{\{u: u < z\}} f_2(u-x) du \right) F_1(dx) = \int_{\{u: u < z\}} \left(\int_{\mathbb{R}^d} f_2(u-x) F_1(dx) \right) du, \end{aligned}$$

так что $F(z)$ обладает плотностью распределения $f(z)$,

$$f(z) = \int_{\mathbb{R}^d} f_2(z-x) F_1(dx).$$

Если сверх того вектор ξ обладает плотностью распределения $f_1(x)$, то

$$f(z) = \int_{\mathbb{R}^d} f_2(z-x) f_1(x) dx. \quad (6)$$

Бинарное отношение (6), ставящее в соответствие двум функциям f_1 и f_2 третью функцию f по формуле (6), называют сверткой функций f_1 и f_2 и пишут

$$f = f_1 * f_2.$$

Оно имеет смысл для любых интегрируемых по Лебегу в \mathbb{R}^d функций f_1 и f_2 , причем функция f также интегрируема в \mathbb{R}^d . Действительно,

$$\begin{aligned} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f_1(z-x)| |f_2(x)| dx \right) dz &= \int_{\mathbb{R}^d} |f_2(x)| dx \int_{\mathbb{R}^d} |f_1(z-x)| dz = \\ &= \int_{\mathbb{R}^d} |f_2(x)| dx \int_{\mathbb{R}^d} |f_1(y)| dy < \infty. \end{aligned}$$

Следовательно, функция $f = f_1 * f_2$ интегрируема и, в частности, почти для всех z конечна. Как и свертка распределений, свертка функций также коммутативна

$$f_1 * f_2 = f_2 * f_1.$$

Обратим внимание на то, что операции свертки распределений и свертки функций определяются различным образом, но обозначаются одинаково. Из контекста должно быть ясно, о какой свертке в данном месте идет речь.

Функция распределения частного двух независимых случайных величин. Пусть ξ и η независимые случайные величины с функциями распределения $F_1(x)$ и $F_2(x)$ соответственно, $P(\eta = 0) = 0$. Найдем функцию распределения случайной величины $\gamma = \frac{\xi}{\eta}$.

Рассмотрим множество

$$B_z = \left\{ (x, y) : \frac{x}{y} < z \right\} = \{ (x, y) : x < yz, y > 0 \} \cup \{ (x, y) : x > yz, y < 0 \}.$$

Пусть $I_z(x, y)$ — индикатор множества B_z . Имеем, используя формулу (3),

$$\begin{aligned} F_1(z) &= P\left\{ \frac{\xi}{\eta} < z \right\} = M I_z(\xi, \eta) = M \{ M I_z(\xi, y) / y = \eta \} = \\ &= \int_0^{\infty} F_1(yz) dF_2(y) - \int_{-\infty}^0 [1 - F_1(yz)] dF_2(y). \end{aligned} \quad (8)$$

Если случайные величины ξ и η имеют плотности распределения $f_1(x)$ и $f_2(x)$, то случайная величина γ имеет плотность распределения $f_1(z)$, причем

$$f_1(z) = \int_{-\infty}^{\infty} |y| f_1(yz) f_2(y) dy. \quad (9)$$

Это следует из соотношения (8).

Распределение Стьюдента. Предположим, что случайные величины ξ, ξ_1, \dots, ξ_n независимы и имеют нормальное распределение $N(0, 1)$. *Распределением Стьюдента** (или *t-распределением*) с n степенями свободы называется распределение случайной величины

$$t = \frac{\xi}{\sqrt{\frac{1}{n} \sum_{k=1}^n \xi_k^2}}. \quad (10)$$

Вычислим плотность распределения случайной величины t . Пусть

$$\eta = \sqrt{\frac{1}{n} \sum_{k=1}^n \xi_k^2}.$$

* Стьюдент — псевдоним английского статистика Госсета, впервые нашедшего это распределение.

Из результатов § 8 (гл. III) следует, что плотность распределения случайной величины η равна

$$f_{\eta}(x) = \begin{cases} \frac{\frac{n}{2}}{2^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2}\right)} x^{n-1} e^{-\frac{nx^2}{2}}, & x > 0, \\ 0, & x < 0. \end{cases}$$

Используя формулу (9), получаем, что плотность распределения случайной величины t равна

$$f_t(z) = \frac{\frac{n}{2}}{\sqrt{n} 2^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2}\right)} \int_0^{\infty} y^n e^{-\frac{y^2}{2}(z^2+n)} dy.$$

В последнем интеграле перейдем к новой переменной интегрирования

$$s = \frac{y^2}{2} (z^2 + n).$$

Тогда получим

$$f_t(z) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{z^2}{n}\right)^{\frac{n+1}{2}}}.$$

Распределение Стьюдента играет важную роль в математической статистике. Заметим, что при $n=1$ распределение Стьюдента совпадает с распределением Коши.

F-распределение (распределение Фишера—Снедекора). Пусть $\xi_1, \dots, \xi_m, \eta_1, \dots, \eta_n$ — независимые нормально распределенные $N(0,1)$ случайные величины. *F-распределением (распределением Фишера—Снедекора с (m, n) степенями свободы называется распределение случайной величины*

$$F = \frac{\frac{1}{m} \sum_{k=1}^m \xi_k^2}{\frac{1}{n} \sum_{k=1}^n \eta_k^2}.$$

Из результатов § 8 гл. III следует, что плотность распределения случайной величины

$$\frac{1}{m} \sum_{k=1}^m \xi_k^2$$

равна при $x > 0$

$$\frac{\frac{m}{2}}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} x^{\frac{m}{2}-1} e^{-\frac{1}{2}mx}.$$

Поэтому используя формулу (9), получим после элементарных преобразований, что плотность распределения случайной величины y равна

$$f_Y(x) = \begin{cases} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} x^{\frac{m}{2}-1} \left(1 + \frac{m}{n}x\right)^{-\frac{m+n}{2}}, & x > 0, \\ 0, & x < 0. \end{cases}$$

§ 3. Условные распределения

В настоящем параграфе даны определения и выведены формулы, связанные с условными вероятностями вида $P\{\xi \in B | \eta = x\}$, где ξ и η — случайные векторы. Если $P(\eta = x) > 0$, то применимо определение условной вероятности события

$$P(\xi \in B | \eta = x) = \frac{P(\{\xi \in B\} \cap \{\eta = x\})}{P(\eta = x)}.$$

Эта формула теряет смысл при $P(\eta = x) = 0$. Общее определение условных вероятностей относительно классов событий, имеющих вероятность нуль, использует аппарат абстрактной теории меры и здесь приводиться не будет. Предположим, что векторы ξ и η (со значениями в R^d и R^m) обладают совместной плотностью распределения $f(x, y)$, $x \in R^d$, $y \in R^m$. Пусть для некоторого куба K_ε , $K_\varepsilon \subset R^m$ с центром в точке y и длиной ребра 2ε имеем $P(\eta \in K_\varepsilon) > 0$ для всех $\varepsilon \in (0, \varepsilon_0)$. Тогда

$$P(\xi \in B | \eta \in K_\varepsilon) = \frac{\int_{K_\varepsilon} \int_B f(x, y) dx dy}{\int_{K_\varepsilon} \int_{R^d} f(x, y) dx dy}.$$

Пусть $\varepsilon \rightarrow 0$. При этом почти для всех y (относительно лебеговой меры в R^m) предел в правой части равенства существует и равен:

$$\frac{\int_B f(x, y) dx}{\int_{R^d} f(x, y) dx} = \frac{\int_B f(x, y) dx}{f_Y(y)}, \quad (1)$$

где $f_{\eta}(y)$ — плотность распределения случайного вектора η . Естественно принять полученное выражение в качестве определения условной вероятности.

Определение. Условной вероятностью $P\{\xi \in B | \eta = y\}$ события $\xi \in B$ при гипотезе $\eta = y$ называют величину (1),

$$P(\xi \in B | \eta = y) \stackrel{\text{Def}}{=} \frac{\int_B f(x, y) dx}{f_{\eta}(y)}, \quad (2)$$

где

$$f_{\eta}(y) = \int_{R^d} f(x, y) dx.$$

Выражение в правой части равенства (2) теряет смысл, если $f_{\eta}(y) = 0$. Однако вероятность того, что вектор η попадает во множество $N_{\eta} = \{y : f_{\eta}(y) = 0\}$, равна нулю,

$$P(\eta \in N_{\eta}) = \int_{N_{\eta}} f_{\eta}(y) dy = 0.$$

Это обстоятельство позволяет игнорировать неопределенность формулы (2). Условимся считать в дальнейшем $P(\xi \in B | \eta = y) = 0$, если $f_{\eta}(y) = 0$.

Формулу (2) можно интерпретировать следующим образом: если пара случайных векторов (ξ, η) имеет плотность распределения $f(x, y)$, то существует условная плотность $f_{\xi|\eta}(x|y)$ распределения вектора ξ при гипотезе $\eta = y$, и она определяется формулой

$$f_{\xi|\eta}(x|y) = \frac{f(x, y)}{f_{\eta}(y)}. \quad (3)$$

Из формулы (3) следует, что

$$f(x, y) = f_{\xi|\eta}(x|y) f_{\eta}(y) = f_{\eta|\xi}(y|x) f_{\xi}(x).$$

Поэтому

$$f_{\xi|\eta}(x, y) = \frac{f_{\eta|\xi}(y|x) f_{\xi}(x)}{f_{\eta}(y)} = \frac{f_{\eta|\xi}(y|x) f_{\xi}(x)}{\int_{R^d} f_{\eta|\xi}(y|x) f_{\xi}(x) dx} \quad (4)$$

при $f_{\eta}(y) > 0$. Если $f_{\eta}(y) = 0$, то в соответствии с предыдущим соглашением, $f_{\xi|\eta}(x|y) = 0$. Формулу (4) называют формулой Байеса для апостериорных плотностей распределения.

Условная плотность $f_{\xi|\eta}(x|y)$ является плотностью некоторого распределения в R^d :

$$F_{\xi|\eta}(A|\eta) = \int_A f_{\xi|\eta}(x|y) dx; \quad \int_{R^d} f_{\xi|\eta}(x|y) dx = 1 \quad \text{при } y \notin N.$$

По отношению к этому распределению можно ввести понятия условного математического ожидания, условной дисперсии и др., причем общие свойства математических ожиданий и дисперсий, установленные ранее, имеют место и в случае условных распределений.

Определение. Условным математическим ожиданием $M(\xi|\eta=y)$ случайной величины ξ при гипотезе $\eta=y$ называют величину

$$M(\xi|\eta=y) \stackrel{\text{Def}}{=} \int_{-\infty}^{\infty} x f_{\xi|\eta}(x|y) dx,$$

а условной дисперсией

$$D(\xi|\eta=y) \stackrel{\text{Def}}{=} M((\xi - M(\xi|\eta=y))^2|\eta=y).$$

Если $M|\xi| < \infty$, то условное математическое ожидание $M(\xi|\eta=y)$ определено для всех $y \in N_1$, где N_1 — такое множество, что $P(\eta \in N_1) = 0$. Действительно,

$$M(\xi|\eta=y) = \frac{\int_{-\infty}^{\infty} x f(x, y) dx}{f_{\eta}(y)},$$

причем

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x| f(x, y) dx dy = \int_{-\infty}^{\infty} |x| f_{\xi}(x) dx < \infty,$$

так что

$$\left| \int_{-\infty}^{\infty} x f(x, y) dx \right| < \infty \text{ почти для всех } y.$$

Условные математические ожидания как случайные величины.

Следующий важный шаг в теории условных математических ожиданий и условных вероятностей связан с изменением точки зрения на природу этих величин, а именно, вместо функций $f_{\xi|\eta}(x|y)$ и $M(\xi|\eta=y)$ аргумента y рассматривают случайные величины

$$f_{\xi}(x|\eta) \stackrel{\text{Def}}{=} f_{\xi|\eta}(x|y)|_{y=\eta}, \quad M(\xi|\eta) \stackrel{\text{Def}}{=} M(\xi|\eta=y)|_{y=\eta},$$

их называют условной плотностью распределения и условным математическим ожиданием относительно случайного вектора η . При этом то обстоятельство, что функция $f_{\xi|\eta}(x|y)$ на множестве $\{y: y \in N_{\eta}\}$ определена произвольно, не играет роли, так как $P\{\eta \in N_{\eta}\} = 0$. Напомним, что равенство случайных величин означает их эквивалентность.

Из определения величины $M(\xi|\eta)$ вытекает, что $M(\xi|\eta) = g(\eta)$, где $g(y)$ — некоторая борелевская функция аргумента $y \in R^m$.

Она обладает следующим свойством:

для любой ограниченной борелевской функции $h(y)$, $y \in R^m$.

$$Mh(\eta) \xi = Mh(\eta) g(\eta). \quad (5)$$

действительно,

$$\begin{aligned} Mh(\eta) g(\eta) &= \int_{R^m} h(y) g(y) f_{\eta}(y) dy = \int_{R^m} \int_{R^d} h(y) x f(x, y) dx dy = \\ &= Mh(\eta) \xi. \end{aligned}$$

Важно отметить, что равенство (5) однозначно определяет случайную величину $g(\eta)$. Это означает, что если для всех ограниченных измеримых функций $h(y)$ равенство (5) выполняется для двух функций $g(y)$ и $\tilde{g}(y)$, то $g(\eta) = \tilde{g}(\eta)$ с вероятностью 1. Докажем это.

Имеем

$$Mh(\eta) \xi = Mh(\eta) g(\eta) = Mh(\eta) \tilde{g}(\eta),$$

откуда $Mh(\eta) (g(\eta) - \tilde{g}(\eta)) = 0$. Положим $h(\eta) = [g(\eta) - \tilde{g}(\eta)] \chi_c$, где $\chi_c = 1$ при $|g(\eta)| \leq C$ и $|\tilde{g}(\eta)| \leq c$ и $\chi_c = 0$ в противном случае. Получим $M\chi_c |g(\eta) - \tilde{g}(\eta)|^2 = 0$, откуда $\chi_c |g(\eta) - \tilde{g}(\eta)| = 0$ с вероятностью 1.

Переходя в этом равенстве к пределу при $c \rightarrow \infty$, получим $g(\eta) = \tilde{g}(\eta)$ с вероятностью 1.

Тот факт, что равенство (5) однозначно определяет случайную величину $\xi = g(\eta)$, позволяет дать следующее общее определение величины $M(\xi | \eta)$.

Определение. Условным математическим ожиданием $(\xi | \eta)$ случайной величины ξ ($M|\xi| < \infty$) относительно случайного вектора η называют случайную величину вида $g(\eta)$, где $g(y)$ — борелевская функция, такая, что для произвольной ограниченной борелевской функции $h(y)$ выполнено равенство (5).

Можно показать, что при $M|\xi| < \infty$ условное математическое ожидание $g(\eta) = M(\xi | \eta)$ всегда существует. На доказательстве этого факта мы не будем останавливаться, так как оно довольно сложно. Из предыдущего вытекает, что новое определение условного математического ожидания совпадает с первоначальным в тех случаях, когда первоначальное определение применимо. Более того, если для случайных вектора ξ и η имеют совместную плотность распределения $f(x, y)$, то для любой борелевской функции $\varphi(x)$, $x \in R^d$, для которой $M|\varphi(\xi)| < \infty$, имеет место формула

$$M(\varphi(\xi) | \eta) = \int_{R^d} \varphi(x) f_{\xi|\eta}(x | \eta) dx. \quad (6)$$

Для доказательства достаточно проверить, что выражение правой части этой формулы удовлетворяет вышеприведенному определению.

Обозначим правую часть равенства (6) через $g(\eta)$. Имеем

$$\begin{aligned} Mh(\eta) g(\eta) &= \int_{R^m} h(y) g(y) f_{\eta}(y) dy = \int_{R^m} \int_{R^d} h(y) \varphi(x) \times \\ &\times f_{\xi|\eta}(x|y) f_{\eta}(y) dy dx = \int_{R^m} \int_{R^d} h(y) \varphi(x) f(x, y) dy dx = Mh(\eta) \varphi(\xi). \end{aligned}$$

Таким образом, величина $g(\eta)$ удовлетворяет определению условного математического ожидания случайной величины $\varphi(\xi)$. Полагая в равенстве (5) $h(y) = 1$, получим

$$MM(\xi|\eta) = M\xi. \quad (7)$$

Укажем ряд свойств условного математического ожидания, в общем случае вытекающих из его определения.

Теорема 1. *Предположим, что все, встречающиеся ниже, случайные величины ξ_i , $i = 1, 2, \dots$ имеют конечные математические ожидания. Пусть c_i , $i = 1, 2$ — константы. Тогда:*

- а) $M(c_1\xi_1 + c_2\xi_2|\eta) = c_1M(\xi_1|\eta) + c_2M(\xi_2|\eta)$;
- б) если $\xi_1 \geq 0$, то $M(\xi_1|\eta) \geq 0$ (в частности, если $\xi_1 \leq \xi_2$, то $M(\xi_1|\eta) \leq M(\xi_2|\eta)$);
- в) если $\xi_1 \leq \xi_2 \leq \dots \leq \xi_n \leq \dots$ и $M \lim \xi_n < \infty$, то

$$\lim M(\xi_n|\eta) = M(\lim \xi_n|\eta)$$

с вероятностью 1.

Приведенные утверждения легко следуют из определения. Действительно, если $M(\xi_i|\eta) = g_i(\eta)$, то из формулы (5) получаем

$$Mh(\eta) (c_1\xi_1 + c_2\xi_2) = Mh(\eta) (c_1g_1(\eta) + c_2g_2(\eta)),$$

откуда, согласно определению, вытекает утверждение а). Аналогично с помощью предельного перехода под знаком математического ожидания для монотонных последовательностей случайных величин получим утверждение в). При этом следует заметить, что равенство (5) достаточно проверить для ограниченных неотрицательных функций $h(y)$. Для доказательства утверждения б) положим в (5) $h(y) = 0$, если $g_1(y) \geq 0$ и $h(y) = 1$, если $g_1(y) < 0$. Тогда $h(\eta)g_1(\eta) \leq 0$. Кроме того, согласно (5), $Mh(\eta)g_1(\eta) = Mh(\eta)\xi_1 \geq 0$, если $\xi_1 \geq 0$.

Таким образом, $h(\eta)g_1(\eta) = 0$ с вероятностью 1 и $P(g_1(\eta) < 0) = 0$.

Ряд важных свойств условных математических ожиданий приведен в следующей теореме.

Теорема 2. 1) Если векторы ξ и η независимы, $M|\varphi(\xi)| < \infty$, то

$$M(\varphi(\xi)|\eta) = M\varphi(\xi). \quad (8)$$

2) Для борелевских функций $\varphi(x)$ и $\psi(y)$, $x \in R^d$, $y \in R^m$,

$$M(\varphi(\xi)\psi(\eta)|\eta) = \psi(\eta) M(\varphi(\xi)|\eta) \quad (9)$$

(при условии $M|\varphi(\xi)\psi(\eta)| < \infty$).

3) Пусть $M|\varphi(\xi)| < \infty$, (η, ξ) — «составной» вектор, состоящий из компонент вектора η и вектора ξ . Тогда

$$M(\varphi(\xi)|\eta) = M(M(\varphi(\xi)|(\eta, \xi))|\eta). \quad (10)$$

Доказательство. 1) Поскольку $Mh(\eta)\varphi(\xi) = Mh(\eta)M\varphi(\xi) = M(h(\eta)(M\varphi(\xi)))$, то по определению в качестве $M(\varphi(\xi)|\eta)$ можно принять постоянную величину $M\varphi(\xi)$.

2) Пусть $g(\eta) = \psi(\eta)M(\varphi(\xi)|\eta)$. Тогда $Mh(\eta)g(\eta) = Mh(\eta) \times \times \psi(\eta)M(\varphi(\xi)|\eta)$. Допустим сначала, что функция $\psi(y)$ ограничена. Тогда $h_1(y) = h(y)\psi(y)$ тоже ограниченная функция и, согласно формуле (5), $Mh(\eta)\psi(\eta)M(\varphi(\xi)|\eta) = Mh(\eta)[\psi(\eta)\varphi(\xi)]$. По определению вытекает, что

$$M(\psi(\eta)\varphi(\xi)|\eta) = g(\eta).$$

Утверждение 2) теоремы доказано для ограниченных функций $\psi(y)$. Доказательство в общем случае можно получить с помощью предельного перехода. Пусть $\psi_c(y) = \varphi(y)$ при $|\psi(y)| \leq c$ и $\psi_c(y) = 0$ при $|\psi(y)| > c$. По доказанному к функции $\psi(y) = \psi_c(y)$ применимо утверждение 2) теоремы. В равенстве $Mh(\eta)\psi_c(\eta)M(\varphi(\xi)|\eta) = Mh(\eta)[\psi_c(\eta)\varphi(\xi)]$ можно перейти к пределу при $c \rightarrow \infty$ (по теореме о мажорируемой сходимости). Получим равенство, из которого следует (9) в общем случае.

3) Положим $g(\eta) = M[M(\varphi(\xi)|(\eta, \xi))|\eta]$ и пусть $h(y)$ — произвольная ограниченная борелевская функция. Дважды воспользовавшись равенством (9), получим

$$\begin{aligned} h(\eta)g(\eta) &= M[h(\eta)M(\varphi(\xi)|(\eta, \xi))|\eta] = \\ &= M[M(h(\eta)\varphi(\xi)|(\eta, \xi))|\eta]. \end{aligned}$$

Воспользовавшись теперь дважды равенством (7), имеем

$$\begin{aligned} Mh(\eta)g(\eta) &= MM[M(h(\eta)\varphi(\xi)|(\eta, \xi))|\eta] = \\ &= M[M(h(\eta)\varphi(\xi)|(\eta, \xi))] = Mh(\eta)\varphi(\xi). \end{aligned}$$

Таким образом, $g(\eta) = M(\varphi(\xi)|\eta)$.

Функция регрессии. Условное математическое ожидание играет важную роль в математической статистике. Эта роль связана с одним экстремальным свойством условного математического ожидания, появляющимся при решении следующей задачи.

Рассматриваются случайные величины $\eta, \xi_1, \xi_2, \dots, \xi_m$. В эксперименте может быть измерен вектор $\xi = (\xi_1, \xi_2, \dots, \xi_m)$, а нас интересует значение величины η . Требуется наилучшим образом оценить (предсказать или предвидеть) значение η в данном эксперименте по наблюдаемому значению ξ . Это означает, что нужно найти функцию $g(x) = g(x_1, \dots, x_m)$, для которой случайная

величина $t| \wedge (1)$ как можно меньше отличается от величины t^* . Чтобы уточнить задачу, нужно еще указать, как характеризуется или измеряется точность приближенного равенства $1] \quad t)$. В математической статистике широко применяется оценка точности приближенных равенств, основанная на среднем квадратическом отклоне $Н И И$. Положим

Величину 5^{\wedge} называют среднеквадратическим отклонением величины $z \setminus$ от $t]$. Принимают следующую точку зрения: чем меньше величина $б^{\wedge}$, тем точнее приближенное равенство $m \setminus$. Конечно, возможны и другие способы оценки точности равенства

Предложенный способ часто приводит к более простым решениям задачи, чем другие, и обладает интуитивной наглядностью. В подробных руководствах по математической статистике рассматривают общую теорию с произвольными способами измерения точности или качества приближенных формул. Некоторые аспекты возникающих здесь вопросов рассматриваются в гл. IX и X. Задачу, которую мы ставим, можно сформулировать следующим образом.

Пусть $L1|1|]^{\wedge} < \infty$. В классе H всех борелевских функций $H(x) \wedge H(X1, \dots, x^{\wedge})$ в $I^{\wedge m}$. для которых $PA \setminus H \setminus I \setminus]^{\wedge} < \infty$, нужно найти такую $\wedge(x)$, что

Эта задача может быть сформулирована на геометрическом языке. Как обычно, пусть $\wedge-5$ — гильбертово пространство всех случайных величин для которых $M |||^{\wedge} < \infty$, со скалярным произведением $(1 \setminus \underline{y} \wedge M \wedge 1^{\wedge} \underline{z} \text{ и } с \text{ расстоянием } p(|:, \quad = |||, - \quad || = = - ? 2 P \quad 3 И — линейное подпространство \quad состоящее из случайных величин \wedge \text{ вида } I = \quad \dots, \quad \text{Требуется найти элемент } \setminus \setminus^{\wedge} И, \text{ находящийся на кратчайшем расстоянии от элемента } z \setminus.$

Теорема. Наилучшая оценка $\underline{z} = \wedge(|1, \dots, \quad \text{величины } t]$, $L1t]^{\wedge} < \infty$, в смысле минимума среднеквадратической ошибки дается функцией

$$\wedge(-\wedge 1 \dots x;)^{\wedge} \wedge \{ \setminus I I = X_i \bullet \bullet \bullet > I m = X m \setminus -$$

Доказательство. Имеем

$$+ \quad m(\wedge(I) - l(I))^{\wedge}$$

причем, в силу формул (7) и (9) получим

то $M(\eta - g(\xi))(g(\xi) - h(\xi)) = 0$. Следовательно,
 $M|\eta - h(\xi)|^2 = M|\eta - g(\xi)|^2 + M|g(\xi) - h(\xi)|^2 \geq M|\eta - g(\xi)|^2$.

Замечание. Важно отметить геометрическую интерпретацию доказанной теоремы. Она показывает, что величина $\tilde{\eta} = M(\eta|\xi)$ является элементом из H , находящимся на кратчайшем расстоянии от $\eta \in L_2$, т. е. условное математическое ожидание $M(\eta|\xi)$ ($My^2 < \infty$) является ортогональной проекцией в L_2 элемента η на подпространство H .

Определение. Поверхность $y = g(x_1, \dots, x_m)$, $g(x_1, \dots, x_m) = M(\eta|\xi_1 = x_1, \dots, \xi_m = x_m)$ называют поверхностью (функцией) регрессии величины η на ξ .

Функцию регрессии часто интерпретируют следующим образом. Предположим, что между некоторыми физическими величинами y, x_1, \dots, x_m существует функциональная связь $y = g(x_1, \dots, x_m)$. Проводятся некоторые эксперименты, в которых величины x_1, \dots, x_m принимают случайные значения ξ_1, \dots, ξ_m , и измеряется величина $\eta = \xi + \Delta$, где $\xi = g(\xi_1, \dots, \xi_m)$, а Δ — ошибка измерения величины y , обладающая свойством «несмещенности»,

$$M\{\Delta|\xi_1 = x_1, \dots, \xi_m = x_m\} = 0 \quad \forall (x_1, \dots, x_m) \in R^m.$$

Тогда $M(\eta|\xi_1 = x_1, \dots, \xi_m = x_m) = g(x_1, \dots, x_m)$. При этом задача отыскания функции регрессии совпадает с задачей определения функциональной зависимости $y = g(x_1, \dots, x_m)$, которая может быть частично или полностью неизвестной. Напомним, что другая (точнее исходная) интерпретация функции регрессии состоит в следующем: величины $\eta, \xi_1, \dots, \xi_m$ случайны и между ними существует только статистическая (часто говорят стохастическая) зависимость. Функция $g = g(x_1, \dots, x_m)$ — это условное среднее значение величины η при заданных значениях $\xi_1 = x_1, \dots, \xi_m = x_m$.

§ 4. Слабая сходимость распределений

Сходимость в основном и слабая сходимость распределений. Предельные теоремы для распределений играют важную роль в теории вероятностей. Прежде чем переходить к этой теме, нужно ввести топологию в пространстве распределений. Простейшие предельные теоремы — предельная теорема Муавра—Лапласа и предельная теорема Пуассона — встречались раньше. В случае интегральной теоремы Муавра—Лапласа утверждалось, что функция распределения $F_n(x)$ случайной величины $\frac{v - np}{\sqrt{npq}}$ (v — число «успехов» в задаче Бернулли, n — число экспериментов, p — вероятность успеха в одном эксперименте, $q = 1 - p$) при каждом x сходится к нормальной функции распределения $\Phi(x)$. Казалось бы, что и в общем случае можно определить сходимость последовательности функций распределения $F_n(x)$ к пределу $F(x)$, как сходимость $F_n(x)$ к $F(x)$ при любом значении x . Однако такое опре-

деление в общем случае было бы неудобным. Поясним это на простейшем примере. Пусть $a_n, n = 1, 2, \dots$, — монотонно убывающая последовательность чисел и $a_n \uparrow 0 = a_\infty$. Можно рассматривать числа a_n как случайные величины с функциями распределения $F_n(x) = 0$ при $x \leq a_n$ и $F_n(x) = 1$ при $x > a_n$. Тогда $F_n(0) = 1 \neq F_\infty(0)$, так как $F_\infty(0) = 0$. С другой стороны, определение сходимости распределений было бы естественным, если бы из сходимости последовательности случайных величин ξ_n с вероятностью 1 к пределу ξ_∞ вытекала бы сходимость распределений величин ξ_n к распределению ξ_∞ . Определение сходимости распределений как поточечная сходимость $F_n(x)$ к $F_\infty(x)$, как мы видели, не обладает таким свойством. Заметим все же, что в приведенном примере $F_n(x) \rightarrow F_\infty(x)$ для всех x , кроме точек разрыва функции $F(x)$. Это обстоятельство оказывается типичным и для более сложных случаев. Учитывая это, можно дать следующее определение.

Определение. Последовательность функций распределения $F_n(x), x \in R^1$, сходится в основном к функции распределения $F(x)$, если $F_n(x) \rightarrow F(x)$ при $n \rightarrow \infty$ в каждой точке x , в которой функция $F(x)$ непрерывна.

Приведем сейчас определение «слабой сходимости распределений». На первый взгляд оно кажется более сложным и менее естественным, чем определение «сходимости в основном». На самом деле эти определения оказываются эквивалентными, но второе более приспособлено к обобщению на многомерный случай и даже на более абстрактные ситуации. Поэтому сформулируем его сразу для распределений в R^d .

Определение. Последовательность распределений $F_n(\cdot), n = 1, 2, \dots$, в R^d называют слабо сходящейся к распределению $F_0(\cdot)$,

$$F_n(\cdot) \Rightarrow F_0(\cdot),$$

если для произвольной ограниченной и непрерывной функции $g(x), x \in R^d$,

$$\int_{R^d} g(x) F_n(dx) \rightarrow \int_{R^d} g(x) F_0(dx). \quad (1)$$

Отметим прежде всего, что соотношение (1) единственным образом определяет предельное распределение $F_0(\cdot)$, если такое существует.

Теорема 1. Слабый предел последовательности распределений $F_n(\cdot), n = 1, 2, \dots$, если он существует, единственен.

Доказательство. Если последовательность $F_n, n = 1, 2, \dots$, имеет два слабых предела F' и F'' , то равенство

$$\int_{R^d} g(x) F'(dx) = \int_{R^d} g(x) F''(dx) \quad (2)$$

выполняется для всех непрерывных функций. Поэтому теорема 1 непосредственно вытекает из следующей леммы.

Лемма 1. Если F' и F'' — две меры на σ -алгебре борелевских множеств в R^d и равенство (2) выполняется для произвольных ограниченных непрерывных функций, то $F'(A) = F''(A)$ для всех борелевских A .

Доказательство. Обозначим через K класс всех измеримых функций, для которых выполняется равенство (2). Этот класс:

а) линейный, т. е., если $g_i(x) \in K$, $i = 1, 2$, то $c_1 g_1(x) + c_2 g_2(x) \in K$;

б) замкнут относительно ограниченной поточечной сходимости: если $f_n \in K$, $|f_n(x)| \leq c$, $\forall (n, x \in R^d)$ и

$$\lim f_n(x) = f_0(x),$$

то $f_0(x) \in K$ (это свойство класса K вытекает из теоремы о мажорируемой сходимости);

в) K содержит все непрерывные ограниченные функции. По минимальный класс функций, обладающий свойствами а), б), в), — это класс борелевских функций. Следовательно, K содержит все борелевские функции. Взяв в равенстве (2) в качестве $g(x)$ индикатор $\chi_A(x)$ борелевского множества A , получим $F'(A) = F''(A)$ для каждого борелевского множества A .

Заметим, что определение слабой сходимости относится к распределениям, а не к случайным величинам. Если случайные величины ξ_n , $n = 1, 2, \dots$, заданы на разных вероятностных пространствах, то нельзя ставить вопрос о сходимости в каком-либо смысле этой последовательности. Однако, и в этом случае можно рассматривать слабую сходимость распределений.

В терминах математических ожиданий условие слабой сходимости (1) распределений F_n можно записать так:

$$Mg(\xi_n) \xrightarrow{n \rightarrow \infty} Mg(\xi_0)$$

для любой ограниченной непрерывной функции $g(x)$. Здесь ξ_n — случайный вектор с распределением F_n , $n = 0, 1, \dots$.

Теорема 2. Если последовательность случайных векторов ξ_n , $n = 1, 2, \dots$, заданных на одном и том же вероятностном пространстве, сходится по вероятности к ξ_0 , то последовательность распределений F_n векторов ξ_n слабо сходится к распределению F_0 вектора ξ_0 .

Доказательство. Пусть $g(x)$ — произвольная ограниченная непрерывная функция, $|g(x)| \leq c$ и K_a — такой куб с центром в начале координат с длиной стороны $2a$, что $P\{\xi_0 \in K_a\} < \frac{\varepsilon}{2c}$, где $\varepsilon > 0$. Обозначим через $\chi(x)$ — индикатор куба K_a , $\chi'(x) = 1 - \chi(x)$. Имеем

$$|Mg(\xi_n) - Mg(\xi_0)| \leq |M(g(\xi_n) - g(\xi_0))\chi(\xi_0)| + |M(|g(\xi_n)| + |g(\xi_0)|)\chi'(\xi_0)|.$$

Функция $g(x)$ равномерно непрерывна в кубе K_{a+1} . Поэтому можно найти такое $\delta > 0$, что $|g(x) - g(y)| < \varepsilon$ при $x, y \in K_{a+1}$. Тогда

$$|M(g(\xi_n) - g(\xi_0)) \chi(\xi_0)| \leq \varepsilon + 2cP(|\xi_0 - \xi_n| > \delta) \quad (\delta < 1),$$

$$M(|g(\xi_n)| + |g(\xi_0)|) \chi'(\xi_0) \leq 2cP(\xi_0 \in K_0) \leq \varepsilon.$$

Мы видим, что при достаточно большом $n \geq n_0(\varepsilon)$ для любого $\varepsilon > 0$

$$|Mg(\xi_n) - Mg(\xi_0)| < 3\varepsilon,$$

т. е.

$$\lim Mg(\xi_n) = Mg(\xi_0).$$

Если бы в соотношении (1) вместо $g(x)$ можно было подставить $\chi_A(x)$ (индикатор множества $A \in \mathfrak{B}^d$, где \mathfrak{B}^d — σ -алгебра борелевских множеств в R^d), то в качестве следствия из слабой сходимости мы получили бы

$$\lim F_n(A) = F_0(A). \quad (3)$$

Но соотношение (1) по определению выполняется только для непрерывных функций, и поэтому равенство (3), вообще говоря, не имеет места. Тем не менее справедливо следующее утверждение.

Обозначим через $[A]$ замыкание множества A , а через $\text{Int } A$ множество всех внутренних точек множества A . Имеем $\text{Int } A \subset A \subset [A]$, $[A] \setminus \text{Int } A = \partial A$ — граница множества A .

Теорема 3. Если последовательность F_n , $n = 1, 2, \dots$, слабо сходится к F_0 , то для любого $A \in \mathfrak{B}^d$

$$F_0(\text{Int } A) \leq \liminf F_n(A) \leq \overline{\lim} F_n(A) \leq F_0([A]). \quad (4)$$

Доказательство. Пусть $\rho(x, A)$ — расстояние от точки x до множества A . Функция $\rho(x, A)$ непрерывна. Функции $g_N(x) = e^{-N\rho(x, A)}$ неотрицательны, непрерывны и образуют монотонно убывающую последовательность. При этом $\lim g_N(x) = \chi_{[A]}(x)$. Имеем

$$\overline{\lim} F_n(A) \leq \overline{\lim} \int_{R^d} g_N(x) F_n(dx) = \int_{R^d} g_N(x) F_0(dx).$$

Воспользовавшись теоремой о мажорируемой сходимости, перейдем в полученном неравенстве к пределу при $N \rightarrow \infty$. Получим

$$\overline{\lim} F_n(A) \leq \lim_{N \rightarrow \infty} \int_{R^d} g_N(x) F_0(dx) = \int_{R^d} \chi_{[A]}(x) F_0(dx) = F_0([A]).$$

Это доказывает одну часть утверждения теоремы. Другая является следствием этой. Действительно, применим доказанное неравенство к множеству $R^d \setminus A$. Заметим, что $[R^d \setminus A] = R^d \setminus \text{Int } A$. Получаем

$$\overline{\lim} F_n(R^d \setminus A) \leq F_0(R^d \setminus \text{Int } A), \text{ или } \overline{\lim} (1 - F_n(A)) \leq 1 - F_0(\text{Int } A),$$

$$\lim F_n(A) \geq F_0(\text{Int } A).$$

Определение. Множество A из \mathfrak{B}^d назовем множеством непрерывности меры F , если $F(\partial A) = 0$.

Теорема 4. Для слабой сходимости F_n к F_0 необходимо и достаточно, чтобы $\lim F_n(A) = F_0(A)$ для всякого множества непрерывности A меры F_0 .

Необходимость. Если A — множество непрерывности F_0 , то $F_0(\text{Int } A) = F_0(A) = F_0(\{A\})$ и из (3) следует

$$F_0(A) \leq \lim F_n(A) \leq \overline{\lim} F_n(A) < F_0(A),$$

так что $\lim F_n(A)$ существует и $\lim F_n(A) = F_0(A)$.

Достаточность. Пусть $g(x)$ — ограниченная непрерывная функция $E_a = \{x: g(x) = a\}$. Число тех значений a , при которых $F_0(E_a) \geq \frac{1}{n}$ не больше чем n . Действительно, при разных a множества E_a не пересекаются, и если $F_0(E_{a_k}) \geq \frac{1}{n}$ при $k = 1, \dots, N$, то $F_0(E_{a_1} \cup E_{a_2} \cup \dots \cup E_{a_N}) = \sum_{k=1}^N F_0(E_{a_k}) \geq \frac{N}{n}$, так что $N \leq n$.

Следовательно, множество значений a , для которых $F_0(E_a) = 0$ не более чем счетно, и в любом сколь угодно малом интервале $(c, c+h)$ найдется такое a ($a \in (c, c+h)$), что $F_0(E_a) = 0$. Возьмем произвольное $\varepsilon > 0$ и построим последовательность чисел b_k , $k = 0, 1, \dots, r$ так, чтобы $F_0(E_{b_k}) = 0$,

$$b_0 \leq \inf \{g(x), x \in \mathbb{R}^d\}, \quad b_r > \sup \{g(x), x \in \mathbb{R}^d\}, \\ b_k - b_{k-1} \leq \varepsilon, \quad k = 1, \dots, r.$$

Положим

$$B_k = \{x: b_{k-1} \leq g(x) < b_k\}, \quad g_\varepsilon(x) = \sum_{k=1}^r b_{k-1} \chi_{B_k}(x).$$

Тогда $0 \leq g(x) - g_\varepsilon(x) < \varepsilon$. Поскольку функция $g(x)$ непрерывна, то

$$[B_k] = \{x: b_{k-1} \leq g(x) \leq b_k\}, \quad \text{Int } B_k \subset \{x: b_{k-1} < g(x) < b_k\}$$

и граница множества B_k $\partial B_k \subset E_{b_{k-1}} \cup E_{b_k}$, так что $F_0(\partial B_k) = 0$ и $F_n(B_k) \rightarrow F_0(B_k)$ при $k = 1, \dots, r$. Следовательно,

$$\lim \left| \int_{\mathbb{R}^d} g(x) F_0(dx) - \int_{\mathbb{R}^d} g(x) F_n(dx) \right| \leq \lim \left| \int_{\mathbb{R}^d} (g(x) - g_\varepsilon(x)) \times \right. \\ \left. \times F_0(dx) + \sum_{k=1}^r b_{k-1} [F_0(B_k) - F_n(B_k)] + \int_{\mathbb{R}^d} [g_\varepsilon(x) - g(x)] F_n(dx) \right| \leq 2\varepsilon.$$

Переходя к пределу при $\varepsilon \downarrow 0$, получим

$$\lim \int_{R^d} g(x) F_n(dx) = \int_{R^d} g(x) F_0(dx).$$

Рассмотрим теперь связь между поточечной сходимостью функций распределения и слабой сходимостью распределений.

Теорема 5. Предположим, что последовательность $F_n(x)$ сходится к некоторой функции распределения $F_0(x)$ для всех $x \in D$, где D — всюду плотное множество в R^d . Тогда соответствующие распределения F_n слабо сходятся к распределению F_0 .

Прежде чем перейти к доказательству теоремы, сделаем несколько замечаний. Пусть $F(x)$ — некоторая функция распределения. Легко доказать неравенство

$$|F(x_1, x_2, \dots, x_d) - F(y_1, y_2, \dots, y_d)| \leq \sum_{j=1}^d |F(+\infty, \dots, +\infty,$$

$$x_j, +\infty, \dots, +\infty) - F(+\infty, \dots, +\infty, y_j, +\infty, \dots, +\infty)|.$$

Отсюда вытекает, что если точка x является точкой разрыва функции $F(x)$, то хотя бы одна из ее координат, например x_j , такова, что $F(+\infty, \dots, +\infty, x_j + 0, +\infty, \dots, +\infty) - F(+\infty, \dots, +\infty, x_j, +\infty, \dots, +\infty) > 0$. Множество значений x_j , для которых выполняется последнее соотношение, не более чем счетно. Пусть $x_j^k, k = 1, 2, \dots$, — все такие значения. Через H_k^j обозначим гиперплоскость в R^d ,

$$H_k^j = \{x = (x_1, \dots, x_j, \dots, x_d) : x_j = x_j^k\} \text{ и } H = \bigcup_{j=1}^d \bigcup_k H_k^j.$$

Тогда $F(x)$ непрерывна в точке x , если $x \notin H$. Итак, в R^d существует не более счетного множества гиперплоскостей, параллельных координатным гиперплоскостям $x_j = 0, j = 1, \dots, d$, вне которых функция $F(x)$ непрерывна. Предположим теперь, что последовательность функций распределения $F_n(x) \rightarrow F(x)$, на некотором всюду плотном в R^d множестве D значений x , где $F(x)$ — тоже функция распределения. Тогда $F_n(x) \rightarrow F(x)$ в каждой точке x , в которой $F(x)$ непрерывна. Действительно, пусть x — точка непрерывности $F(x)$. Найдутся последовательности $x^{(n)} \in D$ и $\bar{x}^{(n)} \in D, n = 1, 2, \dots$, такие, что $x^{(n)} \geq x, \bar{x}^{(n)} \leq x, x^{(n)} \rightarrow x, \bar{x}^{(n)} \rightarrow x$. Так как $F(x) - F_n(x) \leq F(x^{(n)}) - F_n(x^{(n)})$, то $\limsup (F(x) - F_n(x)) \leq F(x^{(n)}) - F(\bar{x}^{(n)})$. При $r \rightarrow \infty$ получим $\limsup_{n \rightarrow \infty} (F(x) - F_n(x)) \leq 0$. Аналогично показываем, что $\lim_{n \rightarrow \infty} (F(x) - F_n(x)) \geq 0$. Итак, $F(x) = \lim F_n(x)$.

Доказательство теоремы. Из сделанных замечаний вытекает следующее предположение: $F_n(x) \rightarrow F_0(x)$ для всех

$x \in H$, где H — сумма не более чем счетного множества гиперплоскостей, параллельных координатным. Зададим произвольно малое $\varepsilon > 0$. Пусть a и b выбраны так, что $a \in H$, $b \in H$ и $F_0(a, b) > 1 - \varepsilon$. При этом все вершины параллелепипеда (a, b) не лежат в H . Следовательно, $\lim F_n[a, b] = F_0[a, b] > 1 - \varepsilon$. Пусть $g(x)$ — произвольная ограниченная непрерывная функция. $c = \sup_{x \in R^d} |g(x)|$, $\delta = \delta(\varepsilon) > 0$ выбрано так, что $|g(x) - g(x')| < \varepsilon$ для всех x и x' из (a, b) , для которых $|x - x'| \leq \delta$. Разобьем $[a, b]$ на N параллелепипедов $[a_k, b_k]$, вершины которых не лежат в H , диаметра не более δ , попарно без общих точек, и пусть x_k — некоторая точка из $[a_k, b_k]$. Положим $g_\varepsilon(x) = 0$ при $x \in [a, b]$, $g_\varepsilon(x) = g(x_k)$ при $x \in [a_k, b_k]$. Тогда $|g(x) - g_\varepsilon(x)| < \varepsilon$ $\forall x \in [a, b]$, имеем

$$\begin{aligned} \overline{\lim} \left| \int_{R^d} g(x) F_0(dx) - \int_{R^d} g(x) F_n(dx) \right| &\leq \overline{\lim} \left(\left| \int_{R^d} g(x) [F_0(dx) - F_n(dx)] \right| + \right. \\ &\quad \left. + \left| \int_{[a, b]} (g(x) - g_\varepsilon(x)) (F_0(dx) - F_n(dx)) \right| + \right. \\ &\quad \left. + \left| \int_{[a, b]} g_\varepsilon(x) F_0(dx) - \int_{[a, b]} g_\varepsilon(x) F_n(dx) \right| \right) \leq \\ &\leq 2c\varepsilon + \varepsilon + \overline{\lim} \sum_{k=1}^N \varepsilon (F_0[a_k, b_k] - F_n[a_k, b_k]) = 2(c+1)\varepsilon. \end{aligned}$$

Переходя в крайней части неравенств к пределу при $\varepsilon \rightarrow 0$, получим

$$\overline{\lim} \int_{R^d} g(x) F_n(dx) = \int_{R^d} g(x) F_0(dx).$$

Из теорем 4 и 5 вытекает следствие.

Следствие. Последовательность распределений F_n слабо сходится к распределению F_0 тогда и только тогда, когда $F_n(x) \rightarrow F_0(x)$ на некотором счетном всюду плотном в D множестве значений x .

Слабая компактность распределений.

Определение. Множество вероятностных распределений $\{F_\varepsilon, \varepsilon \in E\}$ в R^d называется слабо компактным, если из произвольной последовательности $\{F_{\varepsilon_n}, n = 1, 2, \dots\}$ можно выделить слабо сходящуюся подпоследовательность.

Теорема 6. Для слабой компактности множества распределений $\{F_\varepsilon, \varepsilon \in E\}$ необходимо и достаточно, чтобы для любого $\varepsilon > 0$ нашлся такой d -интервал $[a, b] \rightarrow$, что

$$\sup_{\varepsilon} F_\varepsilon[a, b] \geq 1 - \varepsilon. \quad (5)$$

При доказательстве этой теоремы будет использована следующая теорема.

Теорема 7. (Теорема Хелли о выборе). *Какова бы ни была последовательность функций распределения F_n , $n = 1, 2, \dots$, в R^d , из нее можно выделить подпоследовательность F_{n_k} так, чтобы $F_{n_k}(x) \rightarrow F(x)$ в каждой точке непрерывности функции $F(x)$, где $F(x)$ — некоторая непрерывная слева монотонно неубывающая функция d -переменных, и*

$$\Delta_{h_1}^1 \Delta_{h_2}^2 \dots \Delta_{h_d}^d F(x) \geq 0 \quad \forall x, \quad h_1, h_2, \dots, h_d > 0.$$

Доказательство. Пусть $\{x_n, n = 1, 2, \dots\}$ — расположенное в некотором порядке счетное множество J точек R^d , координаты которых имеют вид $\pm \frac{j}{2^k}$, где j и k — целые числа

(или 0). Рассмотрим последовательность чисел $F_n(x_1)$. Так как она ограничена ($0 \leq F_n(x_1) \leq 1$), то из нее можно выделить подпоследовательность $F_{n_k}(x_1)$, сходящуюся к некоторому пределу $F_0(x_1)$, $F_{n_k}(x_1) \rightarrow F_0(x_1)$. Для удобства обозначим индекс n_k через $(k, 1)$. Из последовательности индексов $(k, 1)$ можно выделить новую подпоследовательность, обозначим ее $(k, 2)$, так чтобы $F_{(k, 2)}(x_2)$ сходилась к некоторому пределу $F_0(x_2)$. Этот процесс можно неограниченно продолжать. Тогда для любого натурального n будет найдена последовательность (k, n) , $k = 1, 2, \dots$, являющаяся подпоследовательностью последовательности $(k, n - 1)$, $k = 1, 2, \dots$, такая, что $F_{(k, n)}(x_n)$ сходится к некоторому пределу $F_0(x_n)$ при $k \rightarrow \infty$. Рассмотрим теперь «диагональную» последовательность индексов (n, n) . При каждом n , начиная с члена (n, n) , все ее члены принадлежат последовательности (k, n) , $k = 1, 2, \dots$, и поэтому для любого целого j имеем $\lim_{n \rightarrow \infty} F_{(n, n)}(x_j) = F_0(x_j)$. Последовательность чисел $F_0(x_j)$ будем

рассматривать как функцию, определенную на счетном, всюду плотном в R^d множестве J . Очевидно, она неотрицательна, $F(x_j) \leq 1$, монотонно не убывает, т. е. из $x_j \leq x_k$ следует $F(x_j) \leq F(x_k)$, и $\square_k F(x) \geq 0$ ($\square_k F(x) = \Delta_{h_1}^1 \Delta_{h_2}^2 \dots \Delta_{h_d}^d F(x)$). Определим

теперь для всех $x \in R^d$ функцию $F(x)$, положив $F(x) = \lim_{y \rightarrow x} F_0(y)$.

Под пределом $\lim_{y \rightarrow x} F_0(y)$ следует понимать предел последовательности $F_0(y_n)$, где $y_n \in J$, $y_1 \leq y_2 \leq \dots \leq y_n \leq \dots < x$ и $\lim y_n = x$. Этот предел существует и не зависит от выбора последовательности $y_n \rightarrow x$. Функция $F(x)$ монотонно не убывает и $\square_k F(x) \geq 0$ ($k \geq 0$). Кроме того, она непрерывна слева. Действительно, пусть $x_1 \leq x_2 \leq \dots$ и $\lim x_j = x$. Тогда для каждой точки x_j найдем точку $y_j < x_j$ такую, что $y_j \in J$, $y_1 < y_2 < \dots$ и $|x_j - y_j| \rightarrow 0$. При этом $0 \leq F(x) - F(x_j) \leq F(x) - F(y_j) \leq F(x) - F_0(y_j)$, так как $F(y_j) \leq F_0(y_j)$. Кроме того, $F(x) - F_0(y_j) \rightarrow 0$ при $j \rightarrow \infty$ по определению $F(x)$. Итак, $F(x)$ — непрерывна слева.

Пусть теперь x — точка непрерывности функции $F(x)$. Для любого $\varepsilon > 0$ можно тогда найти такие точки $y', y'' \in J$, $y' < x < y''$, что $F(x) - \varepsilon < F_0(y') \leq F_0(y'') < F(x) + \varepsilon$. Кроме того,

$$F_0(y') \leq \liminf F_{(n, n)}(x) \leq \limsup F_{(n, n)}(x) \leq F_0(y'').$$

Следовательно, $F(x) - \varepsilon \leq \liminf F_{(n, n)}(x) \leq \limsup F_{(n, n)}(x) \leq F(x) + \varepsilon$, т. е.

$$\lim F_{(n, n)}(x) = F(x).$$

Замечание. Очевидно, $0 \leq F(x) \leq 1$. С другой стороны, в теореме не утверждалось, что $F(x)$ — функция распределения, так как может оказаться, например, что $\lim_{x \rightarrow \infty} F(x) < 1$ или $\lim_{x \rightarrow -\infty} F(x) > 0$.

Доказательство теоремы 6. Необходимость. Предположим, что условие (5) не выполнено. Тогда для некоторого $\varepsilon > 0$ и любого $n = (n, \dots, n)$ найдется e_n такое, что $F_{e_n}[-n, n] < 1 - \varepsilon$. Предположим, что из последовательности F_{e_n} можно выбрать подпоследовательность $F_j = F_{n_j}$, слабо сходящуюся к некоторому распределению F_0 . Для каждого k найдем такой d -интервал $[a_k, b_k]$, что $[-k, k] \subset [a_k, b_k] \subset [-(k+1), k+1]$ и вершины d -интервала $[a_k, b_k]$ не лежат в гиперплоскостях, в которых F_0 имеет разрывы. Так как при $n_j < k$ $F_j[a_k, b_k] < 1 - \varepsilon$, то и $F_0[a_k, b_k] < 1 - \varepsilon$. Перейдем в последнем неравенстве к пределу при $k \rightarrow \infty$. Получим $F_0(R^d) < 1 - \varepsilon$, что противоречит предположению. Таким образом, если (5) не выполнено, семейство распределений не может быть слабо компактным.

Достаточность. Пусть F_{e_n} , $n = 1, 2, \dots$ — некоторая последовательность распределений. Выберем из нее подпоследовательность F_j такую, чтобы функции распределения $F_j(x)$ сходились к некоторому пределу $F(x)$, в соответствии с теоремой Хелли о выборе. Из неравенства (5) тогда следует, что для любого n найдется d -интервал $[a_n, b_n]$, такой, что $F[a_n, b_n] > 1 - \frac{1}{n}$. Из монотонности функции F вытекает, что $F(x) \geq F[a_n, b_n] > 1 - \frac{1}{n}$, как только $x > b_n$. Итак, $F(x) \rightarrow 1$ при $x \rightarrow \infty$. Аналогично, если $\min(x_1, \dots, x_d) \rightarrow -\infty$, то $F(x) \rightarrow 0$. Следовательно, $F(x)$ является функцией распределения. По теореме 5 последовательность распределений F_j слабо сходится к F .

§ 5. Характеристические функции

Определения и элементарные свойства. При решении многих задач теории вероятностей характеристические функции распределений играют важную роль.

В ряде случаев оказывается целесообразнее задавать распределения не мерами или плотностями, а характеристическими

функциями. Кроме того, ряд аналитических свойств и преобразований распределений легче описывать с помощью тех же характеристических функций. Случаи характеристических функций одномерных распределений уже рассматривались в § 9, гл. II.

Определение. *Характеристической функцией $\varphi(u) = \varphi_{\xi}(u)$, $u \in R^d$, распределения случайного вектора ξ в R^d (или короче случайного вектора ξ) называют функцию*

$$\varphi(u) \stackrel{\text{Def}}{=} M e^{i(u, \xi)}. \quad (1)$$

Напомним, что (u, ξ) — скалярное произведение векторов u и ξ . Поскольку $|e^{i(u, \xi)}| = 1$, то математическое ожидание в правой части равенства (1) существует для любого случайного вектора ξ и $u \in R^d$. Если распределение вектора ξ обладает плотностью $f(x)$, то в соответствии с формулой (4), § 1,

$$\varphi(u) = \int e^{i(u, x)} f(x) dx = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{i \sum_{k=1}^d u_k x_k} f(x_1, \dots, x_d) dx_1 \dots dx_d. \quad (2)$$

Напомним следующее определение:

Определение. *Для произвольной интегрируемой функции $f(x)$ функция $\varphi(u)$, определяемая равенством (2), называется преобразованием Фурье функции $f(x)$.*

Таким образом, характеристическая функция распределения является преобразованием Фурье плотности распределения, если плотность существует. При широких предположениях функцию, независимо от того, является ли она плотностью распределения, можно выразить через $\varphi(u)$ с помощью «обратного преобразования Фурье»

$$f(x) = \frac{1}{(2\pi)^d} \int_{R^d} e^{-i(u, x)} \varphi(u) du. \quad (3)$$

В одномерном случае это имеет, например, место, если функция $f(x)$ интегрируема и дифференцируема в окрестности данной точки x . Преобразование Фурье является важным методом математического анализа и его теория изучается в так называемом гармоническом анализе. Краткие сведения из этой теории, в том числе теоремы об обращении преобразования Фурье, будут приведены ниже.

В дискретном случае (случайный вектор ξ принимает значения $x^{(1)}, x^{(2)}, \dots, x^{(n)}, \dots, x^{(n)} \in R^d$, с вероятностями p_n , $p_n = P(\xi \in x^{(n)})$) для характеристической функции получаем следующее выражение:

$$\varphi(u) = \sum_{j=1}^{\infty} e^{i(u, x^{(j)})} p_n. \quad (4)$$

Тригонометрический ряд (4) сходится абсолютно, так как $|e^{i(u, x^{(n)})} p_n| = p_n$, $\sum_1^\infty p_n = 1$. Умножив равенство (4) на $\exp[-i(u, x^{(r)})]$ и проинтегрировав по d -мерному кубу $\{u_k \in [-T, T], k = 1, \dots, d\}$, получим

$$\frac{1}{(2T)^d} \int_{-T}^T \dots \int_{-T}^T e^{-i(u, x^{(r)})} \varphi(u) du = p_r + \sum_{n \neq r} p_n \frac{1}{(2T)^d} \int_{-T}^T \dots \int_{-T}^T e^{i(u, x^{(n)} - x^{(r)})} du = p_r + \sum_{n \neq r} \prod_{k=1}^d \frac{\sin T(x_k^{(n)} - x_k^{(r)})}{T(x_k^{(n)} - x_k^{(r)})}.$$

Переходя к пределу при $T \rightarrow \infty$, учитывая абсолютную сходимость ряда в правой части равенства, имеем

$$p_r = \lim_{T \rightarrow \infty} \frac{1}{(2T)^d} \int_{-T}^T \dots \int_{-T}^T e^{i(u, x^{(r)})} \varphi(u) du. \quad (5)$$

Последняя формула показывает, как в дискретном случае вероятности p_r выражаются через характеристическую функцию распределения. Покажем теперь в общем случае, что характеристическая функция однозначно определяет распределение случайного вектора. Этот факт непосредственно вытекает из указанной ниже общей формулы обращения, но мы сначала приведем простое непосредственное доказательство.

Теорема 1. *Характеристическая функция однозначно определяет распределение случайного вектора.*

Доказательство. Пусть ξ_1 и ξ_2 — два случайных вектора, имеющие одну и ту же характеристическую функцию $\varphi(u)$. Обозначим через K класс функций $f(x)$, для которых

$$Mf(\xi_1) = Mf(\xi_2).$$

Класс K линейный. В этот класс входят функции вида $\exp\{i \sum_{k=1}^d u_k x_k\}$, $-\infty < u_k < \infty$ и всевозможные линейные комбинации этих функций. Пусть $f(x)$ — произвольная ограниченная непрерывная функция, $|f(x)| < c$, ε — произвольное положительное число, C — такой куб в R^d , что $P(|\xi_i| \in C) < \frac{\varepsilon}{C}$, $i = 1, 2$. По теореме Вейерштрасса можно найти такую непрерывную и ограниченную функцию $g_\varepsilon(x)$ ($|g_\varepsilon(x)| < C$) вида $g_\varepsilon(x) = \sum_{k=1}^n a_k \times \exp\{i(u^{(k)}, x)\}$, что $|f(x) - g_\varepsilon(x)| < \varepsilon$, $\forall x \in C$. Тогда

$$|M(f(\xi_i) - g_\varepsilon(\xi_i))| \leq M\chi_C |f(\xi_i) - g_\varepsilon(\xi_i)| + M(1 - \chi_C) |f(\xi_i) - g_\varepsilon(\xi_i)| \leq \varepsilon + CP(\xi_i \in C) \leq 2\varepsilon,$$

где $\chi_C = \chi_C(\xi_i)$, $\chi_C(x)$ — индикатор куба C . Далее

$$|Mf(\xi_1) - Mf(\xi_2)| \leq |M(f(\xi_1) - g_\varepsilon(\xi_1))| + |Mg_\varepsilon(\xi_1) - Mg_\varepsilon(\xi_2)| + |M(g_\varepsilon(\xi_2) - f(\xi_2))| \leq 4\varepsilon,$$

поскольку $Mg_\varepsilon(\xi_1) = Mg_\varepsilon(\xi_2)$. Так как левая часть неравенства не зависит от ε , то отсюда вытекает, что $Mf(\xi_1) = Mf(\xi_2)$, т. е. $f \in K$. Таким образом, K содержит все непрерывные функции. Из леммы 1, § 4, следует что распределения векторов ξ_1 и ξ_2 совпадают.

Отметим ряд простых свойств характеристических функций:

1) $\varphi(0) = 1$, $|\varphi(u)| \leq 1$ (это свойство непосредственно вытекает из определения);

2) $\varphi(-u) = \overline{\varphi(u)}$ (здесь $\overline{\varphi}$ обозначает число, комплексно сопряженное с φ);

3) Пусть A — линейный оператор, отображающий R^d в R^m , b — вектор из R^m , тогда

$$\varphi_{A\xi+b}(u) = e^{i(b, u)} \varphi_\xi(A^*u), \quad (6)$$

где A^* — оператор, сопряженный с A .

Доказательство формулы (6) довольно простое:

$$\begin{aligned} \varphi_{A\xi+b}(u) &= Me^{i(A\xi+b, u)} = Me^{i(A\xi, u)} e^{i(b, u)} = \\ &= e^{i(b, u)} Me^{i(\xi, A^*u)} = e^{i(b, u)} \varphi_\xi(A^*u). \end{aligned}$$

4) Если ξ' — вектор, состоящий из первых r компонент вектора ξ , $\xi = (\xi_1, \dots, \xi_d)$, $\xi' = (\xi_1, \dots, \xi_r)$, $r < d$, $u = (u_1, \dots, u_d)$, $u' = (u_1, \dots, u_r)$, то

$$\varphi_{\xi'}(u') = \varphi_\xi(u_1, u_2, \dots, u_r, 0, \dots, 0).$$

Доказательство вытекает из определения характеристической функции.

$$5) |\varphi(u + \Delta u) - \varphi(u)| \leq \sqrt{2(1 - \operatorname{Re} \varphi(\Delta u))}. \quad (7)$$

Действительно,

$$\begin{aligned} |\varphi(u + \Delta u) - \varphi(u)| &= |M[e^{i(\xi, u + \Delta u)} - e^{i(\xi, u)}]| \leq \\ &\leq M|e^{i(\xi, \Delta u)} - 1| \leq \sqrt{M|e^{i(\xi, \Delta u)} - 1|^2}. \end{aligned}$$

Кроме того, $|1 - e^{i\alpha}|^2 = 2(1 - \cos \alpha)$.

Таким образом,

$$|\varphi(u + \Delta u) - \varphi(u)| \leq \sqrt{2M(1 - \cos(\xi, \Delta u))},$$

откуда и следует (7).

6) Функция $\varphi(u)$ равномерно непрерывна в R^m .

На основании неравенства (7) достаточно показать, что функция непрерывна в одной точке $u = 0$. Поскольку $|e^{i(\xi, \Delta u)}| = 1$ и $\lim_{\Delta u \rightarrow 0} e^{i(\xi, \Delta u)} = 1$ для всех ξ , то по теореме о мажорируемой сходимости $Me^{i(\xi, \Delta u)} \rightarrow 1$ при $\Delta u \rightarrow 0$.

7) Если векторы ξ и η независимы, то

$$\varphi_{\xi+\eta}(u) = \varphi_{\xi}(u) \varphi_{\eta}(u).$$

Действительно, из независимости векторов ξ и η вытекает независимость случайных векторов $e^{i(u, \xi)}$ и $e^{i(u, \eta)}$. Поэтому

$$\varphi_{\xi+\eta}(u) = Me^{i(u, \xi+\eta)} = Me^{i(u, \xi)} e^{i(u, \eta)} = Me^{i(u, \xi)} Me^{i(u, \eta)}.$$

Характеристическая функция и моменты распределения. Для случайного вектора ξ со значениями в R^d момент C_{j_1, \dots, j_d} определяется формулой

$$C_{j_1, \dots, j_d} = \overline{M} (\xi_1)^{j_1} (\xi_2)^{j_2} \dots (\xi_d)^{j_d},$$

если выражение в правой части равенства имеет смысл. Величину $|j| = j_1 + j_2 + \dots + j_d$ называют порядком момента. В дальнейшем нас будет интересовать только тот случай, когда j_1, j_2, \dots, j_d — целые числа. Заметим, что если $M|\xi_k|^p < \infty$, $k = 1, \dots, d$, то произвольные моменты C_{j_1, \dots, j_d} порядка $|j| \leq p$ существуют. Действительно, из неравенства между средним арифметическим и средним геометрическим следует

$$\prod_{k=1}^d |\xi_k|^{j_k} = \prod_{k=1}^d |\xi_k|^{|j| \frac{j_k}{|j|}} \leq \sum_{k=1}^d \frac{j_k}{|j|} |\xi_k|^{|j|},$$

откуда

$$M \prod_{k=1}^m |\xi_k|^{j_k} \leq \sum_{k=1}^m \frac{j_k}{|j|} M |\xi_k|^{|j|} \leq \sum_{k=1}^m \frac{j_k}{|j|} (M |\xi_k|^p)^{\frac{|j|}{p}} < \infty.$$

Моменты распределения можно вычислять, зная характеристическую функцию, с помощью дифференцирования. Дифференцируя формулу (1), получаем

$$\frac{\partial^{|j|} \varphi(u)}{\partial u_1^{j_1} \partial u_2^{j_2} \dots \partial u_d^{j_d}} = M (i\xi_1)^{j_1} \dots (i\xi_d)^{j_d} \exp \{i(u, \xi)\}. \quad (8)$$

Дифференцировать под знаком математического ожидания можно, если $M|\xi_k|^{|j|} < \infty$, $k = 1, \dots, d$. Полагая в формуле (8) $u = 0$, получим

$$C_{j_1, \dots, j_d} = (-i)^{|j|} \frac{\partial^{|j|} \varphi(u)}{\partial u_1^{j_1} \dots \partial u_d^{j_d}} \Big|_{u=0}. \quad (9)$$

Таким образом, если $M|\xi_k|^p < \infty$, $k = 1, \dots, d$, то характеристическая функция $\varphi(u)$ имеет непрерывные частные производные порядка не более p , и моменты распределения порядка не более p могут быть определены с помощью формулы (9).

В ряде случаев бывает известна характеристическая функция и желательно установить существование моментов распределения. Однако обращение предыдущего утверждения не всегда имеет место. Все же оно имеет место, если существуют частные производные вида $\frac{\partial^{2j}}{\partial u_k^{2j}} \varphi(u)$ при $u = 0$. Для доказательства можно ограничиться, не уменьшая общности, одномерным случаем, так как

$$\left. \frac{\partial^{2j}}{\partial u_1^{2j}} \varphi(u) \right|_{u=0} = \left. \frac{\partial^{2j}}{\partial u_1^{2j}} \varphi(u_1, 0, \dots, 0) \right|_{u_1=0} = \left. \frac{d^{2j}}{du^{2j}} \varphi_{\cdot 1}(u) \right|_{u=0}.$$

Имеем, например, при $d = 1$

$$\begin{aligned} -\left. \frac{d^2}{du^2} \varphi(u) \right|_{u=0} &= \lim_{h \rightarrow 0} -\frac{\varphi(h) - 2 + \varphi(-h)}{h^2} = \\ &= \lim_{h \rightarrow 0} M2(1 - \cos h\xi) = \lim_{h \rightarrow 0} M \frac{4}{h^2} \sin^2 \frac{h\xi}{2}. \end{aligned}$$

Используя лемму Фату, получим

$$-\left. \frac{d^2}{du^2} \varphi(u) \right|_{u=0} \geq M \lim_{h \rightarrow 0} \frac{4}{h^2} \sin^2 \frac{h\xi}{2} = M\xi^2.$$

Таким образом, $M\xi^2 < \infty$ и по ранее доказанному $M\xi^2 = -\left. \frac{d^2}{du^2} \varphi(u) \right|_{u=0}$. Аналогично рассматривается случай, когда существует $\frac{d^{2j}}{du^{2j}} \varphi(u)$ в точке $u = 0$.

Теорема 2. Если существуют производные $\frac{\partial^{2p} \varphi(u)}{\partial u_k^{2p}}$, $k = 1, 2, \dots, d$, то существуют моменты $C_{j_1 \dots j_d}$ при $j_1 + \dots + j_d \leq 2p$ и они задаются формулой (9).

ПРИМЕРЫ ВЫЧИСЛЕНИЯ ХАРАКТЕРИСТИЧЕСКИХ ФУНКЦИЙ

1. Мультиномиальное распределение. Оно определяется следующим образом: производится n испытаний, в каждом из которых с одинаковой вероятностью, не зависящей от результатов других испытаний, происходит одно из событий E_1, E_2, \dots, E_d . Вероятность того, что в данном испытании произойдет событие E_k , равна p_k , $k = 1, 2, \dots, d$, $p_1 + p_2 + \dots + p_d = 1$. Найдем вероятность того, что в произведенных испытаниях событие E_k произойдет k_j раз $j = 1, \dots, d$, $k_1 + k_2 + \dots + k_d = n$. Сформулированная задача является обобщением задачи Бернулли и сводится к ней, если $d = 2$. Искомая вероятность, обозначим ее через $p_{k_1 k_2 \dots k_d}$, очевидно, равна

$$p_{k_1 k_2 \dots k_d} = C p_1^{k_1} p_2^{k_2} \dots p_d^{k_d},$$

где константа C равна числу способов, которыми можно образовать из n объектов, каждый из которых может быть объектом одного из d различных

видов, последовательностей, содержащих точно k_1 объектов первого вида, k_2 — второго, и так далее, k_d — объектов последнего вида. Из основного правила комбинаторики сразу вытекает, что

$$C = C_n^{k_1} C_{n-k_1}^{k_2} \dots C_{n-k_1-\dots-k_{d-2}}^{k_{d-1}} = \\ = \frac{n!}{k_1! (n-k_1)!} \cdot \frac{(n-k_1)!}{k_2! (n-k_1-k_2)!} \dots \frac{(n-k_1-\dots-k_{d-2})!}{k_{d-1}! k_d!} = \\ = \frac{n!}{k_1! k_2! \dots k_d!}.$$

Итак,

$$p_{k_1 k_2 \dots k_d} = \frac{n!}{k_1! k_2! \dots k_d!} p_1^{k_1} p_2^{k_2} \dots p_d^{k_d}.$$

Набор вероятностей $p_{k_1 k_2 \dots k_d}$ ($0 < k_j \leq n$, $\sum_{j=1}^d k_j = n$) определяет распределение d -мерной дискретной величины ξ , принимающей значения вида (k_1, k_2, \dots, k_d) . Это распределение называют мультиномиальным.

Его характеристическая функция

$$\varphi(u) = \sum_{\substack{k_1, \dots, k_d \\ \sum_{j=1}^d k_j = n}} \frac{n!}{k_1! k_2! \dots k_d!} \prod_{j=1}^d p_j^{k_j} e^{i \sum_{j=1}^d k_j u_j}.$$

Кроме того, если ξ^j — вектор с компонентами $(\chi_1^j, \chi_2^j, \dots, \chi_d^j)$, где $\chi_k^j = 1$ или 0 в зависимости от того, произошло ли событие E_k в j -м испытании или нет, то

$$\xi = \xi^{(1)} + \dots + \xi^{(n)}.$$

Векторы ξ^j — взаимно независимы и одинаково распределены. Характеристическая функция $\varphi_0(u)$ вектора $\xi^{(j)}$ имеет вид

$$\varphi_0(u) = \text{Me}^{i(\xi^{(j)}, u)} = \sum_{k=1}^d p_k e^{i u_k}.$$

Следовательно,

$$\varphi(u) = [\varphi_0(u)]^n = \left(\sum_{k=1}^d p_k e^{i u_k} \right)^n. \quad (10)$$

Сравнивая последнее выражение для $\varphi(u)$ с ранее полученным, приходим к мультиномиальной теореме

$$\left(\sum_{k=1}^d a_k \right)^n = \sum_{\substack{k_1, \dots, k_d \\ \sum_{j=1}^d k_j = n}} \frac{n!}{k_1! k_2! \dots k_d!} a_1^{k_1} \dots a_d^{k_d}. \quad (11)$$

Эта формула установлена для произвольных чисел вида $a_k = p_k e^{i\alpha_k}$, но легко заметить, что она остается справедливой и для любых a_k (в общем случае достаточно положить $p_k = \frac{|a_k|}{\sum |a_k|}$, $e^{i\alpha_k} = \frac{a_k}{|a_k|}$).

Ряд выражений для характеристических функций многомерных распределений можно получить из формул для одномерного случая, когда компоненты случайного вектора независимы.

2. d -мерное равномерное распределение. Пусть $\xi = (\xi_1, \dots, \xi_d)$ — вектор, равномерно распределенный в d -мерном кубе $C = \{x: -a < x < a\}$, $a = (a, a, \dots, a)$. Тогда существует плотность распределения

$$f(x) = \frac{1}{(2a)^d}, \quad x \in C,$$

$$f(x) = 0, \quad x \notin C,$$

компоненты вектора ξ независимы и характеристическая функция распределения равна (см. § 9, гл. II):

$$\varphi(u) = \prod_{k=1}^d \frac{\sin ax_k}{ax_k}.$$

3. d -мерная изотропная гауссовская величина. Пусть компоненты вектора ξ независимы и нормально распределены с параметрами $(0, \sigma^2)$. Плотность распределения равна:

$$f(x) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left\{-\frac{\|x\|^2}{2\sigma^2}\right\}, \quad (12)$$

где $\|x\|$ — норма вектора x , $\|x\| = \sqrt{\sum_{k=1}^d x_k^2}$, а характеристическая функция

$$\varphi(u) = \exp\left\{-\frac{\sigma^2 \|u\|^2}{2}\right\}. \quad (13)$$

4. d -мерное распределение Коши. При рассмотрении этого примера нам понадобятся некоторые вычисления с n -кратными интегралами.

Напомним, что в одномерном случае распределением Коши (центрированным) называют распределение с плотностью

$$f(x) = \frac{1}{\pi} \frac{a}{a^2 + x^2}.$$

Заметим, что

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-a|u|} e^{-iux} du = \frac{1}{2\pi} \left(\int_0^{\infty} e^{-(a+x)u} du + \int_0^{\infty} e^{-(a-x)u} du \right) = \frac{1}{\pi} \frac{a}{a^2 + x^2}.$$

Из формул для преобразования Фурье следует, что

$$\int_{-\infty}^{\infty} \frac{1}{\pi} \frac{a}{a^2 + x^2} e^{iux} dx = e^{-a|u|}, \quad (14)$$

т. е. $\varphi(u) = e^{-a|u|}$ — характеристическая функция плотности Коши.

Назовем d -мерным распределением Коши распределение с характеристической функцией

$$\varphi(u) = e^{-a\|u\|}, \quad \|u\| = \sqrt{\sum_{k=1}^d u_k^2}.$$

Еще предстоит доказать, что $\varphi(u)$ действительно является характеристической функцией. Заодно будет найдено выражение для плотности распределения Коши. С этой целью вычислим обратное преобразование Фурье функции $\varphi(u)$. Заметим, что

$$\begin{aligned} e^{-a\|u\|} &= \frac{a}{\pi} \int_{-\infty}^{\infty} \frac{1}{a^2 + t^2} e^{iut} dt = \frac{a}{\pi} \int_{-\infty}^{\infty} \left(\int_0^{\infty} e^{-(a^2 + t^2)x} dx \right) e^{iut} dt = \\ &= \frac{a}{\pi} \int_0^{\infty} e^{-a^2 x} dx \int_{-\infty}^{\infty} e^{-t^2 x} e^{iut} dt. \end{aligned}$$

Внутренний интеграл с точностью до множителя является характеристической функцией нормальной плотности $\left(\frac{x}{4\pi}\right)^{\frac{1}{2}} e^{-t^2 x}$. Следовательно,

$$\int_{-\infty}^{\infty} e^{-t^2 x} e^{iut} dt = \sqrt{\frac{\pi}{x}} e^{-\frac{u^2}{4x}}$$

и

$$e^{-a\|u\|} = \frac{a}{\sqrt{\pi}} \int_0^{\infty} \frac{1}{\sqrt{x}} e^{-a^2 x - \frac{u^2}{4x}} dx. \quad (15)$$

Поэтому

$$\frac{1}{(2\pi)^d} \int_{R^d} \varphi(u) e^{-i(u, t)} du = \frac{a}{\sqrt{\pi} (2\pi)^d} \int_0^{\infty} \frac{1}{\sqrt{x}} e^{-a^2 x} \int_{R^d} e^{-\frac{\|u\|^2}{4x} - i(u, t)} du dx.$$

Из формул примера 2 следует

$$\int_{R^d} e^{-\frac{\|u\|^2}{4x} - i(u, t)} du = (2\pi \cdot 2x)^{\frac{d}{2}} e^{-x\|t\|^2},$$

т. е.

$$\begin{aligned} \frac{1}{(2\pi)^d} \int_{R^d} e^{-a\|u\|} e^{-i(u, t)} du &= \frac{a}{\sqrt{\pi} (2\pi)^d} \int_0^{\infty} \frac{(4\pi x)^{\frac{d}{2}}}{\sqrt{x}} e^{-(a^2 + \|t\|^2)x} dx = \\ &= \frac{a}{\pi^{\frac{d+1}{2}}} \int_0^{\infty} x^{\frac{d-1}{2}} e^{-(a^2 + \|t\|^2)x} dx = \frac{\Gamma\left(\frac{d+1}{2}\right)}{\pi^{\frac{d+1}{2}}} \frac{a}{(a^2 + \|t\|^2)^{\frac{d+1}{2}}}. \end{aligned}$$

Из формул для преобразования Фурье (2) — (3) вытекает

$$\int_{R^d} \frac{c_d a}{(a^2 + \|t\|^2)^{\frac{d+1}{2}}} e^{i(u, t)} dt = e^{-a\|u\|}, \quad c_d = \frac{\Gamma\left(\frac{d+1}{2}\right)}{\pi^{\frac{d+1}{2}}}. \quad (16)$$

Полагая здесь $u=0$, получим

$$\int_{R^d} \frac{c_d a}{(a^2 + \|t\|^2)^{\frac{d+1}{2}}} dt = 1.$$

Мы видим, что функция $\exp(-a\|u\|)$ действительно является характеристической функцией плотности распределения $k(x)$ в R^d , названной ранее плотностью Коши, причем

$$k(x) = \frac{c_d a}{(a^2 + \|x\|^2)^{\frac{d+1}{2}}}. \quad (17)$$

В доказательстве использована формула обращения Фурье. Отметим, что в предыдущих выкладках было получено равенство

$$\frac{1}{(2\pi)^d} \int_{R^d} e^{-a\|u\| - i(u, x)} du = \frac{ac_d}{(a^2 + \|x\|^2)^{\frac{d+1}{2}}}, \quad (18)$$

не использующее формул обращения для интеграла Фурье. Его можно интерпретировать следующим образом: функция $\frac{1}{c_d} \left(\frac{a}{2\pi}\right)^d e^{-a\|u\|}$ является плотностью распределения с характеристической функцией $\left(1 + \left\|\frac{x}{a}\right\|^2\right)^{\frac{d+1}{2}}$.

Предельная теорема для характеристических функций. Следующая теорема показывает, что характеристические функции можно использовать для исследования слабой сходимости распределений.

Теорема 3. (Теорема Леви). Для того чтобы последовательность распределений F_n , $n = 1, 2, \dots$ в R^d слабо сходилась к некоторому распределению F_0 , необходимо и достаточно, чтобы характеристические функции $\varphi_n(u)$ распределений F_n сходились при каждом u к функции $\psi(u)$, непрерывной в точке $u=0$. В этом случае $\psi(u)$ является характеристической функцией распределения F_0 .

Доказательство. Если $F_n \Rightarrow F_0$, то для любой непрерывной и ограниченной функции $g(x)$

$$\int_{R^d} g(x) F_n(dx) \rightarrow \int_{R^d} g(x) F_0(dx).$$

В частности, это справедливо для функции $g(x) = e^{i(u, x)}$, так что $\varphi_n(u) \rightarrow \varphi_0(u)$ при любом u и предельная функция $\varphi_0(u)$ непрерывна. *Необходимость* условий теоремы доказана. Докажем их *достаточность*. Покажем сначала, что последовательность F_n , $n = 1, 2, \dots$ слабо компактна.

Пусть $A = (a, a, \dots, a) \in R^d$, $a > 0$, $A' = \left(\frac{2}{a}, \frac{2}{a}, \dots, \frac{2}{a}\right)$. Имеем

$$\begin{aligned} \frac{1}{(2\pi)^d} \int_{[-A, A]} (1 - \varphi_n(u)) du &= \frac{1}{(2\pi)^d} \int_{[-A, A]} \int_{R^d} (1 - e^{i(u, x)}) F_n(dx) du = \\ &= \int_{R^d} \left(1 - \prod_{k=1}^d \frac{\sin ax_k}{ax_k}\right) F_n(dx) \geq \frac{1}{2} \int_{[-A', A']^c} F_n(dx) = \frac{1}{2} F_n([-A', A']^c), \end{aligned} \quad (19)$$

так как $\left|\frac{\sin ax_k}{ax_k}\right| \leq 1$ и если $|x_k| > \frac{2}{a}$, то $\left|\frac{\sin ax_k}{ax_k}\right| < \frac{1}{2}$ так, что при $x \in [-A', A']^c$ имеем $1 - \prod_{k=1}^d \frac{\sin ax_k}{ax_k} > \frac{1}{2}$. Учитывая, что $|1 - \varphi_n(u)| \leq 2$ и $\varphi_n(u) \rightarrow \psi(u)$, получим, переходя в неравенстве (19) к пределу при $n \rightarrow \infty$,

$$\overline{\lim} F_n([-A', A']^c) \leq \frac{2}{(2\pi)^d} \int_{[-A, A]} (1 - \psi(u)) du. \quad (20)$$

Поскольку $\psi(0) = 1$ и $\psi(u)$ непрерывна в точке $u = 0$, то интеграл в правой части неравенства (20) стремится к 0 при $a \rightarrow 0$. Таким образом, для любого $\varepsilon > 0$ найдется такое $a > 0$, что для всех $n \geq n_0$ $F_n([-A', A']^c) < \varepsilon_n$. Из теоремы 6, § 4, следует слабая компактность последовательности распределений F_n , $n = 1, 2, \dots$. Пусть F_{n_j} , $j = 1, 2, \dots$ — какая-либо слабо сходящаяся подпоследовательность. Из первой части теоремы следует, что характеристические функции $\varphi_{n_j}(u)$ сходятся к $\psi(u)$ и $\psi(u)$ является характеристической функцией предельного распределения. Так как характеристическая функция однозначно определяет распределение, то любая слабо сходящаяся подпоследовательность распределений $\{F_{n_j}\}$ сходится к одному и тому же распределению F_0 . Отсюда вытекает, что вся последовательность F_n , $n = 1, 2, \dots$ слабо сходится к F_0 . Действительно, допустим противное. Тогда найдется ограниченная и непрерывная функция $f_0(x)$, число $\varepsilon > 0$ и подпоследовательность F_{n_j} такие, что

$$\left| \int_{R^d} f(x) F_{n_j}(dx) - \int_{R^d} f_0(x) F_0(dx) \right| > \varepsilon. \quad (21)$$

Из подпоследовательности F_{n_l} согласно слабой компактности последовательности F_n , $n = 1, 2, \dots$, можно выделить новую подпоследовательность F_{m_j} , слабо сходящуюся, по доказанному выше, к распределению F_0 . Но это противоречит неравенству (21). Таким образом, $F_n \Rightarrow F_0$.

§ 6. Многомерное нормальное распределение

Рассмотрим случайный вектор $\xi = (\xi_1, \xi_2, \dots, \xi_d)$, компоненты которого независимы и нормально распределены с параметрами (a_k^0, σ_k^2) , $\sigma_k > 0$, $k = 1, \dots, d$. Напомним, что плотность и характеристическая функция распределения нормальной случайной величины с параметрами (a, σ^2) имеют вид

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad \varphi(u) = e^{iau - \frac{1}{2}\sigma^2 u^2}.$$

Поэтому плотность распределения вектора ξ как совместная плотность распределения независимых случайных величин ξ_1, \dots, ξ_d равна

$$\begin{aligned} f(x_1, \dots, x_d) &= \prod_{k=1}^d \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{-\frac{1}{2} \frac{(x_k - a_k^0)^2}{\sigma_k^2}\right\} = \\ &= \frac{1}{(2\pi)^{\frac{1}{2}} \prod_{k=1}^d \sigma_k} \exp\left\{-\frac{1}{2} (D^{-1}(x - a^0), x - a^0)\right\}, \end{aligned} \quad (1)$$

где $a^0 = (a_1^0, \dots, a_d^0)$, D^{-1} — диагональная матрица с элементами σ_k^{-2} на диагонали. Характеристическая функция этого распределения

$$\begin{aligned} \varphi_\xi(u) &= M \exp\left\{i \sum_{k=1}^d u_k \xi_k\right\} = \prod_{k=1}^d \exp\left\{ia_k^0 u_k - \frac{\sigma_k^2 u_k^2}{2}\right\} = \\ &= \exp\left\{i(a^0, u) - \frac{1}{2}(Du, u)\right\}, \end{aligned}$$

где D — матрица, обратная D^{-1} , $D = \{\delta_{ik}\sigma_k^2\}$ ($\delta_{ik} = 1$ при $i = k$ и $\delta_{ik} = 0$ при $i \neq k$). Пусть A — произвольное линейное преобразование R^d . Поскольку в R^d фиксирован ортонормированный базис, то буква A будет также обозначать матрицу преобразования в данном базисе. Положим $\eta = A\xi$. Характеристическая функция $\varphi(u)$ распределения вектора η имеет вид

$$\varphi(u) = \varphi_\xi(A^*u) = \exp\left\{i(a, A^*u) - \frac{1}{2}(DA^*u, A^*u)\right\},$$

$$\varphi(u) = \exp \left\{ i(a, u) - \frac{1}{2} (Bu, u) \right\}, \quad (2)$$

где $a = Aa^0$, $B = ADA^*$. Заметим, что матрица B симметрична и неотрицательно определена

$$B^* = (ADA^*)^* = (A^*)^* D^* A^* = ADA^*, \\ (B^*u, u) = (ADA^*u, u) = (Dv, v) \geq 0.$$

Пусть дан произвольный вектор a и симметричная неотрицательно определенная матрица B . Как известно, существует ортогональная матрица U и диагональная матрица D с неотрицательными диагональными элементами, такие, что $B = UDU^*$. При этом диагональные элементы (обозначим их σ_k^2 , $k = 1, \dots, d$) матрицы D являются собственными числами матрицы B , а матрица U соответствует преобразованию данного базиса в базис, состоящий из собственных векторов матрицы B . Выражение $\exp \left\{ i(a, u) - \frac{1}{2} (Bu, u) \right\}$ можно рассматривать как характеристическую функцию случайного вектора $\eta = U\xi$, где ξ имеет характеристическую функцию (2) $a^0 = U^*a$. Если все $\sigma_k^2 > 0$, то вектор ξ имеет плотность распределения (1). Если некоторые $\sigma_k^2 = 0$, например, $\sigma_{s+1} = \sigma_{s+2} = \dots = \sigma_d = 0$, а $\sigma_k > 0$ при $k \leq s$, то компоненты (ξ_1, \dots, ξ_s) вектора ξ имеют совместную плотность распределения

$$f(x_1, \dots, x_s) = \prod_{k=1}^s \frac{1}{\sqrt{2\pi} \sigma_k} e^{-\frac{(x_k - a_k^0)^2}{2\sigma_k^2}},$$

а остальные компоненты являются константами: $\xi_k = a_k^0$ при $k = s+1, \dots, d$.

Определение. d -мерным нормальным (гауссовским) распределением называют распределение, характеристическая функция которого имеет вид (2).

Предыдущие рассуждения показывают, что данное определение имеет смысл, т. е. выражение (2) при любых $a \in R^d$ и симметрической неотрицательно определенной матрице B действительно является характеристической функцией некоторого распределения; если η — d -мерный нормальный (имеющий нормальное распределение) вектор, то всегда найдется ортогональное преобразование U такое, что вектор $\xi = U^{-1}\eta = U^*\eta$ (для ортогональной матрицы $U^{-1} = U^*$) имеет независимые гауссовские компоненты (константу будем рассматривать как гауссовскую случайную величину с дисперсией 0).

Предположим, что матрица B невырождена. Это будет тогда и только тогда, когда все ее собственные числа $\sigma_k^2 > 0$. Найдем явное

выражение для плотности нормального распределения. Пусть C — произвольное борелевское множество, $C \subset R^d$. Тогда

$$\begin{aligned} P(\eta \in C) &= P(U\xi \in C) = P(\xi \in U^{-1}C) = \\ &= \int_{U^{-1}C} \frac{1}{(2\pi)^{\frac{d}{2}} \prod_{k=1}^d \sigma_k} \exp \left\{ -\frac{1}{2} (D^{-1}(x - a^0), x - a^0) \right\} dx. \end{aligned}$$

Сделаем замену переменных $x = U^{-1}y = U^*y$. Поскольку U^* ортогональное преобразование, то $dx = dy$. Таким образом,

$$\begin{aligned} P(\eta \in C) &= \frac{1}{(2\pi)^{\frac{d}{2}} \prod_{k=1}^d \sigma_k} \int_C \exp \left\{ -\frac{1}{2} (D^{-1}(U^*y - U^*a), \right. \\ &\left. (U^*y - U^*a) \right\} dy = \frac{1}{(2\pi)^{\frac{d}{2}} \prod_{k=1}^d \sigma_k} \int_C \exp \left\{ -\frac{1}{2} (B^{-1}(y - a), y - a) \right\} dy, \end{aligned}$$

ибо $UD^{-1}U^* = (UDU^*)^{-1} = B^{-1}$. Далее, пусть $\text{Det } A$ — определитель матрицы A . Имеем $\text{Det } B = \text{Det } U \text{Det } D \text{Det } U^* = \text{Det } D = \prod_{k=1}^d \sigma_k^2$. Таким образом, для плотности распределения вектора η получаем следующее выражение:

$$f(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\text{Det } B}} \exp \left\{ -\frac{1}{2} (B^{-1}(x - a), x - a) \right\}. \quad (3)$$

Теорема 1. Для любого вектора $a \in R^d$ и неотрицательно определенной симметрической матрицы B функция $\varphi(u)$,

$$\varphi(u) = \exp \left\{ i(a, u) - \frac{1}{2} (Bu, u) \right\},$$

является характеристической функцией некоторого распределения.

Если матрица B невырождена, то это распределение имеет плотность распределения (3).

Если матрица B вырождена, то распределение сосредоточено в гиперплоскости, размерность которой равна рангу матрицы B .

Выясним теоретико-вероятностный смысл параметров, определяющих d -мерное нормальное распределение. Для этого найдем производные первого и второго порядков функции $\ln \varphi(u)$, которой в ряде случаев проще пользоваться, чем функцией $\varphi(u)$. Заметим, что

$$\frac{\partial \ln \varphi(u)}{\partial u_k} = \frac{1}{\varphi} \frac{\partial \varphi}{\partial u_k}, \quad \frac{\partial^2 \ln \varphi}{\partial u_k \partial u_r} = \frac{1}{\varphi} \frac{\partial^2 \varphi}{\partial u_k \partial u_r} - \frac{1}{\varphi^2} \frac{\partial \varphi}{\partial u_k} \frac{\partial \varphi}{\partial u_r}.$$

Следовательно,

$$\frac{\partial \ln \varphi(u)}{\partial u_k} \Big|_{u=0} = i c_k, \quad \frac{\partial^2 \ln \varphi}{\partial u_k \partial u_r} \Big|_{u=0} = -c_{kr} + c_k c_r,$$

где

$$c_k = M\eta_k, \quad c_{kr} = M\eta_k \eta_r,$$

$\eta = (\eta_1, \dots, \eta_d)$ — случайный вектор с характеристической функцией $\varphi(u)$. В рассматриваемом случае вычисления особенно просты, $\ln \varphi(u) = i(a, u) - \frac{1}{2}(Bu, u)$, так что

$$a_k = M\eta_k, \quad b_{kr} = M(\eta_k - c_k)(\eta_r - c_r) = c_{kr} - c_k c_r. \quad (4)$$

Таким образом, вектор $a = (a_1, \dots, a_d)$ — это вектор из математических ожиданий компонент вектора η , или среднее значение случайного вектора η , а элементы матрицы B равны вторым центральным моментам компонент вектора η . Ее называют корреляционной матрицей вектора η .

Введем коэффициенты корреляции ρ_{kr} между компонентами ξ_k и ξ_r вектора ξ и дисперсии σ_k^2 величин ξ_k . Тогда

$$b_{kr} = \rho_{kr} \sigma_k \sigma_r.$$

В качестве примера выпишем плотность двумерного распределения. Аргументы этой функции обозначим через x и y (вместо x_1 и x_2). Имеем $\text{Det } B = b_{11}b_{22} - b_{12}^2 = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$, где $\rho = \rho_{12}$,

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-a_1)^2}{\sigma_1^2} - \frac{2\rho(x-a_1)(y-a_2)}{\sigma_1\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2} \right] \right\}.$$

Остановимся на ряде свойств нормального распределения.

1) Если вектор $\eta = (\eta_1, \dots, \eta_d)$ имеет нормальное распределение, то «укороченный» вектор $\bar{\eta} = (\eta_1, \dots, \eta_s)$, $s < d$, также имеет нормальное распределение.

Действительно, если $\bar{u} = (u_1, \dots, u_s)$, $\bar{a} = (a_1, \dots, a_s)$, то

$$\begin{aligned} \varphi_{\bar{\eta}}(\bar{u}) &= \varphi_{\eta}(u_1, \dots, u_s) = \varphi_{\eta}(u_1, \dots, u_s, 0, \dots, 0) = \\ &= \exp \left\{ i \sum_1^s a_k u_k - \frac{1}{2} \sum_{k,j=1}^s b_{kj} u_k u_j \right\} = \exp \left\{ i(\bar{a}, \bar{u}) - \frac{1}{2}(\bar{B}\bar{u}, \bar{u}) \right\}, \end{aligned}$$

где \bar{B} — матрица с элементами b_{kj} , $k, j = 1, \dots, s$.

2) Если векторы $\eta = (\eta_1, \dots, \eta_d)$ и $\xi = (\xi_1, \dots, \xi_s)$ независимы и каждый из них имеет нормальное распределение, то составной вектор $\bar{\eta} = (\eta_1, \dots, \eta_d, \xi_1, \dots, \xi_s)$ также имеет нормальное распределение.

Действительно, пусть $\bar{u} = (u_1, \dots, u_d, u_{d+1}, \dots, u_{d+s})$, $u = (u_1, \dots, u_d)$, $v = (u_{d+1}, \dots, u_{d+s})$. Тогда характеристическая функция $\varphi_{\bar{\eta}}(\bar{u}) = \varphi_{\eta}(u) \varphi_{\zeta}(v)$ имеет вид

$$\varphi_{\bar{\eta}}(\bar{u}) = \exp \left\{ i \sum_{k=1}^{d+s} \bar{a}_k u_k - \frac{1}{2} \sum_{k,j=1}^{d+s} \bar{b}_{kj} u_k u_j \right\}$$

и, следовательно, является характеристической функцией нормального распределения.

3) Если случайный вектор $\xi = (\xi_1, \xi_2, \dots, \xi_d)$ имеет гауссовское распределение и случайные векторы $\xi' = (\xi_1, \dots, \xi_r)$, $\xi'' = (\xi_{r+1}, \dots, \xi_d)$ некоррелированы, то векторы ξ' и ξ'' независимы.

Доказательство. Из некоррелированности ξ' и ξ'' вытекает, что

$$b_{kj} = M \xi_k \xi_j - M \xi_k M \xi_j = 0, \quad k = 1, \dots, r; \quad j = r+1, \dots, d.$$

Поэтому

$$\begin{aligned} \varphi(u) = \exp \left\{ i \langle a', u' \rangle + i \langle a'', u'' \rangle - \frac{1}{2} \sum_{j,k=1}^r b_{kj} u_k u_j - \right. \\ \left. - \frac{1}{2} \sum_{j,k=r+1}^d b_{kj} u_k u_j \right\}, \end{aligned}$$

где $a' = (a_1, a_2, \dots, a_r)$, $a'' = (a_{r+1}, \dots, a_d)$, $u' = (u_1, \dots, u_r)$, $u'' = (u_{r+1}, \dots, u_d)$.

Из предыдущей формулы следует

$$\varphi(u) = M e^{i \langle u', \xi' \rangle + i \langle u'', \xi'' \rangle} = M e^{i \langle u', \xi' \rangle} M e^{i \langle u'', \xi'' \rangle} = \varphi'(u') \varphi''(u''),$$

где $\varphi'(u')$ и $\varphi''(u'')$ — характеристические функции векторов ξ' , ξ'' . Так как характеристическая функция однозначно определяет распределение, то векторы ξ' и ξ'' независимы.

Пусть $A = \|a_{jk}\|$, $j = 1, \dots, n$, $k = 1, \dots, d$ — произвольная прямоугольная матрица и $\eta = A\xi$, т. е. $\eta = (\eta_1, \dots, \eta_n)$,

$$\eta_j = \sum_{k=1}^d a_{jk} \xi_k, \quad j = 1, \dots, n.$$

Вектор η является линейным преобразованием вектора ξ .

4) При линейном преобразовании случайных векторов гауссовские распределения переходят в гауссовские.

Доказательство. Пусть $\varphi_{\eta}(u)$ — характеристическая функция вектора η . Тогда

$$\begin{aligned} \varphi_{\xi}(A'u) &= \exp \left\{ i \langle A'u, a \rangle - \frac{1}{2} \langle BA'u, A'u \rangle \right\} = \\ &= \exp \left\{ i \langle u, Aa \rangle - \frac{1}{2} \langle ABA'u, u \rangle \right\}, \end{aligned} \quad (5)$$

где A' — транспонированная матрица A . Таким образом, вектор η имеет гауссовское распределение с математическим ожиданием Aa и с корреляционной матрицей $B_\eta = ABA'$.

Пусть ξ и η два случайных вектора с совместным нормальным распределением. Найдем условную плотность распределения $f_{\xi|\eta}(x)$. Она существует и является отношением двух гауссовских плотностей. Непосредственное вычисление условной плотности связано с громоздкими вычислениями и мы воспользуемся другим способом.

Пусть $a_\xi = M\xi$, $a_\eta = M\eta$, R_η — корреляционная матрица вектора η , $R_{\xi\eta}$ и $R_{\eta\xi}$ — взаимные корреляционные матрицы соответствующих пар векторов. При этом $R'_{\xi\eta} = R_{\eta\xi}$. Предположим, что матрица R_η не вырождена. Тогда существует R_η^{-1} .

Положим

$$\tilde{\xi} = \xi - a_\xi - R_{\xi\eta}R_\eta^{-1}(\eta - a_\eta).$$

Имеем

$$M\tilde{\xi} = 0. \quad M\tilde{\xi}(\eta - a_\eta)' = R_{\xi\eta} - R_{\xi\eta}R_\eta^{-1}R_\eta = 0.$$

Таким образом, векторы $\tilde{\xi}$ и η некоррелированы и при этом имеют совместное нормальное распределение, так как являются линейными комбинациями векторов ξ и η . Следовательно, $\tilde{\xi}$ и η — независимы. Корреляционная матрица $R_{\tilde{\xi}\tilde{\xi}}$ равна

$$\begin{aligned} R_{\tilde{\xi}\tilde{\xi}} = M\tilde{\xi}\tilde{\xi}' &= R_\xi - R_{\xi\eta}R_\eta^{-1}R_{\eta\xi} - R_{\xi\eta}R_\eta^{-1}R'_{\xi\eta} + \\ &+ R_{\xi\eta}R_\eta^{-1}R_\eta R_\eta^{-1}R'_{\xi\eta} = R_\xi - R_{\xi\eta}R_\eta^{-1}R_{\eta\xi}. \end{aligned} \quad (6)$$

При этом использованы равенства $R'_{\xi\eta} = R_{\eta\xi}$, $(R_\eta^{-1})' = R_\eta^{-1}$. Найдем условную характеристическую функцию

$$\varphi_{\xi|\eta}(u) \stackrel{\text{Def}}{=} M(e^{i(u, \xi)} | \eta).$$

Поскольку существует условная плотность распределения, то $\varphi_{\xi|\eta}(u)$ является характеристической функцией этой плотности

$$\varphi_{\xi|\eta}(u) = \int e^{i(u, x)} f_{\xi|\eta}(x) dx.$$

Так как $\xi = \zeta + \tilde{\xi}$, где $\zeta = a_\xi + R_{\xi\eta}R_\eta^{-1}(\eta - a_\eta)$ и $\tilde{\xi}$ не зависит от η , то

$$\begin{aligned} \varphi_{\xi|\eta}(u) &= M(e^{i(u, \zeta)} e^{i(u, \tilde{\xi})} | \eta) = e^{i(u, \zeta)} M(e^{i(u, \tilde{\xi})} | \eta) = \\ &= e^{i(u, \zeta)} M e^{i(u, \tilde{\xi})}. \end{aligned}$$

В силу того что вектор $\tilde{\xi}$ имеет гауссовское распределение со средним 0,

$$\varphi_{\xi|\eta}(u) = \exp \left\{ i(u, a_\xi + R_{\xi\eta}R_\eta^{-1}(\eta - a_\eta)) - \frac{1}{2} (R_{\tilde{\xi}\tilde{\xi}} u, u) \right\}. \quad (7)$$

Из формул (6) и (7) вытекает следующая теорема.

Теорема 2. Если совместные распределение векторов ξ и η гауссовское, то условная плотность распределения вектора ξ при данном η также является гауссовским со средним

$$M(\xi | \eta) = a_\xi + R_{\xi\eta} R_\eta^{-1} (\eta - a_\eta)$$

и корреляционной матрицей $R_\xi - R_{\xi\eta} R_\eta^{-1} R_{\eta\xi}$.

Замечание. По теореме 2 получаем, что условное среднее $M(\xi | \eta)$ является линейной функцией η , а условная корреляционная матрица вообще от η не зависит.

В случае, когда вектор ξ одномерный, то полученным формулам для условного среднего и условной дисперсии можно придать следующий вид. Пусть R — корреляционная матрица составного вектора (ξ, η) размерности $n+1$ с элементами b_{kj} , $k, j = 0, 1, \dots, n$. Тогда $b_{kj} = (R_\eta)_{kj}$ при $k, j = 1, 2, \dots, n$ и $b_{0k} = M(\xi - a_\xi)(\eta_k - a_{\eta_k})$. Алгебраические дополнения матрицы R обозначим через c_{kj} ; $k, j = 0, 1, \dots, n$. Произведение $R_{\xi\eta} R_\eta^{-1}$

является вектор-строкой с элементами $\sum_{j=1}^n b_{0j} \frac{c_{jk}}{c_{00}}$, $k = 1, \dots, n$, где c_{jk} — алгебраические дополнения элементов матрицы R_η .

Заметим, что $\sum_{j=1}^n b_{0j} c_{jk} = -c_{0k}$. Таким образом,

$$M(\xi | \eta) = a_\xi - \sum_{k=1}^n \frac{c_{0k}}{c_{00}} (\eta_k - a_{\eta_k}),$$

или

$$M(\xi | \eta) = a_\xi + \sum_{k=1}^n \rho_{0k} (\eta_k - a_{\eta_k}), \quad \rho_{0k} = -\frac{c_{0k}}{c_{00}}. \quad (8)$$

Величины ρ_{0k} называют коэффициентами регрессии.

Для условной дисперсии $D(\xi | \eta)$ величины ξ получим следующее выражение (пусть $R_\xi = D\xi = \sigma^2 = b_{00}$):

$$D(\xi | \eta) = R_\xi + \sum_{k=1}^n \frac{c_{0k}}{c_{00}} b_{k0},$$

или

$$D(\xi | \eta) = \frac{c}{c_{00}}, \quad c = \text{Det } R. \quad (9)$$

Для условной плотности $f_{\xi|\eta}(x)$ получаем такое выражение:

$$f_{\xi|\eta}(x) = \sqrt{\frac{c_{00}}{2\pi c}} \exp \left\{ -\frac{c_{00}}{2c} (x - M(\xi | \eta))^2 \right\}, \quad (10)$$

где $M(\xi | \eta)$ задается формулой (8).

§ 7. Элементы гармонического анализа

Рассмотрим ряд вопросов, связанных с теорией преобразований Фурье. При этом нас будут интересовать преобразования Фурье не только плотностей распределения, но и произвольных интегрируемых (и не только интегрируемых) функций. Чтобы получить такие обобщения, часто бывает достаточным представить интегрируемую функцию $f(x)$ в виде линейной комбинации плотностей $f_i(x)$, $i = 1, 2$,

$$f(x) = \alpha f_1(x) + \beta f_2(x),$$

где $f_1(x) = \alpha^{-1} f^+(x)$, $f_2(x) = -\beta^{-1} f^-(x)$, $\alpha = \int_{\mathbb{R}^d} f^+(x) dx$, $\beta = \int_{\mathbb{R}^d} f^-(x) dx$.

Кроме того, когда оперируют с распределениями случайных величин, аналитические факты получают наглядную теоретико-вероятностную интерпретацию. Например, свертка двух интегрируемых функций f и g в том случае, когда f и g являются плотностями распределений, дает плотность распределения суммы независимых случайных векторов с плотностями распределений f и g . Поскольку характеристическая функция суммы независимых случайных векторов равна произведению характеристических функций слагаемых, то преобразование Фурье свертки двух интегрируемых функций равно произведению преобразований Фурье. В этом можно убедиться и непосредственно. Пусть $h(x) = f \times g(x)$, $f, g \in L_1$. Поскольку функция $f(x-y)g(y)$ интегрируема как функция переменных x и y , то по теореме Фубини

$$\begin{aligned} \int_{\mathbb{R}^d} e^{i(u, x)} h(x) dx &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{i(u, x-y) + i(u, y)} f(x-y) g(y) dy dx = \\ &= \int_{\mathbb{R}^d} e^{i(u, y)} g(y) dy \int_{\mathbb{R}^d} e^{i(u, x-y)} f(x-y) dx = \\ &= \int_{\mathbb{R}^d} e^{i(u, y)} g(y) dy \int_{\mathbb{R}^d} e^{i(u, t)} f(t) dt = \varphi(u) \psi(u), \end{aligned}$$

где $\varphi(u)$, $\psi(u)$ — преобразования Фурье функций $f(x)$, $g(x)$. Сформулируем полученный результат.

Теорема 1. Преобразование Фурье свертки двух интегрируемых функций равно произведению преобразований Фурье этих функций.

Пусть $\varphi(u)$ — характеристическая функция случайного вектора ξ с распределением $F(x)$. В равенстве

$$e^{-i(u, y)} \varphi_\xi(u) = M e^{i(u, \xi - y)}$$

вместо u поставим η , где η — случайный вектор, не зависящий от ξ , с распределением $G(x)$. Вычисляя математическое ожидание

от обеих частей равенства и воспользовавшись формулой 3, § 2, получим

$$Me^{-i(y, \eta)} \varphi(\eta) = M(Me^{i(u, \xi-y)})|_{u=\eta} = Me^{i(\eta, \xi-y)} = \\ = M(e^{i(\eta, x-y)})|_{x=\xi},$$

или

$$Me^{-i(y, \eta)} \varphi(\eta) = M\psi(\xi - y), \quad (1)$$

где $\psi(u)$ — характеристическая функция вектора η .

Отметим частный случай полученного равенства, носящий название формулы умножения. Предположим, что ξ и η обладают плотностями $f(x)$ и $g(x)$ соответственно. Формула (1) запишется в виде

$$\int_{R^d} e^{-i(y, u)} \varphi(u) g(u) du = \int_{R^d} \psi(u - y) f(u) du. \quad (2)$$

Полагая здесь $y = 0$, имеем

$$\int_{R^d} \varphi(u) g(u) du = \int_{R^d} f(u) \psi(u) du. \quad (3)$$

Класс функций f и g , для которых формула (3) справедлива, является линейным. При этом φ , ψ обозначают преобразование Фурье функции $f(x)$ и $g(x)$. Таким образом, формула (3) выполняется для произвольных интегрируемых функций f и g . При этом $\varphi(u)$ и $\psi(u)$ — преобразования Фурье функций f и g . Подставим в формулу (2) в качестве $g(u)$ плотность распределения случайной величины вида $e\eta$, где η — d -мерное стандартное гауссовское распределение или распределение с плотностью $[c_d(2\pi)^d]^{-1} \times \exp(-\|u\|)$. Заметим, что если $g_0(x)$ — плотность вектора η , то g_ε , где $g_\varepsilon(x) = \varepsilon^{-d} g_0\left(\frac{x}{\varepsilon}\right)$ — плотность распределения вектора $e\eta$, $\varepsilon > 0$. Получим

$$\frac{1}{(2\pi)^d} \int_{R^d} e^{-i(y, u)} e^{-\frac{\varepsilon^2 |u|^2}{2}} \varphi(u) du = \frac{1}{(\sqrt{2\pi} \varepsilon)^d} \int_{R^d} e^{-\frac{|y-u|^2}{2\varepsilon^2}} f(u) du \quad (4)$$

и

$$\frac{1}{(2\pi)^d} \int_{R^d} e^{-i(y, u)} e^{-\varepsilon |u|} \varphi(u) du = c_d \int_{R^d} \frac{\varepsilon f(u) du}{(\varepsilon^2 + |y-u|^2)^{\frac{d+1}{2}}}. \quad (5)$$

В правых частях полученных формул стоят свертки плотностей распределений $g_\varepsilon(x)$ и $f(x)$. В общем случае, когда существова-

ние плотности распределения вектора ξ не предполагается, те же подстановки в формуле (1) приводят к равенствам

$$M \frac{1}{(V 2\pi e)^d} e^{-\frac{|\xi - y|^2}{2e^2}} = \frac{1}{(2\pi)^d} \int_{R^d} e^{-i(y, u)} e^{-\frac{\varepsilon^2 |u|^2}{2}} \varphi(u) du, \quad (6)$$

$$M \frac{c_d e^{\frac{c_d^2}{2}}}{(e^2 + |y - \xi|^2)^{\frac{d+1}{2}}} = \frac{1}{(2\pi)^d} \int_{R^d} e^{-i(y, u)} e^{-\varepsilon |u|} \varphi(u) du. \quad (7)$$

В левых частях этих равенств стоит плотность распределения вектора $\xi + \varepsilon \eta$, где η имеет одно из указанных выше распределений.

Теорема 2. Свертка $F \star G$ двух распределений является слабо непрерывной операцией.

Под слабой непрерывностью операции $F \star G$ следует понимать, что если $G = G_n$ слабо сходится при $n \rightarrow \infty$ к G_0 , то распределение $F \star G_n$ слабо сходится к $F \star G_0$.

Доказательство. Пусть $h(x)$ — произвольная ограниченная непрерывная функция, F — распределение случайного вектора ξ , G_n — распределение вектора γ_n , $n = 0, 1, 2, \dots$ и $H_n = F \star G_n$. Тогда

$$\int_{R^d} h(x) H_n(dx) = Mh(\xi + \gamma_n) = Mh_1(\gamma_n),$$

где $h_1(x) = Mh(\xi + x)$. Имеем $|h_1(x)| \leq \sup |h(x)|$ и при $x_n \rightarrow x$ $h_1(x_n) = Mh(\xi + x_n) \rightarrow h_1(x)$ на основании того, что $h(\xi + x_n) \rightarrow h(\xi + x)$ при любом ξ и теоремы о мажорируемой сходимости. Итак, $h_1(x)$ — непрерывная ограниченная функция. Поэтому $Mh_1(\gamma_n) \rightarrow Mh_1(\gamma_0)$, т. е.

$$\int_{R^d} h(x) H_n(dx) \rightarrow \int_{R^d} h(x) H_0(dx).$$

Применим доказанную теорему к равенствам (6) и (7). Так как $\varepsilon \eta \rightarrow 0$ по вероятности при $\varepsilon \rightarrow 0$, то распределение суммы $\xi + \varepsilon \eta$ слабо сходится к распределению $F(\cdot)$ вектора ξ . Мы получили следующую теорему:

Теорема 3. Функции

$$f_\varepsilon(y) = \frac{1}{(2\pi)^d} \int_{R^d} e^{-i(y, u)} e^{-\frac{\varepsilon^2 |u|^2}{2}} \varphi(u) du,$$

$$\bar{f}_\varepsilon(y) = \frac{1}{(2\pi)^d} \int_{R^d} e^{-i(y, u)} e^{-\varepsilon |u|} \varphi(u) du$$

являются плотностями распределений, которые при $\varepsilon \rightarrow 0$ слабо сходятся к $F(\cdot)$. В частности, если B множество непрерывности распределения $F(\cdot)$, то

$$F(B) = \lim_{\varepsilon \rightarrow 0} \frac{1}{(2\pi)^d} \int_B \left(\int_{\mathbb{R}^d} e^{-i(y, u)} - \frac{\varepsilon^2 |u|^2}{2} \varphi(u) du \right) dy, \quad (8)$$

$$F(B) = \lim_{\varepsilon \rightarrow 0} \frac{1}{(2\pi)^d} \int_B \left(\int_{\mathbb{R}^d} e^{-i(y, u)} - \varepsilon |u| \varphi(u) du \right) dy. \quad (9)$$

Мы получили две формулы, позволяющие определить распределение случайного вектора по его характеристической функции.

Следствие. Если характеристическая функция интегрируема, то распределение случайного вектора ξ обладает непрерывной плотностью распределения $f(y)$ и

$$f(y) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(y, u)} \varphi(u) du. \quad (10)$$

Последняя формула является формулой обращения преобразования Фурье. В рассматриваемом частном случае она следует из того, что когда функция $\varphi(u)$ из равенства (8) интегрируема, то можно перейти к пределу при $\varepsilon \rightarrow 0$ под знаком интеграла. Получим

$$F(B) = \int_B \left(\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(y, u)} \varphi(u) du \right) dy,$$

откуда и следует (10). При этом интеграл в правой части равенства (10) является непрерывной функцией от y .

Покажем, что соотношение (10) однозначно определяет характеристическую функцию.

Теорема 4. Если $h(u)$ — непрерывная ограниченная и интегрируемая функция и

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(x, u)} h(u) du,$$

где $f(x)$ — плотность распределения, то $h(u) = \varphi(u)$.

Доказательство. Воспользуемся формулой (1), положив в ней $-y$ вместо y , $\xi = \varepsilon\eta$, где η — стандартный гауссовский случайный вектор и пусть η имеет плотность распределения $f(x)$. Получим

$$\begin{aligned} M\varphi(\varepsilon\eta + y) &= M \exp \left\{ i(y, \eta) - \frac{\varepsilon^2}{2} |\eta|^2 \right\} = \\ &= \int_{\mathbb{R}^d} \exp \left\{ i(y, x) - \frac{\varepsilon^2}{2} |x|^2 \right\} f(x) dx = \end{aligned}$$

$$= \frac{1}{(2\pi)^d} \int_{R^d} \exp \left\{ i(y, x) - \frac{\varepsilon^2}{2} |x|^2 \right\} dx \int_{R^d} e^{-i(u, x)} h(u) du.$$

Так как под знаком интеграла в правой части равенства стоят абсолютно интегрируемые функции, то в ней можно изменить порядок интегрирования. Следовательно,

$$\begin{aligned} M\psi(\varepsilon y + y) &= \int_{R^d} h(u) du \int_{R^d} \frac{1}{(2\pi)^d} \exp \left\{ -i(y - u, x) - \frac{\varepsilon^2}{2} |x|^2 \right\} dx = \\ &= \int_{R^d} \frac{1}{(2\pi)^{\frac{d}{2}} \varepsilon^d} h(u) e^{-\frac{|y-u|^2}{2\varepsilon^2}} du = \int_{R^d} \frac{1}{(2\pi)^{\frac{d}{2}} \varepsilon^d} h(u + y) e^{-\frac{|u|^2}{2\varepsilon^2}} du = \\ &= Mh(\varepsilon y + y). \end{aligned}$$

Перейдем к пределу при $\varepsilon \rightarrow 0$. Поскольку $\varepsilon y \rightarrow 0$, то $Mh(\varepsilon y + y) \rightarrow h(y)$ и $M\psi(\varepsilon y + y) \rightarrow \psi(y)$, откуда $h(y) = \psi(y)$, $\forall y \in R^d$.

Можно получить формулы обращения преобразования Фурье плотностей распределения и без предположения об абсолютной интегрируемости характеристических функций. Приведем две теоремы, в одной из которых речь идет о сходимости в L_1 , а в другой — о поточечной сходимости. Эта теорема является простым следствием формул (4) и (5) и следующей теоремы:

Теорема 5. Пусть $g(\cdot) \in L_1(R^d)$, $f(\cdot) \in L_p(R^d)$, причем $\int_{R^d} g(x) dx = 1$, $p \geq 1$.

Тогда $\|g_\varepsilon * f - f\|_p \rightarrow 0$ при $\varepsilon \rightarrow 0$, где $g_\varepsilon(x) = \varepsilon^{-d} g\left(\frac{x}{\varepsilon}\right)$.

Доказательство. Проверим, что $g_\varepsilon * f$ имеет смысл и $g_\varepsilon * f \in L_p$. Заметим, что

$$\left\| \int_{R^d} f(x, y) g(y) dy \right\|_p \leq \int_{R^d} \|f(\cdot, y)\|_p |g(y)| dy. \quad (11)$$

Если функция $f(x, y)$ имеет вид $f(x, y) = \sum f_k(x) h_k(y)$, то неравенство (11) вытекает из неравенства Минковского

$$\left\| \sum c_k f_k \right\|_p \leq \sum c_k \|f_k\|_p, \quad c_k > 0.$$

В общем случае доказательство неравенства (11) получим с помощью предельного перехода. Применим (11) к свертке $h(x) = f * g$. Тогда

$$\|h\|_p \leq \int_{R^d} \left(\int_{R^d} |f(x-y)|^p dx \right)^{\frac{1}{p}} |g(y)| dy,$$

или

$$\|f \times g\|_p \leq \|f\|_p \|g(\cdot)\|_1.$$

Таким образом, $f \times g_\varepsilon \in L_p$. Поскольку $\int_{R^d} g_\varepsilon(x) dx = 1$, то, снова воспользовавшись неравенством (11), получим

$$\begin{aligned} \|f \times g_\varepsilon - f\|_p &\leq \int_{R^d} \left(\int_{R^d} |f(x-y) - f(x)|^p dx \right)^{\frac{1}{p}} |g_\varepsilon(y)| dy = \\ &= \int_{R^d} \left(\int_{R^d} |f(x-\varepsilon y) - f(x)|^p dx \right)^{\frac{1}{p}} g(y) dy. \end{aligned}$$

Величину $\omega_p(h, f) = \left(\int_{R^d} |f(x+h) - f(x)|^p dx \right)^{\frac{1}{p}}$ называют L_p -модулем непрерывности функции $f(x)$. Нетрудно заметить, что для любой функции $f \in L_p$ $\omega_p(h, f) \rightarrow 0$ при $h \rightarrow 0$. Это очевидно для любой непрерывной функции $f(x)$, обращающейся в нуль вне некоторого ограниченного множества. Так как произвольную функцию из L_p можно аппроксимировать последовательностью функций, равных нулю вне некоторого компакта, то $\omega_p(h, f) \rightarrow 0$ при $h \rightarrow 0$ и для любых $f \in L_p$. Мы получили неравенство

$$\|f \times g_\varepsilon - f\|_p \leq \int_{R^d} \omega_p(\varepsilon, y) |g(y)| dy,$$

где $\omega_p(\varepsilon, y) \rightarrow 0$ при $\varepsilon \rightarrow 0$, $\forall y \in R^d$ и $\omega_p(\varepsilon, y) \leq 2\|f\|_p$. По теореме о мажорируемой сходимости $\|f \times g_\varepsilon - f\|_p \rightarrow 0$.

Теорема 6. Для каждой $f \in L_1$

$$f(x) = \lim_{\varepsilon \rightarrow 0} \int_{R^d} \lambda(\varepsilon u) e^{-i(x, u)} \varphi(u) du$$

в смысле сходимости в L_1 . Здесь $\lambda(u) = e^{-|u|}$ или $\lambda(u) = e^{-|u|^2}$.

Как уже отмечалось, эта теорема непосредственно вытекает из равенств (4) — (5) и теоремы 5. При этом следует иметь в виду, что формулы (4) и (5) справедливы для произвольных интегрируемых функций. Действительно, если $f \in L_1$, то $f = \alpha_1 f_1 + \alpha_2 f_2$, где f_1 и f_2 — некоторые плотности распределения. Так как формулы (4) и (5) выполняются для f_1 и f_2 , то они выполняются и для f .

Для доказательства теоремы о поточечной обратимости преобразования Фурье понадобится понятие точки Лебега.

Определение. Пусть $f(x)$, $x \in R^d$, измеримая (по Лебегу) функция, интегрируемая на произвольном ограниченном мно-

жесткие (локально интегрируемая). Точка x_0 называется точкой Лебега функции $f(\cdot)$, если

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{V_\varepsilon} \int_{|y| < \varepsilon} |f(x-y) - f(x)| dy = 0,$$

где V_ε — объем d -мерного шара радиуса ε .

Очевидно, каждая точка непрерывности функции является ее точкой Лебега.

Теорема 7. Множество точек, не являющихся точками Лебега локально интегрируемой функции $f(x)$, имеет меру нуль.

Доказательство. Из теории функций известно, что l — почти для всех x (l — мера Лебега в R^d)

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{V_\varepsilon} \int_{|y| < \varepsilon} f(x-y) dy = f(x).$$

Пусть E_ρ обозначает то множество точек $x \in R^d$, для которых не выполняется следующее равенство:

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{V_\varepsilon} \int_{|y| < \varepsilon} |f(x-y) - \rho| dy = |f(x) - \rho|.$$

Из сказанного выше следует, что $l(E_\rho) = 0$. Положим $E = \bigcup E_\rho$, где сумма \bigcup распространена на все рациональные $\rho > 0$. Тогда $l(E) = 0$. Покажем, что если $x \notin E$, то x — точка Лебега функции $f(\cdot)$. Действительно, возьмем произвольно малое $\delta > 0$ и выберем рациональное ρ так, чтобы $|f(x) - \rho| < \frac{\delta}{2}$. Тогда для всех достаточно малых $\varepsilon > 0$,

$$\begin{aligned} \frac{1}{V_\varepsilon} \int_{|y| < \varepsilon} |f(x-y) - f(x)| dy &\leq \frac{1}{V_\varepsilon} \int_{|y| < \varepsilon} (|f(x-y) - \rho| + \\ &+ |\rho - f(x)|) dy \leq \frac{2\delta}{3} < \delta, \end{aligned}$$

так как при $\varepsilon \rightarrow 0$ на основании того, что $x \notin E$

$$\frac{1}{V_\varepsilon} \int_{|y| < \varepsilon} |f(x-y) - \rho| dx \rightarrow |f(x) - \rho| < \frac{\delta}{3}.$$

Таким образом,

$$V_\varepsilon^{-1} \int_{|y| < \varepsilon} |f(x-y) - f(x)| dy \rightarrow 0 \text{ при } \varepsilon \rightarrow 0.$$

Теорема 8. Пусть $f(x) \in L_1$, а функция $g(x)$ удовлетворяет следующим условиям:

$$g(x) \in L_1, \quad \overline{g}(x) \stackrel{\text{Def}}{=} \sup_{|y| > |x|} |g(y)| \in L_1.$$

Положим $g_\varepsilon(x) = \varepsilon^{-d} g\left(\frac{x}{\varepsilon}\right)$. Тогда для любой точки Лебега функции $f(x)$

$$\lim_{\varepsilon \rightarrow 0} f \times g_\varepsilon(x) = f(x).$$

Доказательство. Пусть x — точка Лебега функции $f(x)$ и $\delta > 0$. Найдем такое ε_0 , чтобы неравенство

$$V_\varepsilon^{-1} \int_{|y| < \varepsilon} |f(x-y) - f(x)| dx < \delta \quad (12)$$

выполнялось для всех $\varepsilon \in (0, \varepsilon_0]$. Имеем

$$|(f \times g_\varepsilon)(x) - f(x)| \leq J_1 + J_2,$$

где

$$J_1 = \int_{|y| < \varepsilon_0} |f(x-y) - f(x)| g_\varepsilon(y) dy,$$

$$J_2 = \int_{|y| > \varepsilon_0} |f(x-y) - f(x)| g_\varepsilon(y) dy.$$

Оценим интегралы J_1 и J_2 . Заметим сначала, что $\bar{g}(x) = \bar{g}_0(|x|)$, где $\bar{g}_0(r)$, $r \geq 0$, монотонно невозрастающая функция. Так как

$$2^{-d} v_1 r^d \bar{g}_0(r) \leq \int_{\frac{r}{2} < |x| < r} \bar{g}(x) dx \rightarrow 0$$

(v_1 — объем единичного шара) при $r \rightarrow 0$ или при $r \rightarrow \infty$, то функция $r^d \bar{g}_0(r) \rightarrow 0$ при $r \rightarrow 0$ или $r \rightarrow \infty$.

Обозначим через $\bar{f}(r)$ следующий интеграл по поверхности сферы в R^d радиуса 1

$$\bar{f}(r) = \int_{|t|=1} |f(x-rt) - f(x)| dS,$$

где dS — элемент площади поверхности единичной сферы. Тогда

$$\begin{aligned} J_1 &= \int_{|y| < \varepsilon_0} |f(x-y) - f(x)| g_\varepsilon(y) dy \leq \\ &\leq \int_0^{\varepsilon_0} dr \int_{|t|=1} |f(x-rt) - f(x)| \varepsilon^{-d} \bar{g}_0\left(\frac{r}{\varepsilon}\right) dS r^{d-1} dr = \\ &= \int_0^{\varepsilon_0} r^{d-1} \bar{f}(r) \varepsilon^{-d} \bar{g}_0\left(\frac{r}{\varepsilon}\right) dr. \end{aligned}$$

Положим

$$\mathfrak{F}(a) = \int_0^a r^{d-1} \bar{f}(r) dr.$$

Заметим, что неравенство (12) эквивалентно неравенству

$$\mathfrak{F}(a) \leq \delta v_1 a^d, \quad \forall a \leq \varepsilon_0.$$

Имеем теперь

$$\begin{aligned} J_1 &\leq \mathfrak{F}(r) \varepsilon^{-d} \bar{g}_0\left(\frac{r}{\varepsilon}\right) \Big|_{\varepsilon}^{\varepsilon_0} - \int_0^{\varepsilon_0} \mathfrak{F}(r) d\left(\varepsilon^{-d} \bar{g}_0\left(\frac{r}{\varepsilon}\right)\right) \leq \\ &\leq \delta v_1 \left(\frac{\varepsilon_0}{\varepsilon}\right)^d \bar{g}_0\left(\frac{\varepsilon_0}{\varepsilon}\right) - \int_0^{\frac{\varepsilon_0}{\varepsilon}} \mathfrak{F}(\varepsilon r) \varepsilon^{-d} d\bar{g}_0(r) \leq \\ &\leq \delta A - \int_0^{\frac{\varepsilon_0}{\varepsilon}} \delta v_1 r^d d\bar{g}_0(r) \leq \delta \left(A - v_1 \int_0^{\infty} r^d d\bar{g}_0(r) \right), \end{aligned}$$

где A — некоторая константа. Далее,

$$- \int_0^{\infty} r^d d\bar{g}_0(r) = d \int_0^{\infty} r^{d-1} g_0(r) dr = \frac{d}{|S_1|} \int_{R^d} \bar{g}(x) dx,$$

где $|S_1|$ — площадь единичной сферы в R^d . Следовательно, существует постоянная A_1 , не зависящая от ε и ε_0 , такая, что $|J_1| \leq \leq \delta A_1$ при всех $\varepsilon \in (0, \varepsilon_0]$. Оценка интеграла J_2 получается проще. Имеем

$$\begin{aligned} J_2 &\leq \int_{|t| > \frac{\varepsilon_0}{\varepsilon}} |f(x - \varepsilon t) - f(x)| g(t) dt \leq \bar{g}_0\left(\frac{\varepsilon_0}{\varepsilon}\right) \times \\ &\times \int_{R^d} |f(x - y) - f(x)| dy \leq \left(\int_{R^d} |f(y)| dy + |f(x)| \right) \bar{g}_0\left(\frac{\varepsilon_0}{\varepsilon}\right), \end{aligned}$$

причем $\bar{g}_0\left(\frac{\varepsilon_0}{\varepsilon}\right) \rightarrow 0$ при $\varepsilon \rightarrow 0$. Следовательно, если ε_0 уже выбрано (скажем, из требования $|J_1| < \frac{\delta_1}{2}$, $\delta_1 > 0$), то при достаточно малом $\varepsilon' = \varepsilon'(\varepsilon_0, \delta_1)$ и $\varepsilon < \varepsilon'$ имеем $J_2 < \frac{\delta_1}{2}$, т. е. $|f \times g_\varepsilon(x) - f(x)| < \delta_1$ при $\varepsilon < \varepsilon'$.

Следствие 1. Если x — точка Лебега интегрируемой функции $f(x)$, $\varphi(u)$ — преобразование Фурье $f(x)$, то

$$f(x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{(2\pi)^d} \int_{R^d} \lambda(\varepsilon u) e^{-i(x, u)} \varphi(u) du, \quad (13)$$

где $\lambda(\varepsilon u) = e^{-\varepsilon|u|}$ или $\lambda(\varepsilon u) = e^{-\frac{\varepsilon|u|^2}{2}}$.

Отметим одно следствие общей формулы (13). Пусть $f(x)$ непрерывна в точке $x=0$. Тогда

$$f(0) = \lim_{\varepsilon \rightarrow 0} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \lambda(\varepsilon u) \varphi(u) du.$$

Предположим, что $\varphi(u) \geq 0$. Из леммы Фату тогда следует

$$\int_{\mathbb{R}^d} \varphi(u) du \leq \liminf_{\varepsilon \rightarrow 0} \int_{\mathbb{R}^d} \lambda(\varepsilon u) \varphi(u) du < \infty.$$

Итак, функция $\varphi(\cdot) \in L_1$. Отсюда вытекает следствие.

Следствие 2. Пусть $f \in L_1$ и $\varphi \geq 0$. Если функция $f(x)$ непрерывна при $x=0$, то $\varphi(\cdot) \in L_1$ и

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(x, u)} \varphi(u) du \quad (14)$$

почти для всех x . В частности, равенство (14) выполняется в каждой точке непрерывности функции $f(x)$.

Действительно, поскольку $\varphi(\cdot) \in L_1$, то по теореме о мажорируемой сходимости в равенстве (13) можно перейти к пределу при $\varepsilon \rightarrow 0$. При этом $\lim_{\varepsilon \rightarrow 0} \lambda(\varepsilon u) = 1$.

Определение преобразования Фурье с помощью интеграла (2) неприменимо к функциям $f(x) \in L_2$, поскольку функция $e^{i(u, x)} f(x)$, вообще говоря, неинтегрируема. Тем не менее, и в этом случае преобразование Фурье может быть естественным образом определено. Это определение основано на следующей теореме:

Теорема 9. Если $f(x) \in L_1 \cap L_2$, то

$$\int_{\mathbb{R}^d} |\varphi(u)|^2 du = (2\pi)^d \int_{\mathbb{R}^d} |f(x)|^2 dx, \quad (15)$$

где $\varphi(u)$ — преобразование Фурье функции $f(x)$.

Доказательство. Пусть $g(x) = \bar{f}(-x)$, $\psi(u)$ — преобразование Фурье функции $g(x)$. Тогда

$$\psi(u) = \int_{\mathbb{R}^d} e^{i(u, x)} \bar{f}(-x) dx = \int_{\mathbb{R}^d} e^{-i(u, x)} \bar{f}(x) dx = \overline{\varphi(u)}.$$

Положим $h(u) = f \times g(x)$, $\gamma(u)$ — преобразование Фурье $h(x)$. По теореме 1 $\gamma(u) = |\varphi(u)|^2 \geq 0$. Покажем, что функция $h(x)$ непрерывна. Действительно,

$$\begin{aligned} |h(x + \Delta x) - h(x)|^2 &= \left| \int_{\mathbb{R}^d} [f(x + \Delta x - y) - f(x - y)] \times \right. \\ &\times g(y) dy \left. \right|^2 \leq \int_{\mathbb{R}^d} |f(x + \Delta x - y) - f(x - y)|^2 dy \int_{\mathbb{R}^d} |g(y)|^2 dy = \\ &= \int_{\mathbb{R}^d} |g(y)|^2 dy \int_{\mathbb{R}^d} |f(x + \Delta x) - f(x)|^2 dx. \end{aligned}$$

Уже отмечалось, что модуль непрерывности $\omega_2(\Delta x, f) \rightarrow 0$ при $\Delta x \rightarrow 0$, $f \in L_2$ (см. доказательство теоремы 5). Таким образом, функция $h(x)$ равномерно непрерывна. Из следствия 2 теоремы 8 вытекает, что $\gamma(u)$ — интегрируемая функция и

$$h(0) = \frac{1}{(2\pi)^d} \int_{R^d} \gamma(u) du = \frac{1}{(2\pi)^d} \int_{R^d} |\varphi(u)|^2 du.$$

Кроме того,

$$h(0) = \int_{R^d} |f(y)|^2 dy.$$

Введем оператор $\mathfrak{F}(\cdot) = \frac{1}{(2\pi)^{d/2}} \int_{R^d} e^{i(u, x)} f(x) dx$, отличающийся от преобразования Фурье функции $f \in L_1 \cap L_2$ множителем $(2\pi)^{-d/2}$. Он определен на $L_1 \cap L_2$, линеен, на основании равенства (15) — изометричен (сохраняет норму элемента в L_2):

$$\int_{R^d} |f(x)|^2 dx = \int_{R^d} |\mathfrak{F}f(u)|^2 du. \quad (16)$$

Подпространство $L_1 \cap L_2$ всюду плотно в L_2 . Действительно, пусть f — произвольный элемент L_2 , K_a — куб с центром в начале координат и длиной стороны $2a$, $f_a(x) = f(x)$ при $x \in K_a$ и $f_a(x) = 0$ при $x \notin K_a$. Тогда $\|f_a\|_2 \leq \|f\|_2$, $\|f_a\|_1 \leq (2a)^{d/2} \|f\|_2$ (согласно неравенству Коши — Буняковского), т. е. $f_a \in L_1 \cap L_2$. Кроме того, $\|f - f_a\| \rightarrow 0$ при $a \rightarrow \infty$ по теореме о мажорируемой сходимости. Это доказывает, что $L_1 \cap L_2$ всюду плотно в L_2 .

Пусть теперь H — произвольное гильбертово пространство.

Лемма. *Линейный изометрический оператор A , определенный на подпространстве H' , всюду плотном в H , $A: H' \rightarrow H$, может быть расширен на все H (и притом единственным образом) с сохранением линейности и изометричности.*

Доказательство. Нужно расширение получается с помощью доопределения «по непрерывности». Пусть f — произвольный элемент H и f_n , $n = 1, 2, \dots$, $f_n \in H'$ — последовательность, сходящаяся к f , $\|f - f_n\| \rightarrow 0$. Из изометричности и линейности A следует $\|Af_n - Af_{n+m}\| = \|f_n - f_{n+m}\| \rightarrow 0$ при $n \rightarrow \infty$, т. е. Af_n — фундаментальная последовательность элементов H . Следовательно, существует предел $\lim Af_n$. Положим $Af = \lim Af_n$. Элемент Af зависит только от f , а не от выбора последовательности f_n . Действительно, если $g = \lim Af_n$ и $g' = \lim Af'_n$, $f_n \rightarrow f$, $f'_n \rightarrow f$, то

$$\|g - g'\| = \lim \|Af_n - Af'_n\| = \lim \|f_n - f'_n\| = 0,$$

т. е. $g = g'$. Итак, Af определено единственным образом $\forall f \in H$. Очевидно, данное определение сохраняет линейность и изометричность оператора A .

$$A(f + f') = Af + Af', \quad \|Af\| = \|f\|.$$

Из леммы вытекает, что оператор $\mathfrak{F}f$ может быть определен на все L_2 . При этом $\mathfrak{F}(f) = \text{l.i.m.} \mathfrak{F}(f_n)$, где $f_n \in L_1 \cap L_2$, $\|f_n - f\|_2 \rightarrow 0$. Иногда бывает удобным в качестве f_n брать введенные выше функции f_a . Тогда

$$\mathfrak{F}(f) = \text{l.i.m.}_{a \rightarrow \infty} \frac{1}{(2\pi)^{\frac{1}{2}} K_a} \int e^{i(a, x)} f(x) dx. \quad (17)$$

Напомним, что линейный изометрический оператор H называется унитарным, если он отображает H на H .

Теорема 10. (Теорема Планшереля). Преобразование Фурье является унитарным оператором \mathfrak{F} в L_2 . Обратный к нему оператор \mathfrak{F}^{-1} имеет вид

$$(\mathfrak{F}^{-1}\varphi)(x) = (\mathfrak{F}\varphi)(-x). \quad (18)$$

Доказательство. Множество функций L' , являющихся преобразованием Фурье функций из L_2 , есть линейным и замкнутым.

Действительно, если $\varphi_n \in L'$ и $\varphi_n = \mathfrak{F}f_n$, $n = 1, 2, \dots$, то $\alpha\varphi_1 + \beta\varphi_2 = \mathfrak{F}(\alpha f_1 + \beta f_2)$, так что $\alpha\varphi_1 + \beta\varphi_2 \in L'$ для любых констант α и β . Если далее $\|\varphi - \varphi_n\|_2 \rightarrow 0$, то $\|f'_n - f_n\|_2 = \|\varphi'_n - \varphi_n\|_2 \rightarrow 0$ и поэтому существует $\lim f_n = f$, причем по определению оператора \mathfrak{F} , $\mathfrak{F}(f) = \varphi$, так что $\varphi \in L'$, и L' замкнуто. Если бы L' не совпадало с L_2 , то в L_2 нашелся бы элемент g , ортогональный L' . Из формулы умножения (3) следует

$$\int_{R^d} g(u) \varphi(u) du = \int_{R^d} \psi(u) f(u) du = 0 \quad \forall f \in L_2,$$

то есть $\psi(u) = (\mathfrak{F}g)(u) = 0$, по $\|\psi\|_2 = \|g\|_2 = 0$, так что $g \equiv 0$. Итак, $L' = L_2$ и оператор \mathfrak{F} — унитарен. При этом следует иметь в виду, что формула (3), установленная ранее для интегрируемых функций f и g , с помощью предельного перехода легко переносится и на функции f, g из L_2 . Для любой $g \in L_2$ существует $f \in L_2$, так что $\varphi = \mathfrak{F}f$. При этом функция f единственная. Действительно, если $\varphi = \mathfrak{F}(f_1) = \mathfrak{F}(f_2)$, то $\mathfrak{F}(f_1 - f_2) = 0$ и $0 = \|\mathfrak{F}(f_1 - f_2)\|_2 = \|f_1 - f_2\|_2$, т. е. $f_1 = f_2$ в L_2 . Мы показали, что однозначно определен обратный оператор $\mathfrak{F}^{-1}\varphi$. Покажем теперь, что

$$\mathfrak{F}^{-1}\varphi(x) = \frac{1}{(2\pi)^{\frac{1}{2}} K_d} \int e^{-i(x, u)} \varphi(u) du. \quad (19)$$

Обозначим правую часть этой формулы через \tilde{f} . Пусть сначала g и $\varphi \in L_1 \cap L_2$. Рассмотрим скалярное произведение в L_2 :

$$(g, h) = \int_{R^d} g(x) h(x) dx.$$

Так как функции g и φ интегрируемы, то $(g, \tilde{f}) = \frac{1}{(2\pi)^{d/2}} \int_{R^d} g(x) \times$
 $\times \int_{R^d} e^{-i(x, u)} \varphi(u) du = \frac{1}{(2\pi)^{d/2}} \int_{R^d} \int_{R^d} g(x) e^{-i(x, u)} dx \varphi(u) du = (\mathfrak{F}g, \varphi) = (\mathfrak{F}g,$
 $\mathfrak{F}f)$, где $\varphi = \mathfrak{F}f$. Учитывая, что $(\mathfrak{F}g, \mathfrak{F}f) = (g, f)$ согласно уни-
 тарности \mathfrak{F} , получим $(g, \tilde{f}) = (g, f)$ или $(g, \tilde{f} - f) = 0$. Посколь-
 ку множество $L_1 \cap L_2$ всюду плотно в L_2 , последнее равенство вы-
 полняется для всех $g \in L_2$. Отсюда $\tilde{f} = f$. Это равенство доказано
 для всех таких f , для которых $\varphi = \mathfrak{F}f \in L_1 \cap L_2$. В частности, оно
 выполняется для функций φ вида $\varphi = \mathfrak{F}f_a$, где $f(x)$ — произволь-
 ная функция из L_2 . Отсюда следует, что равенство $\tilde{f} = f$ имеет
 место для всех $f \in L_2$. Итак, $\mathfrak{F}^{-1}\mathfrak{F}f = f, \forall f \in L_2$.

Положительно определенные функции. Ранее уже отмечалось,
 что характеристическая функция $\varphi(u)$, $u \in R^d$, некоторого распре-
 деления непрерывна и положительно определена.

Определение. Комплекснозначную функцию $g(u)$, $u \in R^d$,
 называют положительно определенной, если для любого целого n ,
 векторов $u^k \in R^d$ и комплексных чисел $z_k, k = 1, \dots, n$,

$$\sum_{k, r=1}^n g(u^{(k)} - u^{(r)}) z_k \bar{z}_r \geq 0. \quad (20)$$

Оказывается, что любая непрерывная положительно определенная
 функция, удовлетворяющая еще условию $g(u) = 1$, является ха-
 рактеристической функцией некоторого распределения. Эта теорема была доказана Бохнером. Отметим, что положи-
 тельно определенные функции $g(u)$ обладают следующими свойст-
 вами:

$$1) g(0) \geq 0; \quad (21)$$

$$2) g(u) = \overline{g(-u)}; \quad (22)$$

$$3) |g(u)| \leq g(0); \quad (23)$$

$$4) |g(u^{(1)}) - g(u^{(2)})|^2 \leq 2g(0) [g(0) - \operatorname{Re} g(u^{(1)} - u^{(2)})]. \quad (24)$$

Из свойства 3) вытекает, что если $g(0) = 0$, то $g(u) \equiv 0$, а из
 4) — что из непрерывности функции $g(u)$ в точке $u = 0$ следует
 равномерная непрерывность $g(u)$ в R^d . Неравенство (21) является
 частным случаем (20) при $n = 1$. Если взять $n = 2, u^{(1)} = u,$
 $u^{(2)} = 0$, то (20) приводит к неравенству

$$g(0) |z_1|^2 + g(u) z_1 \bar{z}_2 + g(-u) \bar{z}_1 z_2 + g(0) |z_2|^2 \geq 0. \quad (25)$$

В частности, при любых z_1, z_2 выражение $g(u) z_1 \bar{z}_2 + g(-u) \bar{z}_1 z_2$
 имеет действительные значения. Это возможно только тогда, когда
 $g(-u) = \overline{g(u)}$, что доказывает 2).

Неравенство (25) можно записать в виде:

$$|\sqrt{g(0)} z_1 + \frac{\bar{g}(u)}{\sqrt{g(0)}} z_2|^2 + \left(g(0) - \frac{|g(u)|^2}{g(0)} \right) |z_2|^2 \geq 0.$$

Отсюда вытекает неравенство (23). Полагая теперь в (20) $n = 3$, $u_3 = 0$, $z_1 = -z_2$, получим

$$2(g(0) - \operatorname{Re} g(u^{(1)} - u^{(2)})) |z_1|^2 + (g(u^{(1)}) - g(u^{(2)})) z_1 \bar{z}_3 + \\ + (g(-u^{(1)}) - g(-u^{(2)})) \bar{z}_1 z_3 + g(0) |z_3|^2 \geq 0.$$

Из положительной определенности полученной квадратической формы, поступая аналогично предыдущему, получим неравенство (24).

Теорема II (Теорема Бохнера). Для того чтобы функция $g(u)$, $u \in R^d$, была характеристической функцией некоторого распределения, необходимо и достаточно, чтобы $g(u)$ была положительно определенной функцией, непрерывной в точке $u = 0$ и $g(0) = 1$.

Доказательство. Из непрерывности функции $g(u)$ в точке $u = 0$ следует равномерная непрерывность $g(u)$ в R^d . Кроме того, $|g(u)| \leq 1$. Заметим, что для произвольной непрерывности и интегрируемой в R^d функции $z(u)$

$$\int_{R^d} \int_{R^d} g(u-v) z(u) \overline{z(v)} du dv. \quad (26)$$

Это вытекает из того, что интеграл в левой части неравенства (26) можно представить как предел суммы вида

$$\sum_{k,r} g(u^{(k)} - u^{(r)}) z_k \bar{z}_r.$$

Из неравенства (26), в частности, имеем

$$\int_{R^d} \int_{R^d} g(u-v) \exp\{-\varepsilon^2(|u|^2 + |v|^2) - i(u-v, x)\} du dv \geq 0.$$

Введем замену переменных $u-v = t$, $u+v = s$. Получим

$$\int_{R^d} \int_{R^d} g(t) \exp\left\{-\frac{\varepsilon^2 |t|^2}{2} - i(t, x)\right\} dt \geq 0,$$

откуда

$$h(x) = \int_{R^d} g(t) \exp\left\{-\frac{\varepsilon^2 |t|^2}{2} - i(t, x)\right\} dt \geq 0. \quad (27)$$

Функция $h(x)$ — неотрицательная и является преобразованием Фурье непрерывной функции. Следовательно (следствие 2 теоремы 8), функция $h(x)$ — интегрируема и

$$g(t) \exp\left\{-\frac{\varepsilon^2 |t|^2}{2}\right\} = \frac{1}{(2\pi)^d} \int_{R^d} e^{i(t, x)} h(x) dx.$$

Функцию $\varphi_\varepsilon(t) = g(t) \exp\left\{-\frac{\varepsilon^2 |t|^2}{2}\right\}$ можно рассматривать, как характеристическую функцию некоторого распределения. При $\varepsilon \rightarrow$

$\rightarrow 0$ она стремится к непрерывной функции $g(t)$ и, следовательно, (теорема 3, § 6) сама также является характеристической функцией некоторого распределения $H(\cdot)$.

$$g(t) = \int_{R^d} e^{i(t, x)} H(dx).$$

В ряде задач теории вероятностей встречаются положительно определенные функции дискретного аргумента. Для них имеет место аналог теоремы Бохнера, ранее установленный Герглотцем.

Определение. Последовательность чисел (действительных или комплексных) c_n , $n = 0, \pm 1, \pm 2, \dots$, называют положительно определенной, если для любого n и любых комплексных z_1, \dots, z_n

$$\sum_{k, j=1}^n c_{k-j} z_k \bar{z}_j \geq 0. \quad (28)$$

Теорема 12. Для того чтобы последовательность чисел $\{c_n, n = 0, \pm 1, \pm 2, \dots\}$ была положительно определенной, необходимо и достаточно, чтобы она допускала представление

$$c_n = \int_{-\pi}^{\pi} e^{inx} dF(x), \quad (29)$$

где $F(x)$ — некоторая мера на отрезке $[-\pi, \pi]$.

Доказательство. Если последовательность допускает представление (29), то

$$\begin{aligned} \sum_{k, j=1}^n c_{k-j} z_k \bar{z}_j &= \int_{-\pi}^{\pi} \left(\sum_{k, j=1}^n e^{i(k-j)x} z_k \bar{z}_j \right) dF(x) = \\ &= \int_{-\pi}^{\pi} \left| \sum_{k=1}^n e^{ikx} z_k \right|^2 dF(x) \geq 0, \end{aligned}$$

т. е. она является положительно определенной. Допустим теперь, что последовательность $\{c_n\}$ положительно определена. Тогда

$$f_N(\theta) \stackrel{\text{Def}}{=} \frac{1}{2\pi(N+1)} \sum_{k, j=1}^{N+1} c_{k-j} e^{-i(k-j)\theta} = \frac{1}{2\pi} \sum_{r=-N}^N c_r \left(1 - \frac{|r|}{N+1} \right) e^{-ir\theta} \geq 0.$$

Имеем

$$\int_{-\pi}^{\pi} e^{ik\theta} f_N(\theta) d\theta = c_k \left(1 - \frac{|k|}{N+1} \right),$$

т. е. последовательность $c_k \left(1 - \frac{|k|}{N+1}\right)$ при $k = 0, \pm 1, \dots, \pm N$ допускает представление

$$c_k \left(1 - \frac{|k|}{N+1}\right) = \int_{-\pi}^{\pi} e^{ik\theta} dF_N(\theta), \quad (30)$$

где $F_N(\cdot)$ — некоторая мера на отрезке $[-\pi, \pi]$, причем

$$F_N[-\pi, \pi] = \int_{-\pi}^{\pi} dF_N(\theta) = c_0.$$

Последовательность ограниченных мер на ограниченном отрезке слабо компактна. Следовательно, можно найти такую подпоследовательность целых чисел N_k , $N_k \rightarrow \infty$, что F_{N_k} будут слабо сходиться к некоторой мере F на $[-\pi, \pi]$, для которой $F[-\pi, \pi] = c_0$. Переходя в равенстве (30) к пределу при $N = N_k \rightarrow \infty$, получим равенство (29).

§ 8. Центральная предельная теорема

Простейший вариант теоремы о сходимости к гауссовскому распределению. Настоящий параграф посвящен одному из самых замечательных результатов теории вероятностей: при широких условиях суммы большого числа независимых малых случайных слагаемых имеют распределение, близкое к гауссовскому. Значение этого результата выходит далеко за рамки теории вероятностей. Он является теоретической основой применения гауссовского распределения при решении многих практических задач. Всегда, когда можно считать, что рассматриваемая величина является суммой малых независимых слагаемых, ее распределение мало отличается от гауссовского. Таковыми являются, например, ошибки измерений в измерительных приборах, продолжительность исправной работы прибора, когда выход его из строя является следствием постоянно действующих (усталостных) изменений, скорости движения молекул газа, получаемых вследствие неоднородных столкновений с соседними молекулами, и т. п.

Приведем сначала простейший вариант центральной предельной теоремы, относящийся к суммам одинаково распределенных слагаемых. Сначала отметим несколько неравенств, используемых в дальнейшем.

Из геометрической интерпретации числа $e^{i\theta}$ следует, что

$$|e^{i\theta} - 1| \leq \min(2, |\theta|).$$

Отсюда

$$\left| \int_0^{\theta} (e^{i\theta} - 1) d\theta \right| \leq \min\left(\frac{|\theta|^2}{2}, 2|\theta|\right),$$

$$|e^{i\theta} - 1 - i\theta| \leq \min\left(\frac{|\theta|^2}{2}, 2|\theta|\right). \quad (1)$$

Таким образом,

$$e^{i\theta} = 1 + i\theta - \frac{\theta^2}{2}(1 + \varepsilon(\theta)),$$

где $\varepsilon(\theta)$ — ограниченная величина для всех $\theta \in (-\infty, \infty)$ и $\varepsilon(\theta) \rightarrow 0$ при $\theta \rightarrow 0$.

Пусть ξ — случайная величина, $M\xi = 0$, $D\xi = \sigma^2$. По теореме о мажорируемой сходимости

$$\varphi(u) = Me^{i\xi u} = 1 - \frac{u^2}{2}(\sigma^2 + \varepsilon_1(u)), \quad (2)$$

где $\varepsilon_1(u) \rightarrow 0$ при $u \rightarrow 0$.

Теорема 1. Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — последовательность независимых одинаково распределенных случайных величин $M\xi_n = 0$, $D\xi_n = 1$ и $F_n(x)$ — функция распределения суммы

$$\xi_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{\sqrt{n}}.$$

При $n \rightarrow \infty$ $F_n(x)$ слабо сходится к гауссовскому распределению с параметрами $(0, 1)$.

Доказательство. Пусть $\varphi(u)$ — характеристическая функция величины ξ_n . Из равенства (2) следует

$$\varphi_{\frac{\xi_k}{\sqrt{n}}}(u) = \varphi\left(\frac{u}{\sqrt{n}}\right) = 1 - \frac{u^2(1 + \varepsilon_n)}{2n},$$

где $\varepsilon_n \rightarrow 0$ при $n \rightarrow \infty$. Далее

$$\varphi_{\xi_n}(u) = \prod_{k=1}^n \varphi_{\frac{\xi_k}{\sqrt{n}}}(u) = \left(1 - \frac{u^2(1 + \varepsilon_n)}{2n}\right)^n.$$

При $n \rightarrow \infty$

$$\lim \varphi_{\xi_n}(u) = \lim e^{-\frac{u^2(1 + \varepsilon_n)}{2n}} = e^{-\frac{u^2}{2}}.$$

Таким образом, $\varphi_{\xi_n}(u)$ сходится при $n \rightarrow \infty$ к характеристической функции стандартного нормального закона и на основании предельной теоремы для характеристической функции распределение величины ξ_n слабо сходится к стандартному нормальному закону.

Следствие. (Интегральная теорема Муавра — Лапласа). Пусть $\frac{\nu}{n}$ — частота наступления события A в по-

следовательности n независимых опытов, в каждом из которых вероятность появления A равна p . Тогда при $n \rightarrow \infty$

$$P\left\{a < \sqrt{\frac{n}{pq}}\left(\frac{v}{n} - p\right) < b\right\} \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{t^2}{2}} dt.$$

Иными словами, распределение случайной величины

$$\xi_n = \sqrt{\frac{n}{pq}}\left(\frac{v}{n} - p\right)$$

при $n \rightarrow \infty$ слабо сходится к стандартному нормальному распределению.

Для доказательства достаточно заметить, что

$$\xi_n = \frac{1}{\sqrt{n}} \sum \frac{\chi_k - p}{\sqrt{pq}},$$

где χ_k — индикатор события A в k -м опыте, а величины $\frac{\chi_k - p}{\sqrt{pq}}$, $k = 1, 2, \dots, n$ независимы и одинаково распределены

$$M \frac{\chi_k - p}{\sqrt{pq}} = 0, \quad D \left(\frac{\chi_k - p}{\sqrt{pq}} \right) = 1.$$

Предельная теорема для стандартных последовательностей серий. Рассмотрим «последовательность серий» случайных величин

$$\xi_{n1}, \xi_{n2}, \dots, \xi_{nm_n}, \quad n = 1, 2, \dots \quad (3)$$

Будем предполагать, что величины, входящие в одну серию, взаимно независимы. Положим

$$\xi_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nm_n}.$$

Мы покажем, что если при $n \rightarrow \infty$ число слагаемых в сумме ξ_n неограниченно увеличивается, а сами слагаемые в определенном смысле равномерно стремятся к нулю, то распределение величин ξ_n слабо сходится к гауссовскому закону. Грубо говоря, сумма бесконечно большого числа малых независимых слагаемых имеет гауссовское распределение. Это утверждение имеет место при некоторых дополнительных предположениях, которые будут сформулированы ниже, и оказывается, вообще говоря, неверным, если соответствующие предположения не выполнены.

Назовем последовательность серий (3) стандартной, если величины ξ_{nk} обладают конечными моментами 2-го порядка и

$$M \xi_{nk} = 0, \quad D \xi_{nk} = \sigma_{nk}^2, \quad \sum_{k=1}^{m_n} \sigma_{nk}^2 = 1 \quad (4)$$

$$\max_{1 \leq k \leq m_n} \sigma_{nk}^2 \rightarrow 0, \quad n \rightarrow \infty. \quad (5)$$

Из условий (4) и (5) следует, что $m_n \rightarrow \infty$. Заметим, что суммы ξ_n , рассмотренные в теореме 1, являются частным случаем последовательности сумм серий независимых величин. В этом случае последовательность серий является «треугольной». Она имеет вид

$$\begin{array}{c} \frac{\xi_1}{\sqrt{1}}, \\ \frac{\xi_1}{\sqrt{2}}, \frac{\xi_2}{\sqrt{2}}, \\ \dots \\ \frac{\xi_1}{\sqrt{n}}, \frac{\xi_2}{\sqrt{n}}, \dots, \frac{\xi_n}{\sqrt{n}}, \\ \dots \end{array}$$

так что $m_n = n$, $\xi_{nk} = \frac{\xi_n}{\sqrt{n}}$, $M\xi_{nk} = 0$,

$$\sigma_{nk}^2 = D\xi_{nk} = \frac{1}{n}, \quad \sum_{k=1}^n \sigma_{nk}^2 = 1$$

и является стандартной.

Пусть $\varphi_n(u)$ — характеристическая функция случайной величины ξ_n , $\varphi_{nk}(u)$ — величины ξ_{nk} . Имеем

$$\varphi_n(u) = \prod_{k=1}^{m_n} \varphi_{nk}(u).$$

Лемма. Пусть ξ_{nk} , $k = 1, \dots, m_n$, $n = 1, 2, \dots$, — стандартная последовательность независимых (в каждой серии) случайных величин. Тогда

$$\lim [\ln \varphi_n(u) - \sum_{k=1}^{m_n} (\varphi_{nk}(u) - 1)] = 0.$$

Доказательство. Имеем

$$\ln \varphi_n(u) = \sum_{k=1}^{m_n} \ln \varphi_{nk}(u) = \sum \ln [1 + (\varphi_{nk}(u) - 1)].$$

Из равенства

$$\ln(1+z) - z = \int_0^z \left(\frac{1}{1+x} - 1 \right) dx = - \int_0^z \frac{x}{1+x} dx,$$

z — комплексное число, $|z| < 1$) вытекает

$$|\ln(1+z) - z| \leq \frac{|z|^2}{1+|z|}, \quad |z| < 1.$$

Поскольку

$$|\varphi_{nk}(u) - 1| = |M(e^{iu\xi_{nk}} - 1 - iu\xi_{nk})| \leq \frac{u^2 \sigma_{nk}^2}{2}$$

и $\sigma_{nk}^2 \rightarrow 0$, то при достаточно большом n получим $u^2 \sigma_{nk}^2 < 1$ и $|\varphi_{nk}(u) - 1| < \frac{1}{2}$. Следовательно,

$$|\ln \varphi_n(u) - \sum_k (\varphi_{nk}(u) - 1)| \leq \sum_k |\varphi_{nk}(u) - 1|^2 \frac{1}{1 + |\varphi_{nk}(u) - 1|} \leq \\ \leq \sum_{k=1}^{m_n} \frac{u^4 \sigma_{nk}^4}{4} \cdot \frac{1}{1 - \frac{u^2 \sigma_{nk}^2}{2}} \leq \frac{1}{2} u^4 \max_k \sigma_{nk}^2 \sum_{k=1}^{m_n} \sigma_{nk}^2 = \frac{1}{2} u^4 \max_k \sigma_{nk}^2.$$

Таким образом,

$$|\ln \varphi_n(u) - \sum_k (\varphi_{nk}(u) - 1)| \rightarrow 0$$

при $n \rightarrow \infty$.

Рассмотрим сумму $\sum_k (\varphi_{nk}(u) - 1)$. Ее можно представить в виде

$$\sum_k (\varphi_{nk}(u) - 1) = \int_{-\infty}^{\infty} (e^{iux} - 1 - iux) \sum_k dF_{nk}(x).$$

Введем функцию

$$f(x) = \frac{e^{iux} - 1 - iux}{x^2} \quad \text{при } x \neq 0 \quad (6) \\ f(0) = -\frac{u^2}{2}.$$

Функция $f(x)$ будет непрерывной и, на основании неравенства (1) — ограниченной на всей прямой $(-\infty, \infty)$. Положим

$$G_n(x) = \int_{-\infty}^x y^2 dF_{nk}(y).$$

При этом

$$G_n(+\infty) = \sum_{k=1}^{m_n} \int_{-\infty}^{\infty} y^2 dF_{nk}(y) = \sum \sigma_{nk}^2 = 1.$$

Функция $G_n(x)$ является функцией распределения и

$$\sum_{k=1}^{m_n} (\varphi_{nk}(u) - 1) = \int_{-\infty}^{\infty} \frac{e^{iux} - 1 - iux}{x^2} dG_n(x). \quad (7)$$

Из полученной формулы и предельной теоремы для характеристической функции непосредственно вытекает следующая теорема.

Теорема 2. Пусть ξ_{nk} , $k = 1, \dots, m_n$, $n = 1, 2, \dots$ — стандартная последовательность независимых случайных величин и

функция $G_n(x)$ сходится в основном к некоторой функции распределения $G(x)$. Тогда распределение суммы $\xi_n = \xi_{n1} + \dots + \xi_{nm_n}$ слабо сходится к распределению с характеристической функцией

$$\varphi(u) = \exp \left\{ \int_{-\infty}^{\infty} \frac{e^{iux} - 1 - iux}{x^2} dG \right\}. \quad (8)$$

Из теоремы 2 легко получить достаточные условия сходимости распределения величины ξ к гауссовскому закону. Условимся говорить, что последовательность случайных величин η_n , $n = 1, 2, \dots$, асимптотически нормальна (a, σ^2), если функция распределения $F_n(x)$ случайных величин η_n слабо сходится к гауссовскому распределению с параметрами a и σ .

Теорема 3. (Теорема Линдеберга). Пусть последовательность серий независимых случайных величин (3) удовлетворяет условию (4) и

$$L(c) = \sum_{k=1}^{m_n} M\xi_{nk}^2 \bar{\chi}_{nk}(c) \rightarrow 0, \quad \forall c > 0, \quad (9)$$

где $\bar{\chi}_{nk}(c) = 1$ при $|\xi_{nk}| > c$ и $\bar{\chi}_{nk}(c) = 0$ при $|\xi_{nk}| \leq c$. Тогда последовательность сумм $\xi_n = \xi_{n1} + \dots + \xi_{nm_n}$ асимптотически нормальна $(0, 1)$.

Условие (9) называют условием Линдеберга.

Доказательство. Заметим, что на основании условия Линдеберга $\max_k \sigma_{nk}^2 = 0$. Действительно,

$$\sigma_{nk}^2 = M\xi_{nk}^2 = M\xi_{nk}^2 (1 - \bar{\chi}_{nk}(c)) + M\xi_{nk}^2 \bar{\chi}_{nk}(c) \leq c^2 + L_n(c),$$

где c — произвольное положительное число и $L_n(c) \rightarrow 0$ при $n \rightarrow \infty$. Таким образом, при любом $c > 0$ и достаточно большом $n \geq n_0(c)$, $\sigma_{nk}^2 \leq 2c$. Следовательно, к рассматриваемой последовательности серий случайных величин применима теорема 2. Заметим еще, что $G_n(-c) \leq L_n(c)$, поэтому для любого $x < -c$ $G_n(x) \rightarrow 0$ при $n \rightarrow \infty$. Аналогично $1 - G_n(c) \leq L_n(c)$, так что $G_n(x) \rightarrow 1$ при $x \geq c$, где c — произвольное положительное число. Таким образом, $G_n(x) \rightarrow \chi(x)$ для всех $x \neq 0$, где $\chi(x) = 0$ при $x \leq 0$ и $\chi(x) = 1$ при $x > 0$. Поскольку для любой непрерывной функции $f(x)$

$$\int_{-\infty}^{\infty} f(x) d\chi(x) = f(0),$$

то согласно теореме 2 и формуле (6)

$$\lim \varphi_n(u) = \exp \left(-\frac{1}{2} u^2 \right),$$

откуда вытекает, что функция распределения суммы ξ_n сходится в основном (в данном случае для всех x) к гауссовской функции распределения с параметром $(0, 1)$.

Центральная предельная теорема впервые строго была доказана при достаточно общих предположениях Ляпуновым. Условия Ляпунова несколько уже условий Линдеберга, но более удобны для проверки.

Теорема 4. (Теорема Ляпунова). Если последовательность серий случайных величин (3) удовлетворяет условию (4) и условию Ляпунова, для некоторого $\delta > 0$

$$\bar{L}_n(\delta) = \sum_{k=1}^{m_n} M |\xi_{nk}|^{2+\delta} \rightarrow 0, \quad (10)$$

то суммы $\xi_n = \xi_{n1} + \dots + \xi_{nm_n}$ асимптотически нормальны $(0, 1)$.

Для доказательства теоремы Ляпунова достаточно показать, что если выполнено условие Ляпунова, то условие Линдеберга также выполняется. Последнее вытекает из неравенств

$$L_n(c) = \sum_{k=1}^{m_n} M |\xi_{nk}|^2 \bar{\chi}_{nk}(c) \leq \frac{1}{c^\delta} \sum_{k=1}^{m_n} M |\xi_{nk}|^{2+\delta} = \frac{1}{c^\delta} \bar{L}_n(\delta).$$

Рассмотрим такую последовательность серий случайных величин, которая не удовлетворяет условиям нормировки (4). Пусть

$$M\xi_{nk} = a_{nk}, \quad D\xi_{nk} = b_{nk}, \quad \sum_{k=1}^{m_n} b_{nk} = b_n. \quad (11)$$

Если ввести новые случайные величины

$$\eta_{nk} = \frac{\xi_{nk} - a_{nk}}{\sqrt{b_n}},$$

то

$$M\eta_{nk} = 0, \quad \sum_{k=1}^{m_n} D\eta_{nk} = 1.$$

К величинам η_{nk} применимы теоремы 2—4.

Теорема 5. Пусть η_{nk} , $k = 1, 2, \dots, m_n$; $n = 1, 2, \dots$, — последовательность серий случайных величин, независимых в каждой серии и имеющих конечные моменты второго порядка. Если выполнено условие Линдеберга: для любого $c > 0$

$$L_n(c) = \frac{1}{b_n} \sum_{k=1}^{m_n} M (\xi_{nk} - a_{nk})^2 \bar{\chi}_{nk}(c) \rightarrow 0 \quad n \rightarrow \infty, \quad (12)$$

где $\chi_{nk} = 1$ при $|\xi_{nk} - a_{nk}| > c\sqrt{b_n}$ и $\chi_{nk} = 0$ при $|\eta_{nk} - a_{nk}| \leq c\sqrt{b_n}$, то последовательность случайных величин

$$\zeta_n = \frac{\sum_{k=1}^{m_n} \xi_{nk} - a_n}{\sqrt{b_n}}, \quad a_n = \sum_{k=1}^{m_n} a_{nk},$$

асимптотически нормальна $(0, 1)$.

Условие Линдеберга (12) выполняется, в частности, если выполнено условие Ляпунова: при некотором $\delta > 0$

$$\frac{1}{b_n^{1+\frac{\delta}{2}}} \sum_{k=1}^{m_n} M|\xi_{nk} - a_{nk}|^{2+\delta} \rightarrow 0 \text{ при } n \rightarrow \infty. \quad (13)$$

Эта теорема непосредственно вытекает из теорем 3, 4, если их применить к введенным выше величинам η_{nk} .

В ряде случаев целесообразно пользоваться следующим вариантом теоремы 5.

Теорема 6. Если $a_n \rightarrow a$, $b_n \rightarrow b$ и выполняется условие Линдеберга (12), то сумма $\zeta_n = \sum_{k=1}^{m_n} \xi_{nk}$ асимптотически нормальна (a, b) .

Действительно, пусть

$$\tilde{\zeta}_n = \frac{\zeta_n - a_n}{\sqrt{b_n}}.$$

Тогда

$$\varphi_{\zeta_n}(u) = e^{ia_n u} \varphi_{\tilde{\zeta}_n}(\sqrt{b_n} u) = e^{ia_n u} \varphi_{\tilde{\zeta}_n}(\sqrt{b} u) + r_n,$$

где

$$\begin{aligned} r_n &= e^{ia_n u} (\varphi_{\tilde{\zeta}_n}(\sqrt{b_n} u) - \varphi_{\tilde{\zeta}_n}(\sqrt{b} u)), \\ |r_n| &\leq |u| M |\sqrt{b_n} - \sqrt{b}| |\tilde{\zeta}_n| \leq |u| |\sqrt{b_n} - \sqrt{b}| (M \zeta_n^2)^{\frac{1}{2}} = |u| |\sqrt{b_n} - \sqrt{b}|. \end{aligned}$$

Таким образом, $\lim r_n = 0$ и $\lim \varphi_{\zeta_n}(u) = \exp\left(iau - \frac{bu^2}{2}\right)$.

Из теоремы 2 можно получить ряд результатов о сходимости распределения сумм независимых слагаемых к законам, отличным от нормального. Например, из нее непосредственно вытекает теорема о сходимости к центрированному распределению Пуассона.

Теорема 7. Пусть стандартная последовательность серий случайных величин удовлетворяет следующим условиям: для некоторого $a > 0$ и любого $h > 0$:

$$a) \sum_{k=1}^{m_n} \left(\int_{-\infty}^{a-h} + \int_{a+h}^{\infty} \right) y^2 dF_{nk} \rightarrow 0,$$

$$б) \sum_{k=1}^{m_n} P(\xi_{nk} \in (a-h, a+h)) \rightarrow \lambda.$$

Тогда распределение суммы $\xi_n = \sum_{k=1}^{m_n} \xi_{nk}$ слабо сходится к распределению величины $a(\pi - \lambda)$, где случайная величина π имеет распределение Пуассона со средним λ .

Доказательство. Из а) и б) следует, что $G_n(u) \rightarrow G(u) = 0$ при $u < a$ и $G_n(u) \rightarrow G(u) = \lambda a^2$ при $u > a$. Таким образом, характеристическая функция предельного распределения равна:

$$\exp \left\{ \int_{-\infty}^{\infty} (e^{iux} - 1 - iux) \frac{1}{x^2} dG(x) \right\} = \\ = \exp \{ \lambda (e^{iua} - 1 - iua) \} = e^{-i(a\lambda)u} \exp \{ \lambda (e^{iua} - 1) \},$$

что совпадает с характеристической функцией величины $a(\pi - \lambda)$.

Центральная предельная теорема для сумм случайных векторов. Предельные теоремы для сумм случайных векторов можно легко получить из предельных теорем для случайных величин. Эта возможность содержится в выражении для многомерной характеристической функции

$$\varphi_{\xi}(tu) = M e^{it \sum_{k=1}^d u_k \xi_k},$$

где t — действительное число, которое одновременно является характеристической функцией одномерной случайной величины

$$\eta = \sum_{k=1}^d u_k \xi_k = (u, \xi).$$

Пусть

$$\xi_{n1}, \xi_{n2}, \dots, \xi_{nm_n}, \quad n = 1, 2, \dots, \quad (14)$$

последовательность серий случайных векторов $\xi_{nk} = ((\xi_1)_{nk}, (\xi_2)_{nk}, \dots, (\xi_d)_{nk})$, независимых в каждой серии и имеющих конечные моменты второго порядка. Положим

$$M \xi_{nk} = a_{nk}, \quad M (\xi_{nk} - a_{nk}) (\xi_{nk} - a_{nk})' = B_{nk}$$

и пусть:

а) существуют пределы

$$a = \lim_{\text{Def}} a_n, \quad B = \lim_{\text{Def}} B_n,$$

где

$$a_n = \sum_{k=1}^{m_n} a_{nk}, \quad B_n = \sum_{k=1}^{m_n} B_{nk}.$$

б) выполняется условие Линдеберга

$$L_n(c) = \sum_{\text{Def}} \sum_{k=1}^{m_n} \mathbf{M} |\xi_{nk} - a_{nk}|^2 \bar{\chi}(c, \xi_{nk} - a_{nk}) \rightarrow 0, \quad n \rightarrow \infty,$$

где $\bar{\chi}(c, t) = 1$, если $t > c$, $\bar{\chi}(c, t) = 0$ при $t \leq c$ ($t > 0, c > 0$).

Очевидно, матрица B неотрицательно определена. Возьмем некоторый вектор u , $u \in R^d$, $|u| = 1$, и введем последовательность серий случайных величин η_{nk} , положив

$$\eta_{nk} = (\xi_{nk}, u) = \sum_{r=1}^d u_r (\xi_r)_{nk}.$$

Имеем

$$\mathbf{M} \eta_{nk} = (a_{nk}, u), \quad \mathbf{D} \eta_{nk} = \mathbf{M} (\xi_{nk} - a_{nk}, u)^2 = (B_{nk} u, u).$$

Пусть

$$\zeta_n = \sum_{k=1}^{m_n} \xi_{nk}, \quad \tilde{\zeta}_n = \sum_{k=1}^{m_n} \eta_{nk}.$$

Тогда

$$\mathbf{M} \tilde{\zeta}_n = (a_n, u), \quad \mathbf{D} \tilde{\zeta}_n = \sum_{k=1}^{m_n} \mathbf{D} \eta_{nk} = (B_n u, u) \stackrel{\text{Def}}{=} \sigma_n^2$$

и

$$\lim \mathbf{M} \tilde{\zeta}_n = (a, u), \quad \lim \mathbf{D} \tilde{\zeta}_n = (Bu, u) \stackrel{\text{Def}}{=} \sigma^2 \geq 0.$$

Будем считать, что вектор u удовлетворяет еще дополнительному условию $\sigma^2 = (Bu, u) > 0$. Покажем, что тогда величины η_{nk} удовлетворяют условию Линдеберга для скалярного случая. Положим

$$\tilde{\eta}_{nk} = \eta_{nk} - (a_{nk}, u), \quad \bar{\xi}_{nk} = \xi_{nk} - a_{nk}.$$

При достаточно большом n , при котором $\sigma_n^2 > \frac{\sigma^2}{4}$, имеем

$$\begin{aligned} L_n(c, \eta) &= \sum_{\text{Def}} \frac{1}{\sigma_n^2} \sum_{k=1}^{m_n} \mathbf{M} \eta_{nk}^2 \bar{\chi}(c \sigma_n, |\eta_{nk}|) \leq \\ &\leq \frac{1}{\sigma_n^2} \sum_{k=1}^{m_n} \mathbf{M} |\tilde{\eta}_{nk}|^2 \bar{\chi}(c \sigma_n, |\tilde{\eta}_{nk}|) \leq \frac{1}{\sigma_n^2} \sum_{k=1}^{m_n} \mathbf{M} |\bar{\xi}_{nk}|^2 \times \\ &\times \chi\left(\frac{c \sigma}{2}, |\bar{\xi}_{nk}|\right) \leq \frac{1}{\sigma_n^2} L_n\left(\frac{c \sigma}{2}\right), \end{aligned}$$

так что $\lim L_n(c, \eta) = 0$. Аналогично, если последовательность серий (14) удовлетворяет условию Ляпунова: при некотором $\delta > 0$

$$\bar{L}_n(\delta) = \sum_{k=1}^{m_n} \mathbf{M} |\xi_{nk} - a_{nk}|^{2+\delta} \rightarrow 0 \quad \text{при } n \rightarrow \infty, \quad (16)$$

то выполняется условие Ляпунова для скалярных случайных величин η_{nk} . Таким образом, если выполнены условия а) и б) (или условие а) и условие Ляпунова (16), то для любого u , для которого $(Bu, u) > 0$, характеристическая функция суммы $\tilde{\xi}_n = \sum_{k=1}^{m_n} \eta_{nk}$ стремится к пределу

$$\lim \varphi_{\tilde{\xi}_n} = \exp \left\{ i(a, u) t - \frac{\sigma^2 t^2}{2} \right\} = \exp \left\{ i(a, u) t - \frac{t^2 (Bu, u)}{2} \right\}.$$

Это равенство сохраняется и тогда, когда $\sigma^2 = (Bu, u) = 0$, так как в этом случае $M\tilde{\xi}_n \rightarrow (a, u)$, $M(\tilde{\xi}_n - (a, u))^2 \rightarrow 0$ и, следовательно, $\varphi_{\tilde{\xi}_n}(t) \rightarrow \exp \{i(a, u)t\}$. Полагая $v = tu$, получим для любого $v \in R^d$

$$\varphi_{\tilde{\xi}_n}(v) = \varphi_{\tilde{\xi}_n}(t) \rightarrow \exp \left\{ i(a, v) - \frac{(Bv, v)}{2} \right\}.$$

Таким образом, доказана следующая теорема:

Теорема 8. (Теорема о нормальной корреляции). Если последовательность случайных векторов (14) с конечными моментами второго порядка удовлетворяет условиям а) и б), то распределение суммы $\tilde{\xi}_n = \xi_{n1} + \xi_{n2} + \dots + \xi_{nm_n}$ слабо сходится к нормальному закону с характеристической функцией

$$\varphi(v) = \exp \left\{ i(a, v) - \frac{1}{2} (Bv, v) \right\}.$$

Теорема 8 показывает, что сумма бесконечно большого числа равномерно бесконечно малых слагаемых векторов (условие Линдеберга можно интерпретировать как условие равномерной бесконечной малости слагаемых) имеет нормальное распределение, каковы бы ни были индивидуальные распределения слагаемых и корреляционные связи между их компонентами. В частности, уравнение регрессии одной из компонент такой суммы на остальные всегда линейно. Это обстоятельство объясняет, почему в приложениях теории вероятностей часто есть все основания считать, что уравнение регрессии интересующих нас случайных величин линейно.

Предельное распределение для мультиномиального распределения. В качестве примера на применение центральной предельной теоремы для случайных векторов рассмотрим предельное распределение для мультиномиального распределения (см. § 5).

Производится n независимых испытаний, в каждом из которых независимо от результатов других испытаний происходит одно из событий E_1, \dots, E_d с вероятностями p_1, p_2, \dots, p_d . Пусть v_k — число испытаний, в которых произошло событие E_k .

Рассмотрим случайный вектор

$$\xi^{(n)} = \left(\sqrt{\frac{n}{p_1}} \left(\frac{v_1}{n} - p_1 \right), \sqrt{\frac{n}{p_2}} \left(\frac{v_2}{n} - p_2 \right), \dots, \sqrt{\frac{n}{p_d}} \left(\frac{v_d}{n} - p_d \right) \right).$$

Его можно представить в виде

$$\xi^{(n)} = \frac{1}{\sqrt{n}} \sum_{j=1}^n \xi^{(j)},$$

где $\xi^{(j)}$ — случайный вектор с компонентами $\xi_k^{(j)}$

$$\xi_k^{(j)} = \frac{1}{\sqrt{p_k}} (\chi_j(E_k) - p_k), \quad k = 1, \dots, d.$$

$\chi_j(E_k) = 1$, если в j -м испытании событие E_k происходит, и $\chi_j(E_k) = 0$ — в противном случае. Векторы $\xi^{(j)}$ взаимно независимы и одинаково распределены, причем

$$M \xi_k^{(j)} = \frac{1}{\sqrt{p_k}} (M \chi_j(E_k) - p_k) = 0,$$

$$M (\xi_k^{(j)})^2 = \frac{p_k(1-p_k)}{p_k} = 1 - p_k,$$

$$M \xi_k^{(j)} \xi_l^{(j)} = \frac{1}{\sqrt{p_k p_l}} (M \chi_j(E_k) \chi_j(E_l) - p_k p_l) = -\sqrt{p_k p_l}$$

при $k \neq l$, так как $M \chi_j(E_k) \chi_j(E_l) = 0$ при $k \neq l$ и

$$M \chi_j(E_k) = p_k.$$

Теорема 9. При $n \rightarrow \infty$ случайный вектор $\xi^{(n)}$ асимптотически нормален со средним 0 и с корреляционной матрицей

$$B = \begin{pmatrix} 1-p_1 & -\sqrt{p_1 p_2} & \dots & -\sqrt{p_1 p_d} \\ -\sqrt{p_2 p_1} & 1-p_2 & \dots & -\sqrt{p_2 p_d} \\ \dots & \dots & \dots & \dots \\ -\sqrt{p_d p_1} & -\sqrt{p_d p_2} & \dots & 1-p_d \end{pmatrix} = I - z z', \quad (17)$$

где I — единичная матрица, z — матрица, состоящая из одного столбца $(\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_d})$, z' — матрица-строка с теми же элементами.

Матрица B (а, следовательно, и предельное распределение $\xi^{(n)}$) является вырожденной. Действительно, $Bz = z - z(z'z) = 0$, так как $z'z = 1$. Таким образом, 0 является собственным числом, z — соответственно собственным вектором матрицы B .

Найдем еще предельное распределение величины

$$|\xi^{(n)}|^2 = \sum_{k=1}^d \frac{n}{p_k} \left(\frac{v_k}{n} - p_k \right)^2, \quad (18)$$

играющей важную роль в математической статистике.

Поскольку величина $|\xi^{(n)}|^2$ является непрерывной функцией от вектора $\xi^{(n)}$, то ее предельное распределение совпадает с распределением величины $|\xi|^2$, где ξ — d -мерный нормальный вектор со средним 0 и корреляционной матрицей (17).

Введем ортогональную матрицу $V = (v_{ik})$, $i, k = 1, \dots, d$, в которой $v_{1k} = \sqrt{p_k}$, $k = 1, \dots, d$, а остальные элементы выбраны произвольно.

Напомним, что векторы, составленные из элементов строк матрицы V (или столбцов), попарно ортогональны. Положим $\eta = V\xi$.

Тогда η — нормально распределенный вектор, $M\eta = 0$ и корреляционная матрица B_η вектора η равна $B_\eta = MV\xi(V\xi)' = VM\xi\xi'V' = VBV' = I - (Vz)(Vz)'$, так как $V' = V^{-1}$. Из упомянутой ортогональности строк матрицы V следует, что $Vz = (1, 0, \dots, 0)$ и $(Vz)(Vz)'$ — матрица, у которой элемент $\beta_{11} = 1$, а остальные элементы равны 0. Таким образом, элементы b_{ij} матрицы B_η равны:

$$b_{11} = 0, \quad b_{ik} = 0 \text{ при } i \neq k, \\ b_{kk} = 1 \text{ при } k = 2, \dots, d.$$

Следовательно, $\eta_1 = 0$ с вероятностью 1, η_2, \dots, η_d — независимые стандартные гауссовские величины. Так как ортогональное преобразование сохраняет длину вектора, то

$$|\xi|^2 = |\eta|^2 = \sum_{k=2}^d \eta_k^2 = \chi_{d-1}^2,$$

где величина χ_{d-1}^2 имеет χ^2 -распределение с $d-1$ степенью свободы. Мы получили следующую теорему.

Теорема 10. При $n \rightarrow \infty$ распределение величины (18) слабо сходится к χ^2 -распределению с $d-1$ степенью свободы.

Глава VII

ПРОЦЕССЫ С НЕЗАВИСИМЫМИ ПРИРАЩЕНИЯМИ

§ 1. Определения, простейшие свойства

Мы уже встречались со случайными процессами, например марковским процессом со счетным числом состояний. Под случайным процессом понимается семейство случайных величин $\{\xi(t), t \in T\}$, где T — некоторое множество на прямой, называемое областью определения процесса. Рассматриваемые величины принимают значения из одного и того же пространства, которое называется фазовым пространством процесса, эти величины считаются заданными на одном и том же вероятностном пространстве.

Одна из важнейших вероятностных характеристик случайного процесса — его конечномерные (частные, иногда говорят «совместные») распределения, задающие для любого набора t_1, t_2, \dots ,

$t_n \in T$; совместное распределение случайных величин $\xi(t_1), \dots, \xi(t_n)$. Процесс $\xi(t)$ называется процессом с независимыми приращениями, если его фазовое пространство есть конечномерное евклидово пространство и для всех $t_0 < t_1 < \dots < t_n$ из T случайные величины $\xi(t_0), \xi(t_1) - \xi(t_0), \dots, \xi(t_n) - \xi(t_{n-1})$ являются независимыми. Для того чтобы задать конечномерные распределения процесса с независимыми приращениями, достаточно задать распределение значений процесса в одной точке $\xi(t)$ и распределения приращений $\xi(s) - \xi(t)$ при $s > t$.

Мы будем рассматривать процессы, заданные на полубесконечном промежутке $[0, \infty)$. В этом случае для задания конечномерных распределений достаточно задать распределение $\xi(0)$ и распределение приращения $\xi(s) - \xi(t)$ при $s > t$.

Процесс $\xi(t)$ называется стохастически непрерывным, если для всякого t и $\varepsilon > 0$

$$\lim_{s \rightarrow t} P\{|\xi(s) - \xi(t)| > \varepsilon\} = 0,$$

т. е. $\xi(s) \rightarrow \xi(t)$ по вероятности при $s \rightarrow t$. Оказывается, для стохастически непрерывных процессов их распределения могут принадлежать только определенному классу распределений, так называемых безгранично-делимых, если только $\xi(0) = 0$. Мы изучим вид распределений для более узкого класса процессов: однородных процессов с независимыми приращениями.

Процесс $\xi(t)$ называется однородным, если $\xi(0) = 0$ и распределение величины $\xi(t+h) - \xi(t)$, $t \geq 0$, $h \geq 0$ не зависит от t .

Введем характеристическую функцию однородного процесса:

$$\varphi(t, z) = M \exp\{i(z, \xi(t))\}, \quad (1)$$

здесь $i = \sqrt{-1}$, (\cdot, \cdot) — скалярное произведение в пространстве R^m — фазовом пространстве процесса, $z \in R^m$. Функция $\varphi(t, z)$ удовлетворяет следующему функциональному уравнению:

$$\varphi(t+s, z) = \varphi(t, z) \varphi(s, z), \quad t, s > 0. \quad (2)$$

Действительно,

$$\begin{aligned} M \exp\{i(z, \xi(t+s))\} &= M \exp\{i(z, \xi(t+s) - \xi(s))\} \exp\{i(z, \xi(s))\} = M \exp\{i(z, \xi(t+s) - \xi(s))\} M \exp\{i(z, \xi(s))\} = \\ &= M \exp\{i(z, \xi(t))\} M \exp\{i(z, \xi(s))\}. \end{aligned}$$

Пусть $\xi(t)$ — стохастически непрерывный процесс. Тогда $\varphi(t, z)$ — непрерывная функция t . Это вытекает из соотношения (2) и того, что

$$\lim_{s \rightarrow 0} \varphi(s, z) = 1. \quad (3)$$

Для доказательства (3) заметим, что

$$|\varphi(s, z) - 1| \leq M |\exp\{i(z, \xi(s))\} - 1| \leq \varepsilon + 2P\{|(z, \xi(s))| > \varepsilon\}.$$

Последняя вероятность стремится к нулю при $s \rightarrow 0$, так как $\xi(s) \rightarrow \xi(0) = 0$ по вероятности. Из соотношения (3) и равенства

$$\varphi(t, z) = \varphi^n\left(\frac{t}{n}, z\right)$$

вытекает, что $\varphi(t, z) \neq 0$ для всех t . Далее

$$\varphi\left(\frac{k}{n}, z\right) = [\varphi(1, z)]^{\frac{k}{n}}$$

(берется главное значение корня n -й степени из $\varphi(1, z)$). Переходя к пределу при $\frac{k}{n} \rightarrow t$ и обозначая $K(z) = \ln \varphi(1, z)$ главное значение логарифма, имеем

$$\varphi(t, z) = \exp\{tK(z)\}. \quad (4)$$

Функция $K(z)$ называется кумулянтной процесса и может быть вычислена по формуле

$$K(z) = \lim_{h \rightarrow \infty} \frac{1}{h} [\varphi(h, z) - 1]. \quad (5)$$

Пусть $F(h, dx)$ — распределение величины $\xi(h)$. Положим

$$\begin{aligned} K_h(z) &= \frac{1}{h} \int (e^{i(z, x)} - 1) F(h, dx) = \\ &= \int \left(e^{i(z, x)} - 1 - \frac{i(z, x)}{1 + |x|^2} \right) \frac{1 + |x|^2}{|x|^2} G(h, dx) + \\ &\quad + \frac{1}{h} \int \frac{i(z, x)}{1 + |x|^2} F(h, dx), \end{aligned}$$

где $G(h, A) = \frac{1}{h} \int_A \frac{|x|^2}{1 + |x|^2} F(h, dx)$.

Поскольку для всех $z \in R^m$ существует предел (5), то существует и предел

$$\lim_{h \rightarrow 0} \int (\cos(z, x) - 1) \frac{1 + |x|^2}{|x|^2} G(h, dx) = \operatorname{Re} K(z). \quad (6)$$

Из непрерывности $K(z)$ и формулы (4) вытекает, что предел в (5) существует равномерно по z в любом ограниченном множестве. Следовательно, и предел в (6) существует равномерно при $|z| \leq C$. Выберем в R^m базис $\{e_1, \dots, e_m\}$. Тогда

$$\int \sum_{k=1}^m \left(1 - \frac{\sin(x, e_k) \delta}{\delta(x, e_k)} \right) \frac{1 + |x|^2}{|x|^2} G(h, dx) =$$

$$= \frac{1}{2\delta} \int_{-\delta}^{\delta} \sum_{k=1}^m \int (1 - \sin \lambda(x, e_k)) \frac{1 + |x|^2}{|x|^2} G(h, dx) d\lambda \rightarrow$$

$$\rightarrow -\frac{1}{2\delta} \int_{-\delta}^{\delta} \sum_{k=1}^m K(\lambda e_k) d\lambda.$$

Поэтому

$$\overline{\lim}_{h \rightarrow 0} \int_{|x| > \frac{1}{\delta}} \frac{1 + |x|^2}{|x|^2} G(h, dx) \leq$$

$$\leq \overline{\lim}_{h \rightarrow 0} \frac{1}{\inf_{|x| > \frac{1}{\delta}} \sum_{k=1}^m \left(1 - \frac{\sin(x, e_k) \delta}{\delta(x, e_k)}\right)} \frac{1}{2\delta} \int_{-\delta}^{\delta} \sum_{k=1}^m K(\lambda e_k) d\lambda.$$

Обозначим $\inf_{\lambda \geq \alpha} \left(1 - \frac{\sin \lambda}{\lambda}\right) = \psi(\alpha)$. Тогда

$$\inf_{|x| > \frac{1}{\delta}} \sum_{k=1}^m \left(1 - \frac{\sin(x, e_k) \delta}{\delta(x, e_k)}\right) \geq \psi\left(\frac{1}{\sqrt{m}}\right).$$

Следовательно,

$$\lim_{\delta \rightarrow 0} \overline{\lim}_{h \rightarrow \infty} \int_{|x| > \frac{1}{\delta}} G(h, dx) \leq \psi^{-1}\left(\frac{1}{m}\right) \times$$

$$\times \lim_{\delta \rightarrow 0} \frac{1}{2\delta} \int_{-\delta}^{\delta} \sum_{k=1}^m \operatorname{Re} K(\lambda, e_k) d\lambda = 0,$$

так как $K(0) = 0$. Но тогда множество мер $G(h, dx)$ компактно. Пусть $G(h, dx)$ сходится к $G(dx)$. Имеем

$$\lim_{h \rightarrow 0} \int \left(e^{i(z, x)} - 1 - \frac{i(z, x)}{1 + |x|^2} + \frac{1}{2} \cdot \frac{(z, x)^2}{1 + |x|^2} \right) \frac{1 + |x|^2}{|x|^2} G(h, dx) =$$

$$= \int \left(e^{i(z, x)} - 1 - \frac{i(z, x)}{1 + |x|^2} + \frac{1}{2} \cdot \frac{(z, x)^2}{1 + |x|^2} \right) \frac{1 + |x|^2}{|x|^2} G(dx).$$

Поскольку подынтегральная функция в выражении справа равна нулю в точке 0, если доопределить ее по непрерывности, то интеграл справа не изменится, когда меру $G(dx)$ заменить на $\bar{G}(dx)$, где $\bar{G}(A) = G(A - \{0\})$, ($\{0\}$ — множество из одной точки 0).

Поэтому существует и предел

$$\lim_{h \rightarrow 0} \left[-\frac{1}{2} \int \frac{(z, x)^2}{|x|^2} G(h, dx) + \frac{1}{h} \int \frac{i(z, x)}{1 + |x|^2} F(h, dx) \right] =$$

$$= K(z) - \int \left(e^{i(z, x)} - 1 - \frac{i(z, x)}{1 + |x|^2} + \frac{1}{2} \cdot \frac{(z, x)^2}{1 + |x|^2} \right) \frac{1 + |x|^2}{|x|^2} \bar{G}(dx).$$

Но тогда существуют пределы вещественной и мнимой частей. Из того что вещественная часть квадратическая неотрицательная функция относительно z , вытекает существование неотрицательного линейного оператора B_1 , для которого

$$\lim_{h \rightarrow 0} \frac{1}{h} \int \frac{(z, x)^2}{|x|^2} G(h, dx) = (B_1 z, z).$$

Точно так существует $a \in R^m$, такое, что

$$\lim_{h \rightarrow 0} \frac{1}{h} \int \frac{i(z, x)}{1 + |x|^2} F(h, dx) = i(a, z).$$

Поэтому

$$\begin{aligned} K(z) = i(a, z) - \frac{1}{2} (B_1 z, z) + \int \left(e^{i(z, x)} - 1 - \frac{i(z, x)}{1 + |x|^2} + \right. \\ \left. + \frac{1}{2} \frac{(z, x)^2}{1 + |x|^2} \right) \frac{1 + |x|^2}{|x|^2} \bar{G}(dx) = i(a, z) - \frac{1}{2} (Bz, z) + \\ + \int \left(e^{i(z, x)} - 1 - \frac{i(z, x)}{1 + |x|^2} \right) \frac{1 + |x|^2}{|x|^2} \bar{G}(dx), \end{aligned}$$

где

$$(Bz, z) = (B_1 z, z) - \int \frac{(z, x)^2}{|x|^2} \bar{G}(dx).$$

Покажем, что B — положительный оператор. Так как $|\varphi(t, z)| = \exp\{t \operatorname{Re} K(z)\} \leq 1$, то $\operatorname{Re} K(z) \leq 0$. Но

$$\operatorname{Re} K(z) = \frac{1}{2} (Bz, z) + \int (\cos(z, x) - 1) \frac{1 + |x|^2}{|x|^2} \bar{G}(dx).$$

Поскольку второе слагаемое ограничено, а (Bz, z) однородно по z , то

$$-\frac{1}{2} (Bz, z) = \lim_{\lambda \rightarrow \infty} \frac{1}{\lambda^2} \operatorname{Re} K(\lambda, z) \leq 0.$$

Таким образом, доказана следующая теорема.

Теорема. Для всякого однородного стохастически непрерывного процесса с независимыми приращениями $\xi(t)$ со значениями в R^m существуют: 1) вектор $a \in R^m$, 2) неотрицательный симметричный оператор B в R^m , 3) конечная мера на борелевских множествах в R^m $\bar{G}(dx)$, для которой $\bar{G}(\{0\}) = 0$, такие, что характеристическая функция процесса $\varphi(t, z)$ имеет вид

$$\begin{aligned} \varphi(t, z) = \exp \left\{ t \left[i(a, z) - \frac{1}{2} (Bz, z) + \right. \right. \\ \left. \left. + \int \left(e^{i(z, x)} - 1 - \frac{i(z, x)}{1 + |x|^2} \right) \frac{1 + |x|^2}{|x|^2} \bar{G}(dx) \right] \right\}. \end{aligned} \quad (7)$$

Следствие. Для стохастически непрерывного однородного процесса в R^1 существуют такие $a \in R^1$, $b > 0$ и неубывающая

функция ограниченного вариации \bar{G} на R^1 , что характеристическая функция процесса имеет вид

$$\varphi(t, z) = \exp \left\{ t \left[i(a, z) - \frac{1}{2} bz^2 + \int \left(e^{izx} - 1 - \frac{izx}{1+x^2} \right) \frac{1+x^2}{x^2} d\bar{G}(x) \right] \right\}. \quad (8)$$

Замечание. Пусть $G(x)$ — функция,

$$\bar{G}(x) + b\varepsilon(x),$$

где $\varepsilon(x) = 0$ при $x < 0$, $\varepsilon(x) = 1$ при $x > 0$. Если считать, что функция

$$\left(e^{i(z, x)} - 1 - \frac{izx}{1+x^2} \right) \frac{1+x^2}{x^2}$$

доопределена в точке $x = 0$ по непрерывности, т. е. её значение есть $-\frac{z^2}{2}$, то

8) можно записать так:

$$\varphi(t, z) = \exp \left\{ t \left[iaz + \int \left(e^{izx} - 1 - \frac{izx}{1+x^2} \right) \frac{1+x^2}{x^2} dG(x) \right] \right\}. \quad (9)$$

§ 2. Обобщенный процесс Пуассона

Рассмотрим однородный процесс с независимыми приращениями в R^m , для которого выполнены следующие условия:

- 1) $\int \frac{1+|x|^2}{|x|^2} \bar{G}(dx) < \infty$,
- 2) $B = 0$,
- 3) $(a, z) = \int \frac{(z, x)}{|x|^2} \bar{G}(dx)$ при $z \in R^m$,

здесь a , B , \bar{G} — величины, входящие в формулу для характеристической функции процесса (7), § 1. В этом случае

$$\varphi(t, z) = \exp \left\{ tc \int (e^{i(z, x)} - 1) F(dx) \right\}, \quad (1)$$

где

$$c = \int \frac{1+|x|^2}{|x|^2} \bar{G}(dx),$$

$$F(A) = \frac{1}{c} \int_A \frac{1+|x|^2}{|x|^2} \bar{G}(dx)$$

есть распределение вероятностей в R^m . Однородный процесс с независимыми приращениями с характеристической функцией (1) называется обобщенным процессом Пуассона. Это название объясняется следующими соображениями. Предположим сначала, что процесс происходит в R^1 и мера $F(dx)$ сосредоточена в точке 1. Тогда

$$\varphi(t, z) = \exp \{ tc (e^{iz} - 1) \}. \quad (2)$$

Следовательно, приращение процесса $\xi(t+h) - \xi(t)$ имеет распределение Пуассона с параметром tc , так что процесс $\xi(t)$ есть однородный процесс Пуассона с параметром c (см. § 7, гл. II). Покажем теперь, как с помощью процесса Пуассона построить процесс с характеристической функцией (1). Пусть $v(t)$ — процесс Пуассона, для которого характеристическая функция задается правой частью (2). Обозначим через $\xi_1, \xi_2, \dots, \xi_n, \dots$ последовательность одинаково распределенных случайных величин в R^m , имеющих распределение $F(dx)$ и в совокупности не зависящих от процесса $v(t)$. Рассмотрим случайный процесс

$$\eta(t) = \sum_{k=1}^{v(t)} \xi_k, \quad \left(\sum_{k=1}^0 = 0 \right). \quad (3)$$

Это процесс кусочно-постоянный, его скачки происходят в тех точках, где происходят скачки $v(t)$, при этом величина k -го скачка есть ξ_k . Покажем, что $\eta(t)$ — процесс с независимыми приращениями. Для этого найдем совместную характеристическую функцию величин

$$\eta(t_1), \eta(t_2) - \eta(t_1), \dots, \eta(t_n) - \eta(t_{n-1})$$

$$M \exp \{ i(z_1, \eta(t_1)) + i(z_2, \eta(t_2) - \eta(t_1)) + \dots + i(z_n, \eta(t_n) - \eta(t_{n-1})) \}.$$

Подсчитаем сначала условное математическое ожидание при условии, что $v(t_1) = k_1, v(t_2) - v(t_1) = k_2, \dots, v(t_n) - v(t_{n-1}) = k_n$. Тогда

$$\eta(t_1) = \sum_{j=1}^{k_1} \xi_j, \quad \eta(t_2) - \eta(t_1) = \sum_{j=k_1+1}^{k_1+k_2} \xi_j, \quad \dots,$$

$$\eta(t_n) - \eta(t_{n-1}) = \sum_{j=k_1+\dots+k_{n-1}+1}^{k_1+\dots+k_n} \xi_j.$$

Воспользовавшись тем, что ξ_j не зависят от v_t , и независимостью ξ_k , можно записать

$$M(\exp \{ i(z_1, \eta(t_1)) + \dots + i(z_n, \eta(t_n) - \eta(t_{n-1})) \} / v(t_1) = k_1, \dots, v(t_n) - v(t_{n-1}) = k_n) = (M \exp \{ i(z_1, \xi_1) \})^{k_1} \dots$$

$$\dots (M \exp \{ i(z_n, \xi_1) \})^{k_n}.$$

Пусть $\psi(z) = \int \exp \{ i(z, x) \} F(dx)$ — характеристическая функция величины ξ_1 . Поскольку

$$P \{ v(t_1) = k_1, \dots, v(t_n) - v(t_{n-1}) = k_{n-1} \} =$$

$$= \frac{(ct_1)^{k_1}}{k_1!} e^{-ct_1} \dots \frac{[c(t_n - t_{n-1})]^{k_n}}{k_n!} e^{-c(t_n - t_{n-1})},$$

то на основании формулы полного математического ожидания

$$\begin{aligned} M \exp \{i(z_1, \eta(t_1)) + \dots + i(z_n, \eta(t_n) - \eta(t_{n-1}))\} = \\ = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \dots \sum_{k_n=0}^{\infty} \frac{(ct_1)^{k_1}}{k_1!} \psi^{k_1}(z_1) e^{-ct_1} \frac{(c(t_2 - t_1))^{k_2}}{k_2!} \psi^{k_2}(z_2) e^{-c(t_2 - t_1)} \times \\ \times \dots \frac{(c(t_n - t_{n-1}))^{k_n}}{k_n!} \psi^{k_n}(z_n) e^{-c(t_n - t_{n-1})} = \\ = e^{ct_1(\psi(z_1) - 1)} e^{c(t_2 - t_1)(\psi(z_2) - 1)} \dots e^{c(t_n - t_{n-1})(\psi(z_n) - 1)}. \end{aligned}$$

Из этой формулы вытекает, что $\eta(t)$ — процесс с независимыми приращениями и что характеристическая функция величины $\eta(t_2) - \eta(t_1)$ задается выражением

$$e^{c(t_2 - t_1)(\psi(z) - 1)},$$

т. е. распределение $\eta(t_2) - \eta(t_1)$ зависит лишь от $t_2 - t_1$. Поэтому процесс $\eta(t)$ является однородным процессом с независимыми приращениями. Характеристическая функция этого процесса такова:

$$M \exp \{i(z, \eta(t_1))\} = \exp \{ct(\psi(z) - 1)\} = \exp \left\{ ct \int (e^{i(z, x)} - 1) F(dx) \right\},$$

т. е. имеет вид (1).

Покажем, что для всякого обобщенного процесса Пуассона существует процесс $\eta^*(t)$ вида (3), для которого

$$P\{\xi(t) = \eta^*(t)\} = 1$$

каково бы ни было $t \geq 0$. (Два процесса $\xi(t)$ и $\eta^*(t)$, для которых выполнено указанное соотношение, называются стохастически эквивалентными). При построении процесса $\eta^*(t)$ будем использовать тот факт, что конечномерные распределения процессов $\xi(t)$ и $\eta(t)$ совпадают, так как оба эти процесса — однородные процессы с независимыми приращениями, имеющие одинаковые распределения приращений. Определим величины $\xi_1^{(n)} = \xi\left(\frac{t_1}{2^n}\right)$, если $\xi\left(\frac{j}{2^n}\right) = 0$ при $j \leq l_1 - 1$, $\xi\left(\frac{l_1}{2^n}\right) \neq 0$; $\xi_2^{(n)} = \xi\left(\frac{l_2}{2^n}\right) - \xi\left(\frac{l_1}{2^n}\right)$, если $\xi\left(\frac{j}{2^n}\right) - \xi\left(\frac{l_1}{2^n}\right) = 0$ при $l_1 < j \leq l_2 - 1$, $\xi\left(\frac{l_2}{2^n}\right) - \xi\left(\frac{l_1}{2^n}\right) \neq 0$; $\xi_k^{(n)} = \xi\left(\frac{l_k}{2^n}\right) - \xi\left(\frac{l_{k-1}}{2^n}\right)$, если $\xi\left(\frac{j}{2^n}\right) - \xi\left(\frac{l_{k-1}}{2^n}\right) = 0$ при $l_{k-1} < j \leq l_k - 1$, $\xi\left(\frac{l_k}{2^n}\right) - \xi\left(\frac{l_{k-1}}{2^n}\right) \neq 0$ и т. д.

Обозначим $\frac{l_k}{2^n} = \tau_k^{(n)}$. Построим аналогичные величины по процессу $\eta(t)$: $\bar{\xi}_1^{(n)}, \dots, \bar{\xi}_k^{(n)}, \bar{\tau}_1^{(n)}, \dots, \bar{\tau}_k^{(n)}$. Очевидно, совместные распределения величин

$$\begin{aligned} \{(\xi_k^{(n)}, \tau_k^{(n)}), k = 1, \dots; n = 1, 2, \dots\}, \\ \{(\bar{\xi}_k^{(n)}, \bar{\tau}_k^{(n)}), k = 1, \dots; n = 1, 2, \dots\} \end{aligned} \quad (4)$$

одинаковы в силу совпадения распределений у процессов $\xi(t)$ и $\eta(t)$. Из вида процесса $\eta(t)$ легко видеть, что $\xi_k^{(n)}$ при $n \rightarrow \infty$ сходятся по вероятности к ξ_k^* , а $\tau_k^{(n)}$ — к величине τ_k^* — моменту k -го скачка процесса $\nu(t)$. Поэтому и величины $\xi_k^{(n)}$ и $\tau_k^{(n)}$ сходятся по вероятности к некоторым пределам. Обозначим их соответственно через ξ_k^* и τ_k^* . Из одинаковой распределенности групп величин (4) вытекает, что совместное распределение (ξ_k^*, τ_k^*) , $k = 1, 2, \dots$ такое же, как и совместное распределение величин (ξ_k, τ_k) , $k = 1, 2, \dots$. Поэтому две группы величин $\{\xi_k^*, k = 1, 2, \dots\}$, $\{\tau_k^*, k = 1, 2, \dots\}$ независимы, $\xi_1^*, \xi_2^*, \dots, \xi_k^*, \dots$ одинаково распределены и имеют функцию распределения $F(dx)$, а $\tau_1^*, \tau_2^*, \dots, \tau_k^*, \dots$ — моменты скачков однородного процесса Пуассона $\nu^*(t)$, характеристическая функция которого задается правой частью равенства (2). Построим процесс

$$\eta^*(t) = \sum_{k=1}^{\nu^*(t)} \xi_k^*.$$

Согласно построению величин $\xi_k^{(n)}$ и $\eta_k^{(n)}$ справедливо соотношение

$$\sum_{\tau_i^{(n)} < \frac{t}{2^n}} \xi_i^{(n)} = \xi\left(\frac{t}{2^n}\right)$$

для всех t . Поэтому если $\frac{t}{2^n}$ не является значением ни одной из величин $\tau_1^*, \tau_2^*, \dots$ (это происходит с вероятностью 1, так как $\tau_1^*, \tau_2^*, \dots$ имеют непрерывное распределение: τ_k^* есть сумма показательно распределенных случайных величин), то

$$\sum_{\tau_i^{(n)} < \frac{t}{2^n}} \xi_i^{(n)} \rightarrow \sum_{\tau_i^* < \frac{t}{2^n}} \xi_i^* = \eta^*\left(\frac{t}{2^n}\right).$$

Следовательно,

$$P\left\{\xi\left(\frac{t}{2^n}\right) = \eta^*\left(\frac{t}{2^n}\right)\right\} = 1,$$

т. е. $P\{\xi(t) = \eta^*(t)\} = 1$ для всех двоично рациональных t . Используя стохастическую непрерывность обоих процессов, убеждаемся, что последнее равенство выполнено для всех t , т. е. что процессы $\xi(t)$ и $\eta^*(t)$ стохастически эквивалентны.

§ 3. Процесс броуновского движения

Рассмотренный в предыдущем параграфе процесс Пуассона можно рассматривать как непрерывный аналог случайного блуждания — случайного блуждания с непрерывным временем. Однако

амо движение скачкообразно и его «непрерывность» (как случайного блуждания) заключается в том, что скачки могут происходить в любой момент времени и поэтому $\xi(t)$ как функция t оказывается стохастически непрерывной. В то же время имеются примеры движения, в которых положение меняется непрерывно, а само движение достаточно точно описывается процессом независимыми приращениями. Простейшим примером может служить броуновское движение.

Рассмотрим частицу, которая движется в некоторой среде (жидкости или газе) под влиянием взаимодействий с частицами среды. Если среда однородна, то движение частицы из каждой точки среды в какой бы момент мы туда не попали, будет одинаковым в вероятностном смысле, т. е. смещение частицы за время h $\xi(t+h) - \xi(t)$ не должно зависеть ни от поведения до момента t , ни от момента t . Независимость движения после момента t от скорости в момент t противоречит, на первый взгляд, законам механики, однако на самом деле за очень малое время скорость так часто меняет свое направление и величину, что частица успевает «забыть» значение скорости в исходный момент времени. Очевидно также, что траектория движущейся частицы $\xi(t)$ — непрерывная функция t . Непрерывный процесс с независимыми приращениями, на основании приведенных только что соображений, называется процессом броуновского движения. Мы рассмотрим однородное по времени броуновское движение, т. е. однородный непрерывный процесс с независимыми приращениями.

Естественно попытаться построить непрерывный процесс из обобщенного процесса Пуассона с помощью предельного перехода, при котором число скачков процесса будет расти, а сами скачки будут уменьшаться в размере. Пусть $\nu_c(t)$ — пуассоновский процесс с параметром c , а $\xi_1^{(c)}, \dots, \xi_k^{(c)}, \dots$ — последовательность независимых одинаково распределенных величин, не зависящих от $\nu_c(t)$ и имеющих распределение $F^{(c)}(dx)$. Положим

$$\xi_c(t) = \sum_{k=1}^{\nu_c(t)} \xi_k^{(c)}. \quad (1)$$

Исследуем поведение этого процесса при $c \rightarrow \infty$ (c — среднее число скачков в единицу времени, по предположению должно неограниченно расти). Величины $\xi_k^{(c)}$ при этом стремятся к нулю. Вероятность того, что $\xi_c(t)$ на отрезке $[0, t]$ будет иметь максимальный скачок, не превосходящий $\varepsilon > 0$, будет

$$\begin{aligned} & \sum_{n=0}^{\infty} P \{ \max_{k \leq n} |\xi_k^{(c)}| \leq \varepsilon \} P \{ \nu_c(t) = n \} = \\ & = \sum_{n=0}^{\infty} (F^{(c)} \{x: |x| \leq \varepsilon\})^n \cdot \frac{(ch)^n e^{-ch}}{n!} = \exp \{ hc [F^{(c)} \{x: |x| \leq \varepsilon\} - 1] \}. \end{aligned}$$

Для того чтобы эта вероятность стремилась к нулю, нужно чтобы для всякого $\varepsilon > 0$

$$\lim_{c \rightarrow \infty} c F^{(c)} \{x: |x| > \varepsilon\} = 0. \quad (2)$$

Предположим, что можно указать такое ε_c , $\varepsilon_c \rightarrow 0$ при $c \rightarrow \infty$, что $P\{|\xi_j^{(c)}| > \varepsilon_c\} = 0$. Тогда соотношение (2) выполнено для всех $\varepsilon > 0$. Характеристическая функция $\varphi_c(t, x)$ процесса $\xi_c(t)$ имеет вид

$$\varphi_c(t, z) = \exp \left\{ ct \int (e^{i(z, x)} - 1) F^{(c)}(dx) \right\}.$$

Если существует предел $\varphi_c(t, z)$ при $c \rightarrow \infty$, то существует

$$\lim_{c \rightarrow \infty} c \int (e^{i(z, x)} - 1) F^{(c)}(dx). \quad (3)$$

Так как

$$\begin{aligned} \int_{|x| < \varepsilon_c} (e^{i(z, x)} - 1) F^{(c)}(dx) &= \int_{|x| < \varepsilon_c} (e^{i(z, x)} - 1) F^{(c)}(dx) = \\ &= i \int (z, x) F^{(c)}(dx) - \frac{1}{2} \int (z, x)^2 F^{(c)}(dx) + \\ &+ \int \left(e^{i(z, x)} - 1 - i(z, x) + \frac{1}{2} (z, x)^2 \right) F^{(c)}(dx), \end{aligned}$$

а

$$\begin{aligned} \left| \int \left(e^{i(z, x)} - 1 - i(z, x) + \frac{1}{2} (z, x)^2 \right) F^{(c)}(dx) \right| &\leq \\ &\leq \frac{1}{6} \int |z, x|^3 F^{(c)}(dx) \leq \frac{|z|}{6} \varepsilon_c \int (z, x)^2 F^{(c)}(dx), \end{aligned}$$

то

$$\begin{aligned} \lim_{c \rightarrow \infty} c \int (e^{i(z, x)} - 1) F^{(c)}(dx) &= \lim_{c \rightarrow \infty} \left[ci \int (z, x) F^{(c)}(dx) - \right. \\ &\left. - \frac{1}{2} c \int (z, x)^2 F^{(c)}(dx) (1 + O(\varepsilon_c)) \right]. \end{aligned}$$

Поскольку $O(\varepsilon_c) \rightarrow 0$, то последний предел существует, если существует предел вещественной и мнимой частей. Последний будет линейная функция от z , а первый — квадратическая. Таким образом, предел (3) имеет вид

$$i(a, z) - \frac{1}{2} (Bz, z),$$

где $a \in R^m$, B — неотрицательный симметричный оператор в R^m , а характеристическая функция предельного процесса будет

$$\varphi(t, z) = \exp \left\{ t \left[i(a, z) - \frac{1}{2} (Bz, z) \right] \right\}. \quad (4)$$

т. е. она получается из общей формулы (7), § 1, при $\bar{G}(R^m) = 0$. Покажем, что это справедливо для всякого непрерывного процесса.

Однородный процесс с независимыми приращениями, имеющий характеристическую функцию вида (4), называется однородным по времени броуновским движением в R^m . Если в формуле (4) положить $a=0$, $B=I$, где I — единичный оператор, то процесс называется винеровским.

В частности, одномерный винеровский процесс — это однородный процесс с независимыми приращениями, для которого характеристическая функция имеет вид

$$M \exp \{i z w(t)\} = e^{-(t/2)z^2}, \quad z \in R^1, \quad (5)$$

приращение $w(t+h) - w(t)$ имеет нормальное распределение с математическим ожиданием 0 и дисперсией t . Если имеем броуновское движение с характеристической функцией (4), а $z \in R^m$ таково, что $(Bz, z) > 0$, то процесс

$$w_z(t) = \frac{(\xi(t), z) - t(a, z)}{\sqrt{(Bz, z)}} \quad (6)$$

будет одномерным винеровским процессом. Теперь мы покажем, что броуновское движение есть непрерывный процесс в следующем смысле: существует такой стохастически эквивалентный непрерывный процесс $\xi^*(t)$, т. е.

$$P\{\xi(t) = \xi^*(t)\} = 1, \quad t \geq 0.$$

Используя преобразование (6), можно ограничиться случаем одномерного винеровского процесса.

Теорема 2. Для одномерного винеровского процесса $w(t)$ существует стохастически эквивалентный непрерывный процесс $w^*(t)$.

Доказательство. Нам потребуется следующее вспомогательное утверждение: если ξ_1, \dots, ξ_n — независимые симметричные величины, то

$$P\left(\sup_{k \leq n} (\xi_1 + \dots + \xi_k) > a\right) \leq 2P\{\xi_1 + \dots + \xi_n > a\}. \quad (7)$$

Действительно,

$$\begin{aligned} \frac{1}{2} P\{\xi_1 \leq a, \dots, \xi_1 + \dots + \xi_{k-1} < a, \xi_1 + \dots + \xi_k > a\} < \\ < P\{\xi_1 < a, \dots, \xi_1 + \dots + \xi_{k-1} < a, \xi_1 + \dots + \xi_k > a, \\ \xi_{k+1} + \dots + \xi_n \geq 0\} < P\{\xi_1 < a, \dots, \xi_1 + \dots + \xi_{k-1} < a, \\ \xi_1 + \dots + \xi_k > a, \xi_1 + \dots + \xi_n > a\}. \end{aligned}$$

(Поскольку $\xi_{k+1} + \dots + \xi_n$ симметричная величина, то $P\{\xi_{k+1} + \dots + \xi_n \geq 0\} \geq \frac{1}{2}$). Суммируя крайние члены неравенства по k , получим

$$\begin{aligned} \frac{1}{2} P\left(\sup_{k \leq n} (\xi_1 + \dots + \xi_k) > a\right) < P\left\{\left(\sup_{k \leq n} (\xi_1 + \dots + \xi_k) > a\right) \cap \right. \\ \left. \cap \{(\xi_1 + \dots + \xi_n) > a\}\right\} < P\{\xi_1 + \dots + \xi_n > a\}. \end{aligned}$$

Так как величины $w\left(\frac{k}{2^n}\right) - w\left(\frac{k-1}{2^n}\right)$ независимы и симметричны, то

$$\mathbf{P}\left\{\sup_{k < j < l} \left(w\left(\frac{j}{2^n}\right) - w\left(\frac{k}{2^n}\right)\right) > a\right\} \leq 2\mathbf{P}\left\{\left[w\left(\frac{l}{2^n}\right) - w\left(\frac{k}{2^n}\right)\right] > a\right\}.$$

Для любых двух двоично рациональных чисел $\frac{k}{2^n}$ и $\frac{l}{2^n}$ имеем

$$\begin{aligned} & \mathbf{P}\left\{\sup_{m > n} \sup_{\substack{\frac{k}{2^n} < \frac{j}{2^m} < \frac{l}{2^n}}} \left(w\left(\frac{j}{2^m}\right) - w\left(\frac{k}{2^n}\right)\right) > a\right\} = \\ & = \lim_{m \rightarrow \infty} \mathbf{P}\left\{\sup_{\substack{\frac{k}{2^n} < \frac{j}{2^m} < \frac{l}{2^n}}} \left(w\left(\frac{j}{2^m}\right) - w\left(\frac{k}{2^n}\right)\right) > a\right\} \leq \\ & \leq 2\mathbf{P}\left\{w\left(\frac{l}{2^n}\right) - w\left(\frac{k}{2^n}\right) > a\right\}. \end{aligned}$$

Пусть $\mathcal{D} \subset [0, \infty)$ множество всех двоично рациональных чисел. Тогда при $t, s \in \mathcal{D}$

$$\mathbf{P}\left\{\sup_{t < u < s, u \in \mathcal{D}} (w(u) - w(t)) > a\right\} \leq 2\mathbf{P}\{w(s) - w(t) > a\}.$$

Используя симметричность $w(t)$, можно записать

$$\begin{aligned} \mathbf{P}\left\{\inf_{\substack{t < u < s \\ u \in \mathcal{D}}} (w(u) - w(t)) < -a\right\} &= \mathbf{P}\left\{\sup_{t < u < s} (-w(u) + w(t)) > a\right\} \leq \\ &\leq 2\mathbf{P}\{w(t) - w(s) > a\} = 2\mathbf{P}\{w(s) - w(t) > a\}. \end{aligned}$$

Следовательно,

$$\mathbf{P}\left\{\sup_{\substack{t < u < s \\ u \in \mathcal{D}}} |w(u) - w(t)| > a\right\} \leq 4\mathbf{P}\{w(s) - w(t) > a\}.$$

Положим для $h \in \mathcal{D}$

$$\eta(T, h) = \sup_{\substack{t, s \in \mathcal{D} \cap [0, T] \\ |t-s| < h}} |w(t) - w(s)|.$$

Эта величина конечна, так как

$$\eta(T, h) \leq 2 \sup_{kh < T} \sup_{kh < u < (k+2)h} |w(u) - w(kh)|.$$

Из этого неравенства вытекает, что

$$\begin{aligned} \mathbf{P}\{\eta(T, h) > a\} &\leq \sum_{kh < T} \mathbf{P}\left\{\sup_{kh < u < (k+2)h} |w(u) - w(kh)| > \frac{a}{2}\right\} \leq \\ &\leq \frac{4T}{h} \mathbf{P}\left\{w(2h) > \frac{a}{2}\right\}. \end{aligned}$$

В частности,

$$\begin{aligned} P\left\{\eta\left(T, \frac{1}{2^n}\right) > \frac{1}{n}\right\} &\leq 4 \cdot 2^n TP\left\{\omega\left(\frac{1}{2^n}\right) > \frac{1}{2n}\right\} \leq \\ &\leq 4 \cdot 2^n TP\left\{\left|\omega\left(\frac{1}{2^{n-1}}\right)\right| > \frac{1}{2^n}\right\} \leq 4 \cdot 2^n \frac{M\left|\omega\left(\frac{1}{2^n}\right)\right|^4}{\left(\frac{1}{2^n}\right)^4} = \\ &= 2^{n+6} T n^4 3 \left(\frac{1}{2^{n-1}}\right)^2 = \frac{2^8 \cdot 3 \cdot T n^4}{2^n} \end{aligned}$$

(мы воспользовались тем, что $\omega(h)$ — гауссовская величина со средним 0 и дисперсией h , так что $M|\omega(h)|^4 = 3h^2$). Поскольку для всякого $T > 0$

$$\sum P\left\{\eta\left(T, \frac{1}{2^n}\right) > \frac{1}{n}\right\} < \infty,$$

то на основании теоремы Бореля — Кантелли с вероятностью 1

$$\eta\left(T, \frac{1}{2^n}\right) \rightarrow 0.$$

Таким образом, $\omega(t)$ с вероятностью 1 равномерно непрерывно, но $[0, T] \cap \mathcal{D}$, каково бы ни было T . Пусть $\omega^*(t) = \omega(t)$ при $t \in \mathcal{D}$, $\omega^*(t) = \lim_{s \in \mathcal{D}, s \rightarrow t} \omega(s)$. Тогда $\omega^*(t)$ непрерывный процесс. Для $t \in \mathcal{D}$, $P\{\omega(t) = \omega^*(t)\} = 1$. Используя стохастическую непрерывность обоих процессов, убеждаемся, что $\omega^*(t)$ стохастически эквивалентен $\omega(t)$.

Глава VIII

КОРРЕЛЯЦИОННАЯ ТЕОРИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ

§ 1. Элементы анализа в L_2

Непрерывность и дифференцируемость случайного процесса. В предыдущих разделах мы рассматривали примеры случайных процессов (процесс Пуассона, марковские процессы, процессы с независимыми приращениями).

Случайным процессом $\xi(t)$ называют семейство случайных величин, зависящих от параметра t , $t \in [a, b]$ или $t \in (a, b)$, $-\infty \leq a < b \leq +\infty$, который интерпретируется как время. В настоящей главе рассмотрены процессы, принимающие, вообще говоря, комплексные значения, для которых $M|\xi(t)|^2 < \infty$, $\forall t \in [a, b]$. Значение процесса в момент времени t мы интерпретируем, как элемент пространства $L_2 = L_2(\Omega, \mathcal{F}, P)$. С этой точки зрения

процесс $\xi(t)$, $t \in [a, b]$ можно рассматривать как некоторую кривую в пространстве L_2 .

Пусть $\{\xi_\lambda, \lambda \in \Lambda\}$ — некоторое семейство случайных величин, $\xi_\lambda \in L_2$.

Определение. *Функция*

$$B(\lambda, \lambda') = M \xi_\lambda \bar{\xi}_{\lambda'} = (\xi_\lambda, \xi_{\lambda'})$$

называется *ковариацией семейства* (ξ_λ) .

Очевидно, ковариация обладает свойствами

$$B(\lambda, \lambda) = M |\xi_\lambda|^2 \geq 0, \quad B(\lambda, \lambda') = \overline{B(\lambda', \lambda)}, \quad (1)$$

$$|B(\lambda, \lambda')|^2 \leq B(\lambda, \lambda) B(\lambda', \lambda'), \quad (2)$$

$$\sum_{k,r=1}^n c_k \bar{c}_r B(\lambda_k, \lambda_r) \geq 0. \quad (3)$$

Неравенство (2) — это неравенство Коши—Буняковского, а (3) вытекает из того, что

$$\sum_{k,r} c_k \bar{c}_r B(\lambda_k, \lambda_r) = M \left(\sum_k c_k \xi_{\lambda_k} \overline{\sum_r c_r \xi_{\lambda_r}} \right).$$

Функции $B(\lambda, \lambda')$, $\lambda, \lambda' \in \Lambda$, удовлетворяющие (3) при любых n , комплексных c_1, \dots, c_n и $\lambda_1, \dots, \lambda_n$ из Λ , называют положительно определенными ядрами на Λ .

Определение. *Корреляционной функцией семейства случайных величин* $(\xi_\lambda, \lambda \in \Lambda)$ называют функцию

$$R(\lambda, \lambda') = M (\xi_\lambda - m(\lambda)) \overline{(\xi_{\lambda'} - m(\lambda'))},$$

где $m(\lambda) = M \xi_\lambda$ — среднее значение семейства (ξ_λ) .

В настоящем параграфе приведены определения и условия непрерывности, дифференцируемости и интегрируемости случайного процесса, соответствующие топологии пространства L_2 , выраженные с помощью ковариации процесса.

Пусть на множестве Λ определена положительная функция $g(\lambda)$, $g(\lambda) > 0$, такая, что $\inf (g(\lambda), \lambda \in \Lambda) = 0$.

Определение. *Пределом в среднеквадратическом при* $g(\lambda) \rightarrow 0$ *семейства случайных величин* $\{\xi_\lambda, \lambda \in \Lambda\}$, $\xi_\lambda \in L_2$ *называют случайную величину* η , $\eta \in L_2$, *такую, что для любого* $\varepsilon > 0$ *найдется* $\delta = \delta(\varepsilon) > 0$, *такое, что* $M |\eta - \xi_\lambda|^2 < \varepsilon$ *для всех* λ , *для которых* $g(\lambda) < \delta$.

Если η — среднеквадратический предел семейства (ξ_λ) , то пишут $\xi_\lambda \xrightarrow{L_2} \eta$ при $g(\lambda) \rightarrow 0$, или $\eta = \lim_{g(\lambda) \rightarrow 0} \xi_\lambda$. Так же, как и для чисел доказываем, что для того чтобы η было среднеквадратическим пределом семейства (ξ_λ) , необходимо и достаточно, чтобы для любой последовательности λ_n , для которой $g(\lambda_n) \rightarrow 0$, $M |\eta - \xi_{\lambda_n}|^2 \rightarrow 0$ (определение предела по Гейне).

Из полноты пространства L_2 следует также, что для существования среднеекватрического предела семейства (ξ_λ) необходимо и достаточно, чтобы

$$M|\xi_\lambda - \xi_{\lambda'}|^2 \rightarrow 0 \text{ при } g(\lambda) + g(\lambda') \rightarrow 0.$$

Лемма. Для существования среднеекватрического предела η семейства (ξ_λ) необходимо и достаточно, чтобы существовал предел

$$\lim_{g(\lambda)+g(\lambda') \rightarrow 0} B(\lambda, \lambda') = B_0.$$

Если этот предел существует, то $B_0 = M|\eta|^2$.

Доказательство. Необходимость. Если $\eta = \text{l. i. m. } \xi_\lambda$, то $g(\lambda) \rightarrow 0$

$$\begin{aligned} B(\lambda, \lambda') - M|\eta|^2 &= M\xi_\lambda \bar{\xi}_{\lambda'} - M\eta \bar{\eta} = \\ &= M(\xi_\lambda - \eta)(\bar{\xi}_{\lambda'} - \bar{\eta}) + M\eta(\bar{\xi}_{\lambda'} - \bar{\eta}) + M(\xi_\lambda - \eta)\bar{\eta}, \end{aligned}$$

отсюда следует

$$\begin{aligned} |B(\lambda, \lambda') - M|\eta|^2| &\leq (M|\xi_\lambda - \eta|^2 \cdot M|\xi_{\lambda'} - \eta|^2)^{1/2} + \\ &+ (M|\eta|^2 \cdot M|\xi_{\lambda'} - \eta|^2)^{1/2} + (M|\xi_\lambda - \eta|^2 \cdot M|\eta|^2)^{1/2} \rightarrow 0 \end{aligned}$$

при $g(\lambda) \rightarrow 0$.

Достаточность. Из существования предела $B(\lambda, \lambda')$ при $g(\lambda) + g(\lambda') \rightarrow 0$ следует

$$M|\xi_\lambda - \xi_{\lambda'}|^2 = B(\lambda, \lambda) - B(\lambda, \lambda') - B(\lambda', \lambda) + B(\lambda', \lambda') \rightarrow 0,$$

откуда вытекает существование л. i. m. $\xi_\lambda = \eta$.

Из первой части леммы получаем $M|\eta|^2 = B_0$.

Определение. Случайный процесс $\{\xi(t), t \in (a, b)\}$, $\xi(t) \in L_2$ называют:

а) непрерывным в среднеекватрическом в точке t , если

$$\text{l. i. m. } \xi(t') = \xi(t),$$

б) дифференцируемым в среднеекватрическом, если существует предел

$$\text{l. i. m. } \frac{\xi(t+h) - \xi(t)}{h} = \xi'(t).$$

Случайную величину $\xi'(t)$, когда она существует, называют среднеекватрической производной процесса $\xi(t)$ в точке t . Если процесс непрерывен в среднеекватрическом, дифференцируем в среднеекватрическом в каждой точке интервала (a, b) , то его называют непрерывным в среднеекватрическом (в среднеекватрическом дифференцируемым) на (a, b) .

Очевидны те изменения в определениях непрерывности в среднеквадратическом в точке t слева (или справа), или в определении дифференцируемости в точке t слева (или справа), а также в определениях непрерывности в среднеквадратическом (среднеквадратической дифференцируемости) процесса на отрезке $[a, b]$.

Теорема 1. а) Для непрерывности в среднеквадратическом процесса $\xi(t)$ в точке t необходимо и достаточно, чтобы

$$\square B = \square_{(t', t)} B(t, t) = B(t', t') - B(t', t) - B(t, t') + B(t, t) \rightarrow 0$$

при $t' \rightarrow t$.

б) для среднеквадратической дифференцируемости процесса $\xi(t)$ в точке t необходимо и достаточно, чтобы существовал предел

$$1. \text{ i. m. } \frac{\square_{(t', t'')} B(t, t)}{(t' - t)(t'' - t)} = \left. \frac{\partial^2 B(t, s)}{\partial t \partial s} \right|_{s=t}, \quad (4)$$

где

$$\square_{(t', t'')} B(t, t) = B(t', t'') - B(t', t) - B(t, t'') + B(t, t).$$

Доказательство. а) очевидно,

$$\mathbf{M} |\xi(t') - \xi(t)|^2 = B(t', t') - B(t', t) - B(t, t') + B(t, t) = \square B;$$

б) непосредственно вытекает из леммы 1, так как

$$\mathbf{M} \frac{\xi(t') - \xi(t)}{t' - t} \frac{\xi(t'') - \xi(t)}{t'' - t} = \frac{\square_{(t', t'')} B(t, t)}{(t' - t)(t'' - t)}.$$

Замечание. Если в некоторой окрестности точки (t, t) существует смешанная производная $\frac{\partial^2 B(t, s)}{\partial t \partial s}$, то

$$\lim_{t' \rightarrow t, t'' \rightarrow t} \frac{\square_{(t', t'')} B(t, t)}{(t' - t)(t'' - t)} = \left. \frac{\partial^2 B(t, s)}{\partial t \partial s} \right|_{s=t}.$$

Это замечание оправдывает то обозначение, которое введено в формуле (4). В общем случае предел (4), если он существует, называют обобщенной производной $\left. \frac{\partial^2 B(t, s)}{\partial t \partial s} \right|_{s=t}$.

Определение. Взаимной ковариацией (корреляционной функцией) двух процессов $\xi(t)$ и $\eta(t)$, $t \in (a, b)$, называют функцию

$$B_{\xi\eta}(t, s) = \mathbf{M} \xi(t) \overline{\eta(s)}$$

$$(R_{\xi\eta}(t, s) = \mathbf{M} (\xi(t) - m_\xi(t)) (\eta(s) - m_\eta(s))).$$

Найдем вид ковариации производного процесса $\xi'(t)$ и взаимной ковариации процессов $\xi(t)$ и $\xi'(t)$ (в предположении, что $\xi'(t)$ существует).

Пусть $\xi(t)$, $\eta(t)$ — два процесса, $t \in (a, b)$ и процесс $\xi(t)$ дифференцируем в среднеквадратическом. Так как скалярное произведение является непрерывной функцией своих аргументов, то

$$\frac{B_{\xi\eta}(t+h, s) - B_{\xi\eta}(t, s)}{h} = \left(\frac{\xi(t+h) - \xi(t)}{h}, \eta(s) \right) \rightarrow (\xi'(t), \eta(s)).$$

Это означает, что функция $B_{\xi\eta}(t, s)$ дифференцируема по t и

$$B_{\xi\eta}(t, s) = \frac{\partial B_{\xi\eta}(t, s)}{\partial t}. \quad (5)$$

Из тех же соображений вытекает, что когда процесс дифференцируемый в среднеквадратическом, то

$$B_{\xi\eta'}(t, s) = \frac{\partial^2 B_{\xi\eta}(t, s)}{\partial t \partial s}. \quad (6)$$

Процесс с ортогональными приращениями. Процесс $\xi(t)$, $t \in (a, b)$, называют процессом с ортогональными приращениями, если $\xi(t) \in L_2$ и для любых t_1, t_2, t_3, t_4 ($a < t_1 < t_2 \leq t_3 < t_4 < b$)

$$M[\xi(t_4) - \xi(t_3)][\xi(t_2) - \xi(t_1)] = 0.$$

Процесс $\xi(t)$ называют процессом с некоррелированными приращениями, если центрированный процесс $\xi_1(t) = \xi(t) - m_\xi(t)$ является процессом с ортогональными приращениями. Если $\xi(t)$ — процесс с независимыми приращениями и конечными моментами второго порядка, то он является процессом с некоррелированными приращениями.

В геометрической интерпретации процесс с ортогональными приращениями представляет собой такую кривую в L_2 , что приращения радиуса-вектора точки на кривой постоянно ортогональны между собой.

Пусть $\xi(t)$, $t \in (a, b)$, — процесс с ортогональными приращениями и для некоторого $t_0 \in (a, b)$ $\xi(t_0) \equiv 0$. Имеем при $t_0 < s < t < b$

$$B(t, s) = M \xi(t) \overline{\xi(s)} = M (\xi(t) - \xi(s)) \times \\ \times \overline{(\xi(s) - \xi(t_0))} + M |\xi(s)|^2 = M |\xi(s)|^2 = B(s, s).$$

Отсюда, учитывая, что $B(t, s) = \overline{B(s, t)}$, получаем следующее выражение для ковариации $B(t, s)$ ($t \geq t_0, s \geq t_0$):

$$B(t, s) = B(t \wedge s), \quad B(t) = B(t, t) = M |\xi(t)|^2. \quad (7)$$

Заметим, что функция $B(t)$ монотонно не убывает. Действительно, пусть $t > s > t_0$. Тогда, согласно ортогональности приращений процесса $\xi(t)$,

$$B(t) = M|\xi(t)|^2 = M|(\xi(t) - \xi(s)) + (\xi(s) - \xi(t_0))|^2 = \\ = M|\xi(t) - \xi(s)|^2 + M|\xi(s)|^2 = M|\xi(t) - \xi(s)|^2 + B(s).$$

т. е.

$$M|\xi(t) - \xi(s)|^2 = B(t) - B(s) \geq 0. \quad (8)$$

Процесс $\xi(t)$ будет непрерывным в среднее квадратическое тогда и только тогда, когда непрерывна функция $B(t)$. Кроме того, процесс $\xi(t)$ в среднее квадратическое дифференцируем в точке t тогда и только тогда, когда $B'(t) = 0$, так что, вообще говоря, процесс с ортогональными приращениями не дифференцируем в среднее квадратическое. Например, пусть $\xi(t)$ — процесс Пуассона со средним λ . $\xi_1(t) = \xi(t) - \lambda t$ — центрированный процесс Пуассона. Тогда $B_{\xi_1}(t) = M\xi_1^2(t) = \lambda t$. Этот процесс непрерывен в среднее квадратическое, но не дифференцируем в среднее квадратическое. То же самое имеет место для винеровского процесса $w(t)$, для которого $Mw(t) = 0$, $B_w(t) = Mw^2(t) = t$.

Интегрирование случайных процессов.

Определение. *Интегралом*

$$\int_a^b \xi(t) dt \quad (9)$$

от процесса $\xi(t)$, $t \in [a, b]$, $\xi(t) \in L_2$, называют предел интегральных сумм $\lim_{|\lambda| \rightarrow 0} S_\lambda$, где

$$S_\lambda = \sum_{\text{Def } \lambda} \xi(t_k) \Delta t_k.$$

Здесь λ — символ разбиения отрезка $[a, b]$ на частичные с помощью точек t_1, \dots, t_n , $a = t_0 < t_1 < \dots < t_n = b$, $\Delta t_k = t_{k+1} - t_k$, $|\lambda| = \max |\Delta t_k|$. Из леммы 1 вытекает:

для существования интеграла достаточно, чтобы функция $B(t, s)$ была интегрируема по Риману на квадрате $a \leq t, s \leq b$. При этом

$$M \left| \int_a^b \xi(t) dt \right|^2 = \int_a^b \int_a^b B(t, s) dt ds. \quad (10)$$

Действительно, ковариация сумм S_λ имеет вид

$$MS_\lambda \bar{S}_{\lambda'} = \sum_{(\alpha), (\alpha')} B(t_k, t'_r) \Delta t_k \Delta t'_r,$$

т. е. представляет собой интегральную сумму Римана, предел которой равен интегралу в правой части равенства (10).

Если процесс $\xi(t)$ задан на бесконечном отрезке $[a, \infty]$, то определим несобственный интеграл

$$\int_a^\infty \xi(t) dt = \text{l. i. m.} \int_a^b \xi(t) dt.$$

Для его существования достаточно, чтобы функция $B(t, s)$ была интегрируемой по Риману на любом квадрате $[a, b] \times [a, b]$, $b > a$ и существовал несобственный интеграл

$$\int_a^\infty \int_a^\infty B(t, s) dt ds.$$

Аналогично определяется интеграл в пределах от $-\infty$ до $+\infty$.
Рассмотрим теперь неопределенный интеграл

$$\eta(t) = \int_a^t \xi(t') dt', \quad t \in [a, b].$$

Из предыдущего легко вытекает следующее выражение для ковариации:

$$B_\eta(t, s) = \int_a^t \int_a^s B_1(t', t'') dt' dt''.$$

В частности, процесс $\eta(t)$ непрерывен в среднеквадратическом.

Теорема 2. В каждой точке t , в которой функция $B(t, t)$ непрерывна, процесс $\eta(t)$ дифференцируем в среднеквадратическом и $\eta'(t) = \xi(t)$.

Доказательство. Имеем

$$\begin{aligned} \delta(t) &= \underset{\text{Def}}{\eta(t+h) - \eta(t)} - \xi(t) = \\ &= \frac{1}{h} \int_t^{t+h} [\xi(t') - \xi(t)] dt'. \end{aligned}$$

Следовательно,

$$\begin{aligned} M|\delta(t)|^2 &= \frac{1}{h^2} \int_t^{t+h} \int_t^{t+h} M(\xi(t') - \xi(t))(\xi(t'') - \xi(t)) dt' dt'' \leq \\ &\leq \frac{1}{h^2} \left(\int_t^{t+h} (M|\xi(t') - \xi(t)|^2)^{1/2} dt' \right)^2 \leq \frac{1}{h} \int_t^{t+h} M|\xi(t') - \xi(t)|^2 dt. \end{aligned}$$

Из непрерывности функции $B(t, s)$ в точке (t, t) следует, что $M|\xi(t') - \xi(t)|^2 \rightarrow 0$ при $t' \rightarrow t$ и поэтому $\lim_{h \rightarrow 0} M|\delta(t)|^2 = 0$.

стохастические интегралы по процессу с ортогональными приращениями. Во многих вопросах бывает нужным определить понятие интеграла

$$\mathcal{I}(f) = \int_a^b f(t) d\xi(t),$$

где $\xi(t)$ — случайный процесс, аналогичное определению обычного интеграла Стильеса, или Лебега—Стильеса, так чтобы он обладал основными свойствами, присущими обычному интегралу Стильеса. Из этих свойств напомним следующие:

1) если f_i , $i = 1, 2$, — интегрируемые функции (т. е. интеграл $\mathcal{I}(f_i)$ определен и конечен), c_i — константы, то $c_1 f_1 + c_2 f_2$ интегрируемая функция и

$$\mathcal{I}(c_1 f_1 + c_2 f_2) = c_1 \mathcal{I}(f_1) + c_2 \mathcal{I}(f_2),$$

т. е. $\mathcal{I}(f)$ является линейным функционалом, определенным на линейном функциональном пространстве;

2) если $f(t) = I_{(c,d)}(t)$ — индикатор интервала (c, d) и точки c и d являются точками непрерывности функции $\xi(t)$, то

$$\mathcal{I}(I_{(c,d)}) = \xi(d) - \xi(c);$$

3) при определенных условиях можно переходить к пределу под знаком интеграла, т. е. если f_n — некоторая сходящаяся последовательность интегрируемых функций и выполнены некоторые другие предположения, то $\lim f_n$ является интегрируемой функцией и

$$\lim \mathcal{I}(f_n) = \mathcal{I}(\lim f_n).$$

Исходя из этих требований, приведем определение, ограничиваясь случаем, когда интегрируемые функции $f(t)$ — неслучайные, а интегрирующий процесс является процессом с ортогональными приращениями.

Пусть $\xi(t)$, $t \in (a, b)$, $-\infty \leq a < b \leq +\infty$ — процесс с ортогональными приращениями. Положим $(a \leq c < d < b)$

$$F(c, d) = M |\xi(d) - \xi(c)|^2.$$

Функцию интервала $F(c, d)$ назовем структурной функцией процесса $\xi(t)$. Она аддитивна, т. е. если $c_1 < c_2 < c_3$, то

$$F(c_1, c_3) = F(c_1, c_2) + F(c_2, c_3).$$

Действительно, из ортогональности приращений процесса $\xi(t)$ следует

$$M |\xi(c_3) - \xi(c_1)|^2 = M |\xi(c_3) - \xi(c_2)|^2 + M |\xi(c_2) - \xi(c_1)|^2.$$

Поэтому $F(c, d)$, как функция от d , $d > c$, монотонно не убывает, а как функция от c ($c < d$) монотонно не возрастает. Следовательно, множество точек разрыва функции $F(c, d)$ (по каждому аргументу в отдельности) не более чем счетно, и односторонние

пределы $F(c \pm, d \pm)$ существуют при любых c и d . Отсюда вытекает, что при каждом $t \in (a, b)$ существуют пределы в среднеквадратическом

$$\xi(t-) = \text{l. i. m.}_{s \uparrow t} \xi(s), \quad \xi(t+) = \text{l. i. m.}_{s \downarrow t} \xi(s),$$

причем процессы $\xi(t-)$ и $\xi(t+)$ являются процессами с ортогональными приращениями. При этом процессы $\xi(t-)$ и $\xi(t+)$ отличаются от $\xi(t)$ не более чем на счетном множестве значений, и процесс $\xi(t+)$ ($\xi(t-)$) непрерывен в среднеквадратическом справа (слева). Поэтому можно считать, что процесс $\xi(t)$ непрерывен в среднеквадратическом, например, справа. Тогда и функция $F(c, d)$ будет непрерывна справа по каждому из своих аргументов. В дальнейшем будем считать процесс $\xi(t)$ определенным на $(-\infty, \infty)$ (если процесс $\xi(t)$ определен на (a, b) , то его, естественно, можно продолжить на $(-\infty, \infty)$, положив $\xi(t) = \xi(a+)$ при $t \leq a$ и $\xi(t) = \xi(b-)$ при $t \geq b$ и при этом сформулированные ранее предположения о процессе остаются в силе), а функцию $F(c, d)$ — конечной на каждом конечном интервале. Вместо $F(c, d)$ будем писать $F(c, d]$.

Назовем функцию $f(t)$ простой, если она определена на $(-\infty, \infty)$, равна 0 вне некоторого конечного интервала и принимает только конечное число значений c_k , $k = 1, \dots, n$, на конечных полуинтервалах

$$\Delta_k = (a_k, b_k], \quad -\infty < a_k < b_k < +\infty.$$

Такую функцию можно представить в виде

$$f(t) = \sum_{k=1}^n c_k I_{\Delta_k}(t).$$

В соответствии с общими свойствами интеграла 1), 2) положим

$$\mathcal{J}(f) = \int_{-\infty}^{\infty} f(t) d\xi(t) = \sum_{k=1}^n c_k [\xi(b_k) - \xi(a_k)], \quad (11)$$

причем, нетрудно заметить, что это определение однозначно. Пусть даны две простые функции $f_i(t)$, $i = 1, 2$. Их всегда можно записать в виде

$$f_i(t) = \sum_{k=1}^n c_k^i I_{\Delta_k}(t),$$

т. е. считать, что полуинтервалы Δ_k , входящие в формулу (11), одинаковы для обеих функций. При этом не исключается, что некоторые c_k^i равны 0. Очевидно,

$$M\mathcal{J}(f_1) \overline{\mathcal{J}(f_2)} = \sum_{k=1}^n c_k^1 \overline{c_k^2} F(a_k, b_k]. \quad (12)$$

По непрерывной справа аддитивной функции $F(a, b]$ можно построить на $(-\infty, \infty)$ конечную или σ -конечную меру $F(\cdot)$, определенную на пополнении σ -алгебры борелевских множеств прямой $(-\infty, \infty)$, совпадающую с $F(a, b]$ на полуинтервалах $(a, b]$. Равенство (12) можно записать следующим образом:

$$M\mathcal{I}(f_1)\mathcal{I}(f_2) = \int_{-\infty}^{\infty} f_1(t) \overline{f_2(t)} F(dt). \quad (13)$$

Обозначим через $L_0^{\mathcal{E}}$ множество всех случайных величин вида

$$\sum_{k=1}^n c_k [\xi(b_k) - \xi(a_k)], \quad -\infty < a_1 < b_1 < a_2 < \dots < b_n < +\infty,$$

а через L_0 множество всех простых функций на $(-\infty, \infty)$. $L_0^{\mathcal{E}}$ является линейным подпространством пространства L_2 . Обозначим его замыкание в L_2 через $L_2^{\mathcal{E}}$. Оно состоит из тех и только тех случайных величин η , для которых существует последовательность $\eta_n \in L_0^{\mathcal{E}}$, $n = 1, 2, \dots$, такая, что $M|\eta - \eta_n|^2 \rightarrow 0$. $L_2^{\mathcal{E}}$ является гильбертовым пространством со скалярным произведением $(\eta', \eta'') = M\eta'\eta''$. Кроме того, рассмотрим замыкание линейного пространства функций L_0 по норме $\|f\|_F = \sqrt{\int_{-\infty}^{\infty} |f(t)|^2 F(dt)}$. Обозначим его через L_2^F . Как известно, L_2^F состоит из всех \mathfrak{F}_F -измеримых функций $f(t)$, для которых

$$\int_{-\infty}^{\infty} |f(t)|^2 F(dt) < \infty,$$

где \mathfrak{F}_F — пополнение относительно меры F , σ -алгебры борелевских множеств на $(-\infty, \infty)$ и также является гильбертовым пространством. Формула (11) устанавливает линейное соответствие между L_0 и $L_0^{\mathcal{E}}$. Оно взаимно однозначно. Действительно, из (13) следует равенство

$$M|\mathcal{I}(f_1) - \mathcal{I}(f_2)|^2 = \int_{-\infty}^{\infty} |f_1(t) - f_2(t)|^2 F(dt), \quad (14)$$

означающее, что двум разным функциям f_1 и f_2 соответствуют разные случайные величины $\mathcal{I}(f_1)$ и $\mathcal{I}(f_2)$. При этом под равенством $f_1 = f_2$ следует понимать равенство в L_2^F , т. е., что функции f_1 и f_2 совпадают F -почти всюду на $(-\infty, \infty)$. Соответствие $f \rightarrow \mathcal{I}(f)$ по формуле (13) или (14) изометрично. Следовательно, это соответствие можно по непрерывности продолжить до взаимно однозначного отображения L_2^F в $L_2^{\mathcal{E}}$. Возьмем произвольную функцию $f \in L_2^F$ и рассмотрим последовательность $f_n \in L_0$, такую, что

$\|f - f_n\|_F \rightarrow 0$. Тогда последовательность f_n является фундаментальной в L_2^F и такой же будет последовательность $\eta_n = \mathcal{I}(f_n)$ в $L_2^{\Delta k}$. Положим

$$\mathcal{I}(f) = \text{l. i. m. } \mathcal{I}(f_n).$$

Этот предел зависит только от f , т. е. не зависит от выбора конкретной последовательности $\{f_n, n = 1, 2, \dots\}$, аппроксимирующей f .

Определение. Случайную величину $\mathcal{I}(f)$ ($f \in L_2^F$) называют стохастическим интегралом и пишут

$$\mathcal{I}(f) = \int_{-\infty}^{\infty} f(t) d\xi(t). \quad (15)$$

Теорема 3. Стохастический интеграл (15) определен для произвольной функции $f \in L_2^F$, где F — структурная функция непрерывного справа процесса с ортогональными приращениями. Он обладает следующими свойствами:

$$\text{а) } \int_{-\infty}^{\infty} I_{[a,b]} d\xi = \xi(b) - \xi(a), \quad (16)$$

$$\text{б) } \int_{-\infty}^{\infty} [\alpha_1 f_1(t) + \alpha_2 f_2(t)] d\xi = \alpha_1 \int_{-\infty}^{\infty} f_1(t) d\xi + \alpha_2 \int_{-\infty}^{\infty} f_2(t) d\xi, \quad (17)$$

$$\text{в) } \mathbf{M} \int_{-\infty}^{\infty} f_1(t) d\xi \overline{\int_{-\infty}^{\infty} f_2(t) d\xi} = \int_{-\infty}^{\infty} f_1(t) \overline{f_2(t)} F(dt), \quad (18)$$

$$\text{г) если } \int_{-\infty}^{\infty} |f - f_n|^2 dF \rightarrow 0, \text{ то л. i. m. } \mathcal{I}(f_n) = \mathcal{I}(f),$$

и устанавливает взаимно однозначное соответствие между L_2^F и $L_2^{\Delta k}$.

Формулы (17) и (18) ранее были установлены для $f \in L_0$. С помощью предельного перехода они легко переносятся на произвольные функции из L_2^F . Из формулы (18) вытекает формула (14) для произвольных функций $f_i \in L_2^F$, а из формулы (4) вытекает г), так как

$$\|f - f_n\|_F^2 = \mathbf{M} \|\mathcal{I}(f) - \mathcal{I}(f_n)\|^2.$$

§ 2. Слабо стационарные процессы

Важный класс случайных процессов образуют стационарные процессы. Так называют процессы, теоретико-вероятностные характеристики которых не меняются со временем. Можно еще сказать, что стационарные процессы — это процессы, протекающие в неизменяющихся со временем условиях. Более точно это означает следующее.

Определение. Случайный процесс $\xi(t)$, $t \in (a, b)$, со значениями в R^d называют стационарным, если для любого n и любых t_0, t_1, \dots, t_n , таких, что $t + t_k \in (a, b)$, совместное распределение случайных векторов $\xi(t_1 + t), \dots, \xi(t_n + t)$ не зависит от t .

Условие независимости распределения последовательности $\xi(t_1 + t), \dots, \xi(t_n + t)$ от t эквивалентно требованию, чтобы для любой ограниченной непрерывной функции $f(x_1, \dots, x_n)$, $x_k \in R^d$, величина $Mf(\xi(t_1 + t), \dots, \xi(t_n + t))$ не зависела от t . В частности, если компоненты $\xi_j(t)$, $j = 1, \dots, d$, вектора $\xi(t)$ обладают конечными моментами второго порядка, то величины

$$m_j(t) = M\xi_j(t), \quad j = 1, \dots, d,$$

не зависят от t , $m_j(t) = m_j$, а величины

$$B_{jk}(t, s) = M\xi_j(t)\xi_k(s); \quad j, k = 1, \dots, d,$$

$t \geq s$, зависят только от разности $t - s$.

В тех задачах, где имеют значение только моменты первого и второго порядков рассматриваемых процессов, условие стационарности можно ослабить, заменив его условиями, относящимися только к моментам первых двух порядков.

Определение. Случайный процесс $\xi(t) = (\xi_1(t), \dots, \xi_d(t))$, $t \in (-\infty, \infty)$, называют слабо стационарным процессом, если $M|\xi(t)|^2 < \infty$ и

$$M\xi(t) = m = \text{const}, \quad M(\xi(t) - m)(\xi(t) - m)^* = R(t - s),$$

где $R(t)$ — непрерывная матричная функция.

Функцию $R(t)$ называют корреляционной (матричной) функцией процесса $\xi(t)$. Ее элементы

$$R_{kj}(t) = M(\xi_k(t + s) - m_k)(\xi_j(s) - m_j)$$

являются взаимными корреляционными функциями одномерных случайных процессов $\xi_k(t)$ и $\xi_j(t)$.

Слабо стационарные процессы называют еще процессами стационарными в широком смысле или стационарными процессами второго порядка. В дальнейшем мы ограничиваемся определениями и результатами для одномерных комплекснозначных процессов. Большинство этих определений и результатов легко обобщается на векторные процессы.

Требование, чтобы функция $R(t, s)$ зависела только от разности $t - s$, означает, что совместное распределение пары случайных величин $\xi(t)$ и $\xi(s)$ обладает некоторыми свойствами инвариантности относительно сдвига времени.

Обозначим через \mathfrak{L}_2 гильбертово пространство всех комплекснозначных случайных величин η на (Ω, \mathcal{F}, P) , для которых $M|\eta|^2 < \infty$, и в котором скалярное произведение случайных величин η' и η'' определяется соотношением

$$(\eta', \eta'') = M\eta'\bar{\eta''}.$$

Из определения слабой стационарности вытекает, что:

1) норма вектора $\xi(t)$ не зависит от времени

$$\|\xi(t)\| = \sqrt{(\xi(t), \xi(t))} = \sqrt{M|\xi(t)|^2} = R(0),$$

2) угол φ между парой векторов $\xi(t)$ и $\xi(t+h)$ не зависит от t и является непрерывной функцией от h

$$\cos \varphi = \frac{(\xi(t), \xi(t+h))}{\|\xi(t)\| \|\xi(t+h)\|} = \frac{R(h)}{R(0)}.$$

Если интерпретировать $\xi(t)$ как вектор в L_2 , проведенный из точки 0, то стационарному в широком смысле процессу соответствует некоторая кривая $\xi(t)$, $-\infty < t < \infty$, на поверхности сферы в L_2 радиуса $R(0)$, обладающая следующим свойством: за равные промежутки времени радиус-вектор $\xi(t)$ поворачивается на один и тот же угол, т. е. угол между $\xi(t)$ и $\xi(t+h)$ не зависит от t .

ПРИМЕРЫ

1. Колебания со случайными параметрами. Во многих вопросах теории колебаний (механических, колебаниях электрического тока и др.) рассматривают процессы вида

$$\xi(t) = \sum_{k=1}^n \gamma_k e^{i\omega_k t}. \quad (1)$$

Они являются суммами простых гармонических колебаний $\gamma_k(t) = \gamma_k e^{i\omega_k t} = \alpha_k e^{i(\omega_k t + \varphi_k)}$. Здесь α_k — амплитуда, $\gamma_k = \alpha_k e^{i\varphi_k}$ — комплексная амплитуда, φ_k — начальная фаза, ω_k — частота гармонического колебания $\xi_k(t)$. Совокупность частот ω_k , $k = 1, \dots, n$, рассматриваемых как множество точек на прямой $(-\infty, \infty)$, называют спектром колебаний процесса $\xi(t)$.

Введем некоторые термины, используемые при физической интерпретации случайных процессов. Если $|\xi(t)|$ обозначает силу тока в момент времени t и имеется в виду энергия, рассеиваемая этим током на единичном сопротивлении, то естественными являются следующие определения.

Энергией, переносимой случайным процессом $\xi(t)$ в течение промежутка времени (t_1, t_2) , называется величина

$$\int_{t_1}^{t_2} |\xi(t)|^2 dt,$$

а средней мощностью случайного процесса называют предел

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |\xi(t)|^2 dt,$$

если он существует.

В том случае, когда процесс имеет иную физическую интерпретацию, принятая терминология может ей не соответствовать. Все же в дальнейшем эта терминология будет использована.

Нетрудно вычислить среднюю мощность, переносимую процессом $\xi(t)$.

$$\begin{aligned} \frac{1}{2T} \int_{-T}^T |\xi(t)|^2 dt &= \frac{1}{2T} \int_{-T}^T \sum_{k, r=1}^n \gamma_k \bar{\gamma}_r e^{it(u_k - u_r)} dt = \\ &= \sum_{k=1}^n |\gamma_k|^2 + \sum_{\substack{k, r=1 \\ k \neq r}}^n \gamma_k \bar{\gamma}_r \frac{\sin T(u_k - u_r)}{T(u_k - u_r)}. \end{aligned}$$

При $T \rightarrow \infty$ получаем

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |\xi(t)|^2 dt = \sum_{k=1}^n |\gamma_k|^2.$$

Таким образом, средняя мощность, переносимая колебательным случайным процессом $\xi(t)$, равна сумме средних мощностей, переносимых каждой гармонической составляющей процесса.

В дальнейшем будем считать комплексные амплитуды γ_r случайными величинами. Тогда $\xi(t)$ — случайный комплекснозначный процесс, $t \in (-\infty, \infty)$.

Распределение величины $\xi(t)$, даже при специальных предположениях о распределении величины γ_k , является весьма сложным, но моменты первых двух порядков величины $\xi(t)$ получить нетрудно. Предположим, что комплексные амплитуды γ_k имеют средние значения, равные 0, и между собой не коррелированы, т. е.

$$M\gamma_k = 0, M\gamma_k \bar{\gamma}_r = 0, k \neq r, M|\gamma_k|^2 = c_k^2. \quad (2)$$

Тогда

$$\begin{aligned} M\xi(t) &= 0, R(t_1, t_2) = M\xi(t_1) \overline{\xi(t_2)} = M \left[\sum_{k, r=1}^n \gamma_k \bar{\gamma}_r e^{i(u_k t_1 - u_r t_2)} \right] = \\ &= \sum_{k=1}^n c_k^2 e^{i u_k (t_1 - t_2)}. \end{aligned}$$

Таким образом, корреляционная функция процесса $\xi(t)$ зависит только от разности $t_1 - t_2$, $R(t_1, t_2) = R(t_1 - t_2)$, где

$$R(t) = \sum_{k=1}^n c_k^2 e^{i u_k t}. \quad (3)$$

Следовательно, если выполнены условия (2), то процесс $\xi(t)$ является стационарным процессом в широком смысле и его корреляционная функция задается формулой (3). Эту формулу называют спектральной представлением корреляционной функции. Она определяет спектр случайного процесса, т. е. совокупность частот $\{u_k\}$, $k = 1, \dots, n$, гармонических колебаний, составляющих процесс $\xi(t)$ и математические ожидания c_k^2 средних мощностей, переносимых соответствующими составляющими процесса. Величину c_k^2 называют средним значением мощности гармонической составляющей процесса с частотой u_k .

Назовем спектральной функцией $F(u)$ процесса (1) функцию

$$F(u) = \sum_{\substack{k \\ u_k < u}} c_k^2.$$

Ясно, что $F(u)$ равна средней мощности, переносимой гармоническими составляющими процесса $\xi(t)$, частоты которого менее заданного значения u . С помощью спектральной функции корреляционная функция процесса $\xi(t)$ может быть записана в виде

$$R(t) = \int_{-\infty}^{\infty} e^{iut} dF(u). \quad (4)$$

Оказывается, что понятие спектральной функции и представление (4) могут быть обобщены на произвольные случайные процессы.

2. Пусть $F(u)$, $u \in (-\infty, \infty)$, — произвольная монотонно неубывающая непрерывная слева функция и $F(-\infty) = 0$, $F(+\infty) = \sigma^2$. Введем случайную величину ξ с функцией распределения $\frac{1}{\sigma^2} F(x)$ и определим случайный процесс

$$\xi(t) = \sigma e^{i(\xi t + \varphi)},$$

где φ — равномерно распределенная на $(0, 2\pi)$ случайная величина, не зависящая от ξ . Тогда

$$\begin{aligned} M_{\xi}^2(t) &= \int_{-\infty}^{\infty} \int_0^{2\pi} \sigma^2 e^{i(tu + \varphi)} \frac{1}{\sigma^2} F(du) \frac{d\varphi}{2\pi} = 0, \\ R(t_1, t_2) &= M_{\xi}^2(t_1) \overline{\xi(t_2)} = \\ &= \int_{-\infty}^{\infty} \int_0^{2\pi} \sigma^2 e^{i(t_1 - t_2)u} \frac{1}{\sigma^2} F(du) \frac{d\varphi}{2\pi} = R(t_1 - t_2), \end{aligned}$$

где $R(t)$ имеет вид (4), а $F(u)$ — произвольная ограниченная монотонно неубывающая непрерывная слева функция.

Теорема 1. (Хинчина). Для того чтобы функция $R(t)$ была корреляционной функцией непрерывного в среднеквадратическом слабо стационарного процесса, необходимо и достаточно, чтобы она допускала представление (4).

Доказательство. Необходимость. Если $R(t)$ — корреляционная функция слабо стационарного процесса, то $R(t-s)$ является положительно определенным ядром: для любых целых n , действительных t_1, \dots, t_n и комплексных c_1, c_2, \dots, c_n (см. § 7, гл. VI),

$$\sum_{k, r=1}^n R(t_k - t_r) c_k \bar{c}_r \geq 0,$$

т. е. $R(t)$ — положительно определенная функция. Из среднеквадратической непрерывности процесса слева следует (теорема 1, § 1), что

$$\square R(t, t) = 2(R(0) - \operatorname{Re} R(t' - t)) \rightarrow 0 \text{ при } t' \rightarrow t.$$

Отсюда вытекает (см. (24), § 7, гл. VI), что функция $R(t)$ непрерывна. По теореме Бохнера (теорема 11, § 7, гл. VI) функция $R(t)$ является, с точностью до множителя, характеристиче-

ской функцией некоторого распределения, т. е. она имеет представление (4).

Достаточность. Если функция $R(t)$ имеет вид (4), то она является корреляционной функцией некоторого слабо стационарного процесса, например, процесса, рассмотренного в примере 2 (с. 291). Этот процесс непрерывен в среднеквадратическом, так как из непрерывности функции $R(t)$ вытекает непрерывность в среднеквадратическом процесса (теорема 1, § 1, гл. VIII).

Заметим, что функция $F(u)$ в формуле (4) определяется через функцию $R(t)$ однозначно (§ 7, гл. VI). Более того, если a и b точки непрерывности функции $F(u)$, то имеет место следующая формула обращения:

$$F(b) - F(a) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\pi} \int_a^b \int_{-\infty}^{\infty} R(t) e^{-iut - \frac{\varepsilon^2 t^2}{2}} dt du. \quad (5)$$

Определение. Функцию $F(u)$, содержащуюся в формуле (4), называют спектральной функцией слабо стационарного процесса. Если $F(u)$ абсолютно непрерывна, т. е.

$$F(u) = \int_{-\infty}^u f(v) dv,$$

то функцию $f(u)$ называют спектральной плотностью процесса.

Если корреляционная функция $R(t)$ интегрируема на интервале $(-\infty, \infty)$

$$\int_{-\infty}^{\infty} |R(t)| dt < \infty,$$

то можно положить (см. (10), § 7, гл. VI), что

$$f(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iut} R(t) dt,$$

т. е. спектральная плотность совпадает с обратным преобразованием Фурье корреляционной функции.

§ 3. Слабо стационарные последовательности

Понятия стационарности и слабой стационарности непосредственно применимы и к последовательностям случайных величин.

Предположим, что через равные промежутки времени наблюдаются случайные величины ξ_n , $n = 0, 1, \dots$. Здесь n интерпретируется как момент времени, в который производится наблюдение, и для простоты предположим, что промежутки между наблюдениями равны 1 и начинаются с момента времени $n = 0$. Во многих

случаях можно предполагать, что наблюдения производились и в «прошлом», так что рассматривается семейство случайных величин вида

$$\dots, \xi_{-n}, \xi_{-n+1}, \dots, \xi_0, \xi_1, \dots, \xi_n, \dots \quad (1)$$

Пусть n и t — произвольные целые числа ($n \geq 0, t \in (-\infty, \infty)$). Обозначим через $F_{t, t+1, \dots, t+n}$ распределение случайных величин $\xi_t, \xi_{t+1}, \dots, \xi_{t+n}$. Распределения $F_{t, t+1, \dots, t+n}$ называют конечномерными распределениями последовательности случайных величин (1).

Последовательность (1) называют стационарной, если для всех целых $n \geq 0, t = 0, \pm 1, \dots$, распределение $F_{t, t+1, \dots, t+n}$ не зависит от t . Ее называют слабо стационарной (стационарной в широком смысле или стационарной второго порядка), если

$$M\xi_n = m, \quad M(\xi_{t+n} - m)(\xi_t - m) = R_n.$$

При этом случайные величины ξ_n , вообще говоря, комплекснозначные. Последовательность чисел R_n называют корреляционной функцией последовательности (1). Отметим следующее свойство последовательности R_n :

$$\begin{aligned} R_0 &= M|\xi_t - m|^2 \geq 0, \\ |R_n| &\leq R_0, \quad R_{-n} = \bar{R}_n, \\ \sum_{k, j=1}^n R_{k-j} c_k \bar{c}_j &= M \left| \sum_{k=1}^n (\xi_k - m) c_k \right|^2 \geq 0. \end{aligned}$$

Таким образом, последовательность R_n является положительно определенной последовательностью. По теореме 12, § 7, гл. VI, такую последовательность можно записать так:

$$R_n = \int_0^{2\pi} e^{inu} dF(u),$$

где $F(u)$ — непрерывная слева монотонно неубывающая функция. Функцию $F(u)$ называют спектральной функцией слабо стационарной последовательности, а ее производную, если она существует, спектральной плотностью. Если спектральная плотность $f(u)$ существует, то величины R_n являются с точностью до множителя $\frac{1}{2\pi}$ ее коэффициентами Фурье.

§ 4. Линейные преобразования случайных процессов

В настоящем параграфе рассматриваются некоторые понятия, связанные с линейными преобразованиями случайных процессов. Для большей наглядности будем интерпретировать линейное преобразование как некоторую систему Σ (устройство или прибор),

принимаящую внешние сигналы (воздействия) и преобразующую их в новые. Эти сигналы описываются некоторыми функциями, скалярными или векторными. Функцию, воздействующую на систему, называют функцией на входе системы, а преобразованную — функцией на выходе системы или реакцией системы на функцию на входе.

Пусть $f(t)$ — значение функции на входе в момент времени t , а $g(t)$ — на выходе. Соответствие между $f(\cdot)$ и $g(\cdot)$ запишем в виде $g = Tf$ или подробнее

$$g(t) = T(f|t).$$

При этом функции на входе, разумеется, могут быть произвольными, они принадлежат некоторому классу, обозначим его через D , который назовем классом функций, допустимых на входе системы.

Система Σ называется линейной, если:

- класс функций, допустимых на входе, линейен;
- оператор T удовлетворяет принципу суперпозиции

$$T(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1 T f_1 + \alpha_2 T f_2,$$

где α_1 и α_2 — константы.

В дальнейшем будем предполагать, что функции на входе и выходе скалярны (вообще говоря, комплекснозначны). Примером линейного преобразования может быть преобразование, определяемое соотношением

$$g(t) = \int_{-\infty}^{\infty} h(t, s) f(s) ds, \quad (1)$$

в котором класс D состоит из функций, для которых интеграл в правой части равенства имеет смысл. Чтобы получить интерпретацию функции $h(t, s)$, представим себе, что на вход системы поступает мгновенный единичный импульс. Это означает, что функция $f(t)$ отлична от нуля в течение короткого промежутка времени $(s, s + \Delta s)$ и $f(t) = \frac{1}{\Delta s}$ при $t \in (s, s + \Delta s)$. Тогда

$$g(t) \cong h(t, s).$$

Таким образом, $h(t, s)$ — это реакция в момент времени t на единичный импульс, возникший в момент времени s . В соответствии с этим $h(t, s)$ называют импульсной или импульсной переходной функцией системы (преобразования).

Введем операцию сдвига времени θ_a ($-\infty < a < \infty$), положив

$$\theta_a(x(\cdot)|t) = x(t + a).$$

Если внутренние параметры системы Σ не зависят от времени, то естественно ожидать, что при временном сдвиге функции на входе такой же сдвиг произойдет с функцией на выходе, т. е.

$$T(\theta_a x(\cdot)|t) = T(x(\cdot)|t + a).$$

Таким образом, $T\theta_a = \theta_a T \forall a \in (-\infty, \infty)$. Системы, для которых класс D функций, допустимых на входе, инвариантен относительно сдвига времени, $\theta_a D = D$ и $T\theta_a = \theta_a T$ для всех $a \in (-\infty, \infty)$ называют однородными во времени. Если система Σ обладает импульсной реакцией $h(t, s)$, для которой функция на выходе представима формулой (1) и однородна во времени, то

$$h(t, s) = h(t + a, s + a) = h(t - s, 0),$$

т. е. $h(t, s) = h(t - s)$. Функцию $h(t)$ также называют импульсной переходной функцией однородной во времени системы. Соотношение (1) для однородной системы принимает вид

$$g(t) = \int_{-\infty}^{\infty} h(t-s) f(s) ds.$$

Таким образом, для однородной системы функция $g(t)$ на выходе является сверткой импульсной реакции $h(t)$ с функцией на входе $f(t)$, $g = h \times f$.

В дальнейшем рассматриваются в основном линейные однородные системы Σ . Пусть на вход такой системы поступило простое гармоническое колебание $f(t) = e^{iut}$. Тогда для любого s

$$T(f|s) = T(\theta_s f|0) = T|(e^{ius} f(\cdot)|0) = e^{ius} H(iu),$$

где

$$H(iu) = T(e^{iut}|0).$$

Таким образом, отношение функции на выходе к функции на входе, когда функция на входе $f(t) = e^{iut}$ не зависит от времени

$$H(iu) = \frac{T(e^{iut}|s)}{e^{ius}}.$$

Определение. Функцию $H(iu)$ называют частотной характеристикой, или коэффициентом передачи системы.

Частотная характеристика определена для любой линейной однородной системы, для которой функции e^{iut} , $u \in (-\infty, \infty)$, являются допустимыми функциями на входе $e^{iut} \in D$.

Если система обладает интегрируемой импульсной реакцией $h(t)$, то реакция на функцию e^{iut} имеет вид

$$g(t) = \int_{-\infty}^{\infty} h(t-s) e^{ius} ds = e^{iut} \int_{-\infty}^{\infty} h(s) e^{-ius} ds.$$

Таким образом,

$$H(iu) = \int_{-\infty}^{\infty} h(s) e^{-ius} ds. \quad (2)$$

Отсюда, в частности, вытекает, что импульсная реакция, когда она существует, однозначно определяется частотной характеристикой.

В этом случае можно дать следующую интерпретацию частотной характеристике. Пусть $f(t)$ — интегрируемая функция на входе, $g(t)$ — соответствующая функция на выходе, $\tilde{f}(a)$, $\tilde{g}(a)$ — их преобразования Фурье. Тогда $\tilde{g}(u) = \tilde{H}(-iu) \tilde{f}(u)$, и

$$H(-iu) = \frac{\tilde{g}(u)}{\tilde{f}(u)},$$

т. е. отношение преобразований Фурье функции на выходе к преобразованию Фурье функции на входе системы не зависит от функции на входе и равно $H(-iu)$.

В реальных физических устройствах нет возможности предвосхитить будущее. Это означает, что если $f(t) = 0$ при $t \leq t_0$, то и реакция системы $g(t) = 0$ при $t \leq t_0$. Таким образом, импульсная реакция реальной системы должна удовлетворять условию $h(t, s) = 0$ при $t < s$. Это условие называют условием физической осуществимости импульсной реакции. Если система однородна, то условие физической осуществимости принимает вид

$$h(t) = 0 \text{ при } t < 0,$$

а выражение для функции на выходе системы

$$g(t) = \int_{-\infty}^t h(t-s) f(s) ds = \int_0^{\infty} h(s) f(t-s) ds.$$

Если система начинает функционировать в момент времени 0, т. е. $f(t) = 0$ при $t < 0$, то

$$g(t) = \int_0^t f(t-s) h(s) ds.$$

Изучая такие системы, вместо преобразования Фурье удобно пользоваться преобразованием Лапласа. Пусть

$$\check{g}(p) = \int_0^{\infty} e^{-pt} g(t) dt$$

— преобразование Лапласа функции $g(t)$. Заметим, что $\check{h}(p) = H(p)$. Имеем

$$\begin{aligned} \check{g}(p) &= \int_0^{\infty} e^{-pt} \int_0^t f(t-s) h(s) ds dt = \\ &= \int_0^{\infty} \int_s^{\infty} e^{-pt} f(t-s) h(s) dt ds = \int_0^{\infty} e^{-ps} h(s) ds \int_0^{\infty} e^{-pt} f(t) dt, \end{aligned}$$

или

$$\check{g}(p) = H(p) \check{f}(p).$$

Проделанная выкладка закона, если при некотором α функции $e^{-\alpha t}h(t)$ и $e^{-\alpha t}f(t)$ абсолютно интегрируемы и $\operatorname{Re} p > 2$.

Пусть теперь на вход однородной системы с импульсной реакцией $h(t)$ подается слабо стационарный процесс $\xi(t)$, $M\xi(t) = 0$. Процесс на выходе обозначим через $\eta(t)$:

$$\eta(t) = \int_{-\infty}^{\infty} h(t-s) \xi(s) ds. \quad (3)$$

Интеграл в правой части равенства (3) определен в § 1, гл. VIII. Для его существования достаточно, чтобы $\forall \alpha$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B(s, s') ds ds' < \infty,$$

где $B(s, s')$ — ковариация процесса $h(t-s)\xi(s)$,

$$\begin{aligned} B(s, s') &= Mh(t-s)\xi(s)\overline{h(t-s')\xi(s')} = \\ &= h(t-s)R_{\xi}(s-s')\overline{h(t-s')}, \end{aligned}$$

$R_{\xi}(t)$ — корреляционная функция процесса $\xi(t)$.

Воспользуемся спектральным представлением корреляционной функции

$$R_{\xi}(t) = \int_{-\infty}^{\infty} e^{iut} F(du).$$

Получим

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B(s, s') ds ds' &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t-s) \times \\ &\times \int_{-\infty}^{\infty} e^{i(t-s-s')u} F(du) \overline{h(t-s')} ds ds' = \int_{-\infty}^{\infty} |H(iu)|^2 F(du). \end{aligned}$$

Итак, если

$$\int_{-\infty}^{\infty} |H(iu)|^2 F(du) < \infty, \quad (4)$$

то интеграл (3) существует для всех $t > 0$. При этом $M\eta(t) = 0$ и выкладка, аналогичная предыдущей, показывает, что

$$\begin{aligned} R_{\eta}(t, t') &= M\eta(t)\overline{\eta(t')} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t-s) R_{\xi}(s-s') \overline{h(t'-s')} ds ds' = \\ &= \int_{-\infty}^{\infty} e^{iu(t-t')} \int_{-\infty}^{\infty} h(t-s) e^{-iu(s-s')} ds \int_{-\infty}^{\infty} h(t'-s') \times \\ &\times e^{-iu(t'-s')} ds' F(du) = \int_{-\infty}^{\infty} e^{iu(t-t')} |H(iu)|^2 F(du). \end{aligned}$$

Таким образом, если выполнено условие (4), то процесс $\eta(t)$, определяемый формулой (3), имеет смысл и является слабо стационарным с корреляционной функцией $R_\eta(t)$

$$R_\eta(t) = \int_{-\infty}^{\infty} e^{iut} |H(iu)|^2 F(du). \quad (5)$$

Определение. Преобразование T называют допустимым фильтром (или, короче, фильтром) процесса $\xi(t)$, если оно имеет вид (3) и выполнено условие (4), или если оно является среднеквадратическим пределом таких преобразований.

Условие сходимости последовательности преобразований (3) $\eta_n(t) = T_n(\xi|t)$ с импульсными реакциями $h_n(t)$ и частотными характеристиками $H_n(iu)$ состоит в следующем:

$$M |\eta_n(t) - \eta_m(t)|^2 = \int_{-\infty}^{\infty} |H_n(iu) - H_m(iu)|^2 F(du) \rightarrow 0.$$

Это означает, что последовательность $H_n(iu)$ фундаментальна в L_2^F . Напомним, что L_2^F обозначает пространство \mathfrak{F} -измеримых функций $f(u)$, для которых

$$\int_{-\infty}^{\infty} |f(u)|^2 F(du) < \infty.$$

Здесь \mathfrak{F} — пополнение по мере F σ -алгебры борелевских множеств на прямой. Тогда существует предел $H(iu) = \text{l. i. m. } H_n(iu)$ (в L_2^F), который называют частотной характеристикой предельного фильтра, и если $\eta(t) = \text{l. i. m. } \eta_n(t)$, то корреляционная функция предельного фильтра также выражается формулой (5).

Обратно, какова бы ни была функция $H(iu) \in L_2^F$, ее можно аппроксимировать в смысле сходимости в L_2^F функциями, являющимися преобразованиями Фурье абсолютно интегрируемых функций. Таким образом, фильтры удобно задавать частотными характеристиками.

Теорема 1. Для того чтобы функция $H(iu)$ была частотной характеристикой допустимого фильтра для процесса $\xi(t)$ со спектральной функцией $F(u)$, необходимо и достаточно, чтобы $H(iu) \in L_2^F$. Корреляционная функция процесса на выходе фильтра с частотной характеристикой $H(iu)$ задается формулой (5). Если $H_k(iu) \in L_2^F$, $k = 1, 2$,

$$H_3(iu) = H_1(iu) + H_2(iu),$$

$\eta_k(t)$ — процессы на выходе фильтра с частотной характеристикой $H_k(iu)$, $k = 1, 2, 3$, то $\eta_3(t) = \eta_1(t) + \eta_2(t)$, процессы $\eta_1(t)$ и $\eta_2(t)$

стационарно связаны и их взаимная корреляционная функция имеет вид

$$R_{12}(t) = \int_{-\infty}^{\infty} e^{itu} H_1(iu) \overline{H_2(iu)} F(du). \quad (6)$$

Доказательство. Первая часть теоремы уже доказана. Покажем, что корреляционная функция процесса на выходе задается формулой (5). Эта формула доказана для фильтров, импульсная реакция которых определена. Пусть $H_n(iu)$ — последовательность частотных характеристик, сходящихся в L_2^F к $H(iu)$, для которых импульсная реакция $h_n(t)$ существует, $\eta_n(t)$ — функция на выходе соответствующих систем. Тогда, как показано выше, л. и. т. $\eta_n(t) = \eta(t)$ — процесс на выходе системы с частотной характеристикой $H(iu)$ и

$$\begin{aligned} R_{\eta}(t) &= \lim_{n \rightarrow \infty} R_{\eta_n}(t) = \int_{-\infty}^{\infty} e^{iut} |H_n(iu)|^2 F(du) = \\ &= \int_{-\infty}^{\infty} e^{iut} |H(iu)|^2 F(du) + \delta(t), \end{aligned}$$

где

$$\delta(t) = \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} e^{iut} (|H_n(iu)|^2 - |H(iu)|^2) F(du)$$

и

$$\begin{aligned} |\delta(t)|^2 &\leq \lim_{n \rightarrow \infty} \left\{ \int_{-\infty}^{\infty} |H_n(iu) - H(iu)|^2 F(du) \times \right. \\ &\quad \times \left. 2 \int_{-\infty}^{\infty} |H_n(iu) + H(iu)|^2 H(iu) |F(du)| \right\} \rightarrow 0. \end{aligned}$$

Таким образом, формула (5) доказана в общем случае. Найдем взаимную корреляционную функцию $R_{12}(t, s)$ процессов $\eta_1(t)$ и $\eta_2(t)$. Предположим сначала, что для систем определены импульсные реакции $h_1(t)$ и $h_2(t)$. Тогда

$$\begin{aligned} R_{12}(t, s) &= M \eta_1(t) \overline{\eta_2(s)} = M \int_{-\infty}^{\infty} h_1(t-a) \xi(a) da \times \\ &\times \overline{\int_{-\infty}^{\infty} h_2(s-b) \xi(b) db} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1(t-a) R(a-b) \overline{h_2(s-b)} da db = \\ &= \int_{-\infty}^{\infty} e^{it(t-s)u} \int_{-\infty}^{\infty} h_1(t-a) e^{-i(t-a)u} da \int_{-\infty}^{\infty} h_2(s-b) e^{-i(s-b)u} db F(du) = \\ &= \int_{-\infty}^{\infty} e^{it(t-s)u} H_1(iu) \overline{H_2(iu)} F(du). \end{aligned}$$

Таким образом, процессы $\eta_1(t)$ и $\eta_2(t)$ стационарно связаны и их взаимная корреляционная функция задается формулой (6). Так же, как формула (5), равенство (6) переносится на произвольные фильтры. Наконец, равенство $\eta_3(t) = \eta_1(t) + \eta_2(t)$ тривиально для фильтров, обладающих импульсной характеристикой, и с помощью предельного перехода легко переносится на произвольные фильтры.

Теорема 1 показывает, что описание фильтров с помощью частотных характеристик бывает предпочтительней описания с помощью импульсной реакции, так как импульсная реакция фильтра может быть и не определена. Примеры будут приведены ниже.

Спектральное разложение слабо стационарных процессов. Спектральное представление корреляционной функции и результаты, полученные в примере 1 (с. 290), слабо стационарного процесса наталкивают на мысль, что произвольный слабо стационарный процесс можно в каком-то смысле рассматривать как непрерывную сумму некоррелированных простых гармонических колебаний вида γe^{iut} , $u \in (-\infty, \infty)$. Тогда $F(du)$ можно интерпретировать как энергию, переносимую гармоническими колебаниями с частотами в интервале $(u, u + du)$, а формула (5) показывает, что при прохождении случайного процесса $\xi(t)$ через фильтр энергия гармонических составляющих этого процесса с частотами в интервале $(u, u + du)$ умножается на $|H(iu)|^2$.

Докажем теперь ряд результатов, являющихся точной формулировкой этих интуитивных представлений.

Теорема 2. Пусть $\xi(t)$ — слабо стационарный процесс, $M\xi(t) = 0$, со спектральной функцией $F(u)$. Тогда существует процесс с ортогональными приращениями $\zeta(u)$, $u \in (-\infty, \infty)$, такой, что $M|\zeta(u)|^2 = F(u)$ и

$$\xi(t) = \int_{-\infty}^{\infty} e^{iut} \zeta(du). \quad (7)$$

Формулу (7) называют спектральным разложением процесса $\xi(t)$, а функцию интервала $\zeta(u, v) = \zeta(v) - \zeta(u)$, $u < v$ — стохастической спектральной мерой процесса $\xi(t)$.

Доказательство. Идея доказательства основана на следующем замечании. Если имеет место формула (7), то по формуле (5) спектральная мера может быть получена на выходе фильтра в момент времени $t = 0$, частотная характеристика которого $E_v(iu) = 1$ при $u \leq v$ и $E(iu) = 0$ при $u > v$, если на вход фильтра подать процесс $\xi(t)$. Итак, положим $\zeta(v) = \eta_v(0)$, где $\eta_v(t)$ — функция на выходе фильтра с частотной характеристикой $E_v(iu)$, на вход которого подан процесс $\xi(t)$. Очевидно, $E_v(iu) \in L_2^F$. Тогда при $v_1 < v_2 < v_3$

$$M(\zeta(v_3) - \zeta(v_2)) \overline{(\zeta(v_2) - \zeta(v_1))} = M\eta''(0) \overline{\eta'(0)},$$

где $\eta'(t)$ ($\eta''(t)$) — функция на выходе фильтра с частотной харак-

теристикой $E_{v_1} - E_{v_2}$ ($E_{v_1} - E_{v_2}$) (см. теорему 1, с. 298). Так как (согласно теореме 1)

$$\begin{aligned} M \eta''(0) \overline{\eta'(0)} &= \int_{-\infty}^{\infty} (E_{v_1}(iu) - E_{v_2}(iu)) \times \\ &\times \overline{(E_{v_2}(iu) - E_{v_1}(iu))} dF = 0, \end{aligned}$$

то процесс $\xi(v)$ является процессом с ортогональными приращениями. При этом

$$M |\xi(v)|^2 = M |\eta_v(0)|^2 = \int_{-\infty}^{\infty} |E_v(iu)|^2 F(du) = F(v),$$

$$M |\xi(v_2) - \xi(v_1)|^2 = F(v_2) - F(v_1), \quad v_1 < v_2.$$

Положим

$$S_\lambda = \sum_{k=0}^{n-1} e^{i\lambda u_k} [\xi(u_{k+1}) - \xi(u_k)],$$

где λ — символ разбиения прямой $(-\infty, \infty)$ на частичные интервалы точками u_0, \dots, u_n , $u_0 < u_1 < \dots < u_n$. Пусть дано произвольное $\varepsilon > 0$. Выберем точки u_0, \dots, u_n так, чтобы

$$F(u_0) < \frac{\varepsilon}{3}, \quad F(+\infty) - F(u_n) < \frac{\varepsilon}{3},$$

$$\max_{0 \leq k \leq n-1} (u_{k+1} - u_k) < \frac{\varepsilon}{6 |t| F(+\infty)}.$$

Оценим величину

$$\delta_\lambda = M |\xi(t) - S_\lambda|^2 = M |\xi(t)|^2 - 2 \operatorname{Re} M \xi(t) \bar{S}_\lambda + M |S_\lambda|^2.$$

Заметим, что $M |\xi(t)|^2 = F(+\infty)$. Из ортогональности приращений процесса $\xi(u)$ следует

$$M |S_\lambda|^2 = \sum_{k=0}^{n-1} M |\xi(u_{k+1}) - \xi(u_k)|^2 = F(u_n) - F(u_0).$$

Пусть $\eta^{(k)}(t)$ — процесс на выходе фильтра с частотной характеристикой $E_{u_{k+1}}(iu) - E_{u_k}(iu)$. Имеем

$$M \xi(t) \bar{S}_\lambda = \sum_{k=0}^{n-1} e^{-it u_k} M \xi(t) \eta^{(k)}(0) = \sum_{k=0}^{n-1} e^{-it u_k} \times$$

$$\begin{aligned} \times \int_{-\infty}^{\infty} e^{itu} (E_{u_{k+1}} - E_{u_k}) F(du) &= \sum_{k=0}^{n-1} \int_{u_k}^{u_{k+1}} e^{it(u-u_k)} F(du) = \\ &= F(u_n) - F(u_0) + \sum_{k=0}^{n-1} \int_{u_k}^{u_{k+1}} (e^{it(u-u_k)} - 1) F(du), \end{aligned}$$

причем,

$$\begin{aligned} \left| \sum_{k=0}^{n-1} \int_{u_k}^{u_{k+1}} (e^{it(u-u_k)} - 1) F(du) \right| &\leq \sum_{k=0}^{n-1} \left| \int_{u_k}^{u_{k+1}} |it(u-u_k)| \times \right. \\ &\times F(du) \leq \sum_{k=0}^{n-1} |t| |u_{k+1} - u_k| (F(u_{k+1}) - F(u_k)) < \frac{\varepsilon}{6}. \end{aligned}$$

Следовательно,

$$\delta_\lambda \leq F(+\infty) - F(u_n) + F(u_0) + \frac{\varepsilon}{3} < \varepsilon.$$

Таким образом, л. и. м. $S_\lambda = \xi(t)$ и формула (7) доказана.

Лемма 1. Пусть $h(t)$, $t \in (-\infty, \infty)$ — непрерывная интегрируемая функция, $\xi(t)$ — слабо стационарный процесс с спектральным разложением (7). Тогда

$$\int_{-\infty}^{\infty} h(t) \xi(t) dt = \int_{-\infty}^{\infty} H(-iu) \zeta(du), \quad (8)$$

где

$$H(-iu) = \int_{-\infty}^{\infty} e^{itu} h(u) du.$$

Доказательство. Заметим, что в условиях леммы интеграл $\int_{-\infty}^{\infty} h(t) \xi(t) dt$ существует. Действительно, для его существования достаточно, чтобы

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t) R(t-s) \overline{h(s)} dt ds = \int_{-\infty}^{\infty} |H(-iu)|^2 F(du) < \infty.$$

Это условие выполняется, так как функция $H(-iu)$ ограничена. Допустим сначала, что функция $h(t) = 0$ при $|t| > a$. Тогда

$$\begin{aligned} \int_{-a}^a h(t) \xi(t) dt &= \text{l. i. м. } \sum_{k=1}^n h(t_k) \xi(t_k) \Delta t_k = \\ &= \text{l. i. м. } \int_{-\infty}^{\infty} \left(\sum_{k=1}^n e^{it_k u} h(t_k) \Delta t_k \right) \zeta(du). \end{aligned}$$

Пусть $S_n = \sum_{k=1}^n e^{it_k u} h(t_k) \Delta t_k$. По теореме о глассируемой сходимости

$$\begin{aligned} M \left| \int_{-\infty}^{\infty} H(-iu) \xi(du) - \int_{-\infty}^{\infty} S_n \xi(du) \right|^2 = \\ = \int_{-\infty}^{\infty} |H(-iu) - S_n|^2 F(du) \rightarrow 0, \end{aligned}$$

если $\Delta t_k \rightarrow 0$ ($-a = t_0 < t_1 < \dots < t_n = a$), поскольку $H(-iu) - S_n \rightarrow 0$ при каждом u и функции $H(-iu) - S_n$ равномерно ограничены. Итак,

$$\int_{-a}^a h(t) \xi(t) dt = \int_{-\infty}^{\infty} H_a(-iu) \xi(du), \quad H_a(-iu) = \int_{-a}^a h(t) e^{iut} dt.$$

Нетрудно заметить, что в полученном равенстве можно перейти к пределу при $a \rightarrow \infty$ (в смысле сходимости в L_2).

Теорема 3. Пусть слабо стационарный процесс $\xi(t)$ со спектральным разложением (7) подается на вход фильтра с частотной характеристикой $H(iu)$. Тогда процесс $\eta(t)$ на выходе фильтра имеет спектральное разложение

$$\eta(t) = \int_{-\infty}^{\infty} e^{itu} H(iu) \xi(du). \quad (9)$$

Доказательство. Допустим сначала, что фильтр обладает интегрируемой импульсной реакцией $h(t)$, т. е.

$$\eta(t) = \int_{-\infty}^{\infty} h(t-s) \xi(s) ds.$$

На основании леммы 1

$$\begin{aligned} \eta(t) = \int_{-\infty}^{\infty} \tilde{H}(-iu) \xi(du), \quad \text{где } \tilde{H}(-iu) = \int_{-\infty}^{\infty} h(t-s) e^{iut} ds = \\ = e^{iut} H(iu). \end{aligned}$$

В общем случае $\eta(t) = \text{l. i. m. } \eta_n(t)$, где $\eta_n(t)$ — функция на выходе фильтра с частотной характеристикой $H_n(iu)$ и интегрируемой импульсной реакцией, причем $H_n(iu) \rightarrow H(iu)$ в L_2^F . Следовательно,

$$\eta(t) = \text{l. i. m. } \int_{-\infty}^{\infty} e^{itu} H_n(iu) \xi(du) = \int_{-\infty}^{\infty} e^{itu} H(iu) \xi(du),$$

$$\left| \int_{-\infty}^{\infty} e^{itu} H_n(iu) \zeta(du) - \int_{-\infty}^{\infty} e^{itu} H(iu) \zeta(du) \right|^2 = \\ = \int_{-\infty}^{\infty} |H_n(iu) - H(iu)|^2 F(du) \rightarrow 0.$$

Примеры фильтров и их частотных характеристик

1. Тожественный фильтр — устройство, не влияющее на функцию на входе. Частотная характеристика фильтра $H(iu) \equiv 1$. Импульсная реакция не определена. Это ясно из смысла импульсной реакции, а математически это вытекает из свойств преобразования Фурье: если $H(iu)$ — преобразование Фурье интегрируемой функции, то $H(iu) \rightarrow 0$ при $u \rightarrow \pm \infty$.

2. Полосовой фильтр пропускает (не изменяя их) только гармонические составляющие процессов с частотами u , для которых $a < |u| < b$, $a > 0$. Частотная характеристика фильтра равна $H(iu) = I_{(a, b)}(u) + I_{(-b, -a)}(u)$, и фильтр является допустимым для произвольного процесса. Импульсная реакция находится по формуле Фурье

$$h(t) = \frac{1}{2\pi} \left(\int_{-b}^{-a} + \int_a^b \right) e^{itu} du = \frac{\sin bt - \sin at}{\pi t}.$$

3. Фильтр высоких частот подавляет низкие частоты, не изменяя высоких. Он является предельным для полосового фильтра при $b \rightarrow \infty$. Его частотная характеристика $H(iu) = I_{\{|u| > a\}}(u)$, а импульсная реакция не существует.

4. Операция среднеквадратического дифференцирования слабо стационарного процесса. Для существования среднеквадратической производной процесса $\xi(t)$ достаточно существования $R''(0)$. Это условие эквивалентно требованию

$$\int_{-\infty}^{\infty} u^2 F(du) < \infty. \quad (10)$$

Кроме того, если это условие выполнено, то при $h \rightarrow 0$ $\frac{e^{ihu} - 1}{h} \rightarrow iu$ (в L_2^F) и в соотношении

$$\frac{\xi(t+h) - \xi(t)}{h} = \int_{-\infty}^{\infty} e^{itu} \frac{e^{ihu} - 1}{h} \zeta(du)$$

можно перейти к пределу при $h \rightarrow 0$ под знаком стохастического интеграла. Следовательно,

$$\xi'(t) = \int_{-\infty}^{\infty} e^{iut} iu \zeta(du). \quad (11)$$

Таким образом, операции дифференцирования соответствует фильтр с частотной характеристикой iu , который является допустимым для всех стационарных процессов, удовлетворяющих условию (10). Импульсная реакция не существует, но фильтр можно рассматривать как предельный ($\varepsilon \rightarrow 0$) для физически осуществимых фильтров с импульсными реакциями $h_\varepsilon(t) = 0$, если $t < 0$, $h_\varepsilon(t) = \frac{1}{\varepsilon^2}$ при $t \in (0, \varepsilon]$ и $h_\varepsilon(t) = -\frac{1}{\varepsilon^2}$ при $t \in (\varepsilon, 2\varepsilon]$, которым соответствуют частотные характеристики

$$H_\varepsilon(iu) = -e^{-iu\varepsilon} \frac{4 \sin^2 \frac{u\varepsilon}{2}}{iu\varepsilon^2}.$$

5. Операция сдвига времени. Так как

$$\xi(t-s) = \int_{-\infty}^{\infty} e^{iut} e^{-ius} \zeta(du),$$

то операции сдвига времени θ_s , $\theta_s(\xi|t) = \xi(t-s)$, соответствует частотная характеристика $H(iu) = e^{-iut}$. Импульсная реакция не определена.

6. Дифференциальные уравнения. Рассмотрим фильтр, определяемый линейным дифференциальным уравнением с постоянными коэффициентами

$$Q\eta = P\xi, \quad (12)$$

где

$$Q = Q\left(\frac{d}{dt}\right) = \alpha_0 \frac{d^n}{dt^n} + \alpha_1 \frac{d^{n-1}}{dt^{n-1}} + \dots + \alpha_n,$$

$$P = P\left(\frac{d}{dt}\right) = b_0 \frac{d^m}{dt^m} + b_1 \frac{d^{m-1}}{dt^{m-1}} + \dots + b_m.$$

Уравнение (12) имеет смысл только тогда, когда процесс $\xi(t)$ в среднеквадратическом дифференцируем m раз. Тогда мы ищем n раз дифференцируемый в среднеквадратическом стационарный процесс $\eta(t)$, удовлетворяющий уравнению (12). Предположим, что (12) имеет стационарное решение. Представим его в виде

$$\eta(t) = \int_{-\infty}^{\infty} e^{iut} H(iu) \zeta(du).$$

Применяя к процессам $\xi(t)$ и $\eta(t)$ операции P и Q соответственно, получаем

$$\int_{-\infty}^{\infty} e^{iut} Q(iu) H(iu) \xi(du) = \int_{-\infty}^{\infty} e^{iut} P(iu) \xi(du),$$

где $Q(iu) = \sum_{k=0}^n a_k(iu)^{n-k}$, $P(iu) = \sum_{k=0}^m b_k(iu)^{m-k}$. Если $Q(iu)$ не имеет действительных корней, то

$$H(iu) = \frac{P(iu)}{Q(iu)}, \quad (13)$$

и, наоборот, если процесс $\xi(t)$ дифференцируем в среднее квадратическом m раз, то $P(iu) \in L_2^F$, $Q(iu) \neq 0$ ($-\infty < u < \infty$), то процесс

$$\eta(t) = \int_{-\infty}^{\infty} e^{iut} \frac{P(iu)}{Q(iu)} \xi(du)$$

дифференцируем в среднее квадратическое n раз и удовлетворяет уравнению (12). Таким образом, при условии $P(iu) \in L_2^F$, $Q(iu) \neq 0$, существует фильтр, у которого процесс $\eta(t)$ на выходе слабо стационарен и удовлетворяет дифференциальному уравнению (12). Заметим, однако, что можно определить решение уравнения (12) и в более общих случаях. Допустим, что многочлен $Q(iu)$ не имеет действительных корней. Фильтр с частотной характеристикой $\frac{P(iu)}{Q(iu)}$ существует и без требования $P(iu) \in L_2^F$, достаточно только, чтобы $\frac{P(iu)}{Q(iu)} \in L_2^F$. Последнее выполняется всегда, когда степень n многочлена Q не меньше m . Таким образом, при $n \geq m$ фильтр с частотной характеристикой (13), знаменатель которого не обращается в нуль при действительных u , является допустимым для произвольного процесса на входе, и процесс на выходе фильтра будет слабо стационарным решением уравнения (12). Ограничиваясь, по-прежнему, дифференциальными уравнениями, для которых многочлен $Q(x)$ не имеет чисто мнимых корней, выделим из дробно-рациональной функции $\frac{P(x)}{Q(x)}$ целую часть $P_0(x)$ (она отлична от нуля, если $m \geq n$) и остаток разложим на простые дроби. Тогда

$$\frac{P(iu)}{Q(iu)} = P_0(iu) + \sum_{k=1}^{n'} \sum_{s=1}^{l'_k} \frac{c'_{ks}}{(iu - p'_k)^s} + \sum_{k=1}^{n''} \sum_{s=1}^{l''_k} \frac{c''_{ks}}{(iu - p''_k)^s},$$

где $P_0(iu) = \sum_{k=0}^{m-1} a_k(iu)^k$ при $m > n$, и $P_0(iu) = 0$ ($m < n$), $\operatorname{Re} p'_k < 0$ и $\operatorname{Re} p''_k > 0$, p'_k и p''_k являются корнями многочлена $Q(x)$. Поскольку

$$\frac{1}{(iu - p)^s} = \frac{1}{(s-1)!} \frac{d^{s-1}}{dp^{s-1}} \int_0^{\infty} e^{pt} e^{-iat} dt = \int_0^{\infty} \frac{t^{s-1}}{(s-1)!} e^{pt} e^{-iat} dt \quad (\operatorname{Re} p < 0)$$

и

$$\frac{1}{(iu - p)^s} = - \int_{-\infty}^0 \frac{t^{s-1}}{(s-1)!} e^{pt} e^{-iat} dt \quad (\operatorname{Re} p > 0),$$

то процесс $\eta(t)$ на выходе фильтра можно представить в виде

$$\eta(t) = \sum_{k=0}^{m-1} a_k \xi^{(k)}(t) + \int_0^{\infty} \xi(t-s) G_1(s) ds + \int_0^{\infty} \xi(t+s) G_2(-s) ds,$$

где

$$G_1(t) = \sum_{k=1}^{n'} \left(\sum_{l=1}^{l'_k} \frac{c_{kl}^1 t^l}{(s-1)!} \right) e^{p_k^1 t} \quad (t > 0),$$

$$G_2(t) = - \sum_{k=1}^{n''} \left(\sum_{l=1}^{l'_k} \frac{c_{kl}^2 t^l}{(s-1)!} \right) e^{p_k^2 t} \quad (t < 0).$$

Заметим, что если многочлен $Q(x)$ имеет корни с положительной действительной частью, то соответствующий фильтр физически не осуществим.

§ 5. Обобщенные случайные процессы

Обобщенный случайный процесс. Одной из целей теории случайных процессов является построение различных моделей случайных процессов, исходя из простейших процессов. В некоторых отношениях простейшим процессом является процесс, называемый «белым шумом». К сожалению, даваемые определения этого процесса часто, особенно в технической литературе, бывают неудовлетворительными. Сам термин «белый шум» должен означать, что этот процесс является стационарным, все гармонические составляющие которого входят с одинаковой интенсивностью, т. е. если $F(u)$ — спектральная функция процесса, то должно быть $F(du) = \mathcal{I} du$. Однако в этом случае формула, дающая спектральное представление корреляционной функции

процесса, лишена смысла. Часто определяют «белый шум» как действительный стационарный процесс с некоррелированными значениями, полагая

$$M\xi(t)\xi(s) = 0 \text{ при } t \neq s \\ M|\xi(t)|^2 = \infty.$$

Это определение также неудовлетворительно. Иногда говорят, что «белый шум» — это производная от винеровского процесса. Но винеровский процесс недифференцируем. Все же, несмотря на явную логическую противоречивость приведенных определений, они содержат в себе существенную и важную информацию о «белом шуме» и можно дать разумное определение «белого шума», при котором он обладает тремя перечисленными выше качествами. Разумеется, определенный при этом объект не является случайным процессом в обычном смысле этого слова, а обобщенным случайным процессом, обобщающим понятие случайного процесса в том же смысле, в котором понятие обобщенной функции (или распределения по Л. Шварцу) обобщает понятие обычной функции.

В дальнейшем мы ограничимся лишь самыми необходимыми определениями и свойствами обобщенных случайных процессов.

Пусть $\varphi(t)$ — функция, определенная на прямой $R = (-\infty, \infty)$. Носителем функции $\varphi(t)$ называют наименьшее замкнутое множество, вне которого $\varphi(t)$ равно нулю. Введем пространство \mathcal{D} — пространство всех комплекснозначных бесконечно дифференцируемых функций на R с ограниченным носителем. В пространстве \mathcal{D} следующим образом введем топологию:

последовательность φ_n , $\varphi_n \in \mathcal{D}$, $n = 1, 2, \dots$, сходится к $\varphi \in \mathcal{D}$, если: а) все носители функций φ_n содержатся в некотором ограниченном множестве; б) φ_n вместе со всеми производными $\varphi_n^{(k)}$ равномерно сходятся к φ и к соответствующим производным $\varphi^{(k)}$ ($k = 1, 2, \dots$). Подпространство \mathcal{D} , состоящее из действительных функций φ , обозначим \mathcal{D}_r .

Определение. Обобщенным случайным процессом $\Lambda(\varphi)$ или $\langle \Lambda, \varphi \rangle$ называют случайный линейный непрерывный функционал, определенный на \mathcal{D} . Это означает, что каждому $\varphi \in \mathcal{D}$ поставлена в соответствие некоторая случайная величина $\Lambda(\varphi)$ (вообще говоря, комплекснозначная), определенная на фиксированном вероятностном пространстве $\{\Omega, \mathcal{F}, P\}$, так что:

а) $\Lambda(c_1\varphi_1 + c_2\varphi_2) = c_1\Lambda(\varphi_1) + c_2\Lambda(\varphi_2) \forall \varphi_i \in \mathcal{D}$ и констант c_i , $i = 1, 2$,

б) если φ_n сходится к φ в \mathcal{D} , то $\Lambda(\varphi_n)$ сходится по вероятности к $\Lambda(\varphi)$.

В настоящей книге рассматриваются только обобщенные случайные процессы $\Lambda(\varphi)$, удовлетворяющие условиям:

а) для всех $\varphi \in \mathcal{D}$ $M|\Lambda(\varphi)|^2 < \infty$,

б) если $\varphi_n \rightarrow \varphi$ в \mathcal{D} , то $M|\Lambda(\varphi_n) - \Lambda(\varphi)|^2 \rightarrow 0$, т. е. $\Lambda(\varphi) =$ л. и. м. $\Lambda(\varphi_n)$.

Поэтому, говоря в дальнейшем об обобщенном случайном процессе, мы будем всегда предполагать, даже если это явно не выражено, что условия а), б) выполнены.

Пусть $\xi(t)$ непрерывный в среднеквадратическом процесс. Положим

$$\langle \xi, \varphi \rangle = \int_{-\infty}^{\infty} \xi(t) \varphi(t) dt. \quad (1)$$

Этот интеграл имеет смысл, причем (см. § 1).

$$M|\langle \xi, \varphi \rangle|^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(t) B(t, s) \overline{\varphi(s)} dt ds,$$

где $B(t, s)$ — ковариация процесса $\xi(t)$. Если $\varphi_n \rightarrow \varphi$ в \mathcal{D} , то

$$M|\langle \Lambda_\varepsilon, \varphi - \varphi_n \rangle|^2 \leq \int_{-a}^a B(t, t) dt \int_{-a}^a |\varphi - \varphi_n|^2 dt \rightarrow 0,$$

где $(-a, a)$ — интервал, содержащий все носители функций φ_n . Итак, каждый непрерывный в среднеквадратическом случайный процесс порождает обобщенный процесс $\langle \xi, \varphi \rangle$, удовлетворяющий условиям а) и б). Отметим, что обобщенный процесс $\langle \xi, \varphi \rangle$ однозначно определяет процесс $\xi(t)$. Действительно, если существуют два непрерывных в среднеквадратическом процесса $\xi(t)$ и $\eta(t)$, которым соответствует, согласно формуле (1), один и тот же обобщенный процесс $\langle \xi, \varphi \rangle$, то

$$\int_{-\infty}^{\infty} \xi(t) \varphi(t) dt = 0 \text{ для } \varphi \in \mathcal{D}, \quad \xi(t) = \xi(t) - \eta(t)$$

для каждой функции $\varphi \in \mathcal{D}$, где $\xi(t) = \xi(t) - \eta(t)$. Отсюда

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(t) B_\xi(t, s) \overline{\varphi(s)} dt ds = 0 \quad \forall \varphi \in \mathcal{D}. \quad (2)$$

Если при некотором t_0 выполняется неравенство $B_\xi(t_0, t_0) > 0$, то, на основании непрерывности функции $B_\xi(t, s)$, можно найти такое h , что $\operatorname{Re} B(t, s) > 0$ при $|t - t_0| < h$, $|s - t_0| < h$. Существует функция $\varphi_0 \in \mathcal{D}$, такая, что $\varphi(t) \equiv 0$ при $|t - t_0| \geq h$ и $\varphi(t) > 0$ при $|t - t_0| < h$. Тогда

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(t) \operatorname{Re} B_\xi(t, s) \varphi(s) dt ds > 0,$$

что противоречит (2). Итак, $B_\xi(t, t) = 0 \quad \forall t > 0$ и $\xi(t) = \xi(t) - \eta(t) = 0$ с вероятностью 1 при каждом t . Следовательно, обобщенный процесс Λ , если он имеет вид $\langle \xi, \varphi \rangle$, однозначно определяет процесс $\xi(t)$. В дальнейшем мы часто будем отождествлять

взяв непрерывный в среднеквадратическом случайный процесс $\xi(t)$ с порождаемым им обобщенным случайным процессом $\xi(\varphi) = \langle \xi, \varphi \rangle$ и говорить в этом случае, что обобщенный процесс $\xi(\varphi)$ является процессом обычного типа.

Для обобщенного случайного процесса $\Lambda(\varphi)$ введем два функционала на \mathcal{D} . Первый из них

$$m(\varphi) = M\Lambda(\varphi)$$

называют средним значением процесса Λ . Очевидно, он линеен и непрерывен на \mathcal{D} .

Ковариацией обобщенного случайного процесса называют функционал

$$B(\varphi, \psi) = M\Lambda(\varphi)\overline{\Lambda(\psi)}.$$

Он билинеен

$$\begin{aligned} B(a_1\varphi_1 + a_2\varphi_2, \psi) &= a_1B(\varphi_1, \psi) + a_2B(\varphi_2, \psi), \\ B(\varphi, a_1\psi_1 + a_2\psi_2) &= \bar{a}_1B(\varphi, \psi_1) + \bar{a}_2B(\varphi, \psi_2), \end{aligned}$$

симметричен

$$B(\varphi, \psi) = \overline{B(\psi, \varphi)}$$

и непрерывен в \mathcal{D} : из $\varphi_n \rightarrow \varphi$ и $\psi_n \rightarrow \psi$ в \mathcal{D} вытекает

$$B(\varphi_n, \psi_n) \rightarrow B(\varphi, \psi).$$

Для доказательства непрерывности ковариации замечаем, что при $\varphi_n \rightarrow \varphi$, $\psi_n \rightarrow \psi$ (в \mathcal{D})

$$\begin{aligned} |B(\varphi_n, \psi_n) - B(\varphi, \psi)| &= |M(\Lambda(\varphi_n)\overline{\Lambda(\psi_n)} - \\ &- \Lambda(\varphi)\overline{\Lambda(\psi)})| = |M[(\Lambda(\varphi) - \Lambda(\varphi_n))(\overline{\Lambda(\psi) - \Lambda(\psi_n)}) + \\ &+ \Lambda(\varphi)\overline{(\Lambda(\psi) - \Lambda(\psi_n))} + \Lambda(\varphi_n)\overline{\Lambda(\psi)} - \Lambda(\varphi_n)\overline{\Lambda(\psi)}]| \leq \\ &\leq (M|\Lambda(\varphi) - \Lambda(\varphi_n)|^2 M|\Lambda(\psi) - \Lambda(\psi_n)|^2)^{1/2} + (M|\Lambda(\varphi)|^2 \times \\ &\times M|\Lambda(\psi) - \Lambda(\psi_n)|^2)^{1/2} + (M|\Lambda(\varphi_n)|^2 M|\Lambda(\psi)|^2)^{1/2} \rightarrow 0 \end{aligned}$$

на основании свойства непрерывности обобщенного процесса.

Аналогичными свойствами обладает корреляционный функционал $R(\varphi, \psi)$ обобщенного случайного процесса,

$$\begin{aligned} R(\varphi, \psi) &= M(\Lambda(\varphi) - m(\varphi))\overline{(\Lambda(\psi) - m(\psi))} = \\ &= B(\varphi, \psi) - m(\varphi)\overline{m(\psi)}. \end{aligned}$$

Если обобщенный процесс $\Lambda(\varphi)$ является процессом обычного типа, $\Lambda(\varphi) = \langle \xi, \varphi \rangle$, то для ковариационного функционала имеем следующее выражение:

$$B(\varphi, \psi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(t) B(t, s) \overline{\psi(s)} dt ds, \quad (3)$$

где $B(t, s)$ — ковариация процесса $\xi(t)$. Заметим еще, что билинейный симметрический функционал $B(\varphi, \psi)$ однозначно восстанавливается по квадратическому $B(\varphi, \varphi)$.

Характеристический функционал обобщенного случайного процесса. Характеристическим функционалом действительного обобщенного случайного процесса $\Lambda(\varphi)$ называют $\Phi(\varphi)$ в \mathcal{D} , определяемую равенством

$$\Phi(\varphi) = M e^{i\Lambda(\varphi)},$$

где $\varphi \in \mathcal{D}$ — действительная функция ($\varphi \in \mathcal{D}_r$).

Очевидно, характеристический функционал однозначно определяет всевозможные совместные распределения величин

$$\Lambda(\varphi_1), \Lambda(\varphi_2), \dots, \Lambda(\varphi_n). \quad (4)$$

В самом деле, характеристическая функция $\Phi_n(u_1, u_2, \dots, u_n)$ совместного распределения последовательности случайных величин (4) равна $\Phi(u_1\varphi_1 + u_2\varphi_2 + \dots + u_n\varphi_n)$, и она однозначно определяет распределение этих величин.

Действительный обобщенный случайный процесс называют гауссовским, если

$$\Phi(\varphi) = e^{im(\varphi) - \frac{1}{2} R(\varphi, \varphi)},$$

где $m(\varphi)$ — некоторый линейный непрерывный в \mathcal{D} функционал, а $R(\varphi, \psi)$ — действительный билинейный симметричный непрерывный в \mathcal{D} функционал, такой, что $R(\varphi, \varphi) \geq 0 \forall \varphi \in \mathcal{D}$. Если $\xi(t)$ — обычный действительный гауссовский непрерывный в среднеквадратическом процесс с

$$M\xi(t) = m(t), \quad M(\xi(t) - m(t))(\xi(s) - m(s)) = R(t, s),$$

то функции $m(t)$ и $R(t, s)$ непрерывны (и обратно, если $m(t)$ и $R(t, s)$ непрерывны, то процесс $\xi(t)$ непрерывен в среднеквадратическом), и порождаемый им обобщенный процесс $\langle \xi, \varphi \rangle$ является гауссовским.

Дифференцирование обобщенных процессов. Определим понятие производной $\frac{d\Lambda}{dt}$ обобщенного случайного процесса.

Определение. Производной $\frac{d\Lambda}{dt}$ называют обобщенный случайный процесс

$$\left\langle \frac{d\Lambda}{dt}, \varphi \right\rangle = - \left\langle \Lambda, \frac{d\varphi}{dt} \right\rangle, \quad (5)$$

и вообще k -й производной $\frac{d^k\Lambda}{dt^k}$ обобщенного процесса Λ называют процесс

$$\left\langle \frac{d^k\Lambda}{dt^k}, \varphi \right\rangle = (-1)^k \left\langle \Lambda, \frac{d^k\varphi}{dt^k} \right\rangle. \quad (6)$$

Очевидно, если Λ — обобщенный процесс, то все его производные также являются обобщенными случайными процессами.

Предположим, что обобщенный случайный процесс $\Lambda(\varphi) = \langle \xi, \varphi \rangle$ является процессом обычного типа, причем, процесс $\xi(t)$ непрерывно дифференцируем. Нетрудно убедиться, что для рассматриваемых интегралов остается справедливой формула интегрирования по частям

$$\int_{-\infty}^{\infty} \xi(t) \frac{d\varphi}{dt} dt = - \int_{-\infty}^{\infty} \xi'(t) \varphi(t) dt.$$

Следовательно, в рассматриваемом случае

$$\left\langle \frac{d\Lambda_{\xi}}{dt}, \varphi \right\rangle = - \int_{-\infty}^{\infty} \xi'(t) \varphi(t) dt,$$

т. е. обобщенный процесс $\frac{d\Lambda_{\xi}}{dt}$ порождается процессом $\xi'(t)$. Таким образом, определение производной обобщенного случайного процесса согласовано с определением среднеквадратической производной обычного среднеквадратического дифференцируемого случайного процесса. С другой стороны, произвольный обобщенный случайный процесс всегда обладает производными любого сколь угодно высокого порядка, которые также являются обобщенными случайными процессами.

Из определения гауссовского обобщенного процесса непосредственно вытекает, что его производные любого порядка также являются гауссовскими обобщенными процессами.

Приведенное определение обобщенного случайного процесса может быть применено и к тому случаю, когда случайные величины $\Lambda(\varphi)$ вырождаются в константы: $\Lambda(\varphi) = M\Lambda(\varphi) = L(\varphi)$. Такой вырожденный обобщенный случайный процесс называют обобщенной функцией. Итак, обобщенная функция $L = L(\varphi) = \langle L, u \rangle$ — это линейный непрерывный функционал на \mathcal{D} . Если

$$L(\varphi) = \int_{-\infty}^{\infty} l(t) \varphi(t) dt,$$

где $l(t)$ — непрерывная (или даже произвольная измеримая функция, интегрируемая на произвольном конечном отрезке), то $L(\varphi)$ называют обобщенной функцией обычного типа и отождествляют с $l(t)$.

Производные от обобщенных функций определяются равенствами (5) и (6) точно так же, как и производные обобщенных случайных процессов.

Простейшим примером обобщенной функции, не являющейся функцией обычного типа, является δ -функция Дирака, она определяется равенством

$$\delta(\varphi) = \varphi(0).$$

δ -функцией в точке a называют обобщенную функцию δ_a , для которой

$$\delta_a(\varphi) = \varphi(a).$$

Часто обобщенную функцию $\delta_a(\varphi)$ записывают в виде

$$\delta_a(\varphi) = \int_{-\infty}^{\infty} \delta(t-a) \varphi(t) dt. \quad (7)$$

При этом интеграл в правой части равенства имеет чисто символическое значение, так как действительной функции, для которой выполняется равенство (7), не существует. В соответствии с общим определением производных производные от δ -функции определяются равенствами

$$\langle \delta', \varphi \rangle = -\varphi'(0), \quad \langle \delta^{(n)}, \varphi \rangle = (-1)^n \varphi^{(n)}(0).$$

Обозначим через $Y(t)$ единичный импульс, $Y(t) = 0$ при $t < 0$, $Y(t) = 1$ при $t \geq 0$ и через $Y(\varphi)$ соответствующую обобщенную функцию

$$Y(\varphi) = \int_0^{\infty} \varphi(t) dt.$$

Тогда $Y'(\varphi) = Y(\varphi') = \varphi(0)$. Таким образом, $Y' = \delta$.

Возвратимся к случайным процессам.

«Белый шум». Назовем процесс с некоррелированными приращениями стандартным, если он удовлетворяет условиям

$$\xi(0) = 0, \quad M\xi(t) = 0, \quad M|\xi(t)|^2 = t, \quad t \geq 0, \quad (8)$$

При $t < 0$ будем считать $\xi(t) = 0$. Если $t > s > 0$, то $R(t, s) = M\xi(t)\overline{\xi(s)} = M(\xi(t) - \xi(s))\overline{(\xi(s) - \xi(0))} + M|\xi(s)|^2 = s$.

Итак, при $t, s > 0$ $R(t, s) = \min(t, s)$. Выше уже отмечалось, что этот процесс в L_2 не дифференцируем. Найдем корреляционный функционал производной от процесса $\xi(t)$, рассматриваемой как обобщенный процесс. Сначала преобразуем выражение для корреляционного функционала процесса $\xi(t)$. Имеем

$$\begin{aligned} R(\varphi, \varphi) &= \int_0^{\infty} \int_0^{\infty} \min(t, s) \varphi(t) \overline{\varphi(s)} dt ds = \\ &= \int_0^{\infty} \varphi(t) \left(\int_0^t s \overline{\varphi(s)} ds + t \int_t^{\infty} \overline{\varphi(s)} ds \right) dt = \end{aligned}$$

$$= \int_0^{\infty} t \varphi(t) \int_t^{\infty} \overline{\varphi(s)} ds dt + \int_0^{\infty} s \overline{\varphi(s)} \int_s^{\infty} \varphi(t) dt ds.$$

Положим

$$\psi(s) = \int_s^{\infty} \varphi(t) dt.$$

Тогда

$$R(\varphi, \varphi) = - \int_0^{\infty} t (\psi'(t) \overline{\psi(t)} + \overline{\psi'(t)} \psi(t)) dt = - \int_0^{\infty} t d(\psi(t) \overline{\psi(t)}),$$

откуда следует, что

$$R(\varphi, \varphi) = \int_0^{\infty} |\psi(t)|^2 dt.$$

Обозначим через $R_0(\varphi, \varphi)$ корреляционный функционал обобщенного процесса $\frac{d\xi}{dt}$. Тогда

$$R_0(\varphi, \varphi) = M \left| \frac{d\xi}{dt}(\varphi) \right|^2 = M |\xi(\varphi')|^2 = R(\varphi', \varphi')$$

и, следовательно,

$$R_0(\varphi, \varphi) = \int_0^{\infty} |\varphi(t)|^2 dt. \quad (9)$$

Формулу (9) можно записать в виде

$$R_0(\varphi, \varphi) = \int_0^{\infty} \int_0^{\infty} \delta(t-s) \varphi(t) \overline{\varphi(s)} dt ds.$$

Это равенство можно интерпретировать следующим образом: корреляционная функция слабого «белого шума» $R(t, s) = \delta(t-s)$, т. е. слабый белый шум является стационарным процессом с некоррелированными значениями.

Определение. Обобщенный процесс $\frac{d\xi}{dt}$, где ξ — стандартный процесс с некоррелированными приращениями, называют слабым белым шумом. Если $\xi(t)$ — действительный процесс с независимыми приращениями, удовлетворяющий условиям (8), то обобщенный процесс $\frac{d\xi}{dt}$ называют «белым шумом».

В частности, гауссовский белый шум — это производная от винеровского процесса. Она является обобщенным гауссовским процессом с характеристическим функционалом

$$\Phi_0(\varphi) = \exp \left\{ -\frac{1}{2} \int_0^{\infty} \varphi^2(t) dt \right\}, \quad \varphi \in \mathcal{D}. \quad (10)$$

Укажем еще вид характеристического функционала пуассоновского белого шума. По определению, пуассоновский «белый шум» — это производная от центрированного пуассоновского процесса

$$\pi_0(t) = \pi(t) - t \quad t \geq 0,$$

где $\pi(t)$ — пуассоновский процесс со средним $M\pi(t) = t$. Рассмотрим более общий случай процесса

$$\bar{\pi}(t) = \pi(t) - at,$$

где $\pi(t)$ — пуассоновский процесс со средним $M\pi(t) = at$. Заметим, что

$$\int_0^\infty \psi(t) \bar{\pi}(t) dt = -\psi(t) \pi'(t) \Big|_0^\infty + \int_0^\infty \psi(t) d\bar{\pi}(t) = \int_0^\infty \psi(t) d\bar{\pi}(t),$$

где $\psi(t) = \int_t^\infty \varphi(s) ds$. Для характеристического функционала процесса $\bar{\pi}(t)$ получаем, учитывая, что процесс $\bar{\pi}(t)$ является процессом с независимыми приращениями, следующие соотношения:

$$\begin{aligned} \bar{\Phi}(\varphi) &= M \exp \left\{ i \int_0^\infty \varphi(t) \bar{\pi}(t) dt \right\} = M \exp \left\{ i \int_0^\infty \psi(t) d\bar{\pi}(t) \right\} = \\ &= \lim M \exp \left\{ i \sum_{k=1}^n \psi(t_k) \Delta \pi(t_k) \right\} = \lim \prod_{k=1}^n M \exp \{ i \psi(t_k) \Delta \pi(t_k) \}. \end{aligned}$$

При этом надо иметь в виду, что функция $\psi(t)$, начиная с некоторого $t > 0$, обращается в 0, $\psi(t) = 0$ при $t \geq a$, а знак \lim означает предел по последовательности разбиений отрезка $[0, a]$ точками $0 < t_1, t_2, \dots, t_n = a$, $\Delta \pi(t_k) = \pi(t_k) - \pi(t_{k-1})$. Далее, для центрированного пуассоновского процесса имеем

$$M \exp \{ i a [\bar{\pi}(t + \Delta t) - \bar{\pi}(t)] \} = e^{a \Delta t} (e^{i a \Delta t} - 1 - i a \Delta t).$$

Таким образом,

$$\bar{\Phi}(\varphi) = \lim \prod_{k=1}^n \exp \{ i a \Delta t_k (e^{i \psi(t_k)} - 1 - i \psi(t_k)) \},$$

или

$$\bar{\Phi}(\varphi) = \exp \left\{ i a \int_0^\infty [e^{i \psi(t)} - 1 - i \psi(t)] dt \right\}. \quad (11)$$

Из последней формулы непосредственно получаем выражение для характеристического функционала $\Phi_\pi(\varphi)$ пуассоновского «белого шума»

$$\Phi_\pi(\varphi) = \exp \left\{ i \int_0^\infty (e^{i \varphi(t)} - 1 - i \varphi(t)) dt \right\}. \quad (12)$$

Хотя для обобщенного процесса значения процесса в фиксированный момент времени не определено, все же можно дать естественное определение обобщенного процесса с независимыми значениями и показать, что белый шум является таковым.

Определение. Действительный обобщенный процесс $\Lambda(\varphi)$ называют процессом с независимыми значениями, если для любых $\varphi_i \in \mathcal{D}_r$, $i = 1, 2$, для которых $\varphi_1 \varphi_2 \equiv 0$, случайные величины $\Lambda(\varphi_1)$ и $\Lambda(\varphi_2)$ независимы.

Теорема 1. Белый шум является процессом с независимыми значениями.

Доказательство. Пусть $\varphi_1(t) \varphi_2(t) \equiv 0$, $\varphi_i(t) \in \mathcal{D}_r$. Множество точек, где $\varphi_i(t) > 0$, представляет собой сумму не более чем счетного множества интервалов $\Delta_{ik} = (a_{ik}, b_{ik})$, $i = 1, 2$, $k = 1, 2, \dots$ попарно без общих точек. При этом ни один из интервалов Δ_{ik} не имеет общих точек ни с одним из φ_{2k} . Следовательно, $\varphi_i(t) = \sum_k \varphi_{ik}(t)$, $i = 1, 2$, где $\varphi_{ik}(t) \in \mathcal{D}_r$ и $\varphi_{ik}(t) = 0$ при $t \notin \Delta_{ik}$. Обозначим через γ белый шум, $\gamma = \frac{d\xi}{dt}$, где ξ — процесс с независимыми приращениями. Заметим, что

$$\begin{aligned} \xi(\varphi'_{ik}) &= \int_{a_{ik}}^{b_{ik}} \varphi'_{ik}(t) \xi(t) dt = \varphi_{ik}(t) \xi(t) \Big|_{a_{ik}}^{b_{ik}} - \\ &- \int_{a_{ik}}^{b_{ik}} \varphi_{ik}(t) \xi(dt) = \int_{a_{ik}}^{b_{ik}} \varphi_{ik}(t) \xi(dt). \end{aligned}$$

Поэтому семейство $\{\xi(\varphi'_{ik}), i = 1, 2, k = 1, 2, \dots\}$ состоит из независимых случайных величин. Следовательно, величины

$$\xi(\varphi'_1) = \sum_k \xi(\varphi'_{1k}) \text{ и } \xi(\varphi'_2) = \sum_k \xi(\varphi'_{2k})$$

независимы. Учитывая это, имеем

$$\begin{aligned} M \exp \{i(u_1 \gamma(\varphi_1) + u_2 \gamma(\varphi_2))\} &= M \exp \{-i(u_1 \xi(\varphi'_1) + \\ &+ u_2 \xi(\varphi'_2))\} = M \exp \{-iu_1 \xi(\varphi'_1)\} \cdot M \exp \{-iu_2 \xi(\varphi'_2)\} = \\ &= M \exp \{iu_1 \gamma(\varphi_1)\} \cdot M \exp \{iu_2 \gamma(\varphi_2)\}. \end{aligned}$$

Полученное равенство доказывает независимость случайных величин $\gamma(\varphi_1)$ и $\gamma(\varphi_2)$. Таким образом, γ является процессом с независимыми значениями.

Преобразование Фурье обобщенных процессов. Пусть $\xi(t)$ — непрерывный в среднеквадратическом случайный процесс, для которого $\int_{-\infty}^{\infty} \sqrt{B(t, t)} dt < \infty$, где $B(t, s)$ — ковариация процесса $\xi(t)$. Для такого процесса определено преобразование Фурье

$$\tilde{\xi}(t) = \int_{-\infty}^{\infty} e^{it\omega} \xi(u) du. \quad (13)$$

Действительно,

$$\begin{aligned} \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} M e^{itv} \xi(u) e^{-itv} \overline{\xi(v)} du dv \right| &= \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{it(u-v)} B(u, v) du dv \right| \leq \\ &\leq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} V \overline{B(u, u) B(v, v)} du dv = \left| \int_{-\infty}^{\infty} V \overline{B(u, u)} du \right|^2 < \infty, \end{aligned}$$

так что условие существования интеграла (13) выполнено (см. § 1, гл. VIII).

Процесс $\xi(t)$, в свою очередь, непрерывен в среднеквадратическом. Действительно, выкладка, аналогичная предыдущей, дает

$$M |\xi(t+h) - \xi(t)|^2 \leq \left(\int_{-\infty}^{\infty} V \overline{B(u, u)} |e^{ihu} - 1| du \right)^2,$$

причем,

$$M |\xi(t)|^2 \leq \left(\int_{-\infty}^{\infty} V \overline{B(u, u)} du \right)^2.$$

Обобщенный случайный процесс $\tilde{\xi}(\varphi) = \langle \xi, \varphi \rangle$, порожденный процессом $\xi(t)$, определяется формулой

$$\tilde{\xi}(\varphi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{itv} \xi(u) du \varphi(t) dt = \int_{-\infty}^{\infty} \xi(u) \tilde{\varphi}(u) du,$$

или

$$\tilde{\xi}(\varphi) = \xi(\tilde{\varphi}), \quad (14)$$

где $\tilde{\varphi}$ — преобразование Фурье функции $\varphi(t) \in \mathcal{D}$. Полученная формула наталкивает на мысль определить преобразование Фурье произвольного обобщенного процесса с помощью формулы (14). Однако такое определение лишено смысла, так что функция $\tilde{\varphi}(u)$ не принадлежит \mathcal{D} (если $\tilde{\varphi}(u) \not\equiv 0$). Чтобы преодолеть возникшее затруднение, сузим понятие обобщенного случайного процесса.

Определение. Пространство S состоит из тех и только тех комплекснозначных функций $\varphi(t)$, $t \in (-\infty, \infty)$, которые:

а) бесконечно дифференцируемы,
б) на бесконечности убывают вместе со своими производными сколь угодно высокого порядка быстрее, чем $|t|^{-n}$, где n — любое положительное число.

Последнее требование означает, что для любых целых положительных m и n функция $t^n \varphi^{(m)}(t)$ ограничена и интегрируема в любой степени.

Лемма. Если $\varphi(t) \in S$, то и $\tilde{\varphi}(t) \in S$, где $\tilde{\varphi}(t)$ — преобразование Фурье функции $\varphi(t)$.

Действительно, в формуле

$$\tilde{\varphi}(t) = \int_{-\infty}^{\infty} e^{itu} \varphi(u) du$$

можно дифференцировать под знаком интеграла любое число раз,

$$\tilde{\varphi}^{(m)}(t) = \int_{-\infty}^{\infty} (iu)^m e^{itu} \varphi(u) du, \quad (15)$$

так как полученная под знаком интеграла функция непрерывна и интегрируема. Таким образом, функция $\tilde{\varphi}(t)$ бесконечно дифференцируема. С другой стороны, интегрируя n раз по частям, получим равенство

$$\int_{-\infty}^{\infty} e^{itu} \frac{d^n}{du^n} \varphi(u) du = (-it)^n \int_{-\infty}^{\infty} e^{itu} \varphi(u) du = (-it)^n \tilde{\varphi}(t), \quad (16)$$

так как все производные функции $\varphi(u)$ стремятся к 0 при $u \rightarrow \pm \infty$. Последняя формула показывает, что $\tilde{\varphi}(t) \rightarrow 0$ при $t \rightarrow \pm \infty$ быстрее любой степени t . Эти же соображения применимы и к любой производной $\tilde{\varphi}^{(m)}(t)$.

Введем теперь в S топологию.

Определение. Последовательность $\varphi_n(t)$, $\varphi_n \in S$, сходится к φ в S , если для любых целых m и p $(1 + |t|^p) \varphi^{(m)}(t) - \varphi_n^{(m)}(t) \rightarrow 0$ равномерно по t .

Нетрудно заметить, что из $\varphi_n \rightarrow \varphi$ в S вытекает, что преобразования Фурье $\tilde{\varphi}_n$ сходятся к $\tilde{\varphi}$ в S .

Пусть $\{\Omega, \mathcal{F}, \mathbf{P}\}$ — некоторое вероятностное пространство. Предположим, что каждому $\varphi \in S$ поставлена в соответствие случайная величина $\Lambda(\varphi)$, такая, что:

а) $\Lambda(c_1\varphi_1 + c_2\varphi_2) = c_1\Lambda(\varphi_1) + c_2\Lambda(\varphi_2)$, c_i — константы, $\varphi_i \in S$, $i = 1, 2$,

б) из $\varphi_n \rightarrow \varphi$ в S следует л. и. м. $\Lambda(\varphi_n) = \Lambda(\varphi)$. Это соответствие называют умеренным обобщенным случайным процессом. Очевидно, что сужение умеренного обобщенного процесса на \mathcal{D} является обобщенным случайным процессом.

Пусть $\Lambda(\varphi) = \int_{-\infty}^{\infty} \lambda(s) \varphi(s) ds$, где $\lambda(s)$ — обычный непрерывный в среднеквадратическом процесс, $\varphi \in S$. Этот интеграл существует, если ковариация $B(t, s)$ процесса $\lambda(t)$ не очень быстро

растет при $t, s \rightarrow \infty$, например, если для некоторых констант c и m $|B(t, s)| \leq c(1 + |t|^m)$. Действительно, в этом случае

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(t) B(t, s) \overline{\varphi(s)} dt ds < \infty,$$

а выполнение этого условия достаточно для существования рассматриваемого интеграла. Нетрудно заметить, что условие б) в определении умеренного обобщенного процесса выполнено. Слово «умеренный» как раз и означает, что момент второго порядка $B(t, t) = M|\xi(t)|^2$ при $|t| \rightarrow \infty$ должен возрастать не очень быстро.

Определение. Преобразованием Фурье $\tilde{\Lambda}(\varphi)$ умеренного обобщенного процесса называют обобщенный процесс

$$\tilde{\Lambda}(\varphi) = \Lambda(\tilde{\varphi}).$$

Из сказанного выше следует, что это определение имеет смысл, и $\tilde{\Lambda}(\varphi)$ также является умеренным процессом.

Аналогично определим преобразование $\tilde{\tilde{\Lambda}}(\varphi)$, положив

$$\tilde{\tilde{\Lambda}}(\varphi) = \Lambda(\tilde{\tilde{\varphi}}),$$

где

$$\tilde{\tilde{\varphi}}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itu} \varphi(u) du.$$

Так как $\tilde{\tilde{\varphi}} = \tilde{\tilde{\tilde{\varphi}}} = \varphi$, то

$$\tilde{\tilde{\Lambda}}(\varphi) = \tilde{\tilde{\tilde{\Lambda}}}(\varphi) = \Lambda(\varphi),$$

т. е. преобразование $\tilde{\tilde{\Lambda}}$ является обратным к преобразованию $\tilde{\Lambda}$: если $\Lambda_1 = \tilde{\Lambda}$, то $\Lambda = \tilde{\tilde{\Lambda}}_1$.

Эти формулы являются двойственными формулами Фурье для обобщенных случайных процессов.

Обобщенные стационарные процессы. Пусть $\xi(t)$ — среднеквадратический непрерывный слабо стационарный процесс,

$$\xi(t) = \int_{-\infty}^{\infty} e^{itu} \zeta(du)$$

— его спектральное разложение. Заметим, что порождаемый $\xi(t)$ обобщенный процесс $\xi(\varphi)$ можно представить в виде

$$\xi(\varphi) = \int_{-\infty}^{\infty} \tilde{\varphi}(u) \zeta(du). \quad (17)$$

Здесь $\tilde{\varphi}(u)$ — преобразование Фурье функции $\varphi(t)$; $\xi(u)$, $u \in (-\infty, \infty)$ — процесс с ортогональными приращениями, $M|\xi(du)|^2 = F(du)$, $F(-\infty) = 0$, $F(+\infty) < \infty$, $F(u)$ — спектральная функция процесса $\xi(t)$.

Введем понятие слабой стационарности для произвольных обобщенных процессов. Пусть θ_h — оператор сдвига аргумента функции

$$\theta_h \varphi(t) = \varphi(t+h).$$

Если $\varphi \in S(\mathcal{D})$, то $\theta_h \varphi \in S(\mathcal{D})$. Определение оператора сдвига аргумента расширим на обобщенные процессы, положив

$$\theta_h \Lambda(\varphi) = \Lambda(\theta_{-h} \varphi).$$

Назовем обобщенный процесс слабо стационарным, если

$$M\theta_h \xi(\varphi) = m(\varphi), \quad M|\theta_h \xi(\varphi)|^2 = B(\varphi, \varphi),$$

где $m(\varphi)$ и $B(\varphi, \varphi)$ не зависят от h .

Приведем общий прием построения умеренных обобщенных слабо стационарных процессов.

Пусть $\xi(u)$, $u \in (-\infty, \infty)$, — процесс с ортогональными приращениями, со средним $M\xi(u) = 0$ и структурной функцией $F(u)$:

$$M|\xi(u_2) - \xi(u_1)|^2 = F(u_2) - F(u_1) \quad (u_1 < u_2).$$

Мы не предполагаем, что функция $F(u)$ ограничена, но пусть при некотором $p > 0$

$$\int_{-\infty}^{\infty} \frac{F(du)}{1+|u|^p} < \infty. \quad (18)$$

Положим

$$\Lambda(\varphi) = \int_{-\infty}^{\infty} \tilde{\varphi}(u) \xi(du), \quad \varphi \in S, \quad (19)$$

где $\tilde{\varphi}(u)$ — преобразование Фурье функции $\varphi(t)$. Интеграл в правой части равенства существует, причем

$$B(\varphi, \varphi) = M|\Lambda(\varphi)|^2 = \int_{-\infty}^{\infty} |\tilde{\varphi}(u)|^2 F(du) < \infty. \quad (20)$$

Если $\varphi_n \rightarrow \varphi$ в S , то $\tilde{\varphi}_n \rightarrow \tilde{\varphi}$ в S , и поэтому

$$M|\Lambda(\varphi) - \Lambda(\varphi_n)|^2 = \int_{-\infty}^{\infty} |\tilde{\varphi}(u) - \tilde{\varphi}_n(u)|^2 F(du) \rightarrow 0$$

при $u \rightarrow \infty$. Таким образом, $\Lambda(\varphi)$ является умеренным обобщенным процессом. Он слабо стационарен. Действительно, заметим, что

$$\overline{(\theta_{-h}\varphi)}(u) = \int_{-\infty}^{\infty} e^{iut} \varphi(t-h) dt = e^{iuh} \tilde{\varphi}(u).$$

Следовательно,

$$\begin{aligned} M|\theta_{-h}\Lambda(\varphi)|^2 &= M|\Lambda(\theta_{-h}\varphi)|^2 = \\ &= \int_{-\infty}^{\infty} |e^{iuh} \tilde{\varphi}(u)|^2 F(du) = \int_{-\infty}^{\infty} |\tilde{\varphi}(u)|^2 F(du) \end{aligned}$$

не зависит от h . Кроме того, $M\theta_{-h}\Lambda(\varphi) = 0$. Таким образом, для произвольного процесса $\xi(t)$ с ортогональными приращениями, структурная функция которого удовлетворяет условию (16), $M\xi(t) = 0$, т. е. обобщенный процесс $\Lambda(\varphi)$, определяемый формулой (19), является слабо стационарным умеренным обобщенным процессом с корреляционным функционалом (20).

Можно показать и обратное, а именно: произвольный слабо стационарный умеренный обобщенный процесс имеет вид (19), где $\xi(u)$ — процесс с ортогональными приращениями, структурная функция которого удовлетворяет при некотором $p > 0$ условию (18)¹. Формулу (19) при этом называют спектральным разложением процесса, а (20) — спектральным представлением корреляционного функционала, процесс $\xi(u)$ — спектральной стохастической мерой, функция $F(u)$ — спектральной функцией процесса.

Если $F(-\infty) > -\infty$, $F(+\infty) < +\infty$, то $\Lambda(\varphi)$ является процессом обычного типа. Действительно, в этом случае

$$\Lambda(\varphi) = \int_{-\infty}^{\infty} \tilde{\varphi}(a) \xi(du) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{iat} \varphi(t) dt \xi(du) = \int_{-\infty}^{\infty} \varphi(t) \xi(t) dt,$$

где

$$\xi(t) = \int_{-\infty}^{\infty} e^{itd} \xi(du).$$

Процессы с дробно-рациональными спектральными плотностями. Пусть $\xi(\varphi)$ — обобщенный слабо стационарный процесс со спектральным разложением (19), $P(t) = a_0 t^n + at^{n-1} + \dots + a_n$ — некоторый многочлен, $P\left(\frac{d}{dt}\right) = a_0 \frac{d^n}{dt^n} + a_1 \frac{d^{n-1}}{dt^{n-1}} + \dots + a_n$ — соответ-

¹ См. Гельфанд И. Н., Виленькин Н. Я. Некоторые применения гармонического анализа. Оснащенные гильбертовы пространства. М., Физматгиз, 1961.

ствующий этому многочлену дифференциальный оператор. В соответствии с формулами (5), (6) действие оператора $P\left(\frac{d}{dt}\right)$ на процесс $\xi(\varphi)$ определяется формулой

$$P\left(\frac{d}{dt}\right)\xi(\varphi) = \left\langle \xi, P'\left(\frac{d}{dt}\right)\varphi \right\rangle,$$

где

$$P'\left(\frac{d}{dt}\right) = P\left(-\frac{d}{dt}\right) = (-1)^n a_0 \frac{d^n}{dt^n} + (-1)^{n-1} a_1 \frac{d^{n-1}}{dt^{n-1}} + \dots + a_n.$$

Заметим, что из формулы (14) следует

$$P\left(-\frac{d}{dt}\right)\varphi(u) = P(iu)\tilde{\varphi}(u).$$

Поэтому

$$P\left(\frac{d}{dt}\right)\xi(\varphi) = \int_{-\infty}^{\infty} \tilde{\varphi}(u) P(iu) \xi(du). \quad (21)$$

Полученная формула дает спектральное разложение процесса $P\left(\frac{d}{dt}\right)\xi(\varphi)$. Рассмотрим теперь дифференциальное уравнение

$$Q\left(\frac{d}{dt}\right)\eta = \xi, \quad (22)$$

где $Q(t)$ — некоторый полином, $\xi = \xi(\varphi)$ — обобщенный процесс со спектральным разложением (17), $\eta = \eta(\varphi)$ — искомый процесс. Найдем слабо стационарное решение уравнения (20). Предположим, что уравнение (20) имеет решение со спектральным разложением

$$\eta(\varphi) = \int_{-\infty}^{\infty} \tilde{\varphi}(u) \xi_1(du).$$

Тогда

$$\int_{-\infty}^{\infty} \tilde{\varphi}(u) Q(iu) \xi_1(du) = \int_{-\infty}^{\infty} \tilde{\varphi}(u) \xi(du),$$

$$\forall \tilde{\varphi}(u) \in S,$$

где $\xi(du)$ — спектральная стохастическая мера процесса ξ . Это возможно тогда и только тогда, когда

$$Q(iu) \xi_1(du) = \xi(du).$$

Предположим, что полином $Q(t)$ не имеет нулей на действительной прямой. Поэтому последнее равенство будет выполнено, если

$$\xi_1(du) = \frac{1}{Q(iu)} \xi(du), \quad \xi_1(b) - \xi(a) = \int_a^b \frac{1}{Q(iu)} \xi(du).$$

Итак, уравнение (20) имеет решение

$$\eta(\varphi) = \int_{-\infty}^{\infty} \tilde{\varphi}(u) \frac{1}{Q(iu)} \xi(du). \quad (23)$$

Теорема 2. Если многочлен $Q(t)$ не имеет нулей на действительной прямой, процесс $\xi(u)$ имеет спектральное представление (19), то уравнение (22) имеет единственное слабо стационарное решение (21) с ковариационным функционалом

$$B_{\eta}(\varphi, \varphi) = \int_{-\infty}^{\infty} |\tilde{\varphi}(u)|^2 \frac{F(du)}{|Q(iu)|^2}. \quad (24)$$

Если

$$\int_{-\infty}^{\infty} \frac{F(du)}{|Q(iu)|^2} < \infty,$$

то $\eta(\varphi)$ является процессом обычного типа.

Важный класс процессов образуют слабо стационарные решения уравнения (22), в которых процесс ξ имеет вид $\xi = P\left(\frac{d}{dt}\right)\gamma$, где $P(t)$ — многочлен степени m , γ — слабый белый шум. Уравнение (22) тогда принимает вид

$$Q\left(\frac{d}{dt}\right)\eta = P\left(\frac{d}{dt}\right)\gamma. \quad (25)$$

Из формул (23), (24) вытекает следующее представление слабо стационарного решения этого уравнения:

$$\eta(\varphi) = \int_{-\infty}^{\infty} \tilde{\varphi} \frac{P(iu)}{Q(iu)} \xi(du), \quad M|\xi(du)|^2 = du \quad (26)$$

и соответствующего ковариационного функционала:

$$B_{\eta}(\varphi, \varphi) = \int_{-\infty}^{\infty} |\tilde{\varphi}(u)|^2 \left| \frac{P(iu)}{Q(iu)} \right|^2 du. \quad (27)$$

Если степень полинома $P(u)$ меньше степени полинома Q , ($m < n$), то $\int_{-\infty}^{\infty} \left| \frac{P(iu)}{Q(iu)} \right|^2 du < \infty$ и процесс $\eta(\varphi)$ является процессом

обычного типа, порожденным непрерывным среднеквадратическим процессом

$$\eta(t) = \int_{-\infty}^{\infty} e^{itu} \frac{P(iu)}{Q(iu)} \zeta(du) \quad (28)$$

с ковариацией

$$B_{\eta}(t) = \int_{-\infty}^{\infty} e^{itu} \left| \frac{P(iu)}{Q(iu)} \right|^2 du. \quad (29)$$

Эти процессы называются процессами с дробно-рациональной спектральной плотностью.

Глава IX

ОЦЕНИВАНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ

§ 1. Выборочный метод в статистике

Задача математической статистики — получить определенные выводы из экспериментальных данных. При этом предполагается, что к проводимым экспериментам применимы теоретико-вероятностные концепции. Более точно задачи математической статистики можно формулировать с разной степенью общности и с различных точек зрения. Мы ограничимся лишь рассмотрением некоторых наиболее важных и простых задач, не стремясь к большей общности.

Одна из первых задач математической статистики — это задача оценивания параметров распределений. В общих чертах она состоит в следующем.

Предположим, что нас интересует неизвестное значение некоторой величины θ , которая в данных условиях может принимать произвольные значения из некоторого множества возможных значений Θ . Обычно Θ — некоторое множество в конечномерном пространстве R^s . Однако в некоторых важных задачах Θ может быть и бесконечномерным множеством, например, Θ может быть пространством всех непрерывных функций распределения.

Для того чтобы в данной ситуации определить значения величины θ , проводится эксперимент, результат которого описывается некоторым случайным вектором ζ , распределение которого F зависит от интересующей нас величины θ и возможно еще и от других, также неизвестных параметров.

Пространство R^n , в котором принимает значения вектор ζ , в математической статистике называют выборочным прост-

ранством, а сам вектор ξ — выборкой. Часто вектор ξ образован последовательностью n независимых одинаково распределенных случайных величин (или векторов) $\xi_1, \xi_2, \dots, \xi_n$, $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, измеряемых в n независимых друг от друга экспериментах, с распределением G . В этом случае говорят, что $\xi = (\xi_1, \dots, \xi_n)$ есть выборка объема n из генеральной совокупности с распределением G . Использование этой терминологии — дань определенной исторической традиции.

Задача оценивания параметра θ состоит в построении приближенных формул

$$\theta \approx h(\xi), \quad (1)$$

где $h(x)$ — функция, определенная на выборочном пространстве R^n со значениями во множестве Θ . При этом функция $h(x)$ должна быть одной и той же для всех распределений из данного семейства $\mathfrak{G} \{F \in \mathfrak{G}\}$.

Функцию $h(\xi)$ называют статистикой (на выборочном пространстве R^n), а значение $h(\xi)$ в приближенном равенстве $\theta \approx h(\xi)$ — оценкой параметра θ .

Сразу возникают следующие вопросы: как оценивать точность приближенной формулы (1) и какие формулы следует считать «хорошими»? Как строить «хорошие» оценки параметров в конкретных ситуациях?

По-видимому нет исчерпывающего и общего ответа на эти вопросы, но существует ряд подходов, позволяющих во многих случаях выяснить суть дела.

Для того чтобы как-то оценить погрешность приближенного равенства (1), представим себе, что задана некоторая функция $L(\alpha, \theta)$, $\alpha \in \Theta$, $\theta \in \Theta$, называемая функцией потерь, характеризующая величину ошибки, если в качестве истинного значения параметра θ применять значения α . Если $\hat{\theta}$ — оценка параметра θ , то величина

$$\bar{L}(\hat{\theta}, \theta) = M L(\hat{\theta}, \theta)$$

характеризует средний убыток от применения приближенной формулы (1) в большом числе независимых между собой случаев.

Если для двух оценок $\hat{\theta}_1$ и $\hat{\theta}_2$ параметра θ

$$\bar{L}(\hat{\theta}_1, \theta) < \bar{L}(\hat{\theta}_2, \theta),$$

то естественно считать оценку $\hat{\theta}_1$ лучше оценки $\hat{\theta}_2$ (относительно заданной функции потерь $L(\alpha, \theta)$). Положим

$$\bar{L}(\theta) = \inf_{\hat{\theta}} \bar{L}(\hat{\theta}, \theta),$$

где нижняя грань берется по некоторому классу Φ оценок $\hat{\theta}$, представляющих интерес в данной задаче и для которых выраже-

ние $\bar{L}(\check{\theta}, \theta)$ имеет смысл. Если существует оценка из этого класса, для которой

$$\begin{aligned} \bar{L}(\theta^*, \theta) &= \bar{L}(\theta), \\ \bar{L}(\theta^*, \theta) &\leq \bar{L}(\check{\theta}, \theta) \quad \forall \check{\theta} \in \Phi \end{aligned} \quad \text{или}$$

для всех $\theta \in \Theta$, то θ^* является оптимальной оценкой параметра θ в классе Φ , минимизирующей функцию потерь $L(\alpha, \theta)$.

Изложенный подход к сравнению качества двух оценок и определению оптимальной оценки параметра используется довольно часто. Тем не менее, его применение затруднено следующими обстоятельствами. Во-первых, не всегда есть достаточно убедительные основания для точного выбора функции потерь $L(\alpha, \theta)$. Во-вторых, даже если функция $\bar{L}(\check{\theta}, \theta)$ достигает минимума при каждом значении θ , статистика $\check{\theta}$, на которой он достигается, вообще говоря, зависит от θ , в то время как по самому смыслу задачи оцениваются, ищутся оценки, общие для всех значений $\theta \in \Theta$.

Первое из указанных обстоятельств используют, выбирая в качестве функции потерь $L(\alpha, \theta)$ функцию, наиболее удобную для применения математического аппарата, связанного с задачей минимизации, и обладающую интуитивно убедительными свойствами функции потерь. В качестве такой функции принимают квадрат расстояния между векторами α и θ в R^s

$$L(\alpha, \theta) = |\alpha - \theta|^2.$$

Величину

$$\delta^2 = M |\check{\theta} - \theta|^2$$

называют средней квадратической погрешностью оценки $\check{\theta}$.

В настоящей книге в качестве меры точности приближенной формулы (1) используется только среднеквадратическая погрешность.

Если $M \check{\theta} = \theta$, то среднеквадратическая погрешность превращается в сумму дисперсий компонент оценки $\check{\theta}$, а если θ — одномерный вектор, то δ^2 является дисперсией оценки. Требование, чтобы $M \check{\theta} = \theta$, часто называют условием отсутствия «систематической ошибки» и его наглядное содержание состоит в том, что при многократном использовании формулы (1) среднее значение ошибки $\check{\theta} \approx \theta$ равно нулю.

Определение. Оценка $\check{\theta}$ параметра θ называется несмещенной, если $M \check{\theta} = \theta$.

Возвращаясь к средней квадратической погрешности оценки, заметим, что она не дает полной информации об оценке параметра.

Полностью она содержится в распределении вектора θ . К сожалению, это распределение во многих случаях бывает трудно найти и, кроме того, оно зависит от параметра θ . В ряде случаев и для определенных классов распределений удается строить статистики, распределение которых не зависит от неизвестных параметров. В этом случае полезными являются интервальные оценки параметра θ : находятся интервалы $(\check{\theta}', \check{\theta}'')$, где $\check{\theta}' = h'(\xi)$, $\check{\theta}'' = h''(\xi)$ — некоторые статистики, такие, что вероятность события $h'(\xi) < \theta < h''(\xi)$ является достаточно большой.

Более просто и полно о качестве оценок можно судить в случае выборок большого объема. При этом рассматриваются последовательности оценок $\check{\theta}_n = h_n(\xi_1, \dots, \xi_n)$, $n = 1, 2, \dots$, и их асимптотическое поведение при $n \rightarrow \infty$. Одним из важнейших свойств последовательности оценок параметра является состоятельность.

Определение. Последовательность оценок $\check{\theta}_n = h_n(\xi_1, \dots, \xi_n)$, $n = 1, 2, \dots$ параметра θ , где (ξ_1, \dots, ξ_n) — независимые одинаково распределенные случайные величины, называется состоятельной, если $\text{plim } \check{\theta}_n = \theta$. Она называется сильно состоятельной, если $\check{\theta}_n \rightarrow \theta$ с вероятностью 1.

Другие асимптотические свойства последовательностей оценок будут определены и рассмотрены в дальнейшем.

Прежде чем переходить к общим методам теории оценивания параметров, рассмотрим некоторые простейшие задачи математической статистики. К числу их, в первую очередь, относится задача оценивания вероятности некоторого события и оценивание неизвестной функции распределения случайной величины.

Оценивание вероятности события. Простейшим примером оценивания параметра распределения является оценивание вероятности некоторого события A . Пусть I_A — индикатор A . Случайная величина I_A принимает два значения 1 и 0, первое с вероятностью p , второе с вероятностью $1 - p$. Параметр p принимает значения в интервале $(0, 1)$ и его значение считаем неизвестным. Для определения параметра p производятся n независимых испытаний, в каждом из которых событие A происходит с вероятностью p . Представляется естественным, учитывая частотную интерпретацию вероятности, принять в качестве приближенного значения \check{p} вероятности p частоту наступления события A в проведенных экспериментах,

$$p \approx \check{p} = \frac{v}{n}, \quad (2)$$

где v — число экспериментов, в которых событие A наступило. Имеем

$$M\check{p} = p, \quad M|p - \check{p}|^2 = D\check{p} = \frac{pq}{n}.$$

Таким образом, оценка \hat{p} вероятности p является несмещенной. Далее, при $n \rightarrow \infty$ \hat{p} сходится в среднеквадратическом к p ,

$$\text{l.i.m. } \hat{p} = p.$$

Более того, по теореме Бореля $\hat{p} \rightarrow p$ с вероятностью 1. Следовательно, оценки \hat{p} сильно состоятельны.

Асимптотическое распределение ошибок приближенного равенства (2) при больших n дает теорема Муавра — Лапласа. Из этой теоремы следует, что

$$P\left\{|\hat{p} - p| \geq c \sqrt{\frac{p(1-p)}{n}}\right\} \approx 1 - \Phi(c), \quad (3)$$

где

$$\Phi(c) = \frac{2}{\sqrt{2\pi}} \int_0^c e^{-\frac{u^2}{2}} du.$$

Если вероятность $\alpha = 1 - \Phi$ мала, то практически всегда или точнее с большой вероятностью, приблизительно равной $1 - \alpha$, выполняется неравенство

$$|\hat{p} - p| \leq c \sqrt{\frac{p(1-p)}{n}} \quad (4)$$

или

$$\hat{p} - c \sqrt{\frac{p(1-p)}{n}} < p < \hat{p} + c \sqrt{\frac{p(1-p)}{n}}.$$

Приведем несколько значений величины $c = c_\alpha$, соответствующей заданному α

$\alpha = 1 - \Phi(c)$	0,3173	0,05	0,0455	0,01	0,0027	0,001
c	1	1,96	2	2,58	3	3,29

Полученная оценка точности приближенного равенства (2) обладает двумя недостатками:

а) она справедлива при больших n , что является понятием расплывчатым;

б) в правой части неравенства (4) стоит величина, зависящая от p , т. е. от неизвестной величины.

При больших значениях n второе из указанных затруднений может быть устранено с помощью замены p на \hat{p} . Получим

$$\hat{p} - c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

При этом совершается ошибка более высокого порядка малости, чем ошибка, содержащаяся в замене точного распределения величины \check{p} гауссовским. С другой стороны, можно решить неравенство (4) относительно p . Имеем

$$\frac{\check{p}n + \frac{1}{2}c^2 - c \sqrt{\check{p}(1-\check{p})n + \frac{1}{4}c^2}}{n + c^2} \leq p \leq \frac{\check{p}n + \frac{1}{2}c^2 + c \sqrt{\check{p}(1-\check{p})n + \frac{1}{4}c^2}}{n + c^2}.$$

Оценивание дискретного распределения. Пусть $\xi_1, \xi_2, \dots, \xi_n$ — независимые одинаково распределенные случайные величины, принимающие конечное число значений x_1, x_2, \dots, x_s с вероятностями p_1, p_2, \dots, p_s . Для оценки величин p_k принимаем соображения предыдущего пункта $p_k = \frac{v_k}{n} = \check{p}_k$, где v_k — число тех величин ξ_k , которые принимают значение x_k , $v_1 + v_2 + \dots + v_s = n$. Точность оценки p_k при фиксированном k можно определить, как и в случае $s=2$. Однако для определения суммарной точности всех оценок $p_k \approx \check{p}_k$ нужны новые соглашения. Во многих статистических исследованиях в качестве суммарной меры точности приближенных равенств $p_k \approx \frac{v_k}{n}$ принимают величину

$$\zeta_n = \sum_{k=1}^s \frac{n}{p_k} \left(\frac{v_k}{n} - p_k \right)^2.$$

Можно доказать, что при больших n величина ζ_n имеет распределение, близкое к χ^2 -распределению с $s-1$ степенью свободы. Это утверждение можно интерпретировать следующим образом.

Пусть задано некоторое $\gamma \in (0,1)$, близкое к 1. Найдем (например, с помощью таблиц) такое t_1 , что $P\{\chi_{s-1}^2 < t_1\} = \gamma$, где χ_{s-1}^2 — случайная величина, имеющая χ^2 -распределение с $s-1$ степенью свободы. В R^s рассмотрим эллипс V с центром в случайной точке $(\check{p}_1, \check{p}_2, \dots, \check{p}_s)$ и полуосями $\sqrt{\frac{p_k}{n}} t_1 \approx \sqrt{\frac{\check{p}_k}{n}} t_1$. Тогда при большом n точка (p_1, p_2, \dots, p_n) с вероятностью γ лежит внутри сечения эллипса V гиперплоскостью $p_1 + p_2 + \dots + p_s = 1$.

Оценивание непрерывной функции распределения. Предположим, что измеряются n независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_n$ с неизвестной непрерывной функцией распределения $F(x)$. Требуется найти приближенное выражение для $F(x)$.

Эту задачу можно рассматривать как частный случай задачи оценивания параметра, если под параметром θ понимать саму

функцию распределения. Множеством θ в этом случае является пространство всех непрерывных функций распределения.

Так как $F(x) = P(\xi_k < x)$, то в соответствии с приведенным выше способом оценки вероятности события естественно в качестве оценки $F(x)$ принять

$$F(x) \approx \frac{v_n(x)}{n} = \tilde{F}_n(x),$$

где v_n — число тех k , для которых $\xi_k < x$, $k = 1, 2, \dots, n$. Функцию $\tilde{F}_n(x)$ называют эмпирической функцией распределения. График функции $\tilde{F}_n(x)$ можно описать следующим образом. Расположим величины ξ_1, \dots, ξ_n в порядке возрастания

$$\xi_1^*, \xi_2^*, \dots, \xi_n^*. \quad (5)$$

При этом ξ_1^* — наименьшее из значений $\xi_1, \xi_2, \dots, \xi_n$; ξ_2^* — второе по величине и т. д.; ξ_n^* — наибольшее из этих значений. Заметим, что случае совпадения какой-либо пары ξ_k и ξ_l можно пренебречь, так как

$$P(\xi_k = \xi_l) = \int \int_{x_1=x_2} F(dx_1) F(dx_2) = \int (F(x_1) - F(x_1-)) dF(x_1) = 0,$$

поскольку $F(x) = F(x-)$.

Последовательность (5) называется вариационным рядом последовательности случайных величин ξ_1, \dots, ξ_n и ее члены называют порядковыми статистиками выборки.

Очевидно, $\tilde{F}_n(x) = 0$ при $x \leq \xi_1^*$, $\tilde{F}_n(x) = \frac{1}{n}$ при $x \in (\xi_1^*, \xi_2^*]$ и т. д.

Проходя значение аргумента ξ_k^* , функция $\tilde{F}_n(x)$ терпит скачок величины $\frac{1}{n}$, а между значениями ξ_k^* и ξ_{k+1}^* остается постоянной. При $x > \xi_n^*$ имеем $\tilde{F}_n(x) = 1$.

Оценка функции распределения с помощью эмпирической функции распределения не является вполне удачной, хотя бы потому, что $F(x)$ — непрерывная функция, а $\tilde{F}_n(x)$ — разрывная. Однако при больших n скачки эмпирической функции распределения становятся малыми, а по теореме Бореля следует, что $\tilde{F}_n(x) \rightarrow F(x)$ при $n \rightarrow \infty$ с вероятностью 1 для всех $x \in (-\infty, \infty)$.

Для оценки точности равенства $F(x) \approx \tilde{F}_n(x)$ при больших значениях n можно пользоваться одной замечательной предельной теоремой А. Н. Колмогорова. Приведем ее без доказательства.

Теорема Колмогорова. Если $F(x)$ — непрерывная функция распределения, то

$$\lim_{n \rightarrow \infty} P(\sqrt{n} \sup |\tilde{F}_n(x) - F(x)| < a) = K(a),$$

$$K(a) = 1 - 2 \sum_{n=1}^{\infty} (-1)^{n-1} e^{-2n^2 a^2}.$$

Грубо говоря, функция $\tilde{F}_n(x)$ дает равномерную оценку функции $F(x)$ с точностью до величины порядка $\frac{1}{\sqrt{n}}$.

Вариационный ряд. Вариационный ряд последовательности независимых, одинаково распределенных случайных величин с непрерывной функцией распределения имеет важные применения в математической статистике.

Рассмотрим некоторые свойства этого ряда. Ниже мы предполагаем, что распределение $F(x)$ непрерывно. Найдем распределение членов вариационного ряда. Распределения минимального и максимального членов вариационного ряда очевидны:

$$P\{\xi_n^* < x\} = [F(x)]^n,$$

$$P\{\xi_1^* < x\} = 1 - P\{\xi_1^* \geq x\} = 1 - (1 - F(x))^n.$$

Чтобы получить распределение k -й порядковой статистики ξ_k^* , $k = 2, \dots, n-1$, предположим, что отрезок $(-\infty, x]$ разбит на частичные $(-\infty, s_1], (s_1, s_2], \dots, [s_{n-1}, s_n]$ и пусть $\Delta F(s_k) = F(s_k) - F(s_{k-1})$, $s_0 = -\infty$. Тогда $P\{\xi_k^* \in [s_{j-1}, s_j]\}$ равно, с точностью до малых величин высшего порядка малости по сравнению с $\Delta F(s_j)$, вероятности того, что в интервал $(-\infty, s_{j-1})$ попадет $k-1$ выборочных значений, в интервал $[s_{j-1}, s_j]$ — одно, в интервал $[s_j, \infty)$ — $n-k$ выборочных значений. Эта вероятность равна $C_n^{k-1} (n-k+1) [F(s_{j-1})]^{k-1} \Delta F(s_j) [1-F(s_j)]^{n-k}$, где $C_n^{k-1} (n-k+1)$ — число способов, с помощью которых n точек могут быть размещены по трем интервалам так, чтобы в первый попало $k-1$ точка, во второй — одна, а в третий — остальные $n-k$ точек. Суммируя эти вероятности по j и переходя к пределу при измельчении разбиения отрезка $(-\infty, x]$, при котором $\max_i \Delta F(s_j) \rightarrow 0$, получим

$$P\{\xi_k^* < x\} = C_n^{k-1} (n-k+1) \int_{-\infty}^x (F(s))^{k-1} (1-F(s))^{n-k} dF(s),$$

или, после замены переменных $F(s) = t$,

$$P\{\xi_k^* < x\} = C_n^{k-1} (n-k+1) \int_0^{F(x)} t^{k-1} (1-t)^{n-k} dt. \quad (6)$$

Теорема 1. Пусть $\xi, \xi_1, \xi_2, \dots, \xi_n$ — независимые одинаково распределенные случайные величины с непрерывной функцией

распределения, $\xi_1^*, \xi_2^*, \dots, \xi_n^*$ — вариационный ряд, построенный по последовательности $\xi_1, \xi_2, \dots, \xi_n$. Тогда

$$P(\xi_{k-1}^* < \xi < \xi_k^*) = \frac{1}{n+1}, \quad k = 1, 2, \dots, n, \quad (7)$$

$$\xi_0^* = -\infty, \quad \xi_{n+1}^* = +\infty.$$

Доказательство. Пусть $G_k(x)$ — функция распределения величины ξ_k^* . Так как ξ не зависит от величин ξ_1^*, \dots, ξ_n^* , то

$$P(\xi < \xi_k^*) = \iint_{y < x} F(dy) G_k(dx) = \int_{-\infty}^{\infty} F(x) dG_k(x) = \int_{-\infty}^{\infty} C_n^{k-1} (n-k+1) F^k(x) (1-F(x))^{n-k} dF(x) = C_n^{k-1} (n-k+1) \int_0^1 t^k (1-t)^{n-k} dt.$$

Интегрируя по частям k раз, получим

$$P(\xi < \xi_k^*) = C_n^{k-1} (n-k+1) \frac{k!}{(n-k+1) \dots n} \int_0^1 (1-t)^n dt =$$

$$= k \int_0^1 (1-t)^n dt = \frac{k}{n+1}.$$

Итак, $P(\xi < \xi_k^*) = \frac{k}{n+1},$

$$P(\xi_{k-1}^* \leq \xi \leq \xi_k^*) = \frac{1}{n+1}.$$

Замечание. Важным обстоятельством, вытекающим из доказанной теоремы, является то, что найденные вероятности не зависят от функции распределения величин ξ_k .

Пусть $p \in (0, 1)$. Решение x_p уравнения $F(x) = p$ называют p -квантилью распределения $F(x)$. Так как функция $F(x)$ непрерывна, то уравнение $F(x) = p$ имеет решение, однако, возможно, что решения этого уравнения образуют целый отрезок $[\alpha, \beta]$. В этом случае любая точка отрезка $[\alpha, \beta]$, $\alpha < \beta$ является p -квантилью. При $p = \frac{1}{2}$ квантиль называют медианой распределения. Если ξ — случайная величина с функцией распределения $F(x)$, x_1 — медиана распределения, то

$$P(\xi < x_1) = P(\xi \geq x_1) = \frac{1}{2}.$$

Таким образом, медиана характеризует «центральное значение» случайной величины в том смысле, что вероятность принять значение, меньше медианы и вероятность принять значение больше

медианы, равны между собой. Значение удачно выбранных квантилей (например, $p = \frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{2}, \frac{5}{8}, \frac{3}{4}, \frac{7}{8}$) дает наглядное представление о распределении случайной величины.

Порядковые статистики являются хорошими оценками p -квантилей распределения, а именно, для оценки ξ_p p -квантили x_p рассмотрим порядковую статистику $\xi_{k(p)}^*$, где $k(p) = np$, если np — целое число и $k(p) = [np] + 1$, в противном случае. Итак, полагаем $x_p \approx \xi_p = \xi_{k(p)}^*$.

Величину ξ_p называют выборочной p -квантилью. В частности, $\xi_{\frac{1}{2}}$ называют эмпирической медианой.

Теорема 2. Пусть функция распределения $F(x)$ непрерывна и уравнение $F(x) = p$ имеет единственное решение x_p . Тогда порядковая статистика $\xi_p = \xi_{k(p)}^*$ является строго состоятельной оценкой p -квантили x_p .

Если, кроме того, функция $F(x)$ дифференцируема в точке x_p и $F'(x_p) = f(x_p) > 0$, то величина

$$\sqrt{n} (\xi_p - x_p)$$

асимптотически нормальна с параметром $\left(0, \frac{pq}{f^2(x_p)}\right)$, где $q = 1 - p$.

Доказательство. Пусть $I_j(x)$ — индикатор события $\{\xi_j < x\}$, $j = 1, \dots, n$. События $\{\xi_{k(p)}^* < x\}$ и $\left\{\sum_{j=1}^n I_j(x) \geq k(p)\right\}$ происходят одновременно. Случайные величины $I_j(x)$ взаимно независимы и одинаково распределены. При этом $MI_j(x) = F(x)$. На основании усиленного закона больших чисел при $n \rightarrow \infty$ с вероятностью 1

$$\frac{1}{n} \sum_{j=1}^n I_j(x) \rightarrow F(x).$$

Поэтому при $F(x) > p$ событие $\left\{\sum_{j=1}^n I_j(x) \geq k(p)\right\} = \left\{\frac{1}{n} \sum_{j=1}^n I_j(x) \geq \frac{k(p)}{n}\right\}$ происходит при достаточно большом n с вероятностью 1,

а при $F(x) < p$ — с вероятностью 1 $\left\{\sum_{j=1}^n I_j(x) < k(p)\right\}$, начиная с некоторого достаточно большого $n_0 = n_0(\omega)$. Таким образом, $\xi_{k(p)}^* > x$ при достаточно большом n с вероятностью 1 для любого x такого, что $F(x) < p$ и $\xi_{k(p)}^* < x$, если $F(x) > p$. Иными словами, $\lim \xi_{k(p)}^* = x_p$ с вероятностью 1. Строгая состоятельность оценки $x_p \approx \xi_p$ доказана.

Предположим теперь, что $F'(x_p) = f(x_p) > 0$. Положим

$$\eta = f(x_p) \sqrt{\frac{pq}{n}} (\xi_p - x_p).$$

Тогда

$$\begin{aligned} P(\eta < x) &= P\left(\xi_p < x_p + \sqrt{\frac{pq}{n}} \frac{x}{f(x_p)}\right) = \\ &= P\left\{\frac{1}{n} \sum_{i=1}^n I_i\left(x_p + \sqrt{\frac{pq}{n}} \frac{x}{f(x_p)}\right) \geq \frac{k(p)}{n}\right\}. \end{aligned}$$

Величины $\xi_{ni} = I_i\left(x_p + \sqrt{\frac{pq}{n}} \frac{x}{f(x_p)}\right)$ при каждом n одинаково распределены с математическим ожиданием $a_n = F\left(x_p + \sqrt{\frac{pq}{n}} \frac{x}{f(x_p)}\right)$ и дисперсией $\sigma_n^2 = a_n(1 - a_n)$ и равномерно ограниченными моментами третьего порядка. Поэтому к ним применима теорема Ляпунова, по которой величины $\frac{1}{\sigma_n \sqrt{n}} \sum_{i=1}^n (\xi_{ni} - a_n)$ асимптотически нормальны $(0, 1)$. Так как

$$\begin{aligned} P(\eta < x) &= P\left\{\frac{1}{\sigma_n \sqrt{n}} \sum_{i=1}^n (\xi_{ni} - a_n) \geq \frac{\sqrt{n}}{\sigma_n} \left(\frac{k(p)}{n} - \right. \right. \\ &\quad \left. \left. - F\left(x_p + \sqrt{\frac{pq}{n}} \frac{x}{f(x_p)}\right)\right)\right\} \end{aligned}$$

и

$$\begin{aligned} \frac{\sqrt{n}}{\sigma_n} \left[\frac{k(p)}{n} - F\left(x_p + \sqrt{\frac{pq}{n}} \frac{x}{f(x_p)}\right) \right] &= \sqrt{\frac{n}{a_n}} \left(p + o\left(\frac{1}{n}\right) - \right. \\ &\quad \left. - p - \sqrt{\frac{pq}{n}} x + o\left(\frac{1}{\sqrt{n}}\right) \right) = -\frac{\sqrt{pq}}{\sigma_n} x + o(1) \rightarrow -x, \end{aligned}$$

то

$$P(\eta < x) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-x}^{\infty} e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du.$$

Оценивание параметров распределения с помощью эмпирического распределения. Эмпирическая функция распределения может быть основой для получения интуитивно наглядных оценок произвольных параметров распределения. Получаемые при этом оценки часто оказываются сильно состоятельными.

Предположим, что нас интересует значение θ некоторого параметра неизвестного распределения $F(x)$. Будем считать, что

этот параметр однозначно определяется функцией распределения, т. е.

$$\theta = \Phi(F(\cdot)),$$

где Φ — некоторый функционал, определенный на множестве всех функций распределения. Разумеется, легко строить примеры, в которых параметр распределения не определяется функцией распределения, но этими случаями можно пренебречь, так как такие параметры, по существу, не связаны с рассматриваемой статистической задачей. В дальнейшем, говоря о параметре распределения, мы будем считать, что он однозначно определяется распределением.

Предположим, что рассматривается выборка $(\xi_1, \xi_2, \dots, \xi_n)$, где $\xi_1, \xi_2, \dots, \xi_n$ — независимые одинаково распределенные случайные величины с непрерывной функцией распределения $F(x)$ и $\theta = \Phi(F(\cdot))$. Пусть $\check{F}_n(x)$ — эмпирическая функция распределения, построенная по этой выборке. Если функционал $\Phi(F(\cdot))$ непрерывен относительно поточечной сходимости: $\Phi(\check{F}_n(\cdot)) \rightarrow \Phi(F(\cdot))$, когда $\check{F}_n(x) \rightarrow F(x)$ при каждом x , $x \in (-\infty, \infty)$, то $\lim \Phi(\check{F}_n(\cdot)) = \Phi(F(\cdot)) = \theta$ с вероятностью 1. Таким образом, представляется естественным в качестве оценки параметра θ предложить случайную величину

$$\check{\theta}_n \approx \Phi(\check{F}_n(\cdot)). \quad (8)$$

Оценки, полученные по формуле (8), называют выборочными, или эмпирическими значениями параметра θ .

ПРИМЕРЫ

1. Пусть ξ_1, \dots, ξ_n — последовательность независимых одинаково распределенных случайных величин с функцией распределения $F(x)$. Положим

$$G = Mg(\xi_k) = \int g(x) F(dx) \neq \infty.$$

Нас интересует оценка величины G . Выборочное значение величины G имеет вид

$$\check{G}_n = \int g(x) \check{F}_n(dx) = \frac{1}{n} \sum_{k=1}^n g(\xi_k).$$

Эта оценка несмещена

$$M\check{G}_n = \frac{1}{n} \sum_{k=1}^n Mg(\xi_k) = G.$$

Далее по усиленному закону больших чисел она сильно состоятельна

$$\frac{1}{n} \sum_{k=1}^n g(\xi_k) \rightarrow G$$

с вероятностью 1. Если еще $D_g(\xi_k) = \sigma_g^2 < \infty$, то на центральной предельной теореме для одинаково распределенных случайных слагаемых следует, что величина

$$\zeta_n = \frac{\sqrt{n}}{\sigma_g} (\check{G}_n - G) = \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{g(\xi_k) - G}{\sigma_g}$$

асимптотически нормальна $(0,1)$. Отсюда вытекает, что при больших n неравенство

$$\check{G}_n - \alpha \frac{\sigma_g}{\sqrt{n}} < G < \check{G}_n + \alpha \frac{\sigma_g}{\sqrt{n}} \quad (9)$$

выполняется с вероятностью

$$p_\alpha = \frac{2}{\sqrt{2\pi}} \int_0^\alpha e^{-\frac{u^2}{2}} du. \quad (10)$$

Беря p_α достаточно близким к 1, чтобы игнорировать наступление событий, вероятность которых равна $1 - p_\alpha$, и определяя α из формулы (10), получаем случайный интервал $\left(\check{G}_n - \alpha \frac{\sigma_g}{\sqrt{n}}, \check{G}_n + \alpha \frac{\sigma_g}{\sqrt{n}}\right)$, внутри которого практически во всех случаях лежит оцениваемое значение параметра G . Точнее говоря, неравенства (9) в экспериментах не выполняются только с малой вероятностью, равной $1 - p_\alpha$. Построенная интервальная оценка параметра G все же неудобна, так как содержит величину σ_g , которая сама может быть неизвестной. Однако при больших n , согласно тех же соображений,

$$G_g^2 \approx \frac{1}{n} \sum_{i=1}^n (g(\xi_k) - \check{G}_n)^2 = \check{\sigma}_g^2.$$

Заменяя в формуле (9) величину σ_g на $\check{\sigma}_g$, мы совершим ошибку, вообще говоря, более высокого порядка малости, чем та, которая совершается при замене точного распределения величины нормальным. Поэтому предыдущие заключения остаются справедливыми и для неравенств

$$\check{G}_n - \alpha \frac{\check{\sigma}_g}{\sqrt{n}} < G < \check{G}_n + \alpha \frac{\check{\sigma}_g}{\sqrt{n}}.$$

2. Частными случаями рассмотренной выше задачи являются задачи оценки математического ожидания и моментов случайной величины. Если

$$a = M\xi_k,$$

то предыдущие формулы приводят к оценке

$$\check{a} = \frac{1}{n} \sum_{k=1}^n \xi_k.$$

Величину \check{a} называют выборочным (эмпирическим) средним. Если $M\xi_k \neq \infty$, то \check{a} — несмещенная сильно состоятельная оценка. Для моментов m_r случайных величин ξ_k

$$m_r = M\xi_k^r, \quad M\xi_k^r \neq \infty,$$

получаем несмещенные и сильно состоятельные оценки

$$\check{m}_r = \frac{1}{n} \sum_{k=1}^n \xi_k^r.$$

Величины \check{m}_r называют выборочными (эмпирическими) моментами. Оценкой дисперсии σ^2 случайных величин ξ_k является выборочная дисперсия

$$\check{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k - \check{a})^2 = \frac{1}{n} \sum_{k=1}^n \xi_k^2 - \check{a}^2.$$

Вычислим ее математическое ожидание. Поскольку $M(\xi_k - \check{a}) = 0$, то

$$\begin{aligned} M(\xi_k - \check{a})^2 &= D(\xi_k - \check{a}) = D\left[\left(1 - \frac{1}{n}\right)\xi_k - \frac{1}{n} \sum_{r \neq k} \xi_r\right] = \\ &= \left(\frac{n-1}{n}\right)^2 D\xi_k + \frac{1}{n^2} \sum_{k \neq r} D\xi_r = \frac{n-1}{n} \sigma^2. \end{aligned}$$

Итак,

$$M\check{\sigma}^2 = \left(1 - \frac{1}{n}\right) \sigma^2.$$

Полученная формула показывает, что $\check{\sigma}^2$ является смещенной оценкой дисперсии. Однако смещение этой оценки

$$M\check{\sigma}^2 - \sigma^2 = -\frac{\sigma^2}{n} \rightarrow 0 \text{ при } n \rightarrow \infty,$$

так что выборочная дисперсия асимптотически несмещена. Кроме того, легко несколько видоизменить величину $\check{\sigma}^2$ с тем, чтобы видоизмененная оценка уже оказалась несмещенной. Положим

$$s^2 = \frac{1}{n-1} \sum_{k=1}^n (\xi_k - \check{a})^2.$$

Поскольку $s^2 = \frac{n}{n-1} \check{\sigma}^2$, то s^2 уже несмещенная оценка параметра σ^2 . При этом она остается сильно состоятельной.

3. Рассмотрим распределение Коши с плотностью

$$f(x, \theta) = \frac{1}{\pi} \frac{1}{1 + (x + \theta)^2}, \quad x \in (-\infty, \infty).$$

Точка $x = \theta$ является центром симметрии для плотности. Таким образом, значение $x = \theta$ играет роль «среднего значения» (центра расположения) распределения, но математическое ожидание для него не существует. Кажется заманчивым в качестве оценки параметра θ предложить эмпирическое среднее

$$\check{\theta}_n = \frac{1}{n} \sum_{k=1}^n \xi_k.$$

Однако такая оценка неудовлетворительна. Покажем это.

Найдем распределение величины $\check{\theta}_n$. Так как характеристическая функция распределения Коши имеет вид $\varphi(u) = \exp(i\theta u - |u|)$ (см. § 9, гл. III), то для характеристической функции $\varphi_n(u)$ величины θ_n имеем

$$\varphi_n(u) = \left[\varphi\left(\frac{u}{n}\right) \right]^n = \exp(i\theta u - |u|) = \varphi(u).$$

Это означает, что ошибки приближенных равенств $\theta \approx \check{\theta}_n$ и $\theta \approx \xi$ имеют одинаковое распределение, и можно сказать, что статистика $\check{\theta}_n$ несет о параметре θ столько же информации, сколько и одно измерение. Говоря более точно, величина $\check{\theta}_n - \theta$ имеет плотность распределения $f(x, 0)$, не зависящую от n , и оценка $\check{\theta}_n$ несостоятельна.

С другой стороны, значение θ является медианой распределения $f(x, \theta)$:

$$\int_{-\infty}^{\theta} f(x, \theta) dx = \int_0^{\infty} \frac{1}{\pi} \frac{1}{1+x^2} dx = \frac{1}{2},$$

и по теореме 2 оценка $\theta \approx \xi_1$ является сильно состоятельной, причем, величина

$$\frac{2}{\pi} \sqrt{n} \left(\xi_1 - \frac{1}{2} \right)$$

асимптотически нормальна $(0,1)$.

Оценки минимальной дисперсии. Неравенство Крамера — Рао. Предложенный выше метод оценки параметров распределений, как указывалось ранее, часто приводит к состоятельным оценкам. Однако это еще не говорит о том, насколько эти оценки хороши при небольших значениях n и нет ли лучших оценок.

В настоящем пункте будет показано, что для семейства функций распределения $\{F(x, \theta), \theta \in \Theta\}$, достаточно регулярно зависящих от параметра θ , можно указать нижнюю границу дисперсии всех несмещенных оценок параметра. В некоторых случаях эта граница является точной нижней границей и существуют оценки параметра, на которых она достигается. Такие оценки являются оптимальными, если качество оценки характеризовать средней квадратической погрешностью, и их называют эффективными. Сравнение дисперсии данной несмещенной оценки с нижней границей дисперсии несмещенных оценок позволяет судить, насколько данная оценка близка к оптимально возможной.

Мы будем рассматривать выборки фиксированного объема n . Поэтому во многих случаях достаточно ограничиться оценками параметра по одному выборочному n -мерному случайному вектору. Структура выборки, т. е. тот факт, что выборочный вектор ξ образован последовательностью n независимых одинаково распределенных случайных величин, проявляется при рассмотрении асимптотических свойств оценок.

В дальнейшем при изложении теоретических вопросов будем предполагать, что выборочный вектор ξ обладает плотностью распределения $f(x, \theta)$, $x \in R^n$, $\theta \in \Theta$. Результаты, полученные в этом

случае, легко переносится и на дискретные распределения по очевидным изменениям формулировок и условий теорем, а именно, вместо плотности $f(x, \theta)$ в последнем случае следует говорить о вероятности события $\xi = x$, $f(x, \theta) = P_\theta(\xi = x)$. При этом $f(x, \theta) > 0$ только для конечного или счетного множества значений $x = x_i$. В приводимых примерах наряду с абсолютно непрерывными рассматриваются и дискретные распределения.

В последующем часто встречается случайная функция аргумента θ , $f(\xi, \theta)$, $\theta \in \Theta$. Иногда мы будем ее обозначать через Λ , или $\Lambda(\theta)$ и называть функцией правдоподобия. Если выборка образована n независимыми одинаково распределенными случайными величинами ξ_k , $k = 1, \dots, n$, с плотностью распределе-

ния $g(x, \theta)$, то $\Lambda(\theta) = \prod_{k=1}^n \Lambda_k(\theta)$, где $\Lambda_k(\theta)$ — независимые одинаково распределенные случайные величины (при фиксированном θ), $\Lambda_k(\theta) = g(\xi_k, \theta)$.

Перейдем к выводу формулы для нижней границы дисперсии несмещенных оценок некоторого параметра распределения.

Рассмотрим сначала случай одномерного параметра. Пусть Θ — действительная прямая, или интервал действительной прямой, $f(x, \theta)$ — плотность распределения вектора ξ , принимающего значения в R^n , $\theta \in \Theta$. Во избежание недоразумений через M_θ будем обозначать математическое ожидание функций от вектора в том случае, когда плотность распределения вектора ξ равна $f(x, \theta)$.

Лемма. Предположим, что:

а) существуют производные $f'_\theta(x, \theta)$ и $f''_{\theta\theta}(x, \theta)$, причем

$$M_\theta \left| \frac{\partial}{\partial \theta} \ln f(\xi, \theta) \right| < \infty,$$

$$M_\theta \left| \frac{\partial^2}{\partial \theta^2} \ln f(\xi, \theta) \right| < \infty.$$

$$б) M_\theta \left| \frac{\partial}{\partial \theta} \ln f(\xi, \theta) \right|^2 < \infty.$$

Тогда

$$M_\theta \frac{\partial}{\partial \theta} \ln f(\xi, \theta) = 0, \quad \forall \theta \in \Theta, \quad (11)$$

$$M_\theta \left[\frac{\partial}{\partial \theta} \ln f(\xi, \theta) \right]^2 = -M_\theta \frac{\partial^2}{\partial \theta^2} \ln f(\xi, \theta). \quad (12)$$

Доказательство. Продифференцируем по θ равенство

$$1 = \int_{R^n} f(x, \theta) dx.$$

По предположениям леммы можно дифференцировать под знаком интеграла. Получим

$$0 = \int_{R^n} f'(x, \theta) dx = \int_{R^n} \frac{f'(x, \theta)}{f(x, \theta)} f(x, \theta) dx = M_\theta \frac{\partial}{\partial \theta} \ln f(\xi, \theta).$$

Далее

$$M_3 \frac{\partial^2}{\partial \theta^2} \ln f(\zeta, \theta) = \int_{R^n} \frac{f''_{\theta\theta}(x, \theta) f(x, \theta) - [f'_\theta(x, \theta)]^2}{f^2(x, \theta)} f(x, \theta) dx = \\ = \int_{R^n} f''_{\theta\theta}(x, \theta) dx - M_3 \left[\frac{\partial}{\partial \theta} \ln f(\zeta, \theta) \right]^2 = -M_3 \left[\frac{\partial}{\partial \theta} \ln f(\zeta, \theta) \right]^2,$$

так как

$$\int_{R^n} f''_{\theta\theta}(x, \theta) dx = 0.$$

Функцию

$$I(\theta) = M_3 \left[\frac{\partial}{\partial \theta} \ln f(\zeta, \theta) \right]^2$$

называют количеством информации по Фишеру. В определенном смысле ее можно рассматривать как количество информации, содержащееся в векторе ζ , о значении параметра θ . Неравенство Крамера — Рао, которое сейчас будет доказано, является одним из источников такой точки зрения.

Теорема 3. (Неравенство Крамера — Рао). Пусть выполнены предположения а) и б) леммы 1 и $h(\theta)$ — дифференцируемая функция на Θ , для которой существует несмещенная оценка $\tilde{h}(\zeta)$ с конечной дисперсией, удовлетворяющей условию

$$\int_{R^n} \left| \tilde{h}(x) \frac{\partial}{\partial \theta} f(x, \theta) \right| dx < \infty, \quad \forall \theta \in \Theta.$$

Тогда

$$M |\tilde{h}(\zeta) - \theta|^2 \geq \frac{[h'(\theta)]^2}{I(\theta)}, \quad \forall \theta \in \Theta. \quad (13)$$

Знак равенства в (13) достигается тогда и только тогда, когда

$$\frac{\partial}{\partial \theta} \ln f(x, \theta) = c(\theta) [\tilde{h}(x) - h(\theta)], \quad (14)$$

или

$$f(x, \theta) = \exp \{ \psi_1(\theta) \tilde{h}(x) + \psi_2(\theta) + K(x) \}. \quad (15)$$

Доказательство. На основании несмещенности оценки

$$\int_{R^n} \tilde{h}(x) f(x, \theta) dx = h(\theta).$$

Продифференцируем это равенство по θ . Из условия теоремы следует, что правую часть равенства можно дифференцировать под знаком интеграла. Получим

$$h'(\theta) = \int_{R^n} \tilde{h}'(x) f'_\theta(x, \theta) dx = \int_{R^n} \tilde{h}(x) \frac{\partial}{\partial \theta} \ln f(x, \theta) f(x, \theta) dx = \\ = M \tilde{h}(\zeta) \frac{\partial}{\partial \theta} \ln f(\zeta, \theta).$$

Согласно лемме 1,

$$h'(\theta) = M(\check{h}(\xi) - \check{h}(\theta)) \frac{\partial}{\partial \theta} \ln f(\xi, \theta).$$

Воспользовавшись неравенством Коши — Буняковского, имеем

$$[h'(\theta)]^2 \leq I(\theta) M(\check{h}(\xi) - \check{h}(\theta))^2,$$

причем знак равенства здесь возможен тогда и только тогда, когда

$$\frac{\partial \ln f(\xi, \theta)}{\partial \theta} = c(\theta) [\check{h}(\xi) - \check{h}(\theta)] \text{ с вероятностью } 1.$$

Интегрируя по θ последнее равенство, получим для $f(x, \theta)$ выражение вида (15).

Неравенство (13) называют неравенством Крамера — Рао.

В применении к оценке самого параметра θ , $h(\theta) = \theta$, неравенство (13) принимает вид

$$M|\check{h}(\xi) - \theta|^2 \geq \frac{1}{I(\theta)}. \quad (16)$$

Если в неравенстве (13) достигается знак равенства, то оценка $\check{h}(\xi)$ является оптимально несмещенной оценкой в том смысле, что она обладает минимальной дисперсией среди всех несмещенных оценок параметра $h(\theta)$. Такие оценки называют еще эффективными. Заметим, что статистика $\check{h}(\xi)$ для плотностей, удовлетворяющих (14), является несмещенной. Это вытекает из равенства

$$0 = M \frac{\partial}{\partial \theta} \ln f(\xi, \theta) = c(\theta) [M\check{h}(\xi) - h(\theta)].$$

Хотя нижняя граница Крамера — Рао достигается для некоторых классов распределений, вообще говоря, она не является точной нижней границей дисперсий несмещенных оценок параметра.

Сделаем ряд замечаний и дополнений к теореме 3.

Замечания. 1. Если для некоторой функции $h(\theta)$ существует эффективная несмещенная оценка, то для другой функции $h_1(\theta)$, как это непосредственно вытекает из формулы (14), несмещенной эффективной оценки уже не существует.

Это замечание показывает, что не всегда разумно ограничиться только несмещенными оценками.

2. Используя неравенство Крамера — Рао, можно получить оценку нижней границы средней квадратической погрешности не обязательно несмещенных оценок. Пусть $\check{h}(\xi)$ — некоторая оценка параметра θ , $M\check{h}(\xi) = \theta + b(\theta)$, где $b(\theta)$ — смещение оценки $\check{h}(\theta)$.

$$\begin{aligned} \delta^2(\theta) &= M(\check{h}(\xi) - \theta)^2 = M(\check{h}(\xi) - M\check{h}(\xi) + b(\theta))^2 = \\ &= M(\check{h}(\xi) - M\check{h}(\xi))^2 + b^2(\theta). \end{aligned}$$

Предполагая, что функция $b(\theta)$ дифференцируема, и замечая, что $\check{h}(\zeta)$ является несмещенной оценкой параметра $M\check{h}(\zeta) = \theta + b(\theta)$, получим, используя (13),

$$D^2(\theta) = M(\check{h}(\zeta) - \theta)^2 \geq b^2(\theta) + \frac{(1 + b'(\theta))^2}{I(\theta)}. \quad (17)$$

3. Пусть ζ — выборочный вектор, $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$, где ξ_1, \dots, ξ_n — независимые одинаково распределенные случайные величины с плотностью распределения $g(x, \theta)$, и выполнены предположения а) и б) леммы 1. Тогда

$$I(\theta) = -M_\theta \frac{\partial^2}{\partial \theta^2} \ln f(\zeta, \theta) = -nM_\theta \frac{\partial^2}{\partial \theta^2} \ln g(\xi_1, \theta).$$

Формула (13) показывает, что нижняя граница Крамера — Рао при больших n имеет порядок $\frac{1}{n}$

$$M|\check{h}(\zeta) - \theta|^2 \geq \frac{h'(\theta)^2}{-nM_\theta \frac{\partial^2}{\partial \theta^2} \ln g(\xi_1, \theta)}. \quad (18)$$

Для сравнения качества двух несмещенных оценок можно пользоваться понятием относительной эффективности.

4. Относительной эффективностью оценок $\check{\theta}_2$ и $\check{\theta}_1$ параметра называют отношение

$$\text{eff}(\check{\theta}_2 | \check{\theta}_1) = \frac{M|\check{\theta}_1 - \theta|^2}{M|\check{\theta}_2 - \theta|^2}.$$

Пусть $\check{\theta}'_n$ и $\check{\theta}''_n$, $n = 1, 2, \dots$ — две последовательности несмещенных оценок параметра θ по выборке объема n . Относительной асимптотической эффективностью этих последовательностей оценок называют предел, если он существует, относительных эффективностей $\text{eff}(\check{\theta}''_n | \check{\theta}'_n)$.

Последовательность несмещенных оценок $\check{\theta}_n$, $n = 1, 2, \dots$, параметра θ , построенных по выборке объема n , называют асимптотически эффективной, если отношение нижней границы Крамера — Рао дисперсий несмещенных оценок к дисперсии оценок $\check{\theta}_n$ стремится к 1, т. е. если

$$nI(\theta) M|\check{\theta}_n - \theta|^2 \rightarrow 1 \quad \text{при } n \rightarrow \infty.$$

Предположим, что для некоторой функции $h(\theta)$ существует эффективная оценка $\check{h}(\zeta)$, но нас интересует оценка параметра $h_1(\theta) = v(h(\theta))$, где $v(t)$, $t \in (-\infty, \infty)$ — некоторая дифференцируемая функция. Естественно оценить параметр $h_1(\theta)$ с помощью статистики $\check{h}_1(\zeta) = v(\check{h}(\zeta))$. Хотя эта статистика неэффективна, все же она асимптотически несмещена и эффективна.

Действительно, при больших n

$$\begin{aligned} M|\check{h}_1(\zeta) - h_1(\theta)|^2 &= M|v(\check{h}(\zeta)) - v(h(\theta))|^2 \approx \\ &\approx M[v'(h(\theta))]^2 |\check{h}(\zeta) - h(\theta)|^2 = \frac{[v'(h(\theta)) h'(\theta)]^2}{I(\theta)} \end{aligned}$$

и

$$M(\check{h}_1(\zeta) - h_1(\theta)) \approx Mv'(h(\theta)) (\check{h}(\zeta) - h(\theta)) = 0,$$

т. е. $\check{h}_1(\zeta)$ — асимптотически несмещенная оценка величины $h_1(\theta)$ ($M\check{h}_1(\zeta) \rightarrow h_1(\theta)$) и ее средняя квадратическая погрешность при больших n эквивалентна нижней грани дисперсий несмещенных оценок величины $h_1(\theta)$.

ПРИМЕРЫ

1. Параметры нормального распределения

Рассмотрим задачу оценивания параметров нормального распределения с плотностью

$$g(x, a, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Оценивание производится по выборке объема n . Плотность распределения выборки равна:

$$f(x, a, \sigma) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - a)^2\right\}.$$

Предположим, что дисперсия σ^2 известна и оценивается среднее a . Имеем

$$\frac{\partial}{\partial a} \ln f(x, a, \sigma) = \frac{1}{\sigma^2} \sum_{k=1}^n (x_k - a).$$

Таким образом, выполнено условие (14) существования несмещенной эффективной оценки и она имеет вид

$$\tilde{h}(\xi_1, \xi_2, \dots, \xi_n) = \frac{1}{n} \sum_{k=1}^n \xi_k.$$

Следовательно, выборочное среднее $\bar{\xi} = \frac{\xi_1 + \dots + \xi_n}{n}$ является несмещенной оценкой минимальной дисперсии нормального распределения. Дисперсия этой оценки

$$D(\bar{\xi}) = \frac{\sigma^2}{n}.$$

Для всякой другой регулярной несмещенной оценки \check{a} параметра a нормального распределения $D(\check{a}) \geq D(\bar{\xi}) = \frac{\sigma^2}{n}$.

Предположим теперь, что среднее a известно и оценивается дисперсия $\sigma^2 = \theta$. Имеем

$$\frac{\partial}{\partial \theta} \ln f(x, a, \theta) = \sum_{k=1}^n \frac{(x_k - a)^2}{2\theta^2} - \frac{n}{2} \frac{1}{\theta^2}.$$

Условие (14) существования эффективной оценки выполнено. При этом $c(\theta) =$

$= \frac{n}{2\theta^4}$, $h(x) = \frac{1}{n} \sum_{k=1}^n (x_k - a)^2$. Таким образом, оценка

$$\sigma^2 \approx \frac{1}{n} \sum_{k=1}^n (\xi_k - a)^2 = \check{\sigma}^2$$

является эффективной. Ее дисперсия

$$D\check{\sigma}^2 = \frac{1}{n} D(\xi_k - a)^2 = \frac{1}{n} (M(\xi_k - a)^4 - \sigma^4) = \frac{2\sigma^4}{n}.$$

2. Оценка параметра биномиального распределения

Вектор ξ — одномерен, принимает конечное число значений $0, 1, \dots, n$ с вероятностями

$$f(x, \theta) = C_n^x \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \dots, n, \quad \theta \in (0, 1).$$

Имеем

$$\frac{\partial \ln f(\xi, \theta)}{\partial \theta} = \frac{\xi}{\theta} - (n - \xi) \frac{1}{1 - \theta} = \frac{n}{\theta(1 - \theta)} \left(\frac{\xi}{n} - \theta \right).$$

Это равенство показывает, что оценка $\check{\theta} = \frac{\xi}{n}$ параметра θ — вероятности успеха в серии n независимых экспериментов с постоянной вероятностью успеха, где ξ — число успехов в серии, является несмещенной эффективной оценкой. Ее дисперсия

$$D\check{\theta} = M \left(\frac{\xi}{n} - \theta \right)^2 = \frac{\theta(1 - \theta)}{n}.$$

3. Оценка параметра распределения Пуассона

Пусть $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, где ξ_k имеют распределение Пуассона

$$f(x, \theta) = P(\xi = x) = \frac{\theta^x}{x!} e^{-\theta}, \quad x = 0, 1, \dots, \quad \theta > 0.$$

Функция правдоподобия выборки

$$\Lambda(\theta) = \prod_{k=1}^n f(\xi_k, \theta) = \frac{1}{\xi_1! \xi_2! \dots \xi_n!} e^{-n\theta} \theta^{\sum_{k=1}^n \xi_k}.$$

Так как

$$\frac{\partial \ln \Lambda(\theta)}{\partial \theta} = \frac{1}{\theta} \sum_{k=1}^n \xi_k - n = \frac{n}{\theta} \left(\frac{1}{n} \sum_{k=1}^n \xi_k - \theta \right),$$

то снова существует эффективная несмещенная оценка параметра

$$\check{\theta} = \frac{1}{n} \sum_{k=1}^n \xi_k = \bar{\xi}.$$

Дисперсия этой оценки

$$D\check{\theta} = \frac{1}{n} D\xi_1 = \frac{\theta}{n}.$$

4. Распределение Коши

Рассматриваем выборку объема n , компоненты которой независимы и имеют плотность распределения

$$f(x, \theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}.$$

Имеем

$$\frac{\partial \ln \Lambda(\theta)}{\partial \theta} = 2 \sum_{i=1}^n \frac{\xi_i - \theta}{1 + (\xi_i - \theta)^2}.$$

В этом случае ни для какой функции $A(\theta)$ параметра θ не существует эффективной оценки. Заметим, что случайные величины $\frac{\xi_k - \theta}{1 + (\xi_k - \theta)^2}$ имеют среднее нуль и дисперсию

$$D\left(\frac{\xi_k - \theta}{1 + (\xi_k - \theta)^2}\right) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{(x - \theta)^2}{(1 + (x - \theta)^2)^3} dx.$$

Интеграл в правой части равенства вычисляется с помощью повторного интегрирования по частям

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{(x - \theta)^2}{(1 + (x - \theta)^2)^3} dx &= \int_{-\infty}^{\infty} \frac{x^2 dx}{(1 + x^2)^3} = \frac{1}{4} \int_{-\infty}^{\infty} \frac{dx}{(1 + x^2)^2} = \\ &= \frac{1}{4} \left[\int_{-\infty}^{\infty} \frac{dx}{1 + x^2} - \int_{-\infty}^{\infty} \frac{xx dx}{(1 + x^2)^2} \right] = \\ &= \frac{1}{4} \left[\pi - \frac{1}{2} \int_{-\infty}^{\infty} \frac{dx}{1 + x^2} \right] = \frac{\pi}{8}. \end{aligned}$$

Отсюда мы получаем для фишеровского количества информации значение

$$I(\theta) = M\left(\frac{\partial}{\partial \theta} \ln A(\theta)\right)^2 = \frac{n}{2},$$

и нижняя граница Крамера — Рао дисперсии несмещенных оценок параметра θ в плотности Коши равна $\frac{2}{n}$.

Из приведенных выше результатов вытекает, что состоятельными оценками параметра θ является эмпирическая медиана $\xi_{\frac{1}{2}}$. Для дисперсии $\xi_{\frac{1}{2}}$ при больших n имеем

$$D\xi_{\frac{1}{2}} \approx \frac{\pi^2}{4n}.$$

Так как $\frac{\pi^2}{4} > 2$, то оценка $\xi_{\frac{1}{2}}$ не является асимптотически эффективной.

Нижние границы оценок векторного параметра. Пусть плотность распределения $f(x, \theta) = f(x, \theta_1, \dots, \theta_s)$ случайного вектора ξ (со значениями в R^n) зависит от параметра θ , принимающего значения в Θ , где Θ или все R^s или s -мерный параллелепипед.

Предположим, что:

а) производные $\frac{\partial}{\partial \theta} f(x, \theta)$ существуют и

$$M \left| \frac{\partial}{\partial \theta_j} \ln f(\xi, \theta) \right| < \infty, \quad j = 1, \dots, s, \quad \forall \theta \in \Theta;$$

б) матрица $I(\theta) = \{I_{ij}(\theta)\}$ с элементами $I_{ij}(\theta) = M \frac{\partial}{\partial \theta_i} \ln f(\xi, \theta) \frac{\partial}{\partial \theta_j} \ln f(\xi, \theta)$; $i, j = 1, \dots, s$ положительно определена $\forall \theta \in \Theta$.

Матрицу $I(\theta)$ называют информационной матрицей Фишера.

Пусть векторная функция $\check{h}(x) = (\check{h}_1(x), \dots, \check{h}_d(x))$, $x \in R^n$, представляет собой несмещенную оценку $\check{h}(\zeta)$ d -мерного параметра $h(\theta) = (h_1(\theta), \dots, h_d(\theta))$,

$$M\check{h}(\zeta) = h(\theta), \quad (19)$$

причем функции $\check{h}_k(x)$ обладают частными производными $\frac{\partial}{\partial x_j} \check{h}_k(x)$.

В дальнейшем используются матричные обозначения и векторы $h(\theta)$, $\check{h}(\theta)$ рассматриваются как матрицы, состоящие из одного столбца. Через $\nabla = \nabla_\theta$ обозначим векторную (вектор-столбец) операцию — вычисление градиента функции по переменной θ ,

$$\nabla g(\theta) = \left(\frac{\partial g}{\partial \theta_1}, \dots, \frac{\partial g}{\partial \theta_s} \right).$$

Введем знак $*$, который означает переход к транспонированной матрице. Например, $\nabla h^*(\theta)$ — это матрица с элементами

$$a_{kj} = \frac{\partial}{\partial \theta_k} h_j(\theta),$$

$k = 1, \dots, s$; $j = 1, \dots, d$. Заметим, что если a — вектор-столбец, то aa^* — квадратная неотрицательно определенная матрица.

Обозначим

$$D(\theta) = M(\check{h}(\zeta) - h(\theta))(\check{h}(\zeta) - h(\theta))^*,$$

$$H(\theta) = M\nabla \ln f(\zeta, \theta) \check{h}^*(\zeta).$$

Теорема 4. Пусть семейство плотностей $f(x, \theta)$, $\theta \in \Theta$, удовлетворяют условиям а), б), $\check{h}(\zeta)$ — несмещенная оценка векторного параметра $h(\theta)$, для которой

$$\int \check{h}_k(x) \left| \frac{\partial}{\partial \theta_j} f(x, \theta) \right| dx < \infty, \quad (20)$$

$$j = 1, \dots, s; k = 1, \dots, d, \theta \in \Theta.$$

Тогда матрица $D(\theta) - H^*(\theta)I^{-1}(\theta)H(\theta)$ — неотрицательно определена для всех θ .

Доказательство. Рассмотрим неотрицательно определенную матрицу ($H = H(\theta)$, $I = I(\theta)$)

$$B = M(H^*I^{-1}\nabla \ln f(\zeta, \theta) - \check{h}(\zeta) - h(\theta)) \times \\ \times (H^*I^{-1}\nabla \ln f(\zeta, \theta) - \check{h}(\zeta) - h(\theta))^*.$$

Имеем

$$B = H^*I^{-1}M\nabla \ln f(\zeta, \theta)(\nabla \ln f(\zeta, \theta))^*I^{-1}H - H^*I^{-1}M\Delta \ln f \times \\ \times (\zeta, \theta)(\check{h}(\zeta) - h(\theta))^* - M(\check{h}(\zeta) - h(\theta))(\nabla \ln f(\zeta, \theta))^*I^{-1}H + \\ + M(\check{h}(\zeta) - h(\theta))(\check{h}(\zeta) - h(\theta))^*.$$

Заметим, что

$$\int f(x, \theta) \check{h}_j(x) dx = h_j(\theta).$$

Согласно предположению а), можно дифференцировать по θ под знаком интеграла. Получаем

$$\int \nabla f(x, \theta) \check{h}^*(x) dx = \nabla h^*(\theta),$$

или

$$M \nabla \ln f(\zeta, \theta) \check{h}^*(\zeta) = \nabla h^*(\theta). \quad (21)$$

Кроме того, из равенства

$$\int f(x, \theta) dx = 1$$

с помощью дифференцирования имеем

$$M \nabla \ln f(\zeta, \theta) = 0. \quad (22)$$

Из формул (21) и (22) вытекает

$$M \nabla \ln f(\zeta, \theta) (\check{h}(\zeta) - h(\theta))^* = M \nabla \ln f(\zeta, \theta) \check{h}^*(\zeta) = H,$$

откуда

$$B = -H^*(\theta) I^{-1}(\theta) H(\theta) + D(\theta).$$

Теорема доказана.

Выделим два частных случая доказанной теоремы.

а). Пусть $h(\theta_1, \dots, \theta_n)$ — одномерная статистика. Тогда $D(\theta)$ — скалярная величина, $H(\theta)$ — вектор-столбец с элементами

$$H_j(\theta) = M \frac{\partial}{\partial \theta_j} f(\zeta, \theta) h(\zeta), \quad j = 1, \dots, s,$$

и из теоремы 4 вытекает неравенство

$$D(\theta) \geq H^*(\theta) I^{-1}(\theta) H(\theta). \quad (23)$$

Знак равенства в соотношении (23) достигается тогда и только тогда, когда существует вектор $Q(\theta)$ такой, что

$$Q^*(\theta) \nabla \ln f(\zeta, \theta) = \check{h}(\zeta) - h(\theta). \quad (24)$$

В этом случае функция $h(\theta)$ имеет несмещенную оценку $\check{h}(\zeta)$ минимальной дисперсии $H^* I^{-1} H$, где $H = I(\theta) Q(\theta)$.

Действительно, равенство (24) имеет место в том случае, когда матрица $B(\theta)$, введенная при доказательстве теоремы 4, равна 0. В этом случае $H^* I \nabla \ln f(\zeta, \theta) = \check{h}(\zeta) - h(\theta)$ с вероятностью 1. Обратно, если имеет место (24), то, во-первых,

$$M \check{h}(\zeta) - h(\theta) = Q^*(\theta) M \nabla \ln f(\zeta, \theta) = 0,$$

т. е. $\check{h}(\xi)$ является несмещенной оценкой параметра θ , а во-вторых,

$$\begin{aligned} Q^*(\theta) M \nabla \ln f(\xi, \theta) (\nabla \ln f(\xi, \theta))^* = \\ = M (\nabla \ln f(\xi, \theta) \check{h}^*(\xi))^* = H^*(\theta), \end{aligned}$$

или $Q^*(\theta) I(\theta) = H^*(\theta)$, $Q^*(\theta) = H^*(\theta) I^{-1}(\theta)$, так что $B(\theta) = 0$.

б) Пусть $h(\theta) = \theta$. Тогда элементы $H_{kj}(\theta)$ матрицы $H(\theta)$ ($k, j = 1, \dots, s$) равны:

$$H_{kj}(\theta) = M \frac{\partial}{\partial \theta_k} \ln f(\xi, \theta) \check{h}_j(\xi) = \frac{\partial}{\partial \theta_k} M \check{h}_j(\xi) = \frac{\partial}{\partial \theta_k} \theta_j$$

на основании несмещенности оценки $\check{h}(\xi)$ параметра θ . Таким образом, $H_{kj}(\theta) = \delta_{kj}$, $H(\theta)$ — единичная матрица.

Теорема 4 в этом случае утверждает, что матрица $D(\theta) = I^{-1}(\theta)$ неотрицательно определена. Пусть z — произвольный вектор из R^s . Выражение типа $z^* D z$ является квадратической формой: $z^* D z = \sum_{k,j=1}^s D_{kj} z_k z_j$. Неотрицательная определенность матрицы $D(\theta) = I^{-1}(\theta)$ означает, что для всех $z \in R^s$

$$z^* D z \geq z^* I^{-1}(\theta) z. \quad (24)$$

Легко заметить, что знак равенства здесь достигается тогда и только тогда, когда

$$\nabla \ln f(\xi, \theta) = I(\theta) (\check{h}(\xi) - \theta). \quad (25)$$

Оценки $\check{h}_j(\xi)$ компонент θ_j , $j = 1, \dots, s$, параметрического вектора θ , для которых неравенство (24) обращается в равенство $\forall z \in R^s$, т. е. для которых $D(\theta) = I^{-1}(\theta)$, называются совместно эффективными. Для того чтобы существовали совместно эффективные оценки параметров θ_j , $j = 1, \dots, s$, в условиях теоремы 4 необходимо и достаточно, чтобы функция правдоподобия $f(\xi, \theta)$ удовлетворяла равенству (25).

К сожалению, условие (25) является весьма ограничительным и не выполняется в сравнительно простых случаях.

Введем понятие асимптотически совместно эффективных оценок. Пусть $\check{\theta}^{(n)}$ — последовательность несмещенных оценок параметрического вектора θ , построенных по выборке объема n , $I^{(n)}(\theta)$ — матрица количества информации соответствующей выборки, $D^{(n)}(\theta)$ — дисперсионная матрица оценки $\check{\theta}^{(n)}$. Если

$$I^{(n)-1}(\theta) D^{(n)}(\theta) \rightarrow 1 \quad \text{при } n \rightarrow \infty,$$

то последовательности оценок $\check{\theta}^{(n)}$ назовем асимптотически совместно эффективными.

1. Совместные оценки параметров нормального распределения

Пусть оба параметра a и θ нормальной плотности распределения

$$f(x, a, \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-a)^2}{2\theta}}$$

не известны. Оценка этих параметров производится по n независимым и одинаково распределенным случайным величинам ξ_1, \dots, ξ_n с плотностью распределения $f(x, a, \theta)$. Функция правдоподобия имеет вид

$$\Lambda(\theta) = (2\pi\theta)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\theta} \sum_{k=1}^n (\xi_k - a)^2 \right\}.$$

Условия применимости теоремы 4 выполнены. Имеем

$$\frac{\partial \ln \Lambda}{\partial a} = \frac{1}{\theta} \sum_{k=1}^n (\xi_k - a), \quad \frac{\partial \ln \Lambda}{\partial \theta} = \frac{1}{2\theta^2} \left(\sum_{k=1}^n (\xi_k - a)^2 - n\theta \right).$$

Совместно эффективные оценки параметров a и θ отсутствуют, так как равенство (25) не может быть выполнено. Для элементов информационной матрицы получим следующие выражения:

$$M \left(\frac{\partial \ln \Lambda}{\partial a} \right)^2 = \frac{n}{\theta}, \quad M \frac{\partial \ln \Lambda}{\partial a} \frac{\partial \ln \Lambda}{\partial \theta} = 0, \quad M \left(\frac{\partial \ln \Lambda}{\partial \theta} \right)^2 = \frac{n}{2\theta^2},$$

поскольку

$$\begin{aligned} M \left(\sum_{k=1}^n (\xi_k - a) \right)^2 &= D \left(\sum_{k=1}^n (\xi_k - a) \right) = nD(\xi_k - a) = n\theta, \\ M \left[\sum_{k=1}^n (\xi_k - a) \right] \left(\sum_{k=1}^n (\xi_k - a)^2 - n\theta \right) &= M \sum_{k,j=1}^n (\xi_k - a) (\xi_j - a)^2 = 0, \\ M \left[\sum_{k=1}^n (\xi_k - a)^2 - n\theta \right]^2 &= D \left(\sum_{k=1}^n (\xi_k - a)^2 \right) = nD[(\xi_k - a)^2] = 2n\theta^2. \end{aligned}$$

Таким образом,

$$I(\theta) = \begin{pmatrix} \frac{n}{\theta} & 0 \\ 0 & \frac{n}{2\theta^2} \end{pmatrix}.$$

Причем в качестве оценок параметров a и θ несмещенные статистики

$$\bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k, \quad \bar{s}^2 = \frac{1}{n-1} \sum_{k=1}^n (\xi_k - \bar{\xi})^2.$$

Вычислим корреляционную матрицу $D(\theta)$ вектора $(\bar{\xi}, \bar{s}^2)$. Заметим, что

$$\sum_{k=1}^n (\xi_k - a)^2 = \sum_{k=1}^n (\xi_k - \bar{\xi})^2 + n(\bar{\xi} - a)^2. \text{ Поэтому}$$

$$D \left[\sum_{k=1}^n (\xi_k - \bar{\xi})^2 \right] M = \left(\sum_{k=1}^n [(\xi_k - a)^2 - \theta] - [n(\bar{\xi} - a)^2 - \theta] \right)^2 = \\ = M \left(\sum_{k=1}^n (\xi_k - a)^2 - \theta \right)^2 - 2M \left(\sum_{k=1}^n [(\xi_k - a)^2 - \theta] - \right. \\ \left. - \theta \right) (n(\bar{\xi} - a)^2 - \theta) + M [n(\bar{\xi} - a)^2 - \theta]^2.$$

Так как величины $(\xi_k - a)^2 - \theta$ независимы, то

$$M \left(\sum_{k=1}^n (\xi_k - a)^2 - \theta \right)^2 = n D (\xi_k - a)^2 = 2\theta^2 n.$$

Далее,

$$M [n(\bar{\xi} - a)^2 - \theta]^2 = D (n(\bar{\xi} - a)^2) = n^2 D (\bar{\xi} - a)^2 = 2\theta^2$$

на основании того, что $\bar{\xi} - a$ является гауссовской случайной величиной со средним 0 и дисперсией $\frac{\theta}{n}$. Наконец,

$$M \left(\sum_{k=1}^n (\xi_k - a)^2 - \theta \right) (n(\bar{\xi} - a)^2 - \theta) = n M [n(\xi_k - a)^2 (\bar{\xi} - a)^2 - \theta^2], \\ M (\xi_k - a)^2 (\bar{\xi} - a)^2 = \frac{1}{n^2} M (\xi_k - a)^2 \left(\sum_{j=1}^n (\xi_j - a)^2 \right) = \frac{1}{n^2} [M (\xi_k - a)^4 + \\ + \sum_{j \neq k} M (\xi_k - a)^2 (\xi_j - a)^2] = \frac{1}{n^2} [3\theta^2 + (n-1)\theta^2] = \frac{n+2}{n^2} \theta^2.$$

Следовательно,

$$D \left[\sum_{k=1}^n (\xi_k - \bar{\xi})^2 \right] = 2\theta^2 n - 2n \left(\frac{n+2}{n} - 1 \right) \theta^2 + 2\theta^2 = 2\theta^2 (n-1).$$

Корреляция между \bar{s}^2 и $\bar{\xi}$ равна 0. Действительно,

$$M (\bar{s}^2 - \theta) (\bar{\xi} - a) = \frac{1}{n-1} M \left[\sum_{k=1}^n (\xi_k - a)^2 - n(\bar{\xi} - a)^2 \right] (\bar{\xi} - a) = \frac{1}{n-1} \times \\ \times \left(\sum_{k=1}^n M (\xi_k - a)^2 (\bar{\xi} - a) - n M (\bar{\xi} - a)^3 \right) = 0$$

вследствие того, что нечетные моменты гауссовских случайных величин $\xi_k - a$ равны 0. Мы получаем для корреляционной матрицы следующее значение:

$$D(\theta) = \begin{pmatrix} \frac{\theta}{n} & 0 \\ 0 & \frac{2\theta^2}{n-1} \end{pmatrix},$$

причем

$$I(\theta) D(\theta) = \begin{pmatrix} 1 & 0 \\ 0 & \frac{n}{n-2} \end{pmatrix} \rightarrow 0.$$

Таким образом, оценки $(\bar{\xi}, \bar{s}^2)$ параметров (a, θ) нормального распределения являются асимптотически совместно эффективными.

2. Оценка среднего в многомерном нормальном распределении.

Пусть выборочный вектор ξ образован n независимыми нормально распределенными векторами $\xi^{(1)}, \dots, \xi^{(n)}$ с неизвестной невырожденной корреляционной матрицей R и неизвестным средним $\theta = (\theta_1, \dots, \theta_p)$. Функция правдоподобия

$$f(\xi, \theta) = \frac{1}{(2\pi \text{Det } R)^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{k=1}^n (\xi^{(k)} - \theta)^* R^{-1} (\xi^{(k)} - \theta) \right\}.$$

Поскольку

$$\frac{\partial}{\partial \theta} \ln f(\xi, \theta) = R^{-1} \sum_{k=1}^n (\xi^{(k)} - \theta),$$

то для вектора θ существует совместно эффективная оценка

$$\bar{\theta} = \frac{1}{n} \sum_{k=1}^n \xi^{(k)}.$$

При этом

$$I(\theta) = nR^{-1}, \quad D(\bar{\theta}) = \frac{1}{n} R.$$

§ 2. Достаточные статистики

Пусть ξ — случайный вектор со значениями в R^n , плотность распределения которого $f(x, \theta)$ зависит от неизвестного параметра θ , $\theta \in \Theta$, где Θ подмножество в R^p . В ряде задач параметр θ можно считать случайным вектором, принимающим в рассматриваемых реализациях вектора ξ фиксированное значение. Информация, которую несет в себе реализация вектора ξ о значении параметра θ , содержится в условном распределении вектора θ при данном ξ . Предположим, что для вектора θ существует плотность распределения $g(\theta)$.

Плотность $g(\theta)$ называют априорной плотностью распределения параметра θ . Условная плотность распределения параметра θ при гипотезе $\xi = x$ задается формулой Байеса

$$f(\theta/x) = \frac{f(x, \theta) g(\theta)}{\int_{\Theta} f(x, \theta) g(\theta) d\theta}.$$

Заметим, что если плотность $f(x, \theta)$ распределения вектора имеет вид

$$f(x, \theta) = \varphi(T(x), \theta) h(x), \quad (1)$$

где функции $h(x)$ и $T(x)$ не зависят от θ , а $T(x)$ — некоторая измеримая скалярная или векторная функция, то

$$f(\theta/x) = \frac{\varphi(T(x), \theta) g(\theta)}{\int_{\Theta} \varphi(T(x), \theta) g(\theta) d\theta} = \psi(T(x), \theta),$$

т. е. условное распределение параметра θ при любой априорной плотности $g(\theta)$ зависит только от θ и $T(x)$.

Отсюда можно сделать вывод, что статистика $T(\xi)$ несет в себе всю информацию о параметре θ , содержащуюся в векторе ξ . Если размерность статистики T меньше размерности вектора ξ , то статистика T осуществляет «сокращение эмпирических данных», содержащихся в ξ , не уменьшая информации о параметре θ . Чем меньше размерность вектора T , тем это сокращение сильнее. Особое практическое значение такое сокращение играет в тех случаях, когда рассматриваются выборки большого объема n , а последовательности соответствующих статистик T имеют размерность, не зависящую от n .

Нетрудно заметить, что если «апостериорное» распределение $f(\theta/x)$ параметра θ зависит только от θ и некоторой статистики $T(x)$, $f(\theta/x) = K(T(x), \theta)$ для некоторой априорной плотности $g(\theta)$, $g(\theta) > 0$, $\forall \theta \in \Theta$, то плотность распределения $f(x, \theta)$ вектора ξ имеет вид (1). Действительно, в этом случае

$$f(x, \theta) = \frac{1}{g(\theta)} K(T(x), \theta) \int_{\Theta} f(x, \theta) g(\theta) d\theta = \varphi(T(x), \theta) h(x),$$

где

$$\varphi(T, \theta) = K(T, \theta)/g(\theta), \quad h(x) = \int_{\Theta} f(x, \theta) g(\theta) d\theta.$$

Определение. Если семейство плотностей распределения $f(x, \theta)$, $\theta \in \Theta$ допускает представление (1), то вектор $T(\xi)$ называют достаточной статистикой для данного семейства распределений (или для параметра θ).

Понятие достаточной статистики можно разъяснить и не рассматривая параметр θ как случайный вектор, а именно, достаточные статистики характеризуются следующим свойством: статистика $T(\xi)$ достаточна для семейства плотностей $\{f(x, \theta) \mid \theta \in \Theta\}$ вектора ξ тогда и только тогда, когда условное распределение вектора ξ при данном $T(\xi)$ не зависит от θ .

Для доказательства допустим, что $T(x)$ — s -мерная векторная функция, $T(x) = (T_1(x), T_2(x), \dots, T_s(x))$ и существует $n - s$ -мерная векторная функция $K(x) = (K_1(x), \dots, K_{n-s}(x))$ такая, что отображение $x \rightarrow (T_1(x), \dots, T_s(x), K_1(x), \dots, K_{n-s}(x))$ взаимно однозначно и дифференцируемо. При этом плотности распределения вектора ξ и $f_{(T, K)}(t, y)$, $t \in R^s$, $y \in R^{n-s}$ вектора $(T(\xi), K(\xi))$ связаны соотношением

$$f_{\xi}(x) = f_{(T, K)}(T(x), K(x)) |J|, \quad (2)$$

где J — якобиан преобразования $x \rightarrow (T(x), K(x))$.

Пусть $S(t, y)$, $t \in R^s$, $y \in R^{n-s}$ — преобразование, обратное преобразованию $(T(x), K(x))$, $\varphi(x)$ — произвольная ограниченная борелевская функция в R^n , $M\{\varphi(\xi)/t\}$ — условное математическое

ожидание величины $\varphi(\xi)$ при гипотезе $T = t$, $p_T(t)$ — плотность распределения статистики $T(\xi)$:

$$p_T(t) = \int_{R^{n-s}} f_{(T, K)}(t, y) dy.$$

Функция $M\{\varphi(\xi)/t\}$ удовлетворяет соотношению

$$M\{\varphi(\xi)/t\} \chi(T) = \int_{R^s} \chi(t) M\{\varphi(\xi)/t\} p_T(t) dt$$

для любых ограниченных и измеримых функций $\chi(t)$.

Кроме того,

$$\begin{aligned} M\{\varphi(\xi)/t\} \chi(T) &= \int_{R^n} \varphi(x) \chi(T(x)) f_{\xi}(x, \theta) dx = \\ &= \int_{R^n} \varphi(s(t, y)) \chi(t) f_{\xi}(s(t, y)) |J| dt dy = \\ &= \int_{R^n} \varphi(s(t, y)) \chi(t) f_{(T, K)}(t, y) dt dy = \\ &= \int_{R^s} \left(\int_{R^{n-s}} \varphi(s(t, y)) \frac{f_{(T, K)}(t, y)}{p_T(t)} dy \right) \chi(t) p_T(t) dt. \end{aligned}$$

Так как функция $\chi(t)$ — произвольная, то

$$M\{\varphi(\xi)/t\} = \int_{R^{n-s}} \varphi(s(t, y)) \frac{f_{(T, K)}(t, y)}{p_T(t)} dy. \quad (3)$$

Пусть теперь плотность $f_{\xi}(x) = f(x, \theta)$ имеет вид (1). Тогда на основании равенства (2) $f_{(T, K)}(t, y) = \psi(t, \theta) h(t, y)$ и, следовательно, $p_T(t) = \psi(t, \theta) h_1(t)$, где $\psi(t, \theta)$, $h(t, y)$, $h_1(t)$ — некоторые функции, зависящие от указанных аргументов. Следовательно, условная плотность $f_{(T, K)}(t, y)/p_T(t)$ не зависит от θ . Поэтому $M\{\varphi(\xi)/t\}$ не зависит от θ . Так как $\varphi(x)$ произвольна, то условное распределение вектора ξ при гипотезе $T = t$ не зависит от параметра θ . Обратно, пусть это условное распределение, а вместе с ним и $M\{\varphi(\xi)/t\}$ для любой ограниченной борелевской функции $\varphi(x)$ не зависит от параметра θ . Тогда согласно (3) $f_{(T, K)}(t, y, \theta)/p_T(t, \theta) = h_2(t, y)$, где $h_2(t, y)$ не зависит от θ . Таким образом, $f_{(T, K)}(t, y, \theta) = p_T(t, \theta) h_2(t, y)$ и из (2) следует $f(x, \theta) = p_T(T(x), \theta) \tilde{h}(x)$. Это завершает доказательство нашего утверждения. Мы показали, что аналитическое представление функции $f(x, \theta)$ в виде (1) эквивалентно следующему теоретико-вероятностному свойству статистики $T(\xi)$: условное распределение вектора ξ (а, следовательно, и любой статистики от ξ) при заданном $T(\xi)$ не зависит от параметра θ .

Формулу (1) называют факторизационным предположением плотности распределения. Отметим следующее важное свойство достаточной статистики.

Пусть $h(\xi)$ — несмещенная оценка параметра θ и $T = T(\xi)$ — достаточная статистика. Тогда существует несмещенная оценка параметра θ , являющаяся функцией от достаточной статистики и имеющая не меньшую эффективность, чем $h(\xi)$.

Теорема 1. Пусть T — достаточная статистика для семейства плотностей $\{f(x, \theta), \theta \in \Theta\}$, $h(\xi)$ — несмещенная оценка параметра θ . Положим $M\{h(\xi)/T\} = K(T)$. Тогда $K(T)$ — несмещенная оценка параметра θ

$$M(K(T) - \theta)^2 \leq M(h(\xi) - \theta)^2,$$

причем знак равенства здесь имеет место только тогда, когда $h(\xi) = h_1(T)$ с вероятностью 1.

Доказательство. Действительно, на основании свойства условного математического ожидания

$$MK(T) = MM(h(\xi)/T) = Mh(\xi) = \theta,$$

так что $K(T)$ является несмещенной оценкой параметра θ .

Далее, используя неравенство Коши — Буняковского $(M\xi)^2 \leq M\xi^2$ и упомянутое выше свойство условных математических ожиданий, имеем

$$\begin{aligned} M(K(T) - \theta)^2 &= M(M\{(h(\xi) - \theta)/T\})^2 \leq \\ &\leq MM\{(h(\xi) - \theta)^2/T\} = M(h(\xi) - \theta)^2, \end{aligned}$$

причем, $M(K(T) - \theta)^2 = M(h(\xi) - \theta)^2$ тогда и только тогда, когда $M\{(h(\xi) - \theta)/T\} = h(\xi) - \theta$, т. е.

$$h(\xi) = M\{h(\xi)/T\} = h_1(T).$$

Следствие. Если для параметра θ существует оценка минимальной дисперсии, то она является функцией от достаточной статистики.

ПРИМЕРЫ

1. Пусть $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, ξ_k — гауссовские случайные величины с параметрами $a = M\xi_k$, $\theta = D\xi_k$. Функция правдоподобия для вектора ξ имеет вид

$$\begin{aligned} f(\xi, a, \theta) &= (2\pi\theta)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\theta} \sum_{k=1}^n (\xi_k - a)^2\right\} = \\ &= (2\pi\theta)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\theta} \left[\sum_{k=1}^n (\xi_k - \bar{\xi})^2 - n(\bar{\xi} - a)^2\right]\right\}, \quad \bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k. \end{aligned}$$

При неизвестных параметрах (α, θ) семейство распределений $f(\xi, \alpha, \theta)$ имеет двумерную достаточную статистику $(\bar{\xi}, ns^2)$, $ns^2 = \sum_{k=1}^n (\xi_k - \bar{\xi})^2$. Это непосредственно вытекает из факторизационного равенства (1). Отсюда также вытекает, что при известном значении θ достаточной статистикой для параметра α является $\bar{\xi}$, а при известном значении α достаточной статистикой для параметра θ является $\sum_{k=1}^n (\xi_k - \alpha)^2$, но не $\sum_{k=1}^n (\xi_k - \bar{\xi})^2$.

2. Пусть ξ_k , $k = 1, \dots, n$ имеют равномерное распределение на отрезке $(0, 1)$. Тогда

$$f(\xi, \theta) = \frac{1}{\theta^n} I(\theta - \xi(n)),$$

где $\xi(n)$ — максимальный член последовательности $\xi_1, \xi_2, \dots, \xi_n$, $I(x) = 1$ при $x > 0$ и $I(x) = 0$ при $x < 0$. Таким образом, статистика $\xi(n) = \max_{1 \leq k \leq n} \xi_k$

достаточна для параметра θ .

Аналогичные соображения показывают, что если ξ_k имеют равномерное распределение на интервале (α, β) , концы которого неизвестны, то пара статистик $(\xi(1), \xi(n))$, $\xi(1) = \min_{1 \leq k \leq n} \xi_k$, является достаточной для пары параметров (α, β) .

§ 3. Метод максимального правдоподобия

Пусть наблюдается случайный вектор ξ , плотность распределения которого $f(x, \theta)$ зависит от параметра θ , принимающего значения из некоторого множества Θ , $\Theta \in R^s$, $\theta = (\theta_1, \theta_2, \dots, \theta_s)$.

Метод оценки параметра θ , получивший название метода максимального правдоподобия, основан на следующем интуитивном представлении: в большей части случаев в эксперименте наблюдается то значение вектора ξ , при котором плотность вероятности $f(x, \theta)$ близка к максимальному значению.

Таким образом, в качестве оценки $\hat{\theta}$ параметра θ естественно принять решение уравнения

$$f(\xi, \hat{\theta}) = \max_{\theta} f(\xi, \theta), \quad (1)$$

если это уравнение имеет смысл и его решение существует.

Решения уравнения (1) называются оценками максимального правдоподобия.

Во многих случаях вместо функции $f(x, \theta)$ проще рассматривать функцию $\ln f(x, \theta)$, достигающую максимума в тех же точках, что и $f(x, \theta)$. Если функция $f(x, \theta)$ непрерывно дифференцируема по θ , то для решения уравнения (1) можно сначала найти стационарные точки функции $f(\xi, \theta)$, а затем сравнить значения функции $f(\xi, \theta)$ в этих и граничных точках множества Θ и выбрать точку максимума $f(x, \theta)$. Для отыскания стационарных точек имеем уравнения $\frac{\partial \ln f(\xi, \theta)}{\partial \theta_i} = 0$, $i = 1, 2, \dots, s$.

Эти уравнения называют уравнениями максимального правдоподобия.

Метод максимального правдоподобия в некоторых случаях обладает оптимальными свойствами. В случае большого числа n независимых наблюдений при достаточной регулярности плотности распределения он асимптотически оптимален, но практические применения метода максимального правдоподобия часто приводят к весьма сложным вычислениям. И к тому же при небольших значениях n не всегда ясно, насколько этот метод целесообразен.

Отметим несколько свойств оценок максимального правдоподобия.

а) Если для семейства плотностей $f(x, \theta)$ существует достаточная статистика, то решение уравнения правдоподобия (когда оно существует и единственно) является функцией от достаточной статистики.

Действительно, если $f(x, \theta) = h(x) g(T(x), \theta)$, то уравнение (1) принимает вид

$$g(t, \tilde{\theta}) = \max_{\theta} g(t, \theta),$$

откуда $\tilde{\theta} = \varphi(t) = \varphi(T(\xi))$.

б) Если выполнены условия регулярности теоремы 1 и существует эффективная оценка $\check{h}(\xi)$ одномерного параметра θ , то $\check{h}(\xi)$ является также оценкой максимального правдоподобия.

Действительно, если выполнены условия теоремы 3, § 1, и существует эффективная оценка $\check{h}(\xi)$ параметра θ , то на основании этой теоремы для функции $\frac{\partial}{\partial \theta} \ln f(\xi, \theta)$ имеет место равенство (14), § 1. Уравнение максимального правдоподобия записывается следующим образом:

$$\frac{\partial}{\partial \theta} \ln f(\xi, \theta) = c(\theta) [\check{h}(\xi) - \theta] = 0.$$

Его решение $\theta = \tilde{\theta}$ имеет вид $\tilde{\theta} = \check{h}(\xi)$, т. е. $\tilde{\theta}$ является эффективной оценкой.

в) Пусть функция $h(\theta)$, $\theta \in \Theta$ взаимно однозначно отображает Θ в Q , где Q некоторое множество в R^s . Если $\check{\theta}$ — оценка параметра θ семейства плотностей $\{f(x, \theta), \theta \in \Theta\}$, то $h(\check{\theta})$ является оценкой максимального правдоподобия параметра $h(\theta)$.

Доказательство. Пусть $\check{\theta} = v(q)$, где $v(q)$ — функция, обратная $h(\theta)$. Тогда $\max_{\theta \in \Theta} f(x, \theta) = \max_{q \in Q} f(x, v(q))$, причем если

максимум по θ функции $f(x, \theta)$ достигается в точке $\check{\theta} = \check{\theta}(x)$, то максимум по q функции $f(x, v(q))$ достигается в точке \check{q} , являющейся решением уравнения $v(\check{q}) = \check{\theta}$, т. е. в точке $\check{q} = \check{h}(\theta)$.

1. Найденные ранее в примерах 1—3 (с. 343) эффективные оценки параметров распределений являются одновременно оценками максимального правдоподобия. Мы не станем проводить вычисления, подтверждающие это следствием свойства б) оценок максимального правдоподобия.

2. Найдем оценки максимального правдоподобия параметров $a = M\xi_k$, $\theta = D\xi_k$ выборки объема n из нормальной генеральной совокупности. Функция правдоподобия

$$\Delta(\theta) = (2\pi\theta)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2} \sum_{k=1}^n (\xi_k - a)^2 \right\}.$$

Уравнения правдоподобия

$$\frac{\partial \ln \Delta}{\partial a} = \frac{1}{\theta} \sum_{k=1}^n (\xi_k - a) = 0,$$

$$\frac{\partial \ln \Delta}{\partial \theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{k=1}^n (\xi_k - a)^2 = 0$$

имеют решения

$$\hat{a} = \frac{1}{n} \sum_{k=1}^n \xi_k = \bar{\xi}, \quad \hat{\theta} = \frac{1}{n} \sum_{k=1}^n (\xi_k - \bar{\xi})^2.$$

Как было показано выше, оценка \hat{a} является несмещенной эффективной оценкой среднего a , а вторая — асимптотически несмещенной и асимптотически эффективной оценкой дисперсии σ^2 .

3. Оценивание параметров лог-нормального распределения. Лог-нормальным распределением называют распределение положительной случайной величины ξ , если $\eta = \ln \xi$ имеет нормальное распределение.

Пусть a и θ — параметры этого распределения. Тогда

$$P(\xi < x) = P(\ln \xi < \ln x) = \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^{\ln x} \exp \left\{ -\frac{1}{2\theta} (u - a)^2 \right\} du.$$

Следовательно, плотность $g(x, a, \theta)$ случайной величины ξ имеет вид

$$g(x, a, \theta) = \frac{1}{\sqrt{2\pi\theta}} \exp \left\{ -\frac{1}{2\theta} (\ln x - a)^2 \right\} \frac{1}{x}, \quad x > 0.$$

Пусть m — математическое ожидание, D — дисперсия величины ξ . Тогда

$$m = M\xi = Me^{\eta} = \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\theta} (x - a)^2 \right\} dx = e^{a + \frac{\theta}{2}}.$$

$$M\xi^2 = Me^{2\eta} = \frac{1}{\sqrt{2\pi\theta}} \int_{-\infty}^{\infty} \exp \left\{ 2x - \frac{1}{2\theta} (x - a)^2 \right\} dx = e^{2(a + \theta)} = m^2 e^{\theta},$$

$$D = m^2 (e^{\theta} - 1).$$

Пусть $\xi = (\xi_1, \dots, \xi_n)$ — выборка объема n из лог-нормальной генеральной совокупности. Имеем:

$$\ln f(\xi, a, \theta) = -\frac{n}{2} \ln 2\pi\theta - \sum_{k=1}^n \ln \xi_k - \frac{1}{2\theta} \sum_{k=1}^n (\ln \xi_k - a)^2,$$

$$\frac{\partial \ln f}{\partial a} = \frac{1}{\theta} \sum_{k=1}^n (\ln \xi_k - a), \quad \frac{\partial \ln f}{\partial \theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{k=1}^n (\ln \xi_k - a)^2.$$

Решение уравнения правдоподобия имеет вид

$$\tilde{a} = \frac{1}{n} \sum_{k=1}^n \ln \xi_k = \overline{\ln \xi}, \quad \tilde{\theta} = \frac{1}{n} \sum_{k=1}^n (\ln \xi_k - \overline{\ln \xi})^2.$$

Представим себе, что плотность распределения величины ξ_k выражена как функция от параметров m и D и нас интересует оценка этих параметров. Вследствие свойства г) метода максимального правдоподобия, оценки \tilde{m} и \tilde{D} математического ожидания и дисперсии величины ξ_k имеют вид:

$$\tilde{m} = \exp\left(\tilde{a} + \frac{\tilde{\theta}}{2}\right) = \left(\prod_{k=1}^n \xi_k\right)^{\frac{1}{n}} \exp\left\{\frac{1}{2n} \sum_{k=1}^n (\ln \xi_k - \overline{\ln \xi})\right\},$$

$$\tilde{D} = \tilde{m}^2 \left(\exp\left\{\frac{1}{n} \sum_{k=1}^n (\ln \xi_k - \overline{\ln \xi})^2\right\} - 1 \right).$$

Из теорем, которые приводятся ниже, вытекает, что эти оценки асимптотически эффективны.

Асимптотические свойства оценок максимального правдоподобия. Предположим, что оценка параметра θ распределения $g(x, \theta)$ производится по выборке объема n . Рассмотрим асимптотическое поведение оценок $\tilde{\theta} = \tilde{\theta}_n$ параметра θ , полученных по методу максимального правдоподобия. Покажем, что при довольно широких предположениях оценки максимального правдоподобия состоятельны, а при несколько более жестких — асимптотически эффективны и асимптотически нормальны. Сделаем несколько предварительных замечаний о функции правдоподобия.

Если $\xi_1, \xi_2, \dots, \xi_n$ — компоненты выборочного вектора ξ , то функция правдоподобия имеет вид

$$\Lambda(\theta) = \prod_{k=1}^n g(\xi_k, \theta).$$

Предположим, что область значений Θ параметра θ компактна и истинное значение θ параметра лежит внутри Θ . Рассмотрим отношение $\frac{\Lambda(\lambda)}{\Lambda(\theta)}$, $\lambda \in \Theta$. Так как функция $\ln x$, $x > 0$, строго выпукла, то из неравенства Йенсена следует

$$M_\theta \ln \frac{\Lambda(\lambda)}{\Lambda(\theta)} \leq \ln M_\theta \frac{\Lambda(\lambda)}{\Lambda(\theta)} = 0, \quad (2)$$

так как

$$M_0 \frac{\Lambda(\lambda)}{\Lambda(\theta)} = \int \prod_{k=1}^n g(x_k, \lambda) dx_1 \dots dx_n = 1.$$

При этом, если $\Lambda(\lambda) \neq \Lambda(\theta)$ с положительной вероятностью, то

$$M_0 \ln \frac{\Lambda(\lambda)}{\Lambda(\theta)} < 0 \text{ и } M_0 \ln \Lambda(\lambda) < M_0 \ln \Lambda(\theta)$$

с вероятностью 1, если

$$I(\lambda) = M_0 \ln g(\xi_k, \lambda) = \int \ln g(x, \lambda) g(x, \theta) dx \neq \infty.$$

Перейдем к доказательству состоятельности оценок метода максимального правдоподобия. Положим

$$I(\theta, V_\lambda) = M_0 \left\{ \inf_{\lambda' \in V_\lambda} \ln \frac{g(\xi, \theta)}{g(\xi, \lambda')} \right\},$$

где V_λ — некоторая окрестность точки $\lambda \in \Theta$. Предположим, что выполнены условия:

1) если $\theta \neq \lambda$, то $g(x, \theta) \neq g(x, \lambda)$ на множестве положительной лебеговой меры;

2) функция $g(x, \theta)$ непрерывна по θ ;

3) для каждой точки $\lambda \in \Theta$, $\lambda \neq \theta$ существует окрестность V_λ такая, что $I(\theta, V_\lambda) > -\infty$.

Теорема 1. Если выполнены условия 1), 2) и 3), то оценки максимального правдоподобия сильно состоятельны.

Доказательство. Пусть $V_\theta = \{\theta' : |\theta' - \theta| < \varepsilon\}$, где $\varepsilon > 0$ произвольно заданное малое число, V_θ — дополнение к V_θ . Покажем, что

$$p_n = P_\theta \left\{ \frac{\prod_{k=1}^n g(\xi_k, \theta)}{\sup_{\lambda \in V_\theta} \prod_{k=1}^n g(\xi_k, \lambda)} > 1 \right\} \rightarrow 1 \text{ при } n \rightarrow \infty.$$

Это неравенство показывает, что максимум функции $\prod_{k=1}^n g(\xi_k, \lambda)$ при достаточно большом n достигается в окрестности V_θ с вероятностью, большей $1 - \varepsilon$, т. е. что произвольная последовательность оценок $\hat{\theta}_n$ максимального правдоподобия сходится по вероятности к θ .

Положим для любого $\mathcal{D} \subset \Theta$

$$\eta_k(\mathcal{D}) = \ln \frac{g(\xi_k, \theta)}{\sup_{\lambda \in \mathcal{D}} g(\xi_k, \lambda)}.$$

$$\frac{\prod_{k=1}^n g(\xi_k, \theta)}{\sup_{\lambda \in \bar{U}_\theta} \prod_{k=1}^n g(\xi_k, \lambda)} = \prod_{k=1}^n \frac{g(\xi_k, \theta)}{\sup_{\lambda \in \bar{U}_\theta} g(\xi_k, \lambda)},$$

то

$$\rho_n \geq P_\theta \left\{ \inf_{\lambda \in \bar{U}_\theta} \sum_{k=1}^n \ln \frac{g(\xi_k, \theta)}{g(\xi_k, \lambda)} > 0 \right\}.$$

Пусть $U_\lambda^{(n)}$ — убывающая последовательность окрестностей точки $\lambda \in \Theta$, стягивающая в точку λ_0 . Тогда $\eta_k(U_\lambda^{(n)})$ — монотонно возрастающая последовательность случайных величин и вследствие непрерывности по λ функции $g(x, \lambda)$ $\eta_k(U_\lambda^{(n)}) \rightarrow \ln \frac{g(\xi_k, \theta)}{g(\xi_k, \lambda)}$ с вероятностью 1, а в силу теоремы Лебега имеем

$$M_\theta \eta_k(U_\lambda^{(n)}) \rightarrow M_\theta \ln \frac{g(\xi_k, \theta)}{g(\xi_k, \lambda)} = I(\theta, \lambda).$$

Функцию $I(\theta, \lambda)$ иногда называют информационной функцией Кульбака — Ляйблера. Заметим, что согласно неравенству Иенсена

$$\begin{aligned} I(\theta, \lambda) &= M_\theta \ln \frac{g(\xi_k, \theta)}{g(\xi_k, \lambda)} = -M_\theta \ln \frac{g(\xi_k, \lambda)}{g(\xi_k, \theta)} \geq \\ &\geq -\ln M_\theta \frac{g(\xi_k, \lambda)}{g(\xi_k, \theta)} = -\ln \int_R g(x, \lambda) dx, \end{aligned}$$

т. е.

$$I(\theta, \lambda) \geq 0,$$

причем $I(\theta, \lambda) = 0$ тогда и только тогда, когда функция $\frac{g(\xi_k, \lambda)}{g(\xi_k, \theta)}$ (как функция от ω) с вероятностью 1 константа. Вследствие условия 1), $I(\theta, \lambda) > 0$ для всех $\lambda \neq \theta$. Следовательно, для каждой точки λ найдется такая окрестность U_λ , что $M_\theta \eta_k(U_\lambda) > 0$ для $k = 1, \dots, n$. Из этих окрестностей выберем конечное покрытие $U_{\lambda_1}, \dots, U_{\lambda_m}$ множества \bar{U}_θ . Тогда

$$\begin{aligned} \inf_{\lambda \in \bar{U}_\theta} \sum_{k=1}^n \ln \frac{g(\xi_k, \theta)}{g(\xi_k, \lambda)} &\geq \min_{1 \leq i \leq m} \inf_{\lambda \in U_{\lambda_i}} \sum_{k=1}^n \ln \frac{g(\xi_k, \theta)}{g(\xi_k, \lambda)} \geq \\ &\geq \min_{1 \leq i \leq m} \sum_{k=1}^n \eta_k(U_{\lambda_i}). \end{aligned}$$

Поэтому

$$\rho_n = P_\theta \left(\min_{1 \leq i \leq m} \sum_{k=1}^n \eta_k(U_{\lambda_i}) > 0 \right).$$

По усиленному закону больших чисел

$$\min_j \frac{1}{n} \sum_{k=1}^n (\eta_k(U_{\lambda_j})) \rightarrow \min_j M \eta_k(U_{\lambda_j}) > 0$$

с вероятностью 1. Следовательно, $p_n \rightarrow 1$, и, как отмечалось ранее, любая последовательность оценок максимального правдоподобия состоятельна. Более того, событие

$$A_\varepsilon = \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ \frac{\prod_{k=1}^n g(\xi_k, \theta)}{\sup_{\lambda \in \bar{U}_\theta} \prod_{k=1}^n g(\xi_k, \lambda)} > 1 \right\}$$

означает, что найдется такой номер N , что все оценки максимального правдоподобия $\hat{\theta}_n$ при $n \geq N$ лежат в ε -окрестности точки θ . Имеем

$$P(A_\varepsilon) \geq P_0 \left(\liminf_{N \rightarrow \infty} \min_{n \geq N} \inf_{1 \leq l \leq m} \frac{1}{n} \sum_{k=1}^n \eta_k(U_{\lambda_l}) > 0 \right) = 1$$

на основании тех же соображений, которые использованы при предыдущем доказательстве. Таким образом, любая последовательность оценок максимального правдоподобия $\hat{\theta}_n$ с вероятностью 1 сходится к θ . Итак, оценки максимального правдоподобия сильно состоятельны.

При более сильных предположениях о регулярности функции $\ln g(x, \theta)$ можно показать, что последовательность оценок максимального правдоподобия $\hat{\theta}_n$ при $n \rightarrow \infty$ асимптотически нормальна и асимптотически эффективна.

Рассмотрим случай одномерного параметра.

Теорема 2. Пусть θ — внутренняя точка Θ и выполнены условия:

а) существует производная $\frac{\partial^k}{\partial \theta^k} g(x, \theta)$, $k = 1, 2, 3$, причем

$$M \left| \frac{\partial}{\partial \theta} g(x, \theta) \right| + M \left| \frac{\partial^2}{\partial \theta^2} g(x, \theta) \right| < \infty;$$

б) $\left| \frac{\partial^3}{\partial \theta^3} g(x, \theta) \right| \leq H(x)$, где $H(x)$ не зависит от θ и $MH(\xi_k) < \infty$;

в) $0 < M \left(\frac{\partial}{\partial \theta} \ln g(\xi, \theta) \right)^2 < \infty$.

Тогда произвольная сильно состоятельная последовательность $\hat{\theta}_n$ оценок максимального правдоподобия асимптотически нормальна и асимптотически эффективна.

Доказательство. Прежде всего заметим, что по предыдущей теореме $\check{\theta}_n \rightarrow \theta$ с вероятностью 1. Воспользовавшись разложением Тейлора

$$\sum_{k=1}^n \frac{\partial}{\partial \theta} \ln g(\xi_k, \check{\theta}) = \sum_{k=1}^n \left[\frac{\partial}{\partial \theta} \ln g(\xi_k, \theta) + \frac{\partial^2}{\partial \theta^2} \ln g(\xi_k, \theta) (\check{\theta}_n - \theta) + \frac{1}{2!} \frac{\partial^3}{\partial \theta^3} \ln g(\xi_k, \theta_n^*) (\check{\theta}_n - \theta)^2 \right],$$

где θ_n^* — точка, лежащая между θ и $\check{\theta}_n$, и замечая, что для оценки $\check{\theta}_n$ максимального правдоподобия $\sum_{k=1}^n \frac{\partial}{\partial \theta} \ln g(\xi_k, \check{\theta}_n) = 0$, получим

$$\alpha_n + \beta_n \sqrt{n} (\check{\theta}_n - \theta) + \gamma_n \sqrt{n} (\check{\theta}_n - \theta)^2 = 0,$$

где

$$\alpha_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{\partial}{\partial \theta} \ln g(\xi_k, \theta), \quad \beta_n = \frac{1}{n} \sum_{k=1}^n \frac{\partial^2}{\partial \theta^2} \ln g(\xi_k, \theta),$$

$$\gamma_n = \frac{1}{2n} \sum_{k=1}^n \frac{\partial^3}{\partial \theta^3} \ln g(\xi_k, \theta_n^*).$$

Таким образом,

$$\sqrt{n} (\check{\theta}_n - \theta) = \frac{\alpha_n}{-\beta_n - \gamma_n (\check{\theta}_n - \theta)}.$$

Величина $\sqrt{n} \alpha_n$ является суммой независимых одинаково распределенных величин с конечными моментами второго порядка. При этом

$$M \frac{\partial}{\partial \theta} \ln g(\xi_k, \theta) = 0, \quad M \left(\frac{\partial}{\partial \theta} \ln g(\xi_k, \theta) \right)^2 = I(\theta) > 0.$$

Следовательно, величина α_n асимптотически нормальна $(0, I(\theta))$. По закону больших чисел величина β_n сходится с вероятностью 1 к величине $M \frac{\partial^2}{\partial \theta^2} \ln g(\xi_k, \theta) = I(\theta)$ согласно лемме 1. Что касается величины γ_n , то

$$|\gamma_n| \leq \frac{1}{2n} \sum_{k=1}^n H(\xi_k).$$

Из закона больших чисел вытекает, что величина справа сходится к конечному пределу. Поэтому величина γ_n стохастически ограничена. Так как $\check{\theta}_n - \theta$ сходится по вероятности к 0, то $\gamma_n (\check{\theta}_n - \theta) \rightarrow 0$ по вероятности. Следовательно, распределение величины

$\frac{\alpha_n}{-\beta_n - \gamma_n(\theta_n - \theta)}$ сходится к нормальному закону со средним 0 и дисперсией $I^{-1}(\theta)$. Итак, величина $\sqrt{n}(\hat{\theta}_n - \theta)$ асимптотически нормальна $(0, I^{-1}(\theta))$.

Далее, если рассмотреть отношение средней квадратической ошибки этой оценки к нижней границе Крамера — Рао, то получим $\frac{M(\hat{\theta}_n - \theta)^2}{\frac{1}{nI(\theta)}} \rightarrow 1$ при $n \rightarrow \infty$, т. е. оценка $\hat{\theta}_n$ асимптотически эффективна.

ПРИМЕРЫ

1. Рассмотрим выборку $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, в которой величины ξ_k , $k = 1, \dots, n$, имеют Γ -распределение с плотностью

$$g(x, \lambda) = \frac{1}{\Gamma(\lambda)} x^{\lambda-1} e^{-x}, \quad x > 0,$$

и неизвестным значением параметра λ . Для логарифма функции правдоподобия получаем следующее выражение:

$$\begin{aligned} \ln \Lambda(\lambda) &= \sum_{k=1}^n \ln g(\xi_k, \lambda) = -n \ln \Gamma(\lambda) + \\ &+ (\lambda - 1) \sum_{k=1}^n \ln \xi_k - \sum_{k=1}^n \xi_k, \end{aligned}$$

поэтому

$$\begin{aligned} \frac{\partial}{\partial \lambda} \ln \Lambda(\lambda) &= -n \frac{d}{d\lambda} \ln \Gamma(\lambda) + \sum_{k=1}^n \ln \xi_k, \\ \frac{\partial^2}{\partial \lambda^2} \ln \Lambda(\lambda) &= -n \frac{d^2}{d\lambda^2} \ln \Gamma(\lambda). \end{aligned} \quad (2)$$

Оценкой максимального правдоподобия $\tilde{\lambda}$ параметра λ является единственный корень уравнения

$$\frac{1}{n} \sum_{k=1}^n \ln \xi_k = \frac{d}{d\lambda} \ln \Gamma(\tilde{\lambda}). \quad (3)$$

При этом по теореме 2 величина $\tilde{\lambda}$ асимптотически эффективна и асимптотически нормальна со средним λ и дисперсией

$$\begin{aligned} \frac{1}{nI(\tilde{\theta})} &= \left[nM \left(\frac{\partial}{\partial \lambda} \ln g(\xi_k, \lambda) \right)^2 \right]^{-1} = \left(-nM \frac{\partial^2}{\partial \lambda^2} \ln g(\xi_k, \lambda) \right)^{-1} = \\ &= \left[-n \frac{d^2}{d\lambda^2} \ln \Gamma(\lambda) \right]^{-1}. \end{aligned}$$

Величина $\tilde{\lambda}$, являющаяся решением уравнения (3), трудно вычислима. Поскольку

$$M\xi_k = \frac{1}{\Gamma(\lambda)} \int_0^\infty x^\lambda e^{-x} dx = \frac{\Gamma(\lambda+1)}{\Gamma(\lambda)} = \lambda,$$

то, в соответствии с методом эмпирического распределения, для оценки λ можно предположить более простую оценку $\bar{\lambda}$

$$\bar{\lambda} = \frac{1}{n} \sum_{k=1}^n \xi_k.$$

Для дисперсии этой оценки имеем следующее выражение:

$$D(\bar{\lambda}) = \frac{1}{n} D(\xi_k) = \frac{1}{n} (M\xi_k^2 - (M\xi_k)^2),$$

причем

$$M\xi_k^2 = \frac{1}{\Gamma(\lambda)} \int_0^{\infty} x^{\lambda+1} e^{-x} dx = \frac{\Gamma(\lambda+2)}{\Gamma(\lambda)} = \lambda(\lambda+1),$$

откуда

$$D(\bar{\lambda}) = \frac{\lambda}{n}.$$

С одной стороны, оценка $\bar{\lambda}$ параметра λ не является эффективной, а с другой стороны, отношение $\frac{1}{nI(\theta)} : D(\bar{\lambda}) \left[-\lambda \frac{d^2}{d\lambda^2} \ln \Gamma(x) \right]^{-1}$ не зависит от n и оценка $\bar{\lambda}$ не может быть и асимптотически эффективной.

2. Рассмотрим опять выборку из генеральной совокупности с Γ -распределением. Введем масштабный параметр α , а параметр λ будем считать известным. Плотность распределения равна:

$$g(x, \lambda, \alpha) = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, \quad x > 0, \quad \alpha > 0.$$

Имеем:

$$\ln \Lambda(\alpha) = n(\lambda \ln \alpha - \ln \Gamma(\lambda)) + (\lambda - 1) \sum_{k=1}^n \ln \xi_k - \alpha \sum_{k=1}^n \xi_k,$$

$$\frac{\partial \ln \Lambda(\alpha)}{\partial \alpha} = \frac{n\lambda}{\alpha} - \sum_{k=1}^n \xi_k, \quad \frac{\partial^2 \ln \Lambda(\alpha)}{\partial \alpha^2} = -\frac{n\lambda}{\alpha^2}.$$

Оценка максимального правдоподобия $\check{\alpha}$ параметра α имеет вид

$$\check{\alpha} = \frac{\lambda}{\bar{\xi}}, \quad \bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k.$$

Заметим, что величина $n\check{\xi} = \xi_1 + \dots + \xi_n$ имеет плотность распределения $g(x, n\lambda, \alpha)$ (§ 8, гл. II). Следовательно,

$$M\check{\alpha} = \frac{n\lambda\alpha^{n\lambda}}{\Gamma(n\lambda)} \int_0^{\infty} x^{n\lambda-2} e^{-\alpha x} dx = \frac{n\lambda\alpha^{n\lambda}}{\Gamma(n\lambda)} : \frac{\alpha^{n\lambda-1}}{\Gamma(n\lambda-1)} = \frac{n\lambda\alpha}{n\lambda-1}.$$

Аналогично

$$M\check{\alpha}^2 = \frac{(n\lambda)^2 \alpha^{n\lambda}}{\Gamma(n\lambda)} : \frac{\alpha^{n\lambda-2}}{\Gamma(n\lambda-2)} = \frac{(n\lambda\alpha)^2}{(n\lambda-1)(n\lambda-2)},$$

$$D\check{\alpha} = \left(\frac{n\lambda\alpha}{n\lambda-1} \right)^2 \frac{1}{n\lambda-2}.$$

Оценка α смещена, но ее легко преобразовать в несмещенную. Положив

$$\tilde{\alpha} = \frac{n\lambda - 1}{n\lambda} \alpha,$$

получим

$$M\tilde{\alpha} = \alpha, \quad D\tilde{\alpha} = \frac{\alpha^2}{n\lambda - 2}.$$

Таким образом, оценка $\tilde{\alpha}$ несмещена. Согласно общей теории, она, с одной стороны, одновременно с $\tilde{\alpha}$ асимптотически эффективна и асимптотически нормальна. С другой стороны, нижняя граница Крамера — Рао несмещенных оценок равна:

$$\frac{1}{nI(\theta)} = \frac{1}{M \frac{\partial^2 \ln \Lambda}{\partial \alpha^2}} = \frac{\alpha^2}{n\lambda}$$

и выражение для $D\tilde{\alpha}$ позволяет непосредственно убедиться в асимптотической эффективности оценки $\tilde{\alpha}$.

3. Пусть $\xi = (\xi'_1, \dots, \xi'_n)$, где $\xi'_k = (\xi_k, \eta_k)$ независимые двумерные гауссовские векторы с плотностью распределения

$$g(x, y) = [4\pi^2\theta_1\theta_2(1-\rho^2)]^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \times \right. \\ \left. \times \left[\frac{(x-a_1)^2}{\theta_1} - 2\rho \frac{x-a_1}{\sqrt{\theta_1}} \frac{y-a_2}{\sqrt{\theta_2}} + \frac{(y-a_2)^2}{\theta_2} \right] \right\}.$$

Логарифм функции правдоподобия равен

$$\ln f(\xi, a_1, a_2, \theta_1, \theta_2, \rho) = -\frac{n}{2} \ln 4\pi^2\theta_1\theta_2(1-\rho^2) - \frac{1}{2(1-\rho^2)} - \\ - \sum_k \left[\frac{(\xi_k - a_1)^2}{\theta_1} - 2\rho \frac{(\xi_k - a_1)(\eta_k - a_2)}{\sqrt{\theta_1}\sqrt{\theta_2}} + \frac{(\eta_k - a_2)^2}{\theta_2} \right].$$

Пять уравнений максимального правдоподобия имеют вид

$$\frac{\bar{\xi} - a_1}{\theta_1} - \rho \frac{\bar{\eta} - a_2}{\sqrt{\theta_1\theta_2}} = 0, \quad (I)$$

$$\bar{\xi} = \frac{1}{n} \sum_1^n \xi_k, \quad \bar{\eta} = \frac{1}{n} \sum_1^n \eta_k, \quad \frac{\bar{\eta} - a_2}{\theta_2} - \rho \frac{\bar{\xi} - a_1}{\sqrt{\theta_1\theta_2}} = 0, \quad (II)$$

$$n(1-\rho^2) = \sum_k \left[\frac{(\xi_k - a_1)^2}{\theta_1} - \rho \frac{(\xi_k - a_1)(\eta_k - a_2)}{\sqrt{\theta_1\theta_2}} \right] = 0, \quad (III)$$

$$n(1-\rho^2) = \sum_k \left[-\rho \frac{(\xi_k - a_1)(\eta_k - a_2)}{\sqrt{\theta_1\theta_2}} + \frac{(\eta_k - a_2)^2}{\theta_2} \right] = 0, \quad (IV)$$

$$n(1-\rho^2) = \sum_k \left[\frac{(\xi_k - a_1)^2}{\theta_1} - (1+\rho^2) \frac{(\xi_k - a_1)(\eta_k - a_2)}{\sqrt{\theta_1\theta_2}} + \frac{(\eta_k - a_2)^2}{\theta_2} \right] = 0. \quad (V)$$

Если параметры a_1, a_2, σ_1 и σ_2 известны и подлежит оценке коэффициент корреляции ρ между ξ_k и η_k , то следует решать только уравнение (V). Оно является кубическим уравнением и имеет вид

$$-n\check{\rho}^3 + (\theta_1\theta_2)^{-\frac{1}{2}} \left(\sum_k (\xi_k - a_1)(\eta_k - a_2) \right) (\check{\rho}^2 + 1) + \\ + \left(n - \sum_k \frac{(\xi_k - a_1)^2}{\theta_1} + \frac{(\eta_k - a_2)^2}{\theta_2} \right) \check{\rho} = 0.$$

Если это уравнение имеет три действительных корня, то нужно выбрать тот из них, для которого функция правдоподобия имеет большое значение. Складывая уравнения (III) и (IV) и вычитая (V), получим

$$n(1 - \check{\rho}^2) = \frac{1 - \check{\rho}^2}{\check{\rho}} \sum_k \frac{(\xi_k - a_1)(\eta_k - a_2)}{\sqrt{\check{\theta}_1 \check{\theta}_2}}, \\ \check{\rho} = \frac{\frac{1}{n} \sum_k (\xi_k - a_1)(\eta_k - a_2)}{\sqrt{\check{\theta}_1 \check{\theta}_2}}.$$

Подставляя полученные значения $\check{\rho}$ в уравнения (III) и (IV) (вместо ρ), получим

$$\check{\theta}_1 = \frac{1}{n} \sum_k (\xi_k - a_1)^2, \quad \check{\theta}_2 = \frac{1}{n} \sum_{k=1}^n (\eta_k - a_2)^2$$

и

$$\check{\rho} = \frac{\sum_k (\xi_k - a_1)(\eta_k - a_2)}{\sqrt{\sum_k (\xi_k - a_1)^2 \sum_k (\eta_k - a_2)^2}}.$$

Если требуется оценить все пять параметров распределения, то к уравнениям (III), (IV), (V) нужно еще присоединить уравнения (I) и (II). Решив эти уравнения относительно a_1 и a_2 , полученные значения \check{a}_1 и \check{a}_2 следует подставить вместо a_1 и a_2 в ранее полученные выражения для $\check{\theta}_1, \check{\theta}_2, \rho$. Из уравнения (I) и (II) имеем

$$\check{a}_1 = \bar{\xi}, \quad \check{a}_2 = \bar{\eta},$$

так что для величин $\check{\theta}_1 = \check{\sigma}_1^2, \check{\theta}_2 = \check{\sigma}_2^2, \check{\rho}$ в этом случае получаем следующие формулы:

$$\check{\theta}_1 = \frac{1}{n} \sum_k (\xi_k - \bar{\xi})^2, \quad \check{\theta}_2 = \frac{1}{n} \sum_{k=1}^n (\eta_k - \bar{\eta})^2, \\ \check{\rho} = \frac{\sum_k (\xi_k - \bar{\xi})(\eta_k - \bar{\eta})}{\sqrt{\sum_k (\xi_k - \bar{\xi})^2 \sum_k (\eta_k - \bar{\eta})^2}}.$$

Все полученные оценки асимптотически эффективны.

Найти асимптотическое выражение корреляционной матрицы оценок можно, вычислив информационную матрицу. Так, для корреляционной матрицы параметров $\theta_1, \theta_2, \theta_3$ ($\theta_3 = \rho$) получаем следующее выражение:

$$I_3(\theta) = - \left(M \frac{\partial^2 \ln f}{\partial \theta_i \partial \theta_k} \right) = \frac{n}{1 - \rho^2} \begin{pmatrix} \frac{2 - \rho^2}{4\sigma_1^4} & \frac{-\rho^2}{4\sigma_1^2\sigma_2^2} & \frac{-\rho}{2\sigma_1^3} \\ \frac{-\rho^2}{4\sigma_1^2\sigma_2^2} & \frac{2 - \rho^2}{4\sigma_2^4} & \frac{-\rho}{2\sigma_2^3} \\ \frac{-\rho}{2\sigma_1^3} & \frac{-\rho}{2\sigma_2^3} & \frac{1 + \rho^2}{1 - \rho^2} \end{pmatrix}.$$

Обратная матрица $I_3^{-1}(\theta)$ дает асимптотическую корреляционную матрицу оценок $\check{\sigma}_1^2, \check{\sigma}_2^2, \check{\rho}$,

$$I_3^{-1}(\theta) = \frac{1}{n} \begin{pmatrix} 2\sigma_1^4 & 2\rho^2\sigma_1^2\sigma_2^2 & \rho(1 - \rho^2)\sigma_1^2 \\ 2\rho^2\sigma_1^2\sigma_2^2 & 2\sigma_2^4 & \rho(1 - \rho^2)\sigma_2^2 \\ \rho(1 - \rho^2)\sigma_1^2 & \rho(1 - \rho^2)\sigma_2^2 & (1 - \rho^2)^2 \end{pmatrix}.$$

Информационная матрица $I_1(\theta)$ параметров a_1 и a_2 и ей обратная $I_1^{-1}(\theta)$ равны:

$$I_1(\theta) = \frac{n}{1 - \rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}, \quad I_1^{-1}(\theta) = \frac{1}{n} \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Для информационной матрицы $I(\theta)$ всей совокупности операторов и ей обратной имеем

$$I(\theta) = \begin{pmatrix} I_1(\theta) & 0 \\ 0 & I_2(\theta) \end{pmatrix}, \quad I^{-1}(\theta) = \begin{pmatrix} I_1^{-1}(\theta) & 0 \\ 0 & I_2^{-1}(\theta) \end{pmatrix}.$$

Матрица $I^{-1}(\theta)$ дает асимптотическое значение корреляционной матрицы статистик $(\bar{\xi}, \bar{\eta}, \check{\sigma}_1^2, \check{\sigma}_2^2, \check{\eta})$. Вычисления, приводящие к указанным формулам, несложны и представляются читателю.

4. Случайные величины $\xi_k, k = 1, 2, \dots, n$ независимы и равномерно распределены на интервале $(\theta - l, \theta + l)$. Значение l известно. Требуется оценить параметр θ . Функция правдоподобия имеет вид

$$l(\zeta, \theta) = \left(\frac{1}{2l} \right)^n I(\theta - l - \max_k \xi_k) I(\min_k \xi_k - \theta + l),$$

где $I(x) = 1$ при $x > 0$, $I(x) = 0$ при $x \leq 0$. Величины $\alpha = \min_k \xi_k, \beta = \max_k \xi_k$ являются достаточными статистиками, а решениями уравнения максимального правдоподобия будет любая статистика $\check{\theta}(\zeta)$, такая, что $\beta - l \leq \check{\theta}(\zeta) \leq \alpha + l$.

С другой стороны, статистика $\bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k$ является несмещенной состоятельной оценкой параметра θ . Ее дисперсия

$$D\bar{\xi} = \frac{1}{n} D\xi_k = \frac{1}{n} \left\{ \frac{1}{2l} \int_{\theta-l}^{\theta+l} x^2 dx - \theta^2 \right\} = \frac{l^2}{3n}.$$

Сравним эту оценку с оценкой максимального правдоподобия

$$\hat{\theta} = \frac{\max_k \xi_k + \min_k \xi_k}{2} = \theta + \frac{\max_k (\xi_k - \theta) + \min_k (\xi_k - \theta)}{2}.$$

Из приведенного равенства вытекает, что величина $\frac{\hat{\theta} - \theta}{l}$ имеет распределение, не зависящее от θ и l . Поэтому при дальнейших вычислениях можно считать $\theta = 0, l = 1$. Функция распределения для величин $\beta = \max_k \xi_k$ имеет вид $F(x) = \left(\frac{x+1}{2}\right)^n$. Следовательно,

$$M\beta = \frac{n}{2^n} \int_{-1}^1 x(x+1)^{n-1} dx = \frac{n}{2^n} \int_{-1}^1 [(x+1)^n - (x+1)^{n-1}] dx = 1 - \frac{2}{n+1}.$$

Отсюда

$$M\alpha = -1 + \frac{2}{n+1}.$$

Поэтому оценка $\hat{\theta}$ — несмещенная, $M\hat{\theta} = 0$. Далее

$$M\beta^2 = \frac{n}{2} \int_{-1}^1 x^2 (x+1)^{n-1} dx = 1 - \frac{4n}{(n+2)(n+1)},$$

$$D\beta = \frac{4n}{(n+1)^2(n+2)}.$$

Очевидно, $M\alpha^2 = M\beta^2, D\alpha = D\beta$.

Легко указать совместную плотность распределения величины α и β . Имеем $(-1 < x < y < 1)$,

$$P(x < \alpha < x + dx, y < \beta < y + dy) = n(n-1) \frac{dx}{2} \frac{dy}{2} \left(\frac{y-x}{2}\right)^{n-2},$$

поэтому

$$M(\min_k \xi_k)(\max_k \xi_k) = \int_{-1}^1 \int_x^1 xy n(n-1) \left(\frac{y-x}{2}\right)^{n-2} \frac{dy}{2} \frac{dx}{2} =$$

$$= \frac{4}{n+1} \left(1 - \frac{1}{n+2}\right) - 1.$$

Отсюда

$$M(\alpha - M\alpha)(\beta - M\beta) = \frac{4}{(n+1)^2(n+2)},$$

$$D(\hat{\theta}) = D\left(\frac{\alpha + \beta}{2}\right) = \frac{2}{(n+1)(n+2)}.$$

Мы видим, что в рассматриваемом случае дисперсия несмещенной оценки $\hat{\theta}$ при больших n имеет порядок n^{-2} , в то время как в регулярных случаях (когда применимо неравенство Крамера — Рао) наилучшие оценки имеют дисперсию порядка n^{-1} . Возвращаясь к общему случаю равномерного распре-

ления на отрезке $(\theta - l, \theta + l)$, заметим, что в качестве оценок параметра θ можно было принять также оценки $\check{\theta}_1 = \beta - l$ или $\check{\theta}_2 = \alpha + l$. Эти оценки асимптотически несмещены и имеют дисперсию $\frac{4nl^2}{(n+1)^3(n+2)}$ примерно в два раза большую, чем дисперсия оценки $\check{\theta} = \frac{\alpha + \beta}{2}$.

§ 4. Доверительные интервалы

Понятие оценки параметра неполноценно, если не охарактеризованы величины возможных ошибок, возникающих при исполнении предлагаемых оценок. Наглядное представление о точности оценок дают доверительные интервалы или доверительные области в случае многомерного параметра. Для краткости в дальнейшем мы говорим, главным образом, о доверительных интервалах. Доверительные интервалы являются случайными интервалами, внутри которых с вероятностью γ , близкой к 1, содержится точное значение оцениваемого параметра. Величина γ называется доверительной вероятностью. Она выбирается исследователем в соответствии с готовностью мириться с возможностью ошибки. При заданном γ длина доверительного интервала характеризует точность, с которой локализовано значение параметра. Поэтому естественно стремление выбрать среди различных доверительных интервалов при заданной доверительной вероятности интервал кратчайшей длины. Такой выбор обеспечивает наиболее точную локализацию параметра. Перейдем к более точным определениям.

Пусть ξ — выборочный вектор со значениями в R^n , P_θ — распределение вектора ξ , зависящее от неизвестного параметра θ , $\check{\theta}(\xi)$ — некоторая оценка параметра ξ .

Одномерный параметр. Рассмотрим две действительные функции $\underline{\theta}(x)$, $\bar{\theta}(x)$ со значениями в R^1 , не зависящие от неизвестных параметров распределения вектора ξ и такие, что $\underline{\theta}(x) < \bar{\theta}(x)$, $x \in R^n$. Если

$$P_\theta \{ \underline{\theta}(\xi) < \theta < \bar{\theta}(\xi) \} \geq \gamma, \quad \forall \theta \in \Theta,$$

то случайный интервал $(\underline{\theta}(\xi), \bar{\theta}(\xi))$ называют γ -доверительным интервалом, число γ ($0 < \gamma < 1$) — доверительным уровнем, $\underline{\theta}(\xi)$ и $\bar{\theta}(\xi)$ — нижней и верхней доверительными границами.

В ряде случаев представляет интерес только нижняя или только верхняя доверительная граница для θ . Они соответствуют выбору $\bar{\theta}(x) = \infty$ или $\underline{\theta}(x) = -\infty$ соответственно. Такие интервалы называют верхним (соответственно нижним) доверительным интервалом. Доверительные интервалы имеют следующий смысл.

Будем пользоваться таким статистическим правилом:

если результаты выборки описываются вектором ξ , то неизвестное значение параметра θ лежит внутри интервала $(\underline{\theta}(\xi), \bar{\theta}(\xi))$. Это правило в некоторых случаях оказывается ошибочным.

Однако, если его применять многократно в разных, не зависящих друг от друга ситуациях (при которых, все же P_θ — распределение выборочного вектора ξ), то предложенное правило окажется ошибочным при γ , близком к 1, только в малом числе случаев, а именно, примерно в $100(1-\gamma)\%$ общего числа случаев применения этого правила.

Доверительные интервалы часто можно строить следующим образом. Предположим для простоты, что распределение статистики $\check{\theta}(\xi)$ непрерывно. Для каждого θ найдем такие величины $\theta_1 = \theta_1(\theta)$ и $\theta_2 = \theta_2(\theta)$, что $P_\theta(\theta_1 < \check{\theta}(\xi) < \theta_2) = \gamma$. Чтобы эта процедура была однозначной, часто выбирают θ_1 и θ_2 , исходя из требований

$$P_\theta(\check{\theta}(\xi) \leq \theta_1) = P_\theta(\check{\theta}(\xi) \geq \theta_2) = \frac{1}{2}(1-\gamma).$$

Обозначим через \mathcal{D} подмножество $\Theta \times \Theta$, $\mathcal{D} = \{(\theta, \theta') : \theta_1(\theta) < \theta_1(\theta') < \theta_2(\theta')\}$. При любом $\theta \in \Theta$ событие $\{(\theta, \check{\theta}(\xi)) \in \mathcal{D}\}$ имеет P_θ -вероятность γ . Рассмотрим теперь при фиксированном θ' множество $\mathcal{D}(\theta') = \{\theta : (\theta, \theta') \in \mathcal{D}\}$. Множество $\mathcal{D}(\check{\theta}(\xi))$ является случайным, и событие $\theta \in \mathcal{D}(\check{\theta}(\xi))$ происходит тогда и только тогда, когда $\check{\theta}(\xi) \in (\theta_1(\theta), \theta_2(\theta))$ и, следовательно, при каждом θ имеет вероятность γ . Если множество $\mathcal{D}(\theta')$ является интервалом ($\forall \theta \in \Theta$), то $\mathcal{D}(\check{\theta}(\xi))$ является γ -доверительным интервалом для параметра θ .

Ясно, что существуют различные способы построения доверительных интервалов. Например, в предыдущем построении они зависели от выбора оценки $\check{\theta}(\xi)$. Естественно желание среди различных доверительных интервалов, соответствующих данному доверительному уровню, выбирать как можно более короткие. Предположим, что рассматриваются несмещенные и приблизительно нормальные оценки θ . Тогда введенные выше интервалы $\mathcal{D}(\check{\theta}(\xi))$ будут тем короче, чем меньше дисперсия оценки. Таким образом, эффективные и асимптотически эффективные оценки приводят к кратчайшим или асимптотически кратчайшим доверительным интервалам.

В случае, когда выборка построена по большому числу независимых одинаково распределенных случайных величин, для приближенного построения доверительных интервалов можно предложить следующий способ. Как уже ранее указывалось, в случае регулярности плотности распределения $f(x, \theta)$ выборочного вектора ξ величина $\frac{\partial}{\partial \theta} \ln f(\xi, \theta)$ распределена асимптотически нормально. Если

$$\eta = \frac{\frac{\partial}{\partial \theta} \ln f(\xi, \theta)}{\sqrt{D\left(\frac{\partial}{\partial \theta} \ln f(\xi, \theta)\right)}},$$

то величина η асимптотически нормальна $(0,1)$. Пусть c_γ определяется из условия $P(|\xi| < c_\gamma) = \gamma$, где ξ — стандартная гауссовская величина. Если величина $c_\gamma = c_\gamma(\theta)$ является монотонной функцией от θ , то, разрешая неравенство $|\eta| \leq c_\gamma$ относительно θ , мы для θ получаем γ -доверительный интервал.

ПРИМЕР

Пусть величины ξ_k , $k = 1, \dots, n$ имеют распределение Пуассона $P(\xi_k = k) = e^{-\theta} \frac{\theta^k}{k!}$. Функция правдоподобия для выборочного вектора $\zeta = (\xi_1, \xi_2, \dots, \xi_n)$ такова:

$$f(\zeta, \theta) = e^{-n\theta} \prod_{k=1}^n \frac{\theta^{\xi_k}}{\xi_k!}.$$

Следовательно, $\left(\bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k\right)$,

$$\frac{\partial}{\partial \theta} \ln f(\zeta, \theta) = \frac{n}{\theta} (\bar{\xi} - \theta), \quad \frac{\partial^2}{\partial \theta^2} \ln f(\zeta, \theta) = -\frac{n}{\theta^2} \bar{\xi}$$

и

$$M\left(\frac{\partial}{\partial \theta} \ln f(\zeta, \theta)\right)^2 = -M \frac{\partial^2}{\partial \theta^2} \ln f(\zeta, \theta) = \frac{n}{\theta}.$$

Таким образом, величина

$$\eta = \sqrt{\frac{n}{\theta}} (\bar{\xi} - \theta)$$

асимптотически нормальна $(0,1)$.

Последнее обстоятельство очевидно и без ссылки на асимптотическое свойство оценок максимального правдоподобия.

Определим, пользуясь нормальным законом распределения, приближенные γ -доверительные границы $\pm t_\gamma$ для величины η . Тогда

$$\sqrt{\frac{n}{\theta}} |\bar{\xi} - \theta| < t_\gamma \quad (1)$$

с доверительной вероятностью γ . Положим $\theta = z^2$, $z > 0$. Решая уравнение (1) относительно z , получим

$$\sqrt{\bar{\xi} + \frac{t_\gamma^2}{4n}} - \frac{t_\gamma}{2\sqrt{n}} < z < \sqrt{\bar{\xi} + \frac{t_\gamma^2}{4n}} + \frac{t_\gamma}{2\sqrt{n}},$$

или

$$\bar{\xi} + \frac{t_\gamma^2}{2n} - t_\gamma \sqrt{\frac{\bar{\xi}}{n} + \frac{t_\gamma^2}{4n^2}} < \theta < \bar{\xi} + \frac{t_\gamma^2}{2n} + t_\gamma \sqrt{\frac{\bar{\xi}}{n} + \frac{t_\gamma^2}{4n^2}}.$$

Из ранее сделанных замечаний следует, что построенный доверительный интервал является асимптотически кратчайшим.

Многомерный параметр. Приведенные выше определения и метод построения доверительных интервалов обобщается на случай многомерного параметра. Пусть выборочный вектор ζ

имеет распределение P_θ , зависящее от d -мерного параметра $\theta = (\theta_1, \theta_2, \dots, \theta_d)$, $\theta \in \Theta$, $\Theta \subset R^d$. Задача может состоять или в совместной оценке всех компонент вектора θ , или в оценке одной из них, скажем θ_1 , с помощью доверительного интервала, не зависящего от остальных параметров.

В первом случае вместо доверительных интервалов определяются доверительные области (или зоны) \mathcal{G} с помощью условия $P_\theta(\theta \in \mathcal{G}) \geq \gamma$ для всех $\theta \in \Theta$.

Пусть $\check{\theta}(\xi)$ — некоторая выборочная статистика. Предположим, что она имеет распределение непрерывного типа. Найдем для данного γ , $0 < \gamma < 1$, такой параллелепипед J_γ , чтобы $P_\theta(\theta(\xi) \in J_\gamma) < \gamma$. Определим в R^{2d} следующее множество \mathcal{D} : точка $(\theta, \theta') \in \mathcal{D}$ тогда и только тогда, когда $\theta' \in J_\gamma$, $\theta \in \Theta$. Рассмотрим сечение множества: $\mathcal{D}(\theta') = \{\theta : (\theta, \theta') \in \mathcal{D}\}$. При любом θ событие $\theta \in \mathcal{D}(\check{\theta}(\xi))$ происходит тогда и только тогда, когда $\check{\theta}(\xi) \in J_\gamma$ и, следовательно, имеет вероятность γ . Таким образом, случайное множество $\mathcal{D}(\check{\theta}(\xi))$ является γ -доверительной областью для параметра θ . Это означает, что на практике целесообразно систематически пользоваться правилом $\theta \in \mathcal{D}(\check{\theta}(\xi))$, причем, это правило редко окажется ошибочным: каждый раз при его применении вероятность, что оно будет ошибочным, равна $1 - \gamma$.

Случай, когда оценивается одна компонента параметрического вектора, а остальные компоненты неизвестны, рассматривается аналогично, но построения более сложные.

Доверительные интервалы для параметров нормального закона. Оценки параметров нормального закона были получены раньше. Найдем теперь для них доверительные интервалы. Рассмотрим разные варианты этой задачи.

а) Доверительные интервалы для математического ожидания нормального распределения при известной дисперсии были указаны ранее. Если $\bar{\xi}$ — выборочное среднее из выборки объема n , то величина $\frac{\sqrt{n}}{\sigma}(\bar{\xi} - a)$ нормальна $(0, 1)$. Поэтому, если $\pm c_\gamma$ — доверительные границы для стандартного нормального распределения, соответствующие доверительному уровню γ , то интервал $(\bar{\xi} - c_\gamma \frac{\sigma}{\sqrt{n}}, \bar{\xi} + c_\gamma \frac{\sigma}{\sqrt{n}})$ является γ -доверительным интервалом для a .

Можно строить доверительный интервал для a и с помощью других оценок. Например, если в качестве оценки a принять выборочную медиану $\xi_{\frac{1}{2}}$, то по теореме 2 величина $2f(a)\sqrt{n}(\xi_{\frac{1}{2}} - a)$ асимптотически нормальна $(0, 1)$. Здесь $f(x)$ — плотность распределения величин ξ_k , $f(a) = (2\pi\sigma^2)^{-\frac{1}{2}}$. Поэтому при боль-

ших n доверительные границы для среднего a имеют вид $\bar{\xi}_1 \pm \pm kc_1 \frac{\sigma}{\sqrt{n}}$, где c_1 принимает прежнее значение и $k = \frac{\sqrt{2\pi}}{2} = 1,2533$. Построенные интервалы оказались в k раз длиннее первых доверительных интервалов, построенных с центром в точке $\bar{\xi}$.

б) Пусть среднее известно и оценивается дисперсия σ^2 с помощью эффективной оценки

$$\check{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k - a)^2.$$

Величина $\frac{n\check{\sigma}^2}{\sigma^2} = \sum_{k=1}^n \eta_k^2$, где $\eta_k = \frac{\xi_k - a}{\sigma}$, η_k независимы и нормально распределены $(0, 1)$, имеет χ^2 -распределение с n степенями свободы (§ 8, гл. II). Доверительные интервалы можно построить следующим образом. Найдем такие числа h'_1 и h''_1 , чтобы $P\{\chi^2 \leq h'_1\} = \frac{1-\gamma}{2}$, $P\{\chi^2 \geq h''_1\} = \frac{1-\gamma}{2}$. Тогда $h'_1 < \frac{n\check{\sigma}^2}{\sigma^2} < h''_1$, или

$$\frac{n\check{\sigma}^2}{h''_1} < \sigma^2 < \frac{n\check{\sigma}^2}{h'_1}$$

с вероятностью γ . Иными словами, $\left(\frac{n\check{\sigma}^2}{h''_1}, \frac{n\check{\sigma}^2}{h'_1}\right)$ является γ -доверительным интервалом для σ^2 . Этот интервал обладает некоторой симметрией, интервалы $\left(\frac{n\check{\sigma}^2}{h''_1}, \infty\right)$, $\left(0, \frac{n\check{\sigma}^2}{h'_1}\right)$ являются $1 - \frac{\gamma}{2}$ — верхним (соответственно нижним) доверительным интервалом. Такой способ построения доверительных интервалов при данной статистике $\check{\sigma}^2$ не обязателен и границы доверительного интервала можно выбирать иначе, лишь бы вероятность покрытия доверительным интервалом оцениваемого параметра была равна γ .

в) Рассмотрим теперь случай, когда оба параметра a и σ^2 неизвестны. В качестве совместных оценок этих параметров примем оценки максимального правдоподобия $\bar{\xi} = \frac{1}{n} \sum_k \xi_k$ и $s^2 = \frac{1}{n} \sum (\xi_k - \bar{\xi})^2$. Для построения доверительных интервалов в данном случае можно воспользоваться тем очевидным фактом, что статистики $\frac{\bar{\xi} - a}{s}$, $\frac{s^2}{\sigma^2}$ имеют распределение, не зависящее от параметров a и σ^2 . Найдем это распределение.

Нам понадобятся некоторые теоремы о распределении линейных и квадратических форм от независимых нормальных случайных величин.

Пусть $\xi_k, k = 1, \dots, n$ — независимые нормальные $(0, 1)$ случайные величины.

Лемма 1. Пусть A — ортогональное преобразование в R^n . Тогда компоненты вектора $\tilde{\xi} = A\xi$ независимы и нормально распределены $(0, 1)$.

Доказательство. Из свойств нормального распределения вытекает, что $\tilde{\xi}$ — нормальный вектор. При этом $M\tilde{\xi} = 0$, а корреляционная матрица $\tilde{\xi}$ равна $M\tilde{\xi}\tilde{\xi}^* = MA\xi\xi^*A^* = AM(\xi\xi^*)A^* = AA^* = E$, так как $M\xi\xi^* = E$ и $AA^* = E$ согласно свойствам ортогональных преобразований.

Теорема 1. Пусть $Q_j = Q_j(x_1, \dots, x_n), j = 1, \dots, k$, — квадратические формы ранга $n_j \geq 1, \sum_{j=1}^k n_j = n$ и $\sum_{j=1}^k Q_j = \sum_{i=1}^n x_i^2$. Тогда случайные величины $Q_j(\xi) = Q_j(\xi_1, \dots, \xi_n)$ имеют χ^2 -распределение с n_j степенями свободы и взаимно независимы. При этом существует ортогональное преобразование B , одновременно приводящее все формы Q_j к сумме квадратов, и если $y = Bx$, то

$$Q_j(x) = y_{N_{j-1}+1}^2 + \dots + y_{N_j}^2, \quad N_j = n_1 + \dots + n_j, \quad N_0 = 0.$$

Доказательство. Так как формы Q_j имеют ранг n_j , то их можно представить в виде суммы n_j квадратов линейных

форм $Q_j(x) = \sum_{i=N_{j-1}+1}^{N_j} \chi_i \left(\sum_{s=1}^n b_{is} x_s \right)^2$, где χ_i равны ± 1 . Таким об-

разом, поскольку $\sum_{j=1}^k n_j = n$, то

$$\sum_{s=1}^n x_s^2 = \sum_{i=1}^n \chi_i \left(\sum_{s=1}^n b_{is} x_s \right)^2 = x^* B^* J B x,$$

где J — диагональная матрица с элементами χ_j . Из равенства двух симметричных квадратических форм вытекает равенство их коэффициентов. Следовательно, $B^* J B = E$. Матрица B невырождена и $J = (B^*)^{-1} B^{-1}$. Поэтому матрица J неотрицательно определена, $(Jx, x) = ((B^{-1})^* B^{-1} x, x) = (B^{-1} x, B^{-1} x) \geq 0$ и $\chi_j = 1$ для $j = 1, \dots, n, J = E$ и $B^* B = E$. Это означает, что матрица B ортогональна, вектор $\eta = B\xi$ имеет независимые нормальные компо-

ненты $(0, 1)$. Так как $Q_j(x) = \sum_{i=N_{j-1}+1}^{N_j} y_i^2, Q_j(\xi) = \sum_{N_{j-1}+1}^{N_j} \eta_i^2$, то ортогональное преобразование B одновременно приводит все квадратические формы к сумме квадратов, матрицы $Q_j(\xi)$ независимы и имеют χ^2 распределение с n_j степенями свободы.

Теорема 2. Пусть ξ_1, \dots, ξ_n — независимы и нормальны $(0, 1)$. Тогда выборочное среднее $\bar{\xi}$ и выборочная дисперсия s^2 независимы и величина $(n-1)s^2$ имеет χ^2 распределение с $n-1$ степенью свободы.

Доказательство. Так как $\sum_{i=1}^n \xi_i^2 = (n-1)s^2 + n(\bar{\xi})^2$, то применима лемма 2. При этом $k=2$, ранг квадратической формы $n(\bar{\xi})^2$ равен 1 и, следовательно, ранг квадратической формы $(n-1)s^2$ равен $n-1$, так что $(n-1)s^2$ имеет χ^2 -распределение с $n-1$ степенью свободы и $(n-1)s^2$ не зависит от $\bar{\xi}$.

Теорема 2 сразу переносится на случайные величины ξ_k , имеющие нормальное распределение с параметрами (a, σ^2) . Рассматривая вместо величины ξ_k величины $\frac{\xi_k - a}{\sigma}$, мы видим, что в этом случае $\bar{\xi}_k$ и $(n-1)s^2$ снова независимы и $\frac{(n-1)s^2}{\sigma^2}$ имеет χ^2 -распределение с $n-1$ степенью свободы. При этом величина $\frac{\sqrt{n}(\bar{\xi} - a)}{\sigma^2}$ имеет нормальное $(0, 1)$ распределение. Следовательно,

величина $\tau = \frac{\sqrt{n}(\bar{\xi} - a)}{\sigma^2} / \sqrt{\frac{(n-1)s^2}{(n-1)\sigma^2}} = \frac{\sqrt{n}(\bar{\xi} - a)}{s}$ имеет распределение Стьюдента (§ 2, гл. VI) с $(n-1)$ степенью свободы.

Теорема 3. Если $\xi_k, k=1, \dots, n$, — независимы и нормальны (a, σ^2) , то величина

$$\tau = \frac{\sqrt{n}(\bar{\xi} - a)}{s} = \frac{\sqrt{n}(\bar{\xi} - a)}{\sqrt{\frac{1}{n-1} \sum_{k=1}^n (\xi_k - \bar{\xi})^2}}$$

имеет распределение Стьюдента с $n-1$ степенью свободы.

Теоремы 2 и 3 позволяют строить доверительные интервалы для величин a и σ^2 . Например, если величина t_γ определена из требования $P\{|\tau| < t_\gamma\} = \gamma$, где τ имеет распределение Стьюдента с $n-1$ степенью свободы, то интервал $\left(\bar{\xi} - t_\gamma \frac{s}{\sqrt{n}}, \bar{\xi} + t_\gamma \frac{s}{\sqrt{n}}\right)$ является γ -доверительным интервалом для параметра a ,

$$P\left\{\left(\bar{\xi} - t_\gamma \frac{s}{\sqrt{n}}, \bar{\xi} + t_\gamma \frac{s}{\sqrt{n}}\right) \in a\right\} = \gamma.$$

Далее, определим h'_γ и h''_γ из условия

$$P(\chi^2 < h'_\gamma) = P(\chi^2 > h''_\gamma) = \frac{1-\gamma}{2},$$

где χ^2 имеет χ^2 -распределение с $n-1$ степенью свободы, получим для дисперсии σ^2 -доверительный интервал

$$\frac{(n-1)s^2}{h'_\gamma} < \sigma^2 < \frac{(n-1)s^2}{h''_\gamma}.$$

Указанные доверительные интервалы построены для величин a и σ^2 в отдельности. Было бы ошибочно считать, что пара значений (a, σ^2) лежит в прямоугольнике

$$\bar{\xi} - t_1 \frac{s}{\sqrt{n}} < a < \bar{\xi} + t_1 \frac{s}{\sqrt{n}}, \quad \frac{(n-1)s^2}{h_1'} < \sigma^2 < \frac{(n-1)s^2}{h_1''}$$

с вероятностью γ , так как величины t и s^2 зависимы. Доверительную область для пары параметров (a, σ^2) можно построить, например, снова используя независимость величин $\bar{\xi}$ и s^2 . Имеем

$$\begin{aligned} P \left\{ -t < \frac{\sqrt{n}(\bar{\xi} - a)}{\sigma} < t, \quad h' < \frac{(n-1)s^2}{\sigma^2} < h'' \right\} = \\ = P \left\{ \frac{|\sqrt{n}(\bar{\xi} - a)|}{\sigma} < t \right\} P \left\{ h' < \frac{(n-1)s^2}{\sigma^2} < h'' \right\}. \end{aligned}$$

Приравнявая каждый из этих множителей в правой части равенства к величине γ и определив соответствующие значения t, h', h'' , получим в плоскости (a, θ) γ -доверительную область для пары (a, θ) , $\theta = \sigma^2$ вида

$$\begin{aligned} -\frac{t}{\sqrt{n}} \sqrt{\theta} < a - \bar{\xi} < \frac{t}{\sqrt{n}} \sqrt{\theta}, \\ \beta' < \theta < \beta'' \left(\beta' = \frac{(n-1)s^2}{h'}, \quad \beta'' = \frac{(n-1)s^2}{h''} \right), \end{aligned}$$

представляющую собой область, ограниченную параболой $\theta = \frac{\sqrt{n}}{t} (a - \bar{\xi})^2$ и двумя прямыми $\theta = \beta'$ и $\theta = \beta''$.

Задача о двух нормальных выборках. Пусть $\xi_k, k = 1, \dots, n$ и $\eta_k, k = 1, \dots, m$ — две независимые нормальные выборки с параметрами (a_1, σ_1^2) и (a_2, σ_2^2) соответственно. Пусть параметры $a_i, i = 1, 2$ неизвестны. На практике нам часто надо будет знать, насколько значима разность между средними в этих выборках. Иными словами, нас будут интересовать доверительные интервалы для разности $a_1 - a_2$. Эту задачу называют задачей о сравнении двух средних. Рассмотрим два варианта этой задачи.

а) Пусть известны дисперсии σ_1^2 и σ_2^2 . Так как величины $\bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k$ и $\bar{\eta} = \frac{1}{m} \sum_{k=1}^m \eta_k$ независимы и нормальны $(a_1, \frac{\sigma_1^2}{n})$, $(a_2, \frac{\sigma_2^2}{m})$, то разность $\bar{\xi} - \bar{\eta}$ асимптотически нормальна со средним $a_1 - a_2$ и дисперсией $\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$. Доверительные интервалы для $a_1 - a_2$ имеют вид

$$\bar{\xi} - \bar{\eta} - t_1 \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} < a_1 - a_2 < \bar{\xi} - \bar{\eta} + t_1 \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}},$$

где $\pm t_\tau$ — симметрические доверительные границы для стандартного нормального закона.

б) Пусть дисперсии выборок неизвестны, но известно, что $\sigma_1^2 = \sigma_2^2$. Положим

$$S_1^2 = \frac{1}{n-1} \sum_{k=1}^n (\xi_k - \bar{\xi})^2, \quad S_2^2 = \frac{1}{m-1} \sum_{k=1}^m (\eta_k - \bar{\eta})^2.$$

Так как

$$\begin{aligned} n \left(\frac{\bar{\xi} - a_1}{\sigma_1} \right)^2 + m \left(\frac{\bar{\eta} - a_2}{\sigma_2} \right)^2 + \frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2} = \\ = \sum_{k=1}^n \left(\frac{\xi_k - a_1}{\sigma} \right)^2 + \sum_{k=1}^m \left(\frac{\eta_k - a_2}{\sigma} \right)^2 \end{aligned}$$

и квадратическая форма $(n-1)S_1^2 + (m-1)S_2^2$ имеет ранг $n + m - 2$, то по теореме 1 величина

$$\frac{1}{\sigma^2} [(n-1)S_1^2 + (m-1)S_2^2]$$

имеет χ^2 -распределение с $n + m - 2$ степенями свободы и не зависит от величин $\bar{\xi}$ и $\bar{\eta}$. Учитывая, что $(\bar{\xi} - a_1) - (\bar{\eta} - a_2)$ является нормальной $\left(0, \sigma^2 \left(\frac{1}{n} + \frac{1}{m}\right)\right)$ случайной величиной, видим, что статистика

$$\tau = \frac{\bar{\xi} - \bar{\eta} - (a_1 - a_2) \sqrt{mn}}{\sigma \sqrt{m+n}} \bigg/ \sqrt{\frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2}},$$

или

$$\tau = \frac{\bar{\xi} - \bar{\eta} - (a_1 - a_2)}{\sqrt{(n-1)S_1^2 + (m-1)S_2^2}} \sqrt{\frac{nm}{m+n}}$$

имеет распределение Стьюдента с $n + m - 2$ степенями свободы. Определив симметричные γ -доверительные границы t_τ для величины τ , $P(|\tau| < t_\tau) = \gamma$, получим γ -доверительный интервал для разности $a_1 - a_2$

$$\begin{aligned} (\bar{\xi} - \bar{\eta}) - t_\tau \sqrt{(n-1)S_1^2 + (m-1)S_2^2} \sqrt{\frac{m+n}{mn}} < a_1 - a_2 < \\ < (\bar{\xi} - \bar{\eta}) + t_\tau \sqrt{(n-1)S_1^2 + (m-1)S_2^2} \sqrt{\frac{m+n}{mn}}. \end{aligned}$$

Другой задачей, связанной с двумя выборками, является задача о сравнении дисперсий σ_1^2 и σ_2^2 . В этом случае интересуются доверительными интервалами для отношения дисперсий.

Оценкой такого отношения естественно принять величину $\frac{S_1^2}{S_2^2}$.

Так как отношение $z = \frac{nS_1^2}{\sigma_1^2} : \frac{mS_2^2}{\sigma_2^2}$ равно отношению двух независимых случайных величин χ_{n-1}^2 и χ_{m-1}^2 , то величина z имеет F распределение с $(n-1, m-1)$ степенями свободы. Найдя для величины z γ -доверительные границы c_1' и c_1'' , $P(z \leq c_1') = \frac{1-\gamma}{2}$, $P(z \geq c_1'') = \frac{1-\gamma}{2}$, получим для отношения $\frac{\sigma_1^2}{\sigma_2^2}$ следующий γ -доверительный интервал

$$\frac{c_1' n S_1^2}{c_1'' m S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{c_1' n S_1^2}{c_1'' m S_2^2}.$$

§ 5. Нормальная линейная регрессия

В практических применениях методов математической статистики часто встречается задача, простейший вариант которой сейчас будет рассмотрен.

Предполагается, что между величинами y, x_1, \dots, x_d существует линейная зависимость вида

$$y = \alpha_0 + \beta_1 x_1 + \dots + \beta_d x_d, \quad (1)$$

называемая линейной регрессией (§ 3, гл. VI). Здесь x_1, \dots, x_d — некоторые неслучайные переменные, а коэффициенты $\alpha_0, \beta_1, \dots, \beta_d$, называемые коэффициентами регрессии, неизвестны и должны быть определены экспериментально.

Проводится n независимых экспериментов. В каждом из них величины x_1, \dots, x_d принимают определенные известные значения x_{1k}, \dots, x_{dk} , $k = 1, \dots, n$, вообще говоря, разные в разных экспериментах. В каждом эксперименте измеряется нормальная случайная величина ξ_k , $k = 1, \dots, n$ со средним значением

$$M\xi_k = \alpha_0 + \beta_1 x_{1k} + \dots + \beta_d x_{dk}, \quad k = 1, \dots, n, \quad (2)$$

и постоянной, но неизвестной дисперсией σ^2 . Таким образом, рассматриваемая выборка $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ образована последовательностью независимых нормальных случайных величин ξ_1, \dots, ξ_n с дисперсией σ^2 и разным средним, имеющим вид (2). Оцениваются коэффициенты $\alpha_0, \beta_1, \dots, \beta_d$.

Рассмотрим сначала случай $d=1$. В этом случае регрессию (1) называют простой. Запишем уравнение регрессии в форме

$$y = \alpha + \beta(x - \bar{x}), \quad \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k.$$

Функция правдоподобия выборки имеет вид

$$f(\xi, \alpha, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{k=1}^n (\xi_k - \alpha - \beta(x_k - \bar{x}))^2 \right\}.$$

Для оценки трех неизвестных параметров α , β , σ^2 воспользуемся методом максимального правдоподобия. Чтобы отыскать максимум функции правдоподобия, минимизируем квадратическую форму

$$Q(\alpha, \beta) = \sum_{k=1}^n (\xi_k - \alpha - \beta(x_k - \bar{x}))^2,$$

а затем находим максимум по θ функции

$$\varphi(\theta) = (2\pi\theta)^{-\frac{n}{2}} \exp \left\{ -\frac{Q^*}{2\theta} \right\}, \quad Q^* = \min_{\alpha, \beta} Q(\alpha, \beta).$$

Так как (положим $\bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k$)

$$\begin{aligned} Q(\alpha, \beta) &= \sum_k [(\bar{\xi} - \alpha) + (\xi_k - \bar{\xi}) - \beta(x_k - \bar{x})]^2 = \\ &= n[(\bar{\xi} - \alpha)^2 + \beta^2 S_1^2 + S_2^2 - 2\beta R_{12}] = \\ &= n \left[(\bar{\xi} - \alpha)^2 + \left(S_1 \beta - \frac{R_{12}}{S_1} \right)^2 + S_2^2 - \frac{R_{12}^2}{S_1^2} \right], \end{aligned}$$

где

$$\begin{aligned} S_1 &= \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2, \quad S_2^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k - \bar{\xi})^2, \\ R_{12} &= \frac{1}{n} \sum_{k=1}^n (\xi_k - \bar{\xi})(x_k - \bar{x}), \end{aligned}$$

то минимум Q^* функции $Q(\alpha, \beta)$ достигается при

$$\alpha = \check{\alpha}, \quad \beta = \check{\beta}, \quad \check{\alpha} = \bar{\xi}, \quad \check{\beta} = \frac{R_{12}}{S_1^2}$$

и равен

$$\begin{aligned} Q^* &= \min_{\alpha, \beta} Q(\alpha, \beta) = \sum_{k=1}^n (\xi_k - \check{\alpha} - \check{\beta}(x_k - \bar{x}))^2 = \\ &= n S_2^2 (1 - \check{\rho}^2), \end{aligned}$$

$$\check{\rho} = \frac{R_{13}}{S_1 S_2} = \frac{\sum_{k=1}^n (\xi_k - \bar{\xi})(x_k - \bar{x})}{\sqrt{\sum_{k=1}^n (\xi_k - \bar{\xi})^2 \sum_{k=1}^n (x_k - \bar{x})^2}}.$$

Для определения значения θ , при котором функция $\varphi(\theta)$ достигает максимума, имеем уравнение $(\ln \varphi(\check{\theta}))' = 0$, или $\frac{n}{2\check{\theta}} - \frac{Q^*}{2\check{\theta}^2} = 0$, откуда $\check{\theta} = \check{\sigma}^2 = \frac{1}{n} Q^*$. Таким образом, в качестве оценки величины σ^2 находим величину

$$\check{\sigma}^2 = S_2^2 (1 - \check{\rho}^2).$$

Мы получили оценки максимального правдоподобия $\check{\alpha}$, $\check{\beta}$, $\check{\sigma}^2$ параметров α , β , σ^2

$$\check{\alpha} = \bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k, \quad \check{\beta} = \frac{\sum_{k=1}^n (x_k - \bar{x})(\xi_k - \bar{\xi})}{\sum_{k=1}^n (x_k - \bar{x})^2}, \quad (3)$$

$$\check{\sigma}^2 = \frac{1}{n} \sum_{k=1}^n (\xi_k - \check{\alpha} - \check{\beta}(x_k - \bar{x}))^2. \quad (4)$$

Укажем доверительные интервалы для этих параметров. Заметим, что

$$M\check{\alpha} = \frac{1}{n} \sum_{k=1}^n M\xi_k = \frac{1}{n} \sum_{k=1}^n (\alpha + \beta(x_k - \bar{x})) = \alpha.$$

Так как $M(\xi_k - \bar{\xi}) = \beta(x_k - \bar{x})$, то

$$M\check{\beta} = \frac{1}{S_1^2} \sum_{k=1}^n (x_k - \bar{x}) M(\xi_k - \bar{\xi}) = \beta.$$

Таким образом, $\check{\alpha}$ и $\check{\beta}$ являются несмещенными оценками параметров α и β . Поскольку величины ξ_k независимы, то

$$D\check{\alpha} = \frac{1}{n^2} \sum_{k=1}^n D\xi_k = \frac{\sigma^2}{n}.$$

Учитывая, что $\check{\beta} = \frac{1}{nS_1^2} \sum_{k=1}^n (x_k - \bar{x})(\xi_k - \bar{\xi}) = \frac{1}{nS_1^2} \sum_{k=1}^n (x_k - \bar{x}) \xi_k$, получим

$$D\check{\beta} = \frac{1}{n^2 S_1^4} \sum_{k=1}^n (x_k - \bar{x})^2 \sigma^2 = \frac{\sigma^2}{n S_1^2}.$$

Так как величины $\eta_k = \xi_k - \alpha - \beta(x_k - \bar{x})$, $k = 1, \dots, n$ независимы и нормально распределены $(0, \sigma^2)$, то квадратическая форма $\frac{1}{\sigma^2} Q(\alpha, \beta)$ имеет χ^2 -распределение с n степенями свободы. Нетрудно проверить справедливость равенства

$$Q(\alpha, \beta) = \sum_{k=1}^n [\xi_k - \check{\alpha} - \check{\beta}(x_k - \bar{x}) + (\check{\alpha} - \alpha) + (\check{\beta} - \beta)(x_k - \bar{x})]^2 = \sum_{k=1}^n (\xi_k - \check{\alpha} - \check{\beta}(x_k - \bar{x}))^2 + n(\check{\alpha} - \alpha)^2 + nS_1(\check{\beta} - \beta)^2.$$

Поэтому величины

$$\frac{1}{\sigma^2} \sum_{k=1}^n (\xi_k - \check{\alpha} - \check{\beta}(x_k - \bar{x}))^2 = \frac{n\check{\sigma}^2}{\sigma^2}, \quad \frac{\sqrt{n}(\check{\alpha} - \alpha)}{\sigma}, \quad \frac{\sqrt{n}S_1(\check{\beta} - \beta)}{\sigma}$$

независимы и первая из них имеет χ^2 -распределение с $n-2$ степенями свободы. Далее, величины

$$\frac{\sqrt{n}(\check{\alpha} - \alpha)}{\sigma}, \quad \frac{\sqrt{n}S_1(\check{\beta} - \beta)}{\sigma}$$

являются линейными комбинациями независимых нормальных величин и, следовательно, также нормальны. Из приведенных выше формул следует, что их среднее равно 0, а дисперсии равны 1. Поэтому отношения

$$\frac{\sqrt{n}(\check{\alpha} - \alpha)}{\sqrt{\frac{n}{n-2}\check{\sigma}^2}} = \frac{\sqrt{n-2}(\check{\alpha} - \alpha)}{\check{\sigma}}, \quad \frac{\sqrt{n-2}S_1(\check{\beta} - \beta)}{\check{\sigma}}$$

имеют распределение Стьюдента с $n-2$ степенями свободы. Полученные результаты можно сформулировать следующим образом:

Теорема 1. Оценки $\check{\alpha}$, $\check{\beta}$, $\check{\sigma}^2$, определяемые формулами (3) и (4), величин α , β , σ^2 в задаче о простой нормальной регрессии, являются несмещенными, независимыми между собой и γ -доверительные интервалы для них имеют следующий вид:

$$\check{\alpha} - t_\gamma \frac{\check{\sigma}}{\sqrt{n-2}} < \alpha < \check{\alpha} + t_\gamma \frac{\check{\sigma}}{\sqrt{n-2}}, \\ \check{\beta} - t_\gamma \frac{\check{\sigma}}{\sqrt{n-2}S_1} < \beta < \check{\beta} + t_\gamma \frac{\check{\sigma}}{\sqrt{n-2}S_1},$$

где t_γ — симметричный γ -доверительный уровень для распределения Стьюдента с $n-2$ степенями свободы $P(|\tau| < t_\gamma) = \gamma$, τ — случайная величина, имеющая распределение Стьюдента с $n-2$ степенями свободы.

Для дисперсии σ^2 имеем доверительный интервал

$$\frac{n\check{\sigma}^2}{h_1^*} < \sigma^2 < \frac{n\check{\sigma}^2}{h_1'},$$

где h_1' и h_1^* — границы γ -доверительного интервала для χ^2 -распределения с $n-2$ — степенями свободы:

$$P(\chi_{n-2}^2 < h_1') = \frac{1-\gamma}{2}, \quad P(\chi_{n-2}^2 > h_1^*) = \frac{1-\gamma}{2}.$$

В некоторых случаях интерес представляет оценка ординаты $Y = \alpha + \beta(X - \bar{x})$ прямой регрессии при данном значении X . Эта оценка задается выражением $\check{Y} = \check{\alpha} + \check{\beta}(X - \bar{x})$. Чтобы найти доверительные интервалы для величины Y , заметим, что из независимости величин $\check{\sigma}^2$, $\check{\alpha}$, $\check{\beta}$ вытекает независимость величин $\check{\alpha} - \alpha + (\check{\beta} - \beta)(X - \bar{x}) = \check{Y} - Y$ и $\check{\sigma}^2$. Далее,

$$M\check{Y} = Y, \quad D\check{Y} = D\check{\alpha} + (X - \bar{x}) D\check{\beta} = \frac{\sigma^2}{n} \left(1 + \frac{(X - \bar{x})^2}{S_1^2} \right).$$

Следовательно, отношение

$$\tau = \frac{\sqrt{n-2}(\check{Y} - Y)}{\check{\sigma} \sqrt{1 + \frac{(X - \bar{x})^2}{S_1^2}}} \quad (5)$$

имеет распределение Стьюдента с $n-2$ степенями свободы. Доверительные границы для величины Y имеют вид

$$\check{\alpha} + \check{\beta}(X - \bar{x}) \pm t_\gamma \frac{\check{\sigma}}{\sqrt{n-2}} \sqrt{1 + \frac{(X - \bar{x})^2}{S_1^2}}, \quad (6)$$

где $\pm t_\gamma$ — симметричные γ -доверительные границы для распределения Стьюдента с $n-2$ степенями свободы.

Множественная регрессия. Обобщим полученные результаты на тот случай, когда среднее значение нормальной случайной величины зависит от d аргументов. Итак, мы имеем n независимых нормальных случайных величин ξ_k , $k = 1, \dots, n$ с одинаковой дисперсией σ^2 и со средним $a_k = M\xi_k$, имеющим вид (2). Параметры распределений этих величин σ^2 , α_0 , β_1, \dots, β_d неизвестны и должны быть оценены. Удобнее формулы (2) записать иначе:

$$M\xi_k = \alpha + \beta_1(x_{1k} - \bar{x}_1) + \beta_2(x_{2k} - \bar{x}_2) + \dots + \beta_d(x_{dk} - \bar{x}_d), \quad (7)$$

где $\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ik}$. Для логарифма функции правдоподобия имеем следующее выражение:

$$\ln f(\xi, \alpha, \beta, \sigma^2) = \pm \frac{n}{2} \ln 2\pi\sigma^2 - \frac{Q(\alpha, \beta)}{2\sigma^2}, \quad (8)$$

$$\xi = (\xi_1, \xi_2, \dots, \xi_n),$$

$$Q(\alpha, \beta) = \sum_{k=1}^n [\xi_k - \alpha - \sum_{j=1}^d \beta_j (x_{jk} - \bar{x}_j)]^2.$$

Введем векторные обозначения. Пусть

$$y_j = (x_{j1} - \bar{x}_1, x_{j2} - \bar{x}_2, \dots, x_{jn} - \bar{x}_n), \quad j = 1, \dots, d,$$

$$\eta = \xi - \alpha \mathbf{1} - \sum_{j=1}^d \beta_j y_j.$$

Здесь $\mathbf{1}$ — вектор в R^n , компоненты которого равны 1. Вектор

$$\eta = (\eta_1, \eta_2, \dots, \eta_n), \quad \eta_k = \xi_k - \alpha - \sum_{j=1}^d \beta_j (x_{jk} - \bar{x}_j)$$

имеет независимые нормально распределенные компоненты с параметрами $(0, \sigma^2)$. Выражение для $Q(\alpha, \beta)$ принимает вид

$$Q(\alpha, \beta) = \|\xi - \alpha \mathbf{1} - \sum_{j=1}^d \beta_j y_j\|^2, \quad (9)$$

где $\|\cdot\|$ — норма в пространстве R^n .

Задача максимизации функции правдоподобия сводится к минимизации квадратической формы $Q(\alpha, \beta)$ и дальнейшей максимизации функции $\ln f(\xi, \alpha, \beta, \sigma^2)$ по σ^2 . Если

$$Q^* = \min_{\alpha, \beta} Q(\alpha, \beta),$$

причем минимум $Q(\alpha, \beta)$ достигается при $\alpha = \check{\alpha}$, $\beta = \check{\beta}$, $Q^* = Q(\check{\alpha}, \check{\beta})$, то

$$f^* = \max_{\alpha, \beta, \sigma^2} f(\xi, \alpha, \beta, \sigma^2) = \max_{\sigma^2} \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left\{ -\frac{Q^*}{2\sigma^2} \right\}.$$

Функция в правой части равенства достигает своего максимума при $\sigma^2 = \check{\sigma}^2$, где

$$\check{\sigma}^2 = \frac{1}{n} Q^* = \frac{1}{n} \left\| \xi - \check{\alpha} \mathbf{1} - \sum_{j=1}^d \check{\beta}_j y_j \right\|^2. \quad (10)$$

Остается найти значения $\check{\alpha}$ и $\check{\beta}$.

Задача минимизации квадратической формы (9) имеет простой геометрический смысл. Пусть H — линейное пространство в R^n , натянутое на $d+1$ вектор $\mathbf{1}, y_1, \dots, y_d$. Минимум Q^* формы Q равен квадрату длины перпендикуляра, опущенного из конца

вектора ξ на подпространство H . Если этот минимум достигается при $\alpha = \check{\alpha}$, $\beta_j = \check{\beta}_j$ ($j = 1, \dots, d$), то $\check{\xi} = \check{\alpha} + \sum_{j=1}^d \check{\beta}_j y_j$ является проекцией случайного вектора ξ на H , а $\xi - \check{\xi}$ — вектором, ортогональным H и $|\xi - \check{\xi}|^2 = Q^*$. Для определения проекции $\check{\xi}$ имеем $d+1$ линейное уравнение (так называемые нормальные уравнения метода наименьших квадратов)

$$(\xi - \check{\xi}, 1) = 0, (\xi - \check{\xi}, y_j) = 0, j = 1, \dots, d. \quad (11)$$

Здесь (x, y) — скалярное произведение вектора x и y в R^n : если $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, то $(x, y) = \sum_{k=1}^n x_k y_k$. Заметим, что вектор 1 ортогонален векторам y_j , $j = 1, \dots, d$,

$$(y_j, 1) = \sum_{k=1}^n (x_{jk} - \bar{x}) = 0, \text{ причем } (1, 1) = n.$$

Из уравнения $(\xi - \check{\xi}, 1) = 0$ получаем

$$\check{\alpha} = \frac{1}{n} (\xi, 1) = \bar{\xi}.$$

Остальные уравнения (11) можно записать в виде

$$\sum_{i=1}^d (y_i, y_j) \check{\beta}_i = (\xi, y_j), j = 1, \dots, d. \quad (12)$$

Предположим, что определитель Γ этой системы

$$\Gamma = \begin{vmatrix} (y_1, y_1) & (y_1, y_2) & \dots & (y_1, y_d) \\ (y_2, y_1) & (y_2, y_2) & \dots & (y_2, y_d) \\ \vdots & \vdots & \ddots & \vdots \\ (y_d, y_1) & (y_d, y_2) & \dots & (y_d, y_d) \end{vmatrix}$$

отличен от 0, $\Gamma \neq 0$. Обозначим через γ_i определитель, который получается из Γ , если в нем заменить i -й столбец правыми частями системы (12). Тогда $\check{\beta}_i = \frac{\gamma_i}{\Gamma}$, $i = 1, \dots, d$. Заметим, что

$(\xi, y_j) = \sum_{i=1}^d \beta_i (y_i, y_j) + (\eta, y_j)$. В соответствии с этими равенствами, заменим величины (ξ, y_j) в выражении для γ_i . Получим $\gamma_i = \beta_i \Gamma + \bar{\gamma}_i$, где $\bar{\gamma}_i$ — определитель, получаемый из Γ заменой i -го столбца на столбец из величин (η, y_j) , $j = 1, \dots, d$. Далее

$$\bar{\gamma}_i = \sum_{j=1}^d \Gamma_{ij} (\eta, y_j) = \sum_{k=1}^n \left(\sum_{j=1}^d \Gamma_{ij} y_{jk} \right) \eta_k,$$

где Γ_{ij} — алгебраические дополнения элементов определителя.

Таким образом, мы получаем для величин $\check{\beta}_i$ следующее представление:

$$\check{\beta}_i = \beta_i + \frac{1}{\Gamma} \sum_{k=1}^n \left(\sum_{j=1}^d \Gamma_{ij} y_{jk} \right) \eta_k, \quad i = 1, \dots, d. \quad (13)$$

Кроме того,

$$\alpha = \frac{1}{n} \sum_{i=1}^n \xi_k = \alpha + \frac{1}{n} \sum_{j=1}^d \beta_j \sum_{k=1}^n y_{jk} + \frac{1}{n} \sum_{k=1}^n \eta_k,$$

или

$$\check{\alpha} = \alpha + \frac{1}{n} \sum_{k=1}^n \eta_k = \alpha - \bar{\eta}. \quad (14)$$

Теперь нетрудно указать моменты первого и второго порядка величин $\check{\alpha}$, $\check{\beta}$. Имеем

$$\begin{aligned} M\check{\alpha} &= \alpha, \quad M\check{\beta}_i = \beta_i, \quad D\check{\alpha} = \frac{\sigma^2}{n}, \\ M(\check{\alpha} - \alpha)(\check{\beta}_j - \beta_j) &= M \frac{1}{n} \sum_{k=1}^n \eta_k \frac{1}{\Gamma} \sum_{k=1}^n \left(\sum_{j=1}^d \Gamma_{ij} y_{jk} \right) \eta_k = \\ &= \frac{\sigma^2}{n\Gamma} \sum_{k=1}^n \sum_{j=1}^d \Gamma_{ij} y_{jk} = 0, \\ M(\check{\beta}_i - \beta_i)(\check{\beta}_r - \beta_r) &= \frac{\sigma^2}{\Gamma^2} \sum_{k=1}^n \sum_{j=1}^d \Gamma_{ij} y_{jk} \left(\sum_{j'=1}^d \Gamma_{rj'} y_{j'k} \right) = \\ &= \frac{\sigma^2}{\Gamma^2} \sum_{j,j'} \Gamma_{ij} \Gamma_{rj'} \sum_{k=1}^n y_{jk} y_{j'k} = \frac{\sigma^2}{\Gamma^2} \sum_j \Gamma_{ij} \sum_{j'} \Gamma_{rj'} (y_j, y_{j'}) = \\ &= \frac{\sigma^2}{\Gamma^2} \sum_j \Gamma_{ij} \Gamma \delta_{rj} = \frac{\sigma^2 \Gamma_{ir}}{\Gamma}. \end{aligned}$$

Обратимся к формуле $|\xi|^2 = |\xi|^2 + |\xi_N|^2$. Она имеет место для произвольных значений вектора $\xi = (z_1, z_2, \dots, z_n)$. При этом $|\xi|^2$ является квадратической формой от z_1, \dots, z_n ранга $d+1$ (вследствие того, что $\Gamma \neq 0$ и вектор 1 ортогонален векторам y_1, \dots, y_d), а $|\xi_N|^2$ — квадратическая форма ранга $n-d-1$, так как ξ_N — произвольный вектор из ортогонального дополнения к H . Поэтому величина $\frac{1}{\sigma^2} |\xi_N|^2$ имеет χ^2 — распределение с $n-d-1$ степенью свободы, а любая линейная комбинация вектора ξ не зависит от ξ_N . Получаем следующий результат:

$$\frac{n\check{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{k=1}^n \left(\check{\xi}_k - \check{\alpha} - \sum_{j=1}^d \check{\beta}_j (x_{jk} - \bar{x}_j) \right)^2$$

имеет χ^2 -распределение с $n-d-1$ степенью свободы и не зависит от $\check{\alpha}$ и $\check{\beta}_j$, $j = 1, \dots, d$;

б) каждая из величин

$$\sqrt{n-d-1} \frac{\check{\alpha} - \alpha}{\sigma^2}, \quad \sqrt{\frac{(n-d-1)\Gamma}{n\Gamma_{ij}}} \frac{\check{\beta}_j - \beta_j}{\check{\sigma}}, \quad j = 1, \dots, d$$

имеет распределение Стьюдента с $n-d-1$ степенью свободы.

Доказанная теорема позволяет строить доверительные интервалы для коэффициентов регрессии $\alpha, \beta_1, \dots, \beta_d$.

Обобщая формулу (6), можно получить γ -доверительный интервал для ординаты y линии регрессии в данной точке.

(X_1, \dots, X_d) , $y = \alpha + \sum_{j=1}^d \beta_j (X_j - \bar{x}_j)$. Он имеет вид

$$\begin{aligned} & \check{\alpha} + \sum_{j=1}^d \check{\beta}_j (X_j - \bar{x}_j) \pm t_\gamma \frac{\check{\sigma}}{\sqrt{n-d-1}} \times \\ & \times \sqrt{1 + \sum_{i,j=1}^d \frac{n\Gamma_{ij}}{\Gamma} (X_i - \bar{x}_i)(X_j - \bar{x}_j)}, \end{aligned}$$

где t_γ определяется как симметричная γ -доверительная граница распределения Стьюдента с $n-d-1$ степенью свободы.

Замечание. В тех случаях, когда распределение ошибок измерений

$$\xi_k' = \xi_k - \left[\alpha + \sum_{j=1}^d \beta_j (x_{jk} - \bar{x}_j) \right]$$

неизвестно, и неизвестно, являются ли величины ξ_k' неизвестными, но сохраняется гипотеза об их некоррелированности и $M\xi_k' = 0$, $D\xi_k' = \sigma^2$, в качестве оценок коэффициентов α и β_j часто (начиная от Гаусса) предлагают пользоваться величинами $\check{\alpha}$ и $\check{\beta}_j$, для которых квадратическая форма

$$\begin{aligned} Q(\alpha, \beta) &= \sum_{k=1}^n \left(\xi_k - \left[\alpha + \sum_{j=1}^d \beta_j (x_{jk} - \bar{x}_j) \right] \right)^2 = \\ &= \left| \xi - \left(\alpha 1 + \sum_{j=1}^d \beta_j y_j \right) \right|^2 \end{aligned}$$

достигает минимума, т. е. оценками, определенными предыдущими формулами.

Метод оценки параметров, основанный на минимизации квадратической формы $Q(\alpha, \beta)$, называют методом наименьших квадратов. Теорию этого метода и более подробное изложение теории регрессии можно найти в ряде учебников и монографий по математической статистике (см., например, Линник Ю. В. Метод наименьших квадратов и основы теории обработки наблюдений. М., Физматгиз, 1958).

Глава X

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

§ 1. Проверка простой гипотезы. Критерий χ^2

Одной из задач математической статистики является установление согласованности данной последовательности наблюдений случайных величин или событий с гипотезами о распределении случайной величины (или вероятности события). Проверяемые гипотезы возникают или на основании теоретических соображений, или в результате статистического исследования других наблюдений.

Опишем основной процесс, который используется при проверке гипотезы. Выбирается некоторая величина, характеризующая отклонение частот наблюдаемых событий от гипотетических вероятностей. Поскольку эта величина является функцией от наблюдений, она является статистикой. Статистики, выбираемые для проверки гипотез (обычно они неотрицательны), называются критериями. Пусть выбран некоторый критерий K . Найдем его распределение в предположении, что распределение наблюдений совпадает с гипотетическим. Фиксируем заранее некоторый «уровень значимости» p , считая, что в единичном эксперименте события с вероятностью, меньшей p , практически не происходят. По p найдем такое число x_p , чтобы

$$P\{K > x_p\} = p$$

(для простоты считаем, что распределение K непрерывно). Пусть \hat{K} — вычисленное по наблюдениям значение критерия. Если $\hat{K} > x_p$, то в предположении справедливости гипотезы произошло «практически невозможное» событие, поэтому гипотеза отвергается. В противном случае считается, что наблюдения не противоречат гипотезе.

Проиллюстрируем общую идею на примере критерия А. Н. Колмогорова. Пусть проверяется гипотеза о том, что данные наблюдения x_1, \dots, x_n имеют непрерывную функцию распределения $F(x)$. В качестве критерия возьмем величину

$$D_n = \sqrt{n} \sup_x |F(x) - F_n^*(x)|, \quad (1)$$

где $F_n^*(x)$ — эмпирическая функция распределения, построенная по величинам x_1, \dots, x_n . Особенностью критерия \mathcal{D}_n есть то, что его распределение не зависит от функции распределения $F(x)$. Действительно, полагая в (1) $x = F^{-1}(t)$, где $t \in [0, 1]$, $F^{-1}(t)$ — обратная к F функция (она существует, поскольку F непрерывна и монотонна), получим

$$\mathcal{D}_n = \sqrt{n} \sup_{0 \leq t \leq 1} |t - \Phi_n^*(t)|,$$

где

$$\Phi_n^*(t) = \frac{1}{n} \sum_{i=1}^n \chi \{x_i < \Phi_0^{-1}(t)\} = \frac{1}{n} \sum \chi \{F^{-1}(x_i) < t\}$$

есть эмпирическая функция распределения, построенная по величинам $F^{-1}(x_i)$, которые (независимо от вида F , если только F непрерывно) имеют равномерное распределение на $[0, 1]$. Независимость распределения \mathcal{D}_n от F делает возможным использование критерия, так как если бы в каждом случае было свое распределение, то соответствующие статистические таблицы были бы необозримы. Кроме того, имеется другая аналогичная трудность: распределение \mathcal{D}_n зависит от n . Следовательно, если изменится число наблюдений, то нужно заново вычислять распределение \mathcal{D}_n . Оказывается, что при достаточно больших n распределение \mathcal{D}_n уже практически не зависит от n , так как существует предельное распределение \mathcal{D}_n при $n \rightarrow \infty$. Это предельное распределение и используется в качестве распределения \mathcal{D}_n уже при $n \geq 20$. Выбрав, например, число λ так, чтобы

$$\lim_{n \rightarrow \infty} P\{\mathcal{D}_n > \lambda\} = \frac{5}{100},$$

мы получим следующее правило проверки гипотезы с 5% уровнем значимости: если \mathcal{D}_n^* величина, вычисленная по наблюдениям, то при $\mathcal{D}_n^* > \lambda$ гипотеза отвергается, а при $\mathcal{D}_n^* \leq \lambda$ — принимается (пока до следующих наблюдений).

Предельное распределение величины \mathcal{D}_n находится с использованием таких методов теории случайных процессов, которые в этой книге не излагались, поэтому мы его подробно не рассматриваем. Наиболее широкое применение в статистике получил критерий χ^2 , который будет рассмотрен подробно.

Этот критерий используется при проверке гипотезы о вероятностях нескольких несовместимых событий (или, что то же самое, о распределении случайной величины, принимающей конечное число значений). Мы уже отмечали преимущества критерия Колмогорова, это независимость распределения от проверяемой гипотезы и предельная независимость от числа наблюдений. Предлагаемый критерий χ^2 обладает предельным распределением, неза-

висимым (в определенной степени) от гипотезы, что также облегчает его применение.

Пусть в эксперименте наблюдается одно из r событий: A_1, \dots, A_r , p_1, \dots, p_r — гипотетические вероятности этих событий. Проведено n наблюдений, при этом событие A_k наблюдалось v_k раз, $k = 1, \dots, r$. Составим величину

$$\chi_n^2 = \sum_{i=1}^r \frac{(v_i - np_i)^2}{np_i}. \quad (2)$$

На основании результатов § 8, гл. VI, предельное распределение χ_n^2 совпадает с χ^2 -распределением с $r - 1$ степенью свободы, т. е. с распределением $\xi_1^2 + \dots + \xi_{r-1}^2$, где ξ_1, \dots, ξ_{r-1} — независимые одинаково распределенные величины, имеющие гауссовское распределение со средним 0 и дисперсией 1. Поэтому при $x > 0$

$$\lim_{n \rightarrow \infty} P\{\chi_n^2 < x\} = \frac{1}{2^{\frac{r}{2}} \Gamma\left(\frac{r}{2}\right)} \int_0^x t^{\frac{r}{2}-1} e^{-\frac{t}{2}} dt. \quad (3)$$

Таким образом, зависимость от гипотетического распределения входит через число возможных событий r . При использовании критерия χ^2 для проверки гипотезы о распределении непрерывной величины возможные значения величины разбивают на конечное число непересекающихся областей $\mathcal{J}_1, \dots, \mathcal{J}_r$ и в качестве событий A_k рассматривают события $\{\xi \in \mathcal{J}_k\}$, где ξ — наблюдаемая величина.

При больших значениях r можно приближенно заменить распределение χ^2 нормальным со средним $r - 1$ и дисперсией $2(r - 1)$.

Пусть λ_p такое значение, что

$$\frac{1}{2^{\frac{r}{2}} \Gamma\left(\frac{r}{2}\right)} \int_{\lambda_p}^{\infty} t^{\frac{r}{2}-1} e^{-\frac{t}{2}} dt = p.$$

Тогда при $\chi_n^2 \leq \lambda_p$ гипотеза считается неопровергнутой, а при $\chi_n^2 > \lambda_p$ она отвергается. Посмотрим, что происходит в том случае, когда истинные значения вероятностей событий A_k не равны p_k . Предположим, что $P\{A_n\} = \bar{p}_n$, причем $\sum (p_n - \bar{p}_n)^2 > 0$. Тогда

$$\chi_n^2 = \sum_{i=1}^r \frac{(v_i - np_i)^2}{np_i} + 2 \sum_{i=1}^r \frac{(v_i - np_i)}{p_i} \frac{p_i - \bar{p}_i}{p_i} + n \sum_{i=1}^r \frac{(p_i - \bar{p}_i)^2}{p_i}.$$

Поскольку $M \sum_{i=1}^r \frac{(v - np_i)^2}{np_i} = r - 1$, то $\frac{1}{n} \sum_{i=1}^r \frac{(v_i - n\bar{p}_i)^2}{np_i} \rightarrow 0$ по вероятности. Далее

$$\begin{aligned} & M \left(\frac{1}{n} \sum_{i=1}^r (v - np_i) \left(\frac{p_i - \bar{p}_i}{p_i} \right) \right)^2 \leq \\ & \leq r \frac{1}{n^2} \sum_{i=1}^r M (v - np_i)^2 \frac{(p_i - \bar{p}_i)^2}{p_i^2} = \\ & = r \frac{1}{n^2} \sum_{i=1}^r n \bar{p}_i (1 - \bar{p}_i) \frac{(p_i - \bar{p}_i)^2}{p_i^2} \rightarrow 0. \end{aligned}$$

Следовательно, и $\frac{1}{n} \sum_{i=1}^r \frac{(v - n\bar{p}_i)(p_i - \bar{p}_i)}{p_i} \rightarrow 0$ по вероятности. Тогда

$$\frac{1}{n} \chi_n^2 \rightarrow \sum_{i=1}^r \frac{(p - \bar{p}_i)^2}{p_i}.$$

Поэтому для больших n

$$\chi_n^2 \sim n \sum_{i=1}^r \frac{(p - \bar{p}_i)^2}{p_i}$$

и с достаточно большой вероятностью $\chi_n^2 > \chi_r$. Таким образом, если гипотеза не верна, то мы, используя критерий χ^2 , отвергаем ее при достаточно большом n . Вероятность отвергнуть гипотезу, если она не является истинной, называется мощностью критерия. Эта вероятность зависит от того распределения вероятностей $(\bar{p}_1, \dots, \bar{p}_r)$, которое является истинным. Любое распределение вероятностей, которое может оказаться истинным, но отличается от гипотетического, называется альтернативным, или альтернативой. Так что мощность критерия зависит от альтернативы. Мы показали, что для критерия χ^2 при любом выборе уровня значимости ρ мощность стремится к 1 при $n \rightarrow \infty$.

§ 2. Задача о выборе из двух гипотез

Эта задача возникает при проверке некоторой гипотезы, но причем, имеется всего одна альтернатива, другая гипотеза H_1 . Будем считать, что обе гипотезы относятся к распределению некоторой случайной величины со значениями в R^n . В частности, эта случайная величина может иметь вид

$$\xi = (\xi_1, \dots, \xi_n),$$

где ξ_i — координаты ξ , независимые одинаково распределенные величины. В этом случае мы проверяем гипотезу о распределении (одномерной) случайной величины по n с независимым наблюдением. Пусть $F_i(dx)$ — распределение величины при гипотезе H_i ($i = 0, 1$).

В данном случае отвержение гипотезы H_0 означает принятие гипотезы H_1 , поэтому задача и называется задачей о выборе из двух гипотез.

Всякое правило характеризуется вероятностью принять ту или другую гипотезу в зависимости от наблюдаемого значения величины. Пусть $\pi(x)$ есть «критическая» вероятность — вероятность отвергнуть «основную» гипотезу H_0 , если наблюдаемая величина приняла значения x , т. е.

$$\pi(x) = P(H_1/x)$$

(вероятность принять гипотезу H_1 при условии, что наблюдалось значение x). Качество правила (оно обычно также называется критерием, как и статистика $\pi(x)$), определяется вероятностями принятия и отвержения каждой из гипотез в зависимости от того, какая из гипотез верна. Обычно его характеризуют вероятностями ошибок. Можно отвергнуть истинную основную гипотезу H_0 . Это ошибка первого рода, она обычно обозначается α

$$\alpha = P(H_1/H_0).$$

Ошибка второго рода — принять основную гипотезу, если верна альтернатива, ее вероятность обозначается β :

$$\beta = P(H_0/H_1).$$

Таким образом, α — уровень значимости критерия, $1 - \beta$ — его мощность. Обычно уровень значимости α выбирается заранее (например, уровень значимости 5% или уровень значимости 1%, при которых $\alpha = 0,05$ или $\alpha = 0,01$). Мощность критерия при этом стараются сделать максимальной, следовательно, вероятность ошибки второго рода β — минимальной. Если выбрана критическая вероятность, то

$$\alpha = \int \pi(x) F_0(dx), \quad \beta = \int (1 - \pi(x)) F_1(dx). \quad (1)$$

Рассмотрим случай, когда существует критерий, достоверно различающий гипотезы: $\alpha = 0$, $\beta = 0$. Пусть распределения F_0 и F_1 взаимно сингулярны, т. е. сосредоточены на непересекающихся множествах \mathcal{I}_0 и \mathcal{I}_1 :

$$F_0(\mathcal{I}_1) = 0, \quad F_1(\mathcal{I}_0) = 0, \quad \mathcal{I}_1 \cap \mathcal{I}_0 = \emptyset.$$

Предположим, что $\pi(x) = 1$ при $x \in \mathcal{J}_1$ и $\pi(x) = 0$ при $x \notin \mathcal{J}_1$. Тогда

$$\begin{aligned}\alpha &= \int \pi(x) F_0(dx) = \int_{\mathcal{J}_1} F_0(dx) = F_0(\mathcal{J}_1) = 0, \\ \beta &= \int (1 - \pi(x)) F_1(dx) = 1 - \int \pi(x) F_1(dx) = \\ &= 1 - \int_{\mathcal{J}_1} F_1(dx) = 0.\end{aligned}$$

Предложенный выше критерий различения гипотез характерен тем, что вероятность $\pi(x)$ может принимать лишь два значения 0 и 1. Такие критерии называются **нерандомизированными**. Нерандомизированный критерий выбора между двумя гипотезами задается множеством W тех x , для которых $\pi(x) = 1$. Это множество называется **критическим** и при $x \in W$ основная гипотеза отвергается. Уровень значимости критерия

$$\alpha = F_0(W),$$

а его мощность

$$1 - \beta = F_1(W).$$

Найдем наиболее мощный критерий с заданным уровнем значимости. Рассмотрим наиболее важный для практики случай, когда мера F_1 абсолютно непрерывна относительно меры F_0 : для всех борелевских множеств A из R^m

$$F_1(A) = \int_A f(x) F_0(dx). \quad (2)$$

Укажем два случая, когда равенство (2) выполнено.

1) Пусть F_1 и F_2 имеют плотности относительно лебеговой меры в R^m , равные $f_1(x)$ и $f_0(x)$ соответственно, причем $f_0 > 0$. Тогда равенство (2) выполняется, если

$$f(x) = \frac{f_1(x)}{f_0(x)}.$$

2) Пусть обе меры $F_i(dx)$ дискретны, т. е. можно указать такую последовательность $x_i \in R^m$, что

$$\sum F_0(\{x_i\}) = \sum F_1(\{x_i\}) = 1$$

и $F_0(\{x_i\}) > 0$ для всех x_i . Тогда $f(x_i) = \frac{F_1(\{x_i\})}{F_0(\{x_i\})}$, значение $f(x)$ при $x \neq x_i$ роли не играет.

Пусть \mathcal{J}_t — такое множество, что $f(x) \geq t$ при $x \in \mathcal{J}_t$, $f(x) \leq t$ при $x \in R^m - \mathcal{J}_t$. Предположим, что $F_0(\mathcal{J}_t) = \alpha$. Тогда нерандомизированный критерий с критическим множеством \mathcal{J}_t является

наиболее мощным при уровне значимости α . Действительно, пусть критерий задается критической функцией $\pi(x)$. Тогда

$$\begin{aligned}\alpha &= F_0(\mathcal{J}_t) = \int \pi(x) F_0(dx), \\ 1 - \beta &= \int \pi(x) F_1(dx) = \int \pi(x) f(x) F_0(dx) = \\ &= \int_{\mathcal{J}_t} \pi(x) f(x) F_0(dx) + \int_{R^m - \mathcal{J}_t} \pi(x) f(x) F_0(dx) = \\ &= F_1(\mathcal{J}_t) + \int_{R^m - \mathcal{J}_t} \pi(x) f(x) F_0(dx) - \int_{\mathcal{J}_t} (1 - \pi(x)) f(x) F_0(dx) \leq \\ &\leq F_1(\mathcal{J}_t) + t \int_{R^m - \mathcal{J}_t} \pi(x) F_0(dx) - t \int_{\mathcal{J}_t} (1 - \pi(x)) F_0(dx) = \\ &= F_1(\mathcal{J}_t) + t \left(\int_{R^m} \pi(x) F_0(dx) - F_0(\mathcal{J}_t) \right) = F_1(\mathcal{J}_t).\end{aligned}\quad (3)$$

(Мы воспользовались тем, что $0 \leq \pi(x) \leq 1$, $0 \leq 1 - \pi(x) \leq 1$, $f(x) \leq t$ при $x \in R^m - \mathcal{J}_t$, $-f(x) \leq -t$ при $x \in \mathcal{J}_t$ и (3)).

Приведенное выше рассуждение показывает, что в непрерывном случае 1) для всякого уровня значимости α существует указанного вида критерий максимальной мощности. Чтобы его построить, рассмотрим убывающую функцию

$$\psi(t) = F_0(\mathcal{D}_t), \quad \mathcal{D}_t = \{x: f(x) > t\}.$$

Пусть t_α таково, что $\psi(t_\alpha + 0) \leq \alpha$, $\psi(t_\alpha - 0) \geq \alpha$. Обозначим $\Gamma_\alpha = \{x: f(x) = t_\alpha\}$. Если $F_0(\Gamma_\alpha) = 0$, то $\psi(t_\alpha + 0) - \psi(t_\alpha - 0) = F_0(\Gamma_\alpha) = 0$, следовательно, $F_0(\mathcal{D}_{t_\alpha}) = \alpha$. Если $F_0(\Gamma_\alpha) > 0$, то

$$F_0(\mathcal{D}_{t_\alpha}) \leq \alpha \leq F_0(\mathcal{D}_{t_\alpha} \cup \Gamma_\alpha)$$

и вследствие непрерывности меры F_0 можно указать такое $\Delta \subset \Gamma_\alpha$, что

$$F_0(\mathcal{D}_{t_\alpha} \cup \Delta) = \alpha. \quad (4)$$

Так как $f(x) \geq t_\alpha$ на $\mathcal{D}_{t_\alpha} \cup \Delta$ и $f(x) \leq t_\alpha$ на $R^m - \mathcal{D}_{t_\alpha} \cup \Delta$, то, выбирая $W = \mathcal{J}_{t_\alpha} = \mathcal{D}_{t_\alpha} \cup \Delta$, получаем наиболее мощный нерандомизированный критерий с уровнем значимости α .

Если распределение F_0 дискретно, то приведенное выше рассуждение не подходит, так как нельзя подобрать такое Δ , чтобы выполнялось равенство (4). Пусть $C_{t_\alpha} = R^n - \mathcal{D}_{t_\alpha} \cup \Gamma_\alpha$. Рассмотрим критерий, для которого

$$\pi(x) = \begin{cases} 1, & x \in \mathcal{D}_{t_\alpha}; \\ p, & x \in \Gamma_\alpha; \\ 0, & x \in C_{t_\alpha}, \end{cases}$$

где p выбрано так, что

$$\alpha = \int \pi(x) F_0(dx) = p F_0(\Gamma_\alpha) + F_0(\mathcal{D}_{t_\alpha}). \quad (5)$$

Покажем, что этот критерий также имеет максимальную мощность. Для любого критерия с критической вероятностью $\bar{\pi}(x)$ имеем

$$\begin{aligned} \int \bar{\pi}(x) F_1(dx) &= \int_{\mathcal{D}_{t_\alpha}} \bar{\pi}(x) f(x) F_0(dx) + \\ &+ \int_{C_{t_\alpha}} \bar{\pi}(x) f(x) F_0(dx) + \int_{\Gamma_\alpha} \bar{\pi}(x) f(x) F_0(dx) \leq \\ &\leq F_1(\mathcal{D}_{t_\alpha}) + p F_1(\Gamma_\alpha) + t_\alpha \left(\int_{C_{t_\alpha}} \bar{\pi}(x) F_0(dx) + \right. \\ &\left. + \int_{\Gamma_\alpha} \bar{\pi}(x) F_0(dx) - \int_{C_{t_\alpha}} (1 - \bar{\pi}(x)) F_0(dx) - p F_0(\Gamma_\alpha) \right), \end{aligned}$$

так как $f(x) < t_\alpha$ при $x \in \mathcal{D}_{t_\alpha}$, $-f(x) < -t_\alpha$ при $x \in C_{t_\alpha}$, $f(x) = t_\alpha$ при $x \in \Gamma_\alpha$ и

$$F_1(\Gamma_\alpha) = t_\alpha F_0(\Gamma_\alpha).$$

Тогда, согласно соотношению (5),

$$\int \bar{\pi}(x) F_1(dx) \leq \int \pi(x) F_1(dx) + t_\alpha \left(\int \bar{\pi}(x) F_0(dx) - \alpha \right)$$

и

$$\int \bar{\pi}(x) F_1(dx) \leq \int \pi(x) F_1(dx),$$

если только для критерия с критической вероятностью $\bar{\pi}(x)$ уровень значимости равен α , т. е.

$$\int \bar{\pi}(x) F_0(dx) = \alpha.$$

Построенный выше наиболее мощный критерий выбора из двух гипотез называется критерием Неймана—Пирсона.

Рассмотрим пример применения критерия Неймана—Пирсона для выбора из двух гипотез относительно некоторого события A . Пусть гипотеза H_i ($i = 0, 1$) заключается в том, что $P(A/H_i) = p_i$. Произведено n экспериментов. Обозначим через ξ — случайный вектор с компонентами $\{e_1, e_2, \dots, e_n\}$, где $e_k = 1$, если в k -м эксперименте наблюдалось событие A , и $e_k = 0$ — в противном случае. Обозначим через x_k точки из R^n , имеющие координаты, состоящие только из нулей и единиц, эти точки занумерованы произвольным образом $k = 1, \dots, 2^n$. Тогда

$$F_i(\{x_k\}) = p_i^{n_k} (1 - p_i)^{n - n_k}, \quad i = 1, \dots, 2^n,$$

где n_k — число координат x_k , равных 1. Следовательно, функции $f(x)$, для которой выполнено соотношение (2), имеет вид

$$f(x_k) = \left[\frac{p_1(1-p_0)}{p_0(1-p_1)} \right]^{n_k} \left(\frac{1-p_1}{1-p_0} \right)^n.$$

Для каждого t в область $\mathcal{D}_t = \{x : f(x) > t\}$ входят такие точки x_k , что

$$\ln f(x_k) = n_k p + n p_1 > \ln t,$$

где $p = \ln \frac{p_1}{1-p_1} - \ln \frac{p_0}{1-p_0}$, $p_1 = \ln \frac{1-p_1}{1-p_0}$, т. е. те точки, для которых

$$n_k > \frac{\ln t - n p_1}{p}.$$

Величина n_k при гипотезе H_0 имеет биномиальное распределение

$$F_0(\{x_k : n_k = m\}) = C_n^m p_0^m (1-p)^{n-m}.$$

Выберем m_α такое, что

$$\sum_{m=m_\alpha}^n C_n^m p_0^m (1-p)^{n-m} \leq \alpha \leq \sum_{m=m_\alpha+1}^n C_n^m p_0^m (1-p)^{n-m}.$$

Положим $\pi(x_n) = 1$ при $n_k > m_\alpha$,

$$\pi(x_k) = \frac{\sum_{m=m_\alpha+1}^n C_n^m p_0^m (1-p)^{n-m} - \alpha}{C_n^m p_0^{m_\alpha} (1-p)^{n-m_\alpha}}$$

при $n_k = m_\alpha$. Это и есть критерий Неймана—Пирсона в данном случае. Его мощность равна:

$$1 - \beta = \sum_{m=m_\alpha+1}^n C_n^m p_1^m (1-p)^{n-m} + \\ + \frac{p_1^{m_\alpha} (1-p)^{n-m_\alpha}}{p_0^{m_\alpha} (1-p)^{n-m_\alpha}} \sum_{m=m_\alpha+1}^n (C_n^m p_0^m (1-p)^{n-m} - \alpha).$$

§ 3. Выбор между двумя гипотезами о среднем нормальной величины

Пусть наблюдается выборка независимых величин ξ_1, \dots, ξ_n на основании которой нужно выбрать одну из гипотез H_i ($i = 0, 1$). По гипотезе H_i величины ξ_n имеют нормальное распределение с дисперсией σ^2 (не зависящей от гипотезы) и средним m_i ($i = 0, 1$). Построим критерий Неймана—Пирсона для выбора гипотезы. Будем считать, что $m_0 = 0$ (этого можно достичь, рассматривая величины $x_n - m_0$) и пусть для определенности $m_1 > 0$. Плотность вектора $\xi = (\xi_1, \dots, \xi_n)$ в R^n при гипотезе H_i будет

$$f_i(x_1, \dots, x_n) = (2\pi\sigma^2)^{n/2} \exp \left\{ - \sum_{k=1}^n \frac{(x_k - m_i)^2}{2\sigma^2} \right\}.$$

Поэтому функция $f(x)$, для которой выполнено соотношение (2), § 2, имеет вид

$$f(x_1, \dots, x_n) = \exp \left\{ \sum_{k=1}^n \left[\frac{(x_k - m_0)^2}{2\sigma^2} - \frac{(x_k - m_1)^2}{2\sigma^2} \right] \right\} = \\ = \exp \left\{ \sum_{k=1}^n \frac{x_k}{\sigma^2} m_1 - \frac{1}{2\sigma^2} n m_1^2 \right\}$$

(использован тот факт, что $m_0 = 0$). Множество

$$\{x: f(x_1, \dots, x_n) = t\} = \left\{ \sum_{k=1}^n x_k = \frac{\frac{1}{2} n m_1^2 + \sigma^2 \ln t}{m_1} \right\}$$

имеет меру F_0 , равную нулю, так как $\sum_{k=1}^n \xi_k$ при гипотезе H_0 имеет нормальное распределение со средним 0 и дисперсией $n\sigma^2$. Далее

$$D_t = \{x: f(x_1, \dots, x_n) > t\} = \left\{ x: \sum_{k=1}^n x_k > \frac{\frac{1}{2} n m_1^2 + \sigma^2 \ln t}{m_1} \right\}.$$

Таким образом, критическая область для критерия Неймана—Пирсона в данном случае такова:

$$\{x: \sum_{k=1}^n x_k > c\} \quad (1)$$

(x_k — координаты x), где c выбирается по уровню значимости α . Поскольку $\sum_{k=1}^n \xi_k$ имеет при гипотезе H_0 нормальное распределение со средним 0 и дисперсией $n\sigma^2$, то

$$F_0 \left(\left\{ x: \sum_{k=1}^n x_k > c \right\} \right) = P \left\{ \frac{1}{\sigma \sqrt{n}} \sum_{k=1}^n \xi_k > \frac{c}{\sigma \sqrt{n}} / H_0 \right\} = \\ = \frac{1}{\sqrt{2\pi}} \int_{c/\sigma\sqrt{n}}^{\infty} e^{-u^2/2} du.$$

Пусть z_α — такое значение, что

$$\alpha = \frac{1}{\sqrt{2\pi}} \int_{z_\alpha}^{\infty} e^{-u^2/2} du.$$

Тогда z_α при уровне значимости α

$$c = z_\alpha \sqrt{n\sigma}.$$

Мощность этого критерия

$$1 - \beta = F_1 \left(\left\{ x : \sum_{k=1}^n x_k > c \right\} \right) = P \left\{ \frac{1}{\sigma \sqrt{n}} \sum_{k=1}^n \xi_k > z_{\alpha/H_1} \right\} = \\ = P \left\{ \frac{1}{\sqrt{n}\sigma} \sum_{k=1}^n (\xi_k - m_1) > z_{\alpha} - \frac{m_1}{\sigma} \sqrt{n/H_1} \right\}.$$

Величина $\frac{1}{\sigma \sqrt{n}} \sum_{k=1}^n (\xi_k - m_1)$ при гипотезе H_1 имеет нормальное распределение со средним 0 и дисперсией 1. Следовательно,

$$1 - \beta = \frac{1}{\sqrt{2\pi}} \int_{z_{\alpha} - \frac{m_1 \sqrt{n}}{\sigma}}^{\infty} e^{-u^2/2} du, \\ \beta = \frac{1}{2\pi} \int_{\frac{m_1 \sqrt{n}}{\sigma} - z_{\alpha}}^{\infty} e^{-u^2/2} du. \quad (2)$$

Пусть, например, $\alpha = 0,01$. Тогда $z_{\alpha} = 2,58$. Для выполнения неравенства $\beta < 0,01$ нужно, чтобы

$$\frac{m_1 \sqrt{n}}{\sigma} - z_{\alpha} \geq z_{\alpha}, \quad \frac{m_1 \sqrt{n}}{\sigma} \geq 2z_{\alpha} = 5,16, \\ n \geq 27 \frac{\sigma^2}{m_1^2}.$$

При уменьшении вероятностей ошибок число необходимых наблюдений меняется незначительно. Так, при $\alpha = \beta = 0,001$. $z_{\alpha} = 3,3$

$$n \geq 44 \frac{\sigma^2}{m_1^2},$$

а при $n \geq 100 \frac{\sigma^2}{m_1^2}$ можно различить гипотезы с вероятностями ошибок

$$\alpha = \beta = 0,0000001.$$

Если ошибка первого рода фиксирована и равна α , то для того чтобы сделать ошибку второго рода, меньшую β , необходимо не менее

$$\frac{\sigma_1^2}{m_1^2} (z_{\alpha} + z_{\beta})^2$$

наблюдений (это вытекает из формулы (2)). Отсюда видно, что при фиксированных ошибках число наблюдений пропорционально дисперсии и обратно пропорционально квадрату разности между средними значениями (так как вследствие сделанных предположений эта разность и равна m_1).

§ 4. Байесовский подход к различению гипотез

Понятие о последовательном анализе

Во многих случаях при наличии двух возможных гипотез имеются соображения, позволяющие каждой из гипотез приписать определенную вероятность, другими словами, имеются априорные вероятности гипотез. В этом случае естественно минимизировать возможный убыток от принятия неправильных решений. Обозначим через $r_i > 0$ убыток, возникающий после принятия гипотезы H_i , если она ложна. Тогда средний убыток, который мы получим, пользуясь некоторым критерием различения гипотез с вероятностями ошибок первого и второго рода α и β соответственно, будет

$$R = P(H_0) P(H_1/H_0) r_1 + P(H_1) P(H_0/H_1) r_0, \quad (1)$$

где $P(H_0)$ и $P(H_1)$ — априорные вероятности гипотез, $P(H_1/H_0) = \alpha$, $P(H_0/H_1) = \beta$. Обозначим $P(H_0) = p_0$, тогда $P(H_1) = 1 - p_0$ и

$$R = p_0 r_1 \alpha + (1 - p_0) r_0 \beta. \quad (2)$$

Покажем, что R будет минимальным, если использовать критерий типа Неймана—Пирсона. Предположим, что использован некоторый критерий с вероятностями ошибок $\bar{\alpha}$ и $\bar{\beta}$. Тогда

$$R = p_0 r_1 \bar{\alpha} + (1 - p_0) r_0 \bar{\beta} \geq p_0 r_1 \bar{\alpha} + (1 - p_0) r_0 \bar{\bar{\beta}},$$

где $\bar{\bar{\beta}}$ — вероятность ошибки второго рода для критерия Неймана—Пирсона с уровнем значимости $\bar{\alpha}$, так как по доказанному в § 3 для любого другого критерия $\bar{\beta} \geq \bar{\bar{\beta}}$. Таким образом, выражение (2) достаточно минимизировать по критериям типа Неймана—Пирсона. Для каждого такого критерия вероятность ошибки второго рода $\bar{\beta}$ однозначно определяется вероятностью ошибки первого рода $\bar{\alpha}$: $\bar{\beta} = \bar{\beta}(\bar{\alpha})$. Из вида $\bar{\beta}(\bar{\alpha})$, найденного в § 3, вытекает, что $\bar{\beta}(\bar{\alpha})$ непрерывная функция. Но непрерывная функция от $\bar{\alpha}$

$$p_0 r_1 \bar{\alpha} + (1 - p_0) r_0 \bar{\beta}(\bar{\alpha})$$

на замкнутом отрезке $[0, 1]$ достигает минимума. Если минимум достигается в точке $\bar{\alpha}$, то, взяв критерий Неймана—Пирсона с уровнем значимости $\bar{\alpha}$, получим критерий, минимизирующий выражение (2).

На примере задачи о выборе гипотезы о нормальном распределении мы видели, как связано число необходимых наблюдений с величинами вероятностей ошибок α и β . Поэтому и минимальный убыток (2) зависит от числа наблюдений. Легко представить ситуацию, при которой наблюдения оказываются такими, что можно «досрочно» принять одну из гипотез с минимальным убытком. Это соображение легло в основу последовательного анализа, при котором вопрос о числе необходимых наблюдений решается уже в процессе произведения наблюдений, а не предварительно. Такие правила (критерии) различения гипотез называются последовательными, они впервые были предложены А. Вальдом. Опишем их более точно.

Мы будем рассматривать бесконечные последовательности (выборки) $\vec{x} = \{x_1, \dots, x_n, \dots\}$. Определим на них целочисленную, неотрицательную измеримую функцию $v(\vec{x})$, называемую числом наблюдений и обладающую следующим свойством: если $v(\vec{x}) = k$, $\vec{x} = \{x_1, \dots, x_k, x_{k+1}, \dots\}$, то $v(\vec{y}) = k$ для всякой последовательности $\vec{y} = \{y_1, \dots, y_k, \dots\}$, если только $x_1 = y_1, \dots, x_k = y_k$. Таким образом, множество $\{\vec{x} = k\}$ определяется значением первых членов последовательности \vec{x} . Пусть Γ_k — множество тех $(x_1, \dots, x_k) \in R^k$, для которых $v(\vec{x}) = k$, если первые координаты \vec{x} будут x_1, \dots, x_k .

На множестве Γ задана вероятность $\pi_k(\vec{x}) = \pi_k(x_1, \dots, x_k)$ отвергнуть гипотезу H_0 , это измеримая функция, $1 - \pi_k(x)$ — вероятность принять гипотезу H_0 . Правило выбора задано, если задана функция $v(\vec{x})$ и вероятности $\pi_k(x_1, \dots, x_k)$. Для характеристики правила выбора между гипотезами задаются следующие величины: убыток $r_i > 0$ ($i = 0, 1$), который возникает при принятии гипотезы H_i , если верна другая гипотеза, и априорные вероятности гипотез $P(H_0) = p_0$, $P(H_1) = 1 - p_0$. Заметим, что $v(\vec{x})$ может быть равно нулю, тогда оно равно нулю тождественно, и π_0 от x не зависит. Правило действует следующим образом: если $v = 0$, то наблюдения не производятся, и с вероятностью π_0 гипотеза H_0 отвергается, а с вероятностью $1 - \pi_0$ — принимается. Пусть $v \neq 0$. Производим первое наблюдение x_1 , если $x_1 \notin \Gamma_1$, второе наблюдение x_2 , если $x_2 \notin \Gamma_2$, третье наблюдение и т. д., пока при некотором k последовательность (x_1, \dots, x_k) не попадает во множество Γ_k . Тогда с вероятностью $\pi_k(x_1, \dots, x_k)$ отвергаем гипотезу H_0 и с вероятностью $1 - \pi_k(x_1, \dots, x_k)$ принимаем ее. Качество правила оценивается убытком от принятия неправильного решения и числа произведенных наблюдений. Считается, что стоимость одного наблюдения есть постоянная величина c . Если было произведено v наблюдений, то стоимость производства наблюдений vc . Применение правила привело к принятию одной из гипотез. Если Δ является тем множеством последовательностей (x_1, x_2, \dots) , на которых при-

нимается (в результате применяемого правила) гипотеза H_i ($i = 0, 1$), то $P\{(x_1, x_2, \dots) \in \Delta_i/H_i\}$ есть вероятность принятия гипотезы H_i при условии, что справедлива гипотеза H_i . Учитывая априорное распределение для гипотез, получаем следующее выражение для среднего убытка:

$$R = r_1 P(\Delta_1/H_0) p_0 + r_0 P(\Delta_0/H_1) (1 - p_0) + c (M(v/H_0) p_0 + M(v/H_1) (1 - p_0)) \quad (3)$$

(через $P(\Delta/H_i)$ обозначена вероятность того, что последовательность $\vec{x} = (x_1, x_2, \dots)$ попадет во множество Δ_i при условии, что справедлива гипотеза H_i).

Задача заключается в выборе такого правила, при котором величина R , определяемая равенством (3), принимает минимальное значение.

Рассмотрим последовательное правило выбора между двумя гипотезами о распределении случайной величины. Пусть при гипотезе H_i ($i = 0, 1$) величина имеет плотность распределения $f_i(x)$. Предположим, что обе плотности всюду положительны. Будем считать заданным числа r_i ($i = 0, 1$) — убыток при принятии ошибочной гипотезы H_i и стоимость наблюдения c . Обозначим через $R(p)$ минимальный убыток по всем последовательным правилам различения гипотез, если априорная вероятность гипотезы H_0 равна p , т. е. $R(p)$ есть точная нижняя грань величины (3), вычисляемой при всех последовательных правилах выбора (они задаются функцией $v(x)$ и вероятностями $\pi_k(x_1, \dots, x_k)$).

Функция $R(p)$ удовлетворяет некоторому соотношению, из которого она может быть однозначно определена. Заметим, что если произведено одно наблюдение и оно равно x , то условные вероятности гипотез таковы: для гипотезы H_0 — $\frac{pf_0(x)}{pf_0(x) + (1-p)f_1(x)}$, для гипотезы H_1 — $(1-p)f_1(x)/[pf_0(x) + (1-p)f_1(x)]$. Обозначим через $R_0(p)$ функцию $R_0(p) = \min[r_1 p; r_0(1-p)]$. Если $R_0(p) < c$, то наблюдений производить не стоит, так как стоимость производства наблюдения больше, чем тот убыток, который мы понесем, приняв гипотезу H_1 при $r_1 p < r_0(1-p)$ или гипотезу H_0 при $r_0(1-p) < r_1 p$ (при $r_0 p = r_1(1-p) \leq c$ можем принять любую из гипотез). Пусть произведено одно наблюдение, тогда уже затрачена величина c на наблюдение и добавочный убыток, который появляется в результате сделанного наблюдения при условии, что это наблюдение было равно x , будет

$$R\left(\frac{pf_0(x)}{pf_0(x) + (1-p)f_1(x)}\right),$$

так как условная вероятность гипотезы H_0 при условии, что было наблюдение x , есть

$$\frac{pf_0(x)}{pf_0(x) + (1-p)f_1(x)}.$$

Безусловное распределение величины x это $pf_0(x) + (1-p)f_1(x)$. Поэтому средний убыток при производстве одного наблюдения будет

$$c + \int R \left(\frac{pf_0(x)}{pf_0(x) + (1-p)f_1(x)} \right) [pf_0(x) + (1-p)f_1(x)] dx.$$

Это наблюдение имеет смысл производить, если этот убыток будет меньше, чем убыток $R_0(p)$, полученный при выборе одной из гипотез без производства наблюдений. Таким образом, убыток $R(p)$ удовлетворяет соотношению

$$R(p) = \min \left[R_0(p); c + \int R \left(\frac{pf_0(x)}{pf_0(x) + (1-p)f_1(x)} \right) \times \right. \\ \left. \times [pf_0(x) + (1-p)f_1(x)] dx \right]. \quad (4)$$

Решение этого уравнения можно получить методом последовательных приближений, полагая при $n \geq 1$

$$R_n(p) = \min [R_0(p); c + \int R_{n-1} \left(\frac{pf_0(x)}{pf_0(x) + (1-p)f_1(x)} \right) [pf_0(x) + (1-p)f_1(x)] dx]. \quad (5)$$

С помощью индукции легко убедиться, что $R_n(p) \leq R_{n-1}(p)$, поэтому существует предел $\lim_{n \rightarrow \infty} R_n(p) = R(p)$, который и будет удовлетворять уравнению (4).

Заметим теперь, что для всякой выпуклой вверх функции $V(t)$ на $(0,1)$

$$\int V \left(\frac{pf_0(x)}{pf_0(x) + (1-p)f_1(x)} \right) [pf_0(x) + (1-p)f_1(x)] dx \quad (6)$$

является выпуклой вверх функцией p . Действительно, такая функция с любой степенью точности может быть аппроксимирована функцией вида

$$V_n(t) = \alpha + \beta t + \sum_k \min [\delta_k t; \varepsilon_k (1-t)]; \quad \delta_k, \varepsilon_k > 0.$$

Достаточно проверить, что выражение (6) будет выпуклой функцией, если $V(t) = \alpha$, $V(t) = t$, $V(t) = \min [\delta_k t; \varepsilon_k (1-t)]$. В первом случае соотношение (6) равно α , во втором — p , наконец,

$$\int \min \left[\delta_k \frac{pf_0(x)}{pf_0(x) + (1-p)f_1(x)}, \varepsilon_k \frac{(1-p)f_1(x)}{pf_0(x) + (1-p)f_1(x)} \right] \times \\ \times [pf_0(x) + (1-p)f_1(x)] dx = \int \min [\delta_k pf_0(x); \varepsilon_k (1-p)f_1(x)] dx$$

есть выпуклая вверх функция p , так как для каждого x функция

$$\min [\delta_k pf_0(x); \varepsilon_k (1-p)f_1(x)]$$

выпуклая вверх функция p . Как предел выпуклых вверх функций функция $R(p)$ также выпукла вверх, а, значит, и функция

$$\hat{R}(p) = c + \int R\left(\frac{pf_0(x)}{pf_0(x) + (1-p)f_1(x)}\right) [pf_0(x) + (1-p)f_1(x)] dx$$

будет выпукла вверх. На рис. 18 приведены примерные графики функций $R_0(p)$, $\hat{R}(p)$ и $R(p)$. Пусть числа p_1 и p_2 выбраны так, что при $p \leq p_1$ функция $R(p) = R_0(p) = r_1 p$, а при $p \geq p_2$ функция $R(p) = R_0(p) = r_1(1-p)$, при $p \in (p_1, p_2)$ имеем $R(p) = \hat{R}(p)$. Тогда, если в начальный момент времени $p \leq p_1$, то принимаем гипотезу H_1 , если $p \geq p_2$, то принимаем гипотезу H_0 , если $p \in (p_1, p_2)$, необходимо произвести наблюдения. Пусть уже проведено n наблюдений: x_1, \dots, x_n (и до момента n еще не была принята ни одна из гипотез). Условная вероятность гипотезы H_0 и H_1 при условии, что получены указанные наблюдения, задается формулами

$$P(H_0/x_1, \dots, x_n) = \frac{pf_0(x_1) \dots f_0(x_n)}{pf_0(x_1) \dots f_0(x_n) + (1-p)f_1(x_1) \dots f_1(x_n)},$$

$$P(H_1/x_1, \dots, x_n) = \frac{(1-p)f_1(x_1) \dots f_1(x_n)}{pf_0(x_1) \dots f_0(x_n) + (1-p)f_1(x_1) \dots f_1(x_n)}.$$

Если $P(H_0/x_1, \dots, x_n) = p_1$, то принимаем гипотезу H_1 , если $P(H_0/x_1, \dots, x_n) = p_2$, то принимаем гипотезу H_0 , а если $P(H_0/x_1, \dots, x_n) \in (p_1, p_2)$, то производим опять наблюдение. Таким образом, правило сводится к следующему. Рассмотрим последовательность величин

$$\theta_0 = p, \quad \theta_k = \frac{pf_0(x_1) \dots f_0(x_k)}{pf_0(x_1) \dots f_0(x_k) + (1-p)f_1(x_1) \dots f_1(x_k)}.$$

Обозначим через v такой номер, при котором θ_v впервые не принадлежит (p_1, p_2) . Тогда в момент v наблюдение прекращается и при $\theta_v \leq p_1$ принимается гипотеза H_1 , а при $\theta_v \geq p_2$ — гипотеза H_0 . Это же правило удобно описать еще и так: пусть

$$\xi_k = \ln \frac{f_1(x_k)}{f_0(x_k)}, \quad \zeta_0 = 0, \quad \zeta_k = \xi_1 + \dots + \xi_k, \quad k \geq 1,$$

тогда v — первый момент, когда ζ_k не принадлежит интервалу

$$\left(\ln \left(\frac{p}{p_2} \cdot \frac{1-p_2}{1-p} \right); \ln \left(\frac{p}{p_1} \cdot \frac{1-p_1}{1-p} \right) \right),$$

если при этом $\zeta_v \leq \ln \left(\frac{p}{p_2} \cdot \frac{1-p_2}{1-p} \right)$, то принимаем гипотезу H_0 ,

если $\zeta_v \geq \ln \left(\frac{p}{p_1} \cdot \frac{1-p_1}{1-p} \right)$, то принимаем гипотезу H_1 . Чтобы убедиться в этом, достаточно заметить, что

$$\theta_k = \frac{p}{p + (1-p)e^{\zeta_k}}.$$

СПИСОК ЛИТЕРАТУРЫ

Теория вероятностей

- Гнеденко Б. В. Курс теории вероятностей. М., Наука, 1967.
Боровков А. А. Теория вероятностей. М., Наука, 1976.
Скороход А. В. Елементи теорії ймовірностей та випадкових процесів. К., Вища школа, 1975.
Феллер В. Введение в теорию вероятностей и ее приложения. Т. 1. М., Иностранная литература, 1964; т. 2, Мир, 1967.
Лозэ М. Теория вероятностей. М., Иностранная литература, 1962.
Дороговцев А. Я., Сильвестров Д. С., Скороход А. В., Яценко М. И. Теорія ймовірностей (збірник задач). К., Вища школа, 1977.

Теория случайных процессов

- Розанов Ю. А. Случайные процессы (краткий курс). М., Наука, 1974.
Гихман И. И., Скороход А. В. Введение в теорию случайных процессов. М., Наука, 1977.
Карлин С. Основы теории случайных процессов. М., Мир, 1971.

Математическая статистика

- Крамер Г. Математические методы статистики. М., Мир, 1975.
Ван дер Варден Б. Л. Математическая статистика. М., Иностранная литература, 1960.
Шметтерер Л. Введение в математическую статистику. М., Наука, 1976.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Аксиомы теории вероятностей 36, 37

Алгебра множеств 25

Альтернатива 390

Блуждание возвратное 150

— невозвратное 150

— нерешетчатое 149

— по целочисленной решетке 149

— простейшее симметричное 152

— случайное 3, 37, 151

Вектор случайный 192

Величина гауссовская 225

— априорная 46

— случайная 49, 51, 59, 75, 96

Вероятность апостериорная 46

— геометрическая 35

— доверительная 369

— перехода 181, 164

— события 3, 22, 23, 25, 37

— стационарная 187

— условная 43, 204

Гамма-доверительный интервал 369

Гамма-распределение 116

Границы доверительные 369

Дельта-функция 313

Дисперсия 90

— равномерного распределения 92

— условная 205

α -интервал 192

Задача Бюффона 35

— о выборе двух гипотез 391

— о сравнении двух средних 376

Закон больших чисел 136, 146

— редких событий 104

Индикатор случайного события 53

Интеграл Стильеса 85

— стохастический 287

— от процесса 282

Интерпретация события 5

Класс непериодический 173

— множеств 26, 27

Ковариация 278, 280, 310

Количество информации 340

Комбинаторика 10

Коэффициенты биномиальные 14

— передачи системы 295

— полиномиальные 16

— регрессии 235

Критерий 387, 391

— Неймана — Пирсона 394

— нерандомизированный 392

Математическое ожидание 73, 74, 78,

82, 83, 85, 205, 206

Матрица вероятностей перехода

161

— информационная 346

— корреляционная 232

Медиана распределения 332

Метод максимального правдоподобия 355

— Монте-Карло 141

— наименьших квадратов 387

— статистических испытаний 141

Мера вероятности 29, 32, 37

— стохастическая 300

Многочлен Бернштейна 89, 139

Множество непрерывности меры

214

— критическое 392

— слабо компактное 216

— упорядоченное 14

— цилиндрическое 196

Модель Бозе — Эйнштейна 21

— вероятностная 23, 37

— Максвелла-Больцмана 20

Момент выборочный 337

— m -го порядка 89

— показательного распределения 93

— центральный 90, 92

— эмпирический 337

Мощность критерия 390

— случайного процесса 289, 290

Неравенство Гельфера 94

— Иенсена 95

— Колмогорова 131

— Ляпунова 95

— Чебышева 82, 85, 90

Объединение событий 5

- Ожидание математическое 73, 74, 78, 82, 83, 85, 205, 206
- Ординарность потока событий 106
- Отклонение среднеквадратическое 208
- Отсутствие последействия 62
- Оценки выборочные 335
 - максимального правдоподобия 355
 - параметра 325
 - эффективные 338, 339, 348
- Пересечение множества 6
 - событий 6
- Перестановки 14
 - с повторениями 16
- Период класса 173
 - случайного блуждания 153
- P -квантиль 333
- Плотность априорная 351
 - распределения 67, 195, 205, 292, 293, 354
- Поверхность регрессии 210
- Положение блуждания 149
- Последовательность оценок 327
 - распределений 211
 - стандартная 253
 - стационарная 293
- Правило умножения 11
- Предел последовательности событий 8, 18
- Представление спектральное 321
- Преобразование Фурье 219, 239, 319
- Прообраз 193
- Пространство вероятностное 37
 - фазовое 263
 - элементарных событий 3
- Процесс восстановления 155
 - гауссовский 311
 - локально регулярный 182
 - марковский 160, 181
 - Пуассона 109, 268
 - с независимыми приращениями 109, 281
 - случайный 109, 180, 263, 277, 279, 288, 318, 320
 - стандартный 313
 - стационарный 287
 - стохастически непрерывный 264
 - эквивалентный 270
- Разложение спектральное 300, 321
- Размещения из n по k 15
- Разность множеств 7
- Распределение безгранично-делимое 264
 - биномиальное 61
 - геометрическое 62
 - гипергеометрическое 63
- d -мерное 225
- Коши 71
- Лапласа 118
- лог-нормальное 357
- Максвелла 117
- мультиномиальное 224
- Распределение нормальное 69, 230
 - показательное 70
 - Пуассона 63, 258
 - равномерное 69
 - решетчатое 129
 - случайного вектора 193
 - случайной величины 60
 - Стьюдента 201
 - треугольное 114
 - Фишера-Снедекора 202
 - χ^2 117
 - Эрланга 116
- Реакция системы 294
- Ряд вариационный 330
- Свойства вероятности 38, 39, 40, 41, 42
- Свертка плотностей 111
 - последовательностей 126
 - функций распределения 111, 199
- α -алгебра 26, 29
- Системы марковские 161
 - однородные во времени 295
 - возвратные 167
 - достоверные 5
- События дизъюнктивные 7
 - наблюдаемые 4
 - несовместимые 7
 - противоположные 7
 - случайные 3, 37, 47
 - стохастически устойчивые 22
 - элементарные 3, 4
- Состояние возвратное 167
 - кулево 177
 - положительное 177
- Сочетания из n по k 12
 - с повторениями 18
- Статистика 325
 - достаточная 352
 - математическая 22
- Спектр колебаний процесса 269
- Сумма множеств 6
 - событий 6
- Сходимость по вероятности 132
 - с вероятностью 1, 141
- Теорема Бернулли 138
 - Бернштейна 139
 - Бореля 147
 - Бореля — Кантелли 143
 - Бохнера 249
 - Каратеодори 32
 - Колмогорова 330

- Лебега 68, 69
- Леви 227
- Линдеберга 256
- Ляпунова 257
- Муавра — Лапласа 258
- Планшереля 247
- Хелли 217
- Хинчина 138, 291
- Чебышева 136
- Штольца 137
- Топология 308
- Точка Лебега 242
 - роста функции 68
- Уравнение Колмогорова — Чепмена 165
- Уравнения дифференциальные 305
 - максимального правдоподобия 356
 - нормальные 384
- Условие физической осуществимости 296
- Фильтр высоких частот 304
 - допустимый 298
 - полосовой 304
 - тождественный 304
- Формула Байеса 46, 204, 351
 - биномиальная 18
 - полной вероятности 45
 - Пуассона 140
 - умножения 44
 - Фурье 319

- Функционал характеристический 311
- Функция борелевская 58
 - импульсная 294, 295
 - корреляционная 2, 80, 288, 293
 - Кульбака — Ляйблера 360
 - на входе системы 294
 - на выходе системы 294
 - непрерывная слева 194
 - обобщенная 312
 - опорная 95
 - потеря 325
- Функция правдоподобия 338, 339
 - производящая 124
 - процесса 284
 - распределения 51, 193, 195
 - случайная 109
 - спектральная 290, 292, 293, 321
 - характеристическая 118, 219, 248, 264
 - \mathcal{F} -измеряемая 192
 - эмпирическая 330
- Характеристика частот 295, 298
- Цели Маркова 161, 166
- Частота события 21
- Шаг блуждания 149
 - распределения 129
 - максимальный 130
- Энергия, переносимая случайным процессом 289
- Эксперимент стохастический 3
- Эффективность 342

ОГЛАВЛЕНИЕ

Глава I. Случайные события

§ 1. Стохастический эксперимент, случайные события	3
§ 2. Основные понятия комбинаторики	10
§ 3. Классическое определение вероятности	21
§ 4. Алгебры и σ -алгебры множеств; теорема о продолжении меры	25
§ 5. Построение вероятностных моделей для экспериментов с бесконечным числом исходов; геометрические вероятности	33
§ 6. Аксиомы теории вероятностей	36
§ 7. Условные вероятности	42
§ 8. Независимые случайные события	47

Глава II. Случайные величины и функции распределения

§ 1. Случайные величины	49
§ 2. Распределение случайных величин	59
§ 3. Математическое ожидание случайной величины	72
§ 4. Математическое ожидание функции от случайной величины. Моменты. Дисперсия	85
§ 5. Независимые случайные величины	96
§ 6. Предельные теоремы для биномиального распределения	100
§ 7. Процесс Пуассона	105
§ 8. Суммы независимых случайных величин	109
§ 9. Характеристические и производящие функции	118

Глава III. Последовательности случайных величин

§ 1. Неравенство Колмогорова	131
§ 2. Сходимость по вероятности	132
§ 3. Закон больших чисел	136
§ 4. Сходимость с вероятностью 1	141
§ 5. Усиленный закон больших чисел	146
§ 6. Случайные блуждания	149
§ 7. Процесс восстановления	155

Глава IV. Цепи Маркова

§ 1. Определение цепи Маркова. Простейшие свойства	161
§ 2. Однородные цепи Маркова	166
§ 3. Эргодическая теорема для однородных цепей Маркова	175

Глава V. Марковские процессы со счетным множеством состояний

§ 1. Определение марковского процесса с непрерывным временем	180
§ 2. Уравнения Колмогорова	183
§ 3. Применение теории марковских процессов к задачам массового обслуживания	187

Глава VI. Случайные векторы

§ 1. Распределение случайного вектора	192
§ 2. Независимые случайные векторы	197
§ 3. Условные распределения	203

§ 4. Сходимость последовательностей распределений	210
§ 5. Характеристические функции	218
§ 6. Многомерное нормальное распределение	229
§ 7. Элементы гармонического анализа	236
§ 8. Центральная предельная теорема	251
Глава VII. Процессы с независимыми приращениями	
§ 1. Определения, простейшие свойства	262
§ 2. Обобщенный процесс Пуассона	268
§ 3. Процесс броуновского движения	271
Глава VIII. Корреляционная теория случайных процессов	
§ 1. Элементы анализа в L_2	277
§ 2. Слабо стационарные процессы	287
§ 3. Слабо стационарные последовательности	292
§ 4. Линейные преобразования случайных процессов	293
§ 5. Обобщенные случайные процессы	307
Глава IX. Оценивание параметров распределений	
§ 1. Выборочный метод в статистике	324
§ 2. Достаточные статистики	351
§ 3. Метод максимального правдоподобия	355
§ 4. Доверительные интервалы	369
§ 5. Нормальная линейная регрессия	378
Глава X. Проверка статистических гипотез	
§ 1. Проверка простой гипотезы. Критерий χ^2	387
§ 2. Задача о выборе из двух гипотез	390
§ 3. Выбор между двумя гипотезами о среднем нормальной величины	395
§ 4. Байесовский подход к различению гипотез. Понятие о последовательном анализе	398
Список литературы	403

*Носиф Ильич Гихман
Анатолий Владимирович Скороход
Михаил Носифович Ядренко*

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Редактор Г. Ф. Трофимчук
Обложка художника Д. Д. Грибова
Художественный редактор Е. В. Чурий
Технический редактор Л. И. Швец
Корректор Л. Д. Мякоход

Информ. бланк № 2201

Сдано в набор 30.11.78. Подп. в печать 21.02.79. Формат 60×90¹/₁₆. Бумага типогр. № 1. Лит. гарн. Вис. Печать. 25,5 печ. л. 24,74 уч.-изд. л. Тираж 15000 экз. Изд. № 3461. Зак. 8-519 Цена 1 р. 20 к.

Главное издательство издательского объединения «Вища школа», 252054, Киев-54, ул. Гоголевская, 7

Книжная фабрика им. М. В. Фрунзе Республиканского производственного объединения «Полиграфкинг» Госкомиздата УССР, Харьков, Донец-Захаржевская, 6/8.

3.5C
F517

41.208

